SHAPE PRESERVING INTERPOLATORS FOR SEMI-LAGRANGIAN TRANSPORT

David L. Williamson* and Philip J. Rasch
National Center for Atmospheric Research**
Box 3000
Boulder, Colorado 80307–3000

Summary:
In this report we summarize our current work evaluating shape preserving interpolation for use with semi-Lagrangian advection of scalar fields. We have compared the accuracy of a large number of shape-preserving interpolation schemes. The large number results from considering permutations of interpolating forms, derivative estimates and derivative constraints that have been introduced in the recent shape preserving interpolation literature. We compare:

- Five types of piecewise interpolating functions, all of an osculatory form involving data values and derivative estimates at the end points of the interpolation intervals.
- Seven methods of estimating derivatives, all of a local nature involving a limited number of data values surrounding the estimation point.
- Several derivative constraints appropriate to insure shape preservation within the interpolation interval. These constraints vary with interpolation form and include conditions for monotonicity and/or convexity/concavity.

We summarize the results of this intercomparison, then test the most promising interpolation forms with Cartesian one- and two-dimensional semi-Lagrangian advection of several test shapes. We extend the two-dimensional schemes to spherical geometry and illustrate that no serious problems arise there. Finally, we compare four methods of computing the departure point in spherical geometry. We evalute these with several hypothetical wind fields for which we can calculate the departure point exactly.

---

* Currently on collaborative leave with UCAR Visiting Scientist Research Program at National Meteorological Center, Washington, D.C.

** The National Center for Atmospheric Research is sponsored by the National Science Foundation.

1. *INTRODUCTION*

The spectral transform method has proven to be very good for calculating nonlinear, adiabatic fluid flow on the surface of a sphere. As a result, it has been adopted as the basis of many operational global numerical weather prediction and atmospheric general circulation models. The spectral approach, however, has serious deficiencies in physically representing the advection of fields which have large horizontal gradients. Water vapor, usually expressed as mixing ratio or specific humidity, is such a field. The problem is most noticeable as regions of negative mixing ratio which are inconsistent with physical parameterizations. Various computational fixes have been included in most models to eliminate these negative regions but they can result in significant computational transport. Less noticeable, but equally serious, are regions of overshoot. Unlike negative mixing ratios, it is not obvious from the field itself when overshooting has occurred and thus it cannot be corrected or even monitored before the physical parameterizations are invoked. However, the overshoot can erroneously interact with the parameterizations and produce, for example, spurious precipitation.

The over/undershoot can be attributed to three problems with spectral transport: namely, Gibbs phenomena, spectral truncation, and dispersion errors. The familiar Gibbs phenomena is introduced by the need to represent fields with discontinuities or near discontinuities by smooth global basis functions. Additionally, in the spectral transform approach the number of degrees of freedom in the spectral representation is less than in the gridpoint representation. Gridpoint fields reconstituted from the spectral representation are truncated and contain additional areas of over and undershoot. These areas of over/undershoot are especially noticeable in the fields where large gradients are induced by the forcing. They are not eliminated by finer resolution since the source and sink terms tend to remain discontinuous. Finally, even if the field could be represented exactly by the basis set, dispersion errors due to numerical inaccuracies can introduce over/undershooting. This type of over/undershooting is also present in finite difference solutions to the transport equation unless special attention is given to the numerical method.

We are investigating an alternative scalar transport approach to couple with global spectral transform nonlinear dynamical models. Our approach consists of semi-Lagrangian transport with shape preserving interpolation. Semi-Lagrangian transport is chosen because it has no stability condition restriction on the time step and thus does not suffer from the usual pole problem associated with explicit gridpoint schemes nor from phase errors associated with implicit gridpoint schemes. Shape preserving denotes a class of

interpolation methods that maintain certain properties suggested by the data. Such properties include monotonicity and convexity, and provide a means of avoiding the oscillations often seen in polynomial interpolation. These methods reduce or eliminate the overshooting possible with the semi-Lagrangian transport method.

We summarize our current work in this report. More details are included in papers being prepared for journal publication. In Section 2 we describe various shape preserving interpolators and summarize an evaluation of their relative performance when applied to several test functions. In Section 3 we show selected examples from applying the interpolators to one-dimensional semi-Lagrangian transport and summarize our more complete study. We consider uniform translation of specified test shapes. In Section 4 we summarize results from the application of the schemes that performed best in the first tests to two-dimensional semi-Lagrangian advection on a plane. The test case is uniform rotation of specified test shapes. We extend the two-dimensional advection on a plane to advection on the surface of the sphere in Section 5 and show that no serious problems are introduced by that geometry. The test case is solid body rotation of test shapes about an axis rotated with respect to the polar axis of the coordinate system. Finally, in Section 6 we compare several methods for calculating the departure point in spherical geometry.

2. *THE INTERPOLATION PROBLEM*

We begin by defining the grid $\{x_i\}_{i=1}^n$, $x_1 < x_2 < \ldots < x_n$, and the data values $\{f_i\}, f_i = f(x_i)$. It is also convenient to define the *discrete slopes*

$$\Delta_i = (f_{i+1} - f_i)/(x_{i+1} - x_i) \; . \tag{2.1}$$

The data are locally monotonic at x_i if

$$\Delta_{i-1}\Delta_i > 0 \; , \tag{2.2}$$

and locally convex if

$$\Delta_{i-1} > \Delta_i \; . \tag{2.3}$$

For concave data, the inequality in (2.3) is reversed. We define a piecewise interpolant $p \in C^K[x_1, x_n]$, with $K \geq 0$. On each subinterval $[x_i, x_{i+1}]$, let

$$\theta = (x - x_i)/h_i \qquad h_i = x_{i+1} - x_i \tag{2.4}$$

and

$$p(x) = p_i(\theta) \; . \tag{2.5}$$

The interpolant p is constrained to have the following interpolatory properties

$$p(x_i) = f_i\ , \qquad dp(x_i)/dx = d_i\ . \tag{2.6}$$

Here, d_i is some estimate of the derivative of f at the endpoints of the interval. The interpolant is specified on the subinterval in terms of the data f_i, and the derivative estimates d_i at the endpoints of the subinterval, that is

$$p_i(\theta) = p_i(\theta, f_i, f_{i+1}, d_i, d_{i+1}) \tag{2.7}$$

The interpolant thus adheres to the standard osculatory representation, although the functional form of p_i is not necessarily the usual Hermite cubic polynomial form. In order to reduce the number of schemes involved in the intercomparison, only interpolating forms which involve use of *local* information have been included, i.e., d_i is a function of a few surrounding values of f_i. In this fashion we have excluded from consideration many global schemes; for example, the classic C^2 cubic splines which minimize the integral of the curvature of the interpolant over the entire domain, exponential splines under tension (Spath, 1969), and global versions of the monotone, piecewise interpolants of Fritsch and Carlson (1980) and Delbourgo and Gregory (1983). These schemes require information from the entire domain to interpolate within any one subinterval. Following this restriction, schemes which differ from each other in the following three major ways are considered:

- The method of estimating the derivative is varied according to algorithms that have been suggested in the shape preserving literature.
- The type of interpolating function is varied to encompass cubic polynomials, rational functions, and quadratic Bernstein polynomials with extra knots.
- To guarantee monotonicity or concavity/convexity in the interpolating function, certain constraints are imposed on the derivative estimates. The appropriate constraint depends upon the interpolation form.

It is convenient to address these items in reverse order in the following subsections.

2.1 *Constraints on the derivatives*

Certain constraints must be imposed on the derivative estimates used in the interpolation schemes in order for the interpolants to maintain any convexity/concavity or monotonicity present in the data. The constraints are reviewed in this section, proceeding from the least to the most restrictive form. The constraints can be written in terms of restrictions on the derivative estimates d at the endpoints of an interval, as a function of the discrete slope Δ within the interval. Because of this, the constraint

on d_i based on Δ_{i-1} of the interval to the left may be different from that based on Δ_i of the interval to the right. One may choose to constrain the derivative differently for interpolation over the two intervals in which case the interpolant is C^0, or insist that constraints associated with both intervals be satisfied simultaneously, in which case the interpolant is C^1. When the constraint on d_i depends not only on the discrete slopes over the adjacent intervals Δ_{i-1}, Δ_i, but also the slope estimate d_{i-1}, or d_{i+1} at the other end of the interval, the C^1 interpolants become global. Such forms are not considered in this report.

The requirement that the derivative estimates bound the discrete slope for a C^0 interpolant

$$(d_i - \Delta_i)(\Delta_i - d_{i+1}) > 0 \qquad \text{(NCC0)} \qquad (2.8)$$

and lie between the adjacent discrete slopes for a C^1 interpolant

$$(d_i - \Delta_i)(\Delta_{i+1} - d_i) > 0 \qquad \text{(NCC1)} \qquad (2.9)$$

must be true for the interpolant to remain convex/concave. These requirements are identified as *Necessary Conditions(s) for Convexity/Concavity*, C^0 and C^1, respectively.

In order that the interpolating function be monotonic and C^0 the derivatives must satisfy the *Necessary Condition for Monotonicity*, (NCM0)

$$\begin{array}{ll} \text{sign}(d_i) = \text{sign}(\Delta_i) = \text{sign}(d_{i+1}) & \Delta_i \neq 0 \\ d_i = d_{i+1} = 0 & \Delta_i = 0 \end{array} \qquad \text{(NCM0)} \qquad (2.10)$$

that is, the derivative estimate at the end points must have the same sign as the discrete slope on the interval. For a C^1 interpolant

$$\begin{array}{ll} \text{sign}(\Delta_{i-1}) = \text{sign}(d_i) = \text{sign}(\Delta_i) & \Delta_{i-1}\Delta_i > 0 \\ d_i = 0 & \Delta_{i-1}\Delta_i \leq 0\,. \end{array} \qquad \text{(NCM1)} \qquad (2.11)$$

The derivative estimate at a data point must have the same sign as the discrete slopes surrounding it or be zero if the data are not monotonic at this point. This condition is the *Necessary Condition for Monotonicity*, C^1 (NCM1).

For the rational and piecewise quadratic interpolation forms to be discussed below, the necessary conditions, NCM0 and NCM1, are also sufficient conditions for monotonicity. Similarly, the NCC0 and NCC1 are sufficient conditions for convexity with these

interpolants. On the other hand, for Hermite cubic interpolants NCM0 and NCM1 are necessary but not sufficient for monotonicity and must be augmented by additional constraints on the derivatives.

Fritsch and Carlson (1980) have found both necessary and sufficient conditions for monotonicity of Hermite cubic interpolants. Let $\alpha = d_i/\Delta_i$, $\beta = d_{i+1}/\Delta_i$; then for $\Delta_i \neq 0$ the Hermite cubic interpolant will be monotonic if and only if (α, β) lies within the domain M_{ns} defined by the union of two domains

$$M_{ns} = M_e \cup M_b \tag{2.12}$$

where

$$M_e(\alpha,\beta) = \{\alpha,\beta : \phi(\alpha,\beta) \leq 0\} \tag{2.13}$$

$$M_b(\alpha,\beta) = \{\alpha,\beta : 0 \leq \alpha \leq 3, 0 \leq \beta \leq 3\}$$

and

$$\phi(\alpha,\beta) = (\alpha-1)^2 + (\alpha-1)(\beta-1) + (\beta-1)^2 - 3(\alpha+\beta-2)\,. \tag{2.14}$$

If $\Delta_i = 0, d_i = d_{i+1} = 0$ and the necessary condition discussed earlier is also sufficient. Embedded in this domain M_{ns} is the region M_b recognized independently by de Boor and Swartz (1977) which provides a sufficient condition for monotonicity of the Hermite cubic. This sufficient condition

$$0 \leq \alpha \leq 3\ ,\ 0 \leq \beta \leq 3 \qquad \text{(SCM)} \tag{2.15}$$

is easier to apply than the more general necessary and sufficient condition (M_e) in which α and β may be dependent on each other. Throughout the remainder of this report this simpler condition will be referred to as the *Sufficient Condition for Monotonicity* (SCM). As before, we define C^0 and C^1 forms depending on whether the derivatives d are bounded by just Δ of the interval being interpolated or by the Δ of the two adjacent intervals simultaneously.

At an extremum where the data are not monotonic, SCM1 limiting provides a severe restriction as d_i is then zero there. Hyman (1983) has relaxed the SCM1 limiting concept where the data reach a local extremum, and are not monotonic. He proposed the following limit on the derivatives.

$$d_i = \begin{cases} \min[\max(0,d_i), 3(0,\Delta_{\min})] & 0 < \Delta_{\min} \\ \max[\min(0,d_i), 3(0,\Delta_{\max})] & 0 > \Delta_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{2.16}$$

where

$$\Delta_{\min} = \min(\Delta_i, \Delta_{i-1}) \qquad \Delta_{\max} = \max(\Delta_i, \Delta_{i-1}) \ .$$

This allows for overshoot at local discrete extrema and thus is nonmonotonic but does provide some control of the overshoot and, in particular, prevents oscillations at the edges of flat plateaus.

2.2 *Interpolation forms*

Three types of interpolating functions are considered—all have appeared in the recent literature regarding shape preserving interpolation.

- Cubic polynomials (de Boor and Swartz, 1977; Fritsch and Carlson, 1980; Hyman, 1983; Fritsch and Butland, 1981)
- Rational functions (Gregory and Delbourgo, 1982; Delbourgo and Gregory, 1983, 1985)
- Quadratic Bernstein polynomials with extra knots (McAllister *et al.*, 1977; McAllister and Roulier, 1978, 1981).

The Hermite cubic and rational interpolating funtions can be described using the formalism of Delbourgo and Gregory (1985). Consider the function

$$p_i = P_i(\theta)/Q_i(\theta) \tag{2.17}$$

on the interval $0 \leq \theta \leq 1$, equivalently $x_i \leq x \leq x_{i+1}$, where

$$P_i(\theta) = f_{i+1}\theta^3 + (r_i f_{i+1} - h_i d_{i+1})\ \theta^2(1-\theta)$$

and
$$+(r_i f_i - h_i d_i)\ \theta(1-\theta)^2 + f_i(1-\theta)^3 \tag{2.18}$$

$$Q_i(\theta) = 1 + (r_i - 3)\theta(1-\theta) \tag{2.19}$$

We consider four choices of the parameter r_i

- If $r_i = 3$, p_i reduces to the standard *Hermite cubic* polynomal interpolation form. Recall that the interpolant will be monotonic if the d_i lie within the domain M_b.
- If $r_i = 1+(d_i+d_{i+1})/\Delta_i$, then P_i and Q_i reduce to quadratic polynomials, and p_i is identified as a *rational quadratic* interpolant. Delbourgo and Gregory (1985) have shown that provided d_i and d_{i+1} satisfy the NCM, p will be monotonic over the subinterval, otherwise this interpolant is not well defined.

- If $r_i = 1 + \max(C_i/c_i, C_i/c_{i+1})$ where $c_i = \Delta_i - d_i$, $c_{i+1} = d_{i+1} - \Delta_i, C_i = d_{i+1} - d_i$, then P_i and Q_i are cubic polynomials and p_i is identified as the *rational cubic interpolant version 1.*

- If $r_i = 1 + c_{i+1}/c_i + c_i/c_{i+1}$, the P_i and Q_i are again cubic polynomials and p_i is identified as the *rational cubic interpolant version 2.* Delbourgo and Gregory (1985) have shown that if the derivatives satisfy the convexity/concavity constraints NCC0 or NCC1 then both rational cubic versions will be concave/convex. If the derivatives satisfy the monotonicity constaints NCM0 or NCM1 then both versions will be monotonic. Delbourgo and Gregory (1985) have also shown that Version 2 is in general more accurate than Version 1.

The piecewise quadratic Bernstein polynomials with extra knots cannot be described using the previous formalism. This interpolant is constructed by piecing together two quadratic Bernstein polynomials within each interval, with the point of intersection (the extra knot) determined by a rather complex algorithm which cannot be succinctly described with a few equations or figures. Because of this, the reader is referred to the descriptions found in the series of original articles (McAllister *et al.*, 1977; McAllister and Roulier, 1978, 1981). The characteristics of the Bernstein polynomials, together with the algorithms developed for constructing the knot, the value of the interpolant at the knot, and the interpolant derivative at the knot guarantee that the interpolant will be monotonic provided NCM is satisfied, and convex/concave provided NCC is satisfied.

2.3 *Derivative estimation procedures*

Table 1 lists the algorithms used in estimating derivatives at the nodes. Several of the algorithms suggested in the literature for shape preserving interpolation which differ for unequally spaced data reduce to a common form when the data become equally spaced. Our first comparison uses equally spaced data, and therefore common algorithms are grouped together. The table also includes an algorithm identified as Cubic, which does not usually appear as a derivative estimate. This scheme arises by computing a cubic interpolant through the four nearest points. The slope at the nearest two points can then be written as a linear combination of the four surrounding data points. Such a scheme results in an interpolant which is only C^0 continuous. It is included because this form of interpolation is often used in semi-Lagrangian problems. The harmonic mean, geometric mean and Fritsch-Butland derivative estimates automatically satisfy the NCM and NCC constraints. The others generally must be modified to satisfy them.

TABLE 1.

Identifier	Algorithm	
Akima (A70, FC80, H83)	$d_i = \begin{cases} \frac{\alpha\Delta_{i-1}+\beta\Delta_i}{\alpha+\beta} \\ \frac{(\Delta_{i-1}+\Delta_i)}{2} \end{cases}$	$\alpha+\beta \neq 0$ $\alpha+\beta = 0$
	$\alpha = \lvert\Delta_{i+1}-\Delta_i\rvert,\ \beta = \lvert\Delta_{i-1}-\Delta_{i-2}\rvert$	
Arithmetic Mean (FC80, GD82, DG83, H83) *Deficient Spline*	$d_i = \frac{(\Delta_{i-1}+\Delta_i)}{2}$	
Geometric Mean (DG83)	$d_i = \begin{cases} \text{sign}(\Delta_i)\sqrt{\Delta_{i-1}\Delta_i} \\ 0 \end{cases}$	$\Delta_{i-1}\Delta_i \geq 0$ $\Delta_{i-1}\Delta_i < 0$
Harmonic Mean (FB84) *Rational Linear* (GD82) *McAllister-Roulier* (MR81)	$d_i = \begin{cases} \frac{2\Delta_{i-1}\Delta_i}{(\Delta_{i-1}+\Delta_i)} \\ 0 \end{cases}$	$\Delta_{i-1}\Delta_i \geq 0$ $\Delta_{i-1}\Delta_i < 0$
Fritsch-Butland (FB84, H83)	$d_i = \begin{cases} \frac{3\lvert\Delta_{i-1}\rvert\lvert\Delta_i\rvert}{\max(\Delta_{i-1},\Delta_i)+2\min(\Delta_{i-1},\Delta_i)} \\ 0 \end{cases}$	$\Delta_{i-1}\Delta_i \geq 0$ $\Delta_{i-1}\Delta_i < 0$
Cubic	$d_i = \begin{cases} \frac{(2\Delta_{i-1}+5\Delta_i-\Delta_{i+1})}{6} \\ \frac{(-\Delta_{i-2}+5\Delta_{i-1}+2\Delta_i)}{6} \end{cases}$	$x \in (x_i, x_{i+1})$ $x \in (x_{i-1}, x_i)$
Hyman (H83)	$d_i = \frac{\Delta_{i-2}-7\Delta_{i-1}+7\Delta_i-\Delta_{i+1}}{12}$	

Algorithms for derivative estimates as they simplify for evenly spaced data. Reference codes are as follows: A70–Akima (1970), DG83–Delbourgo and Gregory (1983), FB84–Fritsch and Butland (1984), FC80–Fritsch and Carlson (1980), GD82–Gregory and Delbourgo (1982), H83–Hyman (1983), MR81–McAllister-Roulier (1981).

2.4 *Intercomparisons of interpolation schemes*

The accuracies of the interpolation schemes have been tested by applying them to three shapes: a Gaussian, a cosine bell, and a triangle. These shapes were chosen because they

have similar forms, but may be successively more difficult to approximate accurately. The Gaussian is C^∞, the cosine bell is C^1 and the triangle is C^0. Tests using resolutions of ten points and forty points were performed.

The shapes were successively displaced 100 times, by 1/100 of the grid interval, and measurements of the accuracy were made. This was to establish the sensitivity of the representation to the relative position of the grid and test shape. The accuracy of any one scheme varied by at least a factor of five over the 100 realizations. We intercompared the schemes using the accuracy averaged over all realizations. We have tabulated the error statistics as a function of interpolation form, derivative approximation scheme and derivative limiter in Rasch and Williamson (1987). We included the seven derivative approximations, the five interpolation forms described above and various monotonicity and convexity constraints on the derivatives as appropriate for the interpolation form. We also considered the unconstrained versions for reference.

These tables of errors provide a staggering amount of information and the discussion justifying the conclusion about the interpolation from the tables is somewhat tedious. The tables and discussion are not repeated here. Out of the mass of numbers considered, there are logical inferences to be drawn relating the various schemes to each other. These conclusions may not be universal, as definite known properties of particular fields might be used to advantage in the interpolation scheme. Minor exceptions can be found in our tables that might imply some other scheme is ideal for such specific applications.

We begin by itemizing our conclusions regarding the interpolating functions.

- The Hermite cubic and the second version of the rational cubic interpolant appear to be the most useful interpolation formulas. The first version of the rational cubic interpolant is consistently inferior to the second.
- The Bernstein quadratic interpolant is generally of comparable accuracy to the rational form mentioned above. We found it to be somewhat more difficult to program for the various special cases, which results in a corresponding increase in the complexity of computer code and execution time.
- The rational quadratic interpolant is of comparable accuracy to the SCM limited Hermite cubic for monotonic data, but it does not allow the flexibility of the Hermite cubic near extrema, or allow for the concave/convex structure provided by versions of the rational cubic interpolant. For data which have an extremum, this scheme is not recommended, because there is no alternative to assuming the slope goes to zero at a discrete extremum. This results in much larger errors in the vicinity of the extremum, than the cubic, rational cubic and piecewise quadratic spline forms.

Conclusions regarding the derivative estimates are:

- The geometric mean, harmonic mean and Fritsch-Butland derivative estimates are consistently less accurate than the others. Their virtue is their simplicity. While they may result in visually pleasing interpolants they are generally of insufficient accuracy for many applications. The rational-linear derivative estimate, equivalent to the derivative estimate suggested by McAllister and Roulier (1981), and the harmonic mean estimate suggested by Fritsch and Butland (1984) for equally spaced data, is the least accurate of all the slope estimates. The Fritsch-Butland slope is always more accurate than the rational linear slope.
- The Akima approximation performs extremely well for data with small-scale features, but less well for the broader, more rounded shapes. Careful examination of the results suggests the Akima scheme is actually quite accurate in the vicinity of the extrema, and much less accurate over the rest of the domain.
- Except for the intersection of straight lines such as triangular peaks where the Akima estimate shines, the Hyman derivative estimate is the most accurate, followed generally by the cubic, then arithmetic. The disadvantage of the cubic derivative approximation is that it does not provide for a continuous interpolant while the others do.
- Monotonicity constraints generally improve the interpolation of monotonic data and of data approaching a flat plateau. These constraints degrade the interpolation near extrema by not allowing any overshoot that might be implied in the underlying data. The derivative estimate is constrained to be zero in the vicinity of extrema with the C^1 form. The C^0 continuity constraint is less serious in this regard than the C^1.
- Where strict monotonicity is not required, relaxation of the monoticity condition at any extremum seems desirable to allow the interpolant to form an extremum somewhere other than at a data point. Application of Hyman's limiter for the Hermite cubic or the convexity condition for the rational cubic seems desirable to prevent overshooting in the approach to a flat or nearly flat plateau.

3. *ONE-DIMENSIONAL SEMI-LAGRANGIAN ADVECTION*

Based on the evaluation of the various shape-preserving interpolators described in the previous section, we chose those which rated well for further tests in one-dimensional semi-Lagrangian advection by a uniform wind field. The Eulerian form of the evolution equation for the advection by a constant wind field of a scalar field in the absence of sources and sinks is

$$\frac{\partial f(x,t)}{\partial t} + v\frac{\partial f(x,t)}{\partial x} = 0 \tag{3.1}$$

where f is a scalar field such as mixing ratio, t is time and v is the constant advection velocity. Given f at time t, the solution at time $t + \Delta t$ is

$$f(x, t + \Delta t) = f(\hat{x}, t) \tag{3.2}$$

where

$$\hat{x} = x - v\Delta t\ . \tag{3.3}$$

The evaluation in this section focuses exclusively on the interpolation aspect of the solution. Given the departure point $\hat{x}$, an interpolation is made to find $f(\hat{x}, t)$. The interpolated value is then the forecast value $f(x, t + \Delta t)$. We consider only the Hermite cubic and second version of the rational cubic interpolant coupled with arithmetic, cubic, Hyman and Akima derivative approximations. Unmodified and appropriately limited derivative estimates are considered.

We integrate the advection equation with $v\Delta t/\Delta x$ equal to $\pi/12$. We chose an irrational number so that the relative position of the test field and grid would change with time and given a long enough integration would become almost uniformly distributed over the domain. The equation was integrated for 1000 time steps with the solution plotted every 200. The plots are superimposed using a coordinate system that moves with the advecting velocity. The test shapes deform due to errors in the interpolation. The line code in the figures is such that the shorter the pattern members, the later in time the solution.

Rasch and Williamson (1987) present results from cosine bell and square wave initial conditions, with ten nonzero gridpoints. We select a few examples of the square wave tests to illustrate here the general properties observed. The figures show results from arithmetic, cubic, Hyman and Akima derivative estimates ordered top to bottom.

Figure 1 shows results from the Hermite cubic interpolant. The left column shows the unlimited forms of the derivative approximations. Strong oscillations greater than 10% of the true signal are evident with the Hyman and arithmetic derivative estimate. The cubic derivative estimate shows less over/undershoot and the Akima estimate the least. The Akima version seems to be evolving toward a peaked shape. Figure 1 (center column) shows results from the Hermite cubic interpolant when the derivative estimates are modified to satisfy the SCM0. The monotonicity condition has improved the solution by eliminating the overshoot at the expense of increased diffusion and a decrease in the maximum value, at least for the arithmetic and cubic derivatiave estimates. The Hyman

and Akima derivative versions are improved by the limiting procedure, with little or no increase in diffusion of the shape. Imposition of the SCM1 constraints (right column) results in a very substantial increase in the phase error of all versions except the cubic derivative estimate. However, this interpolator is C^1 only where the derivative estimates were modified.

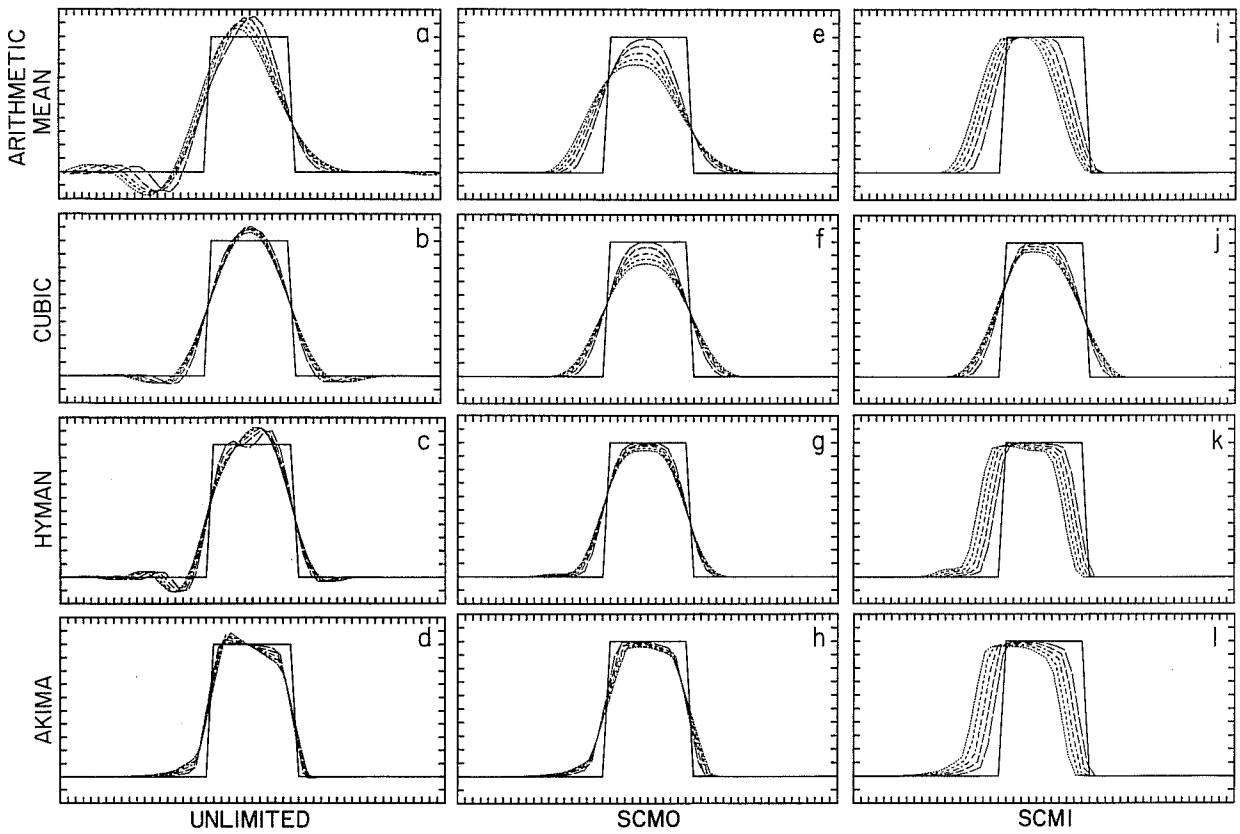

Fig. 1. Semi-Lagrangian advection with Hermite cubic interpolation of an initial square wave shown every 200 time steps (superimposed using a coordinate system that moves with the advecting velocity). The line code is such that the shorter the pattern members, the later in time the solution (the shortest pattern occasionally bleeds into an apparent solid line). Left column is for unlimited derivative estimates, center for estimates modified to satisfy SCM0 and right for SCM1. Top row for arthmetic mean derivative estimate, second row for cubic, third for Hyman and bottom for Akima.

The rational cubic interpolant (Fig. 2) results in a more rounded profile for the unlimited derivative estimates (left column) but provides a substantial reduction in the over/undershooting and dispersion when compared to the Hermite form (Fig. 1, left column) for the arithmetic, cubic and Hyman derivative estimates. Thus we consider this solution to be better. On the other hand, the solution using the Akima derivative estimate is worse with the rational than with the Hermite cubic interpolant. The C^1 necessary condition for monotonicity (Fig. 2, right column) increases the diffusion of the interpolation with the polynomial derivative estimates but leaves that with the Akima estimate relatively unaffected. The C^0 version (not shown) is almost identical to the C^1.

Note that the C^1 monotonic rational interpolant does not show the increase in phase error seen with the C^1 Hermite monotonic interpolant in Fig. 1, right column.

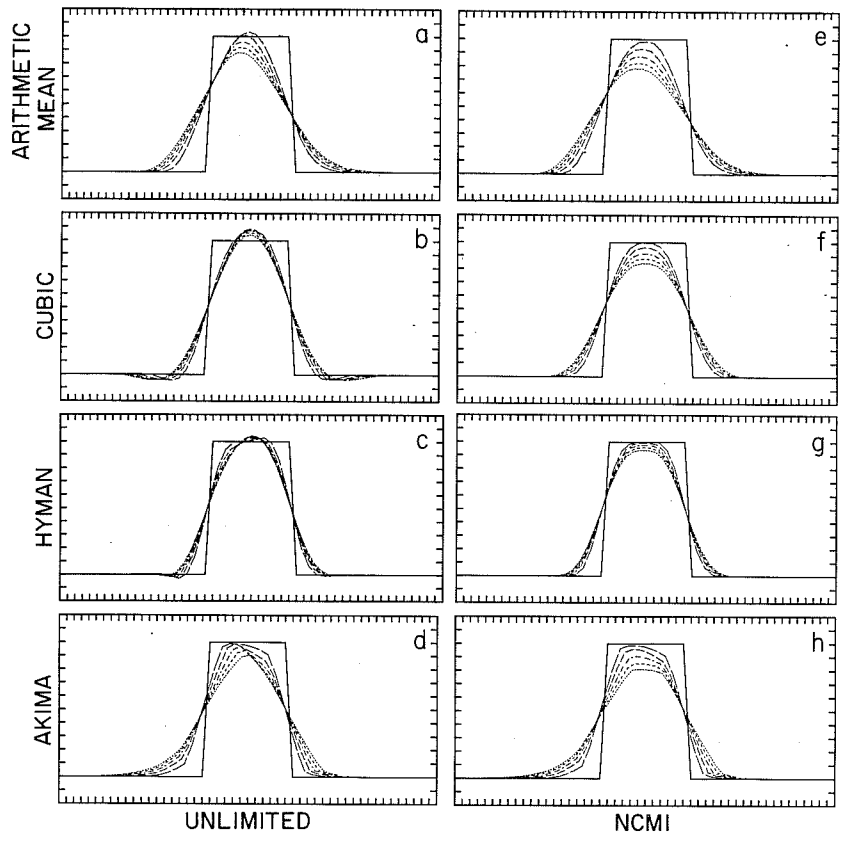

Fig. 2. As in Fig. 1 except with rational cubic interpolation. Left column is for unlimited derivative estimates and right for estimates modified to satisfy NCM1.

4. *TWO-DIMENSIONAL SEMI-LAGRANGIAN ADVECTION IN A PLANE*

We have also evaluated the Hermite and rational cubic interpolants for advection in two-dimensional Cartesian geometry. We consider three methods of applying shape preserving interpolation to the two-dimensional semi-Lagrangian advection problem. The first is to split the two-dimensional advection operator into two one-dimensinal operators via explicit fractional time steps or time-splitting. This approach was used by Purnell (1976) following Strang (1968) to obtain a scheme which was second order in time. Each fractional time step involves one-dimensional interpolation only.

The second method involves monotone piecewise bicubic interpolation. Carlson and Fritsch (1985) have extended their univariate piecewise cubic interpolation (Fritsch and Carlson, 1980) to two dimensions. Their method involves function values and the x, y

and xy derivatives at the data points. They define a two-dimensional monotonicity condition then derive conditions on the derivatives which ensure that the two-dimensional interpolating function is a monotonic surface by their definition within the interpolation rectangle. We add conditions to ensure that over the entire domain the surface is continuous and, if desired, smooth.

The third method considered involves two-dimensional interpolation by tensor products of one-dimensional shape preserving schemes. This provides a straightforward application of the one-dimensional interpolants considered in the previous section. However, the shape preserving interpolants are nonlinear depending on the data themselves, and any analysis of the characteristics of the two-dimensional surface implied is difficult.

These methods were evaluated for solid body rotation of a square block 11 points wide centered 12 grid intervals from the axis of rotation and a cosine bell of radius seven gridpoints centered 15 grid intervals from the center of rotation. The rotation rate was 2.5° per time step with the solution examined after two rotations (288 time steps). Details of the tests are presented in Williamson and Rasch (1987). We illustrate the general results with selected examples of the square block tests.

Figure 3 shows the solution for the two-dimensional Hermite cubic (method 2, upper row) and the time-split Hermite cubic (method 1, lower row). The cubic derivative approximation is used for both. The left side shows the unlimited forms and the right the SCM0 limited form. To economize on space we show the results on one quarter of the domain only, centered on the exact solution. The center of rotation is at the top of the figure. We use a contour interval of 0.1 with the 0.0 and 1.0 contours wider. In addition, we stipple the region which contains the contours from 0.0 to 1.0 of the exact solution, i.e., inside the stippled area the exact solution is 1.0 and outside it is 0.0. The contour routine would spread the contour lines linearly across the stippled area. It is immediately apparent that the time split cubic has larger errors than the Hermite bicubic with both the monotonic and original forms. The time split form is more highly diffused and introduces a position error by shifting the center of the feature away from the axis of rotation. We consider these two errors serious enough to exclude the time split approach from further consideration for two-dimensional advection.

The monotonic approach of Carlson and Fritsch (1985) in the upper right panel produces a reasonable solution. It is properly centered but shows slightly less straight sides than the standard bicubic on the left. Nevertheless, the monotonic form does not have the undesirable over and undershooting. The minimal extra diffusion associated with the monotonic form compared to the standard nonmonotonic form seems acceptable.

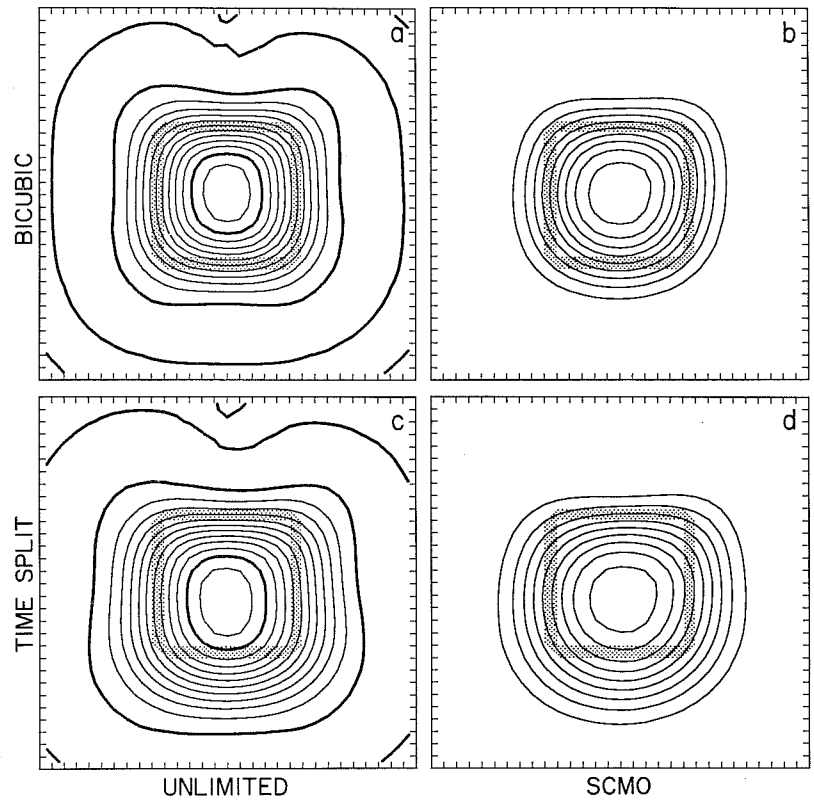

Fig. 3. Two-dimensional semi-Lagrangian advection with Hermite cubic interpolation of an initial square block after two rotations. Top row for bicubic interpolation, bottom for time split forecast. Left column with unmodified derivatives, right with derivatives modified to satisfy SCM0.

Several considerations discourage us from following this constrained surface approach further. Our ultimate goal is to apply these schemes to three-dimensional advection in the atmosphere. Although, it seems possible to extend the Carlson and Fritsch (1985) approach to three dimensions, the constraints on the various derivatives will be even more complicated than the two-dimensional case. In addition, we have seen above that other derivative approximations such as those of Akima (1970) and Hyman (1983) produce more accurate solutions than the cubic. These approximations would have to be extended to the cross-derivatives for the multidimensional case. Yet another detriment is that the shape preserving rational cubic interpolation, which also looks very promising in one-dimensional tests, has not been extended to multidimensional interpolation. Therefore, we now consider the tensor product of one-dimenional monotonic interpolants (method 3).

Figures 4–6 show the Hermite and rational cubic interpolants coupled with the cubic, Hyman and Akima derivative approximations. Figure 4 shows the solutions when the derivatives are unconstrained. Figure 5 shows the C^0 monotonic forms so that the

derivatives are modified to satisfy SCM0 with the Hermite interpolant and NCM0 with the rational. Figure 6 shows the corresponding C^1 forms. These figures are discussed in detail in Williamson and Rasch (1987) along with matching ones from the cosine bell tests. In addition, the conservation characteristics of the solutions are also discussed there. We present only an overview below.

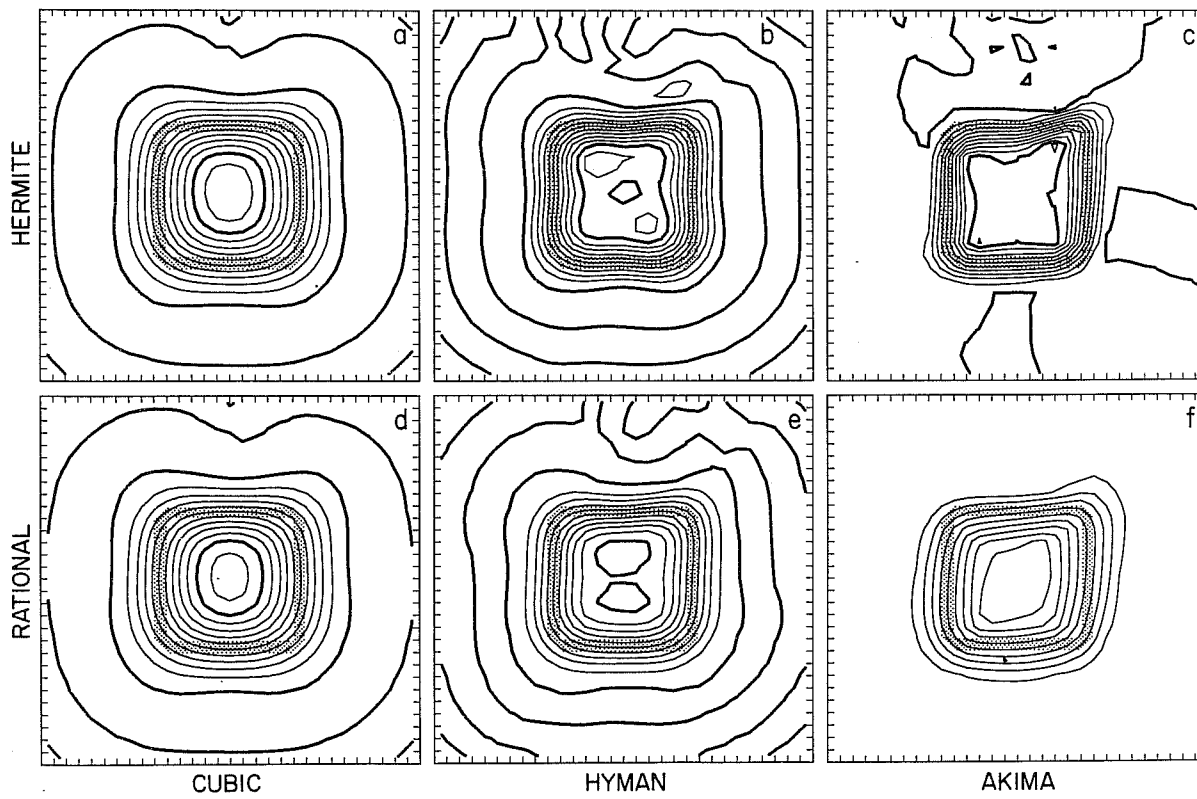

Fig. 4. Two-dimensional semi-Lagrangian advection of initial square block after two rotations with tensor product interpolation forms. Top row for Hermite cubic, bottom for rational cubic. Left column for cubic derivative approximations, center for Hyman and right for Akima.

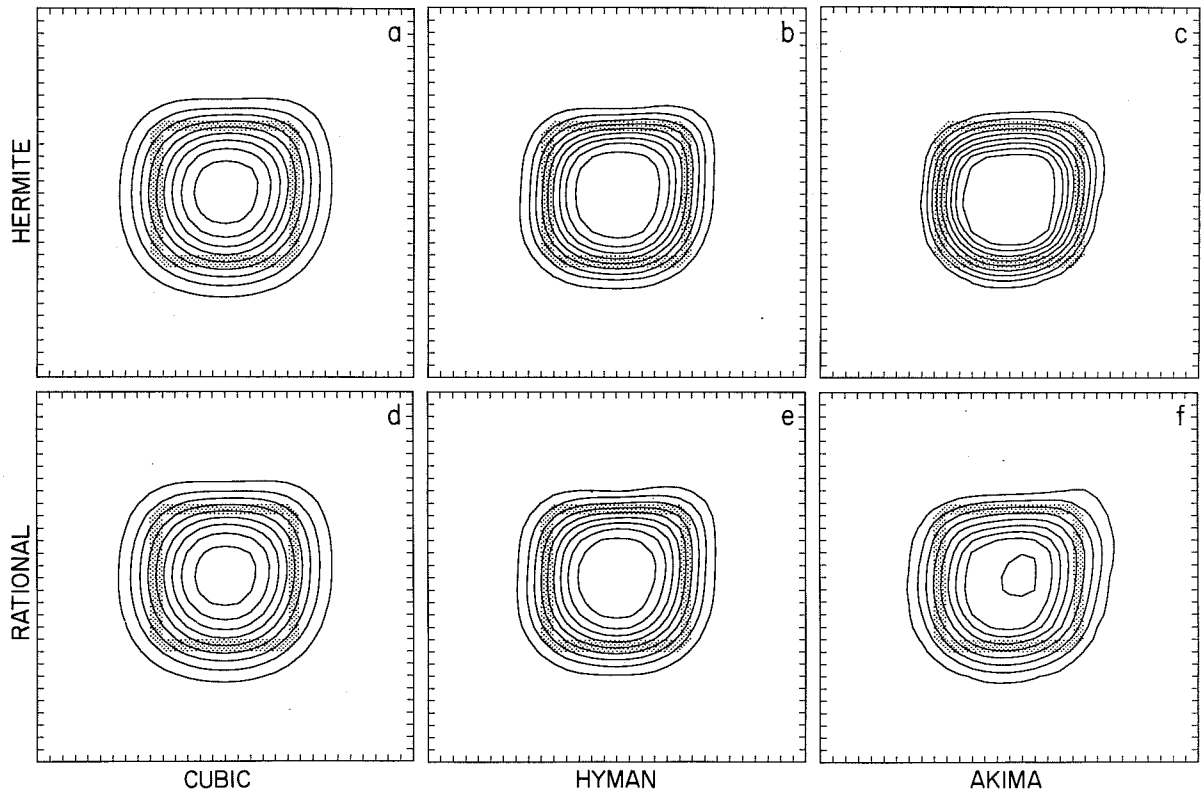

Fig. 5. As Fig. 4 except derivatives modified for C^0 monotonicity, *i.e.*, derivatives for Hermite modified to satisfy SCM0 and those for rational, NCM0.

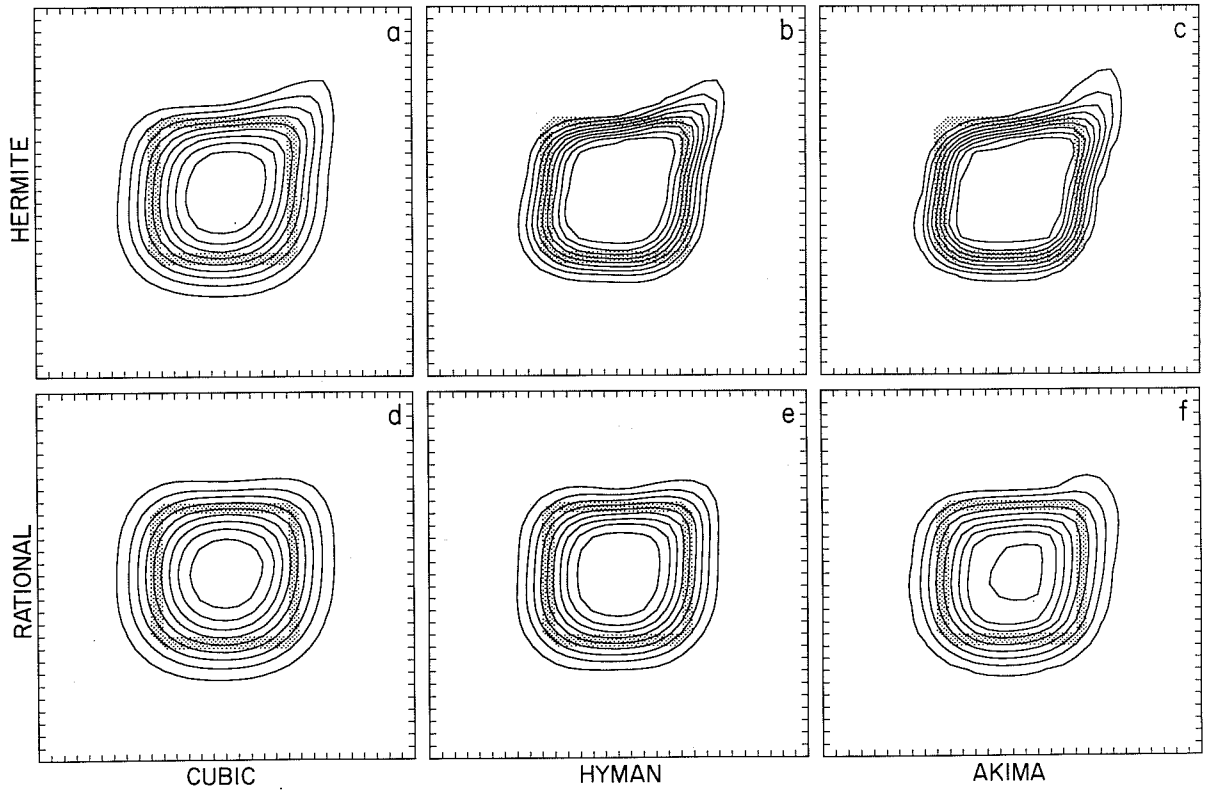

Fig. 6. As Fig. 4 except derivatives modified for C^1 monotonicity, *i.e.*, derivatives for Hermite modified to satisfy SCM1 and those for rational, NCM1.

Visual inspection of the solution shapes shown in Figs. 4–6 suggests that the best monotonic form is the Hermite cubic with Akima derivative approximation modified to satisfy the SCM0. Of the monotonic schemes, this combination produces minimal distortion, sharpest gradients and least diffusion of the original nonzero region. Second best is the rational cubic with the Hyman derivative approximation modified to satisfy NCM0 or NCM1. If C^1 continuity is desired this combination would be best. The C^1 forms of the Hermite cubic show strong distortion of the shape due to differential phase error. If monotonicity were deemed less important than sharp gradients and straight sides then the Hermite cubic with unmodified Akima derivative approximation would be best, but at the expense of distortion due to differential phase error. The Hyman derivative approximation with either the rational or Hermite interpolants does not show such distortion at the expense of less steep gradients. The rational form has less overshoot than the Hermite, but slightly weaker gradients. In general the results from the cosine bell test cases agree with those from the square block.

5. *SEMI-LAGRANGIAN ADVECTION ON A GAUSSIAN GRID*

As mentioned in the introduction our ultimate goal is to couple semi-Lagrangian advection of scalar fields such as moisture with a global spectral transform dynamical model. Therefore, the semi-Lagrangian scheme should be based on a Gaussian grid. Ritchie (1987) has presented a method for semi-Lagrangian advection on a Gaussian grid. We adopt a slightly different approach. In this section we show that no problems arise in adapting the shape preserving interpolation schemes to spherical geometry or to the slightly nonuniform Gaussian grid used in spectral transform models. The interpolation in longitude (λ) is on a uniform grid with periodicity assumed for boundary conditions. The latitudinal (ϕ) Gaussian grid is not uniform but presents no problem as none of the schemes we have considered assumes a uniform grid. Although we presented the derivative approximations in Table 1 for a uniform grid they are easily extendable to nonuniform grids. Near the poles we must be a little careful. Boundary conditions at the poles are implemented by defining gridpoints across the pole, which are needed for the interpolation operators. These points are simply provided by the gridpoints at the appropriate latitude π radians from the longitude of interest (λ_i), i.e.,

$$\psi\left(\lambda_i, \frac{\pi}{2} + \alpha_i\right) = \pm\psi\left(\lambda_i \pm \pi, \frac{\pi}{2} - \alpha_i\right) \tag{5.1}$$

where the latitudes near the pole are rewritten $\phi_i = (\pi/2) - \alpha_i$. The sign in (5.1) is taken to be positive for scalar fields and negative for vector components. In addition we define a pole point $(\lambda_i, \pi/2)$ or row of pole points, not normally part of the Gaussian grid, so that in the limit as interpolation points approach the pole, the interpolated values satisfy

certain physical continuity conditions. These are that the pole be a constant for a scalar field or consist of wavenumber one only for vector components. The pole row is defined by interpolating the zonal average or wavenumber one component, as appropriate, from the adjacent rows. The same interpolation scheme is used to obtain this pole point as is used in the interpolation of the field itself.

With the extensions described above, all the schemes considered in the plane case have been examined in spherical geometry with the test case of solid body rotation of a cosine bell. We have considered geometries in which the axis of rotation is not coincident with the polar axis of the spherical coordinate system. In general the relative properties of the schemes are not affected by the spherical geometry. The results from advecting the cosine bell on the sphere agree with those from the cosine bell on the plane.

We present only two examples here to illustrate that the pole does not present a problem (Fig. 7). They are based on a 64×128 Gaussian grid normally used with T42 spectral resolution. The test shape is a cosine bell of amplitude 1 and radius 7 $(2\pi)/128$ radians centered initially on the equator. We consider solid body rotation over the pole at a rate of $(\pi/64)/2$ radians per time step. The figures present contours of the solution on an orthographic projection with the perspective centered over the center of the true solution, thus the true solution is always in the center of the figure. Our convention is that the test shape is moving toward the top of the figure. The contour interval is 0.1 with 0.0 and 1.0 contours wider. (No 0.0 or 1.0 contours appear on the figures shown here.) The region contained between the 0.1 and 0.9 contours of the true solution is stippled for reference.

The left side of Fig. 7 shows the solution after one complete rotation (256 time steps) from the Hermite cubic interpolant with the Akima derivative estimates modified to satisfy the SCM0. The right side shows that from the rational cubic interpolant with the Hyman derivative estimates modified to satisfy the NCM1. Both tend to elongate the circular structure. The Hermite/Akima combination produces a slightly smaller size than the rational/Hyman accompanied by stronger gradients.

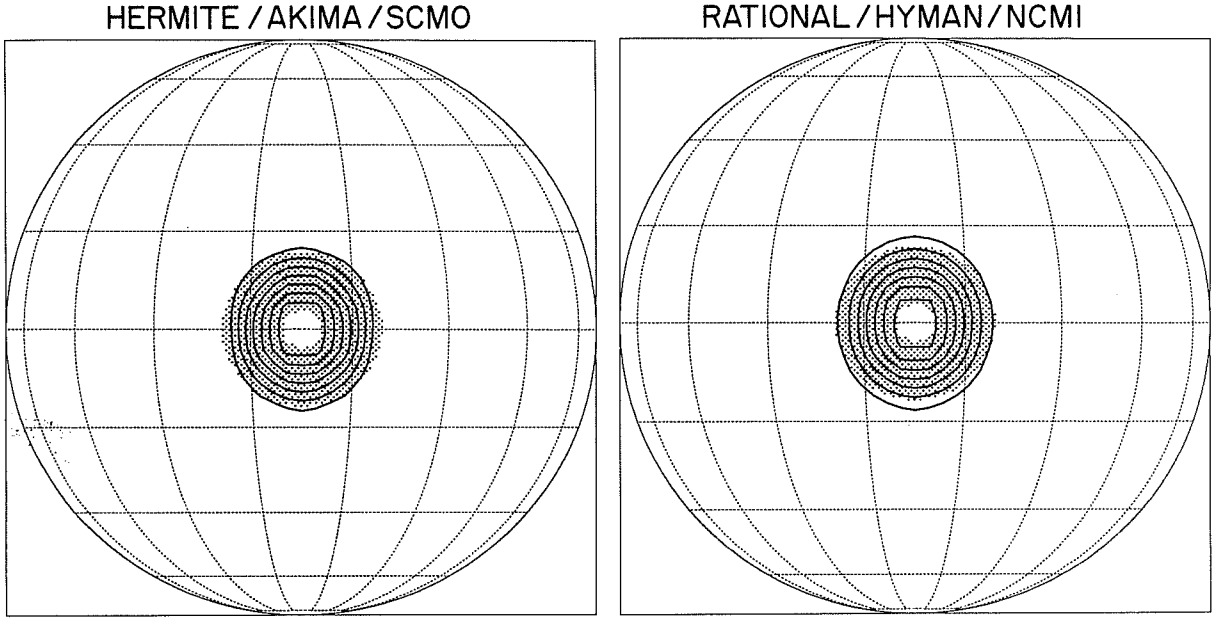

Fig. 7 Solid body rotation of cosine bell after one complete rotation (256 time steps) across both poles. Left side is for Hermite cubic with Akima derivative estimates modified to satisfy SCM0. Right side is for rational cubic with Hyman derivative estimates modified to satisfy NCM1.

6. *DEPARTURE POINT CALCULATION IN SPHERICAL GEOMETRY*

As stated earlier in Section 3, the advection equation for a scalar field like mixing ratio q is just

$$\frac{dq}{dt} = 0 \tag{6.1}$$

In spherical coordinates the semi-Lagrangian form is

$$q(\lambda_A, \phi_A, t + \Delta t) = q(\lambda_D, \phi_D, t) \tag{6.2}$$

where subscript A denotes the arrival point and D the departure point, i.e., the parcel travels from (λ_D, ϕ_D) at time t to (λ_A, ϕ_A) at time $t + \Delta t$. The arrival points (λ_A, ϕ_A) are chosen to be the Gaussian gridpoints and we must estimate the points (λ_D, ϕ_D) from which the parcels departed at time t in order to reach the Gaussian gridpoints at time $t + \Delta t$.

The relationship between departure and arrival points is

$$\lambda_D + \int_{(\lambda_D, \phi_D, t)}^{(\lambda_A, \phi_A, t+\Delta t)} u(\lambda, \phi, t)\, dt = \lambda_A \tag{6.3}$$

$$\phi_D + \int_{(\lambda_D,\phi_D,t)}^{(\lambda_A,\phi_A,t+\Delta t)} v(\lambda,\phi,t)\,dt = \phi_A \tag{6.4}$$

in which the integrals are taken along the trajectory of the parcels and must be approximated by discrete values. We anticipate making the moisture forecast after the momentum forecast, thus we can approximate (6.3) and (6.4) by

$$\lambda_D = \lambda_A - \frac{1}{2}\Delta t\,[u(\lambda_A,\phi_A,t+\Delta t) + u(\lambda_D,\phi_D,t)] \tag{6.5}$$

$$\phi_D = \phi_A - \frac{1}{2}\Delta t\,[v(\lambda_A,\phi_A,t+\Delta t) + v(\lambda_D,\phi_D,t)] \tag{6.6}$$

These are implicit equations for (λ_D,ϕ_D) since they involve u and v at the unknown point, and are solved by iteration starting with some first guess for (λ_D,ϕ_D). The velocity components at (λ_D,ϕ_D) are obtained by iterating.

$$\lambda_D^{k+1} = \lambda_A - \frac{1}{2}\Delta t\,[u(\lambda_A,\phi_A,t+\Delta t) + u(\lambda_D^k,\phi_D^k,t)] \tag{6.7}$$

$$\phi_D^{k+1} = \phi_A - \frac{1}{2}\Delta t\,[v(\lambda_A,\phi_A,t+\Delta t) + v(\lambda_D^k,\phi_D^k,t)] \tag{6.8}$$

We anticipate serious inaccuracies in the vicinity of the poles with (6.7) and (6.8). Near the poles, the velocity components undergo rapid variation over short distances and thus the approximation that the velocity along the trajectory be represented by one or two values breaks down. To avoid this problem we transform to a local geodesic coordinate at each arrival point before calculating the departure point. This local geodesic coordinate denoted (λ',ϕ') is essentially a rotated spherical coordinate system whose equator goes through the arrival point. The rotated system is chosen so that the arrival point in the transformed system (λ'_A,ϕ'_A) is (0,0). In addition, the velocity components are the same there, *i.e.*, $u'(\lambda'_A,\phi'_A) = u(\lambda_A,\phi_A)$ and $v'(\lambda'_A,\phi'_A) = v(\lambda_A,\phi_A)$. Details of the transformations involved are included in Williamson and Rasch (1987). In that paper we compare four approaches to estimating the departure point. All are iterative and require a first guess for the departure point (λ_D,ϕ_D), which we took for the tests to be the arrival point (λ_A,ϕ_A). We compared the error in the four methods for the case of solid body rotation across the pole and for a small asymmetric polar vortex whose wind maximum crosses the pole. This is a more difficult test than the solid body rotation because there is large flow (trajectory) curvature in the region of strong winds. We calculate the true departure point analytically for these cases to establish the error in the approximations. The four methods considered are

Global Spherical: The first scheme involves global spherical coordinates only and consists of iterating the following steps.

1) Interpolate for $u\left(\lambda_D^k, \phi_D^k, t\right)$ and $v\left(\lambda_D^k, \phi_D^k, t\right)$
2) Calculate next estimate of $\left(\lambda_D^{k+1}, \phi_D^{k+1},\right)$ via (6.7 and 6.8)

This is the scheme we expect to have serious errors near the poles.

Local Geodesic: The second scheme uses local geodesic coordinates for the departure point calculation and local geodesic velocity components for the interpolation. The interpolation itself, however, is done in the global spherical coordinate system in which the data grid is separable and the tensor product form applies.

1) Transform spherical components u and v to local geodesic components u' and v' at all points needed for the interpolation
2) Interpolate for $u'(\lambda_D^k, \phi_D^k, t)$ and $v'(\lambda_D^k, \phi_D^k, t)$
3) Calculate $(\lambda_D'^{k+1}, \phi_D'^{k+1})$ in the local geodesic coordinate system via (6.7) and (6.8) with all variables except t replaced with the primed (i.e., local geodesic) versions
4) Transform local geodesic $(\lambda_D'^{k+1}, \phi_D'^{k+1})$ to global spherical $(\lambda_D^{k+1}, \phi_D^{k+1})$

Mixed Spherical-Geodesic: The local geodesic scheme described above requires the transformation of all velocity components used in the interpolation. Since there is a different transformation associated with each arrival point, this requires many transformations. It may be adequate to interpolate the original spherical components and just use the local geodesic coordinate for the approximation of the integral along the trajectory only. Such a scheme is

1) Interpolate for $u(\lambda_D^k, \phi_D^k, t)$ and $v(\lambda_D^k, \phi_D^k, t)$
2) Transform spherical components $u(\lambda_A^k, \phi_A^k, t)$ and $v(\lambda_A^k, \phi_A^k, t)$ to local geodesic components $u'(\lambda_D^k, \phi_D^k, t)$ and $v'(\lambda_D^k, \phi_D^k, t)$;
3) Calculate $(\lambda_D'^{k+1}, \phi_D'^{k+1})$ as in the local geodesic scheme
4) Transform local geodesic $(\lambda_D'^{k+1}, \phi_D'^{k+1})$ to global spherical $(\lambda_D^{k+1}, \phi_D^{k+1})$

Cartesian: The last method is that proposed by Ritchie (1987). It is most like our mixed scheme, but instead of using a local geodesic coordinate, Ritchie uses a three-dimensional Cartesian system to calculate the departure point. Thus step (2) of the mixed scheme is replaced by a transformation to Cartesian components, step (3) by an approximation in Cartesian space and (4) by a transformation from Cartesian coordinates back to spherical.

We anticipate applying these methods to winds obtained from a global spectral transform model, and therefore do not consider monotonicity to be critical for the interpola-

tion of the wind field. Therefore, we have considered bilinear, biquadratic and bicubic interpolation only. Details of the errors are presented in Williamson and Rasch (1987). We summarize the results here.

As expected, the global spherical method has serious errors near the poles and should not be used. The local geodesic, mixed, and Cartesian are basically equally good. The local geodesic is better than the mixed and Cartesian only with linear interpolation coupled with advection over a short distance (small Δt). Cubic interpolation is better than quadratic with the asymmetric polar vortex but they are equally good for solid body rotation over the pole. The linear interpolation has significant maximum errors (20% of latitudinal grid interval) near the pole with the asymetric polar vortex, while the quadratic and cubic seem reasonable (9% and 6%, respectively). The expense of the extra transformations involved in the local geodesic method is not worthwhile. The mixed and Cartesian (Ritchie) methods are equally viable.

7. *COMMENTS*

Semi-Lagrangian transport coupled with shape-preserving interpolation appears to be a promising replacement for moisture transport in global spectral transform atmospheric models. We are currently incorporating the approach in the NCAR Community Climate Model and the NMC global forecast model. We will evaluate it for the most promising interpolation forms and derivative approximations both in very long simulation runs and in the forecast-analysis cycle and medium-range forecasts.

Acknowledgments

We would like to thank E. Boettner for editorial assistance and for preparing the final camera ready copy. This work was partially supported by U.S. Navy Grant N66685686WR86062.

References

Akima, H., 1970: A new method of interpolation and smooth curve fitting based on local procedures. *J. Assoc. Comput. Mach.*, **17**, 589–602.

Carlson, R. E., and F. N. Fritsch, 1985: Monotone piecewise bicubic interpolation. *SIAM J. Numer. Anal.*, **22**, 386–400.

de Boor, C. and B. Swartz, 1977: Piecewise monotone interpolation. *J. Approx. Theory*, **21**, 411–416.

Delbourgo. R., and J. A. Gregory, 1983: C^2 rational quadratic spline interpolation to monotonic data. *IMA J. Numer. Anal.*, **3**, 141-152.

Delbourgo, R., and J. A. Gregory, 1985: Shape preserving piecewise rational interpolation. *SIAM J. Sci. Stat. Comput.*, , **6**, 967–976.

Fritsch, F. N., and J. Butland, 1984: A method for constructing local monotone piecewise interpolants, *SIAM J. Sci. Stat. Comput.*, **5**, 300–304.

Fritsch, F. N., and R. E. Carlson, 1980: Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.*, **17**, 238–246.

Gregory, J. A., and R. Delbourgo, 1982: Piecewise rational quadratic interpolation to monotonic data. *IMA J. Numer. Anal.*, **2**, 123–130.

Hyman, J. M., 1983: Accurate Monotonicity preserving cubic interpolations. *SIAM J. Sci. Stat. Comput.*, **4**, 645–654.

McAllister, D. F., and J. A. Roulier, 1978: Interpolation by convex quadratic splines. *Math. Comp.*, **32**, 1154–1162.

McAllister, D. F., and J. A. Roulier, 1981: Algorithms for computing shape preserving osculatory quadratic splines. *ACM Trans. Math. Software*, **7**, 331–347.

McAllister, D. F., E. Passow and J. A. Roulier, 1977: Algorithms for computing shape preserving spline interpolations to data. *Math. Comp.*, **31**, 717–725.

Purnell, D. K., 1976: Solution of the advective equation by upstream interpolation with a cubic spline. *Mon. Wea. Rev.*, **104**, 42–48.

Rasch, P. J., and D. L. Williamson, 1987: On shape preserving interpolation and semi-Lagrangian transport. *SIAM J. Sci. Stat. Comput.*, in preparation.

Ritchie, H., 1987: Semi-Lagrangian advection on a Gaussian grid. *Mon. Wea. Rev.*, **114**, 135–146.

Spath, 1969: Exponential spline interpolation. *Computing*, **4**, 225–233.

Strang, G., 1968: On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.*, **5**, 506–517.

Williamson, D. L., and P. J. Rasch, 1987: Two-dimensional semi-Lagrangian transport with shape preserving interpolation. *Mon. Wea. Rev.*, in preparation.