Spherical nonparametric estimators applied to the UGAMP model integration for AMIP

Article

Published Version

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(1996)124<2914:SNEATT>2.0.CO;2

www.reading.ac.uk/centaur

Spherical Nonparametric Estimators Applied to the UGAMP Model Integration for AMIP

K. I. HODGES

*Centre for Global Atmospheric Modeling (CGAM) and Environmental Systems Science Centre (ESSC),
University of Reading, Reading, England*

ABSTRACT

The aim of this paper is essentially twofold: first, to describe the use of spherical nonparametric estimators for determining statistical diagnostic fields from ensembles of feature tracks on a global domain, and second, to report the application of these techniques to data derived from a modern general circulation model. New spherical kernel functions are introduced that are more efficiently computed than the traditional exponential kernels. The data-driven techniques of cross-validation to determine the amount of smoothing objectively, and adaptive smoothing to vary the smoothing locally, are also considered. Also introduced are techniques for combining seasonal statistical distributions to produce longer-term statistical distributions. Although all calculations are performed globally, only the results for the Northern Hemisphere winter (December, January, February) and Southern Hemisphere winter (June, July, August) cyclonic activity are presented, discussed, and compared with previous studies. Overall, results for the two hemispheric winters are in good agreement with previous studies, both for model-based studies and observational studies.

1. Introduction

It is important from the point of view of improving the physical parameterizations used in general circulation models (GCMs), and increasing our confidence in the predictions of future climate obtained using GCMs, that these models be validated against the current climate. Current climate models are capable of representing the mean state of the atmosphere fairly well and are also beginning to reproduce the variability observed in the atmosphere. It is therefore important to explore this variability in more depth and to compare with observation.

One approach to this problem is to identify and track interesting features and then to analyze the ensemble of tracks statistically. Early attempts at this type of analysis were performed by identifying features on synoptic charts and tracking the features manually. However, several automatic objective approaches have recently been developed, including those of Murray and Simmonds (1991a) and König et al. (1993). For this study, the techniques of Hodges (1994, 1995) are used, which allow features to be identified and tracked on a spherical domain and hence allow global studies to be performed. The tracking of features and their display on maps is a useful visualization tool both for real data

and for model data; however, it is only when these tracks are analyzed in a quantitative way that they become useful for climatological studies and in particular, for the validation of climate models. The approach is to determine (estimate) a statistic on some grid, then the statistic can be contoured and further analyzed and compared with other climatologies.

The use of statistics to summarize and analyze climate data has been in use for some time. Several hemispheric studies, in particular of the Northern Hemisphere (NH), have been conducted since the early part of this century (e.g., Klein 1957). These early studies all relied on the manual tracking of features. More recently, statistical methods have been applied to tracks derived from model data. For example, Akyildiz (1984) computed cyclone tracks using data from the European Centre for Medium-Range Weather Forecasts (ECMWF) forecast data for the 1981/82 winter for the North Atlantic and constructed statistics of cyclone speed, central pressure, geographical size, and life span. Murray and Simmonds (1991b) applied statistical techniques to cyclone tracks determined from mean sea level pressure (MSLP) data from the Melbourne University GCM and produced statistics of cyclone frequency, cyclogenesis, cyclolysis, and track flux (mean number of tracks within unit distance of an estimation point per unit time) for the SH. The techniques of Murray and Simmonds have been used in several other studies of cyclone–anticyclone climatologies, using MSLP (Murray and Simmonds 1995; Jones and Simmonds 1994), and also in the study of

Corresponding author address: Dr. K. I. Hodges, Environmental Systems Science Center, University of Reading, Whiteknights, P.O. Box 227, Reading RG6 6AB, United Kingdom.
E-mail: kih@mail.nerc-essc.ac.uk

Sinclair (1994), who determined a Southern Hemisphere (SH) cyclone climatology using the geostrophic relative vorticity, computed as the Laplacian of the 1000-mb geopotential. König et al. (1993) have also used the 1000-mb geopotential for their global study, based on tracks of relative minima in the field. There have also been several observational studies, particularly in the high SH latitudes, for example, Streten and Troup (1973), who used satellite imagery, and Taljaard (1967), who used surface maps and 500-mb maps from the International Geophysical Year (IGY).

The Universities Global Atmospheric Modelling Program (UGAMP) model used for this study has been documented by Slingo et al. (1994). It is a 19-level spectral model based on the ECMWF model, cycle 27. New physical parameterizations have been implemented within the model since its initial release, by both ECMWF and UGAMP. These include the radiation scheme of Morcrette (1990), an improved treatment of surface fluxes, the flux-limited vertical advection scheme of Thuburn (1993), and the convective parameterization of Betts and Miller (1993) [as an alternative to the Kuo scheme (Kuo 1974) originally used; either can be used, but the Betts and Miller scheme has been used for AMIP]. The model has been run at resolution T42 for the Atmospheric Model Intercomparison Project (AMIP) with monthly mean observed sea surface temperatures (SST) and sea-ice distributions for the decadal period of 1979–88.

In this paper the 850-mb relative vorticity field is used to determine the distribution of the cyclonic activity in the 10-year integration of the model. However, the term cyclone as normally understood (a closed low pressure center) is not necessarily synomynous with the vorticity centers so that in this paper the term cyclonic activity is preferred (cyclonic vorticity center is a center of cyclonic activity). A further reason for using vorticity is that the Tropics can also be explored, so that a truly global study can be performed on GCM data, although results for the Tropics are not presented here. Sinclair (1994) has outlined several reasons for preferring the vorticity to the MSLP or 1000-mb geopotential field. These are that the center of rotation is better related to vorticity than the pressure minima or maxima and that cyclones do not appear as a pressure minima until after considerable development has occurred. The second point results in many more features to track appearing earlier in their life cycle when vorticity is used.

The statistics produced in previous studies have been predominately computed on some projection, with the distribution of features presented as a frequency. Frequency distributions are often area normalized with some correction for the distortion introduced by the projection. Another aspect is the use by several authors of some form of smoothing of their distributions using subjective ad hoc techniques. In this paper it is shown how the statistics can be computed directly on the sphere, so avoiding the use of projections except for presentation of results. It will also be shown how an objective choice of the smoothing parameters can be determined along with the use of adaptive smoothing of the distributions. Adaptive smoothing is important as it helps resolve some of the problems with the objective determination of the smoothing parameters and also makes the estimators more responsive to the data distribution, so that detail can be retained in high-data-density regions and spurious detail in low-data-density regions can be suppressed. Here, instead of area-normalized frequencies, probability density functions are computed for the distribution of features. The reasons for this will be made clear in the following discussion on statistical estimation. However, the main reasons can be summarized as simplicity and efficiency of computation with minimal systematic bias.

2. Statistical estimation for spherical data

a. General background

In many of the early climatological studies of NH and SH cyclonic activity, statistical estimates were performed by counting the number of occurrences of cyclones within latitude–longitude boxes—for example, cyclone frequency (ratio of the number of events in a box to the number of time steps analyzed) and cyclone density (number of centers per unit area)—or by counting features that crossed latitude–longitude boxes—for example, track frequency/density and track flux. In fact, confusion reigns in these terms as some authors consider cyclone frequency (and its area normalized counterpart) as the number of occurrences (a cyclone is counted n times if it appears n times in the box) within a box normalized by time (and area) (Konig et al. 1993; Lambert 1988; Le Treut and Kalnay 1990), while others consider it to be the number of cyclones crossing a box normalized for time (and/or area) (Hayden 1981; Zishka and Smith 1980; Bell and Bosart 1989). For other diagnostics of cyclone activity, namely cyclogenesis and cyclolysis, it is perhaps a little clearer as these are discrete events in time and are generally computed as frequencies or densities by counting the total number of events within a grid box. In this study we try to avoid these problems of interpretation by considering probability density distributions, for feature density, genesis density, and lysis density. The probability that an event occurs within a small area dA centered on a point X in a domain of interest, in the time period considered, is $P(X)dA$ [where $P(X)$ is the probability density function at X]. Also considered is the track density, which is not strictly a probability density due to the way it is computed, but we will leave discussion of this until later.

Virtually all previous studies have been performed on some projection, which until recently relied on rect-

angular sampling boxes distributed uniformally on a projection. However, this leads to gross shape distortion of the sampling areas on the sphere (unless the sampling areas are relatively small and at the center of the projection) with the resultant introduction of a latitudinal bias for raw frequencies. To correct for this bias, latitude-dependent area normalization is often introduced. Unfortunately this may also lead to bias. Hayden (1981) has compared the use of area normalization and raw frequencies for latitude–longitude cells and found that area normalization leads to systematic bias in the analyses of cyclone frequencies. For the case where cyclone frequency is determined by the number of cyclones passing through a grid box, Taylor (1986) has found that the orientation of the tracks is also important. He found biases due to poleward decrease in grid-box size and due to track orientation of up to 14% (higher values may be possible). To avoid the introduction of bias due to area normalization, several authors have proposed using equal-area grids (Ballenzweig 1959). However, they have not been used to any great degree, and in any case if used on a projection will still introduce some systematic bias, due to the change of shape of the sampling area on the sphere with latitude to maintain the equality of sampling area (conformality and area cannot be jointly preserved on a projection).

The way to eliminate the bias due to area normalization using rectangular grid cells has been known for some time, this is to use equal-area circular cells (Kelsey 1925), in other words, to count within some radial distance of a point rather than within a rectangular area. Changnon et al. (1995) have explored the use of nonoverlapping equal-area circles. The nonoverlapping prevents the double counting of cyclones passing through the circles, at the expense that cyclones may pass between circles and thus not be included in the estimate. Consequently, some corrections still need to be applied to allow comparisons between latitudes. However, if the counting is performed using a projection, systematic bias may still be introduced (a circle on a projection is not necessarily a circle on the sphere). Another option that would reduce some of the problems endemic in using projections for the calculation of cyclone statistics is to use circular cells on the sphere, that is, to count within some arc radius of a point on the sphere. However, this does not appear to have gained much favor in the literature, perhaps because many studies are regional in character rather than global.

The problems discussed above, together with the apparent fact that different authors tend to subjectively choose different grid cell/circle dimensions and use different projections, makes quantitative comparisons difficult. The level of bias introduced, as discussed above, is probably insufficient to change the general

conclusions of these studies, but it is as well to be aware of their shortcomings.

Effectively all of the approaches mentioned above are nonparametric histogram estimates of the frequency or density, or are regressogram estimates of means. The term nonparametric is used in statistics to imply that no assumptions have been made concerning the distribution of the data; thus, histogram techniques are the simplest type of nonparametric estimator, more sophisticated estimators will be discussed shortly. Regression, as used in this context, is not the act of finding a parametric curve or surface, but the process of estimating pointwise values of some underlying surface associated with the spatial data points. The smoothness of histogram estimates depends on the bin width (radius); hence, authors who use different bin widths are applying differing amounts of smoothing. Another, problem with histogram techniques is that the distribution can also be dependent on the bin positions (Silverman 1986). Even so a histogram-type technique leads to an estimate with jumps and zero derivative for a function that typically will be continuous and differentiable.

A statistic can of course be smoothed as it is accumulated from the data by the use of weighted sums, where the weights depend on the separation distance of the data from a grid of estimation points (points at which a statistical estimate is computed). This will lead to a much smoother statistic, but the question of how much smoothness to use arises.

Other techniques of smoothing data on a spherical domain have been used for meteorological data. In particular, physical quantities derived from spectral GCMs are expressed as truncated series of spherical harmonics. Truncation of the series or tapering the series with a sequence of weights provides the smoothing. This approach could also be used as a statistical estimator; however, for large datasets in which the data are randomly distributed in the spherical domain this may be computationally expensive, due to the number of times the associated Legendre polynomials need to be computed.

Techniques of more general applicability for estimating spatial statistics nonparametrically have been in existence for the past decade. In particular, nonparametric kernel estimators for density estimation (Silverman 1986; Scott 1992) and regression estimation (Härdle 1990) have been studied. These are essentially estimators where each data point is assigned a weight, where the weights are usually probability density functions (this is not necessarily the case for density estimation), often termed kernel functions, with the maximum centered on each data point. The kernel functions are functions of the scaled distance from the data point for isotropic kernels and also radial direction for nonisotropic kernels. Thus, the influence of a data point can be adjusted by changing the scaling. In fact, the statistical estimators used by Murray and Simmonds

(1991b) are of this form, although they do not normalize their weighting function (it is not a probability density function) and it is specifically designed for the projection that they have used for their calculations. They have used a rather complicated isotropic weight function dependant on two parameters chosen subjectively to produce a not too noisy plot (low variance) for their statistical estimates on a polar stereographic projection for the SH. The statistics still need to be corrected for the areal distortion, due to using a projection for their calculations.

b. Kernel estimation

Kernel estimators were developed to obviate some of the problems that arise with histogram type methods, namely the discontinuous nature of the estimate and the dependence on the bin positions. The parameters of the kernel function will control its shape and the amount of smoothing applied. These techniques have been mostly applied to data in a Cartesian domain (using Euclidean distance measures). The result of applying a kernel estimator is an estimate at the chosen estimation points, which represents a continuous function (for any nonconstant kernel; the estimate has all the continuity and differentiability properties of the kernel).

For kernel estimators the kernel function is usually defined as a probability density function for the spatial domain D and thus will satisfy the following conditions for any kernel $K(\mathbf{X}, \mathbf{X}_i, C_n, \cdots)$:

$$\int_D K(\mathbf{X}, \mathbf{X}_i, C_n, \cdots) dA = 1, \tag{1}$$

$$K(\mathbf{X}, \mathbf{X}_i, C_n, \cdots) \geq 0; \quad \forall \mathbf{X}, \mathbf{X}_i \in D, \tag{2}$$

where $\mathbf{X}$ is the position vector of an estimation point, $\mathbf{X}_i$ is the position vector of a data point, and C_n is a scaling parameter. The scaling parameter is often called the smoothing parameter, and may be dependant on the size of the dataset n and also the distribution of the data so that the smoothing can be locally varied (see later). The definitions of the various estimators used in this study now follow.

The general definition of the kernel density estimator at a point $\mathbf{X}$ for an isotropic kernel (single smoothing parameter) is given by

$$\hat{f}(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} K(\mathbf{X}, \mathbf{X}_i, C_n). \tag{3}$$

Throughout this paper a circumflex (^) is used to denote a statistical estimate of some quantity. Notice that if the kernel function is a probability density function, then the kernel density estimate will be also.

Expression (3) represents the Rosenblatt–Parzen kernel density estimator. The regression estimator may be defined for both scalar and vector field data; for vector field data $\mathbf{Y}_i$, it is defined as

$$\hat{\mathbf{Y}}(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_i \frac{K(\mathbf{X}, \mathbf{X}_i, C_n)}{\hat{f}(\mathbf{X})}$$

$$= \frac{\sum_{i=1}^{n} \mathbf{Y}_i K(\mathbf{X}, \mathbf{X}_i, C_n)}{\sum_{i=1}^{n} K(\mathbf{X}, \mathbf{X}_i, C_n)}. \tag{4}$$

This is often simplified to

$$\hat{\mathbf{Y}}(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} W_i(\mathbf{X}) \mathbf{Y}_i, \tag{5}$$

where

$$W_i(\mathbf{X}) = \frac{K(\mathbf{X}, \mathbf{X}_i, C_n)}{\hat{f}(\mathbf{X})} \tag{6}$$

are the Nadaraya–Watson weights with the property that $n^{-1} \sum_{i=1}^{n} W_i(\mathbf{X}) = 1$ [thus $\hat{\mathbf{Y}}(\mathbf{X})$ is an unbiased estimator].

Similarly the variance of the field data is estimated as

$$\hat{\sigma}^2(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} [\mathbf{Y}_i - \hat{\mathbf{Y}}(\mathbf{X})]^2 W_i(\mathbf{X}), \tag{7}$$

where $\hat{\sigma}^2(\mathbf{X}) \equiv [\hat{\sigma}_1^2(\mathbf{X}), \hat{\sigma}_2^2(\mathbf{X}), \cdots]^T$ for vectorial field data.

For data distributed in a Cartesian domain with no surface restrictions, an isotropic kernel function is typically a function of the scaled Euclidean distance from the estimation point to the data point; that is,

$$K(\mathbf{X}, \mathbf{X}_i, C_n) = \frac{1}{(C_n)^q} F\left(\frac{|\mathbf{X} - \mathbf{X}_i|}{C_n}\right) \tag{8}$$

for some function $F(\ )$, where q is the dimensionality. Here, C_n, the smoothing parameter, is also called the bandwidth or window width. Thus, C_n determines the spread of the kernel function and range of influence of the data point. It is called a smoothing parameter because as C_n tends to zero the estimates reproduce the data, while as C_n becomes large the estimates smooth out the detail. However, in general, this form of the kernel function is not suitable for estimation on a spherical domain since the separation of points on the sphere are best represented by an angle (or function of an angle). In the next section it will be shown how the idea of kernel estimation can be applied to the sphere.

c. Spherical kernel estimation

For spherical data the unit sphere is taken as the domain, so any attribute/field value is assumed to be already scaled to the requisite units. On the unit sphere a point is specified by the unit vector $\mathbf{X}$ originating at the center of the sphere. The distance between two points on the sphere can now be measured as the angular distance $\psi_i = \arccos(\mathbf{X} \cdot \mathbf{X}_i)$. However, comput-

ing the inverse cosine every time a kernel needs evaluation is inconvenient; hence, the functional dependence of the kernel is usually in terms of the cosine of the angle (the dot product). The result of this is that new types of kernels are required, which reflect the new measure of distance. The smoothing parameter C_n will also require reinterpretation. Typically for our purposes the spherical kernel will be isotropic and unimodal about its polar (mean) direction, and so there will be only the single parameter, C_n.

Some work has been already done on the sphere using spherical kernel functions (Diggle and Fisher 1985; Watson 1983; Hall et al. 1987). Most of this earlier work has been directed at density estimation. For example, Diggle and Fisher (1985) used the Fisher density function defined as

$$K(\mathbf{X}, \mathbf{X}_i, C_n)$$
$$= \frac{C_n}{4\pi \sinh(C_n)} \exp(C_n \mathbf{X} \cdot \mathbf{X}_i); \quad C_n \in (0, \infty) \quad (9)$$

to compute probability densities for geophysical data. The Fisher density function is a special case of a more general class of spherical kernel functions (Kent 1982) based on an exponential functionality. Figure 1a illustrates the effect of the smoothing parameter on the Fisher density. For small values of C_n the distribution is spread more uniformly on the sphere than for larger values where the distribution is more strongly peaked about its mean direction. To relate this to the Cartesian

interpretation of the smoothing parameter and by analogy to filter design, we define an arc bandwidth r_b (ABW) of the Fisher density to be the geodesic distance from $\mathbf{X}_i$ to the half-power point. Thus,

$$r_b = \arccos\left[1 - \frac{1}{C_n} \ln(2) \right]. \quad (10)$$

Hence, as $C_n \to \infty$, then $r_b \to 0$ and $K(\mathbf{X} \cdot \mathbf{X}_i, C_n) \to \delta(\psi_i)$, where $\delta(\psi_i)$ is the Dirac delta function [note: $\delta(\psi) \equiv \delta(1 - \mathbf{X} \cdot \mathbf{X}_i)$]. As $C_n \to 0$, r_b is indeterminate, since $K(\mathbf{X} \cdot \mathbf{X}_i, C_n) \to U_s$ where U_s is the uniform distribution on the sphere. In effect then by analogy with Cartesian kernels the parameter C_n behaves like a reciprocal bandwidth, with r_b a measure of the radial spread of the function on the sphere.

The drawback to distribution functions dependant on exponentials is their computational expense, particularly if the dataset is large. Notice also that these types of distribution are nonlocal in that they are defined on the whole of the sphere so that the influence of a data point covers the whole of the sphere, albeit controlled by the smoothing parameter.

To improve the computational efficiency of kernel estimation on the sphere, a new set of kernels are introduced in this paper. These are much easier and faster to compute since they do not rely on any transcendental functions. They are also defined locally so that the influence of a data point is restricted to a local region.

They are given by the following expressions.

1) Power:

$$K_p(\mathbf{X} \cdot \mathbf{X}_i, \tilde{C}_n) = \begin{cases} \dfrac{(m+1)\tilde{C}_n}{2\pi(\tilde{C}_n - 1)^{m+1}} (\tilde{C}_n \mathbf{X} \cdot \mathbf{X}_i - 1)^m, & \mathbf{X} \cdot \mathbf{X}_i \geq \dfrac{1}{\tilde{C}_n} \text{ and } \tilde{C}_n > 1 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

for some integer index $m \geq 0$.

2) Quadratic:

$$K_q(\mathbf{X} \cdot \mathbf{X}_i, \tilde{C}_n) = \begin{cases} \dfrac{3\tilde{C}_n}{2\pi(\tilde{C}_n - 1)^2(\tilde{C}_n + 2)} [(\tilde{C}_n \mathbf{X} \cdot \mathbf{X}_i)^2 - 1], & \mathbf{X} \cdot \mathbf{X}_i \geq \dfrac{1}{\tilde{C}_n} \text{ and } \tilde{C}_n > 1 \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

3) Biweight:

$$K_b(\mathbf{X} \cdot \mathbf{X}_i, \tilde{C}_n) = \begin{cases} \dfrac{15\tilde{C}_n}{2\pi(\tilde{C}_n - 1)^3(3\tilde{C}_n^2 + 9\tilde{C}_n + 8)} [(\tilde{C}_n \mathbf{X} \cdot \mathbf{X}_i)^2 - 1]^2, & \mathbf{X} \cdot \mathbf{X}_i \geq \dfrac{1}{\tilde{C}_n} \text{ and } \tilde{C}_n > 1 \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Note the two special cases of the power kernel, $m = 0$ a constant spherical cap (estimates will have no continuity, analogous to a histogram) and $m = 1$ (the linear case, continuous but not differentiable), which has been used to produce the results described later in this paper.
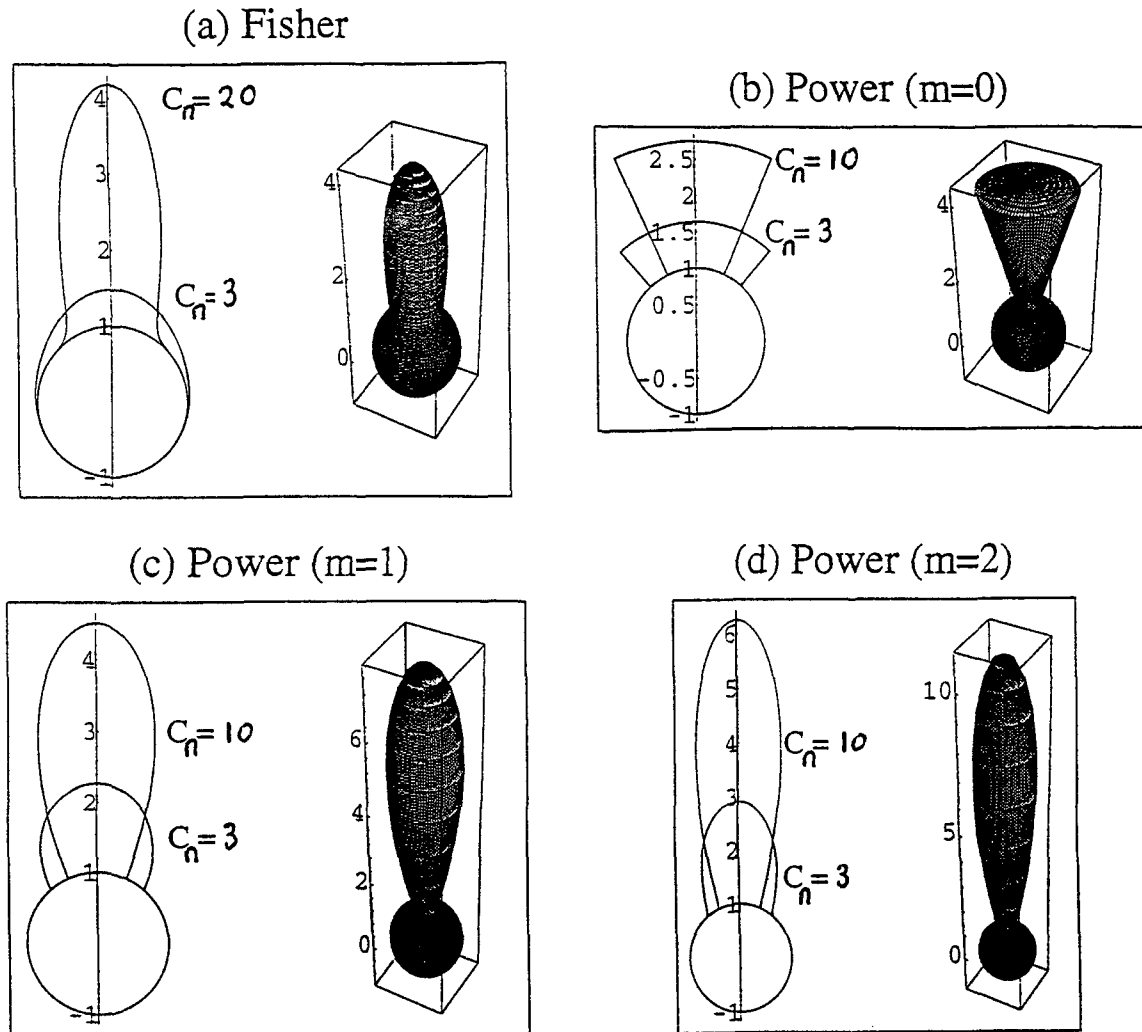
(a) Fisher

(b) Power (m=0)

(c) Power (m=1)

(d) Power (m=2)

FIG. 1. Plots of the cross section of kernels with respect to the unit sphere for two different values of the smoothing parameter C_n, together with a 3D view for a single value of $C_n = 20$ also with respect to the unit sphere, and with the mean direction aligned with the geographical north pole.

A tilde ($\tilde{\ }$) is put on the smoothing parameter symbol for these new kernels to indicate that it is not the same parameter in terms of the limiting behavior of the kernels as that of the Fisher density smoothing parameter. Note that the asymptotic properties for all these new kernels are the same. Namely, that as $\tilde{C}_n \rightarrow \infty$ the kernel encompasses a whole hemisphere centered on its mean direction (this is maximum smoothing), and as $\tilde{C}_n \rightarrow 1$ then $K(\mathbf{X} \cdot \mathbf{X}_i, C_n) \rightarrow \delta(\psi_i)$. To make the smoothing parameter of these new kernels consistent with that of exponential kernels, we rewrite it as

$$\tilde{C}_n = 1 + \left(\frac{1}{C_n}\right); \quad C_n \in (0, \infty). \quad (14)$$

An arc bandwidth r_b (ABW) is also defined as

$$r_b = \left|\arccos\left(\frac{C_n}{1 + C_n}\right)\right| = \arccos\left(\frac{1}{\tilde{C}_n}\right). \quad (15)$$

Now, as with the Fisher density, C_n behaves as an inverse bandwidth. The limiting behavior with respect to C_n is now $\tilde{C}_n \rightarrow 1$, $K(\mathbf{X} \cdot \mathbf{X}_i, C_n) \rightarrow \delta(\psi_i)$, and $r_b \rightarrow 0$ as $C_n \rightarrow \infty$ and that $\tilde{C}_n \rightarrow \infty$ and $r_b \rightarrow 90°$ as $C_n \rightarrow 0$. The ABW now acts as a measure of the radial extent of these new kernel functions on the sphere.

The transformation, (14), also makes the discussion that follows on determining suitable values for C_n consistent across all the spherical kernels so far mentioned. The new kernels are displayed in cross section for two values of C_n and in 3D for a single value of C_n in Fig. 1. Note that, for small values of C_n, the quadratic and biweight kernels are very similar to the power kernel

with $m = 1$ and $m = 2$, respectively, and so are not plotted. Differences occur for larger values of C_n than those used in Fig. 1.

These functions still require some statistical analysis to allow some quantitative assessment of which kernel is best to use. This would entail an in-depth analysis (analytically or numerically) of such statistical quantities as mean integrated square error. This is an involved process and is not presented here. In practical statistical estimation the choice of the kernel is usually based on the type of continuity required and the ease and speed of computation.

Now that a computationally efficient set of kernel functions have been defined it remains to determine, objectively, a suitable value for the smoothing parameter C_n.

d. Choosing the smoothing parameter

Although a subjective choice of smoothing parameter may be made, it is not sensible to choose a value blindly, since an undersmoothed (high variance) or oversmoothed (high bias) estimate may result with no quantification of how good or bad the estimate is. An automatic choice that satisfies some optimal condition is essential to ensure that our estimates are objective; a subjective choice within some interval around the optimum can then be explored. Choosing an optimal smoothing parameter automatically has been extensively studied in the Cartesian domain. Much less has been done in the spherical domain, although Hall et al. (1987) have explored the use of cross-validation for spherical density estimation.

The approach taken in this paper to choosing the smoothing parameter is based on cross-validation. This is a technique that is of general applicability in statistical prediction (Stone 1974) and is applied in both density and regression estimation to determine the smoothing parameter automatically. The idea is to have some statistical predictor, a cross-validation function, that can be optimized with respect to the smoothing parameter.

The use of maximum-likelihood estimation of the smoothing parameter for density estimation is considered first. The cross-validation function is very easy to compute, although there are several drawbacks to its use. Thus, the likelihood is generally expressed in terms of the natural logarithm (log-likelihood) as

$$\Gamma^{(d)}(C_n) = \frac{1}{n} \sum_i^n \log_e[\hat{f}(\mathbf{X}_i, C_n)], \qquad (16)$$

where $\hat{f}(\mathbf{X}_i, C_n)$ is an estimate of the density at the data point $\mathbf{X}_i$ and is dependant on the smoothing parameter C_n. Thus, maximizing the likelihood with respect to C_n should lead to the best value of C_n to use in estimating the density. However, expression (16) is a monotonically increasing function of C_n and if this is used as the

cross-validation function for the density, and is maximized to produce the maximum-likelihood estimate of the smoothing parameter C_n, the likelihood would be found to be infinite (a biased estimate of C_n). The reason for this stems from the self-contribution of a data point to the density estimate at that point. There are essentially two ways around this problem. The first approach is to remove the self-contribution, the so-called leave-one-out approach, the second approach is to use penalty functions. The leave-one-out approach has been used by Diggle and Fisher (1985) in their estimation of densities on the sphere using the Fisher density function, it is the main technique used here for any of the kernel functions mentioned above. The use of penalty function techniques are not discussed further in this paper. The cross-validation function is still the log-likelihood function, but the estimated densities at each of the data points are computed leaving the self-contributions out; thus,

$$\Gamma^{(d)}(C_n) = \frac{1}{n} \sum_{i=1}^n \log_e[\hat{f}_{-i}(\mathbf{X}_i, C_n)], \qquad (17)$$

where

$$\hat{f}_{-i}(\mathbf{X}_i, C_n) = \frac{1}{n-1} \sum_{j \neq i}^n K(\mathbf{X}_j \cdot \mathbf{X}_i, C_n). \qquad (18)$$

The expression (17) is then maximized with respect to the smoothing parameter C_n. Any efficient 1D maximization routine that does not require derivatives can be used, for example, the golden search or parabolic interpolation techniques (Press et al. 1988).

The main drawbacks to cross-validated kernel density estimation with fixed smoothing parameters are that the density estimate can be disproportionately influenced by outlying data points (leading to oversmoothing as the cross validation attempts to accommodate the outlying data), and it can also be disproportionately influenced by small dense clusters of points (leading to undersmoothing). Chow et al. (1982) have suggested that kernels with short tails (zero probability outside of some local region)—for example, our locally defined new spherical kernels— lead to oversmoothing of densities with long tails (nonzero probability outside of the main region of activity), and vice versa. The influence of the tails of the kernels on the density estimate can be reduced by using an adaptive smoothing parameter. This will be discussed shortly in the context of both density and regression estimation.

For regression estimation least squares cross-validation is used, the aim being as with density estimation to determine a value of the smoothing parameter automatically. As with the likelihood cross-validation function for density estimation a leave-one-out approach is used, so the cross-validation function is

$$\Gamma^{(r)}(C_n) = -\frac{1}{n}\sum_{i=1}^{n} |[\mathbf{Y}_i - \hat{\mathbf{Y}}_{-i}(\mathbf{X}_i)]|^2, \quad (19)$$

where

$$\hat{\mathbf{Y}}_{-i}(\mathbf{X}_i) = \frac{1}{n-1}\sum_{j \neq i}^{n} W_j(\mathbf{X}_i)\mathbf{Y}_j. \quad (20)$$

Maximizing expression (19) should provide a suitable choice for the smoothing parameter.

The next section explains how the smoothing can be varied adaptively, so that the parameter C_n can be adjusted to reflect the distribution of the data. This is useful not only in the context of cross-validation as mentioned above, but intuitively, where the data density is high, we want only the nearest data points to contribute significantly to the estimate so that the detailed structure is reproduced. On the other hand, where the data density is low, and if the value of C_n is too large (small amount of smoothing), then spurious bumps will be present. If C_n is changed to smooth the bumps in the low density regions, detail will be lost from the high-density regions.

e. Adaptive smoothing

Adaptive smoothing has been implemented to make the smoothing more responsive to the density of the spherical data for both density and regression estimation. Recall that using a global smoothing parameter C_n, which is only dependant on the data size, can lead to undersmoothing or oversmoothing. To make the smoothing more responsive to the data distribution, the algorithm of Silverman (1986) is easily adapted to the spherical domain. This basically adapts the smoothing parameter according to the local density, the smaller the density the more a data points influence is spread and vice versa (Jones 1990). The algorithm begins by first computing a pilot estimate of the density function at the data points $\mathbf{X}_i$. Experience suggests (Silverman 1986) that the final estimate, at least for density estimation, is relatively insensitive to the choice of pilot estimate. Hence, a pilot estimate constructed using a straightforward subjective choice for the smoothing parameter can be used. However, a little care is required to not choose a pilot estimate with gross oversmoothing or undersmoothing. Using a cross-validation calculation for the pilot estimate is not computationally sensible at this stage. In practice, a somewhat oversmoothed pilot estimate is found to give the best results.

The local smoothing parameters are defined by

$$C_n(\mathbf{X}_i) = C_n^{(g)}\left[\frac{\hat{f}_p(\mathbf{X}_i)}{g}\right]^{\beta}, \quad (21)$$

where $C_n^{(g)}$ is a global smoothing parameter that is locally adapted by the term in brackets. The normaliza-

tion factor g is the geometric mean of the pilot densities ($\hat{f}_p$); that is,

$$\log_e g = \frac{1}{n}\sum_{i=1}^{n} \log_e[\hat{f}_p(\mathbf{X}_i)]. \quad (22)$$

The kernel functions are now defined in terms of the local smoothing parameters as

$$K(\mathbf{X}\cdot\mathbf{X}_i, C_n) \equiv K[\mathbf{X}\cdot\mathbf{X}_i, C_n(\mathbf{X}_i)]. \quad (23)$$

The exponent, $\beta \in [0, 1]$, is a sensitivity parameter that determines the sensitivity of the smoothing to the underlying density, so that for $\beta = 0$ [$C_n(\mathbf{X}_i) = C_n^{(g)}$] global smoothing is applied with no local adaptivity. Notice that there is still a global smoothing parameter $C_n^{(g)}$ to determine. This is achieved within the cross-validation framework previously discussed keeping the pilot estimate unchanged during the maximization procedure as suggested by Silverman (1986). The choice of β is more problematic. In the Cartesian domain there is some evidence to suggest that $\beta = 0.5$ is a suitable value (Silverman 1986), at least in the context of density estimation. The value that has been used for this paper was chosen subjectively as $\beta = 0.6$.

f. Seasonal statistics: Combining distributions

Climatologists are often interested in long-term seasonal statistics, in other words, the statistics for individual seasons over many years, for example, taking all the DJF (December, January, February) seasons. Also, comparisons are often made between different climatological regimes, for example, El Niño–Southern Oscillation (ENSO) and La Niña, so we might require the statistics for all the El Niño occurrences in a 10-year period say.

To pool the data for several long time periods and use the techniques discussed above would almost certainly be computationally prohibitive even using local kernal functions; hence, means to combine the statistical distributions of the individual seasons have been devised.

Distributions can be combined using generalizations of well-known formulas for this purpose. Thus, for k sets of statistical diagnostics, computed on identical estimation grids, the composite distributions are computed using the following formulas:

$$\hat{f}(\mathbf{X}) = \frac{\sum_{i=1}^{k} n_i \hat{f}_i(\mathbf{X})}{\sum_{i=1}^{k} n_i} \quad (24)$$

$$\hat{\mathbf{Y}}(\mathbf{X}) = \frac{\sum_{i=1}^{k} n_i \hat{f}_i(\mathbf{X})\hat{\mathbf{Y}}_i(\mathbf{X})}{\sum_{i=1}^{k} n_i \hat{f}_i(\mathbf{X})} \quad (25)$$

$$\hat{\sigma}^2(\mathbf{X}) = \frac{\sum\limits_{i=1}^{k} n_i \hat{f}_i(\mathbf{X}) \hat{\sigma}_i^2(\mathbf{X})}{\sum\limits_{i=1}^{k} n_i \hat{f}_i(\mathbf{X})}$$

$$+ \frac{1}{2} \frac{\sum\limits_{i=1}^{k} \sum\limits_{\substack{j=1 \\ i \neq j}}^{k} n_i n_j \hat{f}_i(\mathbf{X}) \hat{f}_j(\mathbf{X}) [\hat{\mathbf{Y}}_i(\mathbf{X}) - \hat{\mathbf{Y}}_j(\mathbf{X})]^2}{[\sum\limits_{i=1}^{k} n_i \hat{f}_i(\mathbf{X})]^2}.$$

(26)

These formulas can be evaluated iteratively for convenience and efficiency. Notice that for a fixed smoothing parameter these formulas are exact. However, if differing global smoothing parameters have been used for each individual season or adaptive smoothing has been used, then these combined statistics may not be identical to those obtained by pooling the data for a season over a number of years. To determine if substantial differences occur, the combined DJF statistics for two years of the AMIP integration (each seasonal statistic computed using the same global smoothing parameter and adaptive smoothing) were computed using the above formulas. A second calculation was performed using the pooled data for the DJF's of both years and recomputing the statistics using the same global smoothing parameter and adaptive smoothing. By eyeball comparison of the contoured distributions, the differences between the two sets of distributions were deemed negligible for each of the statistical diagnostic fields.

3. Application to the UGAMP AMIP integration

The dataset used for this study is the field of the vertical component of the relative vorticity at 850 mb. The time step between scenes is 6 h and the data is defined on a T42 Gaussian grid (longitude–latitude grid of size 128 × 64 grid points). The tracking and statistical analysis are performed globally, for the full 10-year period, on the seasonal periods of December, January, February (DJF); March, April, May (MAM); June, July, August (JJA); and September, October, November (SON) for both the positive and negative vorticities. However, only the selected results for the cyclonic activity for the NH DJF (positive) and SH JJA (negative) are presented here.

Before any statistics are computed the track data are filtered to remove tracks that have a lifetime less than a day. The following statistics are considered: feature density, genesis density (cyclogenesis is first appearance of a cyclonic center, provided it does not coincide with the beginning of the data sequence), lysis density (cyclolysis is last appearance of a cyclonic center, provided it does not coincide with the end of the data se-

quence), track density, mean strength (vorticity in this case), standard deviation of strength, mean speed, standard deviation of speed, mean lifetime, and mean velocity. All the densities are probability densities except for the track density. This is because of the way the track density is computed; it is a function not only of the distribution of the data but also a function of the distribution of the estimation points. The reason for this is that the data that are used to compute the track density (and mean lifetime) are the points that correspond to the minimum distance from an estimation point to the set of tracks; thus, the data are different for each estimation point. In principle the distribution can be renormalized by integration of the distribution of the sphere to produce the probability density distribution for tracks passing through a region, but we do not attempt this here. The mean lifetime is interpreted as the mean lifetime of tracks passing through a region. For both the track density and mean lifetime cross-validation is not attempted as this would mean performing a cross-validation calculation for each estimation point. In principle it can be done and the value of $C_n^{(g)}$ could be computed as the average of those values obtained for each estimation point. It has been pointed out that the mean lifetime statistic can be misleading if it is not clearly understood what it represents. This statistic represents the mean lifetime of tracks passing through some local region, as if an observer positioned at some point notes all the systems passing within some distance of him. If someone later tells him the complete history of his observed systems, he can then compute the mean lifetime of those systems for that particular observation station. This is effectively what is done for each grid point.

The statistics are computed using adaptive smoothing with the power kernel, with $m = 1$ (linear), and a subjective choice of the global smoothing parameter. The sensitivity parameter for adaptive smoothing was subjectively chosen to be 0.6. A guide to the choice of the global smoothing parameter is obtained by running the adaptive cross-validation procedures on a single sample of the track data for a single season. A subjective choice is then made, based on the cross-validation values, which are then kept constant for all further analyses; thus, the global amount of smoothing is the same between different analyses although locally the smoothing will vary with the local density of the data. This seems a reasonable approach for the large amounts of data for each three-monthly analysis. The subjective choice is generally a smaller value of $C_n^{(g)}$ (more smoothing) than the cross-validation values due to the propensity of the cross-validation to lead to undersmoothing, particularly for the feature density. It is thought that this may be due to some regions having a large number of relatively short-lived semistationary features resulting in high density clusters of points. Filtering the tracks at longer lifetimes and/or according

to their displacement distance partly remedies this problem, although for this paper all the tracks with lifetimes longer than a day have been retained. However, from what we have seen so far a judicious choice of the global smoothing parameter for use with the adaptive smoothing technique on our large datasets seems to provide a relatively efficient and visually satisfying means of determining the statistics.

The values used for the global smoothing parameter for each statistic are shown in Table 1 along with the cross-validation values (CV) and the values of the ABW.

One further point should be noted before a discussion of the results is given; the plots of mean physical attributes — for example, strength, speed, etc. — should not be considered in isolation, but in conjunction with the feature density. Thus, where the feature density is high, our confidence in the estimated local mean value of a physical attribute will be high; whereas, where the feature density is low, our confidence in the estimated mean attribute will be low. Confidence maps can be computed but are not presented here.

4. Results

The results of applying the techniques described above to the tracks determined from the AMIP integration are now presented and discussed for each hemispheric winter. The standard deviation plots, although mentioned, are not presented to save space and in any case provide only subsidiary information to the main plots of strength and speed.

a. Northern Hemisphere DJF, 1979–88

The results for this period are shown in Fig. 2. The plots in Fig. 2a of the feature density and in Fig. 2b of the track density show, overall, a very similar distribution of cyclonic activity, although they are different in their detail. There are two well-developed storm tracks (baroclinic waveguides, seen more prominently in the track density) in the Pacific and Atlantic with the feature density showing locally high regions of activity within the storm tracks, while the track density is higher in the southern parts of the storm tracks. This difference can be accounted for by the greater speeds in the southern parts of the storm tracks (see Fig. 2d) so that although there are more tracks there are fewer features per unit length of track. Figure 2g shows that the mean direction is west to east with the Atlantic storm track splitting in two over Europe. Mean lifetimes (Fig. 2h) for systems passing through a region are 2–5 days in the Pacific with longer mean lifetimes in the western Pacific. In the Atlantic the mean lifetimes are somewhat less, at 2–3.5 days. The reason lifetimes are greater in the western oceans than the eastern oceans is in part attributable to the generation of secondary cyclonic

TABLE 1. Values for the global smoothing parameter $C_n^{(g)}$ found by cross-validation for each statistical diagnostic and the values chosen subjectively and used to compute the statistics. The ABW for each value of $C_n^{(g)}$ are given in parentheses.

Diagnostic	CV $C_n^{(g)}$	Subjective $C_n^{(g)}$
Mean and STD of strength	722.8 (3.0)	300.0 (4.67)
Mean and STD of speed	886.4 (2.72)	300.0 (4.67)
Feature density	1624.6 (2.0)	300.0 (4.67)
Genesis density	300.9 (4.66)	100.0 (8.07)
Lysis density	328.45 (4.46)	100.0 (8.07)
Mean velocity	641.9 (3.19)	100.0 (8.07)
Track density	too expensive	200.0 (5.72)
Mean lifetime	too expensive	200.0 (5.72)

events associated with the more mature systems. This can be observed both from track histories [e.g., those shown in Hodges (1995)] and in animations of the time series. These secondary systems will generally be much shorter lived due to being spawned from the more mature systems and thus will act to reduce the mean lifetime in the eastern oceans. Mean speeds (Fig. 2d) in the storm tracks are 40–70 km h^{-1} (11–20 m s^{-1}) in the Pacific, and 40–60 km h^{-1} (11–17 m s^{-1}) in the Atlantic. The standard deviation of the strength (not shown) indicates that the greatest variation of the strength to be in the western Pacific and Atlantic where the intensification of systems is greatest, while the strongest systems (Fig. 2c) are in the eastern Pacific and Atlantic and are more mature systems.

The study of Lambert (1988) has compared cyclone densities (cyclones identified by lows in the 1000-mb height field from data sampled once per day) for the Canadian Climate GCM and ECMWF/WMO analyses for the period 1980–84 for both the SH and the NH. The distribution of cyclonic activity obtained for the ECMWF/WMO was compared favorably to that of Klein (1957). The distribution shown in Fig. 2a shows many similarities to Lambert's and Klein's results for the NH. The various local highs of cyclonic activity seem to correspond relatively well. For the Pacific storm track Lambert's distribution of cyclone density (ECMWF/WMO) shows several maxima, in the Gulf of Alaska, the Sea of Okhotsk, and to the east of southern Japan with several weaker maxima at the southern end of Japan and the landward side of the Yellow Sea. Figure 2a shows a somewhat more elongated maxima in the Gulf of Alaska that bears more similarity to Klein's cyclone frequencies for December and January (Klein's southern part is weaker in February). There is also a maxima over Alaska and the Bering Strait, which does not seem to appear in other studies. A maxima that extends from the Sea of Okhotsk and northern end of the Sea of Japan to the south of Kamchatka, a very intense maxima in Mongolia and northeast China, and several weaker maxima in the Yellow Sea and southern China are also present in Fig. 2a. The first of
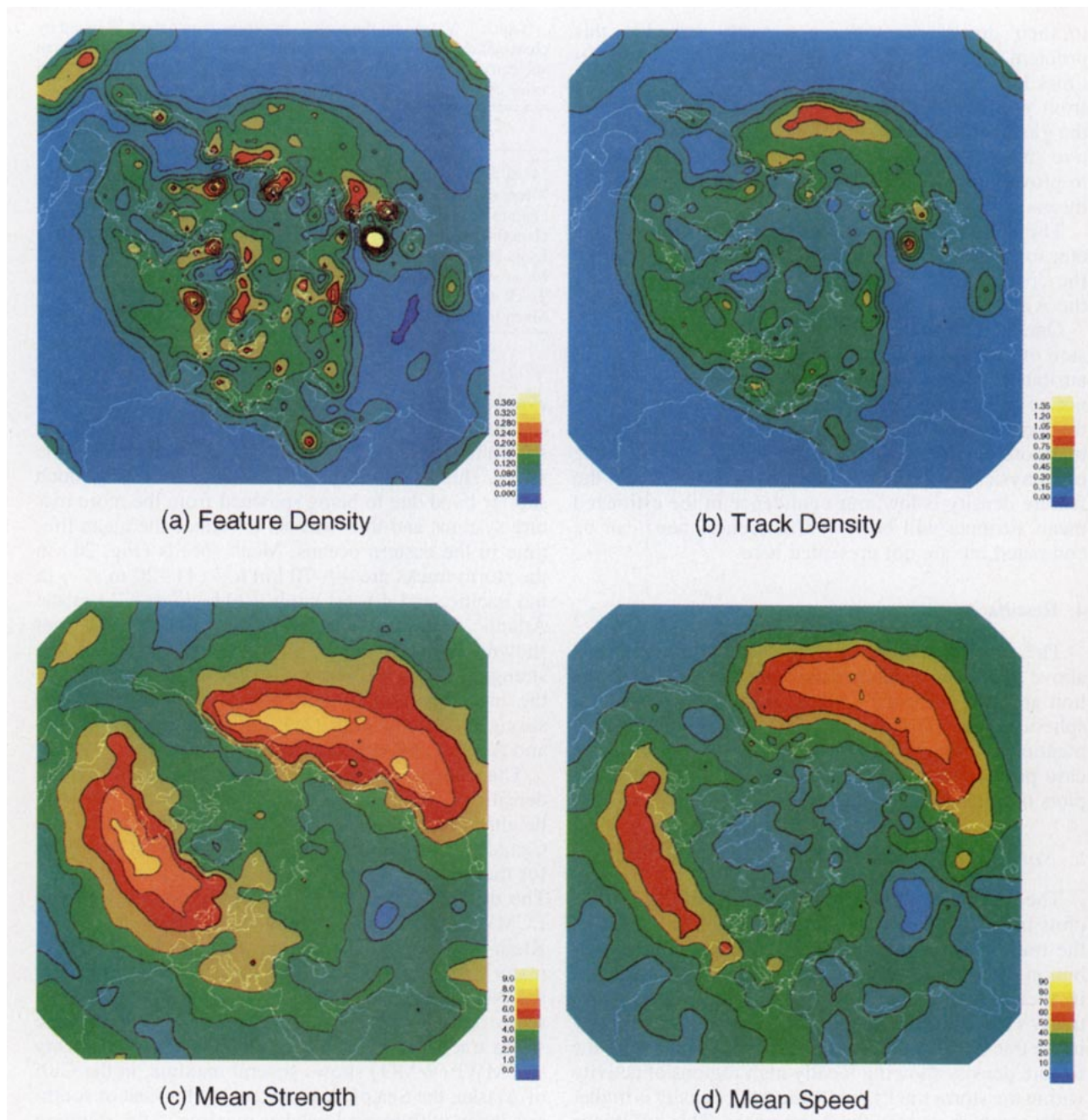
(a) Feature Density

(b) Track Density

(c) Mean Strength

(d) Mean Speed

FIG. 2. Cyclonic statistical diagnostic fields for NH DJF, 1979–88: (a) feature density, (b) track density, (c) mean strength (10^{-5} s^{-1}), (d) mean speed (km h^{-1}),

these areas, the Gulf of Alaska and Alaska itself are also regions of cyclolysis (Fig. 2f), while the later are all regions of cyclogenesis (Fig. 2e). The good agreement in the Gulf of Alaska with Lambert and more particularly Klein is undoubtedly due to the fact that systems in the region are mature (see Fig. 2c) and also moving relatively slowly (Fig. 2d), thus making identification and tracking as easy for relative vorticity as

it is for MSLP and 1000-mb geopotential height. Alaska as a cyclolysis region does not seem to have been previously discussed in the literature, observation of the tracks (e.g., those of Hodges 1995) and the mean velocity (Fig. 2g) indicates that some systems recurve from the main storm track over Alaska and the Bering Strait before disappearing. There is also some indication of the generation of cyclonic activity, which may
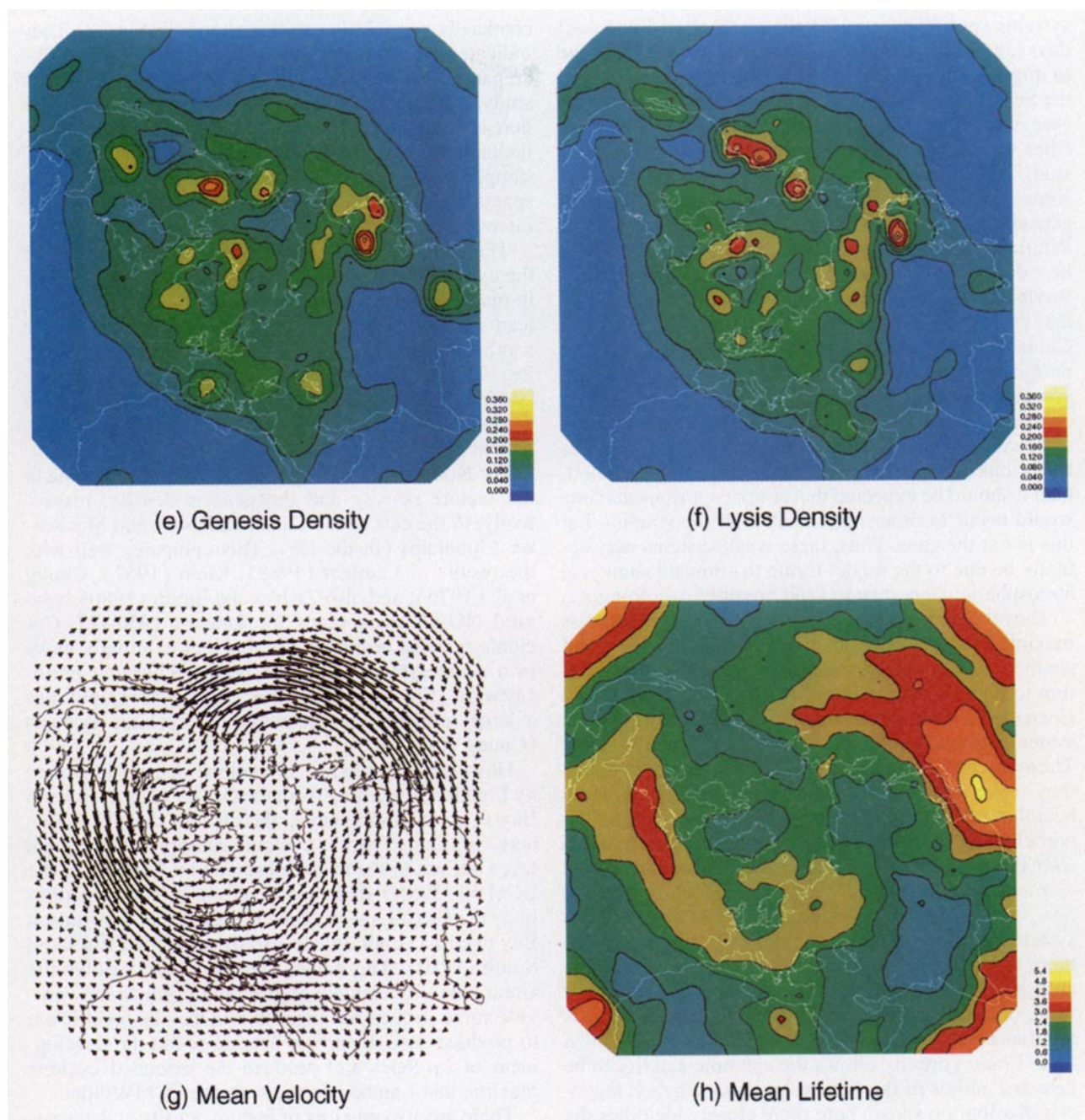
(e) Genesis Density

(f) Lysis Density

(g) Mean Velocity

(h) Mean Lifetime

FIG. 2. (*Continued*) (e) genesis density, (f) lysis density, (g) mean velocity, (h) mean lifetime (days).

be due to secondary cyclonic centers associated with the more mature systems.

For the cyclogenetic regions, the maxima in Mongolia and northeast China (also a region of cyclolysis) does not appear in Klein's distribution (using MSLP) and only appears weakly in Lambert's (using 1000-mb geopotential height). This can be partly associated with the methods of counting, Klein counts only migratory systems; thus, slow-moving systems that remain local to a region and stationary systems are not counted. Whereas Lambert associates the weak activity in his cyclone distribution to poor reduction to pressure levels resulting in extrapolation below the land surface. The fact that the distributions for genesis and lysis shown here have coincident maxima in this region and that the strengths are relatively weak indicates that many of the

systems (not all) are relatively weak, short lived ($\sim$2 days), and fairly immobile (some systems do manage to migrate out of this region into the feature maxima in the Sea of Japan where further development is possible (see track plots of Hodges 1995). However, Boyle and Chen (1992) show some results for this region from a study of Whittaker and Horn (1982), which do show some similarities to the distributions shown here for genesis and feature density in this region. The study of Whittaker and Horn may be more reliable than Klein here due to better and more frequent observational data. Boyle and Chen (1992) discuss this behavior noting that cyclogenesis is different in character in northeast China and Mongolia to that in southern China. For northeast China and Mongolia, they noted that cyclones generally form as secondary cyclones on the trailing cold fronts of occluded lows in western Siberia and that these systems rarely gain any great intensity. If this was a spurious topographic effect, as suggested by Lambert, then it should be expected that similar spurious maxima would occur in other regions of high topography, but this is not the case. Thus, these weak systems may actually be due to the model trying to simulate some real atmospheric phenomenon (and possibly overdoing it).

The other regions of feature density and cyclogenesis maxima (Sakhalin–Kamchatka, Yellow Sea, and southern China) are displaced southwestward in relation to Klein's and Lambert's distributions. Boyle and Chen (1992) note that in southern China many cyclonic systems originate as shallow lows or inverted troughs. These intensify (deepen), rapidly so in many cases, as they move offshore and the cold dry air triggers large heat fluxes from the warm ocean currents. This and the baroclinicity of the region lead to cyclogenesis maxima over the East China Sea and to the east of Japan. The warm ocean currents that penetrate the Yellow Sea and Sea of Japan probably have a similar effect for those systems that originate inland. This has led many au-·thors to not count the cyclonic activity until considerable development has occurred, since they do not count weak or open systems; hence, their regions of genesis are translated northeastward compared to those shown here. Using vorticity allows the cyclonic activity to be detected earlier in the life cycle; hence, the cyclogenesis distribution shown here more closely identifies the source regions. In contrast, previous studies using MSLP or geopotential height with features identified as closed systems identify only regions where the intensification is strong and not the initiation.

The discrepancies between counting systems at initiation or intensification was discussed by Chung et al. (1976). They have used observational data from the IGY to determine the lee cyclogenesis, as initiation events, in the Canadian Rockies and East Asian mountains. Their distribution of cyclogenesis for the East Asia sampling region (albeit for one year only, and for the whole year, and not extending as far north as Kam-

chatka) is very similar to that shown in Fig. 2e. They indicate that sampling strategies may also lead to discrepancies between studies, due to different lengths of study, different data densities (some regions are insufficiently sampled due to lack of observational sites, particularly the early studies), and insufficient temporal sampling (e.g., daily) so that weak cyclones in data sparse regions are not easily identified until they have intensified sufficiently.

The general eastward and northeastward motion of the cyclonic features following genesis in East Asia is in marked contrast to the behavior on the North American continent where weak lee cyclonic features move southeastward before recurving northeastward (Fig. 2g). This has been attributed to the East Asian cyclonic systems being controlled by upper flows moving around the Himalayas, which prevents a southerly component developing (Chung et al. 1976).

For North America Figs. 2a and 2e show maxima in the feature density and the genesis density, respectively, to the east of the Canadian Rocky and Mackenzie Mountains (in the lee). This compares well with the results of Lambert (1988), Klein (1957), Chung et al. (1976), and also Zishka and Smith (1980), who used NOAA track maps for January 1950–77. Cyclonic systems originating in this region initially move in a southeastward direction before recurving northeastward (Fig. 2g), indicating that the orography has a large influence on the weak lee cyclonic systems (Chung et al. 1976).

However, the maxima in cyclone density identified by Lambert, stretching from the Great Lakes to Hudson Bay is not as prominent in the cyclone density shown here (Fig. 2a). There is cyclonic activity over the great lakes but not to the extent observed by Lambert in the ECMWF/WMO data. However, in Klein's distributions the Great Lake maxima is present but the Hudson Bay maxima is not, this is also the case for Zishka and Smith (1980). The model used here does resolve the Great Lakes as a couple of grid points, which may provide some surface forcing, but this may be insufficient to produce enough surface heating to aid the development of cyclones and produce the extended cyclone maxima that Lambert found with the ECMWF data.

There are also maxima of feature density and genesis density in the lee of the southern Rocky Mountains, systems from this region move northwestward (Fig. 2g) some up through the Great Lakes, others toward the American east coast, to join up with the Atlantic storm track. This appears to agree relatively well with other studies (e.g., Zishka and Smith (1980).

Lambert (1988) discusses the penetration of cyclonic events into the Canadian Rocky Mountains from the Gulf of Alaska. Figures 2a and 2b indicate a similar behavior, but this is weak as is the behavior in Lambert's results for the ECMWF/WMO data. This behavior is not apparent in other studies and once again may

be due to the definition of a cyclone (closed systems versus open systems). This indicates that the Canadian Rocky Mountains act as a substantial barrier to the movement of Pacific cyclonic activity into the North American continent. Lambert (1988) also comments on some spurious activity of the coast of Baja California, attributing this to the surface reflection of upper cold lows, which are present during the winter. Figure 2a also shows this type of activity, presumably for the same reason.

In the Atlantic (Fig. 2a), there are several regions of cyclonic activity maxima; Baffin Bay and Davis Strait off the western coast of Greenland; Denmark Strait off the east coast of Greenland, stretching in a weaker band over to the British Isles; the Norwegian Sea; the Baltic Sea as a continuation of the Atlantic storm track. This appears to be in reasonable agreement with both Lambert (1988) and Klein (1957). Figure 2f indicates that feature maxima around Greenland (Baffin Bay and Denmark Strait) are also regions of cyclolysis, with the southern end of the Denmark Strait also a region of cyclogenesis. This has been commented on by Murray and Simmonds (1995), who suggest the coincidence of feature density and genesis density and lysis density may be due to tracking errors due to the coarseness of their model resolution. The data used here is at a higher resolution, but tracking errors cannot be ruled out completely. However, in situ development of cyclonic systems is expected in these regions so that previous studies may not have counted them due to their semistationary nature (Murray and Simmonds 1995). The semistationary nature of systems in these regions around Greenland is indicated by the slow speeds, approximately 10 km h^{-1} (3 m s^{-1}) and for Baffin Bay (a lysis region) the weak strengths (Fig. 2c).

The Atlantic storm track stretches into Europe where the mean speed (Fig. 2d) and strength (Fig. 2c) of systems reduces substantially over the land. Also apparent (Figs. 2a and 2g) is the split in the storm track with one branch directed toward southern Europe and eastward toward the Black Sea and the other directed northeastward into northern Europe and Scandinavia. This behavior is also apparent in the principal tracks of Klein (1957). In southern Europe the lee of the Alps appears as a cyclogenetic region (Fig. 2e) with the cyclonic activity stretching south through Italy into the Mediterranean, in agreement with both Lambert (1988) and Klein (1957).

b. Southern Hemisphere JJA, 1979–88

The results for this period are shown in Fig. 3.

There has been much interest recently in the identification of SH cyclonic activity using objective techniques on model analyses, and its synthesis into statistical diagnostic fields (Murray and Simmonds 1991a,b; Sinclair 1994, 1995). The results of these studies have invariably been compared to the manual observational studies of Taljaard (1967), who used MSLP and 500-mb maps for the IGY, and Streten and Troup (1973), who identified cloud vortices from satellite imagery of the SH.

The results shown here for feature density and track density (Figs. 3a and 3b) show a broad band of cyclonic activity from about 35°S to the Antarctic coast but with some lower-level activity from about 20°S. The mean direction of systems within this band (Fig. 3g) is west to east with the highest speeds attained in a band from 45°–55°S to 20°W–160°E with values of 60–70 km h^{-1} (17–20 m s^{-1}). Within the main band the greatest amount of activity lies close to the Antarctic coast but with a secondary band of enhanced activity at about 50°S south of Australia. The embayment regions of Antarctica (Weddell Sea, Ross Ice Shelf) show several maxima in the feature density. There are also several maxima in the high feature density close to the Antarctic coast, namely the Bellinghausen Sea, north of Novolazarevskaya (10°E), north of Molodezhnaya (45°E), north of Prydz Bay (75°E) and a complicated region stretching from about 90° to 150°E.

Outside of the main band of cyclonic activity in the lower feature density regions of the Pacific, Atlantic, and Indian Oceans (north of 35°S), the activity appears to be associated with relatively strong systems (Figure 3c), moving fairly slowly in a southeastward direction (Figs. 3d and 3g) eventually joining up with the main storm track. In the Atlantic these systems originate over Brazil (Fig. 3e), while in the Indian Ocean they originate over the southern end of South Africa. In the Pacific these systems are generated as initially weak systems in the South Pacific convergence zone (SPCZ), but they can be very long lived (Fig. 3h), intensifying as they move southeastward into the storm track.

The regions of strongest cyclogenesis (Fig. 3e) are close to the Antarctic coast and almost coincident with some of the feature density maxima. There are also weaker maxima over the coastal region of Argentina and South Africa. For the cyclolysis (Fig. 3f), maxima in the distribution also occur close to the Antarctic coast and are also coincident with maxima in the feature density. The coast of Chile also has a lysis maxima due to the Andes providing a barrier to the propagation of systems across South America from the Pacific.

These results are in broad agreement with several previous studies, including those of Lambert (1988) using ECMWF/WMO analyses; Sinclair (1994) using the 1000-mb geopotential height of the twice-daily ECMWF analyses; Murray and Simmonds (1991) using MSLP from the Melbourne University GCM; Taljaard (1967) using surface maps and 500-mb maps for the IGY, and Streten and Troup (1973), using satellite imagery. The observation that many of the maxima in cyclonic activity close to the Antarctic coast coincide with embayments and that breaks in the high density
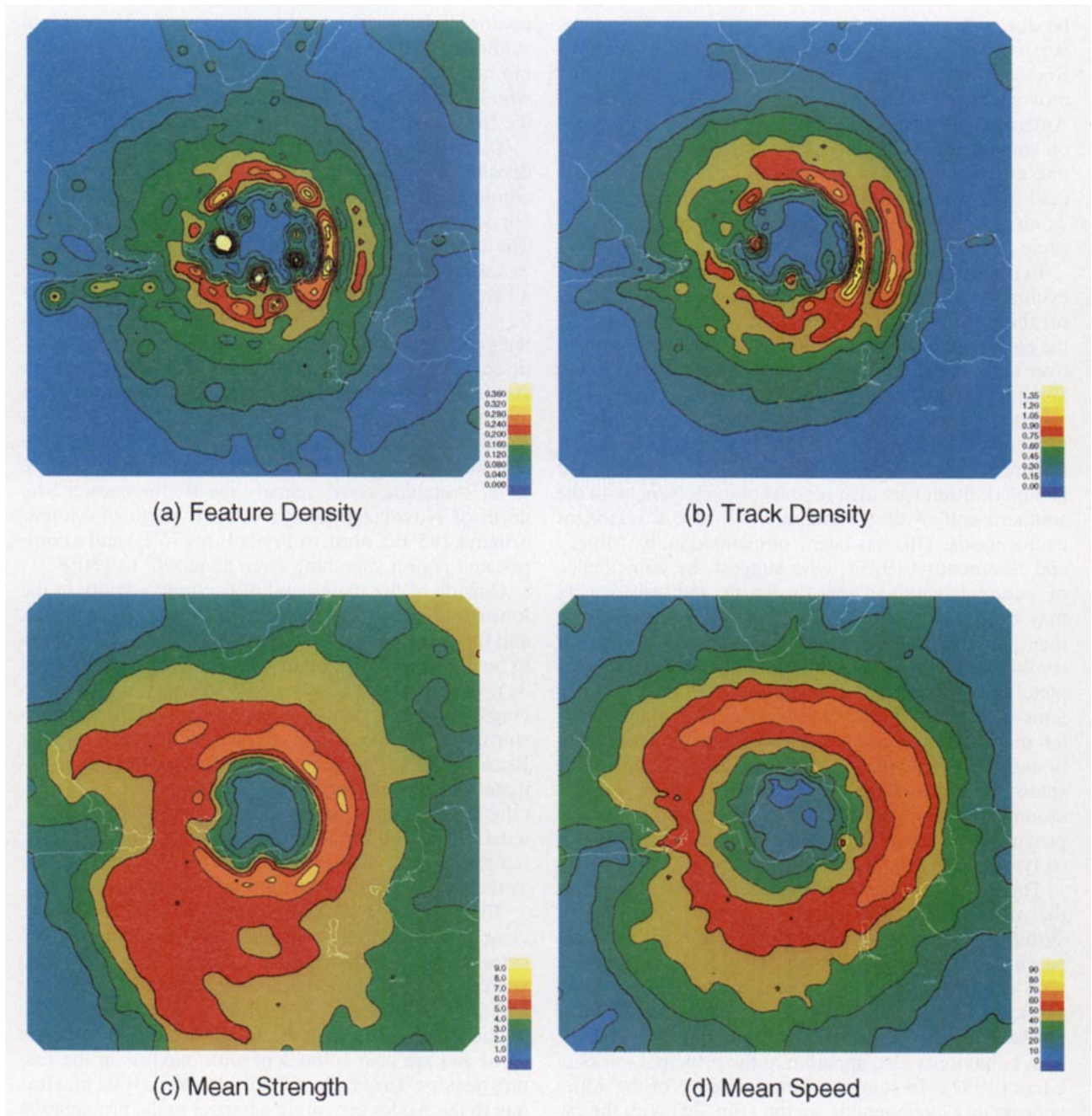
FIG. 3. Cyclonic statistical diagnostic fields for SH JJA, 1979–88: (a) feature density, (b) track density, (c) mean strength (10^{-5} s^{-1}), (d) mean speed (km h^{-1}),

coincide with promontories has been made by several authors (Taljaard 1967; Sinclair 1994). The mean speeds in the high-density regions close to the Antarctic coast are very low in comparison with the central storm track, about 20–40 km h^{-1} (5.5–11 m s^{-1}) with the lower values closest to the coast. This, together with the fact that many of the embayment maxima coincide with genesis and lysis maxima, suggests that many of

the systems that contribute to these maxima are stationary or semistationary systems generated locally due to orographic effects, such as the locally strong katabatic winds, as suggested by Sinclair (1994). However, not all the maxima in the feature density can be attributed to this. Mechoso (1980) points out that mobile cyclonic centers are constrained to the Antarctic coast where they can enhance the baroclinicity, generating
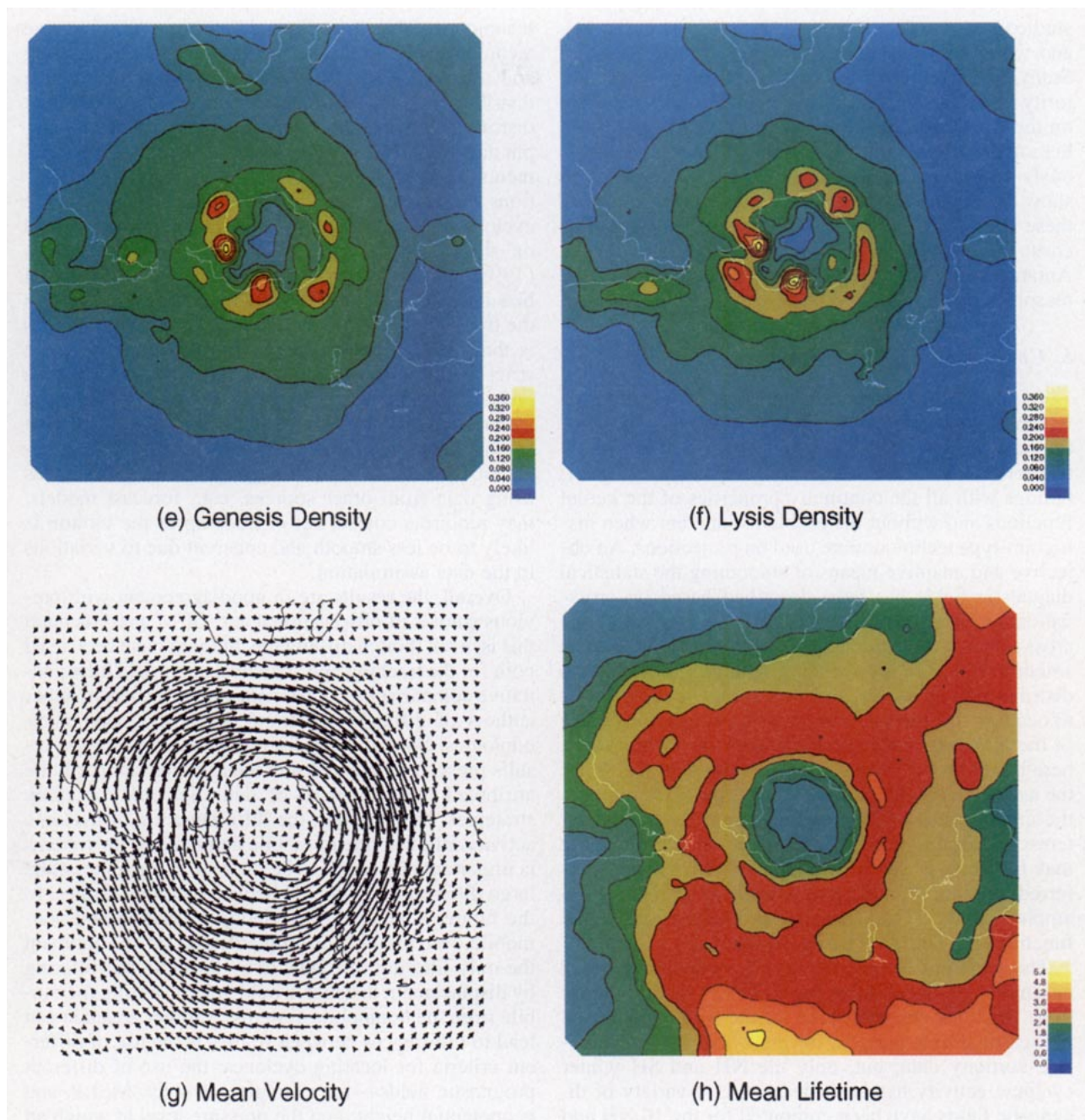
(e) Genesis Density

(f) Lysis Density

(g) Mean Velocity

(h) Mean Lifetime

FIG. 3. (*Continued*) (e) genesis density, (f) lysis density, (g) mean velocity, (h) mean lifetime (days).

strong westerlies over the Antarctic slopes; instabilities then lead to cyclogenesis. The maxima in the lysis density in the Bellingshausen Sea are much stronger than the genesis maxima indicating that this is a region where mobile cyclonic systems coming from midlatitudes die. Some of the other lysis maxima are somewhat stronger than the genesis indicating that not all systems were generated in situ. The Antarctic coast as

a graveyard of mobile midlatitude cyclonic activity has been known for some time (Taljaard 1967).

To study only the mobile cyclonic activity, Sinclair (1994) filtered his tracks according to displacement distance and found that the maxima around Antarctica were much reduced. This is not done here, but future studies undoubtedly should discriminate between mobile and nonmobile features. Sinclair (1995) has also

studied in some detail the genesis and lysis in the SH, and which events should contribute to their statistics. Studying only mobile systems he found that the majority of the genesis occurred at midlatitudes rather than on the Antarctic coast, in contrast to the results shown here and those of the majority of other studies previously mentioned. However, the results shown here do show the cyclogenetic regions at midlatitudes, although these are weak compared to those of the Antarctic coast. This reinforces the conjecture that much of the Antarctic activity is related to in situ generation for the reasons already mentioned.

5. Conclusions and further work

Techniques for estimating statistical properties from feature track data determined objectively have been introduced for a spherical domain. The use of kernel estimators allows the construction of probability distributions with all the continuity properties of the kernel functions and without the problems inherent when histogram-type techniques are used on projections. An objective and adaptive means of smoothing the statistical diagnostic fields has been described based on cross-validation. The adaptivity is useful in relation to the cross-validation, reducing the propensity to undersmooth, and also for reducing spurious detail in the distributions. However, undersmoothing is still deemed to occur to a certain degree, so that a subjective choice of the global smoothing parameters have to be chosen here based on the cross-validated values. In the future the use of penalty functions will be explored to improve the cross-validation calculations. Another area of interest is adaptive nonisotropic spherical kernels, these may be useful where an ensemble of tracks have a preferred direction in a tightly constrained region, for example, in the ITCZ. These types of spherical kernel functions are currently being explored.

The statistical techniques have been applied to track ensembles produced from the UGAMP AMIP integration. The track ensembles are for the whole globe, and for seasonal time periods, for the whole 10 years relative vorticity data; but, only the NH and SH winter cyclonic activity has been discussed. A variety of diagnostic fields have been computed for the 10 NH and SH winters and these have been combined using general formulas to produce a 10-year winter climatology for the two hemispheres. Tropical activity is not discussed, though results are available.

The methodology described for determining the statistics obviates many of the problems that occur with the methodologies used in previous studies. In particular, by working directly on the sphere the introduction of systematic bias can be reduced and the need for latitude-dependent normalization/corrections are made irrelevant. Further, the smoothing is built into the methodology thus obviating the need for ad hoc smoothing

techniques. This, together with the use of objective techniques for determining the amount of smoothing and adaptive smoothing, allows much more reliable distributions to be computed. However, the resultant distributions will reflect only the distribution of the input data, so while the statistical estimators are improvements on previous techniques, the resultant distributions are dependant on the method used to produce the cyclone tracks. In this case the tracks are determined on the spherical domain as described by Hodges (1995), this will prevent the introduction of latitudinal bias into the tracking. The only remaining comment on the tracking that could have an impact on the statistics is that the cyclone tracks were produced using quite strict tracking parameters. This did not appear to be a problem using data from a relatively coarse resolution free running model, where physical fields have been smoothed and the motion is relatively smooth and coherent (as seen in animations). Tracking, of cyclones using data from other sources, e.g., forecast models, may require a coarser set of tracking as the motion is likely to be less smooth and coherent due to variations in the data assimilation.

Overall, the results are in good agreement with previous studies of observational and GCM analysis data; this is, even though different techniques have been used both for the tracking and statistical analysis. Only qualitative comparisons can be made due to the use by other authors of different statistical diagnostics and methodologies. Some of the discrepancies between the results presented here and previous studies can partly be attributed to differences of definition and sampling strategies. In particular, the differentiation of cyclonic activity into that due to mobile and nonmobile systems is important and seems to be the reason for some of the large differences in the SH between various studies. In the future, a better differentiation of mobile and nonmobile systems will be implemented, at the moment the only filtering of tracks is by lifetime, but filtering by displacement distance would help in sorting the mobile from the nonmobile systems. Another area that can lead to differences between studies is the use of different criteria for locating cyclones; the use of different prognostic fields—for example, vorticity, MSLP, and geopotential height; and the pressure level at which an analysis is performed. Using vorticity, a cyclonic center can be identified much earlier than a closed pressure or geopotential height center, also, if a 500-mb prognostic field is used (e.g., Bell and Bosart 1989) a cyclonic system will only appear after significant development has taken place compared with a lower level—for example, 850 mb. Differences between model-based studies and observational studies may also be the result of deficiencies in the quality and quantity of the observational data. In many of the earlier observational studies only daily data was used and data coverage may have been of poor quality in many geographical

regions, for example, the oceans and in particular the SH oceans.

But, perhaps the biggest reason for differences between the results of analyzing different models and observational data are differences between model resolutions and model physical parameterizations. A difference in model resolution may be the reason for the difference in cyclone activity over the Great Lakes and Hudson Bay, in Lambert's NH cyclone distribution and the NH feature density presented here. Increasing the model resolution will also result in the introduction of smaller-scale features with the possibility of increasing the number of secondary systems. Thus, the sensitivity of these types of studies to model resolution needs to be explored further using a model with consistent physical parameterizations.

The physical parameterizations can have a significant impact on the model behavior of cyclonic systems. For example, the convective parameterization and its effect on the tropical circulation can have a significant impact on the extratropics via large-scale teleconnections and also via shorter-scale activity in the form of the poleward migration of tropical cyclones into the storm tracks. Early versions of the ECMWF model used the Kuo convective scheme, while the model integration used here used the Betts–Miller scheme. The Betts–Miller scheme has been shown to be an improvement on the Kuo scheme (Slingo et al. 1994), with respect to the transient characteristics. Using the Kuo scheme results in weak, incoherent disturbances with little poleward migration of tropical systems into the extratropics, while using the Betts–Miller scheme has been shown to result in the poleward migration of tropical systems into the storm track, at least in the western Pacific for a NH winter (Slingo et al. 1994; Hodges 1995). Also, of relevance to the convective parameterization is the possibility of a response in the extratropics to intraseasonal variations in the warm pool of the western Pacific, with a resulting impact on blocking events and the passage of migratory cyclones. The improved radiation scheme in the UGAMP model may also lead to differences due to an improvement in the energetics of midlatitude baroclinic systems (Morcrette 1990). Changes to the vertical diffusion scheme implemented in the ECMWF model in 1988 and the UGAMP model may also affect the blocking frequency for the NH winter (Palmer et al. 1990).

Overall, the comparison with other studies has shown that, at least for the NH and SH winters, GCM models are beginning to produce consistent results for the extratropical climatology, which compares favorably with observation. Thus, modern GCM's with improved parameterizations are beginning to simulate the NH and SH climates reasonably well. While some of the differences between these studies can be attributed to the methodologies used, it is more likely that many of the differences between model-based studies and observation that occur will be the result of different model resolution and parameterization and the quality and quantity of the observational data. Thus, further studies are required to explore the effects on the climatology of changing the resolution at which models are run and changing and improving model parameterizations. Further studies are also required using the large amounts of observational data now available from both geostationary and polar-orbiting satellites. It is only when consistent techniques are used to construct statistics from these diverse data sources that quantitative validation of the models can be carried out.

REFERENCES

Akyildiz, V., 1984: Systematic errors in the behavior of cyclones in the ECMWF operational models. *Tellus,* **37A,** 297–308.

Ballenzweig, E. M., 1959: A practical equal-area grid. *J. Geophys. Res.,* **64,** 647–651.

Bell, G. D., and L. F. Bosart, 1989: A 15-year climatology of Northern Hemisphere 500 mb closed cyclone and anticyclone centres. *Mon. Wea. Rev.,* **117,** 2142–2163.

Betts, A. K., and M. J. Miller, 1993: The Betts–Miller scheme. *The Representation of Cumulus Convection in Numerical Models of the Atmosphere, Meteor. Monogr.,* No. 46, K. A. Emanuel and D. J. Raymond, Eds., Amer. Meteor. Soc., 107–122.

Boyle, J. S., and T. J. Chen, 1992: Synoptic aspects of the winter East Asia monsoon. *Monsoon Meteorology, Oxford Monogr. on Geology and Geophysics,* No. 7, Oxford University Press, 125–160.

Changnon, D., J. J. Noel, and L. H. Maze, 1995: Determining cyclone frequencies using equal-area circles. *Mon. Wea. Rev.,* **123,** 2285–2294.

Chow, Y. S., S. Geman, and L. D. Wu, 1982: Consistent cross-validated density estimation. *Ann. Stat.,* **11,** 25–38.

Chung, Y., K. D. Hage, and E. R. Reinelt, 1976: On lee cyclogenesis and airflow in the Canadian Rocky Mountains and the east Asian mountains. *Mon. Wea. Rev.,* **104,** 879–891.

Diggle, P. J., and N. I. Fisher, 1985: Sphere: A contouring program for spherical data. *Comput. Geosci.,* **11,** 725–766.

Hall, P., G. S. Watson, and J. Cabrera, 1987: Kernel density estimation with spherical data. *Biometrika,* **74,** 751–762.

Härdle, W., 1990: *Applied Nonparametric Regression.* Cambridge University Press, 333 pp.

Hayden, B. P., 1981: Cyclone occurance mapping: Equal area or raw frequencies? *Mon. Wea. Rev.,* **109,** 168–172.

Hodges, K. I., 1994: A general method for tracking analysis and its application to meteorological data. *Mon. Wea. Rev.,* **122,** 2573–2586.

——, 1995: Feature tracking on the unit sphere. *Mon. Wea. Rev.,* **123,** 3458–3465.

Jones, D. A., and I. Simmonds, 1994: A climatology of southern hemisphere anticyclones. *Climate Dyn.,* **10,** 333–348.

Jones, M. C., 1990: Variable kernal density estimates and variable kernal density estimates. *Aust. J. Stat.,* **32,** 361–371.

Kelsey, K., 1925: A new method of charting storm frequency. *Mon. Wea. Rev.,* **53,** 251–252.

Kent, J. T., 1982: The Fisher–Bingham distribution on the sphere. *J. Roy. Stat. Soc. B,* **44,** 71–80.

Klein, W. H., 1957: Principle tracks and mean frequencies of cyclones and anticyclones in the northern hemisphere. Res. Paper No. 40, U.S. Weather Bureau, Washington, DC, 60 pp.

König, W., R. Sausen, and F. Sielmann, 1993: Objective identification of cyclones in GCM simulations. *J. Climate,* **6,** 2217–2231.

Kuo, H. L., 1974: Further studies of the parameterization of the influence of cumulus convection of the large scale flow. *J. Atmos. Sci.,* **31,** 1232–1240.

Lambert, S. J., 1988: A cyclone climatology of the Canadian Climate Centre general circulation model. *J. Climate,* **1,** 109–115.

Le Treut, H., and E. Kalnay, 1990: Comparison of observed and simulated frequency distribution as determined by an objective method. *Atmósfera,* **3,** 57–71.

Mechoso, C. R., 1980: The atmospheric circulation around Antartica: Linear stability and finite amplitude interactions with migrating cyclones. *J. Atmos. Sci.,* **37,** 2209–2233.

Morcrette, J. J., 1990: Impact of changes to the radiation transfer parameterizations plus cloud optical properties in the ECMWF model. *Mon. Wea. Rev.,* **118,** 847–873.

Murray, R. J., and I. Simmonds, 1991a: A numerical scheme for tracking cyclone centres from digital data. Part I: Development and operation of the scheme. *Aust. Meteor. Mag.,* **39,** 155–166.

——, and ——, 1991b: A numerical scheme for tracking cyclone centres from digital data. Part II: Application to January and July general circulation model simulations. *Aust. Meteor. Mag.,* **39,** 167–179.

——, and ——, 1995: Responses of climate and cyclones to reductions in Artic sea ice. *J. Geophys. Res.,* **100**(C3), 4791–4806.

Palmer, T. N., C. Branković, F. Molteni, and S. Tibaldi, 1990: Extended-range predictions with ECMWF models: Interannual variability in operational model integrations. *Quart. J. Roy. Meteor. Soc.,* **116,** 799–834.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1988: *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 735 pp.

Scott, D. W., 1992: *Multivariate Density Estimation Theory, Practice, and Visualization.* John Wiley & Sons, 317 pp.

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 179 pp.

Sinclair, M. R., 1994: An objective cyclone climatology for the Southern Hemisphere. *Mon. Wea. Rev.,* **122,** 2239–2256.

——, 1995: A climatology of cyclogenesis for the Southern Hemisphere. *Mon. Wea. Rev.,* **123,** 1601–1619.

Slingo, J., M. Blackburn, A. Betts, R. Brugge, K. I. Hodges, B. Hoskins, L. Steenman-Clark, and J. Thuburn, 1994: Mean climate and transience in the tropics of the UGAMP GCM. Part I: Sensitivity to convective parameterization. *Quart. J. Roy. Meteor. Soc.,* **120,** 881–922.

Stone, M., 1974: Cross-validatory choice and assessment of statistical predictions. *J. Roy. Stat. Soc. B,* **36,** 111–147.

Streten, N. A., and A. J. Troup, 1973: A synoptic climatology of observed cloud vorticies over the southern hemisphere. *Quart. J. Roy. Meteor. Soc.,* **99,** 56–72.

Taljaard, J. J., 1967: Development, distribution and movement of cyclones and anticyclones in the southern hemisphere during the IGY. *J. Appl. Meteor.,* **6,** 973–987.

Taylor, K. E., 1986: An analysis of the biases in traditional cyclone frequency maps. *Mon. Wea. Rev.,* **114,** 1481–1490.

Thuburn, J., 1993: Use of a flux-limited scheme for vertical advection in a GCM. *Quart. J. Roy. Meteor. Soc.,* **119,** 469–487.

Watson, G. S., 1983: *Statistics on Spheres.* John Wiley & Sons, 238 pp.

Whittaker, L. M., and L. H. Horn, 1982: *Atlas of Northern Hemisphere Extratropical Cyclone Activity; 1958–1977.* University of Wisconsin, 65 pp.

Zishka, K. M., and P. J. Smith, 1980: The climatology of cyclones and anticyclones over North America and surrounding ocean environs for January and July, 1950–77. *Mon. Wea. Rev.,* **108,** 387–401.