UNIVERSITÉ DU QUÉBEC À MONTRÉAL

SIMULATION STRATEGIES FOR OPTIMAL DETECTION OF REGIONAL CLIMATE MODEL RESPONSE TO PARAMETER MODIFICATIONS

DISSERTATION

PRESENTED

AS PARTIAL REQUIREMENT OF THE DOCTORATE

OF EARTH AND ATMOSPHERIC SCIENCES

BY

LEO SEPAROVIC

JANUARY 2012

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

APPROCHES STRATÉGIQUES POUR DÉTECTER DE FAÇON OPTIMALE LA RÉPONSE SIMULÉE PAR UN MODÈLE RÉGIONAL DU CLIMAT SOUMIS À DES MODIFICATIONS DE SES PARAMÈTRES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN SCIENCES DE LA TERRE ET DE L'ATMOSPHÈRE

PAR

LEO SEPAROVIC

JANVIER 2012

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de thèse, Dr. Ramón de Elía, pour sa patience et ses conseils précieux. Ce travail fut d'autant plus agréable grâce à ses nombreux encouragements et le soutient qu'il m'a apporté tout au long du projet.

Mes remerciements vont également à mon co-directeur de thèse, Prof. René Laprise, qui a partagé avec moi ses brillantes intuitions. J'ai aussi grandement apprécié sa gentillesse et sa constante disponibilité.

Je remercie aussi Mme Katja Winger pour son aide technique lors de la production de simulations avec le Modèle Régional Canadien du Climat.

Pour ses encouragements et son assistance morale qui m'ont permis de rédiger cette thèse dans des conditions idéales, je remercie Aleksandra.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AGCM	Atmospheric General Circulation Model
AMIP	Atmospheric Model Intercomparison Project
CGCM	Coupled Global Climate Model
CORDEX	COordinated Regional climate Downscaling Experiment
CRCM	Canadian Regional Climate Model
ERA-40	European Centre for Medium-range Weather Forecasts 40-year re-analysis data
GHG	Greenhouse Gases
LBC	Lateral Boundary Conditions
MYNA	Multi-Year North America
MYSN	Multi-Year Spectral Nudging
NARCCAP	North American Regional Climate Change Assessment Program
PIRCS	Project to Intercompare Regional Climate Simulations
PRUDENCE	Prediction of Regional scenarios and Uncertainties for Defining EuropeaN Climate change risks and Effects
RCM	Regional Climate Model
RPN	Recherche en Prévision Numérique
SN	Spectral Nudging
SYNA	Single-Year North America
SYSN	Single-Year Spectral Nudging
SYDS	Single-Year Domain Size
UKCP	United Kingdom Climate Projections

RÉSUMÉ

Cette thèse vise à rechercher les configurations expérimentales optimales pour étudier la réponse des Modèles Régionaux du Climat à aire limitée (MRC) face à des perturbations de leurs paramètres. Le travail est présenté en deux parties.

La première partie aborde le cas d'une comparaison entre les simulations provenant d'un MRC, où un événement météorologique ou saisonnier est mis à l'échelle dynamiquement à partir de données observées (réanalyses). Cette situation implique l'utilisation de périodes d'intégration relativement courtes. Par conséquent, la réponse obtenue dans les moyennes temporelles des simulations par rapport à des modifications aux paramètres a tendance à être noyée dans le bruit quasi-aléatoire provenant de la dynamique chaotique du MRC. La possibilité d'augmenter le rapport signal-bruit par l'application du pilotage spectral ou par une réduction de la taille du domaine est étudiée. L'approche adoptée consiste à analyser la sensibilité des moyennes saisonnières du MRC Canadien (MRCC) face à des perturbations sur deux paramètres variés un à un. Le premier contrôle la convection profonde tandis que le second régit la condensation stratiforme. Les résultats montrent que l'ampleur du bruit diminue avec la réduction de la taille du domaine ainsi que par l'application du pilotage spectral. Toutefois, la réduction de la taille du domaine produit aussi des altérations statistiquement significatives de certains signaux, ce qui favorise l'utilisation de pilotage spectral.

La deuxième partie de cette thèse aborde le cas d'une comparaison entre deux simulations d'un MRC en termes du climat simulé. À cet effet, un cadre théorique est développé pour le calcul des statistiques de premier et second ordre sur la différence entre les simulations. Les statistiques de la différence sont décomposées en une composante déterministe et reproductible contrainte par les conditions aux frontières et une composante de bruit provenant de la dynamique interne du MRC. Certaines questions reliées à l'estimation de la différence des moyennes temporelles entre les simulations sont développées en détail. Par exemple, un partage optimal des ressources informatiques entre la taille d'un ensemble et la longueur de la période d'intégration, ou encore l'impact de la taille du domaine et du pilotage spectral sur l'estimation de la réponse du modèle. Une application de ces considérations théoriques est illustrée à partir de la réponse des simulations du MRCC dont un paramétre lié à la convection profonde a été perturbé.

Mots-clés: modèle régional de climat, perturbation des paramètres, ensemble, composante reproductible, pilotage spectral, taille du domaine, différence de moyennes.

ABSTRACT

This thesis aims at finding experimental setups and simulation configurations that facilitate studies devoted to quantifying limited-area Regional Climate Model (RCM) response to parameter modification. The work presented in this thesis is divided in two parts.

The first part addresses the case when the researcher attempts to compare the RCM simulations in terms of downscaling a particular weather event or season from the objective analyses. This case implies the use of short integration times, and the response of the time-averaged variables to RCM modification tends to be blurred by the quasi-random noise originating in the RCM chaotic dynamics. The possibility of enhancing the signal-to-noise ratio by the application of spectral nudging or domain-size reduction is studied. The approach adopted to study these issues consists in the analysis of the sensitivity of the Canadian RCM (CRCM) seasonal averages to perturbations of two parameters controlling deep convection and stratiform condensation, perturbed one at a time. Results show that the noise magnitude is decreased both by reduction of domain size and the spectral nudging. However, the reduction of domain size produced statistically significant alterations of some sensitivity signals, which fosters the use of spectral nudging.

The second part of this thesis addresses the case of comparing two RCM simulations in terms of the simulated climate. For this purpose a theoretical framework is developed for calculation of the first- and second-moment statistics of the difference between RCM simulations, such as time means and variances. The statistics of the difference are decomposed into their deterministic, reproducible components, forced by the boundary conditions, and the quasi-random noise originating from RCM internal dynamics. Some issues related to the estimation of the difference of means between control and modified RCM simulations are elaborated in detail, such as the optimal allocation of computational resources between ensemble size and integration time, as well as the impact of spectral nudging and domain size on mean model response estimation. An application of the present theoretical considerations is illustrated by considering the response of CRCM decadal simulations to a perturbation of a deep-convection parameter.

Keywords: Regional climate model, parameter perturbation, ensemble, reproducible component, spectral nudging, domain size, difference of means.

INTRODUCTION

Climate modeling at regional scale

Climate simulations and projections of the future climate at local scale are obtained by coupling separate complex modeling systems. Models of future emission scenarios of greenhouse gases (GHG) and aerosols (Nakicenovic et al. 2000) provide the radiative forcing component in Coupled Global Climate Models (CGCMs) that are the most comprehensive tools for climate studies. CGCMs are comprised of the Atmospheric Global Circulation Models (AGCMs) coupled with the ocean, sea ice and land surface (e.g., Collins et al. 2001). Because of their high complexity and the need to perform very long simulation to stabilize the deep ocean, CGCM simulations are very demanding in computational resources and are performed at relatively coarse horizontal resolution. Development of the adaptation and mitigation strategies require information on spatial scales smaller than those provided by CGCMs.

Regional Climate Models (RCMs) are employed to dynamically downscale CGCM simulations to scales of a few tens of kilometers, using high-resolution representation of the atmospheric dynamics and physics, as well as forcing at the interface between the atmosphere and the other components of the climate system, only over a specified area of the globe. Different approaches to RCM dynamical downscaling have been developed (see, e.g., Laprise 2008, for a review). So far, the most popular strategy has been the one-way nesting based on high-resolution limited-area RCMs. These models are specific because they require prescribed information at the lateral boundaries of their computational domain, with no feedback from the nested model to the driving fields (e.g., Giorgi and Mearns 1999, Rummukainen 2010) .

Climate projections are inherently uncertain because of the fundamental properties of both the climate system and modeling tools. Since the work of Lorenz (1963), it has been acknowledged that due to non-linear chaotic nature of the atmospheric flow, model solutions are unstable with respect to very small perturbations, so that slight

differences in initial states evolve in time into large differences. This phenomenon has been referred to as internal variability. Uncertainties associated with constructing and applying the models are manifold. They can be grouped into: (a) structural, (b) parameterization and (c) parameter uncertainties (Murphy et al. 2007). The structural uncertainties originate in the choices related to model structure and configuration (e.g., choices related to grid, resolution, truncation, numerical integration scheme, the set of processes included). Fine-scale physical processes cannot be resolved explicitly in models due to their high complexity and insufficient resolution, and their bulk effect upon the response of the resolved scales is parameterized. The parameterization uncertainties are related to the fundamental assumptions in the representation of physical processes. Finally, not all parameters that figure in parameterizations can be inferred from first principles of physics or physical experimentation. They may rely on mixture of theoretical understanding and empirical fitting, or may even have no counterpart in the real climate system. The uncertainty, thus, also originates in the presence of a large number of adjustable parameters.

Representing uncertainty in climate modeling

In studies with CGCMs, the uncertainty related to model structure and configuration has been quantified, as initially proposed in Räisänen and Palmer (2001), using multi-model ensembles obtained by combining operational models developed at different research centers. The rationale behind this method lies in the fact that the models grouped in such a way are validated against a large number of observables (e.g., Giorgi and Mearns 2002, 2003; Tebaldi et al. 2005). The operational models are carefully studied and they employ different discretization techniques and include different parameterizations of subgrid processes and other components from a large pool of alternatives (Murphy et al. 2007). However, the method is sometimes criticized for not allowing for an adequate sampling of all possible choices in constructing models, as the models forming multi-model ensembles are assembled on an opportunity basis. The exchange of knowledge between different modeling centers may result in common deficiencies among members of a multi-model ensemble (Tebaldi and Knutti 2007). In addition, it is not clear how to define a space of all possible model configurations of which the members of a multi-model ensemble would be a sample (Murphy et al. 2007).

An alternative approach, based on exploring the uncertain values of adjustable physics parameters in CGCMs has been also developed, mainly as a part of the *climateprediction.net* project (e.g., Murphy et al. 2004, Frame et al. 2005, Piani et al. 2005, Stainforth et al. 2005, Barnett et al. 2006, Forest et al. 2006, Knutti et al. 2006, Sanderson et al. 2008, Ackerley et al. 2009). In this approach, the selection of parameters to perturb and estimation of their range of variation are typically conducted by consulting experts that participated in model development. At least in principle, the perturbed-physics ensembles allow for a systematic sampling of related parameter uncertainties and hence afford a greater control of the experimental design than the multi-model ensembles. The perturbed-physics ensemble approach however does not sample uncertainty arising from all choices that must be made among existing options in order to construct a model, such as those related to model structure, configuration, physical processes included or schemes to parameterize these processes. Despite that, it has been shown that the perturbed-physics CGCM ensembles typically exhibit a spread of members as large as the spread in ensembles of different models collected on the basis of opportunity (Murphy et al., 2004, 2007). The perturbed-physics approach has been sometimes criticized because not all the members of a perturbed-physics ensemble can be expected to offer credible climate simulations, as is the case with multi-model ensembles; some of the members may exhibit substantial departures from the observed climate (Stainforth et al. 2005).

The RCMs are specific because they inherit uncertainty of the driving fields through the lateral boundary conditions (LBC) (e.g., de Elía et al. 2008). There are also specific choices that must be made in configuring RCM simulations, such as the size and position of the computational domain and the nesting technique, which are also sources of uncertainty. Last but not least, similarly to CGCMs, uncertainty in RCMs also originates in the choices related to their structure, parameterization of sub-grid processes and the adjustable parameters. Multi-RCM ensembles have been increasingly employed to quantify the structural uncertainty in RCMs; there has been a large number of internationally coordinated multi-model experiments designed to address uncertainty specifically in RCM integrations (e.g., PIRCS, Takle et al. 1999; PRUDENCE, Christensen et al. 2007; UKCP, Murphy et al. 2007; NARCCAP, Mearns et al. 2009); CORDEX, (Giorgi et al. 2009). On the other hand, the perturbed-physics ensemble approach to quantifying uncertainty in regional climate began to emerge only recently (e.g., Burke et al. 2010). There were, however, a few early attempts to test RCM sensi-

tivity to structural differences, also including perturbations of physics parameters. For example, Yang and Arritt (2002) and Arritt et al. (2004) showed that the differences between RCMs developed at different centers were much larger than the RCM response to perturbations of parameters. This favored multi-model ensembles as a more efficient technique of sampling differences between RCMs. However, the aforementioned studies examined the RCM response to perturbations of only two parameters in a specific part of model parameterization, the convective cloud parameterization, which constitutes only a small subspace of a high-dimensional space of adjustable parameters in RCMs.

The main reason why the RCM parameter space remains largely unexplored lies in the fact that, due to a large number of parameters, a thorough sampling of even a small fraction of the parameter space imposes the need for very large computing resources that are out of reach for the majority of research centers. Quantifying parameter uncertainty in regional modeling necessitates the use of distributed computing network, such as that adopted by the *climateprediction.net*. Due to the large number of variable parameters, it is impossible to run the model for their every plausible combination and to examine all possible skill scores. In addition, a high skill score does not guarantee that it is achieved for the right reasons, since a high skill can be also a consequence of the cancelation of errors. Furthermore cancelation of errors in simulating the present climate does not imply cancelation of errors in simulating a different climate; there may exist several separated regions of the parameter space in which the model can exhibit a relatively good performance in simulating the observed climate, but yield a considerable difference in the projected climate change. Hence, a systematic sampling of model response to parameter perturbations may help to improve the parameters settings, as well as to inform about the uncertainty in the projected climate.

Because of the need to produce more realistic simulations using physically comprehensive models, the increase in computing speed is usually absorbed by adding complexity and increasing resolution rather than by an extensive testing of plausible alternatives in applying and constructing the models. The increasing complexity of models hampers their extensive testing, because of an immense number of combinations of adjustable parameters. Another but not less important difficulty is that the estimation of model response to a modification requires that the signal of model response be distinguished from the noise of internal variability of the modeling system. The most appropriate way to increase the statistical significance of the signal is to devote a sufficiently large

computing time to the climate simulations in order to provide a sufficiently large sample of model states from which robust estimates of the signal can be computed. However, it is typically not possible to generate sufficiently large samples due to the lack of computational resources. The alternative is to employ RCM simulation set-ups that use less computation time. In RCMs this can be done in different ways, which is considered next.

Optimal simulation setup for RCM parameter modification

The following equation illustrates the classical choices that the numerical modeler must confront when conducting an RCM study:

$$\frac{N \times A \times H \times K \times C}{\Delta x^2 \times \Delta h} \propto L \times S, \tag{1}$$

where N represents the number of experiments, A the area covered by the grid, H the total vertical height, K the integration period, C an index of model complexity, Δh vertical resolution, and Δx horizontal resolution. On the right-hand side, S represents computer speed, and L the length of project. Computer speed typically does not evolve much during the lifetime of a project and the length of a project is normally non-negotiable and, at best, limited to a few years. The experimental setup is bounded by these two constraints. In order to increase the number of experiments N, a trade-off has to be made among the other factors on the left hand side of Eq. (1).

The options of reducing model complexity or resolution have been ruled out in this study. Reducing model complexity in order to increase N would considerably change the model structure and make the results very difficult to extrapolate to the operational RCM runs. Changing the horizontal and vertical resolution may also involve issues such as, for example, that the parameter settings in parameterizations of sub-grid processes may be tuned for the specific resolution used in the operational RCM runs. This work rather focuses on the possibility of reducing domain size A and the integration period K, in order to increase the project capacity with respect to the number of modifications N that can be explored. It is important to note that an optimal simulation setup in terms of A and K may not exist. Each setup may require that specific compromises be made with respect to the potential of extrapolating the results to the typical operational RCM continental-scale domains and multi-decadal runs. In the following we discuss the compromises in that respect.

An obvious compromise related to the reduction of the computational area is a loss of generality of the results, since they are confined to a small region of interest. Model response to parameter modification may be very difficult to extrapolate to the geographical regions outside the tested domain, due to different properties of land surface or large-scale atmospheric dynamics. There are also other potential issues with the small RCM domains. For example, Leduc and Laprise (2009) provided evidence that the use of small domains can yield RCM simulations deficient in the fine-scale features when compared to the RCM runs conducted in large, continental-scale domains. This problem raises the concern that model sensitivity to parameter modification can be altered in different ways in small domains, due to the proximity of the lateral boundaries. On the other hand, when large domains are used, the RCM interior large-scale flow is prone to intermittent inconsistencies with the driving large-scale flow, which may result in large spurious gradients at the perimeter of the lateral boundaries (e.g., von Storch et al. 2000). The use of small domains can reduce such inconsistencies.

Reduction of the integration period involves issues related to model internal variability. Because of the chaotic nature of the models and the atmosphere, a negligibly small difference in the initial state eventually yields a large difference in the trajectory in the model phase space. Experiments with climate models point to the fact that the spread between identical integrations that depart from slightly different initial states decreases upon averaging in time and eventually become small when the averaging time is large (e.g., Giorgi and Francisco 2000, de Elía et al. 2008). The internal variability is triggered by any modification in the model, regardless of its magnitude and origin (Giorgi and Bi 2000). This implies that any parameter modification would always produce a change in the evolution of the model states. If the modification is infinitesimally small, the only response will be the noise of internal variability which is filtered by averaging in time. So, in the case of an infinitesimally small modification, such as the perturbation in the initial conditions, the differences in the climate statistics due to the modification are expected to be vanishingly small if the integration time is sufficiently large. If the modification is large however, then it may produce a change in the climate statistics, regardless of the integration time. Hence, in order to distinguish the signal of model response to a modification from the internal variability noise in the climate statistics, larger samples of model states are needed. It follows from these considerations that if the integration period is too short it may be not possible to decide whether

the difference obtained between the time statistics of a control and parameter-modified RCM run is a consequence of the parameter modification or only an apparent difference due to residuals of internal variability. In statistical terminology, we may say that a reduction of the integration period reduces the statistical significance of the signal.

In principle, performing ensemble integrations instead of making inferences from a single integration can solve the problem of the insufficient sample size due to a reduced integration period. Ensemble integrations can be performed, for example, by perturbing initial conditions. The advantage of RCMs is that, unlike in CGCMs, the spread of RCM ensemble members is a controllable parameter. Reduction of the size of the computational domain may considerably reduce the inter-member variance (e.g., Alexandru et al. 2007). An alternative approach to reducing the ensemble spread is the internal nudging of the large-scale components of RCM variables (usually referred to as spectral nudging; von Storch et al. 2000; Biner et al. 2000). This method has been proposed for reducing the intermittent inconsistencies between the internal RCM flow and the driving fields, and has been also employed to reduce the differences between ensemble members (e.g., Weisse and Feser 2003).

Another important issue related to the reduction of the integration period is the representativeness of results with respect to the simulated temporal variability of climate. If the estimates of model response obtained by ensemble integrations over a relatively short period of time are statistically significant, they are so only for this period of time. That is, the results may be statistically significant for given LBC but little representative for some other choice of the LBC. For example, the model response could be different if a different period was chosen from the reanalyses to drive the model. This raises issues such as how much generality of results is lost due to a too short integration period and how the loss of generality depends on the ensemble size and simulation configuration (domain size and nudging). Studying variability in time of the RCM response to parameter modification is not a trivial task, since the time variation of RCM variables is partly enforced by the time variation of the driving fields and partly by the RCM internal dynamics. It is thus of interest to study in a systematic manner how a decrease of internal variability noise by spectral nudging or domain size reduction can affect the temporal variability of the model response to modification.

Objectives and approach

This thesis tries to find optimal RCM simulation set-ups that use less computation time than operational runs while still returning representative results. The overall objective of this work is to analyze RCM response to parameter perturbations with simulation set-ups using different integration periods, model domains and large-scale constraints (spectral nudging) and eventually select the set-up that is the most appropriate for conducting a large number of RCM sensitivity tests in a computationally efficient way. The study is conducted in two parts and is presented in form of scientific papers.

In the first part (Chapter 1) the focus is placed on short RCM ensemble integrations, conducted over a single year. The goal is to find simulation configurations that maximize the statistical significance of the response of seasonal-average RCM variables but are not detrimental to the response. The approach followed consists in the study of the sensitivity of RCM-simulated seasonal averages to perturbations of two parameters controlling deep convection and stratiform condensation, perturbed one at a time. These parameters were selected because they directly influence the formation of precipitation; this variable represents, along with surface air temperatures, the most relevant variable for climate studies. The both parameters also control the atmospheric liquid and solid water content and hence indirectly affect the surface radiative budget and surface temperatures. In addition, the plausible perturbations of these two parameters produce a relatively large signal in precipitation and surface air temperature, which allows for extracting the signal of model response to perturbations among the noise of internal variability at a reasonable computing cost.

The sensitivity to perturbations of the deep-convection and stratiform-condensation parameters is analyzed within three simulation configurations: (a) in a large, continental-scale domain, (b) the same domain as in (a) but with spectral nudging, and (c) in a small domain. In order to sample the contribution of the internal variability noise to model response to parameter modification, for every setting of these parameters, multiple integrations with perturbed initial conditions are also performed. Signal-to-noise ratio is then studied as a function of the three simulation configurations, which allows for quantifying the reduction in the computational cost by spectral nudging and domain size reduction. In addition, we quantify the alterations of the model response to parameter modification (signal) triggered by these noise-reducing methods. Optimally, a noise

reducing method should efficiently reduce the noise but, at the same time, minimally interfere with the signal.

In the first part of this thesis only the RCM integrations performed over a single year are studied and thus the variability of model response to parameter modification in time is neglected. The purpose of the second part of this thesis is to examine how much the results obtained in the single-year study depend on the choice of the simulated year. How the spectral nudging and domain size reduction affect this dependence? It turns out that this is a rather complex issue that requires a specific theoretical framework. Chapter 2 first presents a general theoretical framework for studying temporal variability of RCM response to modification. The theoretical considerations are then employed to analyze perturbed-parameter RCM simulations conducted over a 10-year integration period in different RCM simulations configurations.

CHAPTER I

IMPACT OF SPECTRAL NUDGING AND DOMAIN SIZE IN STUDIES OF RCM RESPONSE TO PARAMETER MODIFICATION

This chapter is presented in the format of a scientific article. It was submitted to the journal *Climate Dynamics* and is now published. The manuscript is entirely based on my work, with the co-authors involved in interpretation of the results and text editing. The detailed reference is:

Separovic, L., de Elía R. and Laprise, R., 2011: Impact of spectral nudging and domain size in studies of RCM response to parameter modification. Climate Dynamics, doi: 10.1007/s00382-011-1072-7

12

Impact of Spectral Nudging and Domain Size in Studies of RCM Response to Parameter Modification

Leo Separovic, Ramón de Elía and René Laprise

Leo Separovic - René Laprise
Centre ESCER (Étude et Simulation du Climat à l Échelle Régionale),
Département des Sciences de la Terre et de l'Atmosphère
Université du Québec à Montréal (UQAM)
B.P. 8888, Succ. Centre-ville
Montréal (Québec) Canada H3C 3P8

Ramón de Elía
Centre ESCER (Étude et Simulation du Climat à l Échelle Régionale),
Consortium Ouranos, 550 Sherbrooke West,
19^{th} floor, West Tower, Montreal, QC H3A 1B9, Canada

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.

Key words: Regional climate models, parameter perturbations, internal variability, spectral nudging, domain size

1.1 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 users judgment (e.g., Laprise et al. 2008). For example, Leduc and Laprise (2009) 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 SN remains an open issue. Alexandru et al. (2009) raised concern that the application of the SN could suppress the 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 SN 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. 2008a; Nikiema and Laprise 2011) and is scale selective (Separovic et al. 2008).

Reduction of domain size or the application of SN 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 2002). 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 Elía 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 RCMs 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 Section 2. The analysis of model sensitivity to modification of parameters within different simulation configurations is carried out in Section 3. Summary and conclusions are provided in Section 4.

1.2 Experimental design

1.2.1 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 (Taylor et al. 2000).

1.2.2 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 1.1. Two experts that participated in CRCM5 development judged the perturbations as being moderate to strong with respect to their range of variation, given the hori-

zontal resolution (B. Dugas and P. Vaillancourt, both from Environment Canada RPN, personal communication).

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 (see Appendix 1.5 for more details); the last column in Table 1.1 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. 1.1 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. 1.1), model versions and number of ensemble members per every model version (Table 1.1); the only difference is that the SN was used. The nudging was only applied to the horizontal wind components, with the truncation at non-dimensional wavenumber 4 (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. 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. 1.1, including the 10-point sponge zone.

1.3 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 1.3.1 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 1.3.2 presents the spatial distribution of the internal variability noise in the three configurations. Sections 1.3.3 to 1.3.5 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 1.3.6 examines the computational cost associated with different simulation configurations in terms of the minimum ensemble size necessary to achieve significant estimates.

1.3.1 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 1.1), 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. 1.2 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. 1.1. The rmsd displayed with coloured marks in Fig. 1.2 are a result of the model response to the parameter perturbations. Internal variability is displayed with black marks in Fig. 1.2. 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 1.2 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 than the reduction of domain size (SYDS); this being true for this case and different configurations of both SN and domain size could yield different results. The plots in Fig. 1.2 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. 1.2a), 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. 1.2b) 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. 1.2 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.

1.3.2 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

$$\sigma = \left[\frac{1}{M_x - 1} \sum_{m=1}^{M_x} \left(\overline{x}_m - \langle \overline{x} \rangle \right)^2 + \frac{1}{M_y - 1} \sum_{m=1}^{M_y} \left(\overline{y}_m - \langle \overline{y} \rangle \right)^2 \right], \tag{1.1}$$

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 1.1. 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. 1.1) is displayed in Fig. 1.3 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. 1.3a) 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 SN and reduced domain size (Fig. 1.3b, c) help to reduce noise level for precipitation in winter to fairly low values. The patterns of the 2 m-temperature in winter (Fig. 1.3d–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. 1.3g) it locally attains 2.5 mm/day over the southern and eastern coastal regions of the continent. SN (Fig. 1.3h) is not very efficient in reducing noise. The domain size reduction (Fig. 1.3i) reduces noise but locally it is still up to 0.6 mm/day. As of 2 m-temperature in summer (Fig. 1.3j–l), noise levels are barely higher than 1°C. SN 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. 1.3g. 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.

1.3.3 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 1.1). 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 2002). 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

$$t = \frac{\langle \overline{y} \rangle - \langle \overline{x} \rangle}{\sqrt{(1/M_x + 1/M_y) S_w^2}}, \tag{1.2}$$

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

$$S_w^2 = \frac{\sum_{m=1}^{M_y} \left(\overline{y}_m - \langle\overline{y}\rangle\right)^2 + \sum_{m=1}^{M_x} \left(\overline{x}_m - \langle\overline{x}\rangle\right)^2}{M_x + M_y - 2} \tag{1.3}$$

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 1.1. The ensemble size of the control version is doubled in order to increase the signal to noise ratio and to estimate well the ensemble variance for at least one model version. Appendix 1.5 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. 1.2). 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 M10 and M00, is presented in Fig. 1.4a–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. 1.4d–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. 1.4a) the strongest and also highly significant (Fig. 1.4d) signal is aligned with the entire Pacific Coast. It reaches locally up to ±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. 1.4a) is a mainly negative signal over the southeast portion of the domain, significant at 95% level. Also note that in several regions in Fig. 1.4d 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. 1.4a. This illustrates the fact that statistical significance does not imply a physically relevant signal.

When the SN is applied (Fig. 1.4b, 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. 1.4a, 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. 1.4a, b appear over the eastern portion of the domain. When the SYDS setup is considered (Fig. 1.4c, 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. 1.4b, c and the signal in the SYNA set shown in Fig. 1.4a. 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 1.6 for details). The resulting fields of statistical significance of the signals differences are shown in Fig. 1.4g (for SYSN-SYNA) and Fig. 1.4h (SYDS-SYNA). The differences between the signals are not shown since they can be inferred from subtracting values from Fig. 1.4b, c from Fig. 1.4a. In Fig. 1.4g 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. 1.4h) is similar to that between the SYSN and SYNA. From Fig. 1.4a, 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. 1.5a–c). It can be seen that the statistical significance levels for the signal are as high as 99% in all configurations (Fig. 1.5d–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°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. 1.5g) displays high local significance levels over the southern and north-central parts of the continent. In the small SYDS domain (Fig. 1.5c) the estimated magnitude of the signal is also somewhat reduced. Figure 1.5h 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.

1.3.4 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. 1.3g–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 2m temperatures (see Fig. 1.6). A widespread cooling is notable over most of the continent, with magnitudes up to 2.2°C (Fig. 1.6a–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. 1.6d–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 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. 1.5a, 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 $-1°C$ occupies more than a half of the continent in the SYNA, extending form the Pacific to the Atlantic coast unlike in SYSN), this difference is not significant when tested in Fig. 1.6g. 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. 1.6c, 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 1.6h 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 models response to the perturbation of the threshold parameter for the onset of deep convection (P01 in Table 1.1). 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. 1.2). Hence, for this perturbation, we focus on the summer months.

1.3.5 Signal P01 in summer

Figure 1.7 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. 1.7b). The perturbation P01 also exhibits a strong effect on summer precipitation in the small SYDS domain (Fig. 1.7c); 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 East Coast, the signal is spatially variable, despite being highly statistically significant (Fig. 1.7e) 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 1.7g, 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. 1.8. 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 in Fig. 1.6a, b. Figure 1.8g shows that the 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. 1.8h, the difference of means is mostly non-significant, and therefore, no evidence is found against the reduction of domain size in the signal P01.

1.3.6 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. 1.9, for precipitation (a) and temperature (b). The plots in Fig. 1.9 represent the square root spatially averaged (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. 1.2) is in fact the signal-to-noise ratio. The numerator in (Eq. 1.2) 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 1.9 displays the 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. 1.1), exclusive of the 10-point sponge zone. Only land points are accounted for in the computation of the rms. In Fig. 1.9 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 simulation configuration. The black step-like line in Fig. 1.9 represents the standard deviation of the difference of ensemble means. It is computed as the rms of the denominator in (Eq. 1.2).

It can be seen in Fig. 1.9 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. 1.9a), 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. 1.9 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

$$t_{\text{eff}} = \frac{\text{RMS}\left[\langle \overline{y} \rangle - \langle \overline{x} \rangle\right]}{\sqrt{\frac{2}{M}}\ \text{RMS}\left[S_w\right]}, \tag{1.4}$$

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. 1.3). For a significance level of 95% the t-statistic (Eq. 1.2) 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_{eff} in (Eq. 1.4) and solving for M gives the proposed rule of thumb for the minimum ensemble size as follows:

$$M_{\text{min}} = 2 \times 1.96^2 \times \left\{ \frac{\text{RMS}\left[S_w\right]}{\text{RMS}\left[\langle \overline{y} \rangle - \langle \overline{x} \rangle\right]} \right\}^2. \tag{1.5}$$

Note that due to the properties of the Students 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_{min}. Due to some vagueness of the concept we rather intend to use M_{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. 1.10 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 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. 1.2b 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. 1.10a) 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. 1.7a). 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. 1.10b) 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. 1.6a, 1.8a); 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. 1.5) for the response to P10 (P01), respectively. Also the noise level in the

SYNA setup in spring is slightly higher than in autumn.

1.4 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 test 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 SN 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 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 SN 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.

Acknowledgements: 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.

1.5 Appendix: 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 $\{y\}_{k=1}^{K}$. We assume that x will be sampled M_x times, while each member of the family $\{y\}_{k=1}^{K}$ 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

$$z_k = \frac{\langle x \rangle - \langle y_k \rangle}{S\sqrt{\frac{1}{M_x} + \frac{1}{M_y}}}, \tag{1.6}$$

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

$$b = \frac{M_x}{M_y}, \tag{1.7}$$

Equation (1.6) can be expressed as

$$z_k = r_k \left(\frac{b}{b+1} \right)^{1/2}, \tag{1.8}$$

where

$$r_k \equiv \frac{\langle x \rangle - \langle y_k \rangle}{S/\sqrt{M_y}}, \tag{1.9}$$

is the signal-to-noise ratio expressed independently of the control sample size M_x. Fig. 1.11 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 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.

1.6 Appendix: 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

$$\delta = \langle \overline{y} \rangle - \langle \overline{x} \rangle, \tag{1.10}$$

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 1.1. It is assumed that $\overline{x}$ and $\overline{y}$ are independent and have equal variances. Under these assumptions it can be shown that the variance of δ can be estimated as

$$S^2 = \frac{1/M_x + 1/M_y}{f}\left[\sum_{m=1}^{M_y}\left(\overline{y}_m - \langle\overline{y}\rangle\right)^2 + \sum_{m=1}^{M_x}\left(\overline{x}_m - \langle\overline{x}\rangle\right)^2\right], \tag{1.11}$$

where f denotes the degrees of freedom and is given as

$$f = M_x + M_y - 2. \tag{1.12}$$

Now we turn our attention to the difference between the signal δ_2 obtained in the SYSN (or SYDS) and δ_2 in SYNA setup. The assumption that the variances of δ_1 and δ_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. 1.3. Thus, the test for the difference of means of variables with unequal variances (von Storch and Zwiers 2002) has to be used to test the difference of the signals obtained in two configurations. The test statistic can be written as follows:

$$t' = \frac{\delta_2 - \delta_1}{\sqrt{S_1^2 + S_2^2}}, \tag{1.13}$$

where S_i^2 is the estimator of the variance of δ_i in the setup $i = 1, 2$, as defined in (Eq. 1.11). Because of unequal variances, the variable t' does not have t-distribution under the null hypothesis that δ_1 and δ_2 have equal means. To solve this problem we employ the Welchs 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:

$$F = \frac{\left(S_1^2 + S_2^2\right)^2}{S_1^4/f_1 + S_2^4/f_2}. \tag{1.14}$$

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:

$$\Pr\left(|t| < t_0\right) = 1 - I_{\frac{F}{t_0^2+F}}\left(\frac{F}{2}, \frac{1}{2}\right), \tag{1.15}$$

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

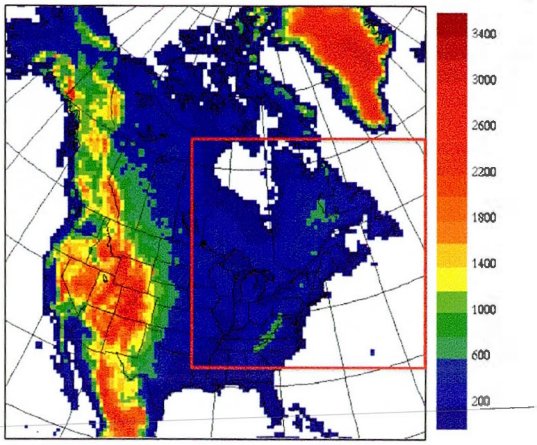

Figure 1.1 Topography of the two CRCM5 computational domains, including the lateral boundary relaxation zone. The large domain is used in the SYNA and SYSN experiments and the smaller domain in the SYDS experiment.

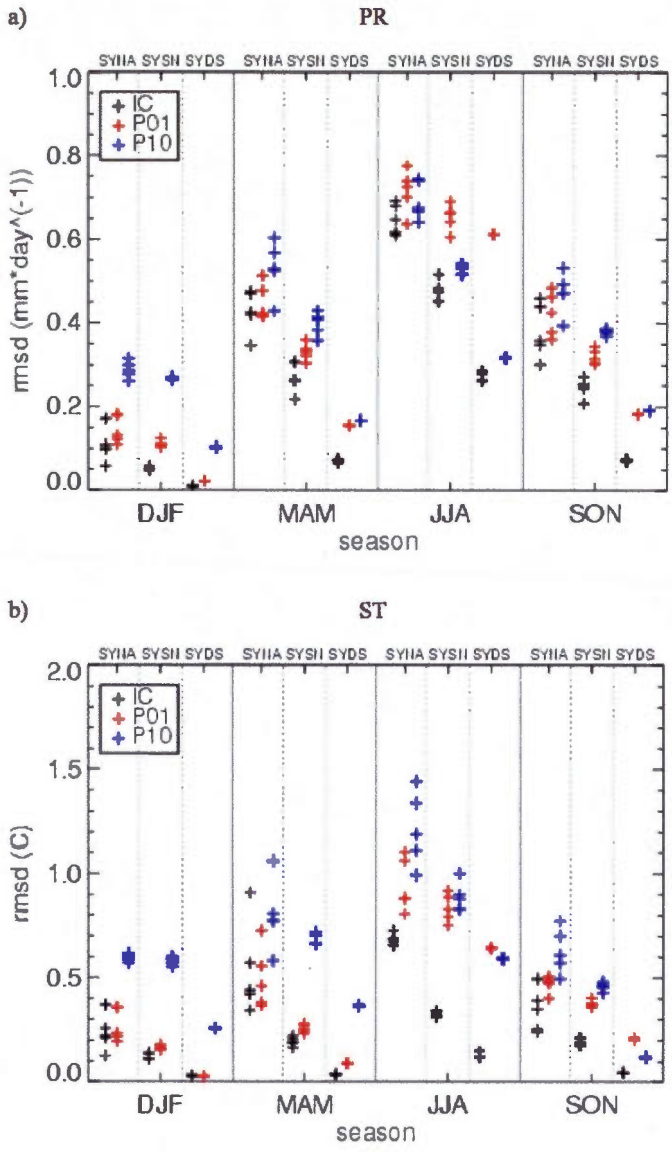

Figure 1.2 The RMS difference between CRCM5 individual simulations for seasonal-average (a) precipitation and (b) 2 m-temperature as a function of the experimental setup and season. The *black* marks display the rmsd in seasonal averages among the ensemble members of the model M00 (Table 1.1); they are 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.

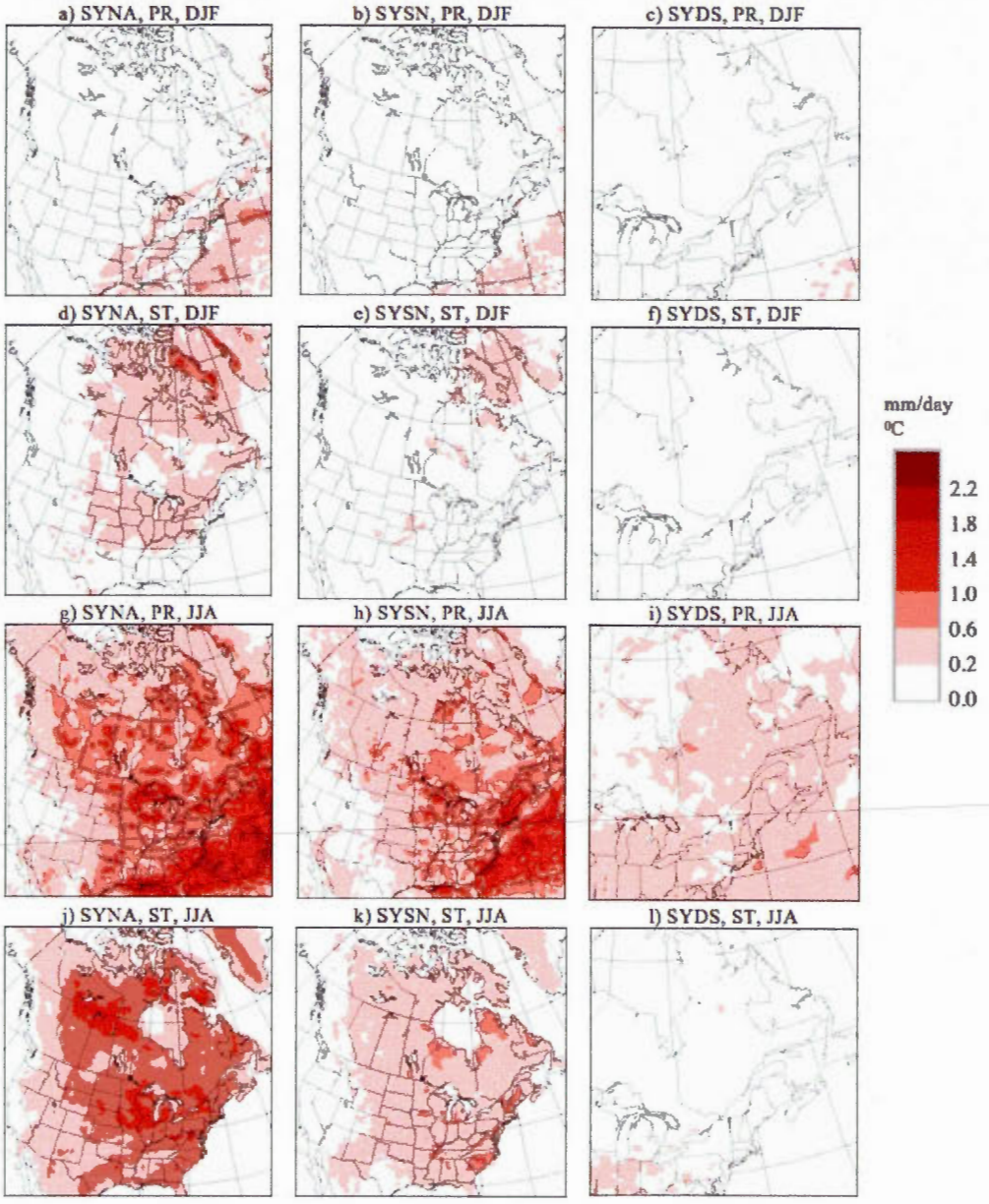

Figure 1.3 Sample standard deviation (Eq. 1.1) of the sensitivity of the CRCM5 seasonal average to parameter perturbation P10 (Table 1.1). The sensitivities are measured as the differences between members of the perturbed-parameter model M10 and members of the control model M00 ensembles, as a function 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.

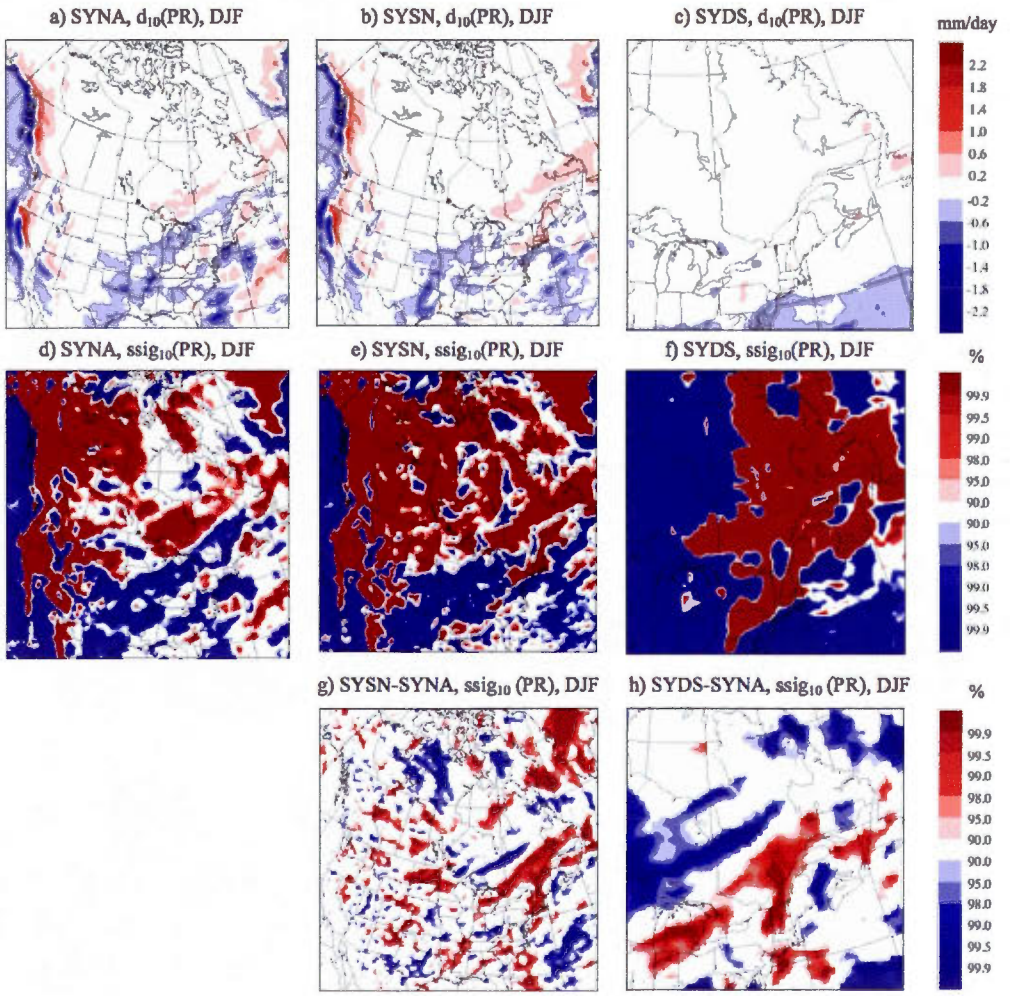

Figure 1.4 Difference of the ensemble mean winter-average (DJF) precipitation (signal) due to the perturbation P10 (Table 1.1) in (a) SYNA, (b) SYSN and (c) SYDS experiments and statistical significance of the responses (d, e, and f, respectively); statistical significance of the difference of the signals (g) between SYSN and SYNA and h between SYDS and SYNA experiments.

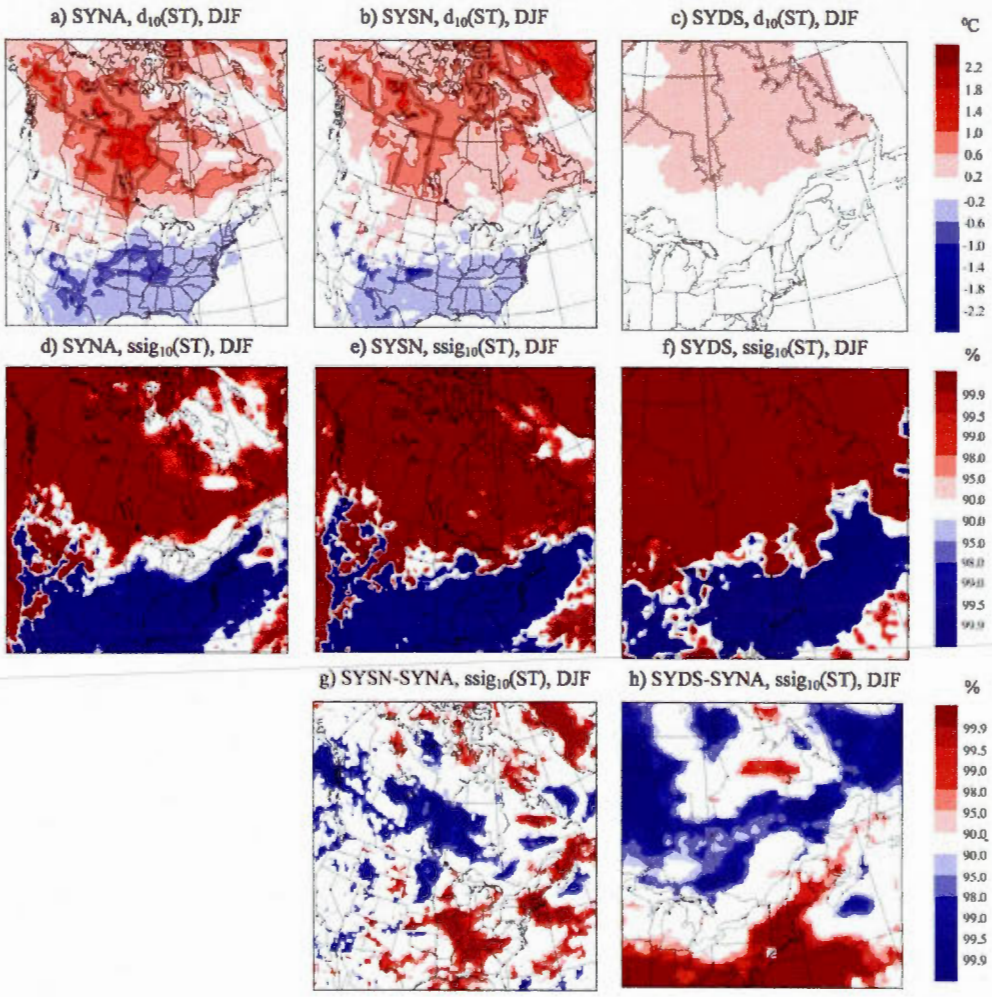

Figure 1.5 Same as in Fig. 4 but for winter 2 m-temperature (DJF).

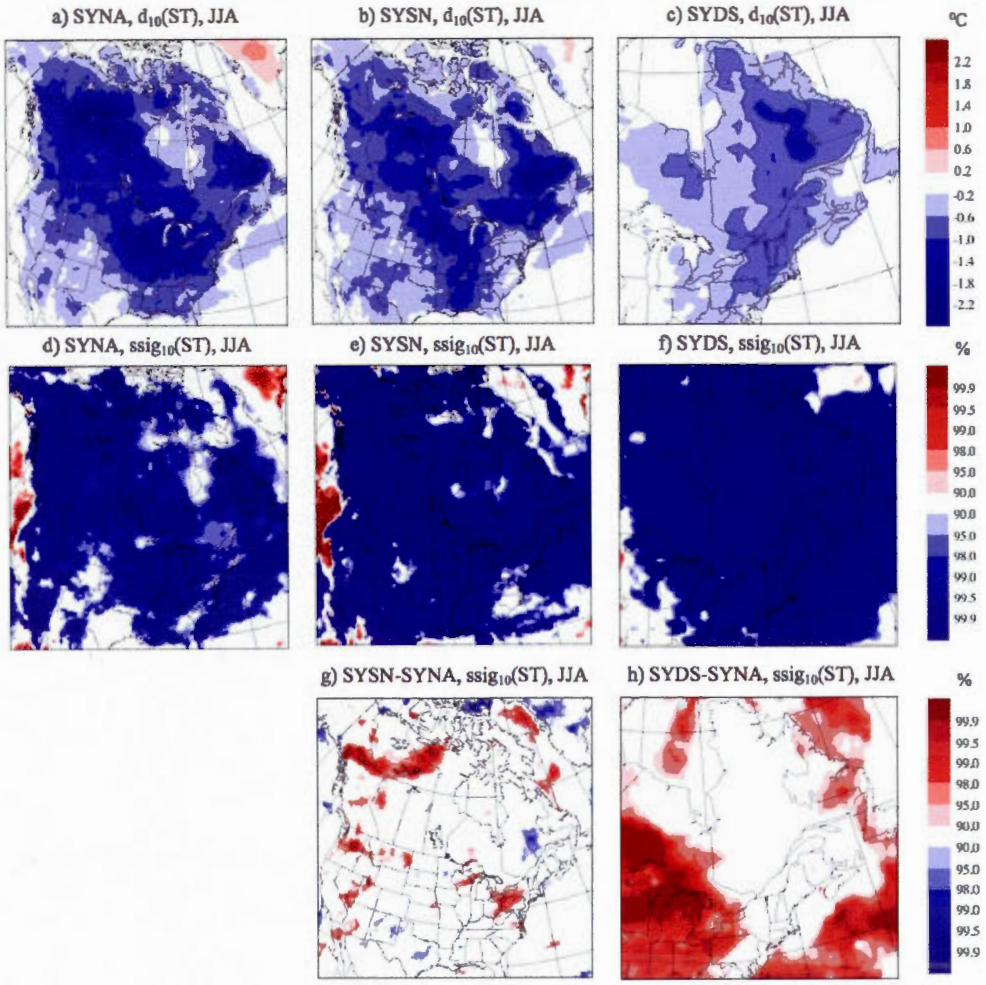

Figure 1.6 Same as in Fig. 4 but for summer 2 m-temperature (JJA).

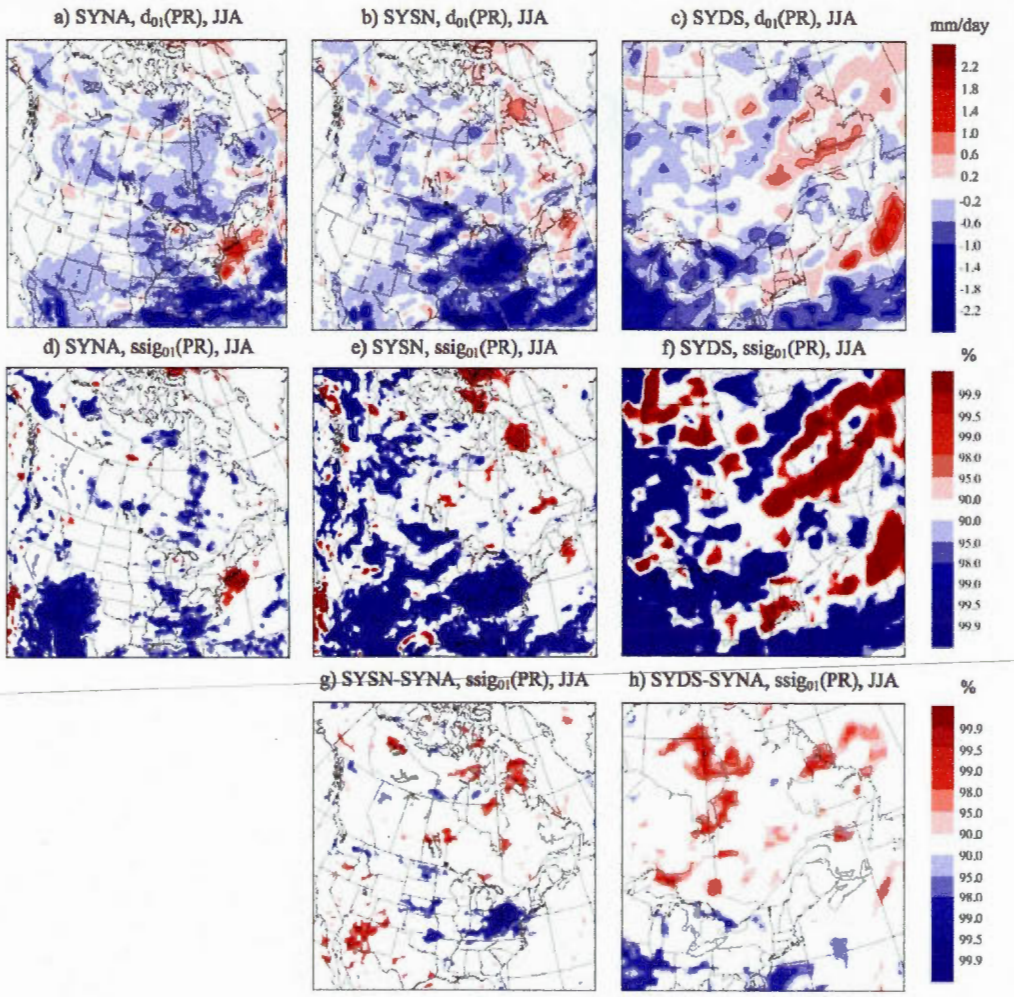

Figure 1.7 Same as in Fig. 4 but for the signals induced by the perturbation P01 for summer precipitation (JJA).

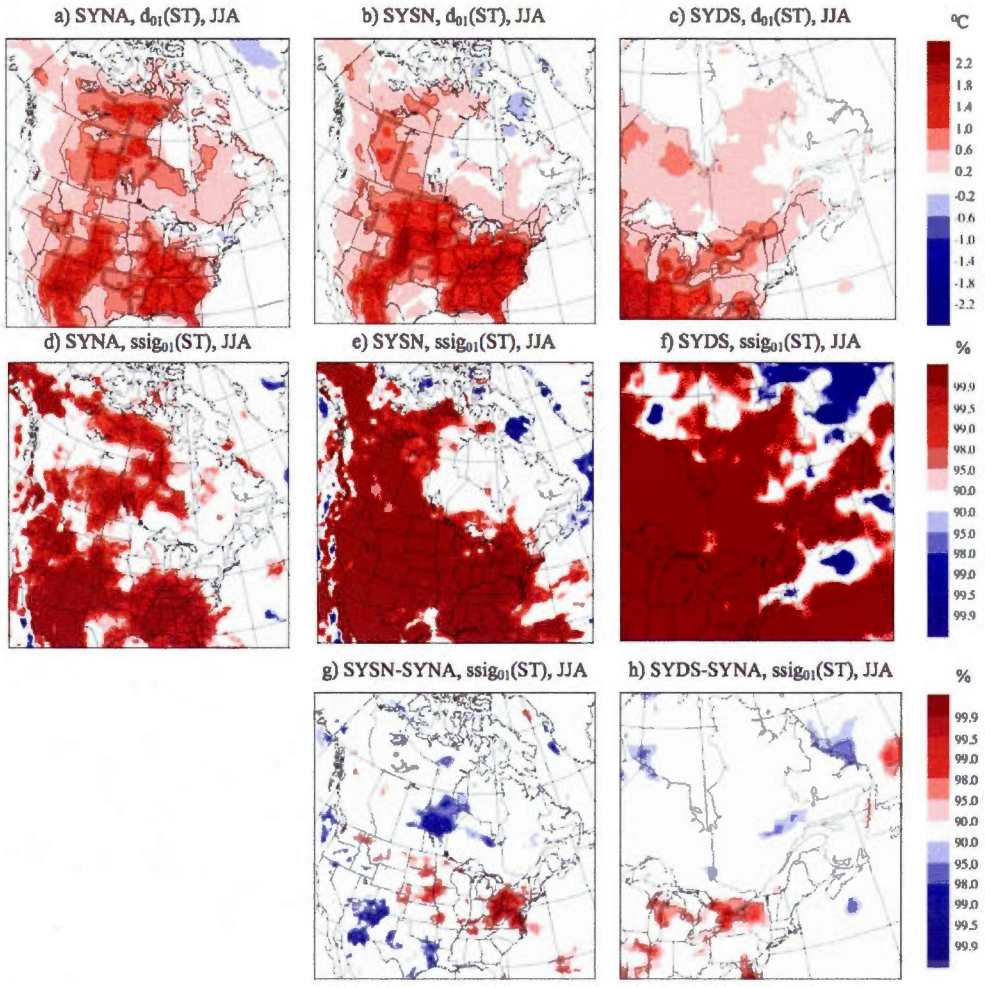

Figure 1.8 Same as in Fig. 4 but for the signals induced by the perturbation P01 for summer 2 m-temperature (JJA).

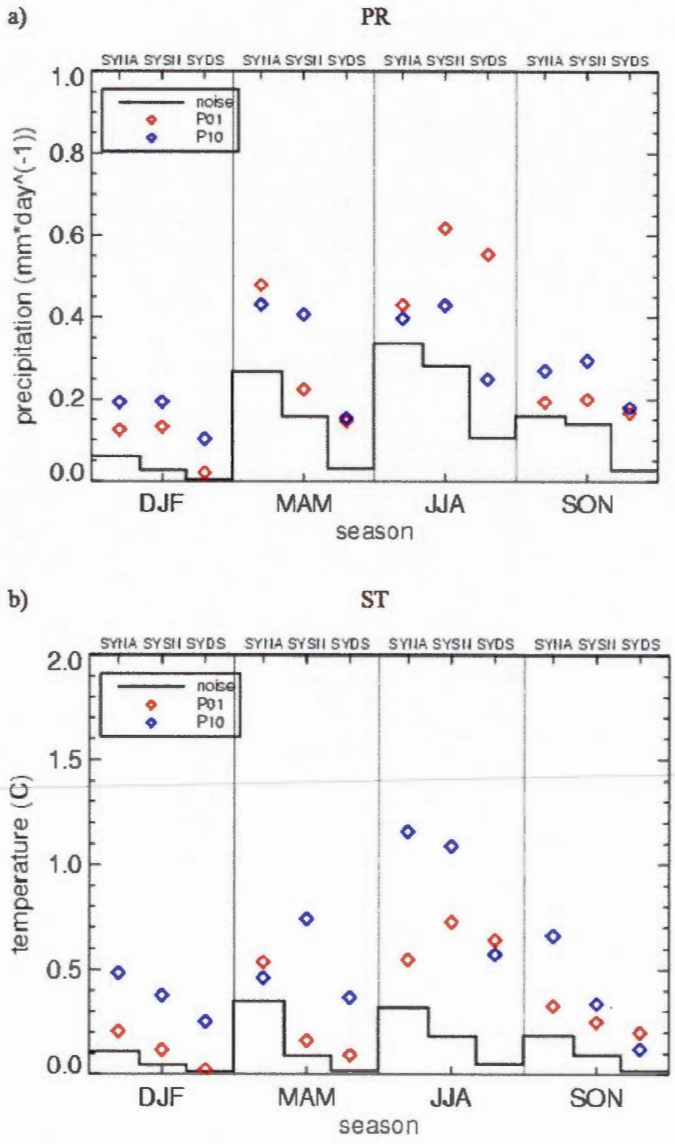

Figure 1.9 The rms signal (*diamonds*) and noise (*step-like line*) as a function of experimental setup and season for seasonal-average (a) precipitation and (b) 2 m-temperature. Signal is estimated as the rms difference of ensemble means of the perturbed-parameter (*red*) M01 and (*blue*) M10 model and control model M00 (Table 1.1). Noise is measured with the standard deviation of the difference of ensemble means.

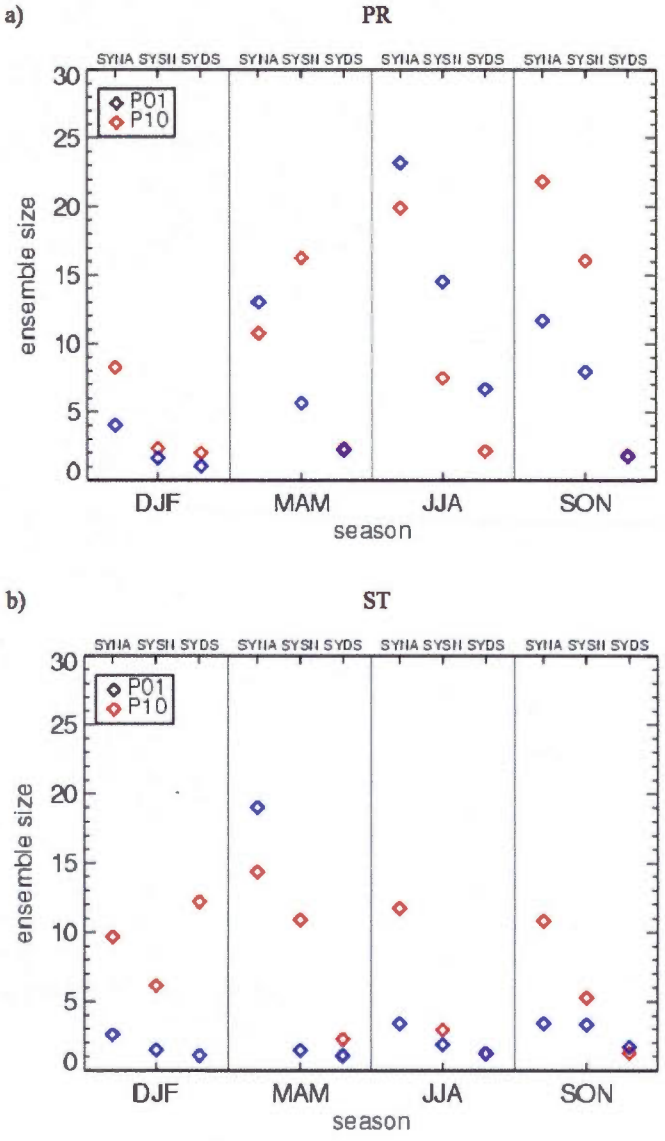

Figure 1.10 Minimal number of ensemble members needed to achieve signicant estimates at 95% level for the signals induced by the perturbations (*red*) P01 and (*blue*) P10, as a function of season and experimental setup, as derived from the rule of thumb in Eq. (1.5); seasonal-average (a) precipitation and (b) 2 m-temperature.

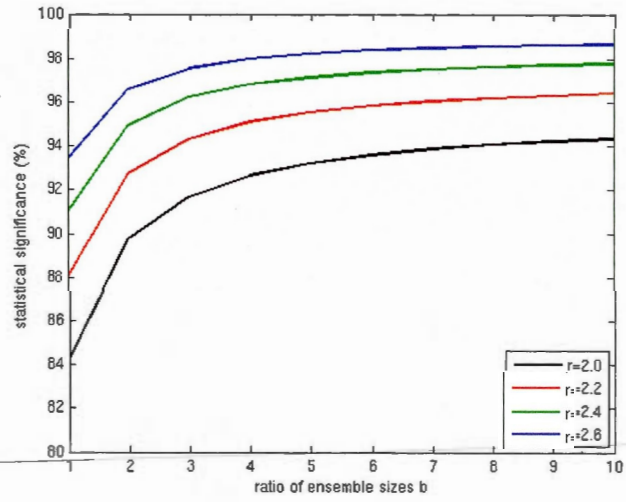

Figure 1.11 Statistical signicance derived from the two-sided test of the difference of means of two samples of unequal sample sizes; the size of the first sample is kept constant and the size of the second is increased by the factor b (abscissa), as in Eq. (1.7). Plots are drawn for selected signal-to-noise ratios (Eq. 1.9).

Table 1.1 Parameters' settings used in different model versions. 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.

Model Version	P10 (hours)	P01 (m/s)	N^o of ensemble simulations
M00	2.8	$3.4E-2$	10
M01	2.8	$6.0E-2$	5
M10	10	$3.4E-2$	5

CHAPTER II

A THEORETICAL FRAMEWORK FOR ANALYSIS OF TEMPORAL VARIABILITY OF RCM RESPONSE TO PARAMETER MODIFICATION

This chapter is presented in the format of a scientific article. It will be submitted to the journal *Climate Dynamics*. This manuscript is entirely based on my work, with the co-authors involved in interpretation of the results and text edits.

A theoretical Framework for Analysis of Temporal Variability of RCM Response to Modification

Leo Separovic, Ramón de Elía and René Laprise

Leo Separovic - René Laprise
Centre ESCER (Étude et Simulation du Climat à l Échelle Régionale),
Département des Sciences de la Terre et de l'Atmosphère
Université du Québec à Montréal (UQAM)
B.P. 8888, Succ. Centre-ville
Montréal (Québec) Canada H3C 3P8

Ramón de Elía
Centre ESCER (Étude et Simulation du Climat à l Échelle Régionale),
Consortium Ouranos, 550 Sherbrooke West,
19^{th} floor, West Tower, Montreal, QC H3A 1B9, Canada

To be submitted to Climate Dynamics (as two papers).

Abstract

A theoretical framework for calculation of the first- and second-order statistics of the difference between Regional Climate Model (RCM) simulations, suitable for quantifying RCM response to code modifications and changes in the simulation setup, is discussed. The approach followed consists of decomposing the RCM response to modification into (1) the response of the deterministic, reproducible components of RCM variables that are forced by the boundary conditions, and (2) the quasi-random noise originating from RCM internal dynamics. Some issues related to the estimation of the difference of means between control and modified RCM simulations are elaborated in detail. This includes the issue of optimal allocation of computational resources between ensemble size and integration time, as well as the impact on mean model response estimation of configuration parameters, such as spectral nudging and the size of the computational domain. An application of the present theoretical considerations is illustrated by considering the response of the Canadian RCM to a perturbation of a deep-convection threshold parameter. The results show that the set-up using the large model domain without spectral nudging was the one being most representative in terms of its inter-annual variability. Using spectral nudging can reduce the error variance of the difference of time means between the control- and modified-model simulations. However, for spectral nudging a possibly reduced sensitivity to deep-convection parameter perturbation was found.

Key words: Regional climate models, internal variability, reproducible components, spectral nudging, domain size, difference of means.

2.1 Introduction

The primary tool for studying the climate system are Coupled Global Climate Models (CGCMs) that consist of Atmospheric General Circulation Models (AGCMs) coupled with ocean, cryosphere and other components. Because of their high complexity and the need to perform very long simulation to achieve dynamical equilibrium and statistical significance, CGCM simulations have to be performed on comparatively low-resolution computational grids and lack fine-scale details in representation of atmospheric dynamics and physics and processes at the interface between the atmosphere and other components of the climate system. One-way nested Regional Climate Models (RCMs) have been widely used to enhance the resolution of climate simulations by downscaling coarse-resolution global fields over a region of interest. In this paradigm, information derived from CGCM simulations or objective analyses provide the initial, lateral boundary and ocean-surface conditions, for integration of atmospheric and land-surface variables over a limited area of the globe, using high-resolution computational grids (e.g., McGregor 1997, Giorgi and Mearns 1999, Wang et al. 2004, Laprise 2008, Rummukainen 2010).

Climate models are initialized from an arbitrary state and integrated forward in time, given prescribed external forcing, such as orbital parameters and solar radiation, and atmospheric composition including trace substances. Due to the chaotic nature of the atmosphere and models, the precise temporal evolution of the atmosphere from an observed initial state is unpredictable beyond a limit of a few days or weeks after the initialization. Climate simulations, hence, have no skill in providing a detailed temporal evolution of the instantaneous weather patterns. Climate models are believed however to have skill in reproducing long-term climate statistics, which is the grounds for climate modeling and climate projections. Essentially, when comparing two climate simulations conducted with two models (or the same model but different simulation setup or external forcing), the issue the researcher faces is the detection of the model response to modification (signal) among the quasi-random noise generated by weather. The model response has to be assessed by estimating the statistics of the difference between two models. The statistical significance of such estimates depends on the size of the sample of differences but also on statistical properties of the noise originating in the natural variability of the system.

The purpose of this manuscript is to develop a theoretical framework for the es-

timation of RCM response to modification of the model code (such as, for example, modification of the horizontal resolution, a new numerical algorithm, perturbed parameter settings, a parameterization, nesting technique) or of the simulation configuration (such as LBC from different sources, spectral nudging option, or size of the computational domain). There are many possible ways of comparing two simulations. Here we focus on first- and second-moment statistics of the difference between simulations, such as the mean, the variance and the correlation.

The reasons for developing a specific framework for RCM studies are multiple. In autonomous CGCMs, differences among identical integrations that depart from slightly different initial conditions (ensemble members) become on average as large as two randomly chosen states of a single CGCM integration, implying that CGCMs' internal variability is as large as the natural variability of the system, discarding for simplicity, long term predictability in CGCMs and non-stationarities, such as those caused by anthropogenic forcing. A distinct feature of RCM simulations is that they typically exhibit some predictive skill of instantaneous weather patterns, conditional on the information prescribed at the Lateral Boundary Conditions (LBC) and ocean surface (e.g., de Elía et al. 2002). The time variability of RCM simulations is partly enforced by the driving fields and partly originates in the RCM internal dynamics. In RCMs simulations, the inter-member variance is typically smaller than in CGCMs (e.g., Giorgi and Bi 2000, Caya and Biner 2004, Alexandru et al. 2007, Separovic et al. 2008, Nikiema and Laprise 2011). Furthermore the magnitude of the time-average inter-member variance is reduced by the application of large-scale spectral nudging (SN) or by decreasing the size of the computational domain (e.g., von Storch et al. 2000, Biner et al. 2000, Juang and Hong 2001, Miguez-Macho et al. 2004, Alexandru et al. 2009). These properties of RCMs have implications on the estimation of the RCM-simulated climate statistics. Understanding the specific nature of RCM-simulated climate variability may facilitate the choice of an appropriate simulation configuration for RCM testing. Moreover, it may help to optimize the computational resources available for an RCM experiment.

This work is organized as follows. In Section 2.2 we develop the theoretical framework for quantifying variability in a single RCM. In Section 2.3 we turn our attention to the analysis of the difference between time series obtained from simulations conducted with two different RCMs. In Section 2.4 we provide some examples from the RCM simulations. Summary is placed in Section 2.5.

2.2 Reproducible and irreproducible components of an RCM simulation

2.2.1 General assumptions

Let us assume a climate model variable ψ that varies as a function of time and is available at discrete instants $t_k \in (1, \ldots, K)$. This variable is also a function of space, but this will be neglected, as we focus on the temporal variability. Depending on the occasion, times t_k will refer to integration time steps, daily or seasonal time averages, etc. Variable ψ is also a function of ensemble realization m. The ensemble realizations can be sampled, for example, by generating runs with different initial conditions, within otherwise identical setup. Such an ensemble experiment with M members will be denoted as Ω. The set ψ_{km} can be thought of as a matrix of the size K by M as shown in Fig. 2.1, in which every column represents a series of realizations of the random variable ψ at times $t_k \in (1, \ldots, K)$, sampled from a single ensemble member, while every row represents the ensemble of M realizations of ψ at a single time t_k.

We will assume that the variable ψ is stationary in the statistical sense, which implies that an initial period of adjustment of the climate model simulation towards a quasi-equilibrium state (in which the initial conditions are forgotten and the spin-up of the differences between ensemble members is completed) is discarded. The adjustment depends on the component of the climate system and may require integration periods from a few years to decades for the spin-up of the land surface, to thousands of years for the deep ocean in CGCMs (Bryan 1998). In the case of RCMs, the spin-up of the atmosphere mainly concerns generation of fine-scale features that are absent from the coarse-resolution initial fields, and this requires only a few days (de Elía et al. 2002).

Fig. 2.1 schematically illustrates typical results of time and ensemble averaging in case of very large K and M, in CGCMs and RCMs. Time average is represented as the row on the upper side of the matrix ψ_{km}, while ensemble average is represented as the column on the right side of the matrix. In ensemble integrations of CGCM in a state of quasi-equilibrium, the ensemble average computed from a large number of members has little or no variability in time (excluding the annual cycle). This is illustrated schematically with the straight line on the right-hand side of Fig. 2.1. To eliminate the effect of

the annual cycle, we can think of the mth member's time series ψ_{km} as consisting of the model output gathered from consecutive years but over only a specific season. Unlike the situation with CGCMs, a considerable amount of temporal variability is enforced in RCMs through the use of the same time-dependent LBC that all members share at time t_k. This is analoguous to AGCMs that are forced by prescribed ocean surface; all members of the ensemble Ω share the same external time-varying lower boundary condition forcing, which results in some temporal variability that subsists ensemble averaging (Zwiers 1996). The LBC forcing in RCMs however is much more powerful than the lower boundary condition forcing in AGCMs.

Experimental evidence shows that in RCM ensemble integrations driven by identical LBC, the differences among members eventually become negligibly small when averaged in time over sufficiently long time periods. Averaging in time over 20 years or more typically reduces the differences between RCM ensemble members to rather small values (e.g., de Elía et al. 2008). Such an assumption is also intuitive in CGCMs, albeit it may require much longer averaging times (Kirtman et al. 2005). This is schematically represented with the two straight lines on the upper side of Fig. 2.1; in both CGCMs and RCMs the differences between members are assumed to vanish when $K \to \infty$.

It is also usually assumed that the statistics of the CGCM output can be retrieved equivalently by sampling ensemble realizations at a single time, or a time series of a single ensemble member; this is referred to as the ergodic assumption (e.g., von Storch and Zwiers 2002). In other words, the simulated natural and inter-member variability are interchangeable terms in CGCMs. The situation is different however in non-autonomous systems such as AGCMs and RCMs, in which part of temporal variability is contributed by external forcing and, hence, may not be retrievable by sampling the output from ensemble members, at a given time. We will now summarize these considerations and examine their consequences in a more rigorous manner.

2.2.2 Definition of the reproducible and irreproducible components

In order to take into account the above considerations, it is convenient to think of the RCM variable ψ as a random variable with values that depend on time and ensemble member. We define the expected value over time as the arithmetic average over time of

a single member realization in the limit when $K \to \infty$, and denote it as E_t. In fact, to omit the annual cycle from consideration, E_t denotes the expected value for a specified season over many years. We also define the expected value over the ensemble as the arithmetic average of all ensemble members at a given time in the limit when $M \to \infty$, and denote it as E_e. The composition of the two expectation operators $E_t[E_e]$ will be denoted by the shorthand E.

We will make the assumption that the time mean of every time series provided by different ensemble members converge to the same value as the integration time increases. According to this assumption, all ensemble members have the same climatological (time) mean, which can be written as follows:

$$\mu_c = E_t[\psi] = E[\psi]. \tag{2.1}$$

For the case of variables with significant trends this may be not the most appropriate definition and, hence, we assume a stationary climate and not the climate change. Next, the ensemble mean is defined as

$$\mu_e(t_k) = E_e[\psi]. \tag{2.2}$$

Note that μ_c is a constant and μ_e is a function of time t_k. The time anomaly of the ensemble mean is defined as

$$f(t_k) = \mu_e(t_k) - \mu_c. \tag{2.3}$$

Also note that f vanishes in the CGCM case due to the ergodic assumption that implies $E_t[\psi] = E_e[\psi]$, but in the RCM case this is not suitable. Further, members' deviations with respect to the ensemble mean are given as

$$\varepsilon(t_k, m) = \psi(t_k, m) - \mu_e(t_k). \tag{2.4}$$

The ensemble deviations are function of member and time.

The previous definitions imply that the variable ψ may be decomposed as follows:

$$\psi(t_k, m) = \mu_c + f(t_k) + \varepsilon(t_k, m). \tag{2.5}$$

Thus, RCM variable ψ is decomposed into its climatological mean μ_c and two compo-

nents of the time anomaly. Variable f will be referred to as the *reproducible anomaly.* It quantifies the component of the variation in time common to all ensemble members and in this sense reproducible. It can be interpreted as the externally forced component of the time variations of ψ, arising in response to the specific lateral-boundary and surface conditions and other external forcing that all ensemble members share at time t_k. Variable ε will be referred to as the *irreproducible anomaly.* It quantifies the component of the variation in time that is distinct in every member and in this sense irreproducible. It can be interpreted as the internal component of the time variations of ψ, arising due to the quasi-random, chaotic deviations in the mth member at time t_k. Note that Eq. (2.5) states that the time anomaly of every ensemble member with respect to its own time mean is quantified by the sum of f and ε; RCM temporal variability is, hence, provided partly in the reproducible and partly in the irreproducible form.

Note that it directly follows from Eq. (2.4) that $E_e[\varepsilon] = 0$, i.e., that the members' deviations at a given time have zero mean. Similarly, it follows from Eqs. (2.1–2.4) that $E_t[\varepsilon] = 0$, which implies that a single member's deviations from the ensemble mean tend to cancel out from its time average. Thus, a single simulation is sufficient to assess the climatological mean μ_c, given a long integration time.

2.2.3 Decomposition of variance

The total variance of RCM variable ψ can be defined as

$$\sigma^2 = E\big[(\psi - E[\psi])^2\big] = E\big[(f+\varepsilon)^2\big]. \tag{2.6}$$

The last expression on the rhs of Eq. (2.6) is derived with the help of Eqs. (2.1) and (2.5). This variance is computed by pooling all members' time series in a single set. We will also specify expressions for the variances in time (given a member of ensemble) and over the ensemble (given a specified time), respectively, as follows:

$$\sigma_t^2 = E_t\big[(\psi - E_t[\psi])^2\big] = E_t\big[(f+\varepsilon)^2\big], \tag{2.7}$$

$$\sigma_e^2 = E_e\big[(\psi - E_e[\psi])^2\big] = E_e\big[\varepsilon^2\big]. \tag{2.8}$$

These statistics will be referred to as time variance and ensemble variance, respectively. Note that, in general, the variance operator σ^2 is a constant, while σ_t^2 is a function of

ensemble member m and σ_e^2 a function of time t_k.

However, since all RCM members share the same LBC and other external forcing, we assume that all members have equal time variances, which implies that, in the limit of very long time series, sample variances of every member's time series converge to the same value. Recall that we also assumed that all members have equal time means (Eq. 2.1). Under these two assumptions, pooling several ensemble members in one grand ensemble only increases the sample size and the statistical robustness of sample statistics, but it does not change the variance of the sampling distribution. These considerations about the variance can be summarized in the following relation:

$$\sigma_t^2 = \sigma^2. \tag{2.9}$$

Specifically, in CGCMs, because of the ergodic assumption, this relation can be further extended to $\sigma_t^2 = \sigma_e^2 = \sigma^2$. Hence in CGCMs, upon neglecting the annual cycle and non-stationary external forcing, all three variances are equal and hence they can be characterized by a single variance, e.g., time variance σ_t^2. The situation however is different for RCMs, where a distinction needs to be made between the total (or, equivalently, time) variance σ_t^2 and the time-dependent ensemble variance σ_e^2. Nested RCMs simulations are not ergodic in the sense that their variability in time cannot be sampled from ensemble members at a given time.

Let us now express the time variance in terms of the reproducible and irreproducible anomalies, f and ε. From Eqs. (2.6) and (2.9) we obtain:

$$\sigma_t^2 = E_t\Big[E_e\big[f^2 + \varepsilon^2 + 2\,f\varepsilon\,\big]\Big]. \tag{2.10}$$

Since $E_e[\varepsilon] = 0$ the expectation E_e of the last term vanishes and we obtain the following decomposition for the temporal variance:

$$\sigma_t^2 = \sigma_f^2 + \sigma_\varepsilon^2, \tag{2.11}$$

where

$$\sigma_f^2 = E_t\big[f^2\big] = E_t\big[(\mu_e - \mu_c)^2\big], \tag{2.12}$$

$$\sigma_\varepsilon^2 = E_t\Big[E_e\big[\varepsilon^2\big]\Big] = E_t[\,\sigma_e^2\,], \tag{2.13}$$

will be referred to, respectively, as the reproducible and irreproducible components of the total (or, equivalently, time) variance. Equation (2.12) shows that the reproducible variance is equal to the time variance of the ensemble mean μ_e. It quantifies the variance of the reproducible, externally forced time anomalies f of variable ψ, with respect to the climatological mean μ_c. Equation (2.13) represents the component of the total (time) variance originating in irreproducible time anomalies ε. This component is quantified by the time-average ensemble variance σ_ε^2.

Because the two components are additive, it is convenient to express the partition in relative terms, using a single parameter. For this purpose, we can introduce the *reproducibility ratio*, similarly as in Separovic et al. (2008), as follows:

$$\rho^2 = \frac{\sigma_f^2}{\sigma_t^2}. \tag{2.14}$$

This parameter takes values between 0 and 1; it is at the minimum when the reproducible component of the time variance vanishes. In CGCMs this term would be vanishingly small beyond their predictability limit, because reproducible anomalies f are negligible. On the time scales of days it also vanishes in AGCMs since the instantaneous weather patterns are unpredictable in AGCMs once the initial conditions are forgotten, but it may be positive on longer time scales due to the prescribed inter-annual variability of the ocean surface. In RCMs it generally takes positive values at all time scales due to the control exerted by the LBC. We now address the case of RCMs in detail.

Note that the time variance σ_t^2 is an observable. When an RCM is driven by reanalysis, one hopes that the temporal variability of the simulations is close to that of the true atmosphere. The sum of the reproducible and irreproducible variances, as defined by Eq. (2.11), can thus be understood as a measure of the simulated natural variability by an RCM, given prescribed lateral boundary and surface forcing.

2.2.4 Redistribution between reproducible and irreproducible variances

The relation (2.11) that states that the sum of the reproducible and irreproducible variance components, σ_f^2 and σ_ε^2, is an observable, has profound consequences in interpreting RCMs experimental results. Under the assumption that the temporal variance σ_t^2 is not

changed by some experimental modification to the model, an increase (decrease) in the irreproducible variance caused by an RCM modification has to be compensated by a decrease (increase) by an equal amount in the reproducible variance. This implies that if a modification to the model code or simulation setup (such as the application of SN or using a smaller computational domain) had a single effect on the RCM simulations, such as reducing the spread of ensemble members, then it would only redistribute the variance between the reproducible and irreproducible components. Therefore, (2.11) can be thought of as a sort of conservation law: with a model that perfectly represents the time variability of the true climate, the sum of the two components has to be a constant determined by the observations.

Since, in practice, the time variance of RCM simulations can somewhat depart from the observed values, RCM modifications may produce an improvement or a deterioration of the skill in reproducing the time variance; large discrepancies with observations would however typically yield to discarding the modification. An example that illustrates how the SN changes the temporal variance will be shown in Section 2.4. In order to take into account the imperfect nature of RCMs, we should understand Eq. (2.11) not as an absolute law but rather as a desired ideal that the sum of the reproducible and irreproducible components should be unaffected by the modification, if the control RCM simulation skilfully reproduces the time variance of the true climate.

Unlike σ_t^2, the sum of the reproducible and irreproducible variances, which is an observable, there are no general expectations about the relative magnitude of the reproducible and irreproducible variances, σ_f^2 and σ_e^2, in RCM simulations. A number of studies show that the time-average ensemble variance (and hence the irreproducible variance σ_ε^2) in RCMs simulations can be controlled by the selection of the simulation configuration, such as SN and the size of the computational domain (e.g., Weisse and Feser 2003, Rapaić et al. 2011). It follows from the above considerations that in RCM simulation configurations characterized with small ensemble variance, such as in spectrally nudged or small-domain simulations, the small irreproducible variance component should be compensated by a larger reproducible variance. However, when the ensemble variance is large, then averaging over members filters out these differences and results in a small reproducible variance component. In the limit when the ensemble variance is as large as the time variance, the reproducible component vanishes. These considerations have important implications on the estimation of the RCM simulation mean from

ensemble integrations. This issue is discussed next.

2.2.5 Estimation of the climatological mean from ensemble integrations

The true model climatological mean μ_c can be estimated from a finite sample of data ψ_{km} of size MK, sampled from M ensemble members at K times, by averaging each member in time and then computing the ensemble average of the members' time-average values. The sample mean obtained in this way is defined as

$$\hat{\mu}_c = \frac{1}{MK} \sum_{k=1}^{K} \sum_{m=1}^{M} \psi_{km}. \tag{2.15}$$

The estimate $\hat{\mu}_c$ is also a random variable whose variance depends on the time series length K, on the ensemble size M but also on the variance of the sampling distribution σ^2, given in Eq. (2.6).

For simplicity we begin the discussion with the CGCM case. Since CGCMs simulations are assumed to be ergodic, it is equivalent to sample their variability over time or over ensemble members. Thus, in the CGCM case, we may think of the matrix ψ_{km} in Fig. 2.1, as the sample of KM variables that, in statistical terminology, all follow the same distribution, where K is the number of independent data in the time series provided by every ensemble member and M the number of ensemble members. The CGCM ensemble members are not correlated in time. The variance of this distribution is σ_t^2, given in Eq. (2.7), since the ensemble variance is saturated at the value equal to the members' time variance and all members have the same time mean. Further, since the variance of the mean of a sample of independent identically distributed data is equal to the ratio of the sample variance and the sample size (e.g., Wilks, 2005), the variance of the estimated climatological mean for CGCM case is given as

$$Var[\hat{\mu}_c]_{\text{CGCM}} = \frac{\sigma_t^2}{T}, \tag{2.16}$$

where $T = MK$ is proportional to the total computing time of making an ensemble of M members simulations of time length proportional to K. This relation shows that, in principle, the error in the CGCM sample mean can be reduced by increasing the total computing time, regardless of how it is partitioned between the ensemble size M and

members' integration time K. Carrying several shorter simulations instead of one long CGCM simulation, which can be very convenient for practitioners, hence, has no effect on the error variance of the estimated CGCM climatological mean.

Appendix 2.6 shows the development of the RCM case. Here we will just discuss the result and elaborate on its consequences. The variance of the estimated climatological mean for RCMs can be quantified as follows:

$$Var[\hat{\mu}_c] = \frac{\sigma_f^2}{K} + \frac{\sigma_\varepsilon^2}{T}, \tag{2.17}$$

where $T = KM$, σ_f^2 and σ_ε^2 are given in Eqs. (2.12) and (2.13). Equation (2.17) shows that increasing the ensemble size M can decrease the variance of $\hat{\mu}_c$ only to a value determined by σ_f^2/K. Given relatively short integration periods, generating more ensemble members can provide the benefit of reducing sampling variance of $\hat{\mu}_c$ only when the spread among the members is large. As implied by Eq. (2.11), then the reproducible part of the variance, σ_f^2, becomes small. As a result, variability among ensemble members approaches in magnitude the variability in time, as in CGCMs. When, on the other hand, the integration period K increases, the sampling uncertainty in $\hat{\mu}_c$ becomes small, regardless of the ensemble size; in the limit when $K \to \infty$, only one simulation suffices to provide robust estimate of the models' mean.

It is worth noting that the total available computing time $T = MK$ for a RCM experiment is typically a non-negotiable constant imposed by external factors, such as the available computational resources and the time frame of the experiment. Equation (2.17) implies that the optimal way to reduce the sampling error in $\hat{\mu}_c$, for an available computing time T, is to use a single ensemble member, $M = 1$, which maximizes the integration time, $K = T$.

As shown in Alexandru et al. (2009), a reduction of the time-average ensemble variance σ_e^2 (or, equivalently, the irreproducible variance σ_ε^2) in RCM simulations can be achieved by means of reduction of the domain size or application of SN. Let us now consider how these methods may affect the error variance in the estimated climatological RCM mean $\hat{\mu}_c$. In order to show this, we will assume, for simplicity, that the application of SN or reduction of domain size does not change the simulated natural variability

measured by the time variance σ_t^2 (Eq. 2.11). We assume that their sole effect is to reduce the irreproducible component σ_ε^2 and to transfer an equal amount of variance to the reproducible component σ_f^2. Under this assumption, differentiating Eq. (2.11) yields $d\sigma_f^2 = -d\sigma_\varepsilon^2$. Hence, we obtain from Eq. (2.17):

$$\frac{dVar[\hat{\mu}_c]}{d\sigma_\varepsilon^2} = -\frac{T-K}{TK}. \tag{2.18}$$

A reduction of the time-average ensemble variance by SN or in a small domain implies a negative value of $d\sigma_\varepsilon^2$ and, hence, $dVar[\hat{\mu}_c]$ is positive, implying that the error variance of $\hat{\mu}_c$ increases. This happens because, in simulation configurations characterized with large ensemble spread, members are more efficient in sampling RCM's temporal variability. A large ensemble variance also implies that the RCM solution is more autonomous with respect to the external forcing. This situation is similar to the CGCM case, when ensemble members are as efficient in sampling simulated natural variability as are the time series of an individual ensemble member. On the other hand, when ensemble spread is reduced by SN or in a small RCM domain, ensemble members are less efficient in sampling temporal variability of the system so the variance of the estimated mean increases. The only exception when reduction of spread does not increase the error variance of $\hat{\mu}_c$ is for $K = T$, i.e., when there is only one ensemble member. It then follows from Eq. (2.18) that $dVar[\hat{\mu}_c]/d\sigma_\varepsilon^2 = 0$.

In order to quantify the efficiency of an RCM ensemble in sampling temporal variability we first note that the ensemble members' time series may be, in general, serially correlated, since their time anomalies have in common the externally-forced reproducible time components f, given in Eq. (2.3). Thus, the effective sample size provided by an ensemble of M members with integration time K may be smaller than the nominal sample size $T = MK$, when the sampling of the RCM-simulated natural variability is considered. The effective sample size T_{eff} can be defined as the sample size that would produce the same error variance in $\hat{\mu}_c$ as the nominal sample size T, but if the ensemble members were not serially correlated. Under this assumption the expression for the error variance of RCM time mean $\hat{\mu}_c$ in Eq. (2.17) would be the same as in CGCMs in Eq. (2.16) and, hence, we can write:

$$Var[\hat{\mu}_c] = \frac{\sigma_t^2}{T_{\text{eff}}}, \tag{2.19}$$

where σ_t^2 is the time variance of RCM variable ψ.

Upon solving Eqs. (2.11), (2.17) and (2.19) for T_{eff} we obtain:

$$T_{\text{eff}} = \frac{T}{1 + \frac{T-K}{K}\rho^2}. \tag{2.20}$$

where ρ^2 is the reproducibility ratio, defined in Eq. (2.14). It can be seen that if the irreproducible variance σ_ε^2 is large with respect to σ_t^2, as in CGCMs, then $\rho^2 \approx 0$ and, hence, from Eq. (2.20) we obtain that the equivalent sample size is at its maximum, $T_{\text{eff}} \approx T$. This means that the effective sample size remains the same regardless of how the computing time T is partitioned between the ensemble size M and the integration period K. On the other hand, if σ_ε^2 is small with respect to σ_t^2, then $\rho^2 \approx 1$ and we obtain from Eq. (2.20) that $T_{\text{eff}} \approx K$, which is its minimum value. In the latter case, all M members become almost identical, as if the ensemble had only one member with integration period K, which yields the effective sample size reduced to K.

According to experimental evidence, the application of SN and reduction of domain size are very likely to reduce the irreproducible variance σ_ε^2 and, thus, increase the reproducibility ratio ρ. According to Eq. (2.20), this results in a reduction of the equivalent sample size T_{eff}. In other words, a spectrally nudged or a small-domain ensemble typically provide a less representative sample of the RCM-simulated natural variability than a non-nudged ensemble of the same size M and the same integration time K, or an ensemble performed over a larger domain. However, when the entire computing time T is allocated to a single member, $M = 1$, then $K = T$ and we obtain from Eq. (2.20) that in this case $T_{\text{eff}} \equiv T$, regardless of the ensemble spread measured by ρ. Consequently, the SN and domain-size reduction decrease T_{eff}, unless the ensemble contains a single member because in this case there is no serial correlation of the data for the trivial reason that there is no multiple members. Also note in Eq. (2.19) that the error variance is inversely proportional to T_{eff} and, hence, SN and domain-size reduction both increase the error variance in the RCM climatological mean, given $M > 1$.

The previous considerations imply that the optimal choice for the estimation of the climatological mean is the use of single member ensembles with the longest integration period permitted by the available computing time. It is worth noting that

this recommendation addresses only RCM ensembles generated with the same LBC and surface conditions, which implies that all members of the ensemble have the same reproducible anomalies f (Eq. 2.3). There are other ways of generating RCM ensembles, the most important example being the ensembles in which a different CGCM or a different member of a CGCM ensemble drives different RCM members. In this case, there would be no reproducible RCM components and, hence, the variance of the estimated RCM climatological mean $\hat{\mu}_c$ would be equivalent to that in CGCMs (Eq. 2.16). In other words, it would be equivalent to sample time variability from members or in time in such an ensemble, implying that the ergodic assumption would be then suitable. When the objective is the éstimation of the climatological mean this would be a more efficient way to increase the sample size than to generate RCM members using a single CGCM simulation.

A distinction should be also made between the cases when the researcher attempts (a) to estimate the long-term RCM mean $\hat{\mu}_c$ and (b) to downscale a particular weather event or season from the objective analyses or some other specified external forcing. In the latter case, what is typically estimated is the ensemble mean μ_e, given in Eq. (2.2), averaged in time over only a specified period K_0, while the long-term time mean is then irrelevant. We denote such an estimator with $\overline{\mu}_e$. It has the same expression as $\hat{\mu}_c$ in Eq. (2.15), with K_0 prescribed. In this context, inter-member variability has to be considered as the only source of sampling error. The following relation is then appropriate for the error variance in the mean $\overline{\mu}_e$:

$$Var[\overline{\mu}_e] = \frac{\overline{\sigma_e^2}^{K_0}}{MK_0}, \tag{2.21}$$

where $\overline{\sigma_e^2}^{K_0}$ is the time-average ensemble variance over period K_0. Evidently, the only way in this case to reduce the error is to generate more ensemble members. Note that in this case it is very convenient to generate ensemble members that all share the same time-varing lateral boundary and surface conditions, since then the time variance is partitioned between the reproducible and irreproducible components, implying that $\overline{\sigma_e^2}$ is smaller than the time variance. In addition, the reduction of the ensemble variance by SN or domain-size reduction would typically reduce the variance of this estimator, by reducing, on average, the ensemble variance σ_e^2.

Now we turn our attention to the difference between simulations conducted with two different RCMs.

2.3 Analysis of the RCM response to modification

The difference between two simulations can also be analyzed in terms of the reproducible and irreproducible components. We consider now two model versions, the control and the modified. We do not need to be very specific about the modification; the modified version can be obtained by changing one or several factors related to the model structure or configuration (e.g., the resolution, a new parameterization is introduced, nesting technique changed). However, it is assumed that the simulations of both models are performed within comparable setups. For simplicity, we will also assume that the control and modified model have the same number of ensemble members M conducted over the common integration period of K integration times. We denote the control and modified model ensembles as Ω_1 and Ω_2, respectively; they can be thought of as matrices of size KM, such as the one shown in Fig. 2.1. Under these assumptions, the differences can be also arranged in a matrix of size KM, denoted with $\Delta\Omega$.

2.3.1 Reproducible and irreproducible components of the RCM response to modification

Similarly as it was done in the single-model case we can define the difference of climatological means of the variable ψ as

$$\Delta\mu_c = E[\psi_2] - E[\psi_1], \tag{2.22}$$

where $E[\cdot] = E_t\big[E_e[\cdot]\big]$. As before this term is assumed to be a constant. Next, the difference of ensemble means can be defined as

$$\Delta\mu_e(t_k) = E_e[\psi_2] - E_e[\psi_1]. \tag{2.23}$$

It is a function of time whose time anomaly can be defined as

$$\Delta f(t_k) = \Delta\mu_e(t_k) - \Delta\mu_c. \tag{2.24}$$

This quantity represents the reproducible anomaly of the difference. On the other hand, similarly to the single-model case, the irreproducible anomaly of the difference is given as

$$\Delta\varepsilon\,(t_k, m) = \Delta\psi(t_k, m) - \Delta\mu_e(t_k), \tag{2.25}$$

where

$$\Delta\psi = \psi_2 - \psi_1, \tag{2.26}$$

The quantity $\Delta\varepsilon$ is a function of member and time. Finally, the difference $\Delta\psi$ between two RCMs can be summarized as follows:

$$\Delta\psi(t_k, m) = \Delta\mu_c + \Delta f(t_k) + \Delta\varepsilon(t_k, m). \tag{2.27}$$

Using these definitions, we begin the analysis of the difference by identifying the relevant parameters that quantify the variability in the RCM response to modification.

2.3.2 Analysis of the variance of response

We begin the analysis by recalling the assumption that, in every RCM ensemble the time statistics of individual members' time series become identical when $K \to \infty$, i.e., that in the limit of long integration time statistics do not depend on the choice of members. As already discussed in Section 2.2, for this reason, the time means $E_t[\psi_1]$ and $E_t[\psi_2]$, as well as the time variances σ^2_{t1} and σ^2_{t2}, of individual members of the corresponding ensembles Ω_1 and Ω_2 are invariant to the choice of members. As an implication of this assumption, the time variance of the difference between any two members given as

$$\sigma^2_{t\Delta} = E_t\big[(\Delta\psi - E_t[\Delta\psi])^2\big], \tag{2.28}$$

is also invariant to the choice of ensemble members. By substituting Eq. (2.26) in Eq. (2.28) and rearranging the terms we obtain:

$$\sigma^2_{t\Delta} = (\sigma_{t2} - \sigma_{t1})^2 + 2\,\sigma_{t1}\,\sigma_{t2}\,(1 - R), \tag{2.29}$$

where the time variances $\sigma^2_{t1,2}$ are defined in Eq. (2.7) and R is the time correlation between any two members from Ω_1 and Ω_2, defined as follows:

$$R = \frac{E_t\big[\big(\psi_1 - E_t[\psi_1]\big)\big(\psi_2 - E_t[\psi_2]\big)\big]}{\sigma_{t1}\,\sigma_{t2}} \tag{2.30}$$

and is also invariant with respect to the selection of members from Ω_1 and Ω_2. Next, we substitute for the two RCM variables ψ_1 and ψ_2 in Eq. (2.30) their decomposed forms given in Eq. (2.5). With the help of Eq. (2.1) we obtain:

$$R = \frac{E_t[f_1 f_2 + f_1\varepsilon_2 + f_2\varepsilon_1 + \varepsilon_1\varepsilon_2]}{\sigma_{t1}\,\sigma_{t2}}. \tag{2.31}$$

It is important to note that the reproducible time anomalies f_1 and f_2 in the two RCMs may be serially correlated, if ensembles Ω_1 and Ω_2 share the same LBC or other external forcing. Their correlation in time may be quantified with the cross-correlation coefficient defined as

$$R_f = \frac{E_t[f_1 f_2]}{\sigma_{f_1}\sigma_{f_2}}, \tag{2.32}$$

where σ_{f_1} and σ_{f_2} are the reproducible time variance components given in Eq. (2.12) and represent the time variances of f_1 and f_2, respectively. The ensemble deviations ε_1 and ε_2 in the control and modified model are independent and, hence, their time covariance $E_t[\varepsilon_1\varepsilon_2]$ vanishes. Finally, for the terms of the form $f_i\,\varepsilon_j$ $(i \neq j)$ in Eq. (2.31) we can write:

$$E_t[f_i\,\varepsilon_j] = E_e\big[E_t[f_i\,\varepsilon_j]\big] = E_t\big[E_e[f_i\,\varepsilon_j]\big] = E_t\big[f_i\,E_e[\varepsilon_j]\big] = 0, \tag{2.33}$$

where we used the assumption that the time statistics do not depend on ensemble members and the fact the $E_e[\varepsilon] = 0$. Upon taking account of these considerations, we can rewrite Eq. (2.31) in a concise form as follows:

$$R = \frac{\sigma_{f_1}}{\sigma_{t1}}\,\frac{\sigma_{f_2}}{\sigma_{t2}}\,R_f = \rho_1\rho_2\,R_f. \tag{2.34}$$

Here, ρ_1 and ρ_2 are the reproducibility ratios of the two models defined in Eq. (2.14). Parameter R_f will be referred to as the *reproducible correlation.* Its value may be zero if the control and modified RCM integrations do not share the same time-dependent external forcing. This would be the case, for example, if the control and modified RCM employed two different CGCM simulations as the LBC. On the other hand, if the LBC in the two RCMs are the same, then R_f will likely be positive.

With the help of Eq. (2.34) we can rewrite Eq. (2.29) as

$$\sigma_{t\Delta}^2 = (\sigma_{t2} - \sigma_{t1})^2 + 2\sigma_{t1}\sigma_{t2}\,(1 - \rho_1\rho_2\,R_f). \tag{2.35}$$

This shows that the time variance of the difference between any two RCM simulations depends on five parameters: their standard deviations in time σ_{t1} and σ_{t2}, the simulations' reproducibility ratios ρ_1 and ρ_2, and reproducible correlation R_f. We now discuss the meaning of this decomposition.

In a hypothetical modeling system that would not be sensitively dependent to initial conditions, infinitesimally small perturbations in the initial conditions would remain infinitesimally small throughout the integration process. If this is so, there would be almost no spread among ensemble members, $\varepsilon \approx 0$, and hence, from Eq. (2.14) we obtain $\rho_1 \approx \rho_2 \approx 1$. In such systems, the difference between the control- and modified-model simulations would be entirely determined by the nature of the modification. The time variance of this difference could be quantified by the reproducible correlation R_f and the change in the time standard deviation σ_t; these parameters quantify the change in the evolution of the system's states due to modification, by quantifying the synchrony and average amplitudes of the two simulations. Thus a minor modification in the model would yield a minor variance of the difference. For this to happen it would be sufficient that the simulations have approximately equal time standard deviations, $\sigma_{t1} \approx \sigma_{t2}$, and that the reproducible correlation $R_f \approx 1$. As it can be seen from Eq. (2.35), it follows that the variance of the response is $\sigma^2_{t\Delta} \approx 0$. Note that, in this case, the two simulations may still have a large difference of means $\Delta\mu_c$, implying a response to modification that is constant in time.

On the other hand, in chaotic systems such as RCMs, $\rho < 1$ and the variance of the difference between two RCM simulations $\sigma^2_{t\Delta}$ is not only due to the modification per se; the modification also triggers internal chaotic variations ε that may become large. Even when the modification to the model tends to be infinitesimally small (which implies $\sigma_{t2} \to \sigma_{t1}$ and $R_f \to 1$), there will still be some variability in the difference between the two simulations, because the chaotic variations are triggered by the model modification: it can be seen in Eq. (2.35) that in this case $\rho < 1$ implies $\sigma^2_{t\Delta} > 0$. This is exactly what occurs when considering two members of the same ensemble, their only difference being a perturbation in the initial conditions. Two members of the same ensemble, by definition, must have identical ensemble means and hence from Eq. (2.32) we obtain $R_f \equiv 1$. They also have the same reproducibility ratios, $\rho_1 = \rho_2 = \rho$. Thus, from Eq. (2.34) we obtain that the time correlation between two ensemble members of the same model is $R = \rho^2$. This shows that the reproducibility ratio quantifies de-correlation in

time between two simulations due solely to the chaotic variations ε, which would be present for even vanishingly small modifications.

The above considerations can be summarized by stating that in the set $\{\rho_1, \rho_2, \sigma_{t1}, \sigma_{t2}, R_f\}$ the first two parameters quantify the contribution of the chaotic nature of the two models to the response and the latter three quantify the contribution of the modification per se. Let us now compare this conclusion to the case of CGCMs. Since CGCMs' reproducibility ratio ρ vanishes, Eq. (2.35) implies that the time variance of the difference between two CGCM simulations is saturated at the value equal to the sum of their time variances, i.e., $\sigma^2_{t\Delta} = \sigma^2_{t1} + \sigma^2_{t2}$. This case can be thought of as the trivial solution to the problem, since CGCM modification has no other effect on the time variance of the response than the change in time variance itself; the time variance of the CGCM response is described with $\{\sigma_{t1}, \sigma_{t2}\}$. The variability in the difference between two CGCM simulations is, thus, entirely generated by their internal chaotic variations.

In fact, the estimation of the parameter R_f is meaningful only when the reproducibilities ρ are relatively high, such as in specific RCM configurations, e.g., when the domain size is not very large. Low reproducibility ratios mean that a small amount of time variance in Eq. (2.11) is partitioned into the reproducible component in Eq. (2.12). This implies that the ensemble mean μ_e has little variability in time and that R_f explains only a small amount of time variance of the difference between two simulations, thus having little physical significance.

2.3.3 Estimation of the reproducible and irreproducible time variance components from two ensemble members

An important consequence of the previous considerations is that the partition of time variance into reproducible and irreproducible parts for a given RCM can be estimated with two ensemble members, given that the integration time is sufficiently long. Let us consider two ensemble members of the same model, say the control model Ω_1, and denote them with ψ' and ψ''. Since these two simulations are drawn from the same ensemble, by definition they must have the same ensemble mean and hence their reproducible correlation be $R_f \equiv 1$. They also have the same reproducibility ratio ρ by definition,

while their time variances σ_t^2 are assumed to be the same. Upon substituting these values in Eq. (2.35) we obtain the following expression for the time variance of the difference between two ensemble members:

$$\sigma^2_{t\,(\psi''-\psi')} = 2\,\sigma_t^2\,(1-\rho^2). \tag{2.36}$$

With the help of Eqs. (2.11) and (2.14) this can be rewritten as follows:

$$\sigma_\varepsilon^2 = \frac{1}{2}\,\sigma^2_{t\,(\psi''-\psi')}, \tag{2.37}$$

which shows that the irreproducible component of the variance, given in Eq. (2.13), is equal to the half the time variance of the difference between two members of the same model.

The above considerations have important implications on the estimation of the irreproducible variance σ_ε^2: it can be estimated by computing the variance of a sample of differences between two ensemble members. It is worth noting that each of the two ensemble members also provides a time series from which the simulations' time variance σ_t^2 can be estimated. Finally, the reproducible variance component σ_f^2, as well as the reproducibility ratio ρ, can be computed from these two estimates using Eqs. (2.11) and (2.14), respectively.

Reproducible correlation R_f can also be estimated if both control and modified RCMs have two ensemble members; once the estimates of reproducibility ratios ρ_1 and ρ_2 and the time correlation R between the control and modified model simulations, defined in Eq. (2.30), are estimated, parameter R_f can be computed from these estimates using Eq. (2.34). In Section 4, when we show results of RCM response to modification of physics parameters, we adopt this approach for estimating these parameters.

Finally, we can also decompose the variance of model response into reproducible and irreproducible components. This can be done by decomposing time variances σ_{t1}^2 and σ_{t2}^2 using Eq. (2.11), which allows us to rewrite Eq. (2.35) as follows:

$$\sigma^2_{t\,\Delta} = \sigma^2_{\Delta f} + \sigma^2_{\Delta\varepsilon}, \tag{2.38}$$

where

$$\sigma^2_{\Delta f} = (\sigma_{f_2} - \sigma_{f_1})^2 + 2\sigma_{f_1}\sigma_{f_2}(1 - R_f), \tag{2.39}$$

$$\sigma^2_{\Delta\varepsilon} = \sigma^2_{\varepsilon_1} + \sigma^2_{\varepsilon_2}, \tag{2.40}$$

are the corresponding reproducible and irreproducible variances of the model response to modification. Equations (2.38)-(2.40) provide the framework for quantifying the reproducible and irreproducible components of RCM transient response to modification.

A complete discussion of the issue of estimation of parameters $\rho_{1,2}$ and R_f requires formal definitions of the corresponding estimators as well as the development of their variances. This would require lengthy derivations and is beyond the scope of this manuscript. Here, we rather intend to show that, given long integrations, estimation of these parameters does not impose the need for large number of ensemble members; two members per model version can provide robust estimates, given that the integration period is sufficiently long. In what follows, we turn our attention to the issue of estimation of the difference of means, which will be elaborated in detail, since it is a key issue in inter-model comparisons.

2.3.4 Estimation of the difference of RCM means

Ensemble experiments Ω_1 and Ω_2 are assumed to contain M time series of RCM output with K independent data. Hence, they also provide $T = MK$ independent realizations of the difference $\Delta\psi_{km}$ between two models' outputs. The difference of the climatological means, $\Delta\mu_c$, can be estimated as the arithmetic average over this set:

$$\Delta\hat{\mu}_c = \frac{1}{MK}\sum_{k=1}^{K}\sum_{m=1}^{M}\Delta\psi_{km}. \tag{2.41}$$

The variance of this estimator can be derived similarly as in Eq. (2.17) as:

$$Var[\Delta\hat{\mu}_c] = \frac{\sigma^2_{\Delta f}}{K} + \frac{\sigma^2_{\Delta\varepsilon}}{T}, \tag{2.42}$$

where $\sigma^2_{\Delta f}$ and $\sigma^2_{\Delta\varepsilon}$ are the reproducible and irreproducible components of the variance of model response to modification, defined in Eqs. (2.39) and (2.40), respectively. As before, $T = MK$ denotes the total computing time.

Let us first consider the CGCM case when the reproducible parts of the control and modified models' variances are negligible, while the irreproducible components are equal to the time variances σ_{t1}^2 and σ_{t2}^2. Hence, the variance of $\Delta\hat{\mu}_c$ for the CGCM case is given as:

$$Var\left[\Delta\hat{\mu}_c\right]_{\text{CGCM}} = \frac{\sigma_{t1}^2 + \sigma_{t2}^2}{T}. \tag{2.43}$$

This shows that the variance of the estimated difference of climatological mean depends on the total computing time T, while the allocation of the computing time between the number of members and their integration period has no effect on this variance.

Now we examine the RCMs in which the reproducible components are, in general, not negligible. Upon substituting Eqs. (2.39) and (2.40) in Eq. (2.42) we obtain the general result:

$$Var\left[\Delta\hat{\mu}_c\right] = \frac{(\sigma_{f_2} - \sigma_{f_1})^2 + 2\,\sigma_{f_1}\sigma_{f_2}(1 - R_f)}{K} + \frac{\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2}{T}. \tag{2.44}$$

It can be seen that this variance depends on the reproducible and irreproducible variance components in each of the two models, but also on the covariance between the reproducible anomalies, quantified with the reproducible correlation R_f.

In order to simplify the analysis of Eq. (2.44), we will define the equivalent sample size, similarly as was done for the single-model case in Eq. (2.19), but this time for the sample of differences between two models. We define the equivalent sample size as the sample size that would yield the same variance of $\Delta\hat{\mu}_c$ as in Eq. (2.44), but assuming that the members within each of the two ensembles are not correlated in time, such as the case for CGCMs in Eq. (2.43). This yields to:

$$Var\left[\Delta\hat{\mu}_c\right] = \frac{\sigma_{t1}^2 + \sigma_{t2}^2}{T_{\text{eff}}}. \tag{2.45}$$

Further, we will assume that (1) the control and modified RCM ensembles have approximately the same total variances $\sigma_{t1}^2 = \sigma_{t2}^2 = \sigma_t^2$ and (2) that their irreproducible components $\sigma_{\varepsilon_1}^2$ and $\sigma_{\varepsilon_2}^2$ are of similar magnitude so that, according to Eqs. (2.11) and (2.14), the two ensembles have also similar reproducible variances $\sigma_{f_1}^2 \approx \sigma_{f_2}^2$, as well as reproducibility ratios $\rho_1 \approx \rho_2 = \rho$. Note that these assumptions do not yield much loss

of generality. Assumption (1) can be thought of as an approximation, since the time variance should be constrained by observations and hence a reasonable modification should not produce a large change in the time variance. As for assumption (2), we note that the ensemble variance σ_e^2 and, thus also the irreproducible variance σ_ε^2, are typically very sensitive to SN and domain size. Hence assumption (2) can be thought of as the assumption that the control- and modified-model simulations are performed using the same computational domain and SN configuration. In any case, keeping unchanged the factors related to the simulation configuration would be a common-sense approach to quantifying RCM response to code modifications.

Under these assumptions and using the definition of the reproducibility ratio in Eq. (2.14), Eq. (2.44) can be rewritten as follows:

$$Var[\Delta\hat{\mu}_c] = 2\,\sigma_t^2 \left[\frac{\rho^2(1-R_f)}{K} + \frac{1-\rho^2}{T} \right] = \frac{2\,\sigma_t^2}{T_{\text{eff}}}. \tag{2.46}$$

Now we can solve Eq. (2.46) for T_{eff}, which leads to the following expression for the equivalent sample size:

$$T_{\text{eff}} = \frac{T}{1 + \frac{T(1-R_f)-K}{K}\rho^2}. \tag{2.47}$$

This quantity is inversely proportional to the error variance of the estimator $\Delta\hat{\mu}_c$. In what follows, we use T_{eff} in order to consider what would be an optimal allocation of computing time between the number of ensemble members M and integration time K, as well as to determine the impact of reducing the ensemble variance by SN and domain-size reduction on the error variance.

We first note that for $R_f = 0$, the equivalent sample size in Eq. (2.47) takes the form identical to Eq. (2.20). This happens when the control and modified model simulations employ the LBC and other external forcing that are uncorrelated in time, such as from two different CGCM simulations. In this special case, the error variance in the estimated time-mean model response $\Delta\hat{\mu}_c$ has identical behavior as the error variance of the single model's climatological mean $\hat{\mu}_c$. We will not pursue this case further, since it was already discussed in Section 2.2.5.

We rather focus on the case when R_f is positive. We now consider the optimal

allocation of resources between members and integration time. By assuming for the moment that K is continuous and computing the partial derivative of Eq. (2.47) in terms of K, it can be shown that T_{eff} is a monotonically increasing function of K. Therefore, the error variance of the estimate $\Delta\hat{\mu}_c$ can be minimized by selecting the largest possible simulations' integration time K, which implies the choice of single-member ensembles for the control and modified model, $M = 1$. This is the same conclusion as that for the estimation of the single-model time mean in Section 2.2.5.

Let us now focus particularly on the estimation of the time mean model response from two single-member ensembles, for which $M = 1$, i.e., $K = T$. In this case we obtain from Eq. (2.47) that

$$T_{\text{eff}} = \frac{T}{1 - R_f \cdot \rho^2}, \tag{2.48}$$

Note the difference between this result that addresses the estimation of the difference of means $\Delta\hat{\mu}_c$ and the corresponding result in the case of estimation of the RCM climatological mean $\hat{\mu}_c$, provided in Eq. (2.20). When $T = K$ we obtain from Eq. (2.20) that $T_{\text{eff}} = T$. On the other hand, it can be seen from Eq. (2.48) that for positive R_f and ρ^2, we obtain $T_{\text{eff}} > T$. Thus, given that the reproducibility ratio is not equal to zero, the effect of the reproducible correlation R_f on the effective sample size is to increase it and hence to reduce the error variance of the estimator $\Delta\hat{\mu}_c$.

It may appear surprising that the effective sample size can be larger than the nominal value, so we will now explain this result. If the reproducible correlation R_f is large, then the reproducible time anomalies f_1 and f_2 vary synchronously in time. According to the above assumptions, they also have similar variances $\sigma^2_{f_1}$ and $\sigma^2_{f_2}$. Hence, the difference Δf tends to be small. If, in addition, the reproducibility ratios ρ are large, then the ensemble deviations ε_1 and ε_2 are of small magnitude, as well as the difference $\Delta\varepsilon$. Therefore, the largest part of the difference $\Delta\psi = \Delta\mu_c + \Delta f + \Delta\varepsilon$ is then contained in the difference of means $\Delta\mu_c$ and, thus, the variance of $\Delta\psi$ is small, yielding also to a small variance of the sample mean difference $\Delta\hat{\mu}_c$. In other words, the externally forced, reproducible component of temporal variability is then largely subtracted out from the model response to modification, implying that the temporal variability of the response is smaller than the temporal variability of each of the two model outputs, from which the response is computed. Hence, given the same integration period, the error variance of the estimator $\Delta\hat{\mu}_c$ is smaller than the error variance of the estimator

$\hat{\mu}_c$, given that R_f is positive. This is then reflected in a larger effective sample size T_{eff} than the nominal size T.

Equation (2.48) shows that, given positive values of the reproducible correlation R_f between the control and modified model, T_{eff} increases (and hence the error variance of $\Delta\hat{\mu}_c$ decreases) with the reproducibility ratio ρ. This implies that the application of SN or reduction of domain size in both the control- and modified-model simulations, can decrease the error variance of $\Delta\hat{\mu}_c$ by increasing the reproducibility ratio of the RCM simulations. Note that this result is fundamentally different of that obtained for the estimated single-model climatological mean $\hat{\mu}_c$. As it can be seen in Eq. (2.20), for an ensemble consisting of a single member, i.e., $T = K$, the effective sample size is $T_{\text{eff}} = T$, so it does not depend on the reproducibility ratio. Furthermore, the time variance of a single RCM simulation should be constrained by the observations and, hence, the redistribution between the reproducible and irreproducible components in Eq. (2.11) should not have a considerable impact on the time variance. Therefore, when the estimate $\hat{\mu}_c$ of the RCM climatological mean from a single-member ensemble is considered, an increase of reproducibility ratio by SN or domain-size reduction should have little influence on its error variance in Eq. (2.19). On the other hand, given that the reproducible correlation R_f is positive, an increase in reproducibility ratio ρ allows for a reduction of the error variance of the time-mean model response $\Delta\hat{\mu}_c$ estimated from single-member ensembles.

The efficiency of SN and domain-size reduction in reducing the error variance depends on the reproducible correlation R_f. It can be seen in Eq. (2.48) that if $R_f \approx 0$, increasing ρ does not yield an increase in the effective sample size T_{eff}. On the other hand, the largest increase in T_{eff} for a given increase in ρ is obtained when $R_f \approx 1$. The reproducible correlation R_f, as discussed earlier, can be maximized by performing the control and modified-model simulations using identical lateral boundary and surface forcing.

The previous considerations point to the fact that the application of SN and domain size reduction can reduce the error variance of the estimated mean model response to modification $\Delta\hat{\mu}_c$ and, thus, help to increase the statistical significances of this estimate. But these methods may have diverse effects on the RCM simulations and may

result in alterations of model response not necessarily restricted to the redistribution between the reproducible and irreproducible components. For example, they can change the mean model response and increase the reproducible correlation R_f, thus reducing the temporal variability of the reproducible component of the model response to modification. Thus, the application of nudging or reduction of domain size with the sole purpose of reducing irreproducible deviations should be considered carefuly.

In order to complete the discussion of the estimation of the difference of means, we also need to consider the RCM difference of means in the case of downscaled anomaly for a specified period, such as a season of a year. In this case what is estimated is the time-average difference of ensemble means, for a given period of time K_0. Analogously to Eq. (2.21) we have:

$$Var\left[\Delta\overline{\mu_e}\right] = \frac{\overline{\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2}^{K_0}}{MK_0}. \tag{2.49}$$

Since the integration period K_0 is now fixed, the only way to reduce the error variance is to increase the ensemble size. Evidently, SN and domain size reduction can decrease the variance given in Eq. (2.49) by reducing the time-average inter-member variance σ_ε^2 in each of the two models.

2.4 Some examples

In this section we will illustrate the previous considerations by performing the analysis of the response of a RCM to a perturbation of parameter.

2.4.1 Model and experiments

The model used in this study is the fifth-generation Canadian Regional Climate Model (CRCM5; Zadra et al. 2008). The model is described in Section 1.2.1 of this thesis. Two sets of experiments, all based on multi-annual CRCM5 simulations, are carried out.

In the first set, denoted as MYNA (multi-year North America), four 10-year sim-

ulations are performed over a continental-scale domain, consisting of 120^2 grid points, covering most of North America. The integration domain, including the 10-point relaxation zone at the perimeter of the lateral boundaries, is shown in Fig. 1.1 (this domain is used in the SYNA experiment, described in Chapter 1). Two CRCM5 versions are used to carry out simulations: the control version (M00) with the standard settings and a perturbed-parameter version (M01), obtained by perturbing the threshold vertical velocity in the trigger function of the deep convection parameterization (Kain and Fritsch 1990). The values of the threshold vertical velocity parameter used in these two model versions are shown in Table 2.1. The parameter perturbation used here is identical to the perturbation P01 in the single-year experiments, discussed in Chapter 1 (see Table 1.1). For each of the two model versions, two 10-year simulations with perturbed initial conditions were performed. The first run is initialized on November 01 1992 at 00UTC and the other starts 24 h apart. They both end on December 01 2002 at 00UTC. This setup provides a two-member ensemble of 10-year simulations per each model version, which yields an integration time of 40 years in total in the MYNA set. All four MYNA simulations share the same lateral boundary conditions, derived from ERA40 reanalysis (Uppala et al. 2005) and ocean surface prescribed from Atmospheric Model Intercomparison Project data (AMIP; Taylor et al. 2000).

The second set of experiments, denoted as MYSN (multi-year spectral nudging), is identical to MYNA set in all terms including the integration domain (Fig. 1.1), model versions and number of ensemble members per model version; the only difference is that the SN was used. Nudging was only applied to the horizontal wind components, with the truncation at non-dimensional wavenumber 4 (1,500 km). The SN configuration is identical to that in the SYSN experiments presented in Chapter 1. It is applied only at the upper half of the model's atmosphere and increases linearly with height, from zero at 500 hPa to 10% of the amplitude of the driving fields per time step at the top level. The present choices of the truncation wavelength and vertical nudging profile reflect our intention to interfere as little as possible with the model own interior dynamics at fine and intermediate spatial scales and in the lower half of the models atmosphere.

The reduced domain size experiments, such as the SYDS in Chapter 1, are not performed because the small domain produced some significant alterations of model response, unlike the SN in the single-year experiments. It is important to note that, unlike in Chapter 1 where we analyzed the seasonal averages, here we will examine the CRCM5

time series of daily-average variables. For simplicity, in what follows we will examine only summer (JJA) daily-average CRCM5 2 m-temperatures. For this purpose, for every simulation, JJA daily-average 2 m-temperatures are gathered in a single time series for the period 1993-2002. The intention here is to examine the transient CRCM5 response to the deep-convection parameter perturbation, rather than the time-average response. We first consider the reproducible and irreproducible components of the temporal variability in the control model. Then we analyze the difference between the control and parameter-perturbed model. The effects of SN on that difference may be manifold; here we focus mainly on the reduction of internal variability in the large-scale nudged simulations in order to provide some illustration for the previous theoretical considerations.

2.4.2 CRCM5 reproducible and irreproducible components

The reproducible and irreproducible components of the time variance of the control CRCM5 version are estimated using the transient difference between its two ensemble members, as specified in Section 2.3.3. Evidently, to find the time variance of an individual ensemble member we could use either of the two members of the ensemble; however, this yield little difference in the final result (not shown) implying a small sampling error in the time variance of 10-year integrations. The results of this decomposition are displayed in Fig. 2.2, as the corresponding standard deviations (std) – σ_t, σ_ε and σ_f, for the non-nudged MYNA experiment (a, b, c) and nudged MYSN simulations (d, e, f) for the CRCM5 2 m-temperatures.

Let us begin with the no-nudging case. The MYNA time std (Fig. 2.2a) exhibits strong spatial variations and is in general much larger over land than over the ocean. Over the continent it takes values of roughly $4-7\,^{\circ}$C with the maximum in the vicinity of the Arctic Ocean coastal regions. Over the most southern part of North America it is smaller and takes values of $2-4\,^{\circ}$C. Its reproducible component (Fig. 2.2b) follows a very similar pattern as the total but it has somewhat smaller values, since a part of the temporal variability is provided in the irreproducible form. It is also worth noting that we display std rather than the variances; only the latter are additive. In any case, the irreproducible component of the time std (Fig. 2.2c) exhibits a notably smaller magnitude than the reproducible components, when the model's 2 m-temperatures are considered. The irreproducible std almost vanishes over the ocean and takes values of

only $0 - 3\,^{\circ}C$ over land. It also vanishes at the perimeter of the lateral boundaries because of the LBC constraint.

In the MYSN experiment (Fig. 2.2d-f), these general properties remain largely the same, with the exception of the irreproducible std (Fig. 2.2f) that is quite a bit smaller than in the no-nudging case (Fig. 2.2c) with values not larger than $2\,^{\circ}C$. This is the consequence of the reduction of inter-member spread in the nudged simulations. Comparison of Figs. 2.2a, b and Figs. 2.2d, e also shows differences, such as that the reproducible component is somewhat larger in the nudging case (Fig. 2.2e). SN produces three main effects: (i) it changes somewhat the total, time variance, (ii) it decreases notably the irreproducible components, and (iii) it transfers time variability from the irreproducible to the reproducible components. To enhance the readability of these results, we present Figs. 2.3 and 2.4.

Figure 2.3 displays the ratio of the time std between the control CRCM5 simulations in the nudging and no-nudging setups. It can be seen that the application of nudging yields a considerable reduction of time standard deviation over a large area adjacent to the Great Lakes and also over Northern Mexico with values as low as 80% of those in the no-nudging simulation. In other regions it remains unchanged or slightly increased by a factor of up to 1.15. To understand the causes of these differences would require a more thorough study of the CRCM5 skill in reproducing surface air temperatures and other climate variables, which is beyond the scope of our present work. We rather intend to illustrate the variations of the time std between different simulations configurations. It is worth noting that the fact that the time variance changes in the case of SN goes against our conjecture of the compensation between the reproducible and irreproducible components (Eq. 2.11). If there were no change in time variance, there would be a full compensation between the two components; a decrease in the reproducible variance would be compensated by an equal increase in the reproducible variance in the SN case. Since the time variance does change somewhat, there is still some compensation, but it is not full, since the sum of the two components (which is equal to the time variance) is not identical in MYNA and MYSN sets.

The effect of SN on the redistribution between the reproducible and irreproducible components is shown in Fig. 2.4, where we display the reproducibility ratio ρ in per-

centage, defined in Eq. (2.14), in the control CRCM5 version for the MYNA (Fig. 2.4a) and MYSN simulations (Fig 2.4b). These figures show the portion of the reproducible components in the time std. It can be seen that at the perimeter of the lateral boundaries reproducibility ratio tends to unity, implying that the daily averages of 2 m-temperature are entirely reproducible there. When the SN is not applied (Fig. 2.4a), roughly $85 - 95\%$ of the time standard deviation is provided in reproducible form over the northern half of the continent. This ratio generally decreases towards the south, where it locally takes values as low as 65%. It is worth noting that in absolute terms (Fig. 2.2c), the irreproducible standard deviation is not large in these areas; it is large in relative terms because the total variability is low there, which results in low values of the reproducibility ratio over the southern part of the continent. It is also worth noting that a similar analysis for daily CRCM5 precipitation results in a much smaller reproducibility ratio (not shown). The reason for selecting 2 m-temperatures for the present analysis is their high reproducibility ratio, or equivalently, small ensemble variance with respect to their time variance, which contributes to the robustness of the results displayed. In the case of precipitation, the integration period of 10 years is still too short for a robust estimation of reproducibility ratio and other statistics and this is the main reason for not showing it.

When SN is applied (Fig. 2.4b) the reproducibility ratio generally increases to values between 90 and 100 % over the largest part of the domain and 80-90 % only over the U.S. southwest and southeast. This implies that a large portion of the variability in daily time series of the control model is provided in the reproducible form due to the reduction of inter-member variance. This result confirms previous findings (e.g., Weisse and Feser 2003, Alexandru et al. 2009).

The above analysis is conducted for the control version of the CRCM5. The results of an analogous analysis for the perturbed-parameter model version is not displayed, because both the reproducibility ratio and the ratio of std obtained in the perturbed-parameter CRCM5 have values very similar to those shown in Figs. 2.3 and 2.4, for the case of 2 m-temperatures.

2.4.3 Response of the CRCM5 time-average to deep-convection parameter perturbation

We now turn our attention to the time series of the difference between the daily-average 2 m-temperatures in the model with perturbed deep-convection parameter and the control model. We begin the analysis by examining the response of the 10-year average summer (JJA) 2 m-temperatures to the deep-convection parameter perturbation. The averaging is performed in time (over ten years) and over the two members of the ensemble for each model version; the average difference between the perturbed and control CRCM5 is then computed. This will be referred to as the signal. Fig. 2.5a shows the signal in the MYNA experiment. It can be seen that the parameter perturbation produced a considerable temperature increase over almost the entire continent of the order of $1-2\,^{\circ}\mathrm{C}$ with the largest values of up to $2.6\,^{\circ}\mathrm{C}$ over the Prairies. Over the ocean, 2 m-temperatures are well constrained by the prescribed sea surface condition and there is little or no response. Similarly, over the Arctic there is little response to perturbation. In order to examine whether these differences are triggered by the deep-convection parameter perturbation or represent the residuals of inter-member differences in the time average (noise), the difference of 10-year means between the two members of the control model in the MYNA set is displayed in Fig. 2.5b. It can be seen that the noise generally produced much smaller differences of means than the parameter perturbation, with values of up to $0.7\,^{\circ}\mathrm{C}$. The exception is the eastern-most parts of North America in Fig. 2.5a, where the parameter perturbation produced smaller responses that may be contaminated with noise; the noise can act both ways, either to increase or decrease the perceived signal.

Next, we consider the signal and noise in the MYSN set (Figs. 2.5c and 2.5d). It can be seen that in the SN case the deep-convection parameter perturbation (Fig. 2.5c) produced the response of a somewhat smaller magnitude over the northern half of the continent than in the no-nudging case (Fig. 2.5a) . Over the southern areas both the magnitude and patterns appear to be similar with and without nudging. Finally, the noise in the nudging case is assessed as the difference of means of the two control model ensemble members, (Fig. 2.5d). It is smaller than its no-nudging counterpart shown in Fig. 2.5b, implying that inter-member variability is somewhat reduced by SN.

Finally, Fig. 2.5e displays the difference between the signals in the MYSN (Fig.

2.5c) and MYNA sets (Fig. 2.5a). The difference is displayed only where it is statistically significant at 95 % or higher levels. It can be seen that in the MYSN experiment, the parameter perturbation generally produced an increase in the mean 2 m-temperature that is smaller by $0.2 - 1.0\,°C$ than in the no-nudging MYNA set, over the large part of the continent. This implies that SN reduces the sensitivity of the CRCM5 time average to the perturbation of the deep-convection parameter; this reduction of sensitivity is also statistically significant. It is interesting to compare Fig. 2.5e with the corresponding result obtained in the single-year study with the same parameter perturbation (Chapter 2). Fig. 1.8g displays the statistical significance of the difference between the signals induced by the same parameter perturbation in the SYSN (Fig. 1.8b) and SYNA sets (Fig. 1.8a). It can be seen in Fig. 1.8g that the regions of statistical significance are small and scattered. The fact that in the present multi-year setup we obtained statistically significant difference between the sensitivities is also the consequence of a larger sample size: we recall that in the SYNA and SYSN experiments the total computing time was 30 years, while the MYNA and MYSN experiments provide 80 simulated years in total.

2.4.4 CRCM5 transient response to deep-convection parameter perturbation

We begin the considerations of the instantaneous difference between the two CRCM5 versions by examining the time variance of the difference between two models, $\sigma^2_{t\Delta}$, given in Eq. (2.38), i.e. the time variance of the model response to the parameter perturbation. The reproducible and irreproducible components of this variance, given in Eqs. (2.39) and (2.40), respectively, are estimated as described in Section 2.3.3. We do not examine the statistical significance of these estimates since the purpose of this section is to illustrate the typical values of various parameters and their relative magnitudes, for the case of model response to perturbations of physics parameters.

The results of these computations for the time series of daily-averaged 2 m-temperatures are presented in Fig. 2.6, as the std ($\sigma_{t\Delta}$, $\sigma_{\Delta f}$ and $\sigma_{\Delta\varepsilon}$) for the MYNA (a-c) and MYSN (d-f) sets, respectively. It can be seen that for both non-nudged and nudged simulations, the irreproducible component of the std of model response (Figs. 2.6 c and f) is now larger than the reproducible component (Figs. 2.6 b and e); this is the opposite of what we obtained when the time variance of the control model output

was decomposed (Fig. 2.2), in which case the reproducible component was dominant.

In addition, it can be seen in Figs. 2.6 a and d that SN considerably reduced the time std of the response. This is much different result from that obtained in Fig. 2.2a and 2.2d that quantified the time std of the model output itself. With no nudging (Fig. 2.6a) the time std of the response is $3-4\,^{\circ}\mathrm{C}$ over most of the continent, while in the nudging case (Fig. 2.6d) it is systematically reduced to values of $1.5-3\,^{\circ}\mathrm{C}$, implying a reduction of up to 50%. The reproducible component of the time std of the response (Figs. 2.6b, e) is small with respect to the total (Figs. 2.6a, d). It takes values of $0-1.5\,^{\circ}\mathrm{C}$ over land and locally up to $2\,^{\circ}\mathrm{C}$ over the southwestern parts of the continent. The SN (Fig. 2.6e) appears to produce little change in the reproducible component of model response when compared to the no-nudging case (Fig. 2.6b).

Note that the reproducible component of the time std represents an estimate of the time std of the difference between the parameter-perturbed and control models' unknown time-dependent ensemble means, $\Delta f = f_2 - f_1$. Thus, it can be thought of as an estimate of the "true" temporal variability of the model response to deep-convection parameter perturbation, obtained when the chaotic variations measured by the irreproducible std (Fig. 2.6c, f) are not accounted for. To put it simply, the reproducible std (Figs. 2.6b, e) quantifies the time variability in the signal, while the irreproducible std of the difference (Figs. 2.6c, f) quantifies the std of noise triggered by chaotic variations ε. Results from Fig. 2.6 do not imply that the signal is small, but rather that the time variability in the signal is small. The signal and noise magnitude can be quantified by the difference of 10-year time averages that are displayed in Fig. 2.5. There it can be seen that the signal-to-noise ratio is actually quite larger than unity. In addition, comparison of Figs. 2.6b (2.6e) with Figs. 2.5a (2.5c) shows that the magnitude of the signal is generally larger than its time std in both non-nudging (nudging) case, respectively. In other words, the deep-convection perturbation produces an increase in daily-average 2 m-temperatures that tends to be rather constant in time. The change due to the parameter perturbation is mainly contained in the change of the time mean μ_c (see Eq. 2.5). Other than this "bias", there is little change in daily-average values due to the parameter perturbation; transient differences between the two model versions are mainly due to the noise generated by their irreproducible time anomalies ε_1 and ε_2. This implies a large time correlation between the externally forced, reproducible anomalies f_1 and f_2 in the two models, which yields a small time variance of the difference

$\Delta f = f_2 - f_1$.

In order to study the CRCM5 2 m-temperature transient response to the parameter perturbation in more detail we will now examine the set of parameters $\{\sigma_{t1}, \sigma_{t2}, R_f, \rho_1, \rho_2\}$. Recall that the first three parameters quantify the contribution of the modification to the response and the latter two quantify the contribution of the chaotic nature of the two models.

Figure 2.7 displays the ratio of time std σ_{t2} and σ_{t1} in the perturbed-parameter and control CRCM5 2 m -temperatures, in the MYNA setup. It can be seen that the parameter perturbation yields an increase of time std over most of the continent, with values of the ratio up to 1.15 over the central part of the continent and as large as 1.25 over the most southeastern regions. The only exception to the general increase in time std are the southwestern U.S. and Mexico where the perturbation produced a decrease in time std to as low as 80% of the values obtained in the control CRCM5. It is difficult to understand the sources of these differences without a thorough study of other CRCM5 variables, which is beyond the scope of this manuscript. We rather intend to show that the temporal variability in a RCM can be considerably changed by the parameter modification, and thus it should be considered in studies of parameter perturbations. The ratio of time std in the MYSN set is very similar to that in MYNA (not shown). This implies that there is little interaction between the parameter perturbation and SN when the ratio of 2 m-temperature time std is considered.

Next we consider the decomposition of the time correlation R into its reproducible part R_f and irreproducible part $\rho_1 \times \rho_2$, following Eq. (2.34). These parameters are estimated following the method discussed in Section 2.3.3.

For the no-nudging MYNA simulations set, the results are displayed in Fig. 2.8a-c. Figure 2.8a displays the time correlation R between the control and perturbed-parameter simulations. It tends to 1 at the perimeter of the domain where the model solution is constrained by the LBC, as well as over the oceans. There are however exceptions such as near the vicinity of the southern boundary. Over land, the correlation R is smaller but still considerable: over the northern half of the continent it has values between 0.75 and 0.85 with values, generally larger towards the western domain

boundary. The lowest temporal correlation between the control and modified model simulations is over the southern half of the continent, where the typical values are between 0.5 and 0.7 with the lowest values of 0.4 obtained over two separated areas (New Mexico and Georgia). Figure 2.8b shows the product of the reproducibility ratios (Eq. 2.14) in the two CRCM5 versions in the MYNA set. This quantity can be interpreted as de-correlation of two simulations due to irreproducible deviations ε. This means that if there was no modification in the model, except, for example, a small perturbation in the initial conditions, then the correlation R in Fig. 2.8a would be approximately equal to the product of reproducibility ratios in Fig. 2.8b. The patterns in Figs. 2.8a and 2.8b are very similar, implying that most of the de-correlation between the control and modified model simulations originates in the irreproducible deviations and not in the modification of the model. Finally, Fig. 2.8c displays the estimated time correlation R_f between the reproducible time anomalies of the control and modified model. The plots in Fig. 2.8c show that the time correlation between the forced responses in the two models is very high, almost equal to one, over most of the domain. The main exception are the values of 0.7-0.8 over the southwestern US and Mexico. These lower values imply that the forced, reproducible part of the model solution is the least controlled by the external forcing in these regions. On the other hand, generally high values of R_f also imply that the reproducible time deviations in the control and modified models appear to be very well synchronized. Note that a more substantial model modification could yield different conclusions.

Figures 2.8d-f display the results for the MYSN simulation setup. The time correlation R between the simulations of the control and modified model with SN (Fig. 2.8d) is higher than without nudging (Fig. 2.8a). The values with nudging are generally about $0.1 - 0.2$ higher while the overall patterns are similar. Further, comparison of the product of reproducibility ratios (Fig. 2.8e) and correlation between the reproducible time deviations of two models (Fig. 2.8f) shows that the change in the total correlation is mainly due to the increase in reproducibility ratio, that is, it is due to the fact that SN reduced internal variability. The correlation between the reproducible components R_f is not much affected by SN (Figs. 2.8c and 2.8f).

In conclusion, Fig. 2.8 shows that the irreproducible anomalies in the CRCM5 simulations appear to be the dominant factor in reducing the correlation between transient 2 m-temperatures, simulated by the control and modified model. SN helps to

increase the time correlation between the two model versions because it decreases the variance of the irreproducible deviations, as shown in Fig. 2.2. At the same time, the model reproducible time anomalies in the two model versions remain highly correlated, in the both spectrally nudged and non-nudged experiments. In addition, there is no evidence that the SN considerably changed the reproducible correlation R_f.

2.5 Summary and discussion

The purpose of this manuscript was to develop a general theoretical framework for the estimation of first- and second-moment statistics of RCM response to modification. The approach followed consists in decomposing the climate variables simulated by an ensemble of RCM integrations into the climatological mean and two components of the time anomaly: (1) the reproducible anomaly enforced by the LBC and ocean surface that represents the time variations of the ensemble mean with respect to the climatological mean and (2) the irreproducible anomaly that represents the spread of ensemble members around the ensemble mean. Theoretical considerations showed that the sum of the variances of these two anomalies represent an observable quantity equal to the time variance of individual RCM simulations in the limit of long integration periods. Since the sum of the reproducible and irreproducible variances is an observable, ideally, whenever the RCM simulation configuration is modified, a change in one of these variances should be compensated by the change of an equal amount but of the opposite sign in the other one. In practice, RCMs do not perfectly reproduce observed temporal variances, and modification in the simulation setup, such as a different domain size or nesting technique, as well as any other modification of the model, may also change somewhat the time variance. However, if the simulated time variance were too different from the observed one, this would indicate a serious problem with the model or simulation configuration.

Thus, in practice, any reasonable modification to the model should not change the sum of the reproducible and irreproducible variances but tend to redistribute the variance between the two components. Apart from this consideration, there are no other general expectations about the relative magnitude of the two components of the time variance, although their relative magnitude in RCMs can be controlled by the SN and

domain size. However, the partition of the variance into the reproducible and irreproducible components may have important implications on the estimation of the RCM time mean as well as on the estimation of the RCM sensitivity to modifications. Selection of an optimal integration period and ensemble size for RCM experiments should also take into account the redistribution between the reproducible components, which can be controlled by the selection of domain size and SN configuration. The purpose of this manuscript was to provide a detailed discussion of this issue and eventually help the practitioners to select an optimal setup for quantifying RCM response to parameter modifications.

The theoretical considerations showed that the error variance in the RCM temporal mean, as estimated from an ensemble of RCM integrations, can also be represented as the contributions of the reproducible and irreproducible components. Of these two parts, only the latter can be reduced equivalently by increasing the number of members or increasing the integration time. The reproducible part of the error variance of the sample mean is irreducible by means of increasing the number of ensemble members. Thus, in order to minimize the error variance of the estimated RCM time mean, there is no other choice than to allocate a larger part of the available computing resources to the integration time, and optimally the ensemble should consist of a single member. An attempt to reduce the ensemble spread by SN or reduction of the size of the computational domain is likely to increase the error variance in the estimated RCM time mean. This happens, because as the ensemble spread is reduced, ensemble members are better constrained by the LBC and less efficient in sampling natural variability of the system so the variance of the estimated mean increases. Of course, this becomes irrelevant when the entire computing time is allocated to a single simulation.

In the case of simulating a particular weather event or a given period, such as a season of a single year, the climatological mean is irrelevant. What is typically estimated is the ensemble mean and deviations with respect to the ensemble mean. The error variance in the RCM ensemble mean is contributed solely by the irreproducible components. In this case, the optimal choice for the simulation setup is to maximize the number of ensemble members. If the inter-member spread is large, SN and a small computational domain do contribute to reducing the error variance in the estimated ensemble mean.

In general, the variance of the RCM response to modification can be described with five parameters: reproducibility ratios $\rho_{1,2}$ in the control and modified model, models' time variances $\sigma^2_{t1,2}$ and time correlation between the reproducible anomalies R_f. The reproducibility ratios quantify the contribution of the quasi-random, chaotic variations triggered by the modification, while the other three parameters quantify the deterministic component of the response. We showed that these parameters can be estimated from two ensemble members per model version, given sufficiently long integration times. The parameter R_f depends to some extent on the selection of the LBC and other external forcing. If the LBC and surface forcing are not correlated in time, then R_f must vanish, which is the case when RCM simulations are driven by different CGCM simulations. The estimation of the parameter R_f is meaningful only when the reproducibilities are relatively high. Low reproducibility ratios mean that a small amount of time variance is partitioned by the reproducible anomalies, which implies that R_f explains only a small amount of time variance of the difference between two simulations, thus having little physical significance.

A decrease of the irreproducible noise as a consequence of SN or domain-size reduction cannot decrease the error variance in the estimated RCM time mean. However, testing the model to modification in a simulation setup with SN or in a small computational domain can reduce the error variance of the estimated difference of time means between two models. Whether this happens or not depends on the reproducible correlation R_f and the number of ensemble members. As in the case of the estimation of the single model time mean, an optimal setup that minimizes the error variance of the estimated difference of time means is obtained by allocating the entire available computing resources to the integration period (i.e., the ensemble should consist of a single member). The efficiency of SN and domain-size reduction in reducing the error variance is proportional to the reproducible correlation R_f. If the models share the same LBC, the correlation R_f may be large, thus allowing for reducing the error variance.

The present experiments with the CRCM5 simulations with perturbed deep-convection parameter showed that, in the case of temperature, this parameter can be very close to unity in the largest part of the computational domain for modest modifications, such as perturbations in the physics parameter. This implies that SN is very efficient in reducing the error variance in the estimates of the difference of temperature time means. However, SN appears to reduce somewhat the estimated sensitivity of the

CRCM5 time JJA mean 2-m temperatures to the deep-convection parameter perturbation. It is also worth noting that the integration time of 10 years in our simulations appears to be too short for an analogous analysis of precipitation.

In summary, using SN can reduce the error variance of the difference of time means between two models. Consequently, the computational cost of the estimation of the difference of models' means can be also reduced by SN. But whether the error variance of the difference of means will be decreased, depends on the nature of modification to the model, as well as on the input data. The error variance reduction is more likely for minor modifications such as perturbing RCM parameters' settings. As of the input data, in order to reduce the error variance it is necessary that the driving fields (LBC) in the control and modified model simulations are the same or at least correlated in time. Using a single set of LBC is a common sense approach in studies of RCM sensitivity to code and parameter modifications. These considerations foster the use of SN in studies of sensitivity of the RCM-simulated climatological mean to parameter modifications. However, caution is necessary with the use of SN, since the results imply that, beside reducing the error variance of the mean response, SN can change the reproducible, deterministic part of the model response to parameter modification.

2.6 Appendix: Variance of the time-ensemble mean

In this section we derive the variance of the ensemble and time average of a sample that consists of KM independent realizations of the model variable ψ_{km}, sampled from M different realizations at each of K times. First we define the operators E_t and E_e as the arithmetic averages of ψ_{km} in the limit $K \to \infty$ and $M \to \infty$, respectively. Next, we define the overall and time means of ψ, respectively, as follows:

$$\mu_c = E_t\big[E_e[\psi]\big]. \tag{2.50}$$

$$f_k = E_e[\psi] - E_t\big[E_e[\psi]\big]. \tag{2.51}$$

Hence, the variable ψ can be decomposed as:

$$\psi_{km} = \mu_c + f_k + \varepsilon_{km}. \tag{2.52}$$

where f and ε are new random variables, such that f is a function of time and ε of both time and member. They satisfy the following condition:

$$E_t[f] = E_e[\varepsilon] = 0. \tag{2.53}$$

The sample mean of KM realizations of variable ψ is given as

$$\hat{\mu}_c = \frac{1}{KM} \sum_{k=1}^{K} \sum_{m=1}^{M} \psi_{km}. \tag{2.54}$$

Upon substituting Eq. (2.52) in Eq. (2.54) we obtain:

$$\hat{\mu}_c = \mu_c + \frac{1}{K} \sum_{k=1}^{K} f_k + \frac{1}{KM} \sum_{k=1}^{K} \sum_{m=1}^{M} \varepsilon_{km}. \tag{2.55}$$

By definition, the variance of $\hat{\mu}_c$ is given as

$$Var[\hat{\mu}_c] = E\Big[\big(\hat{\mu}_c - \mu_c\big)^2\Big]. \tag{2.56}$$

where $E[\cdot] \equiv E_t[E_e[\cdot]]$. From Eqs. (2.55) and (2.56) we obtain:

$$Var[\hat{\mu}_c] = \frac{1}{K^2}\left[\sum_{i,k=1}^{K} E[f_i f_k] + \frac{2}{M}\sum_{i=1}^{K}\sum_{k,m=1}^{M} E[f_i \varepsilon_{km}] + \frac{1}{M^2}\sum_{i,k=1}^{K}\sum_{j,m=1}^{M} E[\varepsilon_{ij}\varepsilon_{km}]\right]. \tag{2.57}$$

From Eqs. (2.51) and (2.53) we find that

$$E[f_i \varepsilon_{km}] = E_t\big[f_k \cdot E_e[\varepsilon_{km}]\big] = 0 \tag{2.58}$$

and from the assumption that the data are independent we have:

$$E_t[f_i f_k] = E_e[\varepsilon_{ij}\varepsilon_{km}] = 0, \quad i, j \neq k, m. \tag{2.59}$$

Therefore, Eq. (2.57) becomes:

$$Var[\hat{\mu}_c] = \frac{1}{K^2}\left[\sum_{k=1}^{K} E_t[f_k^2] + \frac{1}{M^2}\sum_{k=1}^{K}\sum_{m=1}^{M} E[\varepsilon_{km}^2]\right]. \tag{2.60}$$

Since, according to Eq. (2.53), the time mean of variable f is equal to zero, its variance

is defined as

$$\sigma_f^2 = E_t[f^2]. \tag{2.61}$$

Similarly, the ensemble variance of ε at time k is given as

$$\sigma_e^2 = \sigma_e^2(k) = E_e[\varepsilon_{km}^2], \tag{2.62}$$

and its time mean as

$$\sigma_\varepsilon^2 = E_t[\sigma_e^2]. \tag{2.63}$$

With the help of these definitions, the variance of $\hat{\mu}_c$ can be written as

$$Var[\hat{\mu}_c] = \frac{1}{K}\left(\sigma_f^2 + \frac{1}{M}\sigma_\varepsilon^2\right). \tag{2.64}$$

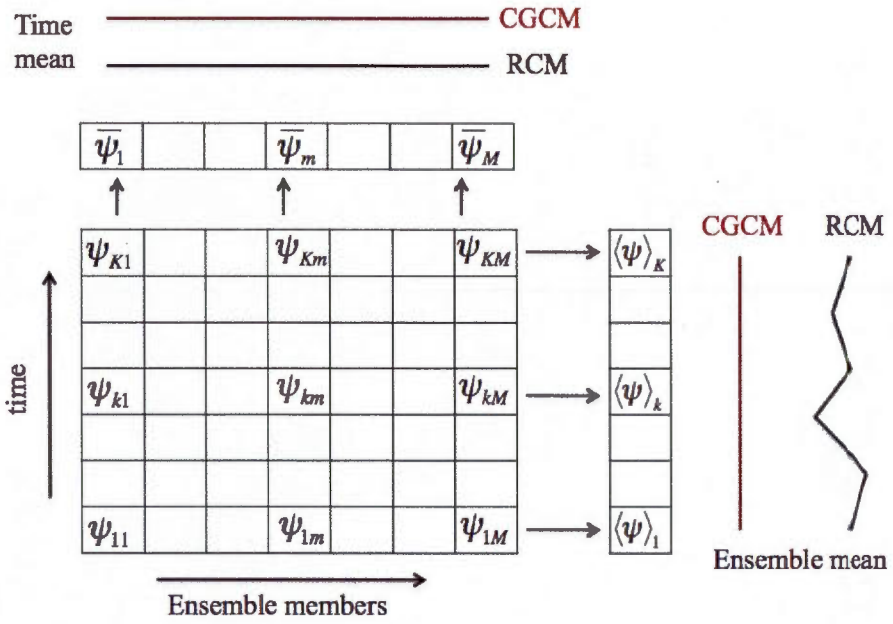

Figure 2.1 Illustration of the time series obtained from an ensemble of climate model simulations. Columns stand for time series obtained from individual members and may represent daily or seasonal averages. Rows represent the realizations obtained from ensemble members at a given time. The only difference between members is in the initial conditions. The lines on the top and right illustrate the typical behaviour of the time mean and ensemble mean, respectively. The black lines are for the case of RCMs and GCMs and red lines represent CGCMs. Note that for the case of ensemble mean the annual cycle is neglected.

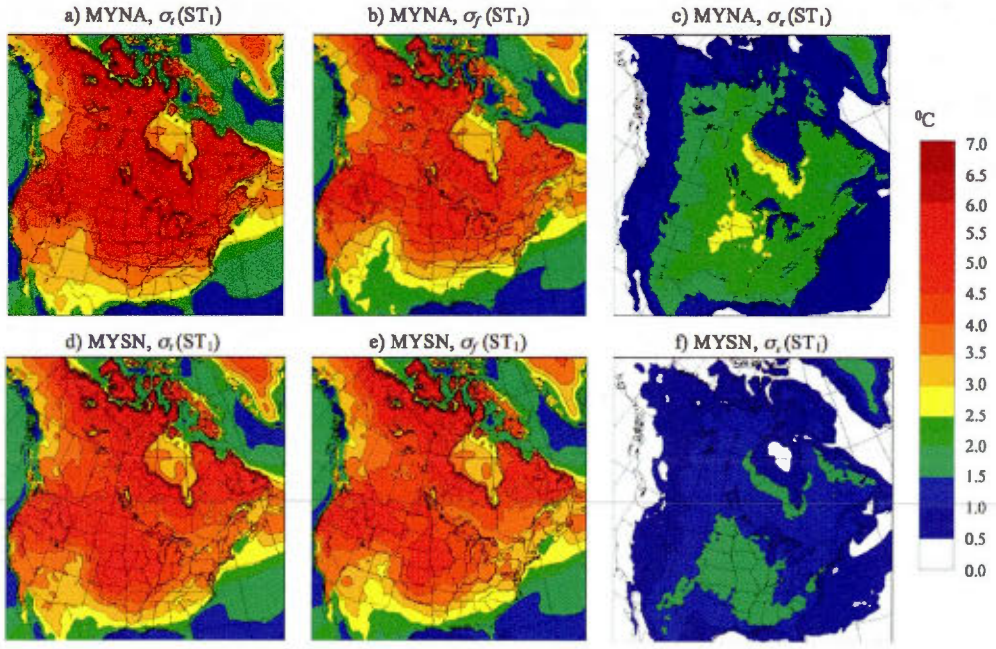

Figure 2.2 Decomposition of the time standard deviation of the control CRCM5 JJA 2 m-temperatures into the reproducible and irreproducible components (Eq. 2.11): time standard deviation σ_t, the reproducible component σ_f and the irreproducible component σ_ε; (a, b, c) MYNA and (d, e, f) MYSN experiment.

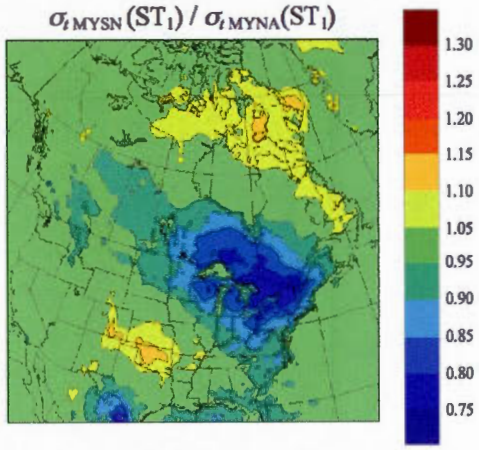

Figure 2.3 Ratio of time standard deviations σ_t of JJA 2 m-temperatures between the control MYSN and control MYNA simulations.

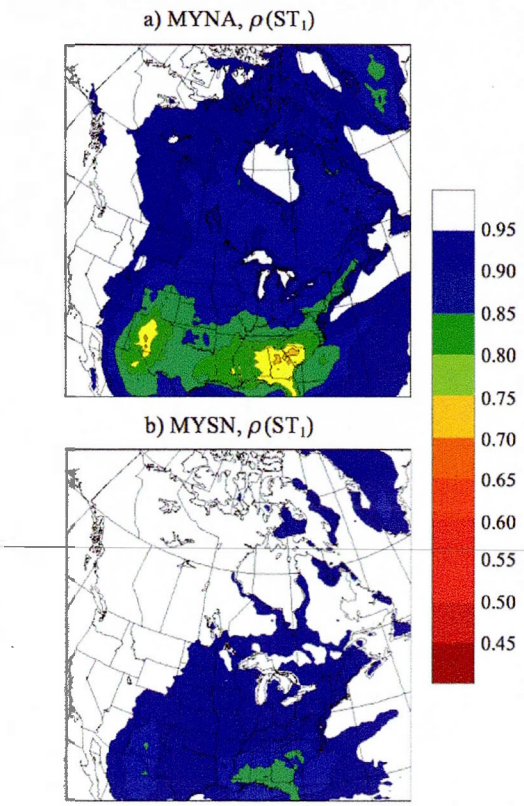

Figure 2.4 Reproducibility ratio of the JJA 2 m-temperatures (Eq. 2.14) in the control CRCM5: (a) MYNA, (b) MYSN.

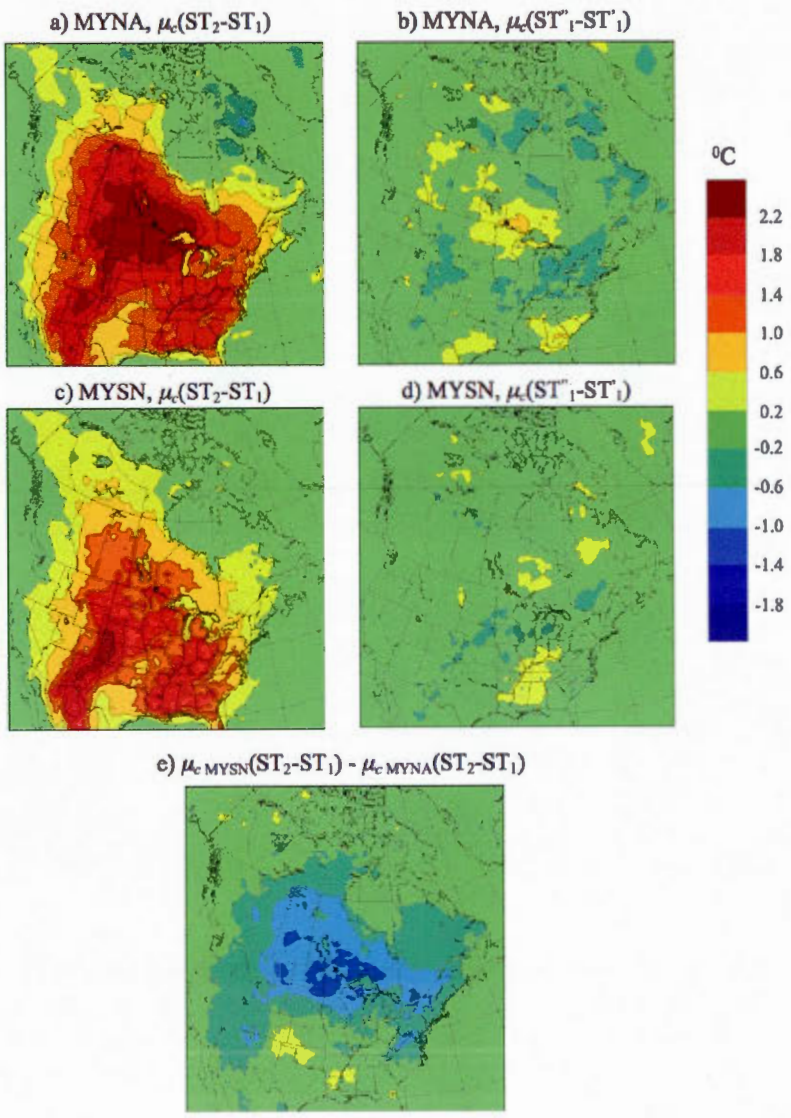

Figure 2.5 Difference of 1993-2002 JJA-average CRCM5 2 m-temperatures in the MYNA experiment, between (a) the parameter-perturbed and control CRCM5 model and (b) between two ensemble members of the control model; (c, d) the same as in (a, b), but for the MYSN experiment. Also shown is the difference between the responses to the parameter perturbation in the MYSN and MYNA sets (e).

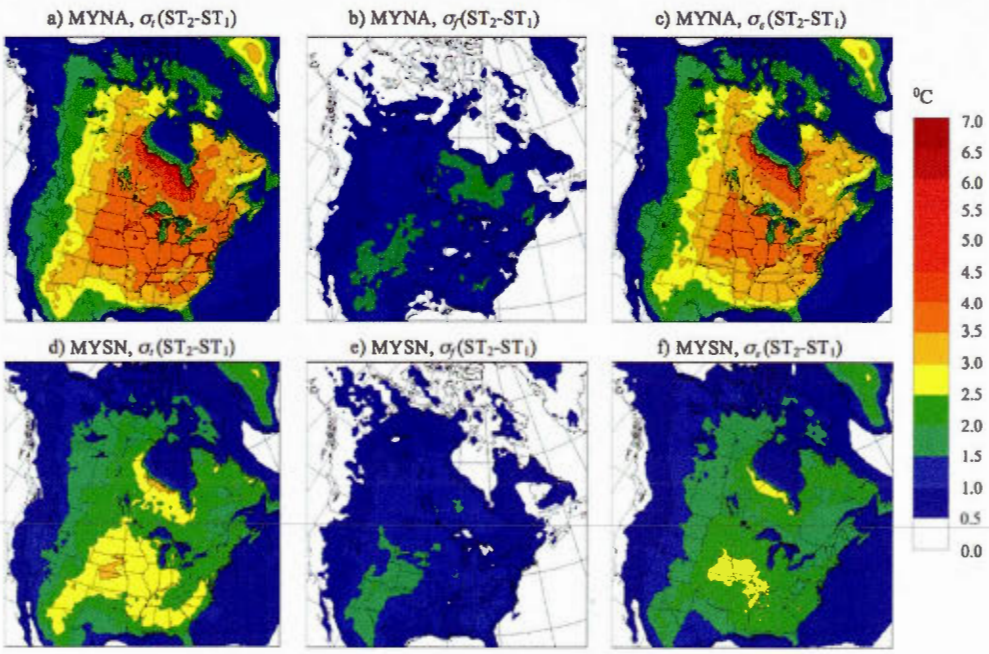

Figure 2.6 Decomposition of the time standard deviation of the difference between the parameter-perturbed and control CRCM5 JJA 2 m-temperatures into the reproducible and irreproducible parts (see Eq. 2.38): standard deviation of the difference $\sigma_{t\Delta}$, its reproducible component $\sigma_{\Delta f}$ and irreproducible component $\sigma_{\Delta\varepsilon}$; (a, b, c) MYNA and (d, e, f) MYSN experiment.

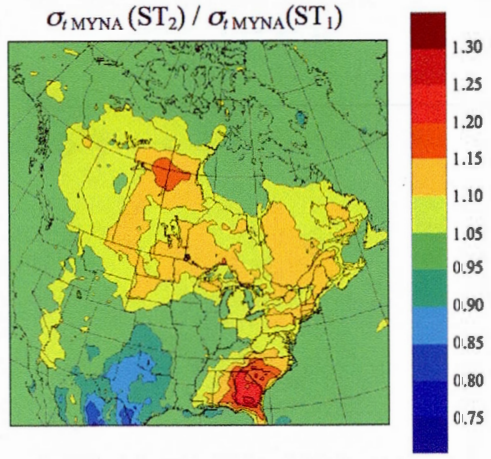

Figure 2.7 Ratio of time standard deviations σ_t of JJA 2 m-temperatures between parameter-perturbed and control CRCM5 runs in the MYNA experiment.

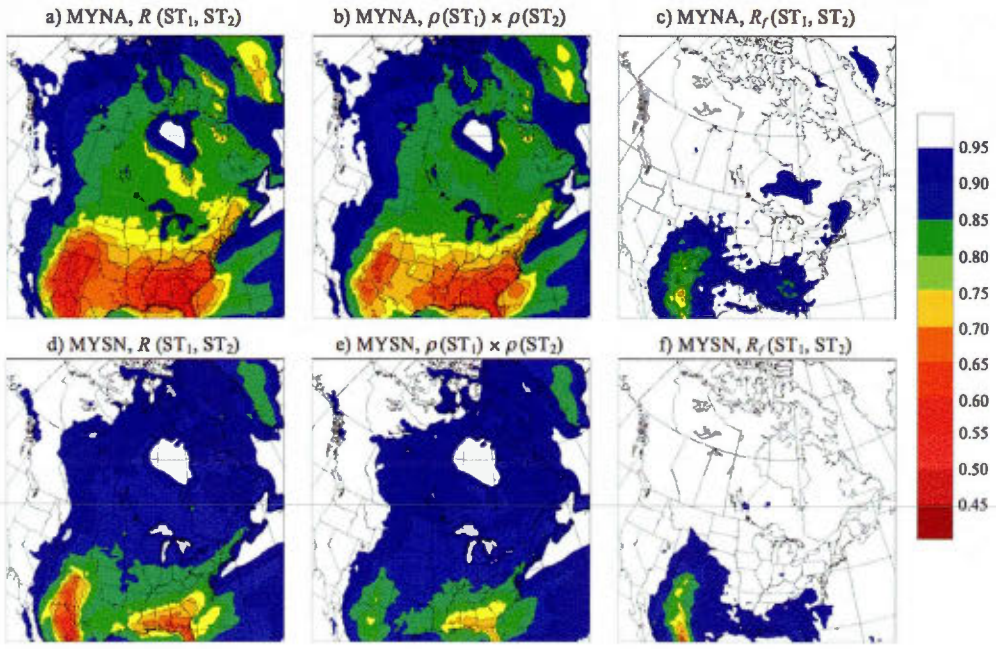

Figure 2.8 Decomposition of the time correlation between control and parameter-perturbed CRCM5 JJA 2 m-temperatures (see Eq. 2.34): time correlation R, product of the reproducibility ratios ρ of the control and parameter perturbed model, and the reproducible correlation R_f; (a, b, c) MYNA and (d, e, f) MYSN.

Table 2.1 The large-scale vertical velocity threshold in the Kain-Fritsch deep convection trigger function used in the two model versions. Also show is the number of ensemble members

Model Version	threshold vertical velocity (m/s)	ensemble size
control	$3.4E-2$	2
perturbed	$6.0E-2$	2

CONCLUSION

The subject matter of this work was to study different RCM simulation configurations, less computationally demanding than the operational runs (in terms of domain size, large-scale SN, inter-member variability and integration period), to be used as test bed for RCM parameter modifications. Ideally, an optimal configuration would yield a model response to parameter perturbations that is representative of typical configurations used in the operational RCM setups. At the same time, it would allow for obtaining of statistically significant estimates by using a reduced computing time.

In the first part of this work we studied the RCM response to parameter perturbations from simulations performed over a single year and driven by the reanalyses. In this part, we focused on the impact of SN and domain size reduction on the model response. Test perturbations were obtained by modifying the values of two parameters related to the deep-convection and large-scale condensation parameterizations. In the perturbed runs, the values of these parameters were set at the limit of their expert-specified range. Therefore, these perturbations should be considered as strong. The model response to the perturbations is compared in three different simulation configurations. The reference configuration included a continental scale domain with no SN. The modified simulation configurations are obtained by (1) applying the large-scale SN in the reference domain and (2) by reducing the size of the reference domain. The reference configuration is selected such that its domain size is similar to those used in operational RCM simulations. Because there is still no general consensus about the application of SN, this is considered as a non-standard option. Finally, for every combination of parameter settings and simulation' configurations, multiple runs with perturbed initial conditions were performed in order to obtain robust estimates of the differences among the model versions.

Both SN and domain-size reduction typically considerably reduce the inter-member variance of ensemble simulations (Alexandru et al., 2007, 2009). In this way, they contribute to the statistical significance of the model response estimated as the difference

of ensemble means of the non-perturbed and perturbed-parameter simulations. In other words, these methods reduce the need for ensemble members for achieving statistically significant estimates of model response and thus help to reduce the computational cost of studying parameter perturbations over a single year. At the same time, SN and domain-size reduction may produce undesirable alterations in the model response to parameter perturbations. Thus the purpose of the above experiments was not only to compare the three simulations configurations in terms of the number of ensemble members necessary to obtain statistically significant estimates of the response to parameter modification but also in terms of their impact on the magnitude and the pattern of the response.

The results obtained in the reference (large-domain, no SN) configuration showed that the deep-convection and stratiform-precipitation parameter perturbations produced fairly large responses in the simulated seasonal-average 2 m-temperatures and precipitation, especially in spring (MAM) and summer (JJA), with magnitudes locally larger than 2°C and 2 mm/day, respectively. Despite a large magnitude, these estimates were not statistically significant over the entire domain, especially in the case of precipitation, because of large spread of ensemble members. SN and domain-size reduction reduced the inter-member variance as expected. As a result, the statistical significance of the estimated model response to perturbations considerably increased in the spectrally nudged and small-domain simulations. Note that the same number of ensemble members was generated in the reference, SN and small-domain configurations. Hence, the SN and small-domain configurations allow for achieving the same levels of statistical significance with a smaller number of ensemble members. On the other hand, the results also showed that SN and domain-size reduction changed the model response to the parameter modifications. However, these changes were statistically significant only in the small-domain configuration.

The above results about the alteration of the response to parameter modification should be carefully interpreted. The statistical significance of the difference between the response to the parameter perturbations in the no-SN and SN setups is (1) proportional to the unknown true difference, (2) proportional to the total number of ensemble members, and (3) inversely proportional to the inter-member spread in the reference and SN configurations. The same holds for the difference between the responses in the small and large domains. As a result of the specific setup of SN parameters in our simulations,

the inter-member spread was quite a bit larger in the SN set than in the small-domain set of simulations. As a consequence, even if SN and domain-size reduction produced the same alteration of the response, this alteration could appear statistically significant in the case of small domain and insignificant in the case of SN, simply because the inter-member spread is larger in the SN case. Thus, to show that in the SN case the alteration of the model response is smaller, we would need more ensemble members. We can thus only conclude that caution is needed when either of the two noise-reducing methods is used, since they both can change the model response to modification.

Throughout the first part of this thesis we studied only the RCM integrations performed over a single year and thus the variability of model response to parameter modification in time was neglected. The purpose of the second part of this thesis was to examine how much the results obtained in the single-year study depend on the choice of the simulated year. In Chapter 2, a general theoretical framework for the estimation of RCM response to modification is developed. The theoretical considerations, albeit lengthy, greatly improve our understanding of model response to modification and thus facilitate the choice of an appropriate simulation configuration for quantifying RCM response. They also help to optimize the computational resources available for an RCM experiment. In what follows we will summarize the main conclusions of the related theoretical considerations.

When in RCM ensemble simulations the inter-member spread is large, the model solution is to a lesser degree controlled by LBC and hence it is less influenced by the selection of the year for which the simulations are performed. This conclusion can be supported by making analogy with CGCM simulations, in which, beyond the predictability limit, the initial conditions are forgotten and there is no information about the time passed from the initialization. In addition, it is equivalent to sample seasonal variables from different CGCM members or from different years of a single CGCM member. Similarly, in RCM simulations, in regions where the inter-member spread is very large, the RCM members would tend to sample some of the inter-annual variability of the RCM solution. On the other hand, when the inter-member spread is very small, such as in our small-domain simulations, the model solution is tightly constrained by LBC and the choice of the simulated year will have considerable impact on the model response to modification. This implies that in the large-domain non-nudged configuration (in which the inter-member spread was the largest among the three configurations) the

estimated model response was the most representative with respect to its inter-annual variability. Note that this is the opposite of what was stated about the statistical significance of the response to parameter modification in the single-year experiment. The largest statistical significance levels were obtained in the small-domain configuration.

In order to approach these issues in a more formal manner, we developed a specific theoretical framework, founded on a distinction between the externally forced, reproducible components of the RCM solution, originating from the RCM response to specific LBC and surface forcing, and irreproducible quasi-random components originating from the RCM chaotic nature. The key property of these two components is that the sum of their variances is equal to the time variance of individual RCM simulations, in the limit when the integration time increases. Since the time variance is an observable, any reasonable modification should not produce large changes in the time variance. This implies that in simulations' configurations characterized with small inter-member spread, the small irreproducible variance component has to be compensated with a large reproducible variance. Conversely, when the inter-member spread is large, the reproducible component of the time variance is small, the time variance being then dominated by the irreproducible components. As a consequence, in our single-year experiments, ensemble members are more efficient in sampling temporal variability in the reference, large-domain, no-nudging configuration, where the member spread was large. Since the SN and small domains reduce the inter-member spread and the irreproducible deviations they also diminish the efficiency of members in sampling temporal variability of the RCM response to parameter modification. When the member spread is very small then the ensemble members are almost identical, as if there was only a single member ensemble.

A pragmatically based approach to the aforementioned issues is to note a distinction between the cases when the researcher attempts to downscale a particular weather event or season from the objective analyses and to estimate the long-term RCM mean. In the downscaling case, the variable of interest is typically the difference between two RCM versions (control and perturbed) but only over a specified period. In this case, what is estimated is the difference between the reproducible components of the modified and control model. The error variance of the difference is contributed solely by the irreproducible components. The only way of reducing the error variance in the estimated difference between two models is, in this case, to generate the ensemble members

in order to sample the irreproducible noise. On the other hand, when the researcher attempts to estimate the difference of long-term (climatological) means between two RCM versions then both the reproducible and irreproducible components of the time anomaly have to be regarded as the source of noise. The irreproducible components can be sampled both ways – in time and from ensemble members. The RCM reproducible components however depend on the specific choice of the LBC and the ocean surface conditions, so that they can be sampled only in time. Thus the optimal way of allocating the computing resources in the case of the estimation of the difference of long-term time means of two model versions is to extend the integration period as much as possible in the control and modified-model integrations. This implies, that the optimal approach is to use single-member ensembles for each model version.

The error variance in the estimated difference of time means of two RCM simulations is, by definition, proportional to the time variance of the difference between two simulations. The theoretical framework presented in this thesis provides a basis to analyze the impact of SN and domain-size reduction on this error variance by considering their influence on the difference's time variance. This quantity can be also decomposed into the reproducible and irreproducible components whose behavior is somewhat different then that of the corresponding components in a single RCM simulation. If the simulations share the same LBC the reproducible components of the two simulations are, in general, correlated in time. If this correlation is large the reproducible components in the two simulations will be largely subtracted out from the difference between these two simulations. Consequently, the difference will be then dominated by the irreproducible components, no matter what is the relative magnitude of the two components in each of the two simulations.

Let us now consider what would be the impact of SN and domain-size reduction in this case. SN and domain-size reduction decrease the irreproducible noise in the RCM simulations. We note from the previous considerations that a decrease of the irreproducible variance has to be compensated by an increase of the reproducible variance in every RCM simulation. However, because the reproducible components of the two simulations are assumed highly correlated, they will be subtracted from the difference between simulations, so the compensation property would have a small impact on the difference. As a result, in the case of large correlations between the reproducible components, SN and domain-size reduction can reduce the time variance of the difference

between two simulations. Consequently, they will also diminish the error variance in the estimated difference of means. This holds for the case when the time correlation between the reproducible components is large.

On the other hand, if the time correlation between the reproducible components of two RCM simulations is vanishingly small, then there will be no subtraction of the reproducible components in the difference between two simulations. In this case, the compensation rule between the reproducible and irreproducible components also holds for the difference between simulations. Consequently, a decrease of the irreproducible components of the difference by SN and domain-size reduction will be compensated by an increase of similar magnitude of its reproducible components. Hence, in this case, SN and domain-size reduction, despite reducing the magnitude of the irreproducible components, cannot reduce the time variance of the difference.

The above considerations show that the efficiency of SN and domain-size reduction in diminishing the error variance of the difference of time means of two simulations depends on the nature of the RCM modification as well as on the driving data in the two simulations. The correlation between the reproducible components is larger if the two simulations share the same LBC and surface forcing. In order to support these considerations with some examples we performed several decadal RCM simulations driven by the reanalyzes. In this experiment, two RCM versions were used, the control and the modified version, obtained by modifying the value of a deep-convection threshold parameter. The model response to this perturbation is compared in two large-domain simulation configurations. The only difference between the two configurations is the SN. Using the theoretical framework developed in this thesis we estimated the reproducible and irreproducible components of the simulations' time variances by performing two-member ensemble simulations with perturbed initial conditions for every combination of model version and configuration. All these simulations shared the same driving fields. The results showed that the reproducible time anomalies of the control and modified-parameter simulations were highly correlated over the entire computational domain. In the SN case, the time variance of the difference between the control and modified-parameter runs was quite a bit smaller than in the no-SN case. A practical implication of these results is that SN can reduce the integration periods necessary to achieve robust estimates of the climatological mean response to parameter modification. However, the results also showed that SN produced statistically significant alterations

of the time-mean model response to the perturbation of the deep convection parameter.

RCM development and applications often require varying model changes and configurations, which implies a large number of sensitivity tests. Because of a large computing cost of such experiments, it is essential to conduct the tests in a computationally efficient way. This study represents an attempt to find optimum RCM simulation set-ups that use less computation time than operational runs while still returning representative results. The mathematical descriptions presented in this thesis may allow for a better understanding and planning of RCM simulation configurations while saving computing time. The results of this work foster the use of SN in RCM sensitivity tests as a way of reducing the computing cost. However, caution is needed because SN can change the the model mean response to parameter modification.

It is worth noting that the modification of a single parameter represents only a minor modification. As a future work, it may be interesting to consider weather the reproducible components of the simulations of two models with major differences, such as those developed at different research centers, but driven with the same LBC, would be as highly correlated in time as they were in the case of the minor modification presented here. It is also worth noting that in all simulations presented here, the SN was configured such that it minimally interfered with the model own interior dynamics at fine and intermediate spatial scales and in the lower half of the model's atmosphere. A stronger SN would further reduce the irreproducible noise in our simulations but might also produce larger alterations in the time-mean response to parameter modification. An optimal SN configuration in that respect has to be determined experimentally, as a part of the future work on the topics considered in this thesis.

REFERENCES

Ackerley, D., Highwood, E. J., and Frame, D. J. (2009). Quantifying the effects of perturbing the physics of an interactive sulfur scheme using an ensemble of GCMs on the climateprediction.net platform. *Journal of Geophysical Research*, 114(D1):D01203+.

Alexandru, A., De Elia, R., and Laprise, R. (2007). Internal Variability in Regional Climate Downscaling at the Seasonal Scale. *Monthly Weather Review*, 135(9):3221–3238.

Alexandru, A., de Elia, R., Laprise, R., Separovic, L., and Biner, S. (2009). Sensitivity Study of Regional Climate Model Simulations to Large-Scale Nudging Parameters. *Monthly Weather Review*, 137(5):1666–1686.

Arakawa, A. and Lamb, V. R. (1977). *Computational design of the basic dynamical processes of the UCLA general circulation model*, volume 17, pages 173–265. Academic Press.

Arritt, R. W., Anderson, C. J., Takle, E. S., and Al, E. (2004). Ensemble methods for seasonal limited-area forecasts. In *20th Conference on Weather Analysis and Forecasting / 16th Conference on Numerical Weather Prediction*.

Barnett, D., Brown, S., Murphy, J., Sexton, D., and Webb, M. (2006). Quantifying uncertainty in changes in extreme event frequency in response to doubled CO2 using a large ensemble of GCM simulations. *Climate Dynamics*, 26(5):489–511.

Biner, S., Caya, D., Laprise, R., and Spacek, L. (2000). Nesting of RCMs by imposing large scales. Technical report, WMO/TD 987.

Bryan, F. O. (1998). Climate Drift in a Multicentury Integration of the NCAR Climate System Model. *J. Climate*, 11(6):1455–1471.

Burke, E. J., Perry, R. H. J., and Brown, S. J. (2010). An extreme value analysis of UK drought and projections of change in the future. *Journal of Hydrology*, 388(1-2):131–143.

Caya, D. and Biner, S. (2004). Internal variability of RCM simulations over an annual cycle. *Climate Dynamics*, 22(1):33–46.

Christensen, J., Carter, T., Rummukainen, M., and Amanatidis, G. (2007). Evaluating the performance and utility of regional climate models: the PRUDENCE project. *Climatic Change*, 81(0):1–6.

Colin, J., Déqué, M., Radu, R., and Somot, S. (2010). Sensitivity study of heavy precipitation in Limited Area Model climate simulations: influence of the size of the domain and the use of the spectral nudging technique. *Tellus A*, 62(5):591–604.

Collins, M., Tett, S. F. B., and Cooper, C. (2001). The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 17(1):61–81.

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A. (1998). The Operational CMCMRB Global Environmental Multiscale (GEM) Model. Part I: Design Considerations and Formulation. *Mon. Wea. Rev.*, 126(6):1373–1395.

Davies, H. C. (1976). A lateral boundary formulation for multi-level prediction models. *Q.J.R. Meteorol. Soc.*, 102(432):405–418.

de Elía, R., Caya, D., Côté, H., Frigon, A., Biner, S., Giguère, M., Paquin, D., Harvey, R., and Plummer, D. (2008). Evaluation of uncertainties in the CRCM-simulated North American climate. *Climate Dynamics*, 30(2):113–132.

de Elía, R., Laprise, R., and Denis, B. (2002). Forecasting Skill Limits of Nested, Limited-Area Models: A Perfect-Model Approach. *Mon. Wea. Rev.*, 130(8):2006–2023.

Forest, C. E., Stone, P. H., and Sokolov, A. P. (2006). Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophysical Research Letters*, 33(1):L01705+.

Frame, D. J., Booth, B. B. B., Kettleborough, J. A., Stainforth, D. A., Gregory, J. M., Collins, M., and Allen, M. R. (2005). Constraining climate forecasts: The role of prior assumptions. *Geophysical Research Letters*, 32(9):L09702+.

Giorgi, F. and Bi, X. (2000). A study of internal variability of a regional climate model. *Journal of Geophysical Research*, 105(D24):29503–29521.

Giorgi, F. and Francisco, R. (2000). Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dynamics*, 16(2):169–182.

Giorgi, F., Jones, C., and Asrar, G. (2009). Addressing climate information needs at the regional level: The CORDEX framework. *WMO Bulletin*, 58(3):175–183.

Giorgi, F. and Mearns, L. O. (1999). Introduction to special section: Regional climate modeling revisited. *Journal of Geophysical Research*, 104(D6):6335–6352.

Giorgi, F. and Mearns, L. O. (2002). Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the Reliability Ensemble Averaging (REA) Method. *J. Climate*, 15(10):1141–1158.

Giorgi, F. and Mearns, L. O. (2003). Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. *Geophysical Research Letters*, 30(12):1629+.

Jones, R. G., Murphy, J. M., and 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(526):1413–1449.

Juang, H.-M. H. and Hong, S.-Y. (2001). Sensitivity of the NCEP Regional Spectral Model to Domain Size and Nesting Strategy. *Mon. Wea. Rev.*, 129(12):2904–2922.

Kain, J. S. and Fritsch, J. M. (1990). A One-Dimensional Entraining/Detraining Plume Model and Its Application in Convective Parameterization. *J. Atmos. Sci.*, 47(23):2784–2802.

Kirtman, B. P., Pegion, K., and Kinter, S. M. (2005). Internal Atmospheric Dynamics and Tropical Indo-Pacific Climate Variability. *J. Atmos. Sci.*, 62(7):2220–2233.

Knutti, R., Meehl, G. A., Allen, M. R., and Stainforth, D. A. (2006). Constraining Climate Sensitivity from the Seasonal Cycle in Surface Temperature. *J. Climate*, 19(17):4224–4233.

Laprise, R. (1992). The Euler Equations of Motion with Hydrostatic Pressure as an Independent Variable. *Mon. Wea. Rev.*, 120(1):197–207.

Laprise, R. (2008). Regional climate modelling. *Journal of Computational Physics*, 227(7):3641–3666.

Laprise, R., de Elía, R., Caya, D., Biner, S., Lucas-Picher, P., Diaconescu, E., Leduc, M., Alexandru, A., and Separovic, L. (2008). Challenging some tenets of Regional Climate Modelling. *Meteorology and Atmospheric Physics*, 100(1):3–22.

Leduc, M. and Laprise, R. (2009). Regional climate model sensitivity to domain size. *Climate Dynamics*, 32(6):833–854.

Livezey, R. E. and Chen, W. Y. (1983). Statistical Field Significance and its Determination by Monte Carlo Techniques. *Mon. Wea. Rev.*, 111(1):46–59.

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *J. Atmos. Sci.*, 20(2):130–141.

Lucas-Picher, P., Caya, D., Biner, S., and Laprise, R. (2008a). Quantification of the Lateral Boundary Forcing of a Regional Climate Model Using an Aging Tracer. *Mon. Wea. Rev.*, 136(12):4980–4996.

Lucas-Picher, P., Caya, D., de Elía, R., and Laprise, R. (2008b). Investigation of regional climate models' internal variability with a ten-member ensemble of 10-year simulations over a large domain. *Climate Dynamics*, 31(7):927–940.

McGregor, J. L. (1997). Regional climate modelling. *Meteorology and Atmospheric Physics*, 63(1):105–117.

Mearns, L. O., Gutowski, W., Jones, R., Leung, R., McGinnis, S., Nunes, A., and Qian, Y. (2009). A regional climate change assessment program for North America. *Eos Trans. AGU*, 90(36):311+.

Miguez-Macho, G., Stenchikov, G. L., and Robock, A. (2004). Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. *Journal of Geophysical Research*, 109(D13):13104–13127.

Murphy, J. M., Booth, B. B. B., Collins, M., Harris, G. R., Sexton, D. M. H., and Webb, M. J. (2007). A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857):1993–2028.

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001):768–772.

Nakicenovic, N., Davidson, O., Davis, G., Grübler, A., Kram, T., La Rovere, E. L., Metz, B., Morita, T., Pepper, W., Pitcher, H., Sankovski, A., Shukla, P., Swart, R., Watson, R., and Dadi, Z. (2000). *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 1 edition.

Nikiema, O. and Laprise, R. (2011). Diagnostic budget study of the internal variability in ensemble simulations of the Canadian RCM. *Climate Dynamics*, 36(11):2313–2337.

Piani, C., Frame, D. J., Stainforth, D. A., and Allen, M. R. (2005). Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophysical Research Letters*, 32(23):L23825+.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1992). *Numerical Recipes in Fortran 77: The Art of Scientific Computing*. Cambridge University Press, 2 edition.

Pudykiewicz, J., Benoit, R., and Mailhot, J. (1992). Inclusion and verification of a predictive Cloud-Water Scheme in a Regional Numerical Weather Prediction Model. *Mon. Wea. Rev.*, 120(4):612–626.

Räisänen, J. and Palmer, T. N. (2001). A Probability and Decision-Model Analysis of a Multimodel Ensemble of Climate Change Simulations. *J. Climate*, 14(15):3212–3226.

Rapaić, M., Leduc, M., and Laprise, R. (2011). Evaluation of the internal variability and estimation of the downscaling ability of the Canadian Regional Climate Model for different domain sizes over the north Atlantic region using the Big-Brother experimental approach. *Climate Dynamics*, 36(9):1979–2001.

Rummukainen, M. (2010). State-of-the-art with regional climate models. *WIREs Clim Change*, 1(1):82–96.

Sanderson, B. M., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D. J., Ingram, W. J., Piani, C., Stainforth, D. A., Stone, D. A., and Allen, M. R. (2008). Constraints on Model Response to Greenhouse Gas Forcing and the Role of Subgrid-Scale Processes. *J. Climate*, 21(11):2384–2400.

Separovic, L., de Elía, R., and Laprise, R. (2008). Reproducible and Irreproducible Components in Ensemble Simulations with a Regional Climate Model. *Mon. Wea. Rev.*, 136(12):4942–4961.

Seth, A. and Giorgi, F. (1998). The Effects of Domain Choice on Summer Precipitation

Simulation and Sensitivity in a Regional Climate Model. *J. Climate*, 11(10):2698–2712.

Sexton, D. and Murphy, J. (2003). Design of coupled atmosphere/ocean mixed-layer model experiments for probabilistic prediction. Technical report, Hadley Centre.

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., and Allen, M. R. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, 433(7024):403–406.

Sundqvist, H., Berge, E., and Kristjánsson, J. E. (1989). Condensation and Cloud Parameterization Studies with a Mesoscale Numerical Weather Prediction Model. *Mon. Wea. Rev.*, 117(8):1641–1657.

Takle, E. S., Gutowski, W. J., Arritt, R. W., Pan, Z., Anderson, C. J., da Silva, R. R., Caya, D., Chen, S.-C., Giorgi, F., Christensen, J. H., Hong, S.-Y., Juang, H.-M. H., Katzfey, J., Lapenta, W. M., Laprise, R., Liston, G. E., Lopez, P., McGregor, J., Pielke, R. A., and Roads, J. O. (1999). Project to Intercompare Regional Climate Simulations (PIRCS): Description and initial results. *Journal of Geophysical Research*, 104(D16):19443–19461.

Taylor, K. E., Williamson, D., and Zwiers, F. (2000). The Sea Surface Temperature and Sea-Ice Concentration Boundary Conditions for AMIP II Simulations. Technical report, Lawrence Livermore National Laboratory, Program for Climate Model Diagnosis and Intercomparison.

Tebaldi, C. and Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857):2053–2075.

Tebaldi, C., Smith, R. L., Nychka, D., and Mearns, L. O. (2005). Quantifying Un-

certainty in Projections of Regional Climate Change: A Bayesian Approach to the Analysis of Multimodel Ensembles. *J. Climate*, 18(10):1524–1540.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J. F., Morcrette, J. J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J. (2005). The ERA-40 re-analysis. *Q.J.R. Meteorol. Soc.*, 131(612):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(1):187–189.

von Storch, H., Langenberg, H., and Feser, F. (2000). A Spectral Nudging Technique for Dynamical Downscaling Purposes. *Mon. Wea. Rev.*, 128(10):3664–3673.

von Storch, H. and Zwiers, F. W. (2002). *Statistical Analysis in Climate Research*. Cambridge University Press.

Wang, Y., Leung, L. R., McGregor, J. L., Lee, D. K., Wang, W. C., Ding, Y. H., and Kimura, F. (2004). Regional climate modeling: Progress, challenges and prospects. *J. Meteor. Soc. Japan*, 82:1599–1628.

Weisse, R. and Feser, F. (2003). Evaluation of a method to reduce uncertainty in wind hindcasts performed with regional atmosphere models. *Coastal Engineering*, 48(4):211–225.

Weisse, R., Heyen, H., and von Storch, H. (2000). Sensitivity of a Regional Atmospheric Model to a Sea StateDependent Roughness and the Need for Ensemble Calculations. *Mon. Wea. Rev.*, 128(10):3631–3642.

Yang, Z. and Arritt, R. W. (2002). Tests of a Perturbed Physics Ensemble Approach for Regional Climate Modeling. *J. Climate*, 15(20):2881–2896.

Zadra, A., Caya, D., Côté, J., Dugas, B., Jones, C., Laprise, R., Winger, K., and Caron, L. P. (2008). The next Canadian regional climate model. *Physics in Canada*, 64:74–83.

Zwiers, F. W. (1996). Interannual variability and predictability in an ensemble of AMIP climate simulations conducted with the CCC GCM2. *Climate Dynamics*, 12(12):825–847.