Feature tracking on the unit-sphere


It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2

Publisher: American Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR
Central Archive at the University of Reading
Reading’s research outputs online
Feature Tracking on the Unit Sphere

K. I. Hodges

Center for Global Atmospheric Modeling (CGAM) and NERC Unit for Thematic Information Systems (NUITS), University of Reading, Reading, England

(Manuscript received 8 December 1994, in final form 26 June 1995)

ABSTRACT

Techniques used in a previous study of the objective identification and tracking of meteorological features in model data are extended to the unit sphere. An alternative feature detection scheme is described based on cubic interpolation for the sphere and local maximization. The extension of the tracking technique, used in the previous study, to the unit sphere is described. An example of the application of these techniques to a global relative vorticity field from a model integration are presented and discussed.

1. Introduction

The objective identification and tracking of moving objects through meteorological data provides a useful diagnostic technique for the analysis of model data derived from general circulation models (GCM). This paper reports an extension to an existing technique to the spherical domain for the identification and tracking of vorticity centers. A recent paper by Hodges (1994) extended the use of technique developed for the tracking of moving objects in video footage by Sethi and Jain (1987) and Salari and Sethi (1990) to the problem of objectively tracking synoptic-scale features in meteorological data. The Hodges paper described techniques for the objective identification of suitable features and the implementation of these, along with the tracking algorithm, in a program called TRACK. An example was given of the application of TRACK to a relative vorticity field from a GCM to highlight the effectiveness of the approach. A further application of the program has since been applied in a study comparing the effects of using different convective parameterizations in the Universities Global Atmospheric Modeling Project (UGAMP) GCM (Slingo et al. 1994).

Previous attempts at tracking meteorological features whether derived from model data or satellites has been performed on some Cartesian projection for the area of interest, for example, the previously mentioned paper of Hodges and the alternative approaches of Murray and Simmonds (1991a), Sinclair (1994), and König et al. (1993). However, with the advent of global datasets, predominantly from models, it makes sense to try to perform global objective analyses of the synoptic activity. A global approach makes the tracking parameters independent of projections. It also negates the introduction of bias into the tracking that may occur on a projection, and which is important for the statistical analysis of the tracks (not reported here).

In this paper an alternative feature detection scheme is described based on interpolation and local maximization. Also described is a formulation of the tracking algorithm for the unit sphere. An example is given of the application of these techniques to a relative vorticity field from the UGAMP GCM.

2. Tracking on the unit sphere

Putting aside the choice of suitable feature points to track until section 3, the tracking algorithm will first be described in general terms, and it will then be shown how it can be applied to the 2D domains of the plane and the sphere (although in principle it can be applied in any domain provided suitable measures for distance and direction can be defined).

The basic tracking technique relies on the minimization of a cost function to solve the correspondence problem (Sethi and Jain 1987) and is defined in terms of "local smoothness" functions. These are constructed from the local deviations in direction and speed computed over three consecutive frames (time steps). Penalties are used to restrict the changes in speed and direction. Phantom feature points, introduced by Salari and Sethi (1990), are used as a computational convenience to cope with the appearance and disappearance of features.

Thus, the cost function \( \Xi \) is defined as

\[
\Xi = \sum_{i=1}^{m} \sum_{k=1}^{n-1} \phi_k^i (P_i^{k+1}, P_i^k, P_i^{k+2}) ,
\]

(1)

© 1995 American Meteorological Society
The distance measure \( \| \mathbf{P}_i^k \mathbf{P}_i^{k+1} \| \) is defined as
\[
\| \mathbf{P}_i^k \mathbf{P}_i^{k+1} \| = \begin{cases} 
\| \mathbf{P}_i^k \mathbf{P}_i^{k+1} \|_e, & \text{if } \mathbf{P}_i^k \text{ and } \mathbf{P}_i^{k+1} \text{ both} \\
\text{true feature points;} & \\
\text{min}, & \text{otherwise,}
\end{cases}
\]
where the subscript \( p = e \) for the plane indicates the Euclidean norm is used, and for the sphere \( p = g \) indicates the geodesic norm is used, and \( \text{min} \) specifies an upper bound on the displacement.

For applications in 2D Cartesian space—for example, a cylindrical projection (Hodges 1994)—\( \mathbf{P}_i^k \mathbf{P}_i^{k+1} \) is the displacement vector from \( \mathbf{P}_i^k \) to \( \mathbf{P}_i^{k+1} \). The distance measure, defined in terms of the Euclidean distance is
\[
\| \mathbf{P}_i^k \mathbf{P}_i^{k+1} \| = | \mathbf{P}_i^{k+1} - \mathbf{P}_i^k |.
\]
The direction \( \hat{\mathbf{T}}_i^{(k+1)} \) is defined as the normalized displacement vector; hence
\[
\hat{\mathbf{T}}_i^{(k+1)} = \frac{\mathbf{P}_i^k \mathbf{P}_i^{k+1}}{\| \mathbf{P}_i^k \mathbf{P}_i^{k+1} \|}; \quad \| \mathbf{P}_i^k \mathbf{P}_i^{k+1} \| > 0.
\]
For application to the unit sphere the spherical data (feature points defined by their latitude—longitude position) need to be converted to 3D Cartesian space. This is the most efficient form of the data for the computation of the vector products. The point \( \mathbf{P}_i^k \) is now a unit radius vector and \( \mathbf{P}_i^k \mathbf{P}_i^{k+1} \) is the directed arc of a great circle (geodesic) from \( \mathbf{P}_i^k \) to \( \mathbf{P}_i^{k+1} \). The distance measure is now defined in terms of the geodesic norm \( \| \|_g \), where
\[
\| \mathbf{P}_i^k \mathbf{P}_i^{k+1} \|_g = \arccos(\mathbf{P}_i^k \cdot \mathbf{P}_i^{k+1}).
\]
To determine the deviation in direction, two unit vectors in the same tangent plane are required. The natural choice is the two unit tangent vectors at the point \( \mathbf{P}_i^k \) for the two directed arcs, \( \mathbf{P}_i^{k-1} \mathbf{P}_i^k \) and \( \mathbf{P}_i^k \mathbf{P}_i^{k+1} \). However, the direct computation of these tangents from the equation of the geodesic for any orientation of the vector pairs \( (\mathbf{P}_i^k, \mathbf{P}_i^{k+1}) \) is not a very efficient method. A more convenient method of finding the two unit tangents is to take the triple vector cross product for each vector pair. The geometry of the problem is illustrated in Fig. 1. Define \( \mathbf{t}_i^{(k,k+1)} \) to be
\[
\mathbf{t}_i^{(k,k+1)} = (\mathbf{P}_i^k \times \mathbf{P}_i^{k+1}) \times \mathbf{P}_i^k = \mathbf{P}_i^{k+1} - (\mathbf{P}_i^k \cdot \mathbf{P}_i^{k+1}) \mathbf{P}_i^k;
\]
this vector expression is easily evaluated and normalized, thus
\[
\hat{\mathbf{T}}_i^{(k,k+1)} = \frac{\mathbf{t}_i^{(k,k+1)}}{\| \mathbf{t}_i^{(k,k+1)} \|} = \frac{\mathbf{t}_i^{(k,k+1)}}{\left[1 - (\mathbf{P}_i^k \cdot \mathbf{P}_i^{k+1})^2\right]^{1/2}}.
\]
with \( P_{i}^{k+1} \cdot P_{i}^{k+1} < 1 \). Then, the values to use in expression (3) for \( \hat{T}_{(k-1)k} \) and \( \hat{T}_{(j,k+1)} \) are given by \( \hat{T}_{(k-1)k} = -t_{(k-1)k} \) and \( \hat{T}_{(j,k+1)} = t_{(j,k+1)} \).

Note, that for both application to the plane and the sphere and also in general, the smoothness function is not defined under certain conditions associated with stationary features. In practice for the data so far explored this does not appear to be a problem, at least at the time intervals of the data currently used and using the off-grid feature points (feature points that are not necessarily coincident with the grid) discussed below, truly stationary features do not occur. The situation of stationary features arises if two or all of the triangle of points \( P_{i}^{k-1}, P_{i}^{k}, P_{i}^{k+1} \) are coincident. If a pair of points are coincident, then the first term of expression (3) is set to zero and \( \psi = w_{5} \) (unless \( P_{i}^{k-1} \) is a phantom point), while if all three points are coincident, then \( \psi \) is set to zero. For the former case of two coincident points, an exchange with a point from another track (during minimization of the cost function) is still possible if it reduces the local value of the cost function, while for the later case of three coincident points no exchange is possible, which will lead to a reduction in the cost function.

The spherical representation of the cost function, as defined above, is minimized in the same way as described by Salari and Sethi (1990) and applied by Hodges (1994). Four parameters must be supplied, the two weights, \( w_{1} \) and \( w_{2} \), and the upper bound values for the geodesic displacement \( d_{\text{max}} \) and the local smoothness \( \psi_{\text{max}} \). The choice of these parameters will depend to a certain degree on the choice of feature points and the time step.

3. Determination of feature points on a sphere

In the previous paper (Hodges 1994), the techniques used assume the data is defined on a rectangular grid. The initial processing of each time step locates the gridpoints maxima or minima for each identified “object.” These may be perfectly adequate for tracking on a high-resolution grid, but on a low-resolution grid—for example, a Gaussian T42 grid—these points may not be sufficient to produce smooth tracks (may result in a staircase effect) and are effectively a limitation on the smoothness that the tracking algorithm can work to. Therefore, these gridpoint extrema are used as initial points in other methods to find alternative feature points. The original approach (Hodges 1994), used an ad hoc clustering technique to determine centers of “mass.” Although this worked reasonably well, resulting in smooth tracks and is computationally quick, it requires a conformal cylindrical projection to work well on the whole globe. It also requires more emphasis to be given to direction rather than speed in the tracking cost function, due to the uncertainty in determining centers of mass for complex objects.

To improve this situation a different approach has been implemented based on interpolation/smoothing and local maximization to compute the off-grid local maxima or minima. This allows the smoothness to be weighted more toward speed in the cost function with the resultant tracks being much smoother overall.

In GCM terms if our data is derived from a spectral model we could use the spectral representation of the field to provide spectrally smoothed field values anywhere on the sphere for the local maximization. However, this is computationally expensive because of the need to recompute the associated Legendre functions for each new point on the sphere (effectively a regridding exercise). So, the data are interpolated or smoothed using the B-spline techniques of Dierckx (1981, 1984). These techniques rely on the data being defined on the latitude–longitude grid. They are capable of producing either a least squares surface with automatic knot positioning or the interpolating surface where the knots are the grid points. The first method (Dierckx 1981) produces the classic surface fit with no constraints imposed by the sphere, while the second method (Dierckx 1984) produces a surface with continuity constraints so that the surface is periodic in longitude and \( \psi \) continuous at the poles. For the data from the UGAMP GCM used here, which has already been spectrally smoothed, the interpolating surface is the most computationally efficient and appropriate to produce. There are two choices that can be made for the interpolation, either the whole field or region of interest can be interpolated, or the interpolation can be performed over the extent of objects (a rectangular region that entirely contains the object). Thus, by considering just the object data only small regions need to be interpolated. For the data used here defined on a coarse T42 Gaussian grid the whole field is interpolated using the second Dierckx algorithm for the sphere. Only the function values and first derivatives are required for the local maximization algorithm and these are evaluated efficiently using the recurrence schemes of de Boor (1972).

The local maximization is achieved using the technique of Goldfarb (1969). This is a constrained conjugate gradient method that uses the gridpoint maxima/minima as starting points. The algorithm returns the position and field value of the extremum point. This then forms a feature point. These are stored in the relevant object record (Hodges 1994).

4. Application to the UGAMP AMIP integration

To illustrate the effectiveness of the tracking on a global domain, the techniques expounded by Hodges (1994) and extended to a global domain in this paper have been applied to data derived from the Atmospheric Model Intercomparison Project (AMIP) integration of the UGAMP GCM. This is a spectral model and has been documented by Slingo et al. (1994). The model has been run, along with the other participating models in the comparison project, for the decadal pe-
period of 1979–88 with climatological sea surface temperatures (SSTs). The data used for this illustration is the field of the relative vorticity at 850 mb for the 1979–80 DJF (December, January, February) period at 6-h time steps on a T42 Gaussian grid (longitude–latitude grid of size 128 × 64 grid points). This field was chosen because synoptic features at all latitudes show up well. Other fields that can be explored for synoptic activity are the pressure at mean sea level (PMSL) (Murray and Simmonds 1991a), the geopotential height (König et al. 1994, 1000-mb height), and the Laplacian of the 1000-mb geopotential height (Sinclair 1994). However, tropical activity—for example, easterly waves—is not easily discernible in these fields. As mentioned in Hodges (1994), the Coriolis factor in the relative vorticity field means that for a global study the method of tracking must be applied to both the positive and negative part of the field. If this is not done or we invert one hemispherical part of the field to deal with only the cyclonic activity say, then the resulting track distribution tends to be discontinuous across the equator.

To obtain the results shown below, the new spherical techniques discussed above have been used with the following parameters. The value of the threshold (Hodges 1994) used to identify “object” field points was set at ±1.0 × 10^{-5} s^{-1}. The upper bound on the angular displacement for a 6-h period was set at 6.5°. The values of the weights chosen for the cost function was 0.2 for direction and 0.8 for speed. The upper bound on the track smoothness was chosen as ψ_{max} = 0.15 (following minimization of the cost function all values of ψ ≤ ψ_{max}).

The results of the tracking are shown in Figs. 2 and 3 for the two signs of the field on a cylindrical projection to highlight the Tropics and any interaction between the Tropics and extratropics. The cyclonic activity is also shown on two polar stereographic projections for comparison with previously published analyses of cyclone tracks, which are predominantly for the extratropics, (e.g., Murray and Simmonds 1991b; Sinclair 1994; König et al. 1993). These are displayed in Figs. 4 and 5. These figures show the tracks (blue lines) overlaid with the strength of the systems at each 6-h position. A brief discussion of these results follows.

5. Discussion
a. Extratropics

In the extratropics the distribution of cyclonic activity for the DJF period displayed in the figures shows considerable difference between the Northern and Southern Hemispheres due to the distribution of continental landmasses. In the Southern Hemisphere the cyclonic tracks are in a band from 40° to 65°S, while in the Northern Hemisphere the main cyclonic tracks appear as two major storm tracks in the Pacific and Atlantic, which are tilted along a northeast axis. This behavior agrees well with that of previous studies for the Southern Hemisphere and with the studies of the Northern Hemisphere (e.g., Klein 1957).

The cyclonic activity can be associated with low pressure centers, which accounts for the close agreement between studies using different physical fields—for example, PMSL, relative vorticity, and geopotential height. However, for the anticyclonic activity, this as-
association (between the anticyclonic activity in a relative vorticity field and high pressure centers) is not so clear. In the results of Jones and Simmonds (1994), and Taljaard (1967) for the Southern Hemisphere for a DJF period, the major anticyclonic activity appears in a broad band between 25° and 42°S. The results shown in Figs. 2 and 3 do show a similar behavior; however, the anticyclonic activity has a more banded appearance than that of Jones and Simmonds (1994), for example, and is generally situated farther south, between 40° and 60°S. Sinclair (1994) noted that additional mobile centers for the cyclonic activity, as compared with PMSL, occur in the band 45° and 55°S, and Fig. 2 shows this to be true for the anticyclonic activity as well. In the subtropics there appear several clusters of weak anticyclonic activity. This is believed to be due to residual high-frequency oscillations derived from the spectral representation of the field due to the discontinuities introduced by the orography.

The cyclonic activity in the Southern Hemisphere is also indicated in Fig. 5. The coastal regions of Antarctica are seen to be strong regions of cyclogenesis and cyclolysis along with Australasia and southern South America. The Antarctic coast as a region of cyclolysis has also been observed in the studies of Taljaard (1967) and Streten and Troup (1973), who used observations of cloud vortices. The observation that the Antarctic coast is also a region for Cyclogenesis has been observed by Le Treut and Kalnay (1990). Systems originating over Australasia, southern South America, and a few from the South Africa region move southeastward, increasing in speed, to join the storm track. The region of the South Pacific convergence zone (SPCZ) also has several features moving south eastward to join the storm track. Overall, these results are in accordance with the observational studies mentioned above and with the results of Murray and Simmonds (1991b) and Sinclair (1994).

For the Northern Hemisphere the cyclonic and anticyclonic activity appear to have similar distributions. The study of Wallace et al. (1988) has a bearing on this behavior. Their study applied high-pass filtering to NMC height analyses for 20 Northern Hemisphere winters. They found that the difference between the distribution of negative and positive height anomalies showed little difference, but that the distribution of cyclones and anticyclones identified in the 1000-mb height field may be quite different. This difference between the anomalies and the actual cyclones and anticyclones was attributed to the background climatology, which biases tracks of mobile features toward semipermanent features over the oceanic regions, for example, the Azores high and the Icelandic low. These same semipermanent features make identifying and tracking synoptic features, in PMSL model data, for example, difficult because they tend to swamp the mobile features. However, the results presented here show that relative vorticity is a better field to use since it is less influenced by the background larger-scale flow.

For the cyclonic activity in the Northern Hemisphere as displayed in Fig. 4, we notice that the eastern continental margins are regions of strong cyclogenesis for systems that eventually form the northern Atlantic and Pacific storm tracks. These systems strengthen and increase in speed as they move northeastward into the oceans. More specifically for the North Pacific, there
appears several regions of cyclogenesis, in the lee of the Tibetan Plateau, and in northeast China and Mongolia (Gobi Desert). Many of those emanating from the first two of these regions propagate eastward and strengthen rapidly as the cold, dry air interacts with the warm water over the East China Sea and the sea of Japan, while the third region generates many weak systems, many of which remain in the region, although several do propagate into the storm track. This type of behavior accords well with observation (Boyle and Chen 1992), with cold surge events being considered the trigger for cyclogenesis rather than lee wave generation. For the Atlantic, the predominate region of cyclogenesis is the North American continent, in particular, in the lee of the Rocky mountains. This again accords well with observation (Zishka and Smith 1980).

The anticyclonic activity for the Northern Hemisphere can be seen in Fig. 3. The strengths are considerably weaker than their cyclonic counterparts. Regions of strong anticyclogenesis are southern Japan, China, and the southern United States. This seems to fit in well with the behavior reported in East Asia by Boyle and Chen (1992). It also seems to agree with the behavior observed by Zishka and Smith (1980) for North America winter anticyclones.

b. Tropics

In the intertropical convergence zone (ITCZ) region of the Pacific, the cyclonic tracks are tightly confined just north of the equator between 0° and 10°, this is a typical situation with the Betts–Miller convective parameterization (Betts and Miller 1993) as used in the model. There are two groups of tracks in this region, the first originating from an area to the west of Central America and propagating westward. The majority die out around 120°W before reaching the second group. Zehnder and Gall (1991) propose that sea surface temperature anomalies are insufficient to explain this behavior in cyclogenesis and that some other initial cyclonic disturbance is required in the form of lee wave generation from the orography, which runs the length of Central America from the Sierra Madre to the Colombian highlands. The second group of tracks begin east of the date line and strengthen as they propagate.
into the warm pool region. Observe in Fig. 2 that several cyclonic systems are deflected northward to join the North Pacific storm track; this is more prevalent in other DJF periods of the AMIP data.

Other regions of cyclogenesis in the Tropics are over Southeast Asia stretching into the Indian Ocean, north of New Guinea, and the Gulf of Guinea. In the Indian Ocean this cyclogenesis appears to emanate from the northern end of Sumatra and may be the result of lee wave generation from this mountainous region. The origin of the few cyclonic systems that originate over the Nigerian region of Africa is not so clear, their number is considerably less than that of the African easterly waves that appear for a JJA (June, July, August) period.

Acknowledgments. The author would like to thank one of the reviewers for their helpful comments on the initial manuscript.

REFERENCES

HODGES


