JOURNAL OF COMPUTATIONAL PHYSICS 89, 95–124 (1990)

Finite-Element Schemes for Extended Integrations of Atmospheric Models

J. STEPPELER

*European Centre for Medium-Range Weather Forecasts, Reading, Berkshire RG2 9AX, England**

I. M. NAVON

Department of Mathematics, Florida State University, and Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida 32306-4052

AND

H.-I. LU

*NASA Marshall Space Flight Center, ES-42, Huntsville, Alabama 35812**

Received June 1, 1988; revised August 10, 1989

The effect of conservation of integral invariants by finite-element discretization schemes of the shallow-water equations as a model for long-term integrations of atmospheric models is investigated. Two finite-element models are used. The first uses rectangular elements and conserves total energy using an intrinsic method. The second model uses triangular elements and a high-accuracy two-stage Numerov–Galerkin method. It well conserves total energy and potential enstrophy by applying a periodical Shuman filter every 48 or 96 time steps. Different conserving and non-conserving versions of these finite-element schemes are compared in terms of their conservation of the integral invariants of the shallow-water equations for long-term integrations (20–100 model days). Critical times for numerical non-linear instability are investigated along with the determination of the critical amount of dissipation required to achieve stable long-term integrations. Comparisons of the two finite-element schemes, namely, the rectangular and the triangular in terms of their relative computational efficiency and accuracy, are also provided. Similarities and differences in the behavior of finite-element schemes and finite-difference schemes for long-term stable integrations of the shallow-water equations are finally discussed.

* This work was done while visiting the Supercomputer Computations Research Institute, Florida State University.

0021-9991/90 $3.00

581/89/1-7

1. INTRODUCTION

Long integrations of the atmospheric equations for climate simulation or medium range forecasts normally use finite-difference schemes which conserve quantities like energy or potential enstrophy. Conservative finite-difference schemes were proposed by Arakawa [1], Janjic [2], Sadourny [3], Arakawa [4], as well as in Arakawa and Suarez [5].

Apart from "a priori" methods to conserve integral invariants, "a posteriori" methods have been used to make schemes conserving. Such methods use non-conserving finite-difference or finite-element schemes and restore the integral invariants by a variational approach. Such schemes were proposed by Sasaki [6], Sasaki *et al.* [7]. Navon [8, 9] investigated a number of "a posteriori" schemes and Kalnay-Rivas *et al.* [10] obtained results on the effect of enstrophy conservation by an "a posteriori" method in a general circulation model. Navon [11] enforced conservation of integral invariants with a finite-element model of the shallow-water equations.

Conservative schemes will help to prevent the non-linear instability first observed by Phillips [12]. This instability can be avoided with most schemes by introducing a spatial smoothing operator according to Shapiro [13]. Conservative schemes will have less artificial energy loss due to such filtering.

In recent years, an increasing number of integrations of atmospheric models have been carried out using Galerkin schemes. The Galerkin method most widely used is the spectral method of Machenhauer [14] and Orszag [15]. Bourke [16, 17] gives an example of a three-dimensional atmospheric model using the spectral method. The spectral method does not formally conserve energy for all dynamic equations where energy is a cubic quantity in the basic variables—like the divergent shallow-water equations or three-dimensional models in σ coordinates. Its suitability for climate integrations, which is generally accepted, is based on two arguments:

(1) The spectral method conserves energy for non-divergent flow fields.

(2) Because of its high accuracy, the spectral method will also approximate the conservation laws with good accuracy.

The relevance of energy conservation for spectral discretization has been discussed in Weigle [51]. The vertical discretizations used with spectral methods for climate or medium range forecast models are normally energy conserving.

Simmons *et al.* [20] gives an example of an energy conserving finite-difference method. The finite-element vertical discretization proposed by Steppeler [21, 22] is energy conserving. Non-conserving finite-element schemes for short range forecasting were developed by Staniforth [23] and Beland *et al.* [24]. In a comparison of an energy conserving finite-element scheme by Steppeler [25] for the vertical discretization with a non-conserving scheme [25], substantial deviations from conservation for the latter scheme were found. These models used finite-element

versions of the model described in [20] using semi-implicit time integration with a time step $\Delta t = 22.5$ min.

For horizontal discretization, finite-element schemes are a possible alternative to the spectral method. Meteorological applications of this method have been reviewed by Navon [26] and Staniforth [27]. According to Cullen [28] and Gresho *et al.* [29], they are much more accurate than second order finite-difference schemes, and are simpler and computationally cheaper than the spectral method, particularly for high resolution models.

This method has been used successfully for short-range forecast models by Cullen [30], Navon [31], and Staniforth *et al.* [32, 33]. Its usefulness for general circulation and medium-range forecast models remains to be seen. Application of finite-elements to mesoscale models have been given by Manton [34], McPherson *et al.* [35], and Mailhot *et al.* [36]. Carson *et al.* [37] evaluated the performance of a finite-element model for short-range forecasting.

For long-term integrations, the conservation of the integral constraints when using discretization schemes has a pivotal importance. In fact, it was shown by Jespersen [39] that the Arakawa schemes [1] can be closely related to a particular form of finite-element schemes. For the finite-element method, Fix [38] showed that energy is conserved for the non-divergent shallow-water equations.

Conserving finite-element schemes for the two-dimensional simulation of inviscid flow Boussinesq equations were discussed by Cliffe [40] and Lee *et al.* [41, 42]. It became clear that standard finite-element schemes will conserve energy only for the cases when energy is a second-order moment in the fields.

In cases when energy is a third order moment, i.e., for variable density, special precautions have to be taken in order to achieve energy conservation. An energy-conserving finite-element scheme for the divergent shallow-water equations was defined in Steppeler [18]. Energy conserving finite-element schemes have not yet been tested with two-dimensional flow models. The schemes of [42, 40, and 18] were tested only by one-dimensional computational examples.

The present paper investigates the application of the energy conserving scheme [18] with linear elements to the two-dimensional divergent shallow-water primitive equations model.

This scheme can be inplemented with both triangular (Navon [31]) and rectangular elements. It was argued by Staniforth [33, 45] that rectangular elements can be implemented much more economically than triangular ones.

The present paper uses both rectangular elements on a regular grid (for which a quasi-regular grid as used in [33] could easily be included), as well as a Numerov–Galerkin high-accuracy scheme with regular triangular elements [11].

The effect of using a minimal amount of dissipation both with and without a constraint restoration by "a posteriori" augmented-Lagrangian methods [9] for energy and potential enstrophy are also investigated using the triangular element model defined in [31, 11]. The effects of critical dissipativity and finite-time "blow-up" in finite-element models is amply discussed.

While such behavior was until now discussed and studied in finite-difference

models of the shallow-water equations (see, for instance, [3] and Fairweather and Navon [48]), there was to our knowledge no similar effort concerning finite-element schemes.

The basic finite-element schemes will be described in Section 2 while the implementation of the different schemes will be described in Section 3. Section 4 describes briefly the Numerov–Galerkin finite-element scheme with linear triangular elements. Section 5 will give the results of test integrations as well as a discussion of the results of the various experiments for long-term integrations using conserving and non-conserving finite-element schemes. Finally, conclusions and directions for further research will be provided in Section 6.

2. The Finite-Element Scheme on Rectangular Elements

For the tests with triangular elements we shall use the Numerov–Galerkin technique described in [11] and documented in [31]. Therefore in the following we describe only the model using rectangular elements.

For a one-dimensional grid X_ν let $e_\nu(X)$ be the piecewise linear function defined by

$$\begin{aligned} e_\nu(X_\nu) &= 1 \\ e_\nu(X_\mu) &= 0 \qquad \text{for} \quad \mu \neq \nu \end{aligned} \tag{1}$$

The representation of a field $\Phi(X, Y)$ in a two-dimensional grid $X_\nu Y_\mu$ is then given by

$$\tilde{\Phi}(X, Y) = \sum_{\nu\mu} \tilde{\Phi}_{\nu\mu}\, \tilde{\Phi}_{\nu\mu} e_\nu(X)\, e_\mu(Y). \tag{2}$$

We define the two-dimensional basis function $b_{\mu\nu}$ by

$$b_{\nu\mu}(X, Y) = e_\nu(X) \cdot e_\mu(Y). \tag{3}$$

The $\Phi_{\nu\mu}$ in (2) are the amplitudes of the field Φ, which are also gridpoint values.

An energy conserving finite-element scheme is defined in terms of Galerkin operators G, which approximate a general field Ψ, represented in (2) by a function of the class defined in (2),

$$\tilde{\Psi} = G\Psi,\ \tilde{\Psi}(X, Y) = \sum_{\nu\mu} \tilde{\Psi}_{\nu\mu} b_{\nu\mu}(X, Y). \tag{4}$$

The amplitudes $\tilde{\Psi}_{\nu\mu}$, which determine $\tilde{\Psi}$ in (4), are defined by

$$(G\Psi, b_{\nu\mu}) = (\tilde{\Psi}, b_{\nu\mu}) = (\Psi, b_{\nu\mu}). \tag{5}$$

Some mathematical properties of the Galerkin operators G are given in the

Appendix. In (5), the scalar product (a, b) of two functions $a(X, Y)$ and $b(X, Y)$ is defined as

$$(a, b) = \int dX\, dY\, a(X, Y)\, b(X, Y)\, w(X, Y). \tag{6}$$

In (6), $w(X, Y)$ is a positive weight function. Different choices of w in (6) will define different Galerkin approximation operators G_1, G_2, ... in (4) and (5). For a more detailed definition of G see [18].

The primitive shallow-water equations to be approximated are

$$\begin{aligned} \dot{U} &= fV - UU_X - VU_Y - H_X \\ \dot{V} &= -fU - UV_X - VV_Y - H_Y \\ \dot{H} &= -(UH)_X - (VH)_Y. \end{aligned} \tag{7}$$

In order to define an energy conserving Galerkin scheme for (7), we use the Galerkin operator G_1 corresponding to $w = H(X, Y)$ and the operator G_2, defined by $w = 1$ in (4), (5), and (6).

An energy conserving finite-element scheme for (7) is defined by

$$\begin{aligned} \dot{U} &= G_1(\eta V - (G_2(\tfrac{1}{2}(U^2 + V^2) + H))_X) \\ \dot{V} &= G_1(-\eta U - (G_2(\tfrac{1}{2}(U^2 + V^2) + H))_Y) \\ \dot{H} &= -G_2((UH)_X + (VH)_Y) \\ \eta &= V_X - U_Y + f. \end{aligned} \tag{8}$$

In (8), G_1 is the Galerkin operation with weight H, and G_2 is the Galerkin operation with weight 1. Here η represents the absolute vorticity.

The proof of energy conservation of scheme (8) is given in [18]. For comparison, we will also use the standard Galerkin finite-element scheme, which is not energy conserving:

$$\begin{aligned} \dot{U} &= G_2(fV - UU_X - VU_Y - H_X) \\ \dot{V} &= G_2(-fV - UV_X - VV_Y - H_Y) \\ \dot{H} &= G_2(-(HU)_X - (HV)_Y). \end{aligned} \tag{9}$$

In (9), G_2 is the Galerkin operator with weight 1. Scheme (9) will be referred to as the non-conserving finite-element scheme.

The computational technique for the computation of G will be given in Section 3. Here we describe the decomposition of G into one-dimensional operators. In analogy to (6), we define

$$(a, b)_w^X = \int dX\, a(X, Y)\, b(X, Y)\, w(X, Y)$$
$$(a, b)_w^Y = \int dY\, a(X, Y)\, b(X, Y)\, w(X, Y), \tag{10}$$

where $(a, b)_w^X$ is a function of Y, and $(a, b)_w^Y$ is a function of X. From (6) we conclude

$$\begin{aligned}(a, b) &= ((a, b)_w^X, 1)_1^Y \\ &= ((a, b)_w^Y, 1)_1^X.\end{aligned} \tag{11}$$

The one-dimensional Galerkin operators G_w^X and G_w^Y are defined in analogy to (5) by

$$\begin{aligned}(G_w^X \Psi(X, Y), e_\nu(X))_w^X &= (\Psi(X, Y), e_\nu(X))_w^X \\ (G_w^Y \Psi(X, Y), e_\mu(Y))_w^Y &= (\Psi(X, Y), e_\mu(Y))_w^Y.\end{aligned} \tag{12}$$

The two-dimensional Galerkin operator G defined in (5) is decomposed into one-dimensional operators by the product

$$G = G_w^X G_w^Y \tag{13}$$

or

$$G = G_w^Y G_w^X. \tag{14}$$

To prove (13), relation (5) has to be demonstrated for $G_w^X G_w^Y$. Using (3),

$$\begin{aligned}(G_w^X G_w^Y \Psi, b_{\nu\mu}) - (\Psi, b_{\nu\mu}) &= (G_w^X G_w^Y \Psi, e_\nu(X)\, e_\mu(Y)) - (G_w^Y \Psi, e_\nu(X)\, e_\mu(Y)) \\ &\quad + (G_w^Y \Psi, e_\nu(X)\, e_\mu(Y)) - (\Psi, e_\nu(X)\, e_\mu(Y)).\end{aligned} \tag{15}$$

Using Eq. (1), we obtain

$$\begin{aligned}(G_w^X G_w^Y \Psi, b_{\nu\mu}) - (\Psi, b_{\nu\mu}) &= ((G_w^X G_w^Y \Psi, e_\nu(X))_w^X, e_\mu(Y))_1^Y \\ &\quad - ((G_w^Y \Psi, e_\nu(X))_w^X, e_\mu(Y))_1^Y \\ &\quad + ((G_w^Y \Psi, e_\mu(Y))_w^Y, e_\nu(X))_1^X \\ &\quad - ((\Psi, e_\mu(Y))_w^Y, e_\nu(X))_1^X.\end{aligned} \tag{16}$$

Using in Eq. (16), the relations

$$(a_1 + a_2, b)_1^X = (a_1, b)_1^X + (a_2, b)_1^X \tag{17}$$
$$(a_1 + a_2, b)_1^Y = (a_1, b)_1^Y + (a_2, b)_1^Y, \tag{18}$$

and applying the definitions in Eq. (12) of G_w^X and G_w^Y, we conclude

$$(G_w^X G_w^Y \Psi, b_{\nu\mu}) - (\Psi, b_{\nu\mu}) = 0. \tag{19}$$

Together with (5) this proves (13).

The decomposition in Eqs. (13) and (14) of the Galerkin operation into one-dimensional ones renders the scheme rather economical and allows the application of direct linear equation solvers. This is the main advantage of rectangular versus triangular elements (See also Staniforth [45]).

3. The Implementation of the Rectangular Finite-Element Shallow-Water Equation Model

The implementation of the triangular scheme **FEUDX** is described in [31]. In the following we will describe the implementation of the rectangular finite-element scheme introduced in Section 2.

Like the spectral method, the finite-element scheme uses a basis function representation for the fields. For the finite-element scheme, the amplitudes $\Phi_{\nu\mu}$ are also gridpoint values. They are represented in the main nodepoint grid, shown in Fig. 1A.

3.1. *Grids and Integration Formulas*

For the spectral method, an efficient computation of the Galerkin operators is possible by using a collocation grid Orszag [44]. The use of a collocation grid with finite-elements has been proposed in [45] as an efficient means to implement a finite-element scheme, and it was used with a vertical finite-element discretization in [22].

According to (4), (5), and (6) the implementation of a Galerkin operator requires the computation of the Galerkin integrals $(\Psi, b_{\nu\mu})$ in (5). These two-dimensional integrals can be reduced to a sequence of one-dimensional integrals (See Eq. (11)).

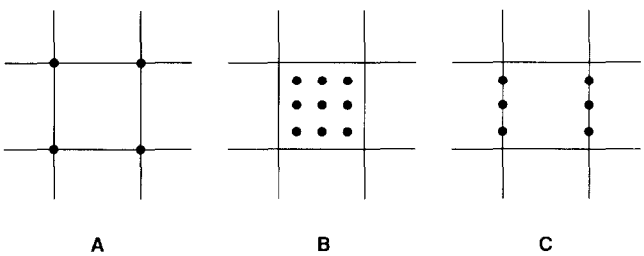

FIG. 1. The grids used for the implementation of the rectangular finite-element scheme, A: main nodepoints, B: collocation grid, C: Y-collocation grid.

According to Eq. (10), the one-dimensional Galerkin integral $(\Psi(X, Y), e_\mu(Y))$ in the Y-direction can be written as

$$(\Psi(X, Y), e_\mu(Y)) = \int_{Y_{\mu-1}}^{Y_{\mu+1}} dY\, \Psi(X, Y)\, e_\mu(Y)\, w(X, Y)$$

$$= \int_{Y_{\mu-1}}^{Y_\mu} dY\, \Psi(X, Y)\, e_\mu(Y)\, w(X, Y)$$

$$+ \int_{Y_\mu}^{Y_{\mu+1}} dY\, \Psi(X, Y)\, e_\mu(Y)\, w(X, Y). \tag{20}$$

For each of the two integrals in (20) the integrand is a polynomial. Therefore, the integral can be evaluated exactly by Gaussian quadrature (see Davis and Rabinowitz [46]). The functions Ψ, as given in (8), are quadratic functions in Y for fixed X, and e_μ and w have a degree not higher than 1. Therefore the integrand can consist of polynomials of degree up to 4. An exact integration in (20) is achieved with three Gaussian points:

$$\int_{Y_\mu}^{Y_{\mu+1}} dY\, \Psi(X, Y)\, e_\mu(Y)\, w(X, Y) = \sum_{\xi=1}^{3} \alpha_\xi \Psi(X, Y_{\mu,\xi})\, e_\mu(Y_{\mu,\xi})\, w(X, Y_{\mu,\xi}). \tag{21}$$

The Gaussian points $Y_{\mu,\xi}$ and weights α_ξ belonging to the interval $(Y_\mu, Y_{\mu+1})$ are given in Abramowitz and Stegun [47].

The integrals on the right-hand side of (20) are evaluated for $X = X_\nu$. The corresponding grid is called the Y-collocation grid. It is shown in Fig. 1C. Gaussian collocation points are also introduced for the X-variable. For each interval X_ν, $X_{\nu+1}$, we obtain three collocation points $X_{\nu,\rho}$, $\rho = 1, 2, 3$. The Y-values belonging to these points are $Y_{\mu,\xi}$. The grid $X_{\nu,\rho}$, $Y_{\mu,\xi}$ is called the collocation grid. It is shown in Fig. 1B.

According to Fig. 1 there are three representations for a function $\Psi(X, Y)$. According to (2), the basic fields U, V, H in (7) are represented in the main nodepoint grid, Fig. 1A. The Galerkin projection $G\Psi$ of any function $\Psi(X, Y)$ is also represented in the main nodepoint grid. The collocation grid, Fig. 1C, will be used to represent functions $\Psi(X, Y)$ occurring in the right-hand side of (8). The Y-collocation grid, Fig. 1B, will be used to represent $G_w^X \Psi$, with G_w^X being defined in (12).

The computational technique is quite similar to that used with the spectral method [44]. All fields are interpolated to the collocation grid, the equations of motion (7) are solved on the collocation grid, and then the Galerkin transformation is used to transform back to the main nodepoint grid.

3.2. *Interpolation Formula*

The interpolation of a field $\Psi(X, Y)$ defined by (2) performed from the main nodepoint grid to the collocation grid, is carried out in two steps. The first step interpolates $\Psi(X, Y)$ to the Y-collocation grid, Fig. 1B:

$$\Psi^{\mathrm{Col}\,Y}_{\nu,\mu,\xi} = \frac{Y_{\mu,\xi} - Y_\mu}{Y_{\mu+1} - Y_\mu}\,\Psi_{\nu,\mu+1} + \frac{Y_{\mu+1} - Y_{\mu,\xi}}{Y_{\mu+1} - Y_\mu}\,\Psi_{\nu,\mu}. \tag{22}$$

$Y_{\mu,\xi}$, $\xi \in \{1, 2, 3\}$ are the Gaussian collocation points for the interval $(Y_\mu, Y_{\mu+1})$, introduced in (21). The Y-derivatives in the Y-collocation grid are defined as

$$\Psi^{\mathrm{Col}\,Y}_{Y\nu,\mu,\xi} = \frac{\Psi_{\nu,\mu+1} - \Psi_{\nu,\mu}}{Y_{\mu+1} - Y_\mu}. \tag{23}$$

In an X-interpolation step, $\Psi^{\mathrm{Col}}_{\nu,\rho,\mu,\xi}$, $\rho, \xi \in \{1, 2, 3\}$ is obtained at the collocation points, Fig. 1B, where ν and μ are the main points and ρ and ξ are indices of the collocation points,

$$\Psi^{\mathrm{Col}}_{\nu,\rho,\mu,\xi} = \frac{X_{\nu,\rho} - X_\nu}{X_{\nu+1} - X_\nu}\,\Psi_{\nu+1,\mu,\xi} + \frac{X_{\nu+1} - X_{\nu,\rho}}{X_{\nu+1} - X_\nu}\,\Psi_{\nu,\mu,\xi}. \tag{24}$$

In (24) $X_{\nu,\rho}$, $\rho \in \{1, 2, 3\}$ are the Gaussian integration points belonging to the interval $(X_\nu, X_{\nu+1})$.

The X-derivatives Ψ_X of $\Psi(X, Y)$ at the collocation points are computed by

$$\Psi^{\mathrm{Col}}_{X\nu,\rho,\mu\xi} = \frac{\Psi_{\nu+1,\mu,\xi} - \Psi_{\nu,\mu,\xi}}{X_{\nu+1} - X_\nu}. \tag{25}$$

The Y-derivatives at the collocation points are computed by (23) and (24) inserting Ψ_Y for Ψ (see Fig. 1).

3.3. *Scalar Products and Galerkin Operations*

In this section we consider the problem of transforming a field $\Psi_{\nu,\rho,\mu,\xi}$, given in the collocation grid to the main nodepoint grid, Fig. 1A, by Galerkin projection. We have to think of $\Psi_{\nu,\rho,\mu,\xi}$ as a rather general function, not necessarily represented as a linear spline (2). In our application $\Psi_{\nu,\rho,\mu,\xi}$ will represent a term in the discretized equation of motion (8).

The Galerkin projection of $\Psi_{\nu,\rho,\mu,\xi}$ will be done according to (14) in two steps. First the projection G^X_w in X-direction will be done. This will lead from the collocation grid, Fig. 1B, to the Y-collocation grid. To obtain G^X_w it is according to (10) and (12) necessary to obtain the Galerkin integrals. We use (10) and evaluate the Galerkin integrals in the X-direction, (12) for $Y = Y_{\mu,\xi}$, $\xi \in \{1, 2, 3\}$,

$$\Psi^{\mathrm{Col}\,Y}_{\nu,\mu,\xi} = \int_{X_{\nu-1}}^{X_{\nu+1}} dX\, \Psi(X, Y_{\mu,\xi})\, e_\nu(X)\, w(X, Y). \tag{26}$$

Using (20) and (21) rewritten for X, rather than Y, we obtain

$$\begin{aligned}\Psi^{\mathrm{Col}\,Y}_{\nu,\mu,\xi} = \sum_{\rho=1}^{3} (&\alpha_\rho\, \Psi^{\mathrm{Col}}_{\nu,\rho,\mu,\xi} e_\nu(X_{\nu,\rho})\, w_{\nu,\rho,\mu,\xi} \\ &+ \alpha_\rho\, \Psi^{\mathrm{Col}}_{\nu-1,\rho,\mu,\xi} e_\nu(X_{\nu-1,\rho})\, w_{\nu-1,\rho,\mu,\xi}).\end{aligned} \tag{27}$$

To evaluate (27) it is necessary to know the weight function w at the collocation points. For our applications with (8), w will be either H or 1. A scheme of roughly the same numerical cost is obtained by using Lobatto integration [47] with four integration points rather than the Gauss formula. If this is reduced to the use of three points, the integration by Simpson's rule is obtained, which is particularly efficient [45]. Three collocation points per grid interval with Gauss integration or four points with Lobatto integration are necessary in order to achieve exact integration in the presence of a weighting function w which is a linear spline. For operational implementation it will be worth investigating if a reduced integration with two collocation points is sufficiently accurate in practice.

The Galerkin transformation G_w^X in the X-direction is defined in (12). $G_w^X \Psi$ is defined by its gridpoint values $(G_w^X \Psi)_{\nu,\mu,\xi}$ in the Y-collocation grid, Fig. 1c. Equation (12) provides a linear equation for the determination of $(G_w^X \Psi)_{\nu,\mu,\xi}$, with $\Psi^{\mathrm{Col}\,Y}_{\nu,\mu,\xi}$ providing the right-hand side. To see this, use (12) with the functional representation for $G_w^X \Psi(X, Y)$:

$$G_w^X \Psi(X, Y) = \sum_\nu (G_w^X \Psi)_\nu (Y)\, e_\nu(X). \tag{28}$$

Note that the coefficients $(G_w^X \Psi)_\nu$ of the basis function e_ν in the X-direction are a function of Y. Inserting (28) into (12) and choosing $Y = Y_{\mu,\xi}$, we obtain the Galerkin equation in the X-direction,

$$\beta^1_{\nu,\mu,\xi}(G_w^X \Psi)_{\nu-1,\mu,\xi} + \beta^2_{\nu,\mu,\xi}(G_w^X \Psi)_{\nu,\mu,\xi} + \beta^3_{\nu,\mu,\xi}(G_w^X \Psi)_{\nu+1,\mu,\xi} = \Psi^{\mathrm{Col}\,Y}_{\nu,\mu,\xi} \tag{29}$$

with

$$\begin{aligned} \beta^1_{\nu,\mu,\xi} &= (e_{\nu-1}, e_\nu)_w^X \\ \beta^2_{\nu,\mu,\xi} &= (e_\nu, e_\nu)_w^X \\ \beta^3_{\nu,\mu,\xi} &= (e_{\nu+1}, e_\nu)_w^X. \end{aligned} \tag{30}$$

The β-term in (30) can be precomputed, if the weight w is fixed. For a weight $w(X, Y)$ variable in time, the β-term can be computed by Gaussian integration, in a similar way as the $\Psi^{\mathrm{Col}\,Y}_{\nu,\mu,\xi}$ in (27).

The scalar products in (30) are defined in (10), and they are evaluated for $Y = Y_{\mu,\xi}$.

Equation (29) can be solved by Gaussian elimination, as described by Ahlberg *et al.* [50]. Note that the index ν, μ has the same value for all terms in (29), and therefore (29) is only an equation in the index ν. The matrix on the left-hand side of (29) is a tridiagonal matrix, and the Gaussian elimination procedure will cost two multiplications for each gridpoint in the Y-collocation grid.

The Galerkin projection G_w^Y in the Y-direction will transform the field $\Phi_{\nu,\mu,\xi}$

$$\Phi_{\nu,\mu,\xi} = (G_w^X \Psi)_{\nu,\mu,\xi}, \tag{31}$$

defined in the Y-collocation grid, into a field $\Phi_{\nu,\mu}$ in the main nodepoint grid. $\Phi_{\nu,\mu}$ represents $G\Psi = G_w^Y G_w^X \Psi$.

In order to compute the result of the application of G_w^Y on $\Phi_{\nu,\mu,\xi}$, we have as a first step to compute the Galerkin integrals of Φ. We evaluate the right-hand side of (12) for $X = X_\nu$ and define

$$\bar{\Phi}_{\nu,\mu}^{Y} = \int_{Y_{\mu-1}}^{Y_{\mu+1}} dY\, \Phi(X, Y)\, e_\mu(Y)\, w(X, Y). \tag{32}$$

Using Eqs. (20) and (21) for $X = X_\nu$, we obtain

$$\bar{\Phi}_{\nu,\mu}^{Y} = \sum_{\xi=1}^{3} (\alpha_\xi \Phi_{\nu,\mu,\xi} e_\mu(Y_{\mu,\xi}) w_{\nu,\mu,\xi} + \alpha_\xi \Phi_{\nu,\mu-1,\xi} e_\mu(Y_{\mu-1,\xi}) w_{\nu,\mu-1,\xi}). \tag{33}$$

Note that in (33) $w_{\nu,\mu,\xi}$ is required on the Y-collocation grid, Fig. 1c.

For $G_w^Y \Phi(X, Y)$ we use the representation

$$G_w^Y \Phi(X, Y) = \sum_\mu (G_w^Y \Phi)_\mu (X)\, e_\mu(Y). \tag{34}$$

Equations (12) and (34) are used for $X = X_\nu$, and we define

$$(G_w^Y \Phi)_\mu (X_\nu) = (G_w^Y \Phi)_{\nu,\mu}. \tag{35}$$

Inserting (34) in (12), we obtain the following linear equations for $(G_w^Y \Phi)$:

$$\gamma_{\nu,\mu}^1 (G_w^Y \Phi)_{\nu,\mu-1} + \gamma_{\nu,\mu}^2 (G_w^Y \Phi)_{\nu,\mu} + \gamma_{\nu,\mu}^3 (G_w^Y \Phi)_{\nu,\mu+1} = \bar{\Phi}_{\nu,\mu}^{Y}. \tag{36}$$

The $\gamma_{\nu,\mu}^i$, $i = 1 \cdots 3$, are defined by

$$\begin{aligned} \gamma_{\nu,\mu}^1 &= (e_{\mu-1}, e_\mu)_w^Y \\ \gamma_{\nu,\mu}^2 &= (e_\mu, e_\mu)_w^Y \\ \gamma_{\nu,\mu}^3 &= (e_\mu, e_{\mu+1})_w^Y. \end{aligned} \tag{37}$$

If the weight w is variable in time, then γ can be determined using the Gaussian points according to Eqs. (20) and (21). Since the index ν has the same value for all terms in (36), it is a linear equation with respect to μ only. The Gaussian elimination procedure (see [50]) costs two multiplications for each gridpoint in the main nodepoint grid.

3.4. *The Computational Algorithm*

The result of the Galerkin projection in (8) is represented in the main nodepoint grid. This will then result in representations for $\dot{U}$, $\dot{V}$, $\dot{H}$ according to (2) and (3); i.e.,

$$\dot{U}(X, Y) = \sum_{\nu,\mu} U_{\nu,\mu} b_{\nu,\mu}(X, Y)$$

$$\dot{V}(X, Y) = \sum_{\nu,\mu} b_{\nu,\mu} b_{\nu\mu}(X, Y) \tag{38}$$

$$\dot{H}(X. Y) = \sum_{\nu,\mu} H_{\nu,\mu} b_{\nu,\mu}(X, Y),$$

where the $\dot{U}$, $\dot{V}$, and $\dot{H}$ are partial derivatives with respect to time.

Defining a quantity

$$e = \tfrac{1}{2}(U^2 + V^2) + H, \tag{39}$$

Eq. (8) can be written as

$$\begin{aligned} \dot{U} &= G_1(\eta V - G_2 e) \\ \dot{V} &= G_1(-\eta U - G_2 e) \\ \dot{H} &= G_2((UH)_X + (VH)_Y) \\ \eta &= V_X - U_Y + f. \end{aligned} \tag{40}$$

The typical algorithm for a time step in the numerical integration of Eq. (40) will be described in the following. At the start $U_{\nu,\mu}$, $V_{\nu\mu}$ and $H_{\nu,\mu}$ are given, and the result of the computation is $U_{\nu,\mu}$, $V_{\nu,\mu}$, and $H_{\nu,\mu}$. In the following, the number of indices of a given field U, V, or H will indicate the grid, on which the field variable is represented. For example, $U_{\nu,\rho,\mu,\xi}$ will mean the U-field, represented on the collocation grid. The following steps are performed in order to compute $\dot{H}_{\nu,\mu}$, $\dot{U}_{\nu,\mu}$, and $\dot{V}_{\nu,\mu}$:

(1) Gridpoint values for U, V, and H are computed on the Y-collocation grid, according to Eq. (22). This step results in coefficients $U_{\nu,\mu,\xi}$, $V_{\nu,\mu,\xi}$, and $H_{\nu,\mu,\xi}$.

(2) The Y-derivatives for U, V, and H are computed on the Y-collocation grid using Eq. (23). We obtain nodepoint values $U_{Y_{\nu,\mu,\xi}}$, $V_{Y_{\nu,\mu,\xi}}$, and $H_{Y_{\nu,\mu,\xi}}$.

(3) Using Eq. (24), the fields U, V, H, U_Y, V_Y, and H_Y are interpolated to the collocation grid. This results in coefficients $U_{\nu,\rho,\mu,\xi}$, $V_{\nu,\rho,\mu,\xi}$, $H_{\nu,\rho,\mu,\xi}$, $U_{Y_{\nu,\rho,\mu,\xi}}$, $V_{Y_{\nu,\rho,\mu,\xi}}$, and $H_{Y_{\nu,\rho,\mu,\xi}}$.

(4) The X-derivatives of U, V, and H in the collocation grid are computed using Eq. (25). This results in nodepoint values $U_{X_{\nu,\rho,\mu,\xi}}$, $V_{X_{\nu,\rho,\mu,\xi}}$, and $H_{X_{\nu,\rho,\mu,\xi}}$.

(5) According to Eq. (3.9), e is evaluated on the collocation grid,

$$e_{\nu,\rho,\mu,\xi} = \tfrac{1}{2}(U^2_{\nu,\rho,\mu,\xi} + V^2_{\nu,\rho,\mu,\xi}) + H_{\nu,\rho,\mu,\xi}. \tag{41}$$

(6) The scalar products e in the X-direction are evaluated in Equation (27) using a weight $w = 1$. The results are represented on the Y-collocation grid.

(7) By Gaussian elimination of Eq. (29) the Galerkin projection in the X-direction, $G_1^X e$ is computed. The result is again represented on the Y-collocation grid by coefficients $e_{\nu,\mu,\xi}$.

(8) The Galerkin integrals of e in the Y-direction are computed using Eq. (33) with a weight $w=1$. This is represented in the main nodepoint grid by coefficients $\bar{e}^Y_{\nu,\mu}$.

(9) Gaussian elimination in Eq. (34) is used to compute the Galerkin projection in the Y-direction. According to Eq. (14), this represents $G_2 e$. The coefficients are $(G_2 e)_{\nu,\mu}$.

(10) Steps 1, 2, 3, and 4 are repeated for $G_2 e$. The result is representative of $(G_2 e)_Y$ and $(G_2 e)_X$ on the collocation grid. These fields are represented by coefficients $(G_2 e)_{Y_{\nu,\rho,\mu,\xi}}$ and $(G_2 e)_{X_{\nu,\rho,\mu,\xi}}$.

(11) The time derivatives are computed on the collocation grid:

$$\begin{aligned}
\eta_{\nu,\mu,\xi} &= V_{X_{\nu,\rho,\mu,\xi}} - U_{Y_{\nu,\rho,\mu,\xi}} + f \\
\dot{U}_{\nu,\rho,\mu,\xi} &= \eta_{\nu,\rho,\mu,\xi} V_{\nu,\rho,\mu,\xi} - (G_2 e)_{X_{\nu,\rho,\mu,\xi}} \\
\dot{V}_{\nu,\rho,\mu,\xi} &= -\eta_{\nu,\rho,\mu,\xi} U_{\nu,\rho,\mu,\xi} - (G_2 e)_{Y_{\nu,\rho,\mu,\xi}} \\
\dot{H}_{\nu,\rho,\mu,\xi} &= -H_{\nu,\rho,\mu,\xi} U_{X_{\nu,\rho,\mu,\xi}} - H_{X_{\nu,\rho,\mu,\xi}} U_{\nu,\rho,\mu,\xi} \\
&\quad - H_{\nu,\rho,\mu,\xi} V_{Y_{\nu,\rho,\mu,\xi}} - H_{Y_{\nu,\rho,\mu,\xi}} V_{\nu,\rho,\mu,\xi}.
\end{aligned} \tag{42}$$

(12) The Galerkin projection steps 6, 7, 8, and 9 are repeated for the fields $\dot{U}$ and $\dot{V}$, resulting in representations on the main nodepoint grid $\dot{U}_{\nu,\mu}$ and $\dot{V}_{\nu,\mu}$. Steps 6 and 7 have to be carried out using the weight $w_{\nu,\rho,\mu,\xi} = H_{\nu,\rho,\mu,\xi}$ in the collocation grid, and steps 8 and 9 have to be carried out using the weight $w_{\nu,\mu,\xi} = H_{\nu,\mu,\xi}$ in the Y-collocation grid. The steps 6, 7, 8, and 9 are then repeated for H, using the weight $w=1$.

The discretization of Eq. (9) is just a subset of steps 1–12.

Once the time derivatives have been computed, the equation system is advanced in time using a leapfrog time differencing scheme. In order to investigate the nonlinear stability, we introduce a second-order smoothing operator with coefficient ε, which reduces both spatially small scales and the computational mode of the leapfrog scheme, acting as a low-pass filter,

$$\begin{aligned}
U^1_{\nu,\mu} &= U^0_{\nu,\mu} + \Delta t \dot{U}^0_{\nu,\mu} \\
V^1_{\nu,\mu} &= V^0_{\nu,\mu} + \Delta t \dot{V}^0_{\nu,\mu} \\
H^1_{\nu,\mu} &= H^0_{\nu,\mu} + \Delta t \dot{H}^0_{\nu,\mu}
\end{aligned} \tag{43}$$

$$\begin{aligned}
U^{n+1}_{\nu,\mu} &= U^{n-1}_{\nu,\mu} + 2\Delta t \dot{U}_{\nu,\mu} + \nabla^2_\varepsilon U \\
V^{n+1}_{\nu,\mu} &= V^{n-1}_{\nu,\mu} + 2\Delta t \dot{V}_{\nu,\mu} + \nabla^2_\varepsilon V \\
H^{n+1}_{\nu,\mu} &= H^{n-1}_{\nu,\mu} + 2\Delta t \dot{H}_{\nu,\mu} + \nabla^2_\varepsilon H.
\end{aligned} \tag{44}$$

A good indication of the numerical cost of the scheme cannot be obtained with the coarse resolution used here. The short vector lengths will result in a rather bad performance on vector machines, which will be uncharacteristic for the behavior with higher resolution. It may, however, be interesting to compare CPU-times for the two schemes Eqs. (8), (9). The code for the energy-conserving scheme was by a factor of 1.3 more expensive than that of the non-conserving scheme, when using the same number of collocation points. Since the scheme of Eq. (9) can be implemented with only two collocation points per grid interval, the non-conserving scheme of Eq. (9) will be, by a further factor of 2, more efficient than that of Eq. (8). The operational feasibility of the non-conserving scheme is proven, since the FE scheme proposed in [33] is more complicated than that of Eq. (9). For operational applications it would be interesting to investigate whether a reduced integration using only two collocation points per interval provides a reasonable approximation for the scheme of Eq. (8).

The smoothing operator ∇^2 is defined as

$$(\nabla_\varepsilon^2 U)_\nu^n = \varepsilon \cdot (-4U_{\nu,\mu}^{n-1} + U_{\nu-1,\mu}^n + U_{\nu+1,\mu}^n + U_{\nu,\mu-1}^n + U_{\nu,\mu+1}^n), \tag{45}$$

where ∇^2 is the discrete Laplacian operator with the difference that the central star grid point is using the previous time level $(n-1)$. Here ε is a coefficient controlling the amount of smoothing. In addition to the filter operation defined in (45) we will also use the Robert filter, as defined by Asselin [58]

$$\Phi_f^n = \Phi^n + \varepsilon_T(\Phi^{n+1} - 2\Phi^n + \Phi_f^{n-1}), \tag{46}$$

with Φ^n representing any field, Φ_f^n being the field after filtering, and ε_T being the filter coefficient. The operation defined in Eq. (46) will remove the computational mode of the leapfrog scheme and leave the physical mode untouched.

4. The Numerov–Galerkin Finite-Element Method with Triangular Elements

In order to test the conservation of total energy as well as potential enstrophy in finite-element method solutions of the non-linear divergent shallow-water equations a two-stage Numerov–Galerkin method described in [11, 31] was also applied using the same initial conditions and similar resolution as the rectangular Galerkin finite-element scheme. The two-stage Numerov–Galerkin technique seeks to obtain high accuracy for the estimation of the non-linear advective term by combining the two-stage Galerkin product advocated by Cullen and Morton [52] with a high-order approximation to the first derivative in the advective term.

In this approach we first calculate the Galerkin approximation to $\partial V/\partial X$ in the advection term, which we call Z as

$$\tfrac{1}{6}Z_{j-1} + \tfrac{2}{3}Z_j + \tfrac{1}{6}Z_{j-1} = \tfrac{1}{2}h^{-1}(V_{j+1} - V_{j-1}), \tag{47}$$

where h is the diameter of the element.

Then we calculate the product $w = U(\partial V/\partial X)$ as

$$\tfrac{1}{6}w_{j+1} + \tfrac{2}{3}w_j + \tfrac{1}{6}w_{j-1} = \tfrac{1}{12}(U_{j-1}Z_{j-1} + U_{j-1}Z_j + U_jZ_{j-1} + U_jZ_{j+1} + U_{j+1}Z_j + U_{j+1}Z_{j+1}) + \tfrac{1}{2}U_jZ_j. \quad (48)$$

The two-stage Galerkin approach results in a lower asymptotic truncation error [52, 31].

In the Numerov–Galerkin method we use a generalized spline to obtain a high-order compact implicit difference approximation of order $O(h^8)$ for the first-derivative (see Swartz and Wendroff [53, 31, and 11].

As the two-stage Numerov–Galerkin method is not conserving integral invariants such as total energy and potential enstrophy, a non-linear constrained optimization approach based on the augmented Lagrangian method (see Navon [54]) was used to enforce "a posteriori" conservation of these integral invariants. However, good conservations also obtained without this variational method and most of the experiments conducted in this research did not use variational "a posteriori" enforcing of conservation.

Also a Shuman [55] low-pass filter designed to filter out short wavelength components at the Nyquist cut-off limit was applied periodically every 6 or 12 h to the meridional V component of the velocity. The smoothing is applied sequentially in the X and Y directions.

Linear triangular elements are used in this method combined with a time-extrapolated Crank–Nicolson time-differencing scheme for integrating in time the system of ordinary differential equations resulting from the application of the Numerov–Galerkin scheme. The resulting system of equations is solved using an iterative method.

5. Test Integrations

5.1. *Test Integrations for the Rectangular Elements Scheme*

Test integrations were carried out with rectangular and triangular linear elements using the models described in Sections 2.3 and 4. The non-conserving version of the rectangular scheme is compared to the version conserving energy by an intrinsic method.

The triangular finite-element model uses the program **FEUDX1** [31]. The non-conserving version [31] of this scheme using only a Schuman low-pass filter is compared to a version [48] which conserves energy and potential enstrophy by restoring these integral constraints after time steps where a conservation threshold is violated. The method uses an augmented Lagrangian non-linearly constrained optimization technique.

We then investigate critical times for "blow-up" as well as the minimal critical dissipativity required in order to obtain stable long-term integrations of the shallow-water equations model.

For numerical schemes to be used for climate simulation or medium-range forecasts, it is considered of primary importance that the finite-time "blow-up" occurs rather late in the forecast and that as little energy loss as possible is incurred when non-linear instability is prevented by using a diffusion term.

Test integrations with the rectangular finite-element scheme were carried out on an 8×8 point main nodepoint grid with

$$\begin{aligned} \Delta X &= \frac{4 \cdot 11}{7} 10^5 m \\ f &= 10^{-4}\ \mathrm{s}^{-1} \\ \Delta t &= 900\ \mathrm{s}. \end{aligned} \tag{49}$$

The initial values were chosen similar to Grammeltvedt [49],

$$h(X, Y) = H_0 + H_1 \tanh \frac{9(Y - Y_0)}{2D} + H_2 \frac{1}{\cosh^2\left(\frac{9(Y - Y_0)}{D}\right)} \sin 2\pi X$$

with

$$\begin{aligned} H_0 &= 20000 m^2\ \mathrm{s}^{-2} \\ H_1 &= 4400 m^2\ \mathrm{s}^{-2} \\ H_2 &= 2660 m^2\ \mathrm{s}^{-2} \\ L = D &= 4400\ \mathrm{km} \\ \Delta X &= \frac{4000}{7}\ \mathrm{km} \\ f &= 10^{-4}\ \mathrm{s}^{-1}. \end{aligned} \tag{50}$$

The amplitude of the perturbation of the zonal flow in Eq. (50) was chosen to be larger by a factor of 2 than in [49], in order to make the stability test more difficult. Using these initial conditions, test integrations with no diffusion ($\varepsilon = 0$) were carried out.

The non-conserving rectangular finite-element scheme, Eq. (9) became unstable after 25 days, whereas the conserving scheme Eq. (8) became unstable after 35 days. Figure 2 shows the total energy (continuous line) for the first 23 days of integration

$$E = \frac{1}{2} \int dX\, dY\, H(U^2 + V^2 + H) \tag{51}$$

as a function of the time step for the non-conserving scheme Eq. (9). The integral in Eq. (51) is evaluated exactly using Eq. (8) or the corresponding interpolation assumption for triangular elements.

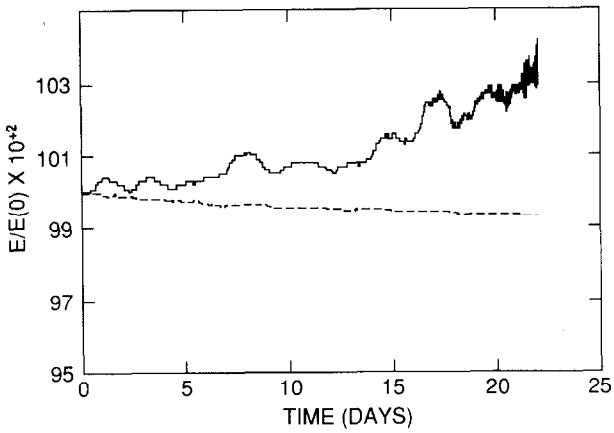

FIG. 2. Energy as function of the time (days) for the non-conserving rectangular finite-element scheme with ε (dissipation coefficient) = 0. $E(0)$ is the initial energy (continuous line) and for the conserving rectangular finite-element scheme (dashed line).

The corresponding diagram for the conserving scheme is shown in Fig. 2 (dashed line).

In the non-conserving case, the energy grows until non-linear instability occurs, whereas for the conserving scheme the energy is more constant and non-linear instability occurs much later.

Potential enstrophy,

$$Z = \int dX\, dY\, H\Theta^2 \tag{52}$$

with

$$\Theta = \frac{V_X - U_Y + f}{H},$$

is not conserved for the rectangular finite-element case. Its time development is shown in Fig. 3 for the non-conserving scheme (continuous line), and by the dashed line in Fig. 3 for the energy conserving scheme. The energy conserving scheme also conserves potential enstrophy better, particularly at the end of the forecast time. The onset of instability is preceded by considerable increases and decreases of potential enstrophy.

A diffusion term ($\varepsilon \neq 0$) as in Eq. (45) will make the time dependence of the energy monotonic and prevent non-linear instability, if ε is sufficiently large. There is a critical minimum value of ε to achieve this. ε values below this value have a small effect on the stability behavior of the scheme. This critical value of ε was found to be $\varepsilon = \frac{1}{4000}$ for the non-conserving scheme and $\varepsilon = \frac{1}{10000}$ for the conserving scheme. Integrations were performed for both schemes up to day 150 (see also [3]).

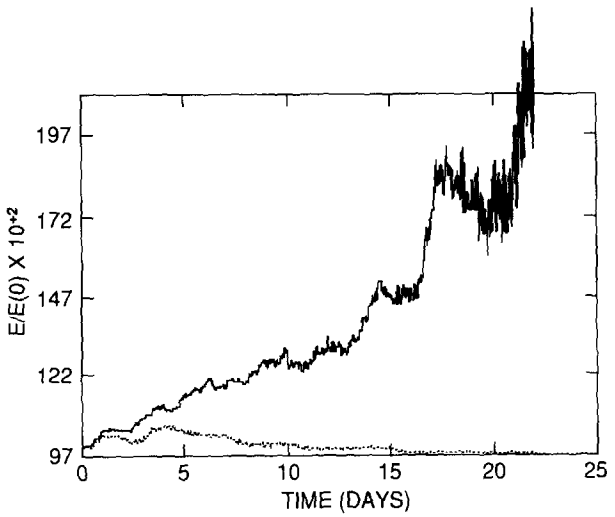

FIG. 3. Potential enstrophy as a function of the time (days) for the non-conserving rectangular finite-element scheme (continuous line), $\Delta t_{\mathrm{RECT}} = 900$ s, and for the energy conserving scheme (dashed line).

The time dependence of the total energy is shown in Fig. 4 for the non-conserving scheme (by the continuous line) and by a dashed line for the energy conserving scheme.

Both schemes suffer from a decrease of energy due to the dissipativity. A comparison of the two curves in Fig. 4 shows that there is a smaller loss of energy for the conserving rectangular finite-element scheme. To put the amount of energy loss into perspective it may be remarked that the non-conserving scheme according to Fig. 4 loses as much energy in the first 10 days of the forecast as the initial amount of the kinetic energy.

Figure 5 (continuous and half tone lines) shows the time diagrams for the potential enstrophy for long-term integrations using the non-conserving and conserving schemes, respectively.

A somewhat better conservation of the energy and enstrophy integral invariants by the energy conserving scheme can be observed.

The filter operator defined in Eq. (45) results in a spatial smoothing and a damping of the computational mode of the leapfrog scheme. It will, however, not damp the computational mode of the $2\Delta X$ wave, which changes sign both at alternate grid points and time levels. Three-dimensional atmospheric models normally contain a Robert filter as defined in Eq. (46) to damp the computational mode and, in addition, a spatial filter operation. In order to investigate the sensitivity of the result to the choice of the filter, tests using alternative smoothing schemes were done. Also, the use of a spatial filter alone is unsuitable with the leapfrog time integration scheme, when not using the Robert filter as given in Eq. (46). The application of a Shuman filter term [55] could not control the computational

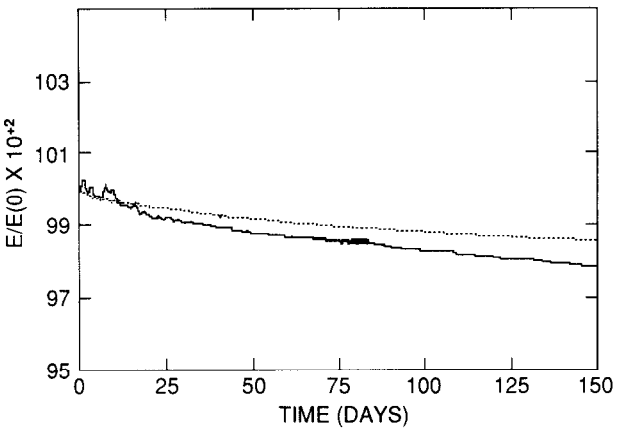

FIG. 4. Time dependence of total energy of the non-conserving scheme using a critical dissipativity coefficient $\varepsilon = \frac{1}{4000}$ (continuous line) and $\varepsilon = \frac{1}{10000}$ (dashed line).

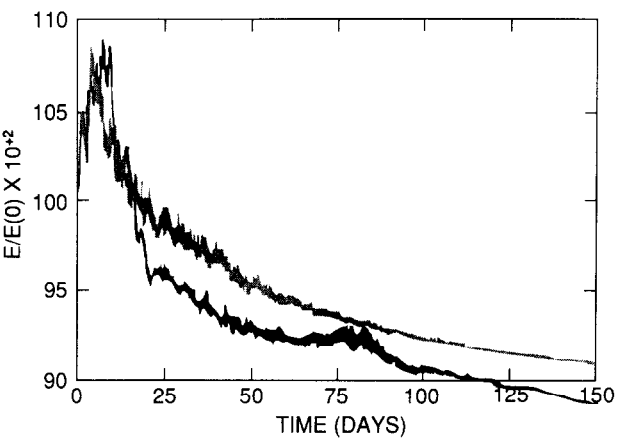

FIG. 5. Potential enstrophy as a function of time using the non-conserving rectangular finite-element scheme in a long-term integration (150 days) with critical dissipativity coefficient of $\varepsilon = \frac{1}{4000}$ (continuous line) and $\varepsilon = \frac{1}{10000}$ (dashed line).

TABLE I

Energy Normalized by the Energy at the Initial Time of the Conserving Model after $N = 2500$ and $N = 15000$ Time Steps for Different Values of ε_T

N	2500	15000
$\varepsilon_T = 0.05$	0.995	0.990
$\varepsilon_T = 0.1$	0.995	0.990

mode, which attained a strong amplitude in long integrations, and a considerable loss of energy occurred.

The application of a time filter according to Eq. (46) would stabilize the integration for the energy conserving scheme without applying any spatial smoothing. The critical value of ε_T to obtain stability is $\varepsilon_T = 0.5$. A further decrease of ε_T to $\varepsilon_T = 0.1$ did not increase the energy dissipation after 2500 and 15,000 timesteps. A comparison with Fig. 4 shows that the energy loss is somewhat smaller than with the smoothing scheme given by Eq. (45).

The non-conserving model could not be stabilized with the application of the Robert filter alone, and values of $\varepsilon_T = 0.1$ or $\varepsilon_T = 0.2$. However, the instability was postponed until after timestep $N = 5000$. The application of a Shuman filter every 100 timesteps was able to eliminate the instability. Energy values are given in Table II. A comparison with Table I shows that the dissipation is stronger than for the conserving scheme.

The experiments indicate that the conserving scheme can be stabilized with less dissipation of energy than the non-conserving scheme.

5.2. *Results of Test Integrations for the Triangular Elements Scheme*

The computations with the triangular elements scheme, using the program **FEUDX1** [31] were carried out on a 12×15 grid, using $\Delta X = 400$ km and a time step of $\Delta t = 1800$ s and initial values as in [49].

The critical "blow-up" time was about 12 days when no dissipation and no "a

TABLE II

Energy Normalized by the Initial Energy of the Non-conserving Scheme after $N = 2500$ and $N = 15000$ Time Steps Using Robert and Shuman Filters

N	2500	15000
E	0.987	0.982

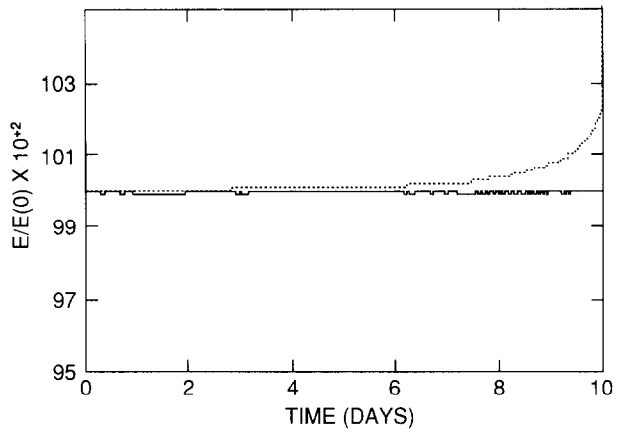

FIG. 6. Energy as a function of time (in days) normalized by the initial energy using the Numerov–Galerkin finite-element scheme with triangular elements with Schuman filter diffusion every 48 h (96 time steps) and constrained restoration (continuous line) and without Shuman filter diffusion (dashed line). The later blowups after about 11 model days.

posteriori" enforcement of conservation were used (see also Fairweather and Navon [48]).

This did not change in a significant way when only the constraint restoration

time steps) was used, both energy and potential enstrophy were conserved very well beyond the time of the "blow-up" as shown in Figs. 6 and 7 (continuous lines).

For comparison, Figs. 6 and 7 (dashed lines) give the same diagrams for the non-conserving version of the triangular finite-element model **FEUDX1** [31].

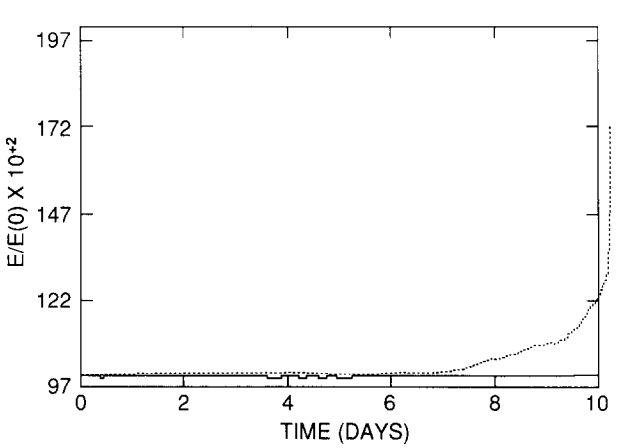

FIG. 7. Same as Fig. 6, but for potential enstrophy. Continuous line stands for the run with Schuman filter applied every 48 h and dashed line stands for the run without Schuman filter. The later run blowups after about 11 model days.

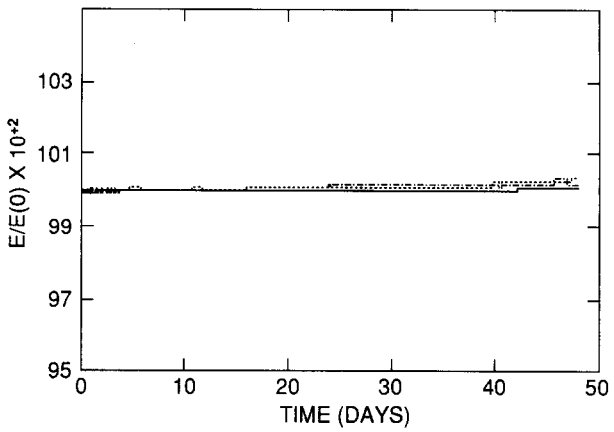

FIG. 8. As Fig. 6, but for a long run of 50 days with both Schuman filter type diffusion every 24 time steps and variational constraint restoration (continuous line), with Shuman filter applied every 96 time steps and no constraint restoration (dashed line) and with Shuman filter applied every 48 time steps (24 h) and no constraint restoration (dot and dash line). $\Delta t_{NG} = 1800$ s.

A moderate Shuman [55] filter term is sufficient in the Numerov–Galerkin triangular element scheme to achieve long-term stability, manifested by quasi-perfect conservation of total energy and enstrophy integral invariants and by a good evolution of the height fields.

Application of the Shuman filter [31], Shuman [55, 19] every 96 steps (48 h) is sufficient to achieve stability.

While variational constraint restoration alone does not prevent finite-time "blow-up," a small amount of dissipation in combination with the variational constraint restoration method [48] achieves very good restoration of constraints as shown in Figs. 8 and 9 (continuous lines).

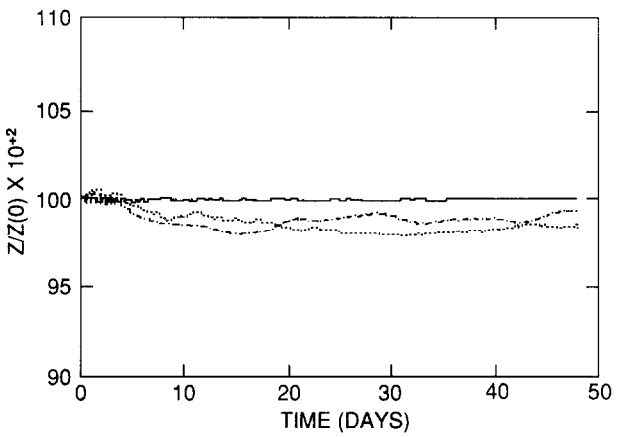

FIG. 9. As Fig. 8, but for potential enstrophy.

The conservation, as apparent in these figures, is much better than the one achieved by the intrinsic energy conserving model.

A number of tests were carried out to find the minimal amount of dissipativity required for the triangular element scheme (see Sadourny [3]). A 50-day run was conducted without constraint restoration applying the Shuman filter to the V-meridional component of the wind velocity every 96 time steps, i.e., every 48 h. A slight decrease in potential enstrophy was observed, while the energy was almost perfectly conserved (Figs. 8 and 9, dashed lines).

Another 50-day run was conducted differing from the first only in that the Shuman filter was applied every 48 time steps, i.e., a larger dissipativity factor is implied.

The results illustrated in Figs. 8 and 9 (dot and dash lines) show a slightly larger decrease in the potential enstrophy invariant (to the level of 98% of its initial value) while the total energy was again perfectly conserved.

However, the threshold of critical dissipativity is reached when one applies the Shuman filter only once every 192 time steps, i.e., every 96 h Then a finite-time "blow-up" occurs after 700 time steps (350 h), i.e., between 14 and 15 days of integration as illustrated in Figs. 10 and 11 (continuous lines).

To assess the impact of the variational constraint restoration, we conducted a model run using variational constraint restoration whenever the normalized potential enstrophy Z/Z_0 varied by more than 10^{-3} or the normalized total energy invariant E/E_0 varied by more than 10^{-3}, where E_0 and Z_0 denote the initial values of the total energy and potential enstrophy, respectively. We first applied the

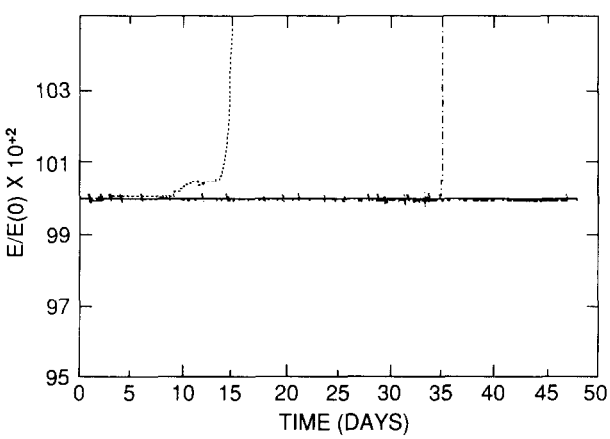

FIG. 10. Energy as a function of time (in days) normalized by the initial energy for the Numerov–Galerkin finite-element scheme with linear triangular elements with Shuman diffusion every 192 time steps and without variational constraint restoration. Finite-time "blow-up" after 700 time steps or about 15 days (continuous line), with variational constraint restoration and Shuman filter applied every 48 time steps (dashed line) and with Shuman filter applied every 96 time steps and with variational constraint restoration (dot and dash line). The later blows-up after 35 model days.

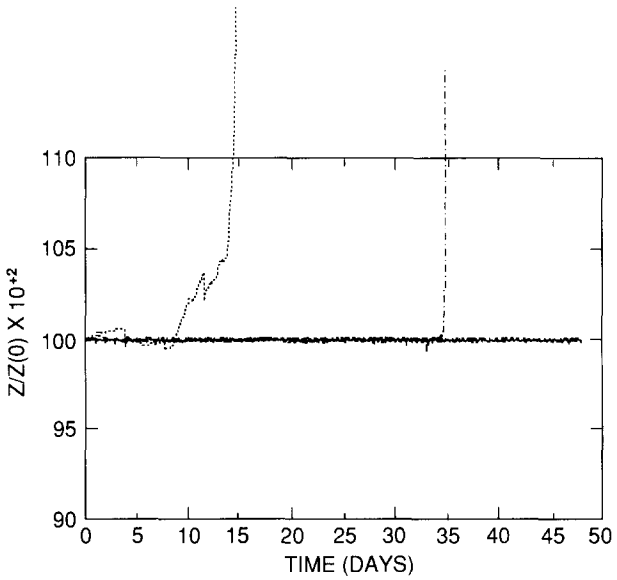

FIG. 11. Same as Fig. 10, but for potential enstrophy.

Shuman filter every 48 time steps (i.e., every 24 h) and a very good conservation was obtained for a 50-day run, as evidenced in Figs. 10 and 11 (dashed lines).

When we applied the Shuman filter (with the variational constraint restoration) every 96 time steps a finite-time "blow-up" occurred after 1750 time steps of 1800 s each, i.e., after about 35 days (see Figs. 10 and 11, dot and dash lines). This is evidence that the critical dissipativity required to stabilize the solution for beyond the critical "blow-up" time T_c is higher for the run where apart from dissipativity, we apply the variational constraint restoration.

The benefit of the variational constraint restoration method is therefore limited only to occurrences when it is applied with a moderate dissipativity. As has been recently observed (Navon [56], Takacs [57]), the constraint restoration approach, due to its being global, cannot control the wave–wave interactions and will allow unphysical energy or potential enstrophy cascades within the resolved spectrum, unless additional constraints on mean-wave number conservation are applied.

5.3. *Accuracy Tests of the Rectangular and Triangular Finite-Element Sschemes*

In order to provide a basis of comparison between the triangular element Numerov–Galerkin scheme and the rectangular element scheme, a very fine mesh (150×111), (i.e., 40 km. mesh-size) run of the Gustafsson [60] non-linear ADI method using finite-differences and a time step of 360 s was taken to give the definitive result. The fine mesh results are denoted by W_{EX}, where

$$W = (u, v, \phi)^T. \tag{53}$$

Representing the triangular Numerov–Galerkin finite-element method by W_{NG} and the rectangular finite-element scheme by W_{RECT}, respectively (or W_{FE} generically), the error for each of the schemes as in Gustafsson [60] is given by

$$G_{FE} = W_{FE} - W_{EX}, \tag{54}$$

while the relative error is defined to be

$$\text{relative error} = \frac{\|G_{FE}\|}{\|W_{EX}\|}, \tag{55}$$

where the norm $\|$ is defined as follows (see Gustafsson [60], Fairweather and Navon [48]):

Define a Hilbert space $\mathscr{H}$ by considering all vector functions of the form

$$W = (u, v, \phi)^T \tag{56}$$

with

$$W_{j,k} = W_{j,N_x+k}, \qquad V_{j,0} = V_{j,N_y} = 0 \tag{57}$$

(i.e., satisfying the boundary conditions).

The inner product of two vector-functions of the form (56)–(57) is defined by

$$(\alpha, \beta) = \Delta x\, \Delta y \sum_{j=1}^{N_x} \left\{ \sum_{k=1}^{N_y - 1} \alpha_{jk}^T \beta_{jk} + \frac{1}{2}(\alpha_{j0}^T \beta_{j0} + \alpha_{jN_y}^T \beta_{jN_y}) \right\} \tag{58}$$

and the norm $\|\cdot\|$ by

$$\|\alpha\|^2 = (\alpha, \alpha). \tag{59}$$

The relative errors for both the triangular and the quadratic f.e.m. schemes after 12 h of numerical integration are displayed in Table III.

The triangular Numerov–Galerkin scheme yields more accurate results but due to the fact that it was run at a higher resolution than the rectangular finite element scheme, this result is not conclusive.

TABLE III

Accuracy Results for the Rectangular and Triangular Numerov–Galerkin Finite-Element Schemes

Rectangular	Triangular Numerov–Galerkin
0.5×10^{-4}	0.21×10^{-4}

TABLE IV

Computational Efficiency of the Two Finite-Element Models in Terms of Respective Run-Times per Full Time Step per Grid Point

Computer	Triangular model (conservative)	Rectangular element model (conservative)
CYBER 205	9.1×10^{-4} s (16 × 12 grid)	8.9×10^{-4} s (8 × 8 grid)
CRAY-XMP-48	4.5×10^{-4} s (16 × 12 grid)	2.5×10^{-4} s (8 × 8 grid)

5.4. *Computational Efficiency*

The quadratic f.e.m. schemes and triangular N-G schemes were compared for computational efficiency by finding the run times in seconds per full time step per grid point. The results are displayed both for a run on a CDC CYBER 205 at Florida State University using automatic vectorization (VAST-2) and FORTRAN compilation under OPT, as well as on a CRAY-XMP-48 at the European Centre for Medium Range Weather Forecasts. The results are given for the conservative versions of both the triangular finite-element scheme as well as for the conservative rectangular model (see Table IV).

Were the runs to be conducted on a higher resolution grid, the computational advantage of the rectangular element model would have become more evident on the CYBER 205. The reason for the difference in the results on the two supercomputers, i.e., the CYBER 205 and the CRAY-XMP-48, respectively, lies in the different startup times for short vectors. The crossover between the two vector supercomputers occurs at a vector length of about 300.

6. Conclusions

Shallow-water equation models using energy conserving and non-conserving finite-element schemes were integrated for extended periods of time with a view to investigate their usefulness for climate models for medium-range forecasting.

While conservation of integral invariants of the shallow-water equations has been investigated for long-term integrations using finite-difference discretization schemes, this research addresses for the first time the issue of the behavior of finite-element schemes when used for long-term integrations of the shallow-water equations. The rectangular finite-element non-conserving scheme already displays a good nonlinear stability for long-term integrations.

The energy conserving scheme increased this further, allowing the reduction of the critical dissipativity parameter necessary to maintain stability in long-term integrations, by a factor of 2.5.

The Numerov–Galerkin triangular element scheme employs a variational constraint restoration scheme in combination with a Shuman filter applied at periodic intervals.

The length of these intervals controls the amount of dissipativity used in the model.

Application of the variational augmented-Lagrangian constraint restoration in the total absence of dissipativity did not increase the long-term non-linear stability and a finite-time "blow-up" occurred after about 12 days.

However, the variational constraint restoration in combination with a very small overcritical dissipativity (every 48 time steps or once every 24 h) yielded excellent conservation of the integral invariants for long-term integration periods exceeding 50 days. In another set of experiments the critical dissipativity required to overcome non-linear instability was determined by switching off the variational constraint restoration. It was then found that application of the Shuman filter every 96 time steps (i.e., every 48 h) still maintained good potential enstrophy and energy conservation as well as good model field behavior for long-term integration periods of more than 50 days. Application of the Shuman filter every 192 time steps (i.e., every 96 h) was below the critical dissipativity required to control non-linear instability at the short wave end of the spectrum, causing a finite-time "blow-up."

By comparison with finite-difference models of the shallow-water, we observe that a similar behavior is obtained as documented in Sadourny [3]. However, the finite-element schemes have intrinsically better conservation properties in the sense that smaller amounts of critical dissipativity are required to maintain long-term non-linear instability.

Future work should address the issues of the use of staggered variables as suggested by Williams and Zienkiewicz [43] as well as the investigation of the impact of resolution on the critical "blow-up" time. Spectral analysis of the energy and enstrophy cascades for finite-element models of the shallow-water equations should also be investigated.

In conclusion, the rectangular finite-element model appears to be more computationally efficient than the triangular finite-element model.

As far as accuracy is concerned, further tests at identical resolution of both models need to be carried out prior to assessing which of the schemes is more accurate.

For operational use in meteorology, both models appear adequate. If computational efficiency is the main concern, one should use the rectangular finite-element model. If accuracy is the main concern, one can use either the Numerov–Galerkin triangular finite-element model or the rectangular finite-element scheme.

Other issues, such as, better representation of geostrophic adjustment and better phase and amplitude error properties should also be addressed (see Neta and Navon [61]).

APPENDIX A: MATHEMATICAL PROPERTIES OF THE GALERKIN OPERATORS G_1 AND G_2

Galerkin operators approximate a general field $\psi(X, Y)$ belonging to the Hilbert space of square integrable functions by a field $\tilde{\psi}(X, Y)$ from a finite-dimensional subspace H_0 of H. The subspaces considered here consist of the $\tilde{\phi}$ defined by Eq. (2).

The defining equation of a Galerkin operator G is (5). If we consider the fields U, V, H used in Eq. (7), two Galerkin operators G_1 and G_2 will be used, whose definition differs in the choice of the weight function w in Eq. (6). Their definition is

$$(G_1\psi, b_{\nu\mu})_1 = (\psi, b_{\nu\mu})_1 \tag{A1}$$

with

$$(\psi, b_{\nu\mu})_1 = \int \psi(X, Y)\, b_{\nu\mu}(X, Y)\, H(X, Y)\, dX\, dY$$

and

$$(G_2\psi, b_{\nu\mu})_2 = (\psi, b_{\nu\mu})_2$$

with

$$(\psi, b_{\nu\mu})_2 = \int \psi(X, Y)\, b_{\nu\mu}(X, Y)\, dX\, dY. \tag{A2}$$

From these definitions it follows immediately that the Galerkin operators G have the properties of a projection operator in Hilbert space:

$$G^2 = G. \tag{A3}$$

$(Ga, b) = (a, Gb)$; $G(\lambda_1 a + \lambda_2 b) = \lambda_1 Ga + \lambda_2 Gb$, with λ_1 and λ_2 being real numbers and a, b, H, and G standing for either G_1 or G_2. The operator G_2, as defined in Eq. (A2) is the standard Galerkin operator as given by Strang and Fix [59]. The use of weights different from one in the definition (A1) of G_1, is equivalent to scalar products with multiplying functions $b_{\nu\mu}(X, Y)\, H(X, Y)$. These are different from the basis functions $b_{\nu\mu}(X, Y)$. The operator G_1 is therefore a Petrov–Galerkin operator.

The proof of energy conservation [18] for the scheme given by Eq. (8) is based on the properties of G given by Eq. (A3).

REFERENCES

1. A. ARAKAWA, *J. Comput. Phys.* **1**, 119 (1966).
2. Z. I. JANJIC, *Mon. Weather Rev.* **102**, 319 (1977).
3. R. SADOURNY, *J. Atmos. Sci.* **32**, 680 (1975).
4. A. ARAKAWA AND V. R. LAMB, *Mon. Weather Rev.* **109**, 18 (1981).

5. A. Arakawa and M. J. Suarez, *Mon. Weather Rev.* **111**, 34 (1983).
6. Y. Sasaki, *J. Comput. Phys.* **21**, 270 (1976).
7. Y. Sasaki and J. N. Reddy, *Int. J. Num. Methods Eng.* **16**, 149 (1980).
8. I. M. Navon, *Mon. Weather Rev.* **109**, 946 (1981).
9. I. M. Navon and R. de Villiers, *Mon. Weather Rev.* **111**, 1228 (1983).
10. Kálnay-Rivas, A. Bayliss, and J. Storch, *Beitr. Phys. Atmos.* **50**, 299 (1977).
11. I. M. Navon, *J. Comput. Phys.* **52**, 313 (1983).
12. N. A. Phillips, in Rossby Memorial Volume, edited by Rockefeller Institute Press, (1959), p. 501.
13. R. Shapiro, *Mon. Weather Rev.* **28**, 523 (1971).
14. B. Machenhauer, *GARP Publ. Ser.* 17, **2**, 124 (1979).
15. S. A. Orszag, *J. Fluid Mech.* **49**, 75 (1971).
16. W. Bourke, *Mon. Weather Rev.* **100**, 683 (1972).
17. W. Bourke, *Mon. Weather Rev.* **102**, 688 (1974).
18. J. Steppeler, *J. Comput. Phys.* **69**, 258 (1987).
19. F. G. Shuman, "A Method of Designing Finite-Difference Smoothing Operators to Meet Specifications," Tech. Memo 7, Joint Numerical Weather Prediction Unit, U.S. Weather Bureau (unpublished).
20. A. J. Simmons and D. M. Burridge, *Mon. Weather Rev.* **109**, 758 (1981).
21. J. Steppeler, *Mon. Weather Rev.* **115**, 1575 (1987).
22. J. Steppeler, A Galerkin finite-element spectral weather forecast model in hybrid coordinates, *Comput. Math. Appl.* **16**, 23 (1988).
23. A. N. Staniforth and R. W. Daley, *Mon. Weather Rev.* **105**, 1108 (1977).
24. Béland, M. J. Côté, and A. Staniforth, *Mon. Weather Rev.* **111**, 2298 (1983).
25. J. Steppeler, "Energy Conserving Finite Schemes for Numerical Weather prediction," in *Finite Element Methods in Flow Problems* (INRIA, Antibes, 1986), p. 359.
26. I. M. Navon, "Survey of Finite-Element Methods in Quasi-Linear Fluid Flow Problems," Report No. 240, National Research Institute for Mathematical Sciences, Pretoria, S. Africa, 1977 (unpublished).
27. A. N. Staniforth, *Int. J. Num. Methods Fluids* **4**, 1 (1984).
28. M. J. P. Cullen, *J. Inst. Math. Applic.* **11**, 15 (1973).
29. P. M. Gresho, R. L. Lee, and R. L. Sani, "Advection-Dominated Flows with Emphasis on the Consequences of Mass Lumping," in *Finite Elements in Fluids*, Vol. 3 (Wiley, New York, 1977).
30. M. J. P. Cullen, *J. Roy. Meteorol. Soc.* **100**, 555 (1974).
31. I. M. Navon, *Comput. Geosci.* **31**, 255 (1987).
32. A. N. Staniforth and H. L. Mitchell, *Mon. Weather Rev.* **105**, 154 (1977).
33. A. N. Staniforth and H. L. Mitchell, *Mon. Weather Rev.* **106**, 439 (1978).
34. M. J. Manton, *Tellus* **30**, 229 (1978).
35. A. K. Macpherson, M. H. Aksel, and P. D. Hilton, *Mon. Weather Rev.* **108**, 1183 (1980).
36. J. Mailhot and R. Benoit, *J. Atmos. Sci.* **39**, 2249 (1982).
37. D. J. Carson and M. J. P. Cullen, *Beitr. Phys. Atmos.* **50**, 1 (1977).
38. G. Fix, *SIAM J. Appl. Math.* **29**, 371 (1975).
39. D. C. Jespersen, *J. Comput. Phys.* **16**, 383 (1974).
40. K. A. Cliffe, *Int. J. Num. Methods Fluids* **1**, 117 (1981).
41. R. L. Lee, P. M. Gresho, S. T. Chan, and R. L. Sani, in *Proceedings, 3rd International Conference on Finite-Elements in Flow Problems, Barniff, Canada*, 1980 (edited by D. H. Norris), p. 431.
42. R. L. Lee, P. M. Gresho, S. T. Chan, R. L. Sani, and M. J. P. Cullen, *Finite Elem. Fluids* **4**, 21 (1981).
43. R. T. Williams and O. C. Zienkiewicz, *Int. J. Num. Methods Fluids* **1**, 81 (1981).
44. S. A. Orszag, *J. Atmos. Sci.* **27**, 890 (1970).
45. A. N. Staniforth and C. Beaudoin, *Int. J. Num. Methods Fluids* **6**, 317 (1986).
46. P. J. Davis and R. Rabinowitz, "Methods of Numerical Integration" (Academic Press, New York, 1975).

47. M. ABRAMOWITZ AND I. A. STEGUN, "Handbook of Mathematical Functions" (Dover, New York, 1970).
48. G. FAIRWEATHER AND I. M. NAVON, *J. Comput. Phys.* **37**, **1** (1980).
49. A. GRAMMELTVEDT, *Mon. Weather Rev.* **97**, 384 (1969).
50. J. H. AHLBERG, E.-N. NILSON, AND J. L. WALSH, *The Theory of Splines and Their Applications* (Academic Press, New York, 1967).
51. W. F. WEIGLE, "Energy Conservation of the Primitive Equations from a One-Layer Fluid," NCAR Cooperative Thesis N26, University of Michigan and Laboratory of Atmospheric Science, NCAR (unpublished).
52. M. J. P. CULLEN AND K. W. MORTON, *J. Comput. Phys.* **34**, 245 (1980).
53. B. K. SCHWATZ AND B. WENDROFF, *SIAM J. Num. Anal.* **11**, 979 (1974).
54. I. M. NAVON, "**FEUDX**, A Two-Stage, High-Accuracy Program for Solving the Shallow-Water Equations," in *Environsoft 86*, edited by P. Zannetti (Comput. Mech., Boston, 1986), p. 487.
55. F. G. SHUMAN, *Mon. Weather Rev.* **85**, 357 (1957).
56. I. M. NAVON, *Atmos. Phys.* **37**, No. 3, 1 (1987).
57. L. TAKACS, *Mon. Weather Rev.* **116**, 525 (1988).
58. R. ASSELIN, *Mon. Weather Rev.* **100**, 487 (1972).
59. G. STRANG AND G. FIX, *An Analysis of the Finite Element Method* (Prentice-Hall, Englewood Cliffs, NJ, 1973).
60. B. GUSTAFSSON, *J. Comput. Phys.* **7**, 239 (1971).
61. B. NETA AND I. M. NAVON, *J. Comput. Phys.* **81**, 277 (1989).