Optimally coherent sets in geophysical flows: A transfer-operator approach to delimiting the stratospheric polar vortex

Naratip Santitissadeekorn,[1] Gary Froyland,[1] and Adam Monahan[2]
[1]*School of Mathematics and Statistics, The University of New South Wales, Sydney, New South Wales 2052, Australia*
[2]*School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada*
(Received 27 June 2010; revised manuscript received 8 August 2010; published 12 November 2010)

The "edge" of the Antarctic polar vortex is known to behave as a barrier to the meridional (poleward) transport of ozone during the austral winter. This chemical isolation of the polar vortex from the middle and low latitudes produces an ozone minimum in the vortex region, intensifying the ozone hole relative to that which would be produced by photochemical processes alone. Observational determination of the vortex edge remains an active field of research. In this paper, we obtain objective estimates of the structure of the polar vortex by introducing a technique based on transfer operators that aims to find regions with minimal external transport. Applying this technique to European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-40 three-dimensional velocity data, we produce an improved three-dimensional estimate of the vortex location in the upper stratosphere where the vortex is most pronounced. This computational approach has wide potential application in detecting and analyzing mixing structures in a variety of atmospheric, oceanographic, and general fluid dynamical settings.

DOI: 10.1103/PhysRevE.82.056311 PACS number(s): 47.27.De, 47.10.Fg, 92.60.Bh

I. INTRODUCTION

Stratospheric ozone in the Southern Hemisphere high latitudes has decreased dramatically since the early 1970s. This long-term trend has been attributed to a combination of natural and anthropogenic factors [1–4]. In particular, it has been discovered that the ozone depletion in the lower stratosphere of the Southern Hemisphere is particularly pronounced, due in part to a strong barrier to meridional transport between middle and high latitudes during the austral winter and early spring [1]. Barriers such as these, which often coexist with turbulent mixing, play a major role in the dynamics of the stratosphere. The polar vortex is a known strong barrier to transport, enclosing a persistent, nondispersive, coherent region over the high latitudes. Our aim in this paper is to precisely determine the spatial location and movements of this coherent region, improving significantly over existing methods of estimation. Our study focuses on the upper stratosphere where the polar vortex is best developed.

It is common meteorological practice to diagnose the polar vortex edge at the position of maximum meridional gradient of potential vorticity (PV). Potential vorticity is a quantity combining measures of circulation and stratification which is materially conserved for adiabatic, inviscid flow (both of which are good approximations in stratospheric flow over time scales of a week or two). It can be shown that strong PV gradients produce a "restoring force" inhibiting meridional motion of air parcels [5]. Nevertheless, PV gradients alone provide only indirect measures of mixing barriers. In contrast, the present study characterizes regions of minimal mixing directly in terms of the transport properties of the observed stratospheric flow. We present a mathematical technique to determine the polar vortex location at different times, directly as coherent structures in observed velocity fields. Lagrangian PV-based measures of the vortex such as those presented in [6] are complicated by the fact that PV is generally a noisy field (as vorticity is the curl of the velocity field). The velocity field is generally much smoother; barriers to mixing estimated from the velocity field can be expected to be less sensitive to (poorly-observed) small-scale features of the flow.

In addition to the PV-based method, techniques based on Lyapunov exponents have been investigated. The application of the finite-time Lyapunov exponent (FTLE) field to atmospheric flows can be traced back to the work of Pierrehumbert [7] who identified and characterized the chaotic mixing in a kinematic flow model; mixing regions appeared to have large exponents whereas small or vanishing exponents indicated a region of transport barriers. The spatial variation of the FTLE field was also used to identify mixing zones and transport barriers in Northern Hemisphere atmospheric data. Mizuta and Yoden [8] used a two-dimensional (2D) barotropic model to study an idealized stratospheric polar vortex. The finite-time Lyapunov exponent field of quasiperiodic and aperiodic solutions for this model reveals a transport barrier, which coincides with the curve of the maximal PV gradient. A similar methodology based on the finite-size Lyapunov exponent (FSLE) was applied by Koh and Legras [9] to the ECMWF wind data on the 500 K isentropic surface. The results showed that the large Lyapunov exponent does not necessarily locate the vortex boundary but rather delimits a region of chaotic mixing called the "stochastic layer" in [9], whereas the vortex "edge" (the curve of the maximal PV gradient) appears as an interior envelope of this mixing region, which approximately follows the curve of the maximal PV gradient. They also compared the structure of large FSLE with that of high hyperbolicity strength, calculated based on the Lagrangian hyperbolicity proposed in Haller [10], and found that the two structures are consistent with one another. Furthermore, Nakamura [11] applied the modified Lagrangian-mean (MLM) diagnostics to identify transport barriers. The method quantifies the barriers based on the degree of transport and mixing using nitrous oxide as a tracer; the edge corresponds to the location of concentrated

tracer gradient. While geometric approaches like FTLE fields have demonstrated some correlation with transport barriers, FTLEs need not identify minimally dispersive regions. For example, Froyland and Padberg [12] have shown that regions enclosed by FTLE ridges do not always represent maximal transport barriers. Other authors (e.g., [13,14]) have noted shortcomings of the FTLE-based approach: potential ambiguity in multiple FTLE "ridges," ambiguity over flow duration for FTLE calculations, and a lack of correspondence between the strength of the ridge and the dispersal of mass across the ridge.

Our approach directly identifies minimal transport structures to high accuracy by quantifying fluid flow, and has a broad range of potential applications to geophysical fluids. For example, transport properties in other long-lived atmospheric coherent structures such as blocking highs are of interest. There is also increasing interest in the transport properties of mesoscale (on the order of 10 to 100 km in diameter) ocean eddies and their influence on biological processes within the upper, sunlit part of the water column [15,16].

II. INPUT DATA AND NONAUTONOMOUS FLOW

Our input data consists of three-dimensional velocity fields obtained from the ECMWF ERA-40 data set (http://data.ecmwf.int/data/index.html). The data are on a three-dimensional grid with 2.5° resolution in the latitude and longitude direction (144 by 73 grid points over the Southern Hemisphere). Vertical coordinates are in units of hPa, with data provided at five pressure levels (3, 5, 7, 10, and 20 hPa). We use 62 days of six-hourly velocity fields from August 1 to September 31 in 1999. The velocity fields will be interpolated linearly in space and in time; thus we can only aim to detect features at the resolution of the data provided. While we recognize that there may be biases in the reanalysis data, particularly in the upper pressure levels near the model's upper boundary, the purpose of this study is to demonstrate the ability of the transfer-operator approach to characterize coherent sets in highly unsteady flows. A more complete climatology of the polar vortex would require a more careful consideration of the data set under consideration.

Our interest is focused on the Lagrangian dynamics in the higher latitudes of Southern Hemisphere. Therefore, we will work on the phase space $X=S^1\times[-90^\circ,-30^\circ]\times D$, where S^1 is a circle parameterized from 0° to 360° and $D=[3,20]$ denotes the range of pressure in hPa. The Lagrangian motion of passive tracers is represented by their trajectories $x(t) := \Phi(x,t;\tau)$, where the flow map $\Phi: X\times\mathbb{R}\times\mathbb{R}\to X$ is a function of time τ and gives the terminal point $\Phi(x,t;\tau)$ in X of a particle initially located at $x\in X$ at time t and flowing for τ time units. The flow map $\Phi(x_0,t_0;\tau)$ is obtained as a solution of the nonautonomous ODE $\frac{dx}{dt}=f(x(t),t)$ with initial condition $x(t_0)=\Phi(x_0,t_0;0)$, where $f(x(t),t)$ in this report is the prescribed velocity data.

III. ALMOST-INVARIANT SETS, COHERENT PAIRS, AND TRANSFER OPERATORS

We shall be interested in finding a pair of sets $A_t, A_{t+\tau}$ at times t and $t+\tau$ so that $\Phi(A_{t+\tau},t+\tau;-\tau)\approx A_t$ (the pullback of $A_{t+\tau}$ from time $t+\tau$ over a duration τ is approximately equal to A_t). One could equivalently state this condition as $\Phi(A_t,t;\tau)\approx A_{t+\tau}$ (the flow forward of A_t at time t for a duration τ is approximately equal to $A_{t+\tau}$. For reasons of mathematical convention (to be consistent with the mathematical definition of measurability) we will work with the pullback approach. Moreover, this pair of sets should retain this property even when some diffusion is added to the system. Let μ be a probability measure that is preserved by the flow at all times. We call $A_t, A_{t+\tau}$ a (ρ_0,t,τ)-*coherent pair* if

$$\rho_\mu(A_t,A_{t+\tau}) := \mu[A_t\cap\Phi(A_{t+\tau},t+\tau;-\tau)]/\mu(A_t)\geq\rho_0, \quad (1)$$

and $\mu(A_t)=\mu(A_{t+\tau})$. The condition on addition of diffusion is important. Clearly, there are many (ρ_0,t,τ)-coherent pairs according to the above definition without diffusion. One may simply select any set $A_t\subset X$ and define $A_{t+\tau}=\Phi(A_t,t;\tau)$ to produce a $(1,t,\tau)$-coherent pair. In chaotic systems, such an image set $A_{t+\tau}$ is likely to be significantly less regular than A_t because of stretching and folding. We are seeking (ρ_0,t,τ)-coherent pairs with *both* sets regular. The requirement that Eq. (1) hold under diffusion acts as a selection principle, removing uncountably many irregular sets, and selecting pairs that are robust to perturbation.

To identify sets satisfying Eq. (1), we use a transfer operator $\mathcal{P}_{t,\tau}: L^1(X,m)\circlearrowleft$ defined by

$$\mathcal{P}_{t,\tau}g(x) := g[\Phi(x,t+\tau;-\tau)]\cdot|\det D\Phi(x,t+\tau;-\tau)|, \quad (2)$$

where m is the normalized Lebesgue measure on X. If $g(x)$ is a density of passive tracers at time t, $\mathcal{P}_{t,\tau}g(x)$ provides their density at time $t+\tau$ induced by the flow Φ. The term $g[\Phi(x,t+\tau;-\tau)]$ "looks back" to time t and takes on the value of the density at the τ-preimage of x. The term $|\det D\Phi(x,t+\tau;-\tau)|$ is a normalization term to ensure that $\mathcal{P}_{t,\tau}g$ is a density if g is a density.

In the autonomous setting, eigenfunctions f of $\mathcal{P}_{t,\tau}$ ($=\mathcal{P}_\tau$ for all t) corresponding to positive eigenvalues $\Lambda\approx 1$ were used to find *almost-invariant sets* [17–20]. The key point of difference between these prior studies and the present work is that the sets studied previously do not move significantly over the time duration studied, while our present work seeks *highly mobile* coherent sets, which are far from being almost-invariant. The theory and numerics we introduce in the next section are specifically designed for nonautonomous or time-dependent systems.

In practical computations, diffusion will be introduced via a finite-rank numerical approximation of the transfer operator $\mathcal{P}_{t,\tau}$. Our technique for identifying coherent pairs, described in Sec. IV, will be based on the spectral properties of these numerical approximations of $\mathcal{P}_{t,\tau}$ (based on singular vectors, rather than eigenvectors, which have been used in prior work). At the operator level, numerical diffusion winnows the spectrum of $\mathcal{P}_{t,\tau}$, reducing an uncountable spectrum to a finite set of spectral points. There is good evidence, both theoretical and numerical, to suggest that this spectral reduction process is very stable, in particular for the large spectral values which are of most interest to us. A precise description of the spectrum of $\mathcal{P}_{t,\tau}$ for general dynamical systems is still an open problem; the main technical difficulty is the development of suitable Banach spaces for the transfer operator to

act upon. For simpler dynamical systems, such expanding interval maps [21] and two-dimensional uniformly hyperbolic maps [22,23], it has been shown that large (isolated) spectral values of $\mathcal{P}$, and their corresponding spectral projectors, are stable under diffusive perturbations, including perturbations resulting from our numerical "Ulam-based" approach (see Sec. IV). In addition, there have been numerous numerical studies (e.g., [12,24,25]) that demonstrate the stability of the eigenvalues and eigenvectors of finite-rank numerical approximations of transfer operators. Thus, there is ample evidence that the selection procedure we will use for identifying coherent pairs that are robust under diffusion is itself relatively independent of the diffusion amplitude.

IV. NUMERICAL APPROACH

We partition X into a grid of boxes $\{B_1,\ldots,B_n\}$. The pressure extents of the boxes are either 3–5, 5–7, 7–10, or 10–20 hPa. Each pressure layer consists of 6605 boxes of approximately equal cross-sectional area in the latitude/longitude directions, leading to $n=26\,420$ boxes in total. To numerically approximate the transfer operator $\mathcal{P}_{t,\tau}$, we construct a finite-dimensional approximation based on Ulam's approach [26]. We introduce an L^1 projection operator that will canonically project L^1 functions onto the finite-dimensional subspace spanned by characteristic functions of the boxes: $\mathcal{B}_n := \mathrm{sp}\{\chi_{B_1},\ldots,\chi_{B_n}\}$. Define $\pi_n: L^1(X)\to\mathcal{B}_n$ by

$$\pi_n(f)=\sum_{i=1}^{n}\left(\frac{1}{m(B_i)}\int_{B_i} f dm\right)\chi_{B_i}. \tag{3}$$

The term in the parentheses [Eq. (3)] is the average value of f over the box B_i; thus $\pi_n(f)$ is the best L^1 approximation in $\mathcal{B}_n$ to f. Our matrix approximation to $\mathcal{P}_{t,\tau}$ will be the matrix representation of $\pi_n\mathcal{P}_{t,\tau}$ restricted to $\mathcal{B}_n$. We will denote this matrix representation by $\mathbf{P}^{(\tau)}(t)$. If one has $m(B_i)=1/n$ for all $i=1,\ldots,n$, then this matrix representation has the form [27],

$$\mathbf{P}^{(\tau)}(t)_{i,j}=\frac{m[B_i\cap\Phi(B_j,t+\tau;-\tau)]}{m(B_i)}, \tag{4}$$

where m is a normalized volume measure in (lat, lon, pressure) coordinates. Partitions with unequal box volumes are also easily treated [36]. The entry $\mathbf{P}^{(\tau)}(t)_{i,j}$ may be interpreted as the probability that a point selected uniformly at random in B_i at time t will be in B_j at time $t+\tau$. The matrix $\mathbf{P}^{(\tau)}(t)$ is row stochastic and thus its leading eigenvalue is 1. If μ is ergodic, with a positive density with respect to Lebesgue, then $\mathbf{P}^{(\tau)}(t)$ is irreducible (Proposition 2.3 [28]). By the Perron-Frobenius theorem (see e.g., Theorem 1.4 [29]), the eigenvalue 1 is simple, and the corresponding left eigenvector p is strictly positive. The vector p, normalized so that $\Sigma_{i=1}^{n}p_i=1$ is the invariant measure of the n-state Markov chain with transition matrix $\mathbf{P}^{(\tau)}(t)$. We will use the vector p to define a probability measure μ_n on X by $\mu_n(B')=\Sigma_{i=1}^{n}m(B'\cap B_i)p_i$ for any measurable $B'\subset X$. Note that $\mu(B_i)=p_i$ for all $i=1,\ldots,n$. The Markov chain governed by $\mathbf{P}^{(\tau)}(t)$ may be considered to be a small random perturbation of the original flow Φ (Lemma 2.2 [30]). The measure μ_n is the natural invariant measure for this slightly perturbed stochastic system.

The discretization naturally produces a diffusion at the level of box diameters. Our boxes are of approximately the same size as the ERA-40 grid resolution (the four pressure levels of our boxes extend between the four ERA-40 pressure levels considered, and on each pressure level there are 6605 boxes compared with 3504 ERA-40 gridded data points). Thus, we do not attempt to resolve structures at resolutions finer than those supported by the ERA-40 gridded data.

We estimate $\mathbf{P}^{(\tau)}(t)_{i,j}$ by

$$\mathbf{P}^{(\tau)}(t)_{i,j}\approx \#\{\ell: y_{i,\ell}\in B_i, \Phi(y_{i,\ell},t;\tau)\in B_j\}/Q, \tag{5}$$

where $y_{i,\ell}$, $\ell=1,\ldots,Q$ are uniformly distributed test points in B_i and $\Phi(y_{i,\ell},t;\tau)$ is obtained via a numerical integration. We set $Q=147$ (seven equidistant sample points in both the longitude and latitude directions and three equidistant sample points in the pressure direction) in our experiments and calculate $\Phi(y_{i,\ell},t;\tau)$ using the standard Runge-Kutta method with step size of 3/4 h. Linear interpolation is used to evaluate the velocity vector of a tracer lying between the data grid points in the longitude-latitude-pressure coordinate. In the temporal direction, the data are affinely interpolated independently in the longitude, latitude, and pressure level directions. The step size of 3/4 h is small enough to guarantee that a tracer will usually not flow to a neighboring data grid set; this limits the numerical integration error.

As before, we assume that the mass density of particles in X is at equilibrium and denote the fractional mass of particles contained in B_i by p_i, so that $\Sigma_{i=1}^{n}p_i=1$. Introduce a weighted inner product $\langle x,y\rangle_p := \Sigma_{i=1}^{n}x_iy_ip_i$. For later use, we note that one has $\langle x\mathbf{P}^{(\tau)}(t),y\rangle_p=\langle x,y\hat{\mathbf{P}}^{(\tau)}(t)\rangle_p$ for all $x,y\in\mathbb{R}^n$, where $\hat{\mathbf{P}}^{(\tau)}(t)_{i,j}=\mathbf{P}^{(\tau)}(t)_{j,i}p_j/p_i$ is the standard reverse time transition matrix from time $t+\tau$ to t (see, e.g., Sec. 6.2 [31]).

Our approach to finding a coherent pair is intuitively based upon seeking a solution to

$$\max_{w\in\{\pm 1\}^n}\frac{\langle w\mathbf{P}^{(\tau)}(t),w\mathbf{P}^{(\tau)}(t)\rangle_p}{\langle w,w\rangle_p}. \tag{6}$$

We think of $A_t := \cup_{i:w_i=1}B_i$ and $A_t^c := \cup_{i:w_i=-1}B_i$ as a coherent partition of X. The parity of the vector w determines the inclusion of boxes in the sets A_t and A_t^c. The numerator in Eq. (6) represents the size of the forward image of the vector w. If there is little transport from A_t to $\Phi(A_t^c,t;\tau)$ and from A_t^c to $\Phi(A_t,t;\tau)$ (so $A_t,\Phi(A_t,t;\tau)$ and $A_t^c,\Phi(A_t^c,t;\tau)$ are both coherent pairs), this numerator will be large. To produce nontrivial partitions (avoiding trivial partitions such as $\{X,\emptyset\}$ or $\{X\backslash A,A\}$ where A is very small in μ-measure), we may need to place lower bounds on the masses of both A_t and A_t^c.

Such a balanced bisection problem is combinatorially hard to solve. Therefore we remove the discrete condition $w\in\{\pm 1\}^n$, allowing w to float freely in $\mathbb{R}^n$. Now the elements of w represent a "fuzzy inclusion;" if w_i is very positive, then B_i is highly likely to be included in A_t, and if w_i is very negative, then B_i is highly likely to be included in A_t^c. If w_i is near to zero, then where the box B_i belongs is less certain and will be determined by an optimization procedure

described later. To effect a balancing of mass between positive and negative components of w, we must insert the condition $\langle w,v\rangle_p=0$, for some nonnegative test vector $v\in\mathbb{R}^n$. The continuous problem now reads

$$\max_{\substack{w\neq 0\\ \langle w,v\rangle_p=0}} \frac{\langle w\mathbf{P}^{(\tau)}(t),w\mathbf{P}^{(\tau)}(t)\rangle_p}{\langle w,w\rangle_p}. \tag{7}$$

Letting $D_{ij}=\delta_{ij}p_i$ and noting that $\langle w,v\rangle_p=\langle wD^{1/2},vD^{1/2}\rangle_2$, one has

$$(7)=\max_{\substack{w\neq 0\\ \langle wD^{1/2},vD^{1/2}\rangle_2=0}} \frac{\langle w\mathbf{P}^{(\tau)}(t)D^{1/2},w\mathbf{P}^{(\tau)}(t)D^{1/2}\rangle_2}{\langle wD^{1/2},wD^{1/2}\rangle_2}$$
$$=\max_{\substack{y\neq 0\\ \langle y,z\rangle_2=0}} \frac{\|yD^{-1/2}\mathbf{P}^{(\tau)}(t)D^{1/2}\|_2^2}{\|y\|_2^2},$$

where $y=wD^{1/2}$ and $z=vD^{1/2}$. From the Courant-Fischer Theorem (see, e.g., Theorem 4.2.11 [32]) we see that setting z to be the leading left singular vector of $D^{-1/2}\mathbf{P}^{(\tau)}(t)D^{1/2}$ (under multiplication on the right), the maximizing y will be the left singular vector of $D^{-1/2}\mathbf{P}^{(\tau)}(t)D^{1/2}$ corresponding to the second largest singular value. As $D^{-1/2}\mathbf{P}^{(\tau)}(t)D^{1/2}$ is nonnegative, if $\mathbf{P}^{(\tau)}(t)\mathbf{P}^{(\tau)}(t)^{\top}$ is irreducible (very likely for chaotic systems), by the Perron-Frobenius Theorem, z will be strictly positive, and thus will serve well as a candidate to enforce the balancing condition $\langle y,z\rangle_2=0$. Having obtained the maximizing y, we can now easily compute the maximizing $w=yD^{-1/2}$, which we write as $w(t)$ to emphasize it describes the fuzzy partitioning at time t. Similarly, we may obtain y', the right singular vector corresponding to the second largest singular value of $D^{-1/2}\mathbf{P}^{(\tau)}(t)D^{1/2}$, and compute $w'=y'D^{-1/2}$, which we write as $w'(t+\tau)$ to emphasize it describes the fuzzy partitioning at time $t+\tau$. We assume that $w(t),w'(t+\tau)$ are normalized so that $\langle w(t),w(t)\rangle_p=1$ and $\langle w'(t+\tau),w'(t+\tau)\rangle_p=1$.

By the properties of the singular value decomposition, one now has

(1) $w(t)\mathbf{P}^{(\tau)}(t)=sw'(t+\tau)$,

(2) $w'(t+\tau)\hat{\mathbf{P}}^{(\tau)}(t)=sw(t)$,

(3) $\langle w(t)\mathbf{P}^{(\tau)}(t),w(t)\mathbf{P}^{(\tau)}(t)\rangle_p=s^2$.

We now extract a coherent pair A_t^* and $A_{t+\tau}^*$ from a pair of vectors $w(t)$ and $w'(t+\tau)$. We create sets that are unions of boxes with w-values above certain thresholds. Define $\hat{A}_t^+(c):=\cup_{i:w_i(t)>c}B_i$ and $\hat{A}_{t+\tau}^+(c):=\cup_{i:w_i'(t+\tau)>c}B_i$, $c\in\mathbb{R}$. Denote $\mu_n[\hat{A}_t^+(c)]=\Sigma_{i:w_i(t)>c}p_i$ and $\mu_n[\hat{A}_{t+\tau}^+(c)]=\Sigma_{i:w_i'(t+\tau)>c}p_i$. Sets A_t and $A_{t+\tau}$ that are arbitrary unions of boxes B_i can be represented as $A_t=\cup_{i\in I_t}B_i$ and $A_{t+\tau}=\cup_{i\in I_{t+\tau}}B_i$, where I_t and $I_{t+\tau}$ are the corresponding sets of box indices. For such $A_t,A_{t+\tau}$, we define

$$\rho_n(A_t,A_{t+\tau})=\sum_{i\in I_t,j\in I_{t+\tau}} p_i\mathbf{P}^{(\tau)}(t)_{ij}/\sum_{i\in I_t} p_i.$$

The quantity ρ_n measures the discretized coherence for the pair $A_t,A_{t+\tau}$. Our procedure is summarized below:

(1) Let $\eta(c)=\arg\min_{c'\in\mathbb{R}}|\mu_n[\hat{A}_t^+(c)]-\mu_n[\hat{A}_{t+\tau}^+(c')]|$. As μ_n is preserved by the approximate Markov chain dynamics [Eq. (4)], the μ_n-sizes of $\hat{A}_t^+(c)$ and $\hat{A}_{t+\tau}^+(c')$ should be constrained to be equal. This necessitates possibly different thresholds c and c' for the "fuzzy inclusion" vectors $w(t)$ and $w'(t+\tau)$, respectively.

(2) Set $c^*=\arg\max_{c\in\mathbb{R}}\rho_n\{\hat{A}_t^+(c),\hat{A}_{t+\tau}^+[\eta(c)]\}$. The value of c^* is selected to maximize the coherence.

(3) Define $A_t^*:=\hat{A}_t^+(c^*)$ and $A_{t+\tau}^*:=\hat{A}_{t+\tau}^+(\eta(c^*))$, as the optimal pair of coherent sets produced by this search heuristic.

The major computational cost is the construction of $\mathbf{P}^{(\tau)}$. The calculation of large singular values and corresponding singular vectors is relatively quick, as $\mathbf{P}^{(\tau)}$ is very sparse and iterative methods for sparse matrices may be used. The construction of $\mathbf{P}^{(\tau)}$ requires numerical integration of $Q\times n$ trajectories for a flow duration of τ time units. The trajectory computations are of course highly parallizable, and further computational savings might be made by reusing already computed trajectory segments to link with new trajectories when the latter pass nearby [33].

V. NUMERICAL RESULTS

We computed the SVD of $D^{-1/2}\mathbf{P}^{(t)}(\tau)D^{1/2}$ at $t=14$ with $\tau=14$ days to obtain the left (respectively, right) singular vectors y (respectively, z) and hence $w(14)$ and $w'(28)$. Figure 1 illustrates the vectors $w(14)$ and $w'(28)$ with the components monotonically rescaled to uniformly distributed values between 0 and 1. The highlighted part of these vectors describes the most coherent pair of sets. We now threshold $w(14)$ and $w'(28)$ using the algorithm described above to extract the corresponding pair of coherent sets; see Fig. 2. We find the optimal coherence ratio is $\rho_n(A_{14},A_{28})\approx 0.7902$; this means that about 21% of the mass in A_{14} on August 14, 1999 falls outside A_{28} on August 28, 1999.

We now compare our coherent pair of sets to sets defined by contours of potential vorticity (PV). A common approach, developed in [34,35] is to define the vortex boundary as the isoline of the largest gradient of PV w.r.t. the *equivalent latitude*. We employ this approach to define potential coherent pairs at $t=14$ and $t=28$. We additionally enforce the constraint that the mass enclosed by a PV isocontour at $t=28$ is approximately equal to the mass of the set enclosed by the determined PV isocontour at $t=14$. The computational cost is relatively low, as the PV values have already been provided by the ERA-40 reanalysis.

The two-dimensional plots of PV-determined coherent pairs at $t=14$ and $t=28$ are compared with the coherent sets in Fig. 2. To estimate the transport of particles from the inside the set at $t=14$ to outside the set at $t=28$, we use a method similar to the contour crossing method introduced in [6]. The tracer particle is considered to be outside the boundary if its potential vorticity is larger than that of the boundary. Note that the contour crossing method is originally developed to estimate the transport on the 2D isentropic surface but we would like to extend its utility to estimate the transport across the boundary surface. Therefore, we interpolate the PV at the final time ($t=28$) to obtain the PV at the par-

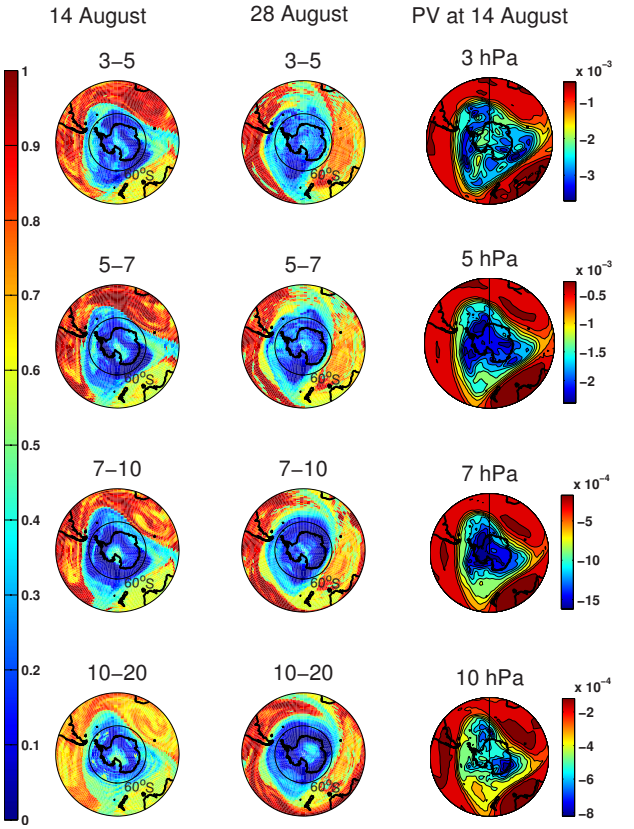

FIG. 1. (Color online) Left column: the vector $w(14)$ shown for the pressure levels 3–5, 5–7, 7–10 and 10–20 hPa. The $4\times6605=26\,420$ components of $w(14)$ have been mapped to the values $1/26\,420,2/26\,420,\ldots,1$, preserving their order. Center column: $w(28)$. Right column: potential vorticity (km^2 kg^{-1} s^{-1}) at levels 3, 5, 7, and 10 hPa on August 14, 1999 ($t=14$).

ticle's final position. We also interpolate the PV of the boundaries of the set at $t=28$ along the pressure coordinate to determine the boundary at the pressure level the advected particle resides in. This calculation shows that the fraction of particles initially inside the surface at $t=14$ remains inside the boundary surface at $t=28$ is approximately 0.7204.

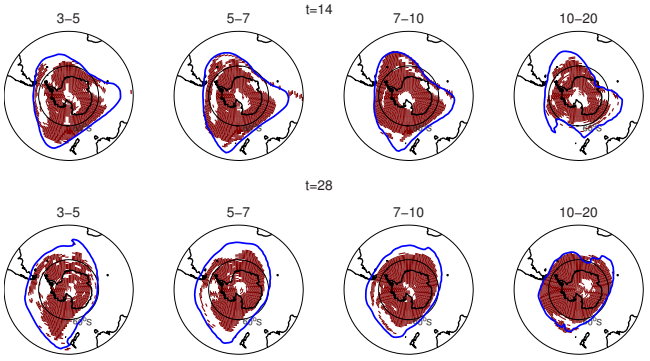

FIG. 2. (Color online) Comparison between coherent pair and PV surface boundary. The optimal coherent set A_{14} and A_{28} (shaded, in red) obtained from thresholding the vectors $w(14)$ and $w'(28)$. The coherent ratio $\rho_\mu(A_{14},A_{28})\approx0.7902$. The blue (solid curve) in each plot shows the PV surface boundary obtained from the maximum PV gradient w.r.t equivalent latitude.

The similarity between the broad outlines of the polar vortex boundaries as diagnosed by the two approaches is striking. Although the transfer-operator approach yields coherent pairs with somewhat (9.69%) greater coherence, these two distinct approaches have yielded similar structures. Note that PV is not exactly conserved for this flow (due to diabatic processes and to irreversible small-scale mixing not captured by the reanalysis fields); an appealing feature of the transfer-operator approach is that it is based strictly on the velocity field without any assumptions beyond mass conservation.

It is important to note that the finer features detected by the transfer-operator approach (relative to the PV approach)—in particular the polar hole—may reflect spurious features of the reanalysis upper-stratospheric velocity field. The analysis presented in this study represents a "proof of concept," illustrating how the transfer-operator approach can diagnose coherent sets in complex flows of geophysical relevance. In particular, we have shown that the essential features of the polar vortex can be diagnosed from two completely distinct methods: the PV approach (which is specific to rotating fluids) and the quite generic transfer-operator approach. More detailed analyses of the polar vortex will make use of more accurate and higher-resolution velocity fields (such as those from the ERA Interim Reanalysis) on isentropic surfaces.

VI. CONCLUSIONS

The Antarctic polar vortex is a well-known feature of the austral wintertime stratosphere separating polar and midlatitude air masses. The strong barrier to transport at the vortex edge plays an important role in ozone dynamics, particularly the development of the Southern Hemisphere ozone hole in austral spring. Diagnosis of the vortex edge from observations is a challenging problem that remains a subject of active research. Previous approaches to this problem have been based on kinematic (following the advection of some tracer) or dynamic (considering gradients of PV) arguments.

We presented a kinematic method of accurately estimating the three-dimensional location of the vortex. This method uses the velocity field to diagnose "optimally coherent pairs" and was able to determine a significantly more accurate estimate of transport barriers, with almost 10% less external transport from the identified vortex region than the PV-based estimate. Future, more detailed studies will include an investigation of the climatology of the polar vortex on isentropic surfaces throughout the stratosphere. Our computational approach for detecting minimal transport structures with high accuracy has a broad range of potential application to studies of transport and mixing in the atmosphere and ocean, and in general fluid dynamics settings.

ACKNOWLEDGMENTS

G.F. is partially supported by the ARC Centre of Excellence for the Mathematics and Statistics of Complex Systems (MASCOS). N.S. is supported by MASCOS. A.M. is supported by NSERC.

[1] M. E. McIntyre, J. Atmos. Terr. Phys. **51**, 29 (1989).
[2] C. R. Webster *et al.*, Science **261**, 1130 (1993).
[3] T. G. Shepherd, J. Meteorol. Soc. Jpn. **85B**, 165 (2007).
[4] T. G. Shepherd, Atmos.-Ocean. **46**, 117 (2008).
[5] G. K. Vallis, *Atmospheric and Oceanic Fluid Dynamics* (Cambridge University Press, Cambridge, England, 2006).
[6] A. H. Sobel *et al.*, J. Atmos. Sci. **54**, 2241 (1997).
[7] R. T. Pierrehumbert, Phys. Fluids A **3**, 1250 (1991).
[8] R. Mizuta and S. Yoden, J. Atmos. Sci. **58**, 2616 (2001).
[9] T. Y. Koh and B. Legras, Chaos **12**, 382 (2002).
[10] G. Haller and G. Yuan, Physica D **147**, 352 (2000).
[11] N. Nakamura and J. Ma, J. Geophys. Res., [Atmos.] **102**, 25721 (1997).
[12] G. Froyland and K. Padberg, Physica D **238**, 1507 (2009).
[13] B. Mosovsky and J. D. Meiss, e-print arXiv:1005.0566.
[14] G. Haller (unpublished).
[15] W. R. Crawford *et al.*, Prog. Oceanogr. **75**, 287 (2007).
[16] F. Nencioli *et al.*, Deep-Sea Res., Part II **55**, 1252 (2008).
[17] M. Dellnitz and O. Junge, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **36**, 491 (1999).
[18] G. Froyland and M. Dellnitz, SIAM J. Sci. Comput. (USA) **24**, 1839 (2003).
[19] G. Froyland, Physica D **200**, 205 (2005).
[20] G. Froyland, K. Padberg, M. H. England, and A. M. Treguier, Phys. Rev. Lett. **98**, 224503 (2007).
[21] G. Keller and C. Liverani, Ann. Scuola Norm. Sup. Pisa, Cl. Sci. **28**, 141 (1999).
[22] G. Froyland, Commun. Math. Phys. **189**, 237 (1997).
[23] M. Blank, G. Keller, and C. Liverani, Nonlinearity **15**, 1905 (2002).
[24] O. Junge, J. E. Marsden, and I. Mezic, *43rd IEEE Conference on Decision and Control* (IEEE, Nassau, 2004), pp. 2225–2230.
[25] P. Koltai, Ph.D. thesis, Technische Universität München, 2009).
[26] S. Ulam, *A Collection of Mathematical Problems* (Interscience Publishers, New York, 1960).
[27] T. Y. Li, J. Approx. Theory **17**, 177 (1976).
[28] G. Froyland, Ph.D. thesis, The University of Western Australia, 1997.
[29] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences* (SIAM, Philadelphia, 1994).
[30] G. Froyland, Random Comput. Dyn. **3**, 251 (1995).
[31] P. Brémaud, *Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues* (Springer, New York, 1999).
[32] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, England, 1985).
[33] S. L. Brunton and C. W. Rowley, Chaos **20**, 017503 (2010).
[34] M. E. McIntyre and T. N. Palmer, J. Atmos. Terr. Phys. **46**, 825 (1984).
[35] E. R. Nash *et al.*, J. Geophys. Res. **101**, 9471 (1996).
[36] In the general case where the volumes of the B_i are different, one replaces $m(B_i)$ in the denominator of Eq. (4) with $m(B_j)$. One can then arrive at the form [Eq. (4)] by the obvious similarity transformation.