

## A Real-Time North American Forecast at 10-km Resolution with the Canadian MC2 Meso-LAM

S. J. THOMAS, M. DESGAGNÉ, AND R. BENOIT

*Recherche en Prévision Numérique, Environnement Canada, Dorval, Quebec, Canada*

(Manuscript received 9 March 1998, in final form 15 October 1998)

### ABSTRACT

The next generation of high-performance computers will be based on clusters of shared-memory symmetric multiprocessor (SMP) nodes interconnected by a low-latency, high-bandwidth network. In this paper, the parallel performance of the nonhydrostatic Mesoscale Compressible Community (MC2) limited-area atmospheric model on clusters of NEC SX-4 symmetric multiprocessor (SMP) nodes is presented. Several hybrid parallel-programming approaches are now possible with the SMP cluster SC-MC2 implementation based on internode MPI message-passing and intranode shared-memory tasking or threads. At total sustained execution rates of between 25 and 30 Gflop  $s^{-1}$  on single-node or multinode clusters, it is now possible for the first time ever to generate a 24–48-h real-time weather forecast over North America at 10-km resolution.

### 1. Introduction

A clear trend is emerging in the procurement of new supercomputers at national weather centers in Europe and North America. The majority of these centers have installed or are migrating toward distributed or distributed-shared-memory parallel computers (Dent and Mozdzynski 1997). The same can be said for climate research and ocean modeling groups around the world. The shared-memory parallel vector processors (PVP) in use at these centers for the past decade or more have now given way to scalable computer architectures. “Scalable” implies that the sustainable performance of a parallel scientific application program increases as the problem size increases with the number of processors. A more precise definition of scalability can be found in the text by Kumar et al. (1994). In the context of real-time weather forecasting, a 24–48-h weather forecast must be produced within a wall-clock time of approximately 1 h in order for the forecast to be useful. Thus, the problem is to solve the largest possible problem or equivalently use the highest spatial resolution over the largest possible geographic area to produce a forecast within in a fixed amount of time.

The overall structure or design of the next generation of supercomputers appears to be converging, and several software standards have also emerged. The basic machine architecture is now a collection of shared-memory

multiprocessor nodes interconnected by a low-latency, high-bandwidth network. Typically, a node consists of 1–128 superscalar RISC or vector processors in a tightly coupled symmetric multiprocessor (SMP) configuration. Peak processor performance will likely range from 0.5 to 8.0 Gflop  $s^{-1}$ . In 1993, the pioneering work of Woodward and a team of scientists from the University of Minnesota demonstrated the capabilities of SMP cluster supercomputing by achieving a sustained performance of 4.9 Gflop  $s^{-1}$  (Gflop = 1 billion floating point operations) on 16 Silicon Graphics (SGI) SMP nodes interconnected with a fiber-optic network in which each SMP node contained 20 R4400 superscalar MIPS processors (Woodward 1996). These results led SGI to introduce the Power Challenge Array as a commercial product. The U.S. Department of Energy’s Accelerated Strategic Computing Initiative (ASCI), in addition, has led to the announcement of an SMP cluster configuration based on the IBM SP-2 and RS/6000 microprocessor. The Fujitsu VPP 700 and Cray T3E distributed-memory massively parallel architectures can be classified also as SMP architectures, with one processor per node and up to 1024 nodes in the case of the T3E. Despite the fact that the peak instruction rate of pipelined RISC microprocessors continues to double approximately every 18 months, highly optimized codes can sustain usually no more than 15%–20% of peak (Hennessy and Patterson 1990). This situation may change as larger secondary cache memories become available; however, SMP clusters based on vector processors capable of sustaining 1 Gflop  $s^{-1}$  or higher are capable currently of 100 Gflop  $s^{-1}$  aggregate performance.

The message-passing interface (MPI) (Gropp et al.

---

*Corresponding author address:* Dr. Stephen Thomas, National Center for Atmospheric Research, Boulder, CO 80803.  
E-mail: thomas@ncar.ucar.edu

1994) has emerged as an international standard for distributed-memory programming, and the majority of weather centers are now developing numerical atmospheric models based on a subset of the Fortran 90 programming language in combination with MPI. An overview of the techniques used to build parallel regional weather models such as MM5, HIRLAM, ARPS, and the German DWD EM/DW is given in Baillie et al. (1997). Several different hybrid programming models are possible on SMP clusters. Internode communication can be based on MPI message passing across the interconnection network. Intranode parallelism can be based on shared-memory tasking mechanisms such as coarse-grain macrotasking or fine-grain loop-level microtasking. Alternatively, lightweight processes or POSIX threads may be viable for small numbers of processors. In this paper, we present the performance of the SMP cluster version of the MC2 atmospheric model based on MPI.

## 2. Nonhydrostatic MC2 model formulation

The nonhydrostatic Mesoscale Compressible Community (MC2) limited-area atmospheric model is used widely in Canada for high-resolution atmospheric research (Benoit et al. 1997). The model is an extension of a fully compressible limited-area model developed in the mid-1980s by Tanguay et al. (1990). The origins of MC2 can be traced to mesoscale and microscale models developed during the past two decades. In particular, Clark (1977) introduced the terrain-following vertical coordinate transformation of Gal-Chen and Somerville (1975) into an anelastic model for small-scale studies:

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

where  $h(X, Y)$  is the height of topography. The constant surfaces  $Z = 0$  and  $Z = H$  represent the bottom and top of the atmosphere, respectively. Following standard conventions, the Jacobian  $G$  and metric coefficients  $G^{ij}$  of the transformation are denoted

$$G = \frac{H - h}{H}, \quad G^{13} = \frac{1}{G} \left[ \frac{Z - H}{H} \right] \frac{\partial h}{\partial X},$$

$$G^{23} = \frac{1}{G} \left[ \frac{Z - H}{H} \right] \frac{\partial h}{\partial Y}.$$

Given a map projection at reference latitude  $\phi_0$ ,  $S = m^2$ ,  $m = (1 + \sin\phi_0)/(1 + \sin\phi)$ , and Coriolis parameter  $f = 2\Omega \sin\phi$ , the compressible governing equations in model  $\mathbf{X} = (X, Y, Z)$  coordinates then become

$$\begin{aligned} \frac{DU}{Dt} &= fV - K \frac{\partial S}{\partial X} - RT \left[ \frac{\partial q}{\partial X} + G^{13} \frac{\partial q}{\partial Z} \right] \\ &\quad - K_U(U - U_e) + F_U \\ \frac{DV}{Dt} &= -fU - K \frac{\partial S}{\partial Y} - RT \left[ \frac{\partial q}{\partial Y} + G^{23} \frac{\partial q}{\partial Z} \right] \\ &\quad - K_V(V - V_e) + F_V \\ \frac{Dw}{Dt} &= -g - \frac{RT}{G} \frac{\partial q}{\partial Z} - K_q(q - q_e) + F_w \\ \frac{DT}{Dt} &= \frac{RT}{c_p} \frac{Dq}{Dt} - K_\Theta \Pi(\Theta - \Theta_e) + \frac{Q}{c_p} \\ \frac{c_v}{c_p} \frac{Dq}{Dt} &= -S \left[ \frac{1}{G} \frac{\partial GU}{\partial X} + \frac{1}{G} \frac{\partial GV}{\partial Y} \right] - \frac{1}{G} \frac{\partial GW}{\partial Z} \\ &\quad - K_q(q - q_e) + \frac{Q}{c_p T} \\ \frac{DM}{Dt} &= F_M \\ \frac{DC}{Dt} &= F_C. \end{aligned} \quad (1)$$

Increased resolution in the planetary boundary layer is achieved by applying a second monotone mapping to  $Z$ , resulting in a generalized terrain-following coordinate with variable vertical resolution. The contravariant vertical velocity  $W$  is related to the covariant velocity components by the equation  $W = G^{-1}w + S(G^{13}U + G^{23}V)$ . Potential temperature is  $\Theta = Te^{-\kappa q}$ , and  $\Pi = (p/p_{00})^\kappa$  is the Exner function, where  $T = \Pi\Theta$ ,  $q = \ln(p/p_{00})$ , and  $p_{00} = 1000$  mb. Variables  $R$  and  $c_p$  are the gas constant and heat capacity for dry air at constant pressure:  $\kappa = R/c_p$ . Variables  $U$ ,  $V$ , and  $w$  are the wind images in projected  $(X, Y, z)$  coordinates, and  $g$  is the gravitational acceleration. The term  $K = (U^2 + V^2)/2$  is the pseudo-kinetic energy per unit mass. Momentum ( $F_U, F_V, F_w$ ), heat  $Q$ , moisture  $F_M$ , and liquid water content  $F_C$  sources or sinks are also included. Up to four different water species may be specified in the optional MC2 microphysics package developed by Kong and Yau (1997).

Lateral boundary conditions are derived either from a global model or as part of a nesting strategy within a limited-area domain. Terms of the form  $K_U(U - U_e)$  represent a Davies-type boundary relaxation scheme in which model variables computed on the interior of the computational domain are relaxed to environmental or driving model values such as  $U_e$  near the lateral boundaries (Davies 1976). Such a scheme is sometimes referred to as a gravity wave absorber, since the overall effect is to damp spurious waves reflections off the lateral boundaries. The compressible governing equations are discretized in three dimensions using centered finite

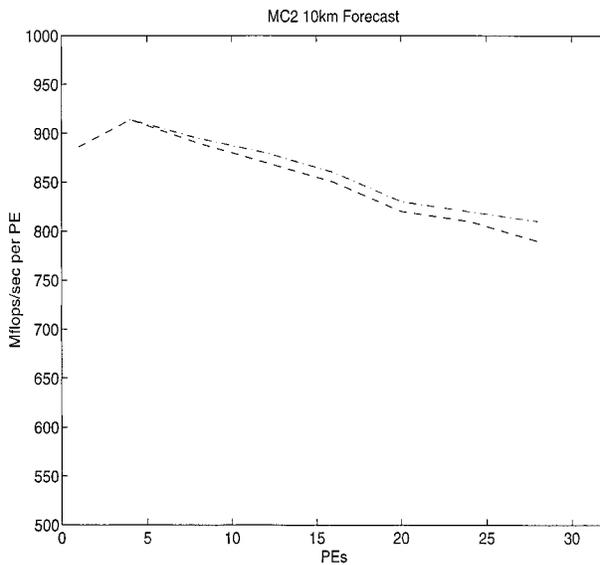


FIG. 1. MC2 performance on SX-4/32M: single node (bottom) vs multinode (top).

differences on a horizontal Arakawa “C” grid and vertical “B” grid. Resolution is uniform, with constant spacing ( $\Delta X$ ,  $\Delta Y$ ,  $\Delta Z$ ). To overcome the severe numerical stability constraints imposed by gravity and acoustic waves in explicit time-stepping schemes, a fully 3D semi-implicit, semi-Lagrangian time discretization is employed. The semi-implicit scheme results in a large, sparse, nonsymmetric system of equations to solve every time step for a log pressure perturbation  $q'$  about an isothermal hydrostatic basic state. To solve the elliptic problem arising in MC2, we have chosen to implement the Generalized Minimal Residual (GMRES) algorithm of Saad and Schultz (1986) (see Thomas et al. 1997). Skamarock et al. (1997) employ the closely related GCR algorithm of Eisenstat et al. (1983) in a semi-implicit formulation of a compressible model that also treats vertical terms  $\partial q/\partial Z$  implicitly. A more implicit formulation of the MC2 model along these lines is described in Thomas et al. (1998).

Krylov subspace methods for nonsymmetric problems are particularly well suited to a distributed-memory, message-passing model of computation, since the methods rely primarily on distributed matrix vector multiplication and an inner product implemented as a global reduction summation. There are now a variety of nonsymmetric Krylov solvers available, and these can be derived from different variational principles (Saad 1996). The convergence rate of a given solver is often problem dependent. Rapid convergence can be achieved only by finding a suitable preconditioner, and this is usually problem dependent. Preconditioners can be based on operator splittings such as the classical iterations (Richardson, Jacobi, SOR, ADI, etc.), or more sophisticated techniques such as multigrid can be employed. Rather than being viewed as competing solver

technology, the excellent convergence rates achieved by multigrid and, more generally, multilevel methods often can be improved by Krylov accelerators such as GMRES (see Smith et al. 1996).

### 3. Multinode parallel programming models

For limited-area atmospheric models such as MC2, the computational domain is discretized using  $N_x \times N_y \times N_z$  points on a rectangular tensor-product grid. Resolution is uniform in the horizontal direction, and the number of grid points in the vertical direction is typically one order of magnitude smaller than in the horizontal. For distributed-memory computation, a domain decomposition of the computational domain across a  $P = P_x \times P_y$  logical processor mesh is employed; thus, each processor (or node) would contain  $N_x/P_x \times N_y/P_y \times N_z$  points. For simplicity, we have assumed that  $N_x/P_x$  and  $N_y/P_y$  are integers; otherwise, the remaining points must be distributed evenly among processors using a load-balance data distribution, for example. Since the number of grid points in the vertical direction is typically one order of magnitude less than in the horizontal and there is also a strong vertical coupling in atmospheric models, partitioning in the vertical direction is generally avoided.

Given the domain decomposition described above, each subdomain contains all grid points in the vertical direction. To achieve the highest possible sustained instruction execution rate on multinode architectures, perhaps the simplest possible strategy is to assign each subdomain to a processor and then optimize the single-processor performance. Once the optimal subdomain dimensions are determined, ideally the execution rate should scale as a function of the problem size (i.e., global grid dimensions), and then, the performance will increase with the number of subdomains. Alternatively, for real-time weather forecasting, the wall-clock execution time should remain constant or grow slowly as the number of processors is increased for a fixed subdomain size (Skälin 1997). The subdomain size for vector processors will be determined by the optimal vector length, whereas for superscalar RISC processors, the subdomain size should be chosen so as to fit into the L2 secondary cache. An SMP node can be assigned fewer subdomains than available processors, and in this case, fine-grain, shared-memory-type parallelism can be exploited to increase the node execution rate. For example, the SX-4 supports traditional shared-memory macrotasking and loop-level microtasking on a single node. In addition, it provides both shared- and distributed-memory implementations of MPI/SX for both internode and intranode message passing. In the case of MC2, loop-level microtasking can be used in the vertical direction in order to make use of the additional processors beyond the number of subdomains assigned to the node. Therefore, to optimize performance on a particular SMP cluster architecture, the parameter space

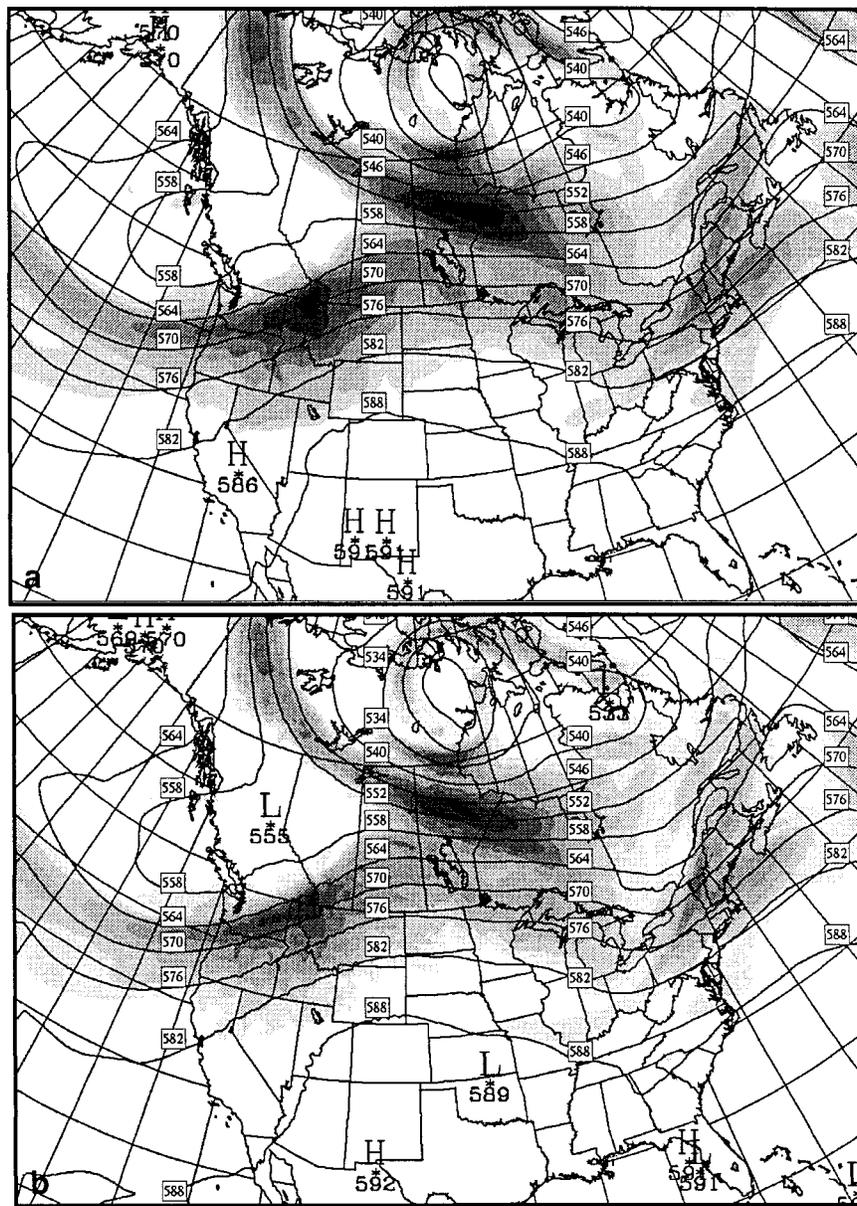


FIG. 2. North American forecast: 500-hPa 24-h forecast, valid at 0000 UTC on 27 June 1997, of geopotential heights (contoured every 6 dam) and wind speed (shades of gray) for (a) GEM 35 km and (b) MC2 10 km.

must be expanded to include the optimal number of subdomains assigned to a node.

We have created an intermediate software layer between the MC2 model and the MPI message-passing library to simplify data interchanges at subdomain boundaries. These routines form an application programmer interface (API) known as message-passing tools for structured grids (MSG). MSG provides a simplified programming model defined in terms of boundary exchanges on a distributed tensor-product grid. It is assumed that the computational domain is partitioned across a logical processor mesh such that every grid

point is assigned to one and only one processor. The computational domain (corresponding to a global array) is partitioned into subdomains (local arrays) assigned to different processors. The size and shape of the halo regions on each processor (containing data communicated from adjacent processors) can then be computed by MSG from the overlap at subdomain boundaries and boundary conditions for the problem domain in each of the three coordinate directions. According to our SMP programming model, a logical processor would correspond to one or more physical processors within an SMP node. In our simulations, we have found that the shared-

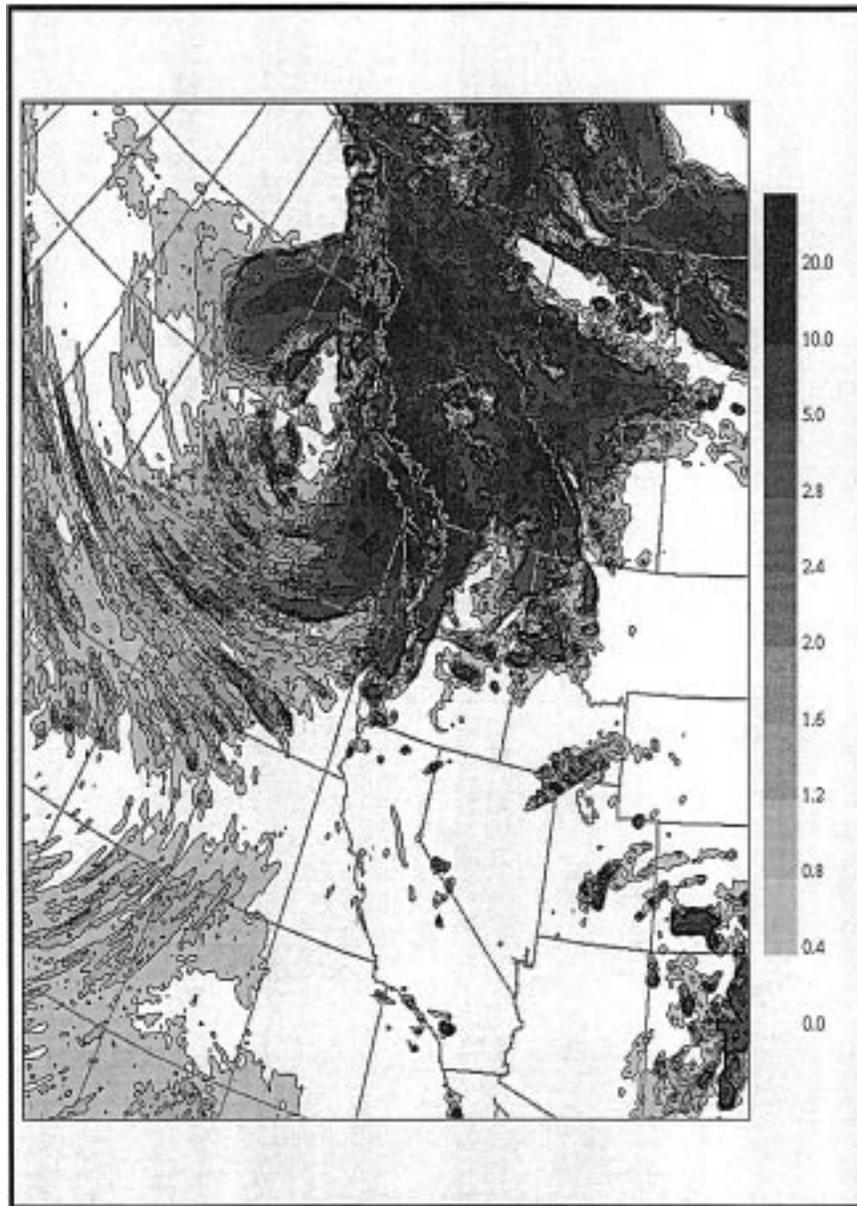


FIG. 3. MC2 North American 10-km forecast: 24-h forecast of total accumulation of precipitation, valid at 0000 UTC on 27 June 1997 (contoured every 0.2 mm).

memory implementation of MPI can be more efficient than macro- and microtasking within a single NEC SX-4/32 node given sufficiently large grain computation.

#### 4. The NEC SX-4 SMP cluster architecture

The multinode NEC SX-4 vector supercomputer is an SMP cluster-type architecture with up to a maximum of 16 nodes. A node can be configured with 4, 8, 16, or 32 processors. The designation SX-4H is given to multinode machines with an HiPPI channel interconnect, whereas, the SX-4M is based on the proprietary NEC IXS crossbar network with fiber-channel interface.

For example, we have benchmarked the SC-MC2 code on an SX-4/32M configuration consisting of two SX-4/16 nodes with IXS interconnect and on a single SX-4/32 SMP node. The SX-4 vector processor is based on low-power CMOS technology with a clock-cycle time of 8 ns (125 MHz). Each processor of the SX-4 contains a scalar unit and a vector unit. The scalar unit is a superscalar architecture. There are 128 64-bit scalar registers per cpu. The vector and scalar units support 32- and 64-bit operands. The scalar unit also supports 8-, 16-, and 128-bit operands. Each processor has hardware support for three different floating-point data formats: IEEE 754, Cray, and IBM. The SX-4 is based on 64-

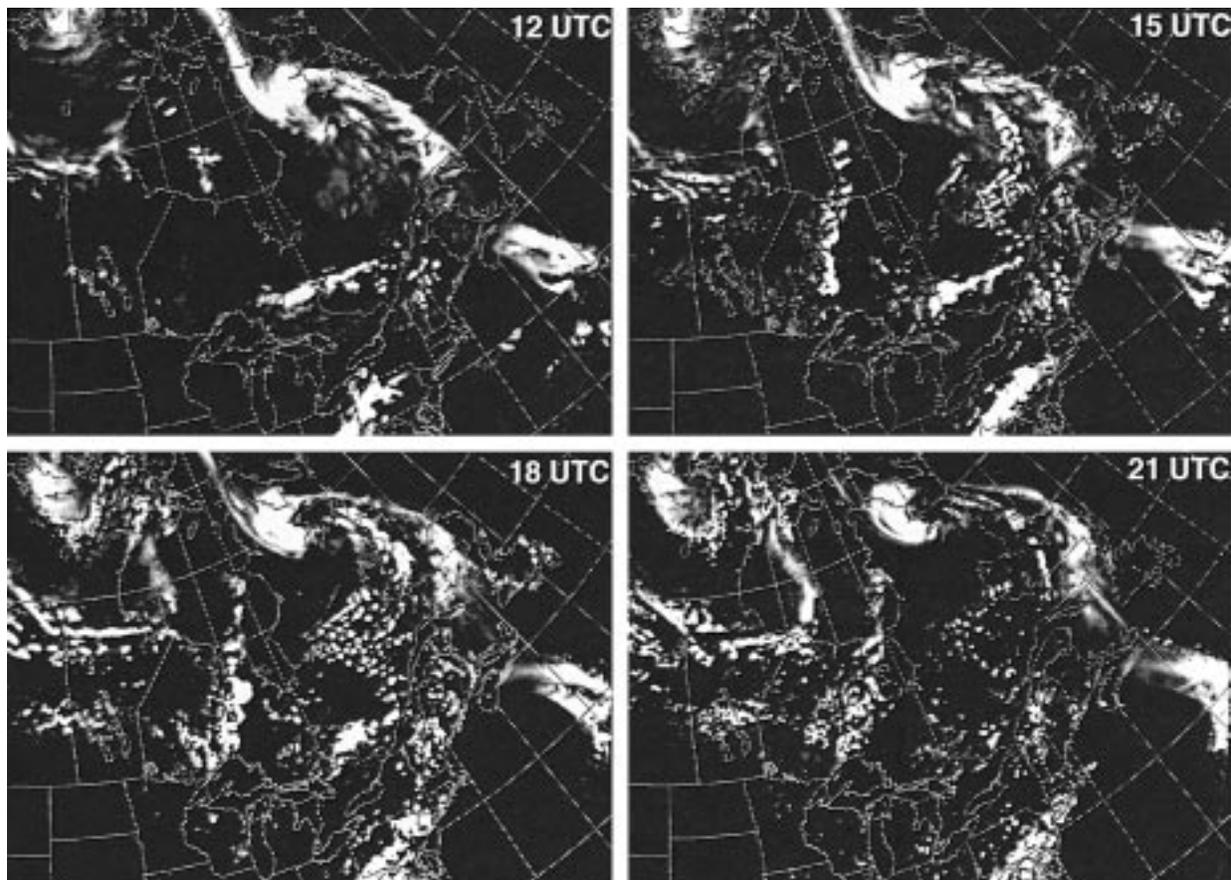


FIG. 4. MC2 North American 10-km forecast: 12–21-h forecasts of total precipitation rate.

bit-wide data paths to memory. The vector unit of each processor consists of eight parallel vector sets of four pipes, one add–shift, one multiply, one divide, and one logical. For each vector unit there are 8 64-bit vector arithmetic registers that are used in arithmetic and logical operations, and there are 64 vector data registers, also 64 bit, used for temporary storage. For a more detailed overview of the SX-4 architecture, see Hammond et al. (1996). The peak performance of a concurrent vector add and vector multiply is 2 Gflop  $s^{-1}$ , and typical scientific codes written in Fortran can sustain 1 Gflop  $s^{-1}$  with sufficiently long vector lengths.

Main memory unit (MMU) configurations for an SX-4 SMP node range from 512 Mbytes to 8 Gbytes of synchronous static random access memory (SSRAM). This memory technology has a 15-ns cycle time coupled with the ability to produce a data item while accepting a new address. In the maximum 8-Gbyte configuration there are 32 main memory cards, consisting of 32 banks of 256 Mbytes, for a total of 1024 banks. Memory bandwidths of 16 Gbytes  $s^{-1}$  per processor are supported, resulting in a total bandwidth of 256 Gbytes  $s^{-1}$ . In the 4-Gbyte MMU configuration, the bandwidth drops to 128 Gbytes  $s^{-1}$ . Supplementing main memory is 16 or 32 Gbytes of extended memory

unit (XMU). This secondary memory [or very fast random access memory (RAM) disk] is made up of dynamic random access memory, having a 60-ns access time. This high-speed device offers peak transfer rates on the order of 4 Gbytes  $s^{-1}$  and is used primarily as a system cache and as high-performance file systems, in which maximum I/O performance is required. To launch a multinode SC-MC2 on an SX-4/32M configuration consisting of two SX-4/16 nodes (ixs0 and ixs1) connected via the IXS crossbar, we use 16 processors on node ixs0 and 16 processors on node ixs1. The command “mpisx -f mpi.hosts” and associated “mpi.hosts” are given below:

```
# mpi.hosts contains SMP cluster configuration
-p 16 -h ixs0 -e ./mc2.Abs
-p 16 -h ixs1 -e ./mc2.Abs.
```

The Mflop  $s^{-1}$  per processor execution rates based on a  $255 \times 40 \times 31$  subdomain are reported in Fig. 1. Each SX-4/16 node in the multinode configuration contained an 8-Gbytes MMU. The per processor performance drop-off observed in Fig. 1 follows closely the bandwidth degradation predicted by Patel (1981) for a processor–memory interconnect such as the SX-4/32 in-

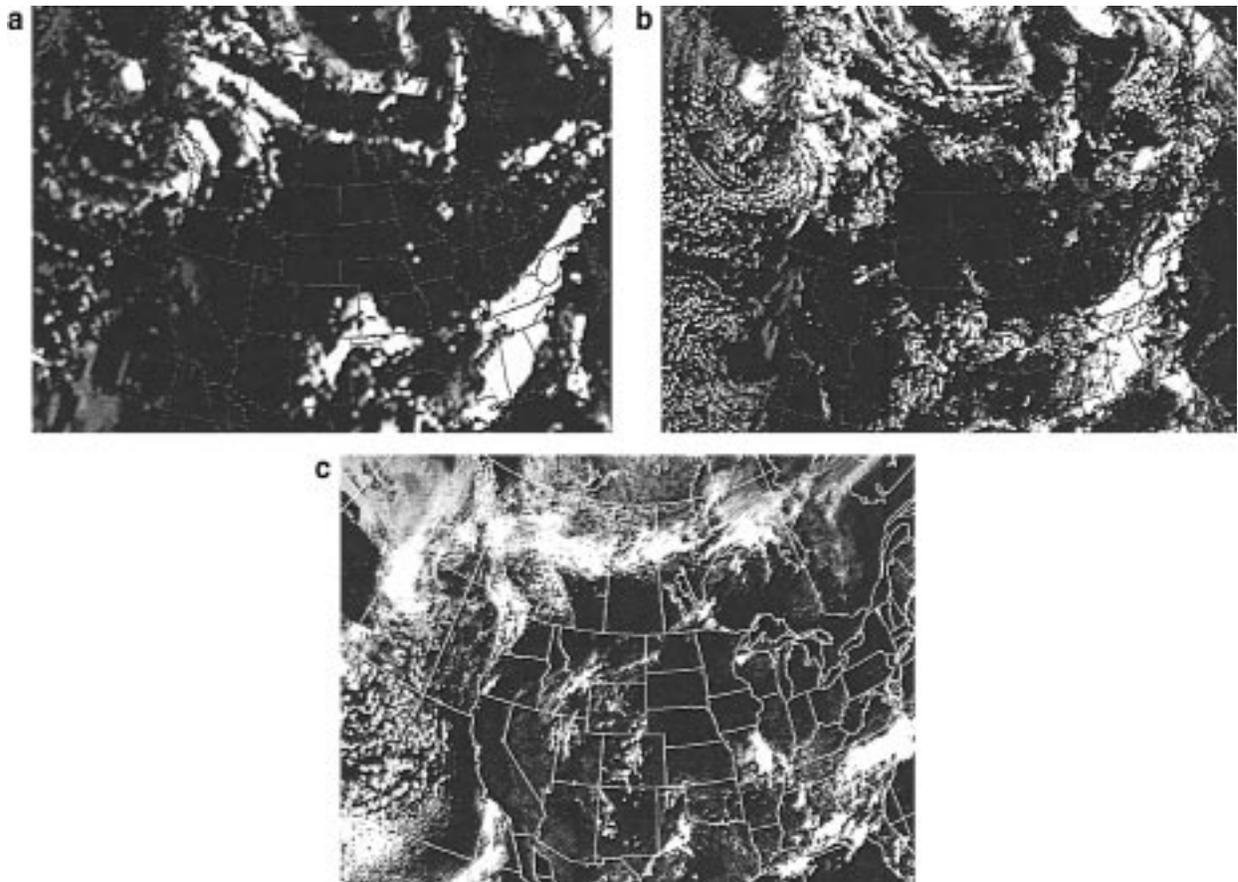


FIG. 5. North American forecast: 18-h forecast of total precipitation rate, valid at 1800 UTC on 26 June 1997, for (a) GEM 35 km and (b) MC2 10 km, compared against (c) cloud albedo from *GOES-8* (Geostationary Operational Environment Satellite) satellite picture, valid at the same time.

ternal crossbar switch. Internode message-passing latency and communication overheads are minimal compared with the degradation in performance due to the internal crossbar network.

### 5. Real-time North American 10-km forecast

In order to build a complete numerical weather prediction model, the adiabatic kernel of the nonhydrostatic MC2 model described above was coupled with version 3.5 of the RPN physical processes parameterization package [detailed documentation for this package is available in Maillhot (1994)]. This package includes a planetary boundary layer scheme based on turbulent kinetic energy, a surface-layer scheme based on similarity theory, solar and infrared radiation, large-scale condensation, convective precipitation, and gravity wave drag schemes. The physics interface was modified also to support input–output to a fast file system (XMU) for memory requirements above the current 8-Gbyte MMU limit on the NEC SX-4/32.

To demonstrate that a real-time weather forecast at 10-km resolution over North America is feasible, the

SMP cluster version of MC2 was run on a single SX-4/32 node at the Canadian Meteorological Centre (CMC). Such a run would, in principle, permit scale interactions to occur across a very large spectral bandwidth (10–7000 km). However, evidence of these interactions (e.g., spectra, PV) were not produced explicitly as part of our simulation. In the initial trial run, no special care was taken to ensure high-quality initial and lateral boundary conditions or to provide the model with high-resolution surface parameters. In fact, initial and lateral boundary conditions were obtained by simple interpolation of output data from a previous run with the operational Canadian Global Spectral Finite Element model initialized at 0000 UTC on 26 June 1997. Although this particular case is very active for the summertime, it was chosen because a dedicated SX-4/32 node was available on this particular day, and the analyses were easily available as part of the operational runs at CMC. The MC2 model was therefore nested within the operational SEF T219L28 to produce a 24-h forecast at 10-km resolution. Boundaries were updated every 6 h, and linear time interpolation was employed

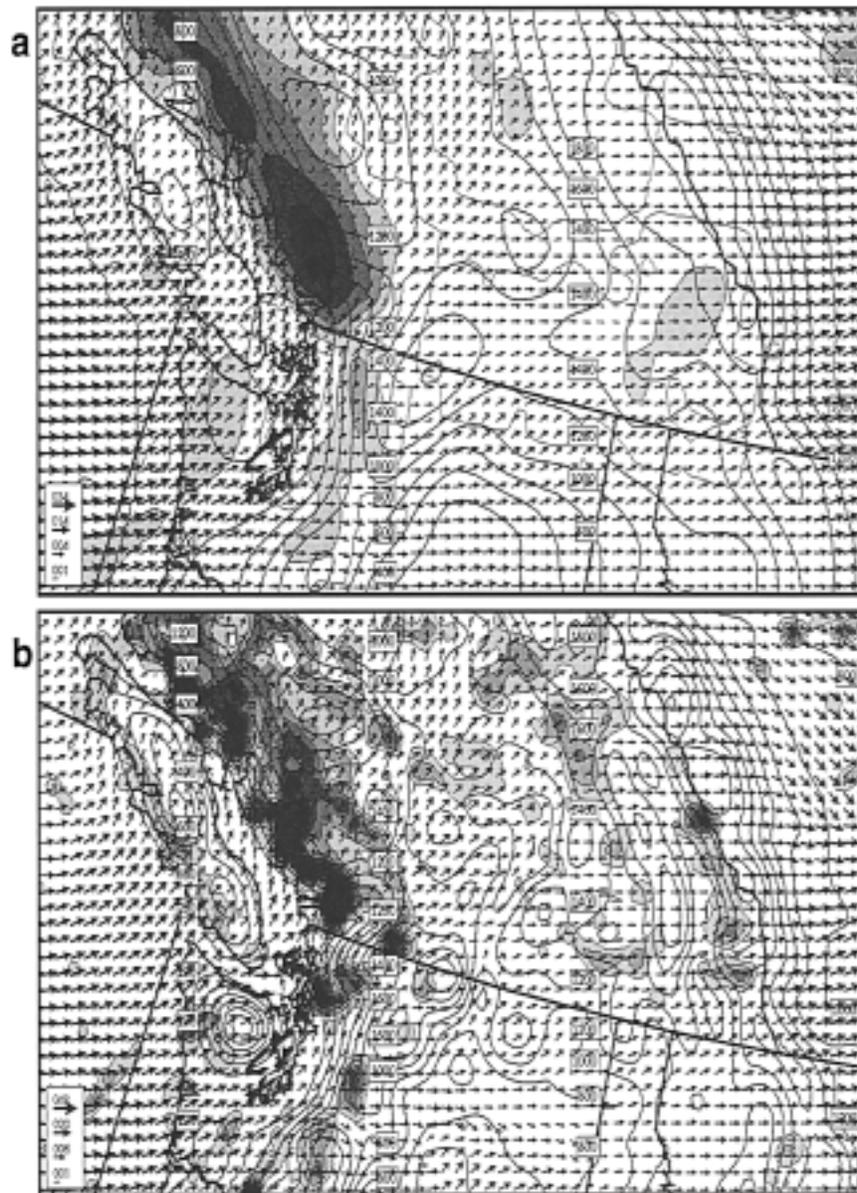


FIG. 6. North American forecast: 24-h forecast, valid at 0000 UTC on 27 June 1997. Total precipitation rate (gray shades) and near-surface flow (arrows with scale in knots in lower left corner) are superimposed over topography (solid contours every 200 m) in the Pacific Northwest for (a) GEM 35 km and (b) MC2 10 km. One arrow every other grid point for each direction is displayed in (b), while interpolation within two consecutive grid points is performed in (a) to provide the equivalent of 1.75 arrows per grid length.

between updates to produce time-dependent boundary conditions at every time step.

The model lid was set at 30 km. Again, here, for simplicity, an explicit Laplacian-type horizontal diffusion was chosen with a diffusion coefficient of  $2000 \text{ m}^2 \text{ s}^{-1}$ . The lateral boundary absorbing layer is described in Benoit et al. (1997) and was set to a width of 10 grid points. Topography was generated using a 500-m digital elevation model on which a Cressman-type filter was applied as an averaging operator. A matching land-sea

mask is then constructed using a 1-km resolution vegetation-type database. Finally, the physics configuration was set to match exactly the current configuration of the operational Canadian Global Environmental Multiscale (GEM) Model, which includes a Kuo-type convection scheme along with a Sundqvist-type scheme for large-scale condensation (Mailhot et al. 1997). The authors acknowledge that a more sophisticated convection scheme such as Fritsch-Chappell would have been preferable, but although available at the time, these schemes

were not performing sufficiently fast enough to be considered for operational purposes. A gravity wave drag parameterization was used for this run but likely had very little effect, since most of this feature is entirely resolved at 10-km resolution.

A  $753 \times 510 \times 31$  grid was required for the North American forecast (the continental United States and Canada) at 10-km resolution. For such a grid, the breakdown of CPU time is 65% dynamics (15% advection and 50% solver), roughly 30% physics, and 5% other tasks, such as I/O, grid nesting, and the Davies scheme. The forecast was completed in 40 min of wall-clock time on an NEC SX-4/32 node. To maximize single cpu performance, a domain decomposition consisting of  $30 = 3 \times 10$  subdomains ( $n_{pes} = n_{px} \times n_{py}$ ) was chosen, resulting in 800 Mflop  $s^{-1}$  per cpu or 24 Gflop  $s^{-1}$  sustained. At 10-km resolution, a 24-h time integration requires 480 time steps of length 180 s. The real-time performance of the SMP cluster MC2 Meso-LAM at high resolution is best characterized by the time to complete a time step: 4 s per time step plus 1.5 s parallel I/O in physics equals 5.5 s. With additional SSRAM memory, the physics will execute in-core, and in this case, the model will run at 4 s per time step, requiring 1920 s or 32 min of wall-clock time, including I/O. Memory requirements for a real-time forecast over North America are quite large: 8 Gbytes total MMU available, consisting of high-density SSRAM chips, each executing process assigned a subdomain requires 240 Mbytes of memory for a total of 7.2 Gbytes. Physics spillover onto 16 Gbytes XMU: 40 Mbytes per process or 1.2 Gbytes total read/write to solid-state RAM disk every time step.

The objective of our simulation was to demonstrate that a real-time 10-km forecast is now possible. Given that normal operational procedures were not followed strictly to initialize the model, it was decided not to perform any objective validation for this particular run. Some qualitative evaluations, however, were performed to assess the usefulness of such a high-resolution run for daily forecasting. For a first validation exercise, the MC2 forecast at 10 km was compared against the regional 35-km forecast produced on the same day by the GEM model. It is clear from Fig. 2 that the large-scale signal from the MC2 is in excellent agreement with the one produced by GEM. This result was expected since the two models employed exactly the same physical parameterization package and analyses. Notice the trough on the Canadian west coast just to the north of Vancouver Island. This trough is associated with a mature low pressure system at lower levels that remained quasi-stationary throughout the entire forecast. A cold front extends westward from this low into the Pacific Ocean, more or less delimiting the northern edge of a large area in which a fingerprint of cloud streets and prefrontal convective rain developed (see Fig. 3). The shadowing effect of the Rockies in this westerly flow can be observed also at the boundary between British

Columbia and Alberta, as well as in Oregon and Washington. Also of some interest is the precipitation tongue moving out to sea in the return flow north of the Queen Charlotte Islands. All of these mesoscale features can be seen clearly in the satellite image of Fig. 5c.

On the eastern side of the continent, a low pressure system over northern Quebec moved northward (Fig. 4). Convective activity can be seen to develop in the afternoon ahead of a cold front extending from this low into central Quebec. Notice, too, the fast-moving convective system moving eastward into southern Quebec and the very well organized frontal systems developing and moving into Ontario from the northwest. Because of their small size, convective cells such as these may sometimes appear as noise (or gridpoint storms) associated with some form of instability in numerical models. This is not the case, since the smallest individual precipitation features depicted in Fig. 4 are resolved by at least 10–15 grid points. However, it must be noted that the afternoon convection in the eastern part of the forecast is more difficult to validate against the satellite picture of Fig. 5c.

Finally, in this attempt to validate qualitatively the mesoscale features of this run, one would expect that a 10-km resolution forecast would more or less add mesoscale details over larger-scale signals obtained from a run at coarser resolution while retaining the integrity of the large-scale features. Referring to Fig. 5, it is our opinion that the MC2 forecast did just that, and in fact, it added details that can be seen clearly in the satellite picture and that could not have been forecast at coarser resolution. However, it is equally important to note that when the large-scale forecast is inaccurate, the added mesoscale features will probably not improve the forecast much. Added mesoscale features can be observed with greater detail in the blowup of the Pacific Northwest, shown in Fig. 6, in which the precipitation rate and near-surface winds from the operational GEM and MC2 are compared. Precipitation responds very sharply to a more detailed topography, and it is seen to fall on the upward slope of mountains, while creating the so-called shadow effect on the lee side. Notice, too, the stronger flow channeling directly south of Vancouver Island in the Fraser River Valley and the deviation of the horizontal flow around sharper topography features such as Mount Adams in the northwest corner of Oregon.

## 6. Conclusions

With SMP cluster architectures now capable of performance in the 25–30-Gflop  $s^{-1}$  range, it is now possible for the first time ever to produce a real-time weather forecast over the entire North American continent at 10-km resolution. We fully expect in the next two years to sustain more than 100 Gflop  $s^{-1}$  at even higher resolutions on up to four SX-4/32 SMP nodes. At these performance levels, it is possible to contemplate an ex-

ploration of the way condensation processes are parameterized (convective and resolved), as we will be in a position to compare against fully explicit simulations with sophisticated microphysics. Also, we hope to investigate the possible limitations of surface processes, since vapor fluxes might not respond to the actual simple formulation in a hurricane.

*Acknowledgments.* The authors would like to thank their RPN colleagues Claude Girard and Peter Bartello for reviewing the manuscript and for their participation in ongoing efforts to improve the MC2 atmospheric model. We also wish to thank Bertrand Denis of the CCCma for reviewing the 10-km forecast products.

#### REFERENCES

- Baillie, C., J. Michalakes, and R. Skälin, 1997: Regional weather modeling on parallel computers. *Parallel Comput.*, **23**, 2135–2142.
- Benoit, R., M. Desgagné, P. Pellerin, S. Pellerin, Y. Chartier, and S. Desjardins, 1997: The Canadian MC2: A semi-Lagrangian, semi-implicit wide-band atmospheric model suited for fine-scale process studies and simulation. *Mon. Wea. Rev.*, **125**, 2382–2415.
- Clark, T. L., 1977: A small-scale dynamic model using a terrain-following coordinate transformation. *J. Comput. Phys.*, **24**, 186–214.
- Davies, H. C., 1976: A lateral boundary formulation for multi-level prediction models. *Quart. J. Roy. Meteor. Soc.*, **102**, 405–418.
- Dent, D., and G. Mozdzyński, 1997: ECMWF operational forecasting on a distributed memory platform: Forecast model. *Proc. Seventh ECMWF Workshop on the Use of Parallel Processors in Meteorology*, Singapore, World Scientific, 36–51.
- Eisenstat, S. C., H. C. Elman, and M. H. Schultz, 1983: Variational iterative methods for nonsymmetric systems of linear equations. *SIAM J. Numer. Anal.*, **2**, 345–357.
- Gal-Chen, T., and R. C. Somerville, 1975: On the use of a coordinate transformation for the solution of the Navier-Stokes equations. *J. Comput. Phys.*, **17**, 209–228.
- Gropp, W., E. Lusk, and A. Skjellum, 1994: *Using MPI: Portable Parallel Programming with the Message-Passing Interface*. The MIT Press.
- Hammond, S., R. Loft, and P. Tannenbaum, 1996: Architecture and application: The performance of the NEC SX-4 on the NCAR benchmark suite. *Supercomputing 96 Proc.*, San Jose, CA, NCAR.
- Hennessy, J. L., and D. A. Patterson, 1990: *Computer Architecture: A Quantitative Approach*. Morgan-Kaufmann Publishers.
- Kong, F., and M. K. Yau, 1997: An explicit approach to microphysics in MC2. *Atmos.–Ocean*, **35**, 257–291.
- Kumar, V., A. Grama, A. Gupta, and G. Karypis, 1994: *Introduction to Parallel Computing: Design and Analysis of Algorithms*. Benjamin/Cummings, 597 pp.
- Mailhot, J., 1994: The Regional Finite Element (RFE) Model Scientific Description. Part 2: Physics. RPN, 307 pp. [Available from RPN, 2121 Trans-Canada, Dorval, QC H9P 1J3, Canada.]
- , R. Sarrazin, B. Bilodeau, N. Brunet, and G. Pellerin, 1997: Development of the 35 km version of the operational regional forecast system. *Atmos.–Ocean*, **35**, 1–28.
- Patel, J. H., 1981: Performance of processor-memory interconnections for multiprocessors. *IEEE Trans. Comput.*, **C-30**, 771–780.
- Saad, Y., 1996: *Iterative Methods for Sparse Linear Systems*. PWS Publishing, 447 pp.
- , and M. Schultz, 1986: GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, **7**, 856–869.
- Skälin, R., 1997: Scalability of parallel gridpoint limited-area atmospheric models. Part II: Semi-implicit time-integration schemes. *J. Atmos. Oceanic Technol.*, **14**, 442–455.
- Skamarock, W. C., P. K. Smolarkiewicz, and J. B. Klemp, 1997: Preconditioned conjugate-residual solvers for Helmholtz equations in nonhydrostatic models. *Mon. Wea. Rev.*, **125**, 587–599.
- Smith, B., P. Bjorstad, W. Gropp, 1996: *Domain Decomposition: Parallel Multilevel Methods for Elliptic PDEs*. Cambridge University Press, 224 pp.
- Tanguay, M., A. Robert, and R. Laprise, 1990: A semi-implicit semi-Lagrangian fully compressible regional forecast model. *Mon. Wea. Rev.*, **118**, 1970–1980.
- Thomas, S. J., A. V. Malevsky, M. Desgagné, R. Benoit, P. Pellerin, and M. Valin, 1997: Massively parallel implementation of the Mesoscale Compressible Community Model. *Parallel Comput.*, **23**, 2143–2160.
- , C. Girard, R. Benoit, M. Desgagné, and P. Pellerin, 1998: A new adiabatic kernel for the MC2 model. *Atmos.–Ocean*, **36**, 241–270.
- Woodward, P. R., 1996: Perspectives on supercomputing: Three decades of change. *IEEE Comput.*, **29**, 99–111.