Regional climate modelling

René Laprise *

Centre ESCER (Étude et Simulation du Climat à l’Échelle Régionale), Université du Québec à Montréal, 550 Sherbrooke West, 19th Floor, Montréal, Que., Canada H3A 1B9

Received 6 April 2006; received in revised form 4 October 2006; accepted 10 October 2006

Available online 30 November 2006

Abstract

A new technique for climate simulations and projections has emerged over the past 15 years that allows to reach unprecedented spatial resolution at an affordable computing cost: regional climate modelling. The technique has matured and several groups around the world are participating in model intercomparison projects and collaborating in coordinated endeavour to produce ensemble climate-change projections at regional scales.

The paper describes the formulation of a specific regional climate model (RCM), the Canadian RCM, commenting on the formulation choices and experience gained during the development and validation process of this model. Current issues in regional climate modelling are also reviewed and discussed.

© 2006 Published by Elsevier Inc.

Keywords: Regional climate model; Climate modelling; Climate change; Numerical methods

1. Introduction

Climate is one of the most challenging geophysical systems to simulate because of the number of interacting components and the wide range of time and spatial scales of relevant processes and their complexity. The components of the climate system are as diversified as the atmosphere (with clouds, aerosols and radiatively active gases including water vapour that transports latent heat), the hydrosphere (oceans, lakes and wet lands), the cryosphere (sea ice, land glaciers and snow), the lithosphere (land–surface processes including soil moisture in its various phases) and the biosphere (vegetation effects on surface albedo, roughness and evapotranspiration, and carbon cycle both over land and in oceans). The physical processes taking place within the individual components and the interactions between the components cover a wide range of temporal scales, owing to vastly different reservoir capacity of the components and the strength of the interactions. Spatial scales also vary greatly, ranging from the micro scale of cloud droplets to the planetary scale of atmospheric and abyssal ocean circulations. The complexity of climate denies a direct mathematical solution, and its global nature precludes the use of direct field experimentation (such as is done for example with cloud seeding for short-term...
weather modification) or laboratory physical models (such as rotating dishpan for Rossby waves or wind tunnel for aerodynamics studies). Numerical modelling constitutes the only approach to further our understanding of the mechanisms responsible for the maintenance of the climate system’s dynamic equilibrium and variability, and to probe its susceptibility and response to changes in its external forcing.

As a consequence of the non-linear nature of the climate system, its behaviour is chaotic, which has profound implications for its numerical modelling. On the one hand, simulations initialised with negligible differences in, say, initial conditions will result in different sequences of weather events. On the other hand, the long-term climate statistics will be the same for a given set of external boundary conditions, and they will be independent of the details of the initial state; this is a benefit given our limited knowledge of the precise state of several of the climate components. The ensuing ergodicity property implies that, under constant external forcing, long-term climate statistics can be generated equivalently by integrating a climate model for very long simulated time periods or by carrying several shorter simulations, a procedure called “ensemble”, which is very convenient for practitioners. Long time-scale variability, however, can only be addressed through long simulations. A climate model is effectively a “weather simulator” whose statistics are studied to determine the simulated climate. Similar to natural variability in the real world, the sequence of weather events in climate simulations is referred to as “weather noise”. The presence of internal variability on a wide range of time scales, however, complicates the determination of the climate statistics, with important consequences for studies of climate changes in response to perturbations in external forcing, with respect to the statistical detection, identification and causal attribution of climate changes referred to as the “signal”.

Coupled global climate models (CGCMs) simulate several aspects of the climate components and their interactions, and thus represent the most sophisticated tools in climate research. Their complexity results in a high computational cost, which is exacerbated by the requirement of ensemble and/or long simulations for statistical significance of the results. As a consequence, the number of computational nodes of climate models employed for century time-scale projections of climate changes, such as those anticipated due to anthropogenic effects, are only of the order of a few millions (about 200 longitudes, 100 latitudes, 50 levels in the vertical, and at least four prognostic variables for the atmosphere and as many in the ocean). In practice, given the currently available computing power of climate research centres, most CGCMs employ computational meshes with grid-point spacing in the horizontal of some hundreds of km. This results in severe truncation in the numerical solution of the differential equations. Owing to the non-linear character of the field equations, the truncation generates a closure problem that calls for parameterising the ensemble effect of subgrid-scale physical processes, such as convection, clouds and precipitation, planetary boundary layer turbulence, interaction of solar and terrestrial electromagnetic radiation with matter, for example. Parameterisations are physically motivated statistical relationships between the state of grid-box mean variables and the forcing due to unresolved physical processes, with some parameters being set for optimal fit. Despite employing sometimes rather crude approximations of the real processes, the parameterisations themselves account for a large part of the computational cost of climate models. Considering the magnitude of numerical errors associated with the discretisation of the equations, which indicates that length scales smaller than several grid points (from 2 to 10, depending on the accuracy standard) are inaccurately handled (e.g. [78,91]), the coarse grids of operational CGCMs preclude resolving with sufficient accuracy several processes that are deemed important for direct application of their projections in climate-change impact studies.

As a pragmatic alternative to the (quasi) uniform coarse resolution of CGCMs, the degrees of freedom can be concentrated over a region of the globe where lies the specific interest of a user: this is called regional climate modelling. However, because the climate system does not know about territorial boundaries, there is clearly a need for maintaining a global aspect at the same time as increasing the resolution over a specific region. Two approaches have been used: nested limited-area models (LAM) and variable-resolution (including stretched-grid) global models (VRGCM).

Following the pioneering work at NCAR in the late 80s ([28,46]), the LAM approach consisting of a high-resolution limited-area model, one-way nested at its lateral boundaries with low-resolution CGCM, has gained wide acceptance. In practice one order of magnitude in (linear) resolution can be gained with the nested approach compared to a GCCM by correspondingly restricting the LAM’s computational domain. One advantage of a LAM is that it can also be driven by (accurate) atmospheric reanalyses rather than by CGCM (with their biases); this feature is very convenient for development and validation purposes. When LAMs are
applied to climate time scales, they are referred to as Regional Climate Models (RCMs); this will be the focus of this paper.

The VRGCM approach consisting of using a variable-resolution global model, with high-resolution over a specific region and lower resolution over the rest of the globe, is gaining popularity at several climate centres (e.g. [69,44,38]). The main advantage of VRGCM is the use of a single model rather than the tandem CGCM-RCM; this avoids potential lack of coherence between the formulation of two models, and it allows naturally two-way interaction between the region of interest and the rest of the globe. While concentrating the resolution over a subset of the Earth’s surface increases computational efficiency, it does not come free of some problems owing to the variation of resolution. The non-uniformity and anisotropy of the computational grid can result in contamination of the solution, as is the case with the convergence of the meridians near the poles on latitude–longitude grids. In both cases the approximation consisting in a local-only increase in resolution comes at the expense of sacrificing some non-linear interactions, as discussed in [65]. In practice the skill of the LAM and VRGCM approaches in reproducing the current climate is very similar (e.g. [39]). The Stretched-Grid Model Intercomparison Project (SGMIP [90,38]) was initiated to study the global variable-resolution (or stretched-grid) approach to regional climate modelling, focusing primarily on the issues of stretching strategy, numerical discretisation, physical parameterisation including its calculation on intermediate uniform resolution or directly on the stretched grid, and performance on parallel supercomputers.

Any regional climate modelling approach affords an increase of resolution over a region of the globe of about one order of magnitude compared to CGCMs, with regional grid-point spacing of a few tens of km in the horizontal, for operational use on climate timescales. It must be borne in mind that regional climate modelling is not a goal in itself but rather a pragmatic approach to circumvent the insufficient computing power required for operationally carrying century-long simulations with high-resolution CGCMs. On the other hand, even when the increase of computing power will allow the operational use of CGCMs at resolution of a few tens of km, the RCM approach could still be useful, allowing to reach resolution of a few km for the same computational load.

Several detailed reviews on regional climate modelling already exist in the specialised literature (e.g. [47,68,105]). Hence this paper will take a rather personal perspective, trying to provide testimony of the steps required for the development, validation and improvement of any RCM, but focussing mainly on the work of the author and his colleagues on a specific RCM, with appropriate references to the work of others as deemed required for completeness. While some aspects of the description will be model-specific, others are shared by several RCMs, thus providing an overview of regional climate modelling. The paper is organised as follows. The next section describes in general terms the formulation of the Canadian RCM, as a specific example of an RCM, and discusses the learning experience of developing such a model. Section 3 provides an overview of some regional climate modelling programmes around the world. Section 4 discusses the current scientific issues that are the focus of attention of RCM research groups. Section 5 provides the concluding remarks.

2. Regional climate model formulation: a Canadian perspective

This section provides an overview of the formulation of the Canadian RCM (CRCM), relating to its dynamics, numerics, nesting and subgrid-scale parameterisations. While a few aspects are model-specific and pertain solely to the CRCM, others are shared by several other RCMs, thus providing a general overview of typical RCMs formulation. Several comments are made about the repercussions of certain choices in the model formulation.

2.1. Dynamical formulation of the CRCM

A noteworthy feature of the CRCM is that it is based on the fully elastic non-hydrostatic field equations, solved by a semi-implicit semi-Lagrangian time marching scheme ([16,17,63,64]). Although non-hydrostatic effects do not come to play at resolutions that can currently be afforded for climate application, this choice stemmed from the desire to make use of a general dynamical kernel developed by the late Robert and colleagues [96], with the claim that it could be used efficiently for atmospheric applications at all spatial scales [62].
The CRCM employs a polar-stereographic conformal projection in the horizontal (e.g. [50]) and a terrain-following scaled-height coordinate in the vertical [41]. Following the traditional approximation in meteorology [77], but without the hydrostatic approximation, the field equations take the following form (see [62] for details on the derivation of these equations):

\[
D_t U + RT' \partial_x q' = f V - K \partial_x S - RT' \partial_y q' - RT' \frac{G_1}{G_0} \partial_z q' + F_X
\]

(1)

\[
D_t V + RT' \partial_y q' = -f U - K \partial_y S - RT' \partial_y q' - RT' \frac{G_2}{G_0} \partial_z q' + F_Y
\]

(2)

\[
D_t w + \frac{RT'}{G_0} \partial_z q' - \frac{g}{T'} T' = -\frac{RT'}{G_0} \partial_z q' + F_Z
\]

(3)

\[
D_t T' - \alpha T' D_t q' + \frac{g q}{R} w = \frac{a T'}{1 - z} \left\{ S(F_1 U + F_2 V) - S(\partial_x U + \partial_y V) - \frac{1}{G_0} \partial_z (G_0 W) + \frac{L}{T} \right\} + L
\]

(4)

\[
(1 - \alpha) \left( D_t q' - \frac{g}{RT'} w \right) + S(\partial_x U + \partial_y V) + \frac{G_0}{G_0} \partial_z (G_0 W) = S(F_1 U + F_2 V) - \frac{1}{G_0} \partial_z (G_0 W) + \frac{L}{T}
\]

(5)

\[
D_t M = E
\]

(6)

\[
w - G_0 W = -S(G_1 U + G_2 V)
\]

(7)

The independent variables of the transformed coordinates are noted by \((X, Y, Z)\) while the physical coordinates on the sphere are noted \((x, y, z)\). The horizontal coordinates are scaled and rotated as follows:

\[
\begin{pmatrix}
\frac{dX}{dY}
\end{pmatrix} = m \begin{pmatrix}
-\sin \lambda & -\cos \lambda \\
\cos \lambda & -\sin \lambda
\end{pmatrix} \begin{pmatrix}
\frac{dx}{dy}
\end{pmatrix}
\]

(8)

where the map scale factor of the polar-stereographic projection is

\[
m = \frac{1 + \sin \varphi_0}{1 + \sin \varphi}
\]

(9)

with \(\lambda\) the longitude and \(\varphi\) the latitude; \(\varphi_0\) is the reference latitude for the projection. The terrain-following scaled vertical coordinate is defined as

\[
Z(X, Y, z) = \left\{ \frac{z - z_0(X, Y)}{H - z_0(X, Y)} \right\} H
\]

(10)

so that \(Z = 0\) on a surface at the height of topography \(z_0\) and \(Z = H\) at the model’s computational lid \(z = H\). The coordinate surfaces rectify linearly with height above the surface, to become horizontal at the height of the computational lid.

Consistent with the coordinate transformation, the physical wind components \((u, v, w)\) are transformed to scaled components \((U, V, W)\). The scaled and rotated horizontal wind components are defined as

\[
\begin{pmatrix}
U \\
V
\end{pmatrix} = m^{-1} \begin{pmatrix}
-\sin \lambda & -\cos \lambda \\
\cos \lambda & -\sin \lambda
\end{pmatrix} \begin{pmatrix}
u
\end{pmatrix}
\]

(11)

and the scaled vertical velocity \(W\) is defined via the relation (7). The introduction of the variable of \(W\) facilitates the formulation and implementation of the physical kinematic lower boundary condition of no flow through topography as a simple homogeneous boundary condition \(W(X, Y, Z = 0) = 0\). For simplicity a similar boundary condition is also applied at the model’s computational lid, \(W(X, Y, Z = H) = 0\), with the understanding that it constitutes a reflective upper boundary condition for vertically propagating internal waves.

The other dependent variables of (1)–(7) are \(T\) the temperature, \(q = \ln(p/p_0)\) with \(p\) the pressure and \(p_0\) a constant, and \(M\) the water vapour specific humidity. The symbols \((F_X, F_Y, F_Z)\) refer to accelerations due to subgrid-scale effects, \(L\) refers to diabatic heating contributions from latent heat released with water phase changes in the atmosphere, radiative flux divergences and other subgrid-scale effects, and \(E\) refers to sources and sinks of water vapour in the atmosphere (more on these terms in Section 2.2). The other variables take their usual meteorological interpretation: \(f = 2\Omega \sin \varphi\) the Coriolis parameter, \(\Omega\) the rotation frequency of
the Earth, \( R \) the gas constant, \( g \) the gravitational acceleration (not to be confused with \( g_0 \) defined below), \( \alpha = \rho^{-1} \) the specific volume, \( \rho \) the density. Auxiliary variables are \( K = (U^2 + V^2)/2 \), \( S = m^2 \), and the factors \( (g_0, G_0, G_1, G_2, F_1, F_2) \) are metric terms associated with the coordinate transformations (see [62] for details). The Lagrangian time derivative can be understood in terms of its local tendency and three-dimensional advection contributions in Eulerian form:

\[
D_t = (\partial_t) + S(U\partial_X + V\partial_Y) + W\partial_Z \tag{12}
\]

In practice (12) is not directly employed because the model uses a semi-Lagrangian transport algorithm as explained later. Throughout this section, all partial derivatives in \( X \), \( Y \) or \( t \) are implicitly understood to be evaluated along constant-\( Z \) surfaces, and partial derivatives in \( Z \) are evaluated along the vertical direction, i.e. at constant \( (X, Y) \).

### 2.2. Numerical formulation of the CRCM

To reduce the presence of computational modes and to ease the implementation of centred finite-difference approximations to spatial derivatives in (1)–(7), the dependent variables are staggered. In the horizontal, variables are arranged following an Arakawa C-grid [5], with the location of the variable \( U \) displaced in \( X \) and that of \( V \) displaced in \( Y \) with respect to the variables \( (q, T, M, w, W) \). When needed the divergence is co-located with \( (q, T, M, w, W) \), and the vorticity are displaced in both \( X \) and \( Y \) with respect to the variables \( (q, T, M, w, W) \). In the vertical the staggering follows a Tokioka B-grid arrangement [100], with the variables \( (T, M, w, W) \) displaced in the vertical with respect to \( (U, V, q) \) – an exception is the lowest level of \( (T, M) \) that lies between \( W \) at the lower boundary and \( q \) above, which facilitates the implementation of physical parameterisations (Fig. 1). Note that this arrangement corresponds to neither of the more commonly used Charney–Phillips (CP) or Lorenz (L) grids (e.g. [58] Section 3.3.5). On the CP-grid, horizontal wind components and temperature are staggered with respect to vertical velocity and pressure (or geopotential); on the L-grid, horizontal wind components, temperature and pressure (or geopotential) are staggered with respect to vertical velocity. The Tokioka B-grid facilitates the implementation of the hydrostatic equilibrium between temperature and geopotential, as does the CP-grid, but it facilitates also the implementation of the geostrophic equilibrium between geopotential and horizontal wind, as does the L-grid; the co-location of temperature and vertical velocity also facilitates the expression of the thermodynamic energy equation. The vertical staggering prevents the occurrence of computational modes [100]; this staggering is close to that judged as the optimal one by [98] who promote the use of potential temperature rather than temperature as dependent variable. In the CRCM the momentum levels are located at the mid-point (in height) between adjacent thermodynamic levels; hence in terms of the notation of the layering schemes (LS) as defined by the relations (5.2) in [59], thermodynamic levels follow LS 2 and momentum levels LS 4.

The rather peculiar form of (1)–(7) is chosen to facilitate the implementation of the semi-implicit semi-Lagrangian time-integration algorithm: on the left-hand side are only retained Lagrangian time derivatives

\[
\begin{align*}
\text{w} &= W = 0 \\
\text{U, V, q} & \quad \text{w, W, T, M} \\
\text{U, V, q} & \quad \text{w, W, T, M} \\
\text{U, V, q} & \quad \text{w, W, T, M} \\
\text{U, V, q} & \quad \text{w, W, T, M} \\
\text{U, V, q} & \quad \text{T, M} \\
\text{W = 0, w} &
\end{align*}
\]

Fig. 1. Staggered layering in the vertical.
and linear terms. Linearisation of the equations is done around a dry, isothermal, hydrostatic environment at rest, noted by an asterisk:

\[
T(X, Y, Z, t) = T^* + T'(X, Y, Z, t)
q(X, Y, Z, t) = q'(z(X, Y, Z)) + q'(X, Y, Z, t)
\]

where \( T^* \) is a constant and \( g' = q_0 - \frac{\rho}{\rho_0} z(X, Y, Z) \) with \( q_0 \) a constant. The semi-Lagrangian transport is calculated following a three-time-level scheme:

\[
D_t \Psi = \frac{\Psi(X, Y, Z, t + \Delta t) - \Psi(X - 2\alpha, Y - 2\beta, Z - 2\gamma, t + \Delta t)}{2\Delta t}
\]

where the displacements \((\alpha, \beta, \gamma)\) are evaluated from the three-dimensional trajectory (Fig. 2)

\[
\alpha(X, Y, Z, t) = \Delta t SU(X - \alpha, Y - \beta, Z - \gamma, t)
\beta(X, Y, Z, t) = \Delta t SV(X - \alpha, Y - \beta, Z - \gamma, t)
\gamma(X, Y, Z, t) = \Delta t W(X - \alpha, Y - \beta, Z - \gamma, t)
\]

Clearly (15) represents a set of three coupled non-linear equations for the displacement in three dimensions; these equations are solved iteratively. Both (14) and (15) require evaluation of fields at upstream positions that do not in general coincide with grid points; the values are estimated based on some spatial interpolation. The choice of interpolation defines the accuracy of the semi-Lagrangian transport scheme. In practice a simple unconstrained cubic interpolations is used in the CRCM. This results in some damping of the smallest scale features in the transported fields and in some spurious ripples (over- and under-shoots) resulting in occasional occurrences of negative concentrations that are simply reset to an ad hoc minimum value. This approach does not garanty mass conservation either; in practice this is not perceived as a major problem in a nested model due to the large lateral-boundary fluxes. Optionally in CRCM a simple mass fixer can be applied to impose domain conservation arbitrarily; this fixer is only activated for water vapour and domain-averaged atmospheric mass (surface pressure).

Consistent with the semi-Lagrangian approach, all the terms are evaluated as spatial averages along the trajectories (15). Following the semi-implicit approach, linear dynamical terms \( L \) on the left-hand side of (1)–(7) are evaluated “implicitly” as \( L\{ \Psi \} \) (the curly parentheses \{\} being used to indicate the variable onto which the operator is applied) using quasi-centred time averages along the trajectory.

Fig. 2. Schematic representation of the semi-Lagrangian transport algorithm, shown in two-dimensions \((x, y)\) for simplicity. The grid point for which the trajectory over two time steps is computed (the arrival point of the trajectory) is noted by the lozenge; the triangle represents the departure point position in the grid; the circles represent the grid points employed to interpolate the value at the departure point when a cubic interpolation is used.
\[
\bar{\Psi} = \frac{(1 + \varepsilon)\Psi(X, Y, Z, t + \Delta t) + (1 - \varepsilon)\Psi(X - 2\alpha, Y - 2\beta, Z - 2\gamma, t - \Delta t)}{2}
\]  
(16a)

where the parameter \( \varepsilon \) controls the degree of off-centring of the average. With \( \varepsilon = 0 \), fast travelling waves are slowed down by the scheme, while with a non-zero value these waves are also damped; this is sometimes beneficial when using very large timesteps to reduce the spurious amplification of forced stationary waves with the scheme (e.g. [52]). In the CRCM with grid spacing of 45 km and a timestep of 15 min, the value \( \varepsilon = 0.05 \) is used in practice.

Following the semi-implicit approach, the remaining non-linear terms on the right-hand side are evaluated “explicitly”. Because a three-time-level approach is used, dynamical contributions \( D \) are evaluated at central time step, \( D\{\bar{\Psi}^{\text{tra}}(t)\} \) as a centred-explicit contribution, and parameterisation contributions \( P \) at past time step, \( P\{\bar{\Psi}^{\text{tra}}(t - \Delta t)\} \) as a forward-explicit contribution, where the spatial average along the trajectory is defined as

\[
\bar{\Psi}^{\text{tra}}(t) = \frac{(1 + \varepsilon)\Psi(X, Y, Z, t) + (1 - \varepsilon)\Psi(X - 2\alpha, Y - 2\beta, Z - 2\gamma, t)}{2}
\]  
(16b)

After substitution of Eqs. (14)–(16) in (1)–(7), the equations take the form of a set of coupled equations

\[
\Psi(X, Y, Z, t + \Delta t) - \frac{\Psi(X - 2\alpha, Y - 2\beta, Z - 2\gamma, t - \Delta t)}{2\Delta t} + \mathbf{L}\{\bar{\Psi}\} = D\{\bar{\Psi}^{\text{tra}}(t)\} + P\{\bar{\Psi}^{\text{tra}}(t - \Delta t)\}
\]  
(17)

Regrouping together on the left-hand side the terms involving variables at time \( (t + \Delta t) \) and on the right-hand side those involving variables at times \( (t) \) and \( (t - \Delta t) \), gives a system of equation of the form

\[
\mathbf{L}_1\{\Psi(X, Y, Z, t + \Delta t)\} = \mathbf{R}(X, Y, Z; t - \Delta t, t)
\]  
(18)

where \( \mathbf{L}_1 \) is a linear (by design) three-dimensional partial differential operator and \( \mathbf{R} \) regroups the explicit contributions from non-linear dynamics and parameterisations. The solution of (18) for dependent variables at time \( (t + \Delta t) \) involves the solution of a boundary-value elliptic equation, which may be written symbolically as a matrix inversion

\[
\Psi(X, Y, Z, t + \Delta t) = \mathbf{L}_1^{-1}\mathbf{R}(X, Y, Z; t - \Delta t, t)
\]  
(19a)

In the CRCM (19a) is solved iteratively by a variant of the alternating-direction implicit method, subject to lateral boundary values of the dependent variables provided at the perimeter of the computational domain.

### 2.3. Nesting formulation of the CRCM

A nested limited-area model as described above constitutes an ill-posed mathematical problem [93]. In essence the problem is over-specified because a set of first-order advective-type equations is solved with a second-order discretisation subject to two-point lateral boundary conditions. In practice problems may develop in the form of strong gradients along the lateral boundaries (analogous to mathematical boundary layers) as a result of mismatch between the solution of the regional model (the inner solution) and the imposed lateral boundary conditions (the outer solution). An approach was proposed by [20] that has since become traditional for nested models, consisting of adding “blending” terms and increased dissipation within a lateral sponge zone of a few grid points, in a ribbon along the perimeter of the domain, as a correction term:

\[
\Psi_f(X, Y, Z, t + \Delta t) = \mathbf{L}_1^{-1}\mathbf{R}(X, Y, Z; t - \Delta t, t)
\]  
(19b)

\[
\Psi(X, Y, Z, t + \Delta t) = \Psi_f(X, Y, Z, t + \Delta t) + \beta\Psi(X, Y, Z)(\Psi_0(X, Y, Z, t + \Delta t) - \Psi_f(X, Y, Z, t + \Delta t))
\]
\[
+ (-1)^\gamma \beta\Psi(X, Y, Z)\nabla^2\Psi(\Psi_0(X, Y, Z, t + \Delta t) - \Psi_f(X, Y, Z, t + \Delta t))
\]  
(20a)

where \( \Psi_0 \) is the value of the driving (nesting) data. The relaxation and diffusion coefficients can vary with dependent variables and their magnitude usually varies from their largest value at the boundary to zero through the sponge zone towards the interior of the domain. With additional terms in (20a), the mathematical problem is well posed, although modified from the original field equations. This procedure was shown by [110] to be an acceptable, simple and pragmatic approach to minimise the lateral boundary problems. It can be
shown that such a sponge both damps and reflects incoming waves; the damping effect is a function of the magnitude of the sponge coefficient, while the internal reflection is a function of its gradient. So it is advantageous to make the sponge both strong at the boundary and wide, which of course increases the overhead associated with this artificial region of the computational domain. The sponge coefficients can also vary with height; when simulating vertically propagating internal waves, this option can be used to absorb incident waves energy before they reach the reflective upper boundary (e.g. [52]). In the CRCM no enhanced diffusion is used, \( \beta_y = 0 \), and the relaxation coefficient \( \gamma_x \) is unity at the boundary and decreases inward as a cosine square over the width of the sponge which is chosen as 10-grid points.

For some applications an additional nesting contribution known as large-scale nudging (e.g. [104,12,72]) is advantageous:

\[
\Psi(X, Y, Z, t + \Delta t) = \Psi_f(X, Y, Z, t + \Delta t) + \gamma_x(X, Y, Z)\left(\Psi_0(X, Y, Z, t + \Delta t) - \Psi_f(X, Y, Z, t + \Delta t)\right) \\
+ (-1)^n \beta_y(X, Y, Z)\nabla^2 \left(\Psi_0(X, Y, Z, t + \Delta t) - \Psi_f(X, Y, Z, t + \Delta t)\right) \\
+ \gamma_{(\Psi,k)}(Z)\mathcal{S}\left\{\Psi_0(X, Y, Z, t + \Delta t) - \Psi_f(X, Y, Z, t + \Delta t)\right\}
\]

(20b)

where \( \mathcal{S} \) is a low-pass filter. In the CRCM a discrete cosine transform is used [23] to separate large scales. The strength \( \gamma_{(\Psi,k)} \) of the nudging term is usually taken to be a function of horizontal scale \( k \), height \( Z \) and dependent variable \( \Psi \). With this technique the large-scale component of RCM fields are nudged towards the corresponding large-scale component of nesting fields, throughout the RCM computational domain; hence the large-scale information of the driving fields is fully used, unlike with the traditional lateral boundary nesting (20a). In applications of CRCM over large domains (larger than about 120 by 120 grid points with a 45-km grid mesh), large-scale nudging is applied, typically to the winds only, for scales larger than 1400 km, at heights above 5 km. The usefulness of this nesting technique will be discussed below.

The forecast timestep cycle is completed by application of a running time filter initially developed by Robert and documented in [7] to prevent the appearance of computational modes resulting in the separation of numerical solutions with the three-level scheme.

2.4. Parameterisation of subgrid-scale physical effects in the CRCM

The CRCM includes a set of complex modules for a detailed parameterisation of the ensemble effect of subgrid-scale processes, the contributions of which were noted by \( \left(\Psi_f, F_y, F_z, L, E\right) \) in (1)–(7). The particular parameterisation package implemented in CRCM allows for any vertical staggering of the dependent variables but not for horizontal staggering. Consequently horizontal wind components are first staggered, averaged as \( \bar{U}^h \) and \( \bar{V}^h \), before being fed to the parameterisations, and tendencies \( F_x \) and \( F_y \) resulting from the parameterisations are averaged as \( \bar{F}_x^h \) and \( \bar{F}_y^h \) before they are passed to the time stepping (19).

Initially the CRCM employed the exact same parameterisation package as the second-generation Canadian GCM [67], the coupled version of which is now referred to as CGCM2 [37]. The reasons behind this choice were purely pragmatic: knowledge of these parameterisations by this author and the desire for a high degree of consistency afforded between the CRCM and its natural driving CGCM. Briefly the parameterisations include all the minimum ingredients deemed necessary for a successful climate simulation: terrestrial and solar radiation including diurnal and seasonal cycles, parameterisation of subgrid-scale cumulus clouds, large-scale precipitation based on water vapour supersaturation, diagnostic cloud water and cloud cover – parameterised in terms of a simple function of local relative humidity and assuming maximum (or random) overlap depending upon whether cloud presence is diagnosed in adjacent layers (or not), turbulent vertical fluxes of momentum, heat and water vapour, subgrid-scale mountain-wave drag, surface energy balance based on force-restore method, and surface hydrology based on the so-called beautified bucket method. The land–surface properties such as background albedo, surface roughness, primary and secondary vegetations, and soil–water field capacity are specified as spatially varying fields. There are four single-level dependent variables over land surfaces: soil temperature, liquid and frozen ground water contents, and snow water-equivalent. The latter affects the surface albedo and heat transfer properties. The reader is referred to [67] for an extensive description of the CGCM2 parameterisations and to [17] for a summary of their initial implementation in CRCM.
Some changes were later made to the CRCM parameterisations, as deemed required due to the increased resolution of CRCM (45-km grid mesh) compared to CGCM2 (T32). The moist convective adjustment scheme was replaced initially by the deep convection scheme of Kain and Fritsch [57] in [63], and later by that of Bechtold [9] in [64]. The cloud-onset function was altered to reduce the excessive cloudiness noted in earlier simulations of CRCM. The prognostic calculation of land–surface temperature was implemented with a backward-implicit scheme to suppress the occasional numerical instabilities associated with the original forward-in-time scheme, as documented in [45]. An interactive mixed-layer lake model has been developed by [49] to simulate the time-dependent evolution of surface water temperature and ice coverage for the lakes that are not resolved by the CGCM, such as the Great Lakes of North America.

Recently the CGCM2 large values of spatially varying soil water capacity have been decreased and set to a constant value of 10 cm in CRCM [81], which improves the representation of summertime surface fluxes (evaporation and precipitation) over land and shortens the time to freeze the ground in the autumn, a problem that had been noted to unrealistically delay the onset of snow cover. The critical “snow masking depth” at which surface elements are assumed to be covered by snow – and the surface albedo changes from the albedo of land to that of snow – has been changed from the spatially varying field used in CGCM2 to a constant value of 3.0 m over land surfaces other than tundra and swamp. Changes have also been made to the solar radiative heating to include more spectral bands and an improved representation of the water vapour continuum, which results in an increase in atmospheric absorption. Diagnostic clouds properties have been changed to increase reflection of incident solar radiation. Turbulent vertical mixing in the boundary layer has been modified to include non-local mixing of heat and moisture under statically unstable conditions. More details can be found in [81].

A regional ocean-ice model is also being coupled with CRCM for application in the Gulf of St. Lawrence [33] and Hudson Bay [87]. Most recently a version of CRCM incorporates the parameterisation package developed for CGCM3, which includes the Canadian land–surface scheme CLASS ([102,103]).

Lateral diffusion is implemented in terrain-following coordinates via a quasi-isotropic two-dimensional bi-harmonic ($V^4$) diffusion, conveniently formulated by the successive application, in alternate directions, of one-dimensional 3-point smoothing and un-smoothing operators (see Appendix of [63]). The constant of diffusion is currently very small, corresponding to a damping factor of 5% per time step for the shortest resolved scales in a 45-km mesh model.

2.5. Formulation issues with the CRCM

The following section discusses a number of issues relating to the formulation of an RCM that have arisen during or since the development of CRCM. This section is motivated by the desire to provide a personal testimony of some of the issues that confront any modeller in the development of an operational model. The following points are discussed: the merit of the non-hydrostatic field equations, the semi-implicit and semi-Lagrangian time-marching schemes, the use of time splitting for incorporating the diabatic contributions to the time tendencies, considerations of interpolations for the lateral boundary conditions, choices of physical parameterisation package, and upgrading the software for taking advantage of upcoming computer architectures.

2.5.1. Fully elastic non-hydrostatic Euler equations

At resolutions that can be afforded today for operational use of RCMs for climate applications, non-hydrostatic effects are unimportant. However, with the foreseen rise in computing power, CGCMs will continue to increase their resolution and consequently the application of RCMs will focus on finer scales. It is possible that within the next decade RCMs will access scales at which non-hydrostatic effects cannot safely be neglected. It has been argued by [96,62] that fully elastic non-hydrostatic Euler equations could be used efficiently, without appreciable computational overhead compared to the hydrostatic equations, for large-scale applications when solved by the semi-implicit semi-Lagrangian time-marching scheme.

In CRCM the Euler equations are cast in terrain-following scaled-height coordinates that redress toward horizontal surfaces linearly with height. An adaptation of this coordinate has been developed by [88] that allows for faster straightening of the coordinate with height, for reduced numerical truncation error, which is favourable in the presence of steep orography (e.g. [48]).
Following the seminal work of Eliassen [31], pressure has been used almost universally as the vertical coordinate for large-scale meteorological models based on the hydrostatic field equations; non-hydrostatic models, however, have usually retained height as vertical coordinate. The formalism of the Euler equations has been extended to hydrostatic-pressure vertical coordinates [60], so that there is a close resemblance between the ensuing non-hydrostatic equations and the hydrostatic equations in pressure coordinate. Several non-hydrostatic models now use this formulation, e.g. the French ALADIN model ([13,10]), the USA Eta model ([42]) and the Canadian GEM model ([111]).

2.5.2. Semi-implicit marching scheme

The use of the semi-implicit time scheme increases computational efficiency by permitting to lengthen the timestep while retaining numerical stability. Stability is achieved at the expense of slowing down (and optionally damping) the fast moving elastic and gravity waves, which may partly falsify the wave energy propagation. Large spatial scales, however, are dominated by Rossby and slow gravity waves for which larger timesteps can safely be used (e.g. [97]). It must be noted that care must be exerted in the choice of dependent variables for the linearisation of the equations in the semi-implicit scheme, as demonstrated by [10] for the particular case of the Euler equations in hydrostatic-pressure vertical coordinates.

2.5.3. Semi-Lagrangian transport scheme

The use of semi-Lagrangian transport scheme allows further lengthening of the timestep, which is no more constrained by the Courant stability criterion for advection, but rather chosen by considerations of the desired accuracy level. In addition to its computational efficiency, the semi-Lagrangian transport scheme has little numerical dispersion, which results in a reduction of spurious ripples in the distribution of transported substances, compared to conventional unconstrained Eulerian transport schemes (e.g. [75,76]). The interpolations required in the semi-Lagrangian scheme result in some damping of the small scales, which could be problematic for quasi-inviscid applications such as orographic internal gravity waves. It has been shown, however, that even simple cubic interpolation provided sufficient precision for an accurate simulation of mountain waves ([79,52]). Simple interpolation, however, does not guaranty monotonicity nor positive definiteness; interpolation constraints can be implemented to ensure these properties (e.g. [109,92]). Similarly the traditional implementation of the interpolative semi-Lagrangian scheme does not guaranty conservation of the transported quantity; finite-volume approaches can be adopted that ensure conservation at the expense of some additional complexity (e.g. [61]).

It must be borne in mind that when the development of the CRCM began some fifteen years ago, the semi-Lagrangian transport had not yet gained wide acceptance (e.g. [75,76,79,52,8]), and several scientists were sceptics about its suitability for climate models. Experience has shown, however, that this marching scheme is competitive in its performance and that it can afford substantial computational savings. For example, a semi-implicit Eulerian spectral CGCM with a triangular truncation at 32 spherical harmonics (T32) – corresponding to roughly 600 km grid mesh – requires a 20-min timestep, while a 45-km version of CRCM runs stably with a 15-min timestep despite more than an order of magnitude higher resolution (e.g. [63,64,81]).

The use of two-time-level semi-implicit semi-Lagrangian marching scheme (rather than three-time-level as used in CRCM) has several attractive features. It can be more economical, by allowing the use of half as many time steps for a given simulated time, although each time step is more costly due to required extra iterations ([111]). It also allows more naturally for the implementation of monotonic conservative transport schemes (e.g. [92]).

2.5.4. Time splitting of diabatic contributions

In the initial phases of the CRCM development, time splitting with respect to processes was used to implement the contributions of physical parameterisation \((F_X, F_Y, F_Z, L, E)\) in \(R\) in (18). Hence (18) was first solved without these terms, and their contributions were added as a second, corrective step. This procedure led to noise generation in regions of intense release of latent heat; it was noted that neglecting the \(L\) contribution in (5) prevented this problem. Later some ways of including the diabatic contribution in (5) was proposed by [99] (see also the penultimate of Section 2b in [63]). Eventually the culprit was identified by [15] to be
the time splitting with respect to processes; the problems disappeared with the proper solution of the elliptic equation including the parameterisation contributions as described in (18).

2.5.5. Lateral boundary data interpolations

The mismatch between the high-resolution RCM and the low-resolution lateral boundary condition (LBC) data, either from CGCMs or reanalyses, is a source of concerns as it can potentially generate numerical artefacts. In addition the LBC data is normally on different grid: either latitude – longitude – pressure for reanalyses data or Gaussian latitude – longitude – hybrid sigma for CGCM data. In the CRCM, the LBC data is interpolated in three dimensions from its original grid onto the RCM grid. As an example of artefact resulting from this data massaging, the separate interpolation of temperature and specific humidity can create spurious super-saturation in the interpolated variables even when the original variables are sub-saturated; in the CRCM this is now avoided by performing the interpolation on the relative humidity variable. As orographic heights usually differ between low- and high-resolution datasets, the lowest boundary of an RCM usually differs from that of the driving data, with repercussions on the placement of computational levels in the vertical; it is then impossible to ensure that lateral boundary fluxes of relevant physical quantities are preserved. Some RCMs choose to define their topographic height as a varying blend of the two definitions of topography near the lateral boundary to lessen this difficulty (e.g. [56]).

There may exist several other inconsistencies between an RCM and the driving data. In the CRCM the vertical velocity \( w \) (not \( \omega = dp/dt \)) is required at the lateral boundary; as this variable is not available in driving data, the crude assumption \( w = 0 \) is assumed at the LBC, and it is hoped that the sponge zone will serve to lessen the detrimental effects of this inconsistent value. Other inconsistencies may occur when an RCM does not employ the same numerical formulation or subgrid-scale physical parameterisation as the model that generated the LBC, since some variables may not be available for defining the LBC of the RCM. For example, cloud water content may not exist as a prognostic variable in a CGCM but may be required in an RCM; often then the LBC is assumed to be cloud free, which exacerbates the spin-up for this variable.

2.5.6. Parameterisations

In the historical development of RCMs, the parameterisation of the ensemble effect of subgrid-scale physical processes often proceeds from the adaptation of existing parameterisation packages either from a numerical weather prediction or mesoscale research model (as for the initial NCAR RCM) or from a CGCM (as for the Canadian RCM described above). The advantage of the former approach is that the parameterisations are adapted to the target resolution of RCMs; the disadvantage is that they may not have been developed with due concerns for climate sensitivity, long-term stability and computational affordability that are required for climate application. The advantage of the latter approach is that the parameterisations of an RCM are compatible with its “parent” CGCM, which eases the nesting process; the disadvantage is that adjustments may be needed to account for the resolution differences between RCMs and CGCMs.

The experience with the CRCM has shown that the use of the same parameterisations in a high-resolution RCM as in a coarse-resolution CGCM is not sufficient to guaranty consistency (e.g. [65]): the same code does not imply the same forcing, owing to its different behaviour under different resolution. This is probably true too for numerical discretisation algorithms that would not perform equivalently at different resolutions. What is needed is a parameterisation package that is “suitable” to the resolution of each model. Lately some centres have developed dynamical kernels and physical parameterisations that are suited for a wide range of resolutions and that can be applied for both numerical weather prediction and climate simulations (e.g. [82]).

2.5.7. Parallel computer code

The CRCM was programmed for single-processor vector or scalar computers. With the current trend towards massively parallel computer architectures, there is a pressing need to upgrade the CRCM to take advantage of available computer platforms. The next generation of CRCM (version 5) now under development, is based on the GEM kernel developed at the Meteorological Service of Canada ([111]), with a LAM option in addition to the standard uniform and stretched-grid configurations.
3. Applications and coordinated RCM experiments

Inspired by the success of coordinated global climate modelling projects such as the Atmospheric Model Intercomparison Project (AMIP, [1,43]) and the Coupled Model Intercomparison Project (CMIP, [19,71]), several coordinated projects have developed involving collaborations between regional climate modelling groups.

3.1. European projects Mercure, Prudence and Ensemble

In Europe the project Modelling European Regional Climate, Understanding and Reducing Errors (MERCURE; 1997–2000), a project financed by the European Commission under contract ENV4-CT97-0485 aimed at identifying the strengths and weaknesses of RCM simulations nested by atmospheric analyses. It paved the way to the project Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects (PRUDENCE; 2001–2004; [83,18]) that investigated the projected climate changes over Europe for 2071–2100. An important goal of the project was to analyse the projections of these models with respect to climate impact studies. PRUDENCE involved some 21 groups and a hierarchy of models were used: three CGCMs, four atmosphere-only GCMs (including a stretched-grid model) and eight RCMs with grid meshes of 50 and 20 km; some of these models were employed to simulate as many as 3-member ensembles. A central element of PRUDENCE was the assessment of uncertainty. Comparing various RCMs’ climate-change projections over Europe allowed contrasting the uncertainty associated with four aspects: (i) natural variability and the limited-time sample, (ii) greenhouse gases (GHG) emissions and concentrations scenarios, (iii) the choice of driving GCM atmospheric and oceanic boundary conditions and (iv) the RCM formulation. For temperature, the most important factors were the GHG concentration and the driving GCM, followed by the RCM formulation, with natural variability as the least important factor. For precipitation, the driving GCM was most important in winter, but the RCM formulation was most important in summer, with the other factors of lesser importance. An example of the results produced by PRUDENCE is shown in Fig. 3 for the projected relative change in surface air temperature. The follow-up project ENSEMBLES (2004–2009, [32]) aims to produce for the first time an objective probabilistic estimate of uncertainty in future climate at the seasonal to decadal and longer timescales, using an ensemble climate prediction system based on state-of-the-art high-resolution global and regional earth system models, validated against quality-controlled observational datasets for Europe.

3.2. Asian project RMIP

In Asia, the Regional Model Intercomparison Project (project RMIP; [86,40]) has been established in 1999 to evaluate and improve RCM simulations for East Asian monsoonal region, where there is high variability in both space and time. RMIP operates under joint support of the Asia-Pacific Network (APN) for global-change research, the global-change SysTem for Analysis, Research and Training (START) Regional Center for Temperate East Asia (RC-TEA), the Chinese Academy of Sciences (CAS) and several participating nations. The project currently involves ten research groups from Australia, China, Japan, South Korea and the USA, as well as scientists from India, Italy, Mongolia, North Korea and Russia. RMIP has three simulation phases, ranging from the annual cycle to investigate the skill in simulating the monsoon behaviour, a decade to examine the skill in simulating the characteristics of climate, and the projection of climate for the 21st century, involving nesting RCMs with CGCMs. The RMIP implementation design is reviewed in [40] with some initial results from the first phase.

3.3. North American projects PIRCS and NARCCAP

In North American, the Project to Intercompare Regional Climate Simulations (project PIRCS, [80,95,2]) was initiated in 1994 to offer a standardised framework for the validation of RCMs driven by atmospheric reanalyses over continental USA. This project paved the way to the coordinated North American Regional Climate Change Assessment Program (NARCCAP; 2005–2008; [73,70]) that will systematically investigate...
Fig. 3. Wintertime screen-level temperature change (°C) following an enhanced greenhouse-gases scenario, as simulated by various models participating in PRUDENCE (figures kindly provided by Dr. Jens Hesselbjerg-Christensen, Head of Programme, Danish Climate Centre, Danish Meteorological Institute). The following codes identify the types of models: “RM” refers to regional models (50-km grid, unless otherwise indicated), “AM” to atmospheric GCMs of intermediate resolution and “CM” to coupled atmosphere-ocean GCMs of low resolution; “CNRM” refers to the French variable-resolution atmospheric GCM. Labels follow one of two formats: GCM_Name for global models and RCM_Name/GCM_Name for regional models, in which case GCM_Name refers to the GCM that provided the lateral and sea-surface boundary conditions to the RCM. “Ensemble” (lower left-hand corner) refers to the ensemble-mean of the following nine regional models driven by HadAM3H-simulated data: HadRM3H, ETHRM, GKSSRM, ICTPRM, KNMIRM, MPIRM, UCMRM, SMHIRM and DMIRM.
the uncertainties in regional-scale high-resolution climate-change projections using six RCMs with grid
meshes of 50 km, two high-resolution atmospheric GCMs and four coupled GCMs, over a domain covering
the conterminous USA, part of Mexico and most of Canada. The recommended RCM domain for NARC-
CAP is shown in Fig. 4.

3.4. Arctic region projects ARCMIP and GLIMPSE

Initiated in 1999, the Arctic Regional Climate Model Intercomparison Project (ARCMIP; [6]) aims to
improve the simulation of the Arctic climate by atmospheric GCMs and RCMs driven by reanalyses.
Model-simulated data are compared and evaluated using observations from satellites, in situ measurements
and field experiments. ARCMIP has been organised under the auspices of the World Climate Research Pro-
gramme (WCRP) Global Energy and Water Experiment (GEWEX) Cloud System Studies Working Group on
Polar Clouds and the ACSYS Numerical Experimentation Group, with funding from the International Arctic
Research Consortium and the European Union project Global Implications of Arctic Climate Process and
Feedbacks (GLIMPSE). Eight different RCMs were integrated over a common domain covering the Beaufort/Chukchi Seas using boundary conditions from ECMWF ([85]). The results show considerable scatter
among the different RCM simulations, indicating rather weak control exerted by LBC despite relatively small
domains. At the same time, RCMs share common biases in temperature and humidity profiles, being too cold
in the low levels in the transition seasons and too warm elsewhere, too dry in the near surface layers and too
wet in the free troposphere. These biases and the scatter amongst models are attributed mainly to parameteri-
sations, which highlights the limitations of current Arctic climate simulations.

Fig. 4. Recommended domain for RCMs participating in North American NARCCAP project. Shown here is the Climate Research Unit (CRU; New et al. [74]) analysed field of precipitation for February 1979.
3.5. Transferability project ICTS

A Transferability Working Group (TWG; [101]) is just being launched, under the auspices of the GEWEX Hydrometeorology Panel (GHP). While most model intercomparison projects are concerned with the application of multiple models to a single region, the TWG aims to evaluate the RCMs contemporary-climate capabilities and challenges. The goal is to evaluate transferability of regional climate models and their components from “native” to other “non-native” regions. Such a project is an antidote against over-tuning of regional models for a specific region. In the Inter Continental-Scale Experiment Transferability Study (ICTS; [54]), RCMs from different Continental Scale Experiments (CSEs) will be transferred from their “home” CSE to other CSEs involved in GEWEX.

3.6. Seasonal prediction

RCMs are also applied to downscale extended-range forecasts and seasonal to interannual climate prediction made with coarse-resolution global models; this is a rapidly growing application of RCMs. Global seasonal forecasts show some skill in predicting planetary-scale anomalies, particularly in the tropics. Coupling to an RCM may allow these anomalies to be translated into more detailed, local weather anomalies such as precipitation. Such use of RCMs can potentially be extremely valuable. To date, the approach has been tested over a variety of regions such as Europe ([30]), North America ([34]), South America ([94]), Asia ([53]), to name a few (see also the Seasonal Climate Prediction for Regional Scales (SCPRS, [89]) and the International Research Institute for Climate and Society (IRICS, [55]). The steps involved in a prediction of seasonal climate anomaly using an RCM–CGCM system consist of first running an ensemble of global simulations to forecast both global sea surface temperature (SST) and atmospheric fields, and then running an ensemble of RCM simulations over the selected region of interest, using the SST and driving large-scale fields from the corresponding CGCM simulations. At present, the weakest link in the regional-scale performance of seasonal prediction systems appears to be prediction of the SST evolution at the seasonal scale.

4. Current issues in regional climate modelling

Several current issues and applications in regional climate modelling have been discussed in [66]. In the following section some of the current hot topics in regional climate modelling are briefly summarised to provide a view of the current understanding of such issues as the process of dynamical downscaling, the spin-up of fine-scale details from low-resolution driving fields, the means of driving a nested model, the downscaling skill, and the added value that can be expected from an RCM.

4.1. Deterministic predictability vs. dynamical downscaling

LAMs were initially designed for short-range numerical weather prediction and mesoscale modelling. For such applications, the best high-resolution analyses are used to initialise the LAM forecast. Coarse-mesh global model forecast fields are used to supply the required time-dependent lateral boundary conditions (LBC) for operational applications; for hind-cast research applications, the LBC can come from reanalyses. In either case it is recognised that LAM forecasts have limited time validity: for typical LAM resolution and domain size, 48 h is often the validity time limit. This stems in part from the contamination of the interior solution by the propagation of low-resolution information from the LBC. The intrinsic predictability limits of small-scale features also contribute to limit the time validity, as shown by [21]. On the other hand, the statistics of the modelled flow are by-and-large successfully reproduced ([22]).

Anthes et al. [3] paved the way to regional climate modelling by stating that: “Although the limit to predictability defines an upper bound on the period of time for which a forecast is useful, there are other potential uses for models where the predictability limit, and the growth of small, initial errors, is less relevant. These uses include ... studies of climatological nature. ... GCM are well known examples. ... a model may have relatively low skill in simulating phenomena of small spatial scale on individual episodes, but ... may be useful for climatological stud-
ies. Another use is nesting of high-resolution regional model with GCM (RCM). While large uncertainties might exist with any one simulation because of small-scale nature of phenomena resolved and the limits of predictability, the statistics of the ensemble could be meaningful and useful...” The earliest RCM ([28]) simulated regional climate through a series of short time slices, i.e. the LAM was only integrated for a few days of simulated time and it was periodically re-initialised with CGCM data. This proved, however, to be unnecessary (e.g. [46]) and nowadays RCMs are often integrated for decades of simulated time without reinitialisation.

It is now widely recognised that high-resolution nested models, equipped with suitable parameterisation of subgrid-scale physical effects and resolving physiographic details, and driven at their lateral boundaries by time-dependent large-scale atmospheric data, will generate meteorologically coherent small-scale features: this is referred to as “dynamical downscaling”. As an example Fig. 5 compares an instantaneous field of cloud cover simulated by a coarse-mesh GCM and a fine-mesh RCM driven by the GCM. The mechanisms responsible for the spontaneous generation of fine-scale details from large-scale LBC include local surface forcing (such as topography, land–surface variations including small lakes, land–sea contrasts, coastal outline), but also purely dynamical effects such as non-linear scale interactions (stretching, stirring and folding) and hydrodynamical instabilities due to the better resolution of gradients. The main aim of RCMs in the context of climate-change projections is to add valuable fine-scale details to, but keep, the large-scale circulation of CGCM integrations. For example, distributions of extreme daily precipitation can become more realistic as grid meshes become finer than current operational CGCMs. One would expect that the trustworthy scales of the driving data to be maintained in a nested model simulation. This should not be interpreted to mean that RCM-simulated data, once aggregated onto the coarse mesh of the CGCM providing the nesting data, should be identical to the CGCM data. Indeed an RCM may resolve physical processes that are not resolvable due to the low resolution of a CGCM. An example is provided in Fig. 6 for the climatological precipitation simulated by the CRCM and CGCM; the high topography of the Rocky Mountains is better resolved in the CRCM, which produces a rain shadow downstream that extends far downstream over the North American Prairies, from the North-West Territories to Texas, in better agreement with analysis of observations.

On the other hand, for the largest resolvable scales in the limited domain of an RCM, there is little reason to expect that the free atmospheric circulation used as LBC to an RCM should be modified by the RCM; and if it were, this would lead to inconsistencies that could result in major LBC difficulties.

Fig. 5. Instantaneous field of clouds simulated by CGCM2 with equivalent grid mesh of 450 km (left panel) and CRCM with grid mesh of 45 km (right panel) (superimposed on the GCM-simulated field, showing the frame of the RCM domain) (figures kindly provided by Dr. Daniel Caya, Chief, Climate Simulation Team, Ouranos Consortium).
4.2. Spin-up

The generation of fine-scale features in an RCM simulation initialised from coarse-resolution fields is not instantaneous, and their spin-up time must be carefully assessed in designing RCMs’ experiments. The spin-up occurs rather fast (1–2 days) for most atmospheric fields ([21]), but it may be much longer for some land–surface processes (such as deep soil hydrology); the spin-up or adjustment period must be removed from the simulation archive before computing climate statistics. More important for climate applications, however, is the spatial spin-up, i.e. the distance from the lateral boundary over which the development of fine-scale features, in an RCM simulation driven at its lateral boundary from coarse-resolution fields, takes place before they

Fig. 6. Twenty-year average of simulated precipitation for the months of December, January and February (in mm day$^{-1}$) under current GHG concentration: top left is the simulation by CGCM2 and top right by version 3.6.3 of CRCM driven by CGCM2 (figures kindly provided by Drs. David Plummer, Research Scientist, Meteorological Service of Canada, and Daniel Caya, Chief, Climate Simulation Team, Ouranos Consortium). Bottom panel shows the CRU (version 2) analysis of observations (over land only).
reach their equilibrium amplitudes. The width of this spin-up or adjustment region is not well defined, but in practice it is certainly larger than the buffer zone. The spin-up width appears to depend on the flow speed (and hence it is much larger in the upper troposphere where winds are strong) and it is wider on the inflow side of the domain. The spin-up width is also a function of the jump between the resolution nesting data and the RCM grid mesh. Although the resolution jump is often between 5 and 10, successful RCM simulations have been performed with much larger jumps (O. Christensen, DMI, personal communication); the key to the success appears to lie in the requirement that the RCM domain spans several grid meshes of the driving data grid mesh. The spin-up width also depends on the strength of the acting downscaling processes, both free (non-linear processes and hydrodynamic instabilities) and forced (by surface processes).

The “physically relevant” region of an RCM domain excludes the spin-up region; it cannot be defined a priori, but rather by careful inspection of the simulation statistics and their sensitivity to changes in domain size. On the one hand, overly small RCM domains (such as 50 by 50 grid points) are computationally cheap but they may not have a physically relevant region due to insufficient space for spinning-up fine scales. On the other hand, with very large domains (larger than some 150 by 150 grid points), the LBC sometimes do not appear to exert sufficient control on the solution developed by an RCM in the interior of its domain ([56]). This point is further developed below.

4.3. Nesting and lateral boundary conditions

The nesting technique is not without potential difficulties as mentioned earlier; [106] discussed the lateral boundary problems and their impact on the simulation in the interior of the domain. An additional issue is the choice of the size of the RCM computational domain.

It has been noted by [26] that the control exerted by LBC on the RCM response varies with season, being stronger in winter for mid-latitude domains. In summer and with very large regional domains, RCM modellers have long noted that the LBC sometimes do not appear to exert sufficient control on the solution developed by an RCM in the interior of its domain; the phenomenon is known as “intermittent divergence in phase space” (IDPS). The occurrence of IDPS results in large gradient that develop particularly at the exit region of the domain, i.e. downstream of the dominant flow; in most severe cases, the RCM solution may develop rather un-meteorological flow to satisfy the LBC. The application of the “large-scale nudging” technique ([104,12,72]) follows (20b): the large-scale components of RCM fields are replaced by (or nudged towards) the corresponding large-scale components of the driving fields; this is done within the whole RCM computational domain. Hence the large-scale information of the driving fields is fully used, unlike with the traditional nesting (20a). The application of large-scale nudging can control and even suppress the occurrence of IDPS ([107,108]) and greatly reduces the occurrence of large gradients at the outflow region of the RCM domain. When driving an RCM with reanalyses, the operation of the large-scale nudging may be considered a kind of “sub-optimal data assimilation” system ([104]). It has been argued [65] that large-scale nudging may be considered a necessary large-scale closure to account for the finite size of an RCM domain, analogous to small-scale closure to account for the finite resolution of any model. The drawback of large-scale nudging is that it may hide systematic biases of an RCM.

4.4. Downscaling skill

As mentioned earlier, high-resolution RCMs do develop fine-scale details when driven at their lateral boundaries by low-resolution atmospheric information. These details appear quite realistic. For example, simulated clouds exhibit filaments similar to those observed on satellite pictures; the distribution of simulated precipitation is far more realistic than that of coarse-mesh GCMs. But the question remains: Beyond appearing realistic and plausible, are the simulated fine-scale features “correct” in some sense? Confidence in RCMs’ simulations and climate-change projections can only be gained through an objective assessment of RCMs’ skill at dynamically downscaling low-resolution information that drives them. One could think of comparing the fine scales simulated by an RCM with observations, but this approach faces several difficulties. The resolution of most climate data analyses is too coarse to resolve the fine scales of interest, and special high-density observations campaigns are often not sufficiently long to establish a climate basis. Furthermore, all models contain
approximations that result in imperfect simulations, and RCMs are no exception; it would be next to impossible to isolate errors arising from the nesting approach of RCMs from others that are the results of modelling approximations as in all models.

The key issue relating to regional climate modelling is whether the climate of a high-resolution RCM simulation driven by low-resolution GCM, is equivalent to the climate of a reference simulation with a GCM with equivalent high resolution. This process is shown in Fig. 7. This is a kind of “perfect-prognosis” approach since a model simulation (from an RCM) is compared to another model simulation (from a GCM). Given the fact, however, that RCMs are first-order sensitive to LBC and that the large scales simulated by the low- and high-resolution GCMs may differ, the comparison of the RCM with high-resolution GCM would not only reflect errors resulting from the nesting method but also the differences in LBC due to the use of low-resolution GCM.

An experimental protocol has been designed to isolate the errors that are specific to the nesting method, independently from errors from the large scales in the LBC used to drive the RCM: the “Big-Brother Experiment” (BBE). The approach consists again in performing first a high-resolution GCM simulation (named the Big Brother, BB) that will serve the dual purposes of (1) defining the reference against which the RCM simulation will be compared and (2) defining the LBC needed to drive the RCM; the RCM simulation is named the Little Brother. [The method is named “Big Brother’ after mentoring programmes providing young people with positive role models who offer support.] As shown in Fig. 8, the LBC are defined by retaining only the large scales of the BB, filtering out fine scales in order to emulate the resolution of typical GCM but without any large-scale errors. With the BBE protocol, the differences in the climates simulated by the RCM and the reference GCM can be attributed unambiguously to the nesting method of the limited-area RCM.

A practical difficulty with either approaches shown in Figs. 7 and 8 is the computational cost of integrating the high-resolution GCM for climate time scales. In order to render the approach computationally more affordable, a “poor-man” version of the BBE has been designed. In this case the high-resolution GCM is replaced by a high-resolution large-domain RCM, as shown in Fig. 9; to the extent that the regional domain is very large, it is hoped that the results would approach those of a global model as on Fig. 8. The BBE protocol has been used for several experiments at UQAM ([24,25,4,29]), for the winter and summer seasons, and for domains over the East and West coast of North America. In general the LB succeeds to reproduce rather

Fig. 7. Flow chart of an idealised RCM validation experiment, comparing the climate simulated by an RCM driven by a low-resolution GCM against a reference climate obtained from a high-resolution GCM simulation.
well the climate statistics of the BB for all simulated fields, for both the large and small scales, and for stationary and transient eddies. These results are comforting in that they indicate dynamical downscaling skill for an RCM driven by low-resolution fields for the case in which the large scales are perfect. More recently [51] have
employed the BBE over a tropical domain to study the skill in seasonal prediction at NCEP; difficulties found in the reproduction of the small-scale precipitation remain largely unexplained and are currently a matter of investigation.

The relative success noted above with the BBE is a necessary but not sufficient test for RCMs to pass, and it should not detract from the fundamental issue that was described at the beginning of this subsection and in Fig. 7: can the climate simulated by a high-resolution RCM nested with LBC from a low-resolution GCM, replicate accurately the climate of a high-resolution GCM? In other words, What is the influence on the RCM-simulated climate of large-scale errors in LBC from imperfect coarse-resolution GCM? One could imagine that some large-scale errors resulting from the use of coarse resolution in a GCM may be partly corrected by an RCM; alternatively errors in the large-scale flow may be amplified when this flow interacts with fine-scale forcing. To investigate this issue, a variant on the BBE has been designed to study the sensitivity of RCM-simulated climate to large-scale errors in LBC: the “Imperfect Big-Brother Experiment” (IBBE, Fig. 10). In the IBBE protocol, the LBC are obtained from simulations of a coarser resolution RCM integrated over different (larger) domains. By varying resolution and domain size, controllable level of errors ensue; these errors are dynamically coherent and are typical of those of coarse-mesh GCMs. The study of [27] with the IBBE for the summer season over an Eastern North American domain indicate that the RCM is rather neutral to LBC errors: large-scale errors are not amplified nor reduced by the RCM. The RCM simulation of small scales is generally poorer in the presence of large-scale errors. An exception is in the case of strong localised surface forcing, which can accurately recreate part of the small scales.

Castro et al. [14] obtained results that are at variance with those obtained in the idealised experimental framework used here; for the particular case they considered, the CSU RAMS RCM did not retain the value of the large scales present in the global reanalysis used as LBC in their nested simulation. There is a need to extend the experimentations with the BBE protocol to other climatic regions, and in particular to Arctic domains where the degree of control exerted by LBC appears to be weaker ([84,85]).

Fig. 10. Flow chart of an “Imperfect Big-Brother Experiment”. Controlled errors are introduced in the Imperfect Big Brother used to drive the Little Brother in order to study the impact of lateral boundary errors on the climate simulated by the Little Brother. Verification is made against the perfect Big-Brother simulation.
4.5. Added value

The purpose of RCMs is to add details beyond the scales provided at the LBC from coarse-mesh GCMs or reanalyses. Hence it makes sense to seek for the added value of RCMs in the scales that are unresolved in the LBC (henceforth referred to as “small scales”) ([65,35,36]). The enhanced resolution of RCMs allows for a better description of mesoscale atmospheric dynamics and fine-scale surface forcing. Where surface forcing are strong, such as near mountains or coastal regions with sharp variation of surface properties, even the time-averaged fields can be substantially improved, especially in the low levels, by the increased resolution, and become substantially different from the driving data from coarser model. The spectral distribution of most atmospheric fields however is such that most of the time-stationary component of the variance is contained in the very large (planetary) spatial scales (in the free atmosphere, away from the surface). At smaller scales the transient-eddy variance usually dominates over that of the stationary component (e.g. [11]). Given that time-averaged fields are not in general the ideal place to seek the added value, except where intense localised forcing is acting, the added value of RCMs is likely to lie mostly in frequency distributions and high-order statistics, reflecting more intense or localised weather events such as intense precipitation events. The scale decomposition of fields (e.g. [23,35,36]) is a useful technique to isolate the added value of RCMs.

Some fields have much flatter spectra than those discussed above, for example precipitation or fields that are strongly influenced by surface forcing. [11] have performed a scale decomposition of the various terms in atmospheric water budget to isolate their respective contributions. This study reinforces the point about the relatively modest contribution of small scales to the time-mean water budget, and a suggestion that the added value of RCMs is contained mostly in the time variability, except again where there is strong localised forcing.

5. Conclusions

Regional climate modelling provides a computationally affordable means of achieving high resolution locally for several applications that would otherwise be out of reach for most research groups, save for the most powerful computing centres. The inherent limitations to the one-way nesting method, relating to the control exerted by lateral boundaries, the lack of two-way feedback and the spin-up of fine scales from low-resolution lateral boundary conditions, require careful attention by practitioners in order to obtain reasonable results and draw fully the expected added value from the method.

Acknowledgments

The author would like to pay tribute to those who have contributed to shape his professional carrier, his professors and senior colleagues: (in rough chronological order) Peter Zwack (RIP, 8 November 2005), Jacques Derome, Philip E. Merilees, George J. Boer, Norman A. McFarlane, W. Richard Peltier, Claude Girard, André J. Robert (RIP, 18 November 1993) and Istzat Zawadzki. The scientific contributions of Akio Arakawa, Akira Kasahara, Paul Queney and Bennert Machenhauer have also exerted a great influence on my research, for the most part unknowingly to them.

References

T.T. Warner, R.A. Peterson, R.E. Treadon, A tutorial on lateral conditions as a basic and potentially serious limitation to regional


