Stratospheric Sudden Warmings as Self-Tuning Resonances. Part I: Vortex Splitting Events

N. JOSS MATTHEWMAN AND J. G. ESLER

Department of Mathematics, University College London, London, United Kingdom

(Manuscript received 17 December 2010, in final form 30 April 2011)

ABSTRACT

The fundamental dynamics of ''vortex splitting'' stratospheric sudden warmings (SSWs), which are known to be predominantly barotropic in nature, are reexamined using an idealized single-layer *f*-plane model of the polar vortex. The aim is to elucidate the conditions under which a stationary topographic forcing causes the model vortex to split, and to express the splitting condition as a function of the model parameters determining the topography and circulation.

For a specified topographic forcing profile the model behavior is governed by two nondimensional parameters: the topographic forcing height M and a surf-zone potential vorticity parameter Ω. For relatively low M, vortex splits similar to observed SSWs occur only for a narrow range of Ω values. Further, a bifurcation in parameter space is observed: a small change in Ω (or M) beyond a critical value can lead to an abrupt transition between a state with low-amplitude vortex Rossby waves and a sudden vortex split. The model behavior can be fully understood using two nonlinear analytical reductions: the Kida model of elliptical vortex motion in a uniform strain flow and a forced nonlinear oscillator equation. The abrupt transition in behavior is a feature of both reductions and corresponds to the onset of a nonlinear (self-tuning) resonance. The results add an important new aspect to the ''resonant excitation'' theory of SSWs. Under this paradigm, it is not necessary to invoke an anomalous tropospheric planetary wave source, or unusually favorable conditions for upward wave propagation, in order to explain the occurrence of SSWs.

1. Introduction

One of the defining characteristics of the stratospheric polar vortices is that they are air masses of elevated Ertel's potential vorticity (PV), relative to the background extratropical ''surf zone'' air that surrounds them. It is well established that a stratospheric sudden warming (SSW) involves a relatively rapid (on the order of a few days) change in the position and/or the structure of the stratospheric polar vortex (e.g., O'Neill 2003), and consequently during an SSW there is a large and rapid disturbance to the PV distribution in the polar region. Because of the ''invertibility'' property of PV, both the rapid increase in stratospheric polar temperatures during an SSW and the decrease in polar zonal wind speeds can be considered to be a consequence of the disturbance to the boundary of the high-PV polar vortex air mass. In other words, the evolution of the three-dimensional boundary of the polar vortex during an SSW largely determines the dynamics of the entire event.

Observed SSWs can be classified into two more or less distinct types [for recent attempts to document and classify observed SSWs see, e.g., Limpasuvan et al. (2004), Charlton and Polvani (2007), Charlton et al. (2007), and Matthewman et al. (2009)]. During a vortex displacement SSW, the polar vortex is displaced from the pole, with the magnitude of the displacement increasing rapidly with altitude (e.g., Manney et al. 1999; Matthewman et al. 2009), indicating that vortex displacement SSWs have a strongly baroclinic structure. During a vortex split SSW, by contrast, the polar vortex splits almost simultaneously with height (e.g., Manney et al. 1994; Matthewman et al. 2009); that is, a vortex split SSW has a predominantly barotropic structure. A further feature of vortex split SSWs, as Matthewman et al. (2009) have concluded from a study of the 14 observed events during 1957–2002, is that the orientation of the vortex split is approximately parallel to the 60°E–120°W meridian (i.e., almost all of the events take place at a more or less fixed

Corresponding author address: N. Joss Matthewman, Department of Mathematics, University College London, 25 Gower Street, London WC1E 6BT, United Kingdom.
E-mail: jmatt@uci.edu

DOI: 10.1175/JAS-D-11-07.1

orientation to the earth's surface). The present work addresses the nonlinear dynamics of the vortex split type of SSW. Our strategy is to investigate arguably the simplest possible model capable of capturing the fundamental fluid dynamics, as the rapid onset of SSWs implies that they can primarily be understood as fluid dynamical events. The key aim is to identify and investigate a plausible fluid dynamical mechanism that can capture the essential dynamics governing the onset of vortex splitting SSWs. The intention is that the present results can then be combined with those from vacillation-type models [following, e.g., Holton and Mass (1976) and Yoden (1987)] of varying complexity (see also, e.g., Chen 1996; Rong and Waugh 2004; Scott and Polvani 2006; Hio and Yoden 2007), which include other relevant physics such as a representation of the effects of solar and longwave radiation and which provide insight into how the SSW onset conditions are approached over a longer time scale.

Of course, this is not the first study to investigate the onset of SSWs from a purely fluid dynamical perspective. Following the suggestion of Matsuno (1971) that SSWs might be the consequence of a resonant excitation of a free Rossby wave mode of the atmosphere by topographic forcing, Tung and Lindzen (1979a,b) investigated the resonant properties of a variety of tropospheric and stratospheric flows in a linear quasigeostrophic β-plane model. They noted that many realistic wind profiles are close to a state of a linear resonance with respect to stationary forcing by the topography, and that linear resonant excitation of free Rossby wave modes is therefore a plausible mechanism for the onset of both tropospheric blocking flows and SSWs. However, as will be discussed further below, there are at least two unsatisfactory aspects to a linear resonance theory. First, observed SSWs are invariably associated with a rapid increase in vortex Rossby wave amplitudes (i.e., they have an abrupt onset). In linear theory, as resonance is approached, Rossby wave amplitudes depend smoothly on the system's controlling parameters. To explain the time scale associated with a sudden increase in wave amplitudes, linear theory requires an additional theory of why the system is suddenly brought close to resonance on that time scale, in contrast with nonlinear theories in which the rapid change can be associated with a bifurcation or catastrophe (e.g., Chao 1985). Second, linear theory cannot give insight into whether nonlinear adjustments to the system due to the amplifying waves will act to intensify or to damp the resonant excitation. Plumb (1981b) addressed the latter issue by deriving a weakly nonlinear evolution equation for the wave amplitude in a β-plane flow with arbitrary vertical shear. Plumb's results revealed that, for an initially somewhat off-resonant forcing, nonlinear effects can act to intensify resonant excitation. Plumb termed this phenomenon ''self-tuning'' resonance and one of the main aims of the present work will be to investigate its relevance in the simple vortex model of SSWs described below.

There have been a number of attempts to investigate Plumb's ideas in more realistic models. Notably, Smith (1989) investigated the possibility of resonance during the February 1979 vortex splitting SSW. A time-dependent quasi-linear calculation revealed that the phase speed of a Rossby wave normal mode approached zero as the SSW developed, hinting at the occurrence of self-tuning resonance.[1] More recently, Esler and Scott (2005) examined in detail the dynamics of a three-dimensional columnar quasigeostrophic vortex forced by a wavenumber-2 lower boundary forcing. It was shown that a vortex split occurred in the model when the frequency of the lower boundary forcing came close to resonance with that of the Rossby wave mode with a height-independent or barotropic vertical structure. The resulting columnar vortex split strongly resembles observed vortex splitting SSWs in its vertical structure (see, e.g., Matthewman et al. 2009). In addition to being an explicit demonstration of resonant forcing leading to SSW-like behavior in a three-dimensional model of the polar vortex, the Esler and Scott (2005) study emphasized that it is only the barotropic or external vertical Rossby wave mode that is plausibly linked to vortex splitting SSW behavior (a different conclusion holds for vortex displacement SSWs). Excitation of the remaining vertical modes, the Charney–Drazin spectrum, generates significant Rossby wave amplitudes only at high altitudes, where Rossby wave breaking and PV filament formation occurs (cf. Polvani and Saravanan 2000). The importance of nonlinearity in contributing to a ''frequency offset'' between the forcing frequency leading to the maximum response, and the frequency of exact linear resonance, is discussed in Esler and Scott (2005), as well as in a subsequent work (Esler et al. 2006), which is focused on the September 2002 Southern Hemisphere vortex splitting SSW. The present work aims to place the qualitative descriptions of the effects of nonlinearity on the vortex development in Esler and Scott (2005) and Esler et al. 2006) on a firmer theoretical footing, albeit in a simpler model.

The paper is structured as follows. In section 2 the model equations to be solved are described, their numerical implementation is detailed, and a set of diagnostics

[1] The present study casts doubt on the accuracy of quasi-linear wave-mean models such as that used by Smith, as it is shown explicitly in section 4d that the interaction of weakly nonlinear vortex Rossby waves with their second harmonic is equally important as their interaction with the zonal mean flow.

used to interpret the results is introduced. In section 3, the results from a set of parameter sweep experiments are presented, and a regime diagram is generated by categorizing the outcome of each experiment based on the qualitative evolution of the vortex. In section 4, the form of the regime diagram is accounted for by making two distinct nonlinear analytical reductions of the equations of motion, each formally valid in a different limit of parameter space. The results allow the construction of a "prototype vortex splitting SSW," in which a vortex undergoing a stable, relatively small (but finite) amplitude oscillation in the numerical model responds to a very small increase in forcing by undergoing a sudden vortex split. This prototype vortex splitting SSW is described in section 5. In section 6 conclusions are given.

2. A single-layer model of polar vortex evolution

The fact that observed vortex splits are planetary-scale events and are primarily barotropic in their vertical structure suggests that the basic dynamics can be captured by a quasigeostrophic, *f*-plane, single-layer shallow water model. The dynamical importance of the vortex boundary for SSWs suggests the use of a model that emphasizes this feature by representing the polar vortex as a single isolated region of elevated PV. The fixed orientation of the SSWs suggests that the tropospheric Rossby wave forcing influencing the stratosphere is primarily stationary with respect to the surface of the planet and can be represented by a stationary topographic forcing in our model.

The above considerations motivate the study of the following system, modeling the physical situation of a rapidly rotating shallow water layer of mean depth H, under gravity g, and with rotation rate $f_0/2$, where f_0 is the Coriolis parameter evaluated at the pole. An isolated vortex with uniform PV $f_0 + Q_i$ and area πR^2 is assumed to evolve in an infinite background with uniform PV $f_0 + Q_0$. The vortex is subject to forcing from a stationary topography of height h. The time evolution of the PV q is then determined by

$$(\partial_t + \mathbf{u}\cdot\nabla)q = 0,$$
$$q = f_0 + \nabla^2\psi - L_R^{-2}\psi + \frac{f_0}{H}h, \qquad (1)$$

where ψ is a streamfunction determining the velocity $\mathbf{u} = -\nabla\times\psi\mathbf{k}$ (where $\mathbf{k}$ is a unit vector along the axis of rotation) and $L_R = \sqrt{gH}/f_0$ is the Rossby radius. For analytical convenience the topography is taken to have the form

$$h(r,\theta) = h_m J_k(lr)\cos k\theta, \qquad (2)$$

where (r, θ) are the usual polar coordinates, $J_k(\cdot)$ is a Bessel function of the first kind, k is an integer azimuthal wavenumber, l is a radial wavenumber, and h_m is a parameter determining the height. The set (1) has been studied by Polvani and Plumb (1992), who focused on small-scale Rossby wave breaking and filament formation. Here, however, the focus will be on the possibility of vortex splitting.

It is convenient at this stage to nondimensionalize the above system as follows. The vortex radius R is chosen as the horizontal length scale, $(Q_i - Q_o)^{-1}$ is the time scale, $R(Q_i - Q_o)$ is the velocity scale, and $(Q_i - Q_o)$ is the PV scale. Waugh and Dritschel (1999) suggest a value to fit observations of $(Q_i - Q_o) = 0.4f_0$, for which one nondimensional time unit ≈ 0.2 days. Working with the nondimensional PV anomaly [i.e., $q \to (q - f_0)/(Q_i - Q_o)$] and elsewhere substituting for all variables and operators with their nondimensional counterparts, the rescaled equations of motion are

$$(\partial_t + \mathbf{u}\cdot\nabla)q = 0,$$
$$q(\mathbf{x}) = \nabla^2\psi - \mathcal{B}^2\psi + MJ_k(\lambda r)\cos k\theta = \begin{cases} 1 + 2\Omega & \mathbf{x}\in\mathcal{D} \\ 2\Omega & \mathbf{x}\notin\mathcal{D}. \end{cases} \qquad (3)$$

Here $\mathcal{D}$ denotes the region, of nondimensional area π, occupied by the vortex. The forcing azimuthal wavenumber k and the four nondimensional parameters

$$M = \frac{h_m}{H}\frac{f_0}{Q_i - Q_o}, \quad \lambda = lR, \quad \mathcal{B} = \frac{R}{L_R}, \quad \text{and}$$
$$\Omega = \frac{Q_o}{2(Q_i - Q_o)} \qquad (4)$$

appear explicitly in the nondimensional system and, together with the initial conditions, fully determine the vortex evolution. They can be interpreted as follows: M is a measure of the amplitude of the topographic forcing, λ measures the ratio of the horizontal scale of the vortex to that of the topography, $\mathcal{B}$ is the ratio of the vortex scale to the Rossby radius (alternatively, the square root of the inverse Burger number of the vortex), and the surf-zone PV parameter Ω controls the strength of the stratospheric jet, relative to the fixed topography. The parameter Ω will be important in the study below, as it allows control over the initial stratospheric "climate" in the simple model (3). For example, different values of Ω might be regarded as being more appropriate to different stages of the seasonal cycle, or alternatively Ω might be regarded as evolving slowly under the control of the changing solar and longwave radiation fields.

For brevity all of the numerical calculations and analysis below will be for a vortex that is initially circular; that is, at $t = 0$,

$$q(\mathbf{x}) = \begin{cases} 1 + 2\Omega & |\mathbf{x}| < 1 \\ 2\Omega & |\mathbf{x}| > 1 \end{cases}. \tag{5}$$

The azimuthal velocity field due to the background PV (nonzero Ω) is easily shown (e.g., Polvani and Plumb 1992) to be

$$U_b(r) = 2\Omega I_1(\mathcal{B}r)/\mathcal{B}, \tag{6}$$

where I_1 is a modified Bessel function of the first kind. The parameter Ω clearly acts to control the magnitude of the azimuthal velocity (i.e., the strength of the stratospheric jet).

We note in passing that previous analyses of the model equations given below in section 4 (Kida 1981; Dritschel 1990) have demonstrated that initializing with other vortex shapes (e.g., elliptical) may lead to a richer variety of outcomes. Investigation of this possibility is postponed to a future study.

a. Numerical implementation

Calculations of the fully nonlinear dynamical evolution of the vortex, under the influence of topographic forcing, are performed using the contour dynamics with surgery algorithm (CD), (Dritschel 1988). In the CD algorithm for the system (3), the PV field q is represented by j distinct regions of constant PV. The velocity at a location $\mathbf{x}$ in the domain is then given by

$$\mathbf{u}(\mathbf{x}) = \frac{1}{2\pi}\sum_j \Delta_j \oint_{\partial\mathcal{D}_j} K_0(\mathcal{B}|\mathbf{x} - \mathbf{x}_j|)\, d\mathbf{x}_j + \mathbf{u}_M(\mathbf{x}),$$

where Δ_j is the PV jump across the bounding contour of the jth region of PV $\partial\mathcal{D}_j$, $\mathbf{x}_j$ are the node locations on $\partial\mathcal{D}_j$, and $\mathbf{u}_M(\mathbf{x})$ is the velocity at $\mathbf{x}$ due to the topographic forcing. The derivation of the expression for $\mathbf{u}(\mathbf{x})$ is discussed in the appendix [see Eq. (A2)].

In the present implementation of the CD algorithm the time step is 0.05 nondimensional time units. The resolution parameter between vortex boundary nodes is chosen to be 0.025, corresponding to an initial 304 nodes on the circular vortex boundary. As the integration proceeds, contour surgery adds intervening nodes when adjacent nodes become too far apart and removes adjacent nodes that become closer than 1.6×10^{-4} vortex radii, corresponding to a dimensional horizontal resolution (for vortex filaments) of approximately 350 m.

b. Vortex diagnostics

To make an objective assessment of the outcome of the numerical experiments, it is helpful to make use of "moment diagnostics" for the vortex (e.g., Melander et al. 1986). Moment diagnostics allow objective statements about the aspect ratio, orientation, and degree of splitting of the vortex during each model run. Since a detailed application of the technique to observations is given in Matthewman et al. (2009), we give only the key definitions below.

The basis for the moment diagnostics are J_{mn}, the centralized moments of PV, which are calculated relative to the vortex centroid $\bar{\mathbf{x}} = (\bar{x}, \bar{y})^{\mathrm{T}}$ using

$$J_{mn} = \int q(\mathbf{x})(x - \bar{x})^m (y - \bar{y})^n\, d\mathbf{x}, \quad \bar{\mathbf{x}} = \frac{\int \mathbf{x} q(\mathbf{x})\, d\mathbf{x}}{\int q(\mathbf{x})\, d\mathbf{x}},$$

where the integrals are over the full domain $\mathbb{R}^2$. In practice, for the case of contour dynamics, J_{mn} can be evaluated by transforming to a sum of integrals around the bounding contours of each patch of PV using Green's theorem in the plane. The orientation $\phi_e(t)$ and aspect ratio $r_e(t)$ of the equivalent ellipse (i.e., the ellipse that is a best fit to the PV distribution of the vortex) are then given by

$$\phi_e(t) = \frac{1}{2}\tan^{-1}\left(\frac{2J_{11}}{J_{20} - J_{02}}\right), \tag{7}$$

$$r_e(t) = \left|\frac{(J_{20} + J_{02}) + \sqrt{4J_{11}^2 + (J_{20} - J_{02})^2}}{(J_{20} + J_{02}) - \sqrt{4J_{11}^2 + (J_{20} - J_{02})^2}}\right|^{1/2}. \tag{8}$$

The statistics of the orientation and aspect ratio of the equivalent ellipse can easily be used in calculations of the "elliptic diagnostics" of Waugh (1997). To diagnose vortex splitting we introduce the "excess kurtosis" here:

$$\mathcal{K}(t) = \pi\frac{J_{40} + 2J_{22} + J_{04}}{(J_{20} + J_{02})^2} - \frac{2}{3}\left[\frac{3r_e^4 + 2r_e^2 + 3}{(r_e^2 + 1)^2}\right]. \tag{9}$$

By construction, $\mathcal{K}(t) = 0$ for an elliptical vortex of uniform PV with any aspect ratio; hence, $\mathcal{K}(t)$ measures the extent of deviation of the vortex from a uniform ellipse. Filament formation leads to positive values of $\mathcal{K}(t)$, and "pinching" (leading to vortex splitting) results in negative values. Therefore, vortex splits can be diagnosed by checking whether $\mathcal{K}(t)$ attains values less than a well-chosen threshold value (see discussion of Figs. 1 and 2 below). Finally, to facilitate direct comparison

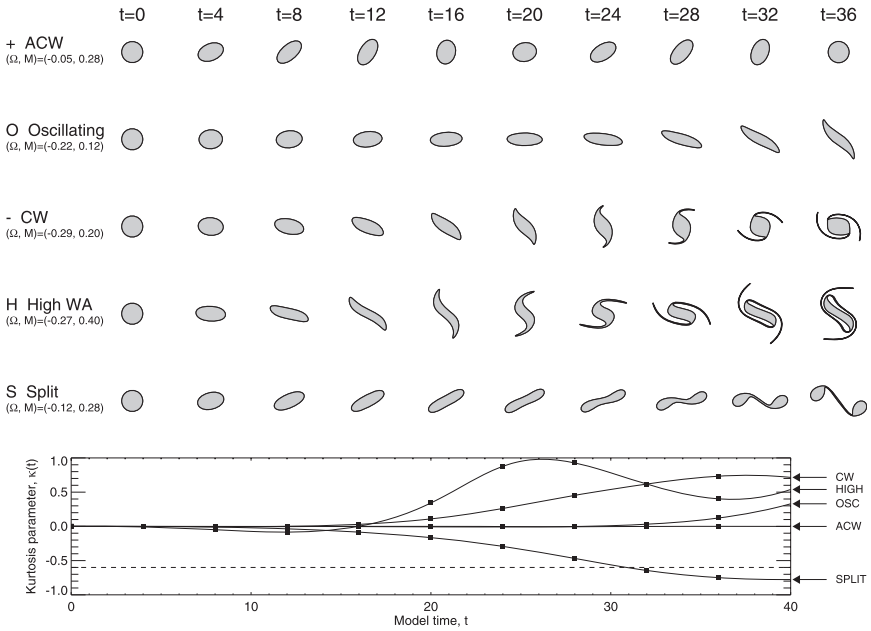

FIG. 1. Snapshots of vortex evolution during experiments representative of each of the five categories ACW, OSC, CW, HIGH, and SPLIT shown in Table 1. Parameter $\mathcal{B} = 0$ is used for all experiments, with Ω and M given in each panel. (bottom) The evolution of $\mathcal{K}(t)$ as a function of time, with box symbols marking times at which snapshots are shown. The dashed line marks the threshold $\mathcal{K}(t) = -0.6$ at which vortex splitting is diagnosed.

with weakly nonlinear theory, an equivalent ''wave amplitude'' diagnostic $a_e(t)$ is defined as

$$a_e(t) = \frac{1}{\sqrt{2}}\left[\frac{2}{\pi}(J_{20} + J_{02}) - 1\right]^{1/2}. \qquad (10)$$

The wave amplitude diagnostic can be recognized as being derived from the angular pseudomomentum (or the excess angular impulse relative to the circular vortex), which is an invariant for the unforced system and is quadratic in wave amplitude (Dritschel and Saravanan 1994). The quantity $a_e(t)$ has been normalized so that, as well as being an invariant of the full nonlinear system in the absence of forcing, it is also equal to the wave amplitude in the linear limit.

3. Nonlinear evolution of the vortex

Two sets of parameter sweeps have been performed to explore the full range of behavior of the nonlinear model. In contrast to Polvani and Plumb (1992), who investigated Rossby wave microbreaking with $k = 1$, here we examine vortex behavior using forcing with $k = 2$. In each case a value of the radial wavenumber parameter $\lambda = 1.162$ is chosen to ensure that the topography has a horizontal scale that is significantly larger than the vortex (the first zero of the J_2 Bessel function is then located at 4.4 vortex radii). Tests reveal that the qualitative behavior is insensitive to the value of λ for $\lambda \lesssim 3$. The two parameter sweeps are distinguished by the value of $\mathcal{B}$. For the first set $\mathcal{B} = 0$, corresponding to the well-known barotropic vorticity equation limit $L_R \to \infty$. In the second set $\mathcal{B} = 1$, a value corresponding to R and L_R being equal (a typical vortex radius for the Northern Hemisphere in midwinter is $R = 2200$ km). A value of $L_R \approx R$ is obtained if it is taken that the relevant dynamical Rossby radius, appropriate if the shallow water model is to describe the dynamics of ''tall'' vortical structures in a stratified atmosphere, is that of the ''external'' Lamb mode. In this case $L_R = NH/f_0\sqrt{\kappa(1-\kappa)}$ for an isothermal compressible stratosphere. Taking the scale height $H = 6.14$ km, Brunt–Väisälä buoyancy frequency $N = 0.022\ \mathrm{s}^{-1}$, Coriolis parameter $f_0 = 4\pi/86\,400\ \mathrm{s}^{-1}$, and $\kappa = 2/7$ yields $L_R \approx 2100$ km $\approx R$ in the Northern Hemisphere.

The remaining parameters M and Ω are varied for each parameter sweep, with $M \in [0.01, 0.48]$ and $\Omega \in [-0.30, 0]$ for experiments with $\mathcal{B} = 0$ and $M \in [0.03, 0.49]$ and $\Omega \in [-0.16, 0.07]$ for experiments with $\mathcal{B} = 1$. The total number of numerical experiments for each parameter sweep is 496 for $\mathcal{B} = 0$ and 360 for $\mathcal{B} = 1$. For each individual model experiment, Eq. (3) is integrated for 40 nondimensional time units [scaled on $(Q_i - Q_o)^{-1}$] for $\mathcal{B} = 0$ and 50 nondimensional time units for $\mathcal{B} = 1$.

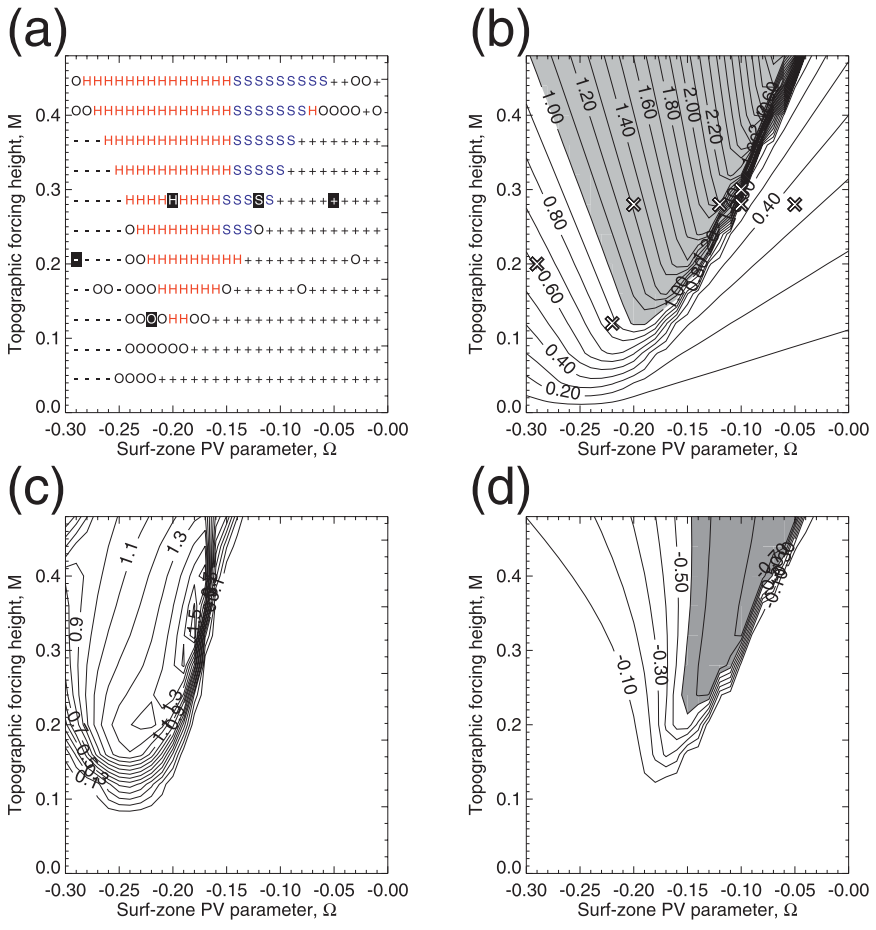

FIG. 2. (a) Classification of vortex behavior as a function of Ω and M in the fully nonlinear numerical model experiments with $\mathcal{B} = 0$. Experiments are classified using the criteria outlined in Table 1 and box shading marks the five experiments shown in Fig. 1. (b) Contours of a_m as a function of Ω and M for the $\mathcal{B} = 0$ sweep in the fully nonlinear model. The contour interval is 0.1 and values > 1 are shaded. Crosses mark experiments shown in Fig. 1 and the parameter values used in the prototype SSW experiment discussed in section 5. (c) As in (a), but with contours of $\mathcal{K}_{\max}$. (d) As in (a), but with contours of $\mathcal{K}_{\min}$. In both (c) and (d), the contour interval is 0.1 and shaded parameter space denotes vortex splitting experiments in which $\mathcal{K}_{\min} < -0.6$. The reader is reminded that $\mathcal{K}_{\max} \geq 0$ and $\mathcal{K}_{\min} \leq 0$ by construction.

For the $\mathcal{B} = 0$ experiments, model runs were classified into five categories as shown in Table 1. In the table, $\dot{\phi}_e$ denotes the time derivative of ϕ_e [Eq. (7)] of the vortex with respect to time. The SPLIT criterion has been carefully chosen so that all of the model runs meeting the criterion clearly result in the vortex splitting into two, whereas those not meeting the criterion do not.

Figure 1 presents snapshots illustrating the typical evolution of the vortex for each of the five categories defined above. The results are taken from the barotropic ($\mathcal{B} = 0$) parameter sweep, although similar qualitative behaviors were found for the $\mathcal{B} = 1$ set. For experiments in the anticlockwise (ACW) regime, an example of which is shown in the row labeled ACW of Fig. 1, the vortex remains almost elliptical throughout its evolution, with the aspect ratio and orientation of the ellipse evolving periodically. The period of a revolution is approximately 18 time units for the experiment shown.

An example of typical vortex behavior for experiments lying in the oscillating regime is shown in the row labeled O. Although the vortex remains almost elliptical

TABLE 1. Model experiment classification criteria.

+	ACW	$\dot{\phi}_e > 0$ throughout the motion
O	OSC	All other vortex behavior
−	CW	$\dot{\phi}_e < 0$ throughout the motion
H	HIGH	$a_m = \max[a_e(t)] > 1$
S	SPLIT	$\mathrm{Min}[\mathcal{K}(t)] < -0.6$

throughout its evolution, instability of disturbances on the vortex edge eventually leads to vortex roll up at the tips of the major axes (see, e.g., Saffman 1992), leading to filaments of PV becoming advected away from the core of the vortex. Despite the appearance of filaments, analysis of the vortex evolution over a longer period reveals a near-periodic oscillation, during which the time derivative of $\phi_e(t)$ changes sign repeatedly, indicating both clockwise and anticlockwise rotation. The period of oscillation is approximately 52.5 time units for the experiment shown (i.e., almost 3 times longer than the period of the ACW experiment). For fixed forcing amplitude M, the maximum aspect ratio throughout the oscillating regime is found to be greater than in the ACW regime. For experiments in the clockwise (CW) regime the vortex behavior is very similar to that seen in the oscillating regime, but with $\dot{\phi}_e < 0$ at all times. The rotation period for the example in row CW of Fig. 1 is approximately 34 time units. We note that, like the oscillating (OSC) experiment shown above, this period is also greater than that of the ACW experiment. In fact, for any given M, the period of rotation or oscillation is generally longer for the CW and OSC experiments when compared to their ACW counterparts, and in all experiments the period increases the closer to the OSC/ACW regime boundary one gets.

The behavior of experiments lying in the highly disturbed regime, an example of which is seen in row H, is similar to that of the oscillating and clockwise rotation regimes, but with larger maximum aspect ratio. The large aspect ratio combined with instability of the vortex structure leads to greatly increased filamentation of PV.

At early times in vortex splitting experiments (e.g., row S) the vortex is near elliptical with the aspect ratio increasing monotonically. For the experiment shown, which is characteristic of experiments located at the right-hand boundary of the vortex splitting regime (see Fig. 2), there is very little change in the orientation of the ellipse as its aspect ratio increases. By $t = 28$ the growth of unstable disturbances on the vortex edge leads to a pinching of the vortex close to its centroid. This pinching initiates a vortex split that occurs at $t \approx 30$. The two daughter vortices resulting from the split have the same area and rotate clockwise around their common centroid for $t > 36$.

Evolution of the vortex splitting diagnostic $\mathcal{K}(t)$ is shown for each experiment at the bottom of Fig. 1. Large negative values of $\mathcal{K}(t)$ are only observed in the splitting experiment, indicating that vortex splitting behavior can be distinguished from all other types of behavior using a criterion based on the $\mathcal{K}(t)$ diagnostic (see also Figs. 2c,d). A detailed comparison over all experiments revealed that $\min[\mathcal{K}(t)] < -0.6$ accurately identified all vortex splitting evolutions but was never satisfied for nonsplitting evolutions.

Figure 2a is a (M, Ω) regime diagram for the $\mathcal{B} = 0$ experiments, with each classification of outcome labeled by the symbols given above. The location of these regimes in parameter space will be discussed further below. Briefly, however, it is noticeable that vortex splits occur only for M larger than a threshold value $M_c = M \approx 0.24$, and that for $M > M_c$ vortex splits occur for only a relatively narrow range of Ω, with the extent of this range increasing with M.

Figure 2b shows the maximum wave amplitude $a_m = \max[a_e(t)]$ attained in the numerical experiments when $\mathcal{B} = 0$, plotted as a function of M and Ω. Regions of parameter space corresponding to ''high'' wave activity ($a_m > 1$, HIGH) are shaded. At low forcing amplitude (small M) the peak response is seen to occur at values of Ω close to that corresponding to resonant excitation of the free vortex Rossby waves by the topographic forcing (linear resonant excitation occurs at a specific value $\Omega = \Omega_0(k, \mathcal{B})$ to be derived in section 4b below). As M increases, the location of the peak in a_m migrates to larger values of Ω, away from linear resonance. Comparison with Fig. 2a reveals that the location of the peak response (for fixed M) in parameter space remains near the regime boundary between the ACW regime and SPLIT/HIGH regimes.

Figures 2c and 2d show the maximum and minimum values of the excess kurtosis diagnostic $\mathcal{K}(t)$—that is, $\mathcal{K}_{\max} = \max[\mathcal{K}(t)]$ and $\mathcal{K}_{\min} = \min[\mathcal{K}(t)]$—during the barotropic ($\mathcal{B} = 0$) model runs. These figures reveal that large positive values of $\mathcal{K}_{\max}$ and large negative values of $\mathcal{K}_{\min}$ occur only in the CW, HIGH, and SPLIT regimes, with values of $\mathcal{K}(t)$ in the ACW regime being close to zero throughout the vortex evolution. In other words, throughout the ACW regime, the vortex remains almost elliptical as it rotates.

4. Analytical approximations to the fully nonlinear behavior

A route to understanding and interpreting the numerical results above is to formulate analytical approximations to the model Eqs. (3). The objective is to obtain a set of approximate equations that will capture the essential ingredients of the behavior of the vortex across as wide a range of parameter space as possible. Three such analytical approximations are presented below, each formally valid in a different limit of $\{M, \Omega, \mathcal{B}, \lambda, k\}$ parameter space. These are

(i) linear theory: valid for $M \ll 1$, $\Omega \neq \Omega_0$, arbitrary $k, \lambda, \mathcal{B}$;

(ii) Kida vortex theory: valid for $\lambda \ll 1$, $k = 2$, $\mathcal{B} = 0$, arbitrary M, Ω; and
(iii) weakly nonlinear theory: valid for $M \sim (\Omega - \Omega_0)^{3/2} \ll 1$, arbitrary $k, \lambda, \mathcal{B}$.

Here $\Omega_0 = \Omega_0(k, \mathcal{B})$ is the value of the surf-zone PV parameter that leads to linear resonance, to be defined below. For readers who do not wish to follow the mathematical treatment in detail, a summary of the findings from each of these three theoretical frameworks, and their relation to the fully nonlinear model results derived from the experiments in section 3, is given in section 4a. The full mathematical analysis then follows in sections 4b–d.

The following analytical device is used as a first step in each approach. The linearity of PV inversion can be used to show that the system (3) can be replaced by the exactly equivalent system (cf. Swanson 2000)

$$\begin{aligned}
&[\partial_t + (\mathbf{u}_R + \mathbf{u}_M)\cdot\nabla]q = 0,\\
&q(\mathbf{x}) = \nabla^2\psi_R - \mathcal{B}^2\psi_R = \begin{cases} 1 + 2\Omega & \mathbf{x} \in \mathcal{D} \\ 2\Omega & \mathbf{x} \notin \mathcal{D}, \end{cases}\\
&\mathbf{u}_R = -\nabla\times\psi_R\mathbf{k},\\
&\mathbf{u}_M = -\nabla\times\psi_M\mathbf{k}, \quad \text{where} \quad \psi_M = M\frac{J_k(\lambda r)\cos k\theta}{\lambda^2 + \mathcal{B}^2}.
\end{aligned} \tag{11}$$

In essence, the above manipulation has replaced the topography in Eq. (3) with an exactly equivalent "topographic velocity" $\mathbf{u}_M$ in Eq. (11). The distribution of the topography has been chosen in Eq. (2) so that the topographic streamfunction ψ_M, which satisfies

$$\nabla^2\psi_M - \mathcal{B}^2\psi_M = -h,$$

takes the simple form given above.

For the case of the vortex patch, the PV advection equation in Eq. (11) can be replaced by the kinematic condition at the vortex edge. For the special case where the vortex edge location is single-valued in the azimuthal coordinate θ—that is, the edge is located at $r = 1 + \epsilon\eta(\theta, t)$, where ϵ is a parameter to be chosen to correspond to the magnitude of the disturbance—the kinematic condition may be written

$$\left[\partial_t + \frac{U_b(1 + \epsilon\eta)}{(1 + \epsilon\eta)}\partial_\theta\right]\eta = (\mathbf{u}_R\cdot\mathbf{n})_c[\eta] + (\mathbf{u}_M\cdot\mathbf{n})_c. \tag{12}$$

Here U_b is the background velocity given in Eq. (6). The vector $\mathbf{n}$ is everywhere perpendicular to the vortex edge, and in terms of the usual radial and azimuthal polar unit vectors $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\phi}}$ is given by

$$\mathbf{n}(\theta) = \hat{\mathbf{r}}(\theta) - \epsilon\frac{\eta_\theta(\theta)}{1 + \epsilon\eta(\theta)}\hat{\boldsymbol{\phi}}(\theta). \tag{13}$$

The subscript c denotes evaluation on the contour defining the vortex edge [i.e., at $r = 1 + \epsilon\eta(\theta, t)$] and the square brackets around η in Eq. (12) is intended to emphasize the functional dependence of $(\mathbf{u}_R\cdot\mathbf{n})_c$ on η. That is, the value of $(\mathbf{u}_R\cdot\mathbf{n})_c$ at a fixed value of the polar angle θ is determined by the entire instantaneous profile of $\eta(\theta, t)$, $(0 \le \theta \le 2\pi)$.

a. Summary of results

Before giving a detailed derivation of each of the three theoretical models introduced above, the main conclusions drawn from each will be summarized, with the focus on the extent to which each explains the various behaviors seen in the fully nonlinear numerical experiments of section 3. A key objective of each theoretical treatment is the derivation of an evolution equation for the vortex Rossby wave amplitude, from which theoretical predictions of $a_m = \max[a_e(t)]$ can be deduced as a function of the surf-zone PV parameter and topographic forcing amplitude. To what extent does each theory predict the various regimes of behavior discussed in section 3 (cf. Figs. 1 and 2)? The main results are summarized in Table 2 and Fig. 3. In the table, the evolution equation(s) for each model are presented, together with the means of determining the vortex Rossby wave amplitude from the solutions and the corresponding predictions for the maximum wave amplitude attained during the evolution. A summary of the qualitative behaviors found is also given. Figure 3 compares the value of the maximum vortex Rossby wave amplitude in the fully nonlinear model (Figs. 3a,b, discussed above) with predictions from linear theory (Fig. 3d), Kida vortex theory (Fig. 3c), and weakly nonlinear theory (Figs. 3e,f).

Comparing maximum wave amplitudes in Fig. 3b with the corresponding predictions of linear theory in Fig. 3d, it is clear that linear theory fails to predict much of the behavior observed in the fully nonlinear model. The linear evolution equation for the complex amplitude A^l [Eq. (17), derived in section 4 below] is that of a forced linear oscillator. When the surf-zone PV parameter takes a value Ω_0 corresponding to resonant linear excitation of the free vortex Rossby wave, the maximum wave amplitude is predicted to be infinite. Therefore, for Ω close to Ω_0, linear theory unsurprisingly overpredicts a_m. Linearity also leads to a response that is symmetric about the line $\Omega = \Omega_0$, again at odds with the asymmetric behavior seen in the fully nonlinear model. Finally, linear theory fails to predict the apparent bifurcation in a_m identified in the fully nonlinear model. That is, a_m varies smoothly with Ω and M throughout parameter space, except for the singularity at $\Omega = \Omega_0$.

TABLE 2. Summary of theoretical model results. For details of the coefficients appearing in each evolution equation, see the discussion surrounding the equation in the relevant subsection.

	Linear theory	Kida vortex theory	Weakly nonlinear theory
Regime of validity	$M \ll 1$, $\Omega \neq \Omega_0$, arbitrary k, λ, $\mathcal{B}$	$\lambda \ll 1$, $k = 2$, $\mathcal{B} = 0$ arbitrary M, Ω	$M \sim (\Omega - \Omega_0)^{3/2} \ll 1$, arbitrary k, λ, $\mathcal{B}$
Evolution equation(s)	Eq. (17): $\frac{dA^l}{dt} + i\omega_0 A^l = iF$	Eqs. (22): $\frac{dr_e}{dt} = 2\Lambda r_e \cos 2\phi_e$, $\frac{d\phi_e}{dt} = -\frac{\Lambda(r_e^2+1)}{r_e^2-1}\sin 2\phi_e + \frac{r_e}{(r_e+1)^2} + \Omega$	Eq. (34): $\frac{dA}{dT} + i\tilde{\omega}_0 A + i\tilde{\omega}_2 A\lvert A\rvert^2 = iF$
Wave amplitude $a_e(t)$ given by:	$M\lvert A^l\rvert$	$\frac{1}{2}(r_e + r_e^{-1} - 2)^{1/2}$	$M^{1/3}\lvert A\rvert$
Max amplitude $a_m = \max[a_e(t)]$	Eq. (21)	From Eq. (24)	Eq. (42)
Regimes of behavior?	No—ACW and CW only	Yes—ACW, OSC, CW, and HIGH in agreement with experiments	Yes—ACW, OSC, CW, and HIGH in agreement with experiments
Amplitude bifurcation	No	Yes—boundary given by Eq. (25)	Yes—boundary given by Eq. (43)
Notable features	Infinite response at resonance	Infinite response in extending regime	Finite response
	Symmetric response	Asymmetric response	Asymmetric response

Comparing maximum wave amplitudes in Fig. 3a with the corresponding predictions of the Kida model in Fig. 3c, it is clear that the Kida model is much more successful in predicting behavior observed in the fully nonlinear model. The Kida vortex evolution Eqs. (25) (derived in section 4) determine the evolution of aspect ratio r_e and orientation ϕ_e of an exactly elliptical vortex under the influence of an idealized forcing. The equations can be solved analytically to find the maximum vortex Rossby wave amplitude, which is qualitatively similar to that observed in the fully nonlinear model throughout much of parameter space. As will be explained in detail below, the Kida model accurately predicts the existence and location of the ACW, CW, and OSC regimes of behavior, although it cannot explicitly identify the SPLIT regime that lies within the OSC regime. The regime boundary between the ACW and OSC regimes marks a bifurcation in a_m, which is seen in Fig. 3 to lie very close in parameter space to the location of the same regime boundary in the fully nonlinear model.

There are, however, some limitations to the Kida model. First, a_m is overestimated where the response is largest, and in particular $a_m \to \infty$ is predicted in an ''extending'' regime in which the vortex becomes infinitely elongated. Second, the Kida theory is valid only for $k = 2$ topographic forcing, and then only in a purely barotropic system where $\mathcal{B} = 0$. Finally, the theory is formally valid only when the horizontal scale of the forcing is much greater than that of the vortex, with the accuracy of the theory breaking down as the scales become comparable.

Comparing maximum wave amplitudes in Figs. 3a and 3b with the corresponding predictions of weakly nonlinear theory in Figs. 3e and 3f, it is clear that the behavior of the fully nonlinear model is best predicted by weakly nonlinear theory. The weakly nonlinear evolution equation for the complex amplitude A [Eq. (34), derived in section 4d] is the generic equation of a forced nonlinear oscillator (e.g., Nayfeh and Mook 1979). Compared with corresponding linear amplitude equation in Table 2, the weakly nonlinear equation has an additional nonlinear term that acts to change the frequency of the nonlinear vortex Rossby waves as they grow and decay, thus providing a mechanism by which the system may be brought into and out of resonance (i.e., Plumb's ''self-tuning resonance''). Similar behavior is found, for example, in the analysis of (the frictionless form of) Duffing's equation. It describes the motion of a mass on a ''soft'' spring—that is, a spring that obeys a nonlinear extension/restoring force relationship (generalized Hooke's law) when the opposite end of the spring is oscillated at a frequency close to the spring's natural frequency. As will be demonstrated explicitly below, solutions of the amplitude Eq. (34) are well known to undergo an amplitude bifurcation similar to that reported above for the Kida model. Equation (34) has been derived by Plumb (1981b) for forced Rossby waves on the β plane, also with SSWs in mind. The generic properties of Eq. (34) clearly suggest it has utility as a conceptual model of SSWs with relevance beyond the single-layer model considered here.

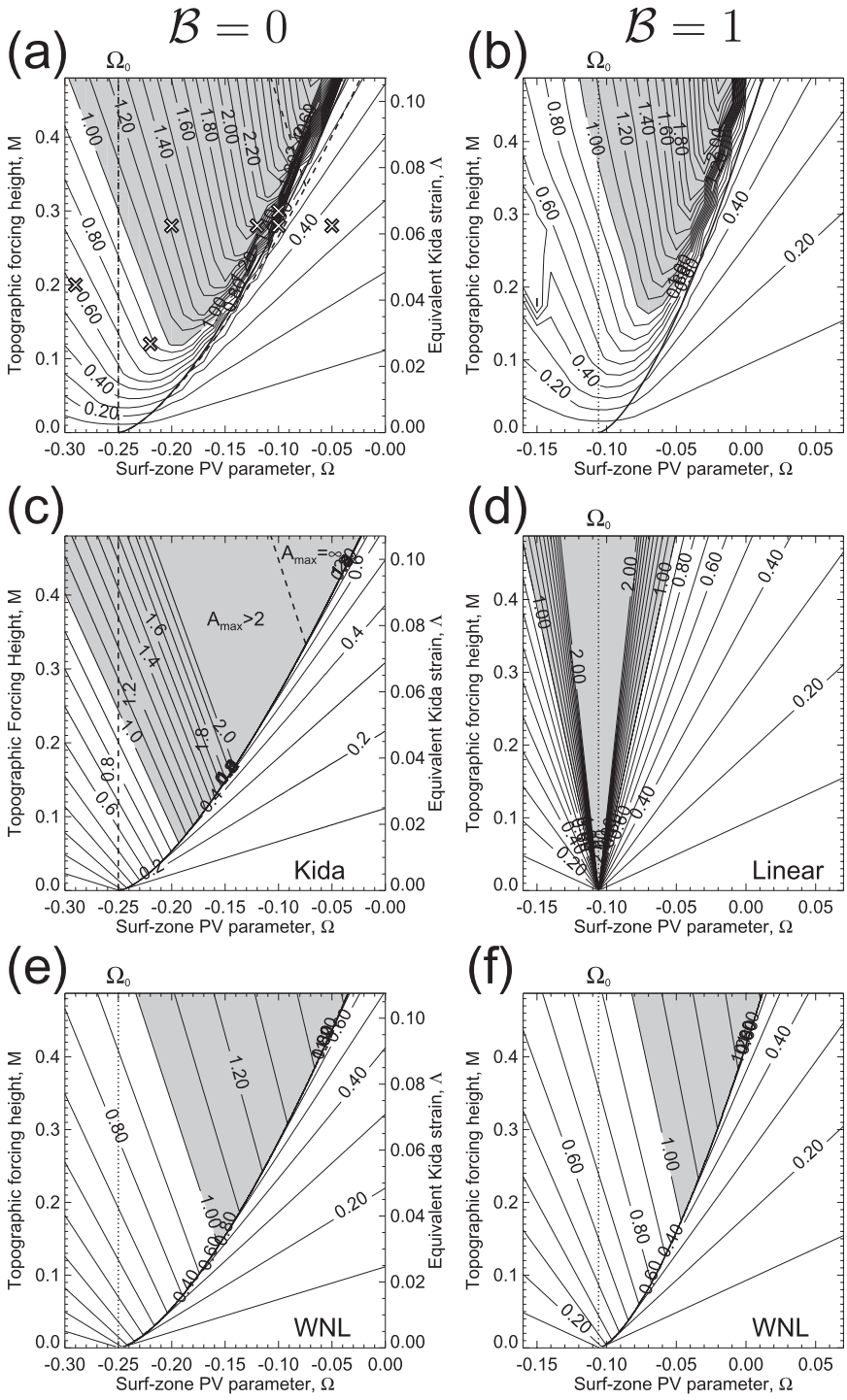

FIG. 3. Plots of a_m as a function of Ω and M for both the fully nonlinear numerical calculations [given by Eq. (11)], and in the corresponding theoretical predictions (see Table 2). (a) Fully nonlinear numerical calculations for the $k = 2$, $\mathcal{B} = 0$ set. (b) Fully nonlinear numerical calculations for the $k = 2$, $\mathcal{B} = 1$ set. (c) Kida model predictions for the $k = 2$, $\mathcal{B} = 0$ set. (d) Linear predictions for the $k = 2$, $\mathcal{B} = 1$ set. (e) Weakly nonlinear prediction for the $k = 2$, $\mathcal{B} = 0$ set. (f) Weakly nonlinear prediction for the $k = 2$, $\mathcal{B} = 1$ set. In all panels, dashed lines denote regime boundaries for the Kida vortex model, dotted lines denote $\Omega = \Omega_0$ (linear resonance), and thick solid lines denote the location of the bifurcation in a_m predicted by the weakly nonlinear theory. The contour interval in each case is 0.1 and values > 1 (i.e., the HIGH regime) are shaded. Contours > 2 are suppressed in (c) and (d). Crosses in (a) mark experiments shown in Fig. 1 and the parameter values used in the prototype SSW experiment discussed in section 5.

The strengths of weakly nonlinear theory are first that predictions of a_m are finite and are of similar amplitude to those seen in the fully nonlinear model throughout parameter space. Second, as with the Kida model, weakly nonlinear theory accurately predicts the ACW, CW, and OSC regimes of behavior seen in the fully nonlinear model. Finally, the regime boundary between the ACW and OSC regimes marks a bifurcation in the amplitude of a_m, which is qualitatively very similar to that seen in the fully nonlinear and Kida models. While strictly only valid when $M \ll 1$, weakly nonlinear theoretical predictions are surprisingly accurate when compared with the fully nonlinear model experiments for M well outside of this range.

It will be argued below that, from the perspective of vortex splitting SSWs, the most important feature of the two nonlinear analytical reductions is that they capture the bifurcation in a_m that coincides with the ACW/OSC regime boundary in both models. The bifurcation follows a path in parameter space along which nonlinear effects act to tune the vortex toward resonance most efficiently. It will be shown below, using a fully nonlinear model experiment, that a vortex split can be generated by crossing the bifurcation curve. Crossing the curve requires only a small change in the controlling parameters (either Ω or M) and can lead to an instantaneous jump from an ACW oscillation with relatively small a_m to an OSC oscillation with large a_m. In the latter case, vortex splitting becomes possible because of the development of Love (1893)-type instabilities on the elongated vortex when $a_e(t)$ is large (see also Dritschel 1986; Mitchell and Rossi 2008).

While resonance is responsible for growth of disturbances in all three idealized models, understanding the role of nonlinearity in tuning the system is essential if one is to accurately predict behavior in the fully nonlinear model. Further, the Kida and weakly nonlinear analyses exhibit a bifurcation in the maximum Rossby wave amplitudes observed during the vortex evolution, indicating a plausible mechanism to explain the rapid onset of a vortex splitting SSW.

b. Analytical reduction I: Linear theory

To appreciate the importance of nonlinearity in determining the behavior of the vortex, it is first helpful to review the linear theory of Rossby wave excitation due to the topography. The treatment here offers an alternative route to results previously given in Swanson (2000), and the linear resonance behavior found is analogous to that found by Tung and Lindzen (1979b) in their study of topographic excitation of Rossby waves in a β channel.

Linear theory is formally valid for $M \ll 1$. For the linear theory the small parameter ϵ is chosen to be M and the kinematic condition (12) is linearized by retaining only those terms that are leading order in ϵ. The resulting equation is

$$[\partial_t + U_b(1)\partial_\theta]\eta = \frac{1}{2\pi}\int_0^{2\pi} [\eta_\theta(\theta + \tilde{\psi}) - \eta_\theta(\theta)\cos\tilde{\psi}]K_0\left(\mathcal{B}\sqrt{2 - 2\cos\tilde{\psi}}\right)d\tilde{\psi} + v_M^l. \tag{14}$$

The appendix gives details of how the leading-order expression for $(\mathbf{u}_R \cdot \mathbf{n})_c$ [i.e., the integral appearing in Eq. (14)] is obtained. The linear topographic forcing term v_M^l is defined by $\epsilon v_M^l = (\mathbf{u}_M \cdot \hat{\mathbf{r}})_{r=1}$ (i.e., the azimuthal component of the topographic forcing velocity evaluated at $r = 1$) and can be evaluated for the Bessel function form given in Eq. (12) as $v_M^l = -F\sin kx$, where the topographic forcing parameter F is given by

$$F(k, \mathcal{B}, \lambda) = \frac{kJ_k(\lambda)}{\lambda^2 + \mathcal{B}^2}. \tag{15}$$

Equation (14) is a forced integro-differential equation. Wavelike solutions, with azimuthal wavenumber k equal to that of the forcing, can be sought using the ansatz

$$\eta(\theta, t) = \mathrm{Re}A^l(t)e^{ik\theta},$$

where $A^l(t)$ is a complex amplitude. The integral expression in Eq. (14) can be evaluated using the following Fourier series expansion

$$K_0\left(\mathcal{B}\sqrt{2 - 2\cos\tilde{\psi}}\right) = I_0(\mathcal{B})K_0(\mathcal{B}) + 2\sum_{m=1}^{\infty} I_m(\mathcal{B})K_m(\mathcal{B})\cos m\tilde{\psi}, \tag{16}$$

which can be obtained as a limiting case of one of the Neumann addition formulas for Bessel functions [Watson 1944, ch. 11, Eq. (8)]. The following equation for the evolution of $A^l(t)$ is obtained:

$$\frac{dA^l}{dt} + i\omega_0 A^l = iF, \tag{17}$$

where

$$\omega_0(k, \mathcal{B}) = k[U_b(1) + I_1K_1 - I_kK_k],$$
$$U_b(1) = 2\Omega I_1/\mathcal{B}, \tag{18}$$

where I_k and K_k are modified Bessel functions evaluated at $\mathcal{B}$ [i.e., $I_k \equiv I_k(\mathcal{B})$, etc.]. The frequency ω_0 has been defined so that Eq. (18) is exactly the nondimensional linear dispersion relation for the unforced system (Swanson 2000, with an additional term here due to allowing $\Omega \neq 0$). Linear resonance occurs when $\omega_0 = 0$, or equivalently when

$$\Omega = \Omega_0 \equiv \frac{\mathcal{B}(I_k K_k - I_1 K_1)}{2I_1}. \tag{19}$$

In the case where $F \neq 0$, the solution of Eq. (17) when no wave is present initially [$A^I(0) = 0$] is easily found to be

$$A^I(t) = \frac{F}{\omega_0}[1 - \exp(i\omega_0 t)]. \tag{20}$$

The above result allows the linear prediction of the maximum wave amplitude to be calculated as

$$a_m = \epsilon \max[|A^I(t)|] = \left|\frac{2MF}{\omega_0}\right|. \tag{21}$$

A representative comparison between the maximum wave amplitudes attained in the fully nonlinear model experiments and the corresponding linear predictions for the $\mathcal{B} = 1$ and $k = 2$ parameter sweep was shown in Figs. 3b and 3d and discussed in section 4a above. The linear predictions shown there are calculated from Eq. (21). Given the shortcomings of linear theory that were identified, it is necessary to turn to nonlinear theories.

c. Analytical reduction II: The Kida vortex limit

The analysis to follow holds only for the special case with both barotropic flow ($\mathcal{B} = 0$) and the azimuthal wavenumber of the forcing $k = 2$. The limit to be considered is that where the horizontal scale of the forcing is much greater than that of the vortex (i.e., $\lambda \ll 1$). It turns out that in this limit, an exact nonlinear solution of the system exists for all values of the remaining parameters (M, Ω), namely the well-known Kida vortex solution (Kida 1981). Kida's solution describes the evolution of an elliptical vortex in a uniform strain and solid body rotation flow: such a vortex remains elliptical for all time.

The Kida solution is recovered for the special case of the topographically forced vortex problem because first, when $\lambda \ll 1$, $J_2(\lambda r) \sim \lambda^2 r^2/8$ and hence ψ_M satisfies

$$\lim_{\lambda \to 0} \psi_M = \lim_{\lambda \to 0} \frac{MJ_2(\lambda r)}{\lambda^2} \cos 2\theta = \frac{\Lambda r^2}{2} \cos 2\theta, \quad \left(\Lambda = \frac{M}{4}\right),$$

which is the streamfunction of a uniform strain flow. Second, when $\mathcal{B} = 0$ a nonzero surf-zone PV (parameter $\Omega \neq 0$) results in a solid body rotation flow with rate Ω. In summary, for the case of our experiments, an initially circular vortex in a barotropic flow ($\mathcal{B} = 0$) with Ω and with $k = 2$ forcing satisfying $\lambda \ll 1$ and with amplitude M will evolve as a Kida vortex in a solid body rotation of strength Ω and a uniform strain flow of strength $\Lambda = M/4$. Comparing the Kida model with the fully nonlinear model results in section 3, where $\lambda \ll 1$ is not satisfied ($\lambda = 1.162$), we approximate the relationship between M and Λ by demanding that the streamfunction at the initially circular vortex edge at $r = 1$ is the same in both cases, giving $\Lambda = 2MJ_2(\lambda)/\lambda^2$. Note that this still gives $\Lambda \to M/4$ as $\lambda \to 0$ (it has been confirmed that either choice of $\Lambda = M/4$ or $\Lambda = 2MJ_2(\lambda)/\lambda^2$ gives similar results to those in Fig. 3c).

Kida showed that the evolution of r_e ($r_e \geq 1$) and ϕ_e of the vortex in such a strain and rotation flow are governed by the pair of coupled ordinary differential equations

$$\begin{aligned} \dot{r}_e &= 2\Lambda r_e \cos 2\phi_e, \\ \dot{\phi}_e &= -\frac{\Lambda(r_e^2 + 1)}{r_e^2 - 1} \sin 2\phi_e + \frac{r_e}{(r_e + 1)^2} + \Omega. \end{aligned} \tag{22}$$

Here dots denote time derivatives. Note that r_e and ϕ_e in this context are entirely consistent with the definitions given in section 2b. The two Eqs. (22) can be combined following the substitution $X = \sin 2\phi_e$ to obtain a linear equation in $X(r_e)$, which can then be integrated once (e.g., Dritschel 1990). After some manipulation, the following "potential form" evolution equation is obtained for the aspect ratio,

$$\dot{r}_e^2 + V_e(r_e; \Omega, \Lambda) = 0, \tag{23}$$

where

$$V_e(r_e; \Omega, \Lambda) = 4r_e^2 \left(\left\{ \frac{r_e}{r_e^2 - 1} \log\left[\frac{(r_e + 1)^2}{4r_e}\right] + \Omega \frac{r_e - 1}{r_e + 1} \right\}^2 - \Lambda^2 \right).$$

The initial condition $r_e(0) = 1$, due to the vortex being circular at $t = 0$, has been used to eliminate the arbitrary constant of integration.

Once written in the potential form [Eq. (23)] the vortex evolution equation becomes relatively straightforward to interpret. Solutions are restricted to values of r_e for which $V_e \leq 0$. As the vortex initially satisfies $r_e = 1$, the maximum possible aspect ratio $r_m = \max[r_e(t)]$ during the evolution of the vortex is given by the smallest root of V_e in $r_e > 1$—that is, where

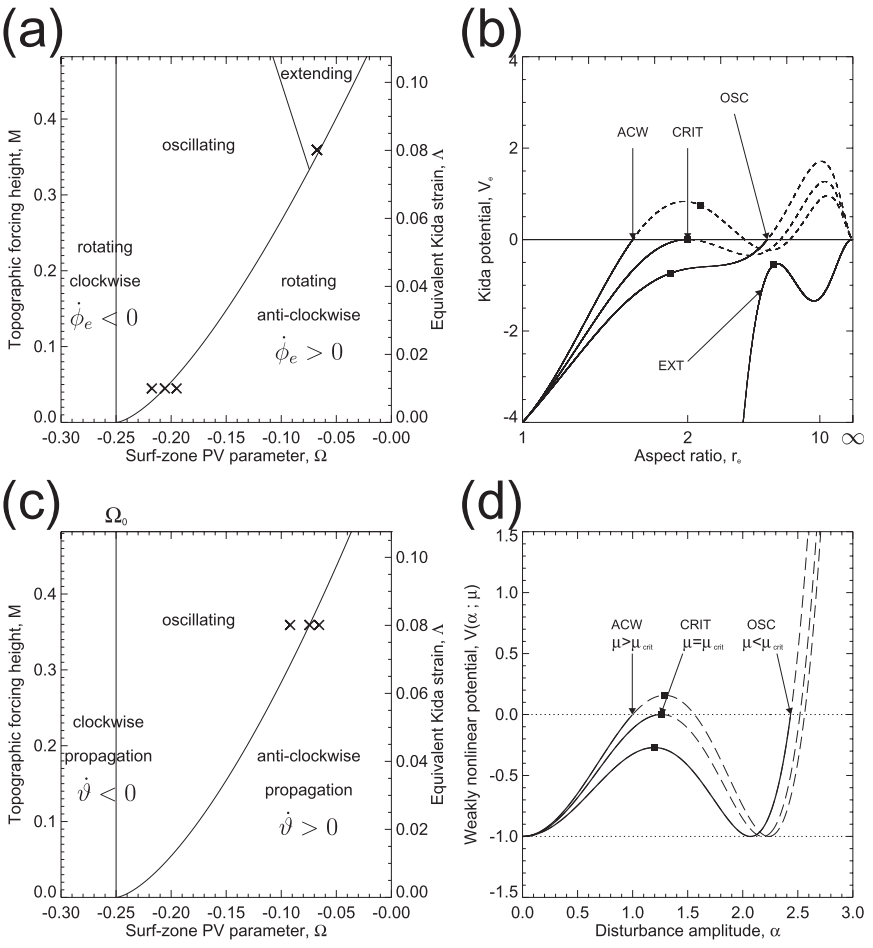

FIG. 4. (a) Location of the four Kida regimes in M and Ω parameter space. Crosses denote the individual cases shown in (b). (b) Plots of the Kida potential function V_e as a function of r_e for the parameters marked in (a). Black boxes mark where $d\phi_e/dt$ changes sign. (c) Location of the three weakly nonlinear regimes in M and Ω parameter space when $\mathcal{B} = 0$, $k = 2$, and $\lambda = 1.162$. Crosses denote the individual cases shown in (d). (d) Plots of the potential function $V(\alpha; \mu)$ as a function of α for three choices of μ corresponding to (Ω, M) marked in (c). Black boxes mark where $d\vartheta/d\tau$ changes sign. In (b) and (d), the solid curve denotes the portion of the potential function encountered during the vortex evolution. Curves marked CRIT correspond to experiments lying on a regime boundary.

$$V_e(r_m; \Omega, \Lambda) = 0. \tag{24}$$

The direction of rotation can be also be deduced from V_e using

$$\dot{\phi}_e = \pm \frac{1}{\sqrt{4\Lambda^2 r_e^2 + V_e}} \left(\frac{1}{2} \frac{dV_e}{dr_e} - \frac{V_e}{r_e} \right).$$

Oscillatory behavior, characterized by a change of sign in $\dot{\phi}_e$, occurs only if the potential function satisfies $dV_e/dr_e = 2V_e/r_e$ for some r_e in the range $1 < r_e < r_m$.

Figure 4a shows the (Ω, Λ) parameter-space regime diagram for the Kida vortex [essentially Fig. 2a of Dritschel (1990)]. The regime boundaries are determined from consideration of the qualitative features of V_e. It is notable that the Kida regimes (ACW, OSC, CW) occupy nearly identical regions of parameter space to those identified for the fully nonlinear model in Fig. 2. At high forcing amplitudes (high strain rate Λ) the Kida vortex exhibits extending behavior, which occurs when the uniform strain flow overcomes the tendency of the vortex to rotate and it becomes stretched out without bound ($r_e \to \infty$). The extending regime (EXT) occupies a subset of the region of parameter space where vortex splits occur in the fully nonlinear model, as will be discussed further below.

The most important regime boundary, as discussed in section 4a above, is the curve separating the OSC regime

and ACW regions of parameter space. In terms of the motion of the vortex, this curve marks the location of an amplitude bifurcation, with relatively low values of r_m in the ACW region and large values in the OSC region. The nature of the bifurcation can be seen in Fig. 4b where V_e is plotted for three values of (Ω, Λ), marked as crosses on Fig. 4a. At the point of transition between the ACW and OSC regimes the potential function V_e has the form illustrated by the curve CRIT, which is distinguished by having a turning point at $r_e = r_m$; that is,

$$\frac{dV_e}{dr_e}(r_m;\Omega_c,\Lambda_c) = V_e(r_m;\Omega_c,\Lambda_c) = 0.$$

The above equations can be used to obtain an expression for the ACW/OSC transition curve in parametric form ($s \geq 1$):

$$\Omega_c(s) = -\frac{1}{2} + \frac{s^2+1}{2(s-1)^2}\log\left[\frac{(s+1)^2}{4s}\right],$$
$$\Lambda_c(s) = -\frac{s-1}{2(s+1)} + \frac{(s+1)}{2(s-1)}\log\left[\frac{(s+1)^2}{4s}\right]. \quad (25)$$

Figure 4b shows that when passing from one side of the transition curve to the other V_e changes from having three roots (ACW region) to just one (OSC region). Hence there is a discrete jump in the value of the lowest root r_m across the regime boundary, corresponding to a jump in the maximum aspect ratio attained by the vortex.

The maximum wave amplitude during the oscillation can be obtained from the vortex aspect ratio using Eq. (10), which reveals that $a_m = (1/2)\sqrt{r_m + r_m^{-1} - 2}$. The maximum aspect ratio can be obtained numerically from Eq. (24) as a function of (Ω, Λ) using a root finding method, again following Dritschel (1990, see his Fig. 3a). The result is shown in Fig. 3c; it has been compared with the results from the numerical model (for $\mathcal{B} = 0$, $k = 2$) shown in Fig. 3a, as discussed above in section 4a. The reason that the contours on Fig. 3c are straight lines is because Eq. (24), which determines r_m and thus a_m, is linear in both Ω and Λ.

The above analysis gives some further insight into the reasons for the inaccuracy of the Kida model at large amplitudes. First, once the vortex in the fully nonlinear model (for which λ is finite) undergoes significant elongation, it becomes influenced by $\mathbf{u}_M$ in regions where it is not well approximated by the uniform strain flow. Second, an elliptical vortex is well known to be unstable to Love (1893)-type instabilities once its aspect ratio exceeds 3. Mitchell and Rossi (2008) have recently shown that the branch of the instability for aspect ratios between 3 and 6 leads only to ejection of filaments, while a vortex split will result if the vortex aspect ratio exceeds 6.046 because of the "negative wave-4" branch of the Love instability. Hence, although vortex splitting is not an explicit prediction of the Kida theory, it can nevertheless be anticipated as a probable outcome when a_m exceeds a given threshold that will depend on the details of the forcing (i.e., the value of λ here).

d. Analytical reduction III: Weakly nonlinear theory

Next, a weakly nonlinear analysis of the forced vortex model is presented that is exact in the joint limit $M \to 0$, $\Omega \to \Omega_0$ for arbitrary k, $\mathcal{B}$, and λ. Note that this limit is quite distinct from the Kida vortex limit discussed above. Scaling analysis suggests the choice $\epsilon = M^{1/3}$, in contrast to $\epsilon = M$ for the linear theory above, for the small parameter. The system must be near resonance for nonlinear effects to be important (otherwise the linear theory described above will remain accurate) and nonlinearity is found to enter at leading order when $|\Omega - \Omega| \sim \epsilon^2$. Consequently, we define

$$\Omega_2 = \frac{\Omega - \Omega_0}{\epsilon^2},$$

with Ω_2 taken to be an order unity quantity in all that follows. The background velocity $U_b(r)$ can then be expanded as

$$U_b(r) = U_0(r) + \epsilon^2 U_2(r), \quad \text{with}$$
$$U_0(r) = \frac{I_k K_k - I_1 K_1}{I_1} I_1(\mathcal{B}r), \quad U_2(r) = \frac{2\Omega_2}{\mathcal{B}} I_1(\mathcal{B}r). \quad (26)$$

The relevant time scale for the motion based on these scalings is $O(\epsilon^{-2})$, hence a "slow" time variable $T = \epsilon^2 t$ is introduced.

Just as for the linear theory above, the kinematic Eq. (12) at the vortex boundary can be expanded in powers of ϵ. Solutions for the disturbance $\eta = \eta(\theta, T)$ are sought that depend only on the slow time variable T; hence, the time derivative ∂t is replaced by $\epsilon^2 \partial_T$. The expansion proceeds, as is shown in the appendix, using the fact that when the vortex boundary is located at $r = 1 + \epsilon\eta$ the advection term can be expanded in powers of ϵ as

$$(\mathbf{u}_R \cdot \mathbf{n})_c = \epsilon(\mathbf{u}_R \cdot \mathbf{n})_c^l[\eta] + \epsilon^2(\mathbf{u}_R \cdot \mathbf{n})_c^q[\eta] + \epsilon^3(\mathbf{u}_R \cdot \mathbf{n})_c^c[\eta] + O(\epsilon^4), \quad (27)$$

where the superscripts l, q, and c denote the linear, quadratic, and cubic functionals of η given by Eqs. (A6)–(A8). Finally, $\eta(\theta, T)$is itself expanded as

$$\eta = \eta_1 + \epsilon\eta_2 + \epsilon^2\eta_3 + O(\epsilon^3),$$

allowing the kinematic Eq. (12) to be expanded. Equating terms at successive powers of ϵ reveals

$$O(\epsilon): \quad \mathcal{L}\eta_1 = 0, \tag{28}$$

$$O(\epsilon^2): \quad \mathcal{L}\eta_2 = (\mathbf{u}_R \cdot \mathbf{n})_c^q[\eta_1] + [U_0(1) - U_{0r}(1)]\eta_1\eta_{1\theta}, \tag{29}$$

$$O(\epsilon^3): \quad \mathcal{L}\eta_3 = -\eta_{1T} + (\mathbf{u}_R \cdot \mathbf{n})_c^c[\eta_1] + (\mathbf{u}_R \cdot \mathbf{n})_c^q[\eta_1, \eta_2] + \upsilon_M^l - \left[\frac{1}{2}U_{0rr}(1) - U_{0r}(1) + U_0(1)\right]\eta_1^2\eta_{1\theta} + [U_0(1) - U_{0r}(1)](\eta_1\eta_{2\theta} + \eta_2\eta_{1\theta}) - U_2(1)\eta_{1\theta}. \tag{30}$$

Here υ_M^l is identical to the corresponding forcing term in the linear problem above, and U_{0r} and U_{0rr} refer to the first and second derivatives of $U_0(r)$ defined in Eq. (26) above. The linear operator $\mathcal{L}$ is defined by

$$\begin{aligned}\mathcal{L}\eta &\equiv U_0(1)\eta_\theta(\theta) - (\mathbf{u}_R \cdot \mathbf{n})_c^l[\eta], \\ &= U_0(1)\eta_\theta(\theta) - \frac{1}{2\pi}\int_0^{2\pi}[\eta_\theta(\theta + \tilde{\psi}) - \eta_\theta(\theta)\cos\tilde{\psi}]K_0\left(\mathcal{B}\sqrt{2 - 2\cos\tilde{\psi}}\right)d\tilde{\psi}.\end{aligned}$$

The leading-order Eq. (28) is identical to that of the unforced linear problem and has the solution

$$\eta_1(\theta, T) = \mathrm{Re}A(T)e^{ik\theta},$$

where $A(T)$ is (at this order) an arbitrary complex amplitude. As the leading-order equation is homogeneous, the function $e^{ik\theta}$ inhabits the kernel of $\mathcal{L}$. In order for the entire solution to depend only on the slow time variable T, at higher orders the forcing terms on the right-hand side of the equation must not project[2] onto the kernel of $\mathcal{L}$. If the higher-order Eqs. (29)–(30) are written schematically as

$$\mathcal{L}\eta_i = \mathcal{F}_i, \quad i = 2, 3, \ldots,$$

the projection or orthogonality condition on $\mathcal{F}_i$ is easily shown to be

$$\int_0^{2\pi}\mathcal{F}_i e^{-ik\theta}d\theta = 0, \quad i = 2, 3, \ldots \tag{31}$$

and must be satisfied for all $i \geq 2$.

At $O(\epsilon^2)$ the quadratic forcing term is shown in the appendix to be given by Eq. (A11) and acts to excite the second harmonic. The orthogonality condition (31) is satisfied automatically. The solution for η_2 is found, using the results in the appendix, to be

$$\eta_2 = \mathrm{Re}\frac{(P_{2k} - P_k + P_1) + \frac{\mathcal{B}}{2}(2P_k' + P_{2k}' - P_1') + [U_0(1) - U_{0r}(1)]}{4(P_k - P_{2k})}A(T)^2e^{2ik\theta} - \frac{1}{4}|A(T)|^2. \tag{32}$$

Here P_k denotes the Bessel function product

$$P_k(\mathcal{B}) = I_k(\mathcal{B})K_k(\mathcal{B}) \tag{33}$$

and P_k' its derivative with respect to $\mathcal{B}$. The properties of the products P_k have been the subject of recent study in the mathematics literature (e.g., Penfold et al. 2007; we follow their notation here) and they are discussed further in the appendix. The final term in Eq. (32) arises because an arbitrary function $g(T)$ is a solution of the homogeneous equation $\mathcal{L}g = 0$. The function $g(T)$ is chosen at this order to be the vortex area correction $g(T) = -|A|^2/4$. This choice for $g(T)$ is essential to ensure that the vortex retains its initial area throughout its evolution in time to $O(\epsilon^2)$ accuracy [with the initial area being determined by the circular initial condition $A(0) = 0$]. Note that the vortex area correction would in fact be enforced at $O(\epsilon^4)$ in the expansion, and the result must be anticipated here in order to obtain the correct area-preserving result at $O(\epsilon^2)$.

[2] Had we retained the dependence on the fast time variable t here, the projection condition would appear as the "condition for the removal of secular terms" at higher orders in the expansion.

At $O(\epsilon^3)$ an amplitude equation follows from application of the orthogonality condition (31). In the appendix all of the relevant forcing terms are evaluated. Applying Eq. (31), the following equation is obtained:

$$\frac{dA}{dT} + i\tilde{\omega}_0 A + i\omega_2 A|A|^2 = iF. \tag{34}$$

Here

$$\tilde{\omega}_0 = 2k\Omega_2 \mathcal{B}^{-1} I_1 = kU_2(1) \tag{35}$$

measures the sense and extent to which the forcing is initially off-resonant for the relevant linear mode;

$$\omega_2 = -\frac{k}{4}\Bigg(2(\mathcal{B}^2 + k^2)P_k - \frac{1}{2}(\mathcal{B}^2 + 4k^2)P_{2k} - \frac{3}{2}(\mathcal{B}^2 + 1)P_1 + \frac{1}{8}P_{2k} - \frac{1}{8}P_k - P_1 + \frac{\mathcal{B}^2}{4}(2P_{2k} - P_{2k+1} - P_{2k-1}) + \frac{\mathcal{B}}{4}(5P_1' - 10P_k' - P_{2k}') - \left[\frac{1}{2}U_{0rr}(1) - 2U_{0r}(1) + 2U_0(1)\right] + \frac{\{P_k - P_{2k} + 2P_1 + \mathcal{B}(2P_k' + P_{2k}' - P_1') + 2[U_0(1) - U_{0r}(1)]\}^2}{8(P_k - P_{2k})}\Bigg) \tag{36}$$

measures the frequency correction to the free mode as the (square of) its amplitude increases; and the forcing $F(k, \mathcal{B}, \lambda)$ is given by Eq. (15). It is notable that Eq. (34) differs from its linear counterpart [Eq. (17)] only by the introduction of the cubic nonlinear $iA|A|^2$ term. In the appendix it is shown that the coefficient ω_2 of this nonlinear term reduces to known results in certain limits, namely the barotropic 2D Euler limit ($\mathcal{B} \to 0$) and the straight contour limit $k \to \infty$, $\mathcal{B} \to \infty$ (with $k/\mathcal{B}$ held constant).

Equation (34) can be rewritten using the substitutions

$$A = \left(\frac{F}{\omega_2}\right)^{1/3} \alpha(\tau) \exp[-i\vartheta(\tau)], \quad \text{and} \quad \tau = (F^2\omega_2)^{1/3} T,$$

where $\alpha(\tau)$ and $\vartheta(\tau)$ are real functions, resulting in the ordinary differential equations

$$\begin{aligned} \dot{\alpha} &= -\sin\vartheta, \\ \dot{\vartheta} &= -\frac{\cos\vartheta}{\alpha} - \mu + \alpha^2, \end{aligned} \tag{37}$$

where $\mu = -\tilde{\omega}_0/(F^2\omega_2)^{1/3}$; the dots here denote differentiation with respect to τ. Equations (37) have a similar form and interpretation to the Kida Eqs. (22) and can be analyzed in a similar way. Changing variables to $Y = \cos\vartheta$ and combining the two equations results in the linear equation

$$\frac{dY}{d\alpha} + \frac{Y}{\alpha} = -\mu + \alpha^2,$$

which can be integrated directly to obtain

$$Y(\alpha) = -\frac{\mu\alpha}{2} + \frac{\alpha^3}{4} + \frac{C}{\alpha}, \tag{38}$$

where the constant of integration C can be set to zero using the initial condition $\alpha(0) = 0$, since the wave amplitude is initially zero.

Using the first equation in Eqs. (37) to substitute for $Y(\alpha)$, a potential form equation for $\alpha(\tau)$ is obtained:

$$(\dot{\alpha})^2 + V(\alpha; \mu) = 0, \tag{39}$$

where the potential $V(\alpha; \mu)$ is a cubic polynomial in α^2:

$$V(\alpha; \mu) = \alpha^2\left(\frac{\mu}{2} - \frac{\alpha^2}{4}\right)^2 - 1. \tag{40}$$

The behavior of the forced nonlinear oscillator consequently depends only on the parameter μ. A bifurcation in the maximum amplitude of the wave, similar to that described above for the Kida model, occurs when the cubic switches from having three roots to one. The ''local maximum'' turning point of $V(\alpha; \mu)$ is easily shown to occur at $\alpha = \sqrt{2\mu/3}$, and the condition for three roots to exist is $V(\sqrt{2\mu/3}; \mu) \geq 0$, which can be rearranged to obtain the condition $\mu \geq \mu_c = 3/2^{1/3} \approx 2.38$.

Figure 4d illustrates how the potential curve changes as μ passes through μ_c and indicates the associated jump in the maximum amplitude as the nearest roots of $V(\alpha; \mu)$ become complex. The curve $\mu = \mu_c$ defines a boundary between a low-amplitude ACW regime and a high-amplitude OSC regime, analogous to the Kida model regime boundary [Eq. (25)]. A rotation rate analogous to that of the Kida model can be expressed as

$$\dot{\vartheta} = \frac{1}{2\sqrt{1 + V}}\frac{dV}{d\alpha} = \frac{\mu}{2} - \frac{3\alpha^2}{4},$$

making it clear that

(i) the vortex will rotate in both directions in the oscillating regime, as $\dot{\vartheta}$ changes sign at each turning point of $V(\alpha; \mu)$, and
(ii) a further regime boundary exists at $\mu = 0$, between the OSC and CW regimes, as $V(\alpha, \mu)$ no longer has any turning points for $\mu < 0$ and in this case $\dot{\vartheta} < 0$ throughout the motion.

The resulting regime diagram, shown in Fig. 4c for the parameter settings corresponding to the fully nonlinear case ($\mathcal{B} = 0, k = 2$, and $\lambda = 1.162$), strongly resembles that of the Kida vortex (Fig. 4a) and the fully nonlinear model (Fig. 2).

The exact value of $\alpha_m = \max[\alpha(\tau)]$ can be calculated using Cardano's formulas for the roots of a cubic. The result is

$$\alpha_m^2 = \begin{cases} \dfrac{4\mu}{3} + \dfrac{2\mu^2}{3(27 - \mu^3 + 3\sqrt{81 - 6\mu^3})} + \dfrac{2}{3}\left(27 - \mu^3 + 3\sqrt{81 - 6\mu^3}\right)^{1/3}, & \mu < \mu_c \\ \dfrac{4\mu}{3} - \dfrac{4\mu}{3}\cos\left[\dfrac{1}{3}\arccos\left(\dfrac{\mu^3 - 27}{\mu^3}\right)\right] \end{cases} \tag{41}$$

The maximum of $\alpha_m(\mu)$ occurs at $\mu = \mu_c$ on the OSC side of the regime boundary, and the amplitude is exactly halved (from $\alpha_m = 2^{4/3}$ to $\alpha_m = 2^{1/3}$) as the boundary is crossed into the ACW regime.

Rescaling to facilitate comparison with the fully nonlinear model, the maximum vortex Rossby wave amplitude a_m can be written in terms of functions of the original parameters of the problem as

$$a_m = -\left[\frac{MF(k,\mathcal{B},\lambda)}{\omega_2(k,\mathcal{B})}\right]^{1/3}\alpha_m(\mu), \quad \text{where}$$
$$\mu = -\frac{2k[\Omega - \Omega_0(k,\mathcal{B})]I_1(\mathcal{B})}{\mathcal{B}F(k,\mathcal{B},\lambda)^{2/3}\omega_2(k,\mathcal{B})^{1/3}M^{2/3}}, \tag{42}$$

and $\Omega_0(k,\mathcal{B})$ is given by Eq. (19), $\omega_2(k,\mathcal{B})$ is given by Eq. (36), and $F(k, \mathcal{B}, \lambda)$ is given by Eq. (15). Equation (42) is the basis for the weakly nonlinear prediction for a_m for the $k = 2$, $\mathcal{B} = 0$, and $\mathcal{B} = 1$ cases shown in Figs. 3e and 3f, respectively. The success of these predictions is discussed in section 4a above. It should be noted that where very high amplitudes are predicted, the good agreement is probably serendipitous, as weakly nonlinear theory ought not to be valid. The key prediction of the location of the amplitude bifurcation in parameter space also appears to be accurate far beyond amplitudes for which it should be expected. The amplitude bifurcation curve, defined by $\mu = \mu_c$, can be expressed in terms of the original full model parameters as

$$M = \frac{4}{3\sqrt{3}}\frac{(kI_1)^{3/2}}{\mathcal{B}^{3/2}\sqrt{-\omega_2(k,\mathcal{B})}F(k,\mathcal{B},\lambda)}[\Omega - \Omega_0(k,\mathcal{B})]^{3/2}, \tag{43}$$

which explains the simple shape of the curve in the (Ω, M) regime diagrams. The close correspondence between the curve [Eq. (43)] and its Kida model equivalent [Eq. (25)] (see curves on Fig. 3a) is notable given that the two approaches are valid in different limits of the full model.

5. Prototype stratospheric sudden warming experiments

In this section, the results presented above will be used to construct a simulation of a "prototype" vortex splitting SSW using the simple nonlinear vortex model described in section 2. In previous studies, SSWs have often occurred as a direct response to a switch-on or dramatic increase in forcing or are uncontrolled spontaneous events in the midst of a seasonal cycle or long model experiment. Our simulation differs from those of previous studies, and from our experiments in section 3, in that the timing of the SSW will be controlled by an incremental change in the model boundary conditions. This approach is similar to that of Chao (1985), who used catastrophe theory to explain the transition from steady flow to vacillating regime in a Holton and Mass (1976) model framework, with the "catastrophic" transition being an SSW-like event. However, the focus here is on quantitatively predicting this transition in our model using the self-tuning resonant theory described in section 4.

The fully nonlinear numerical model is initialized in the same way as the experiments in section 3 with parameters $\mathcal{B} = 0$, $k = 2$, and $\lambda = 1.162$. Throughout the experiment, the surf-zone PV parameter is kept at $\Omega = 0.10$. Between model times of $t = 0$ and $t = 215.5$ (~42.9 days, based on taking $Q_i - Q_o = 0.4f_0$, as is typical of the NH winter stratosphere) the topographic forcing is kept constant at $M = M_1 = 0.28$ ($h_m = 688$ m, using a reference height of $H = 6.14$ km). At model time $t = 215.5$, when the vortex patch has returned to its initially circular shape, the topographic forcing is instantaneously increased by approximately 7% to $M = M_2 = 0.3$ ($h_m = 737$ m). The forcing is maintained at $M = M_2$ until the end of the experiment at $t = 300$. The locations in parameter space of (Ω, M_1) and (Ω, M_2) are shown in the middle-right panel of Fig. 5 (as well as Figs. 2b and 3a), with the initial parameters (Ω, M_1) lying in the ACW

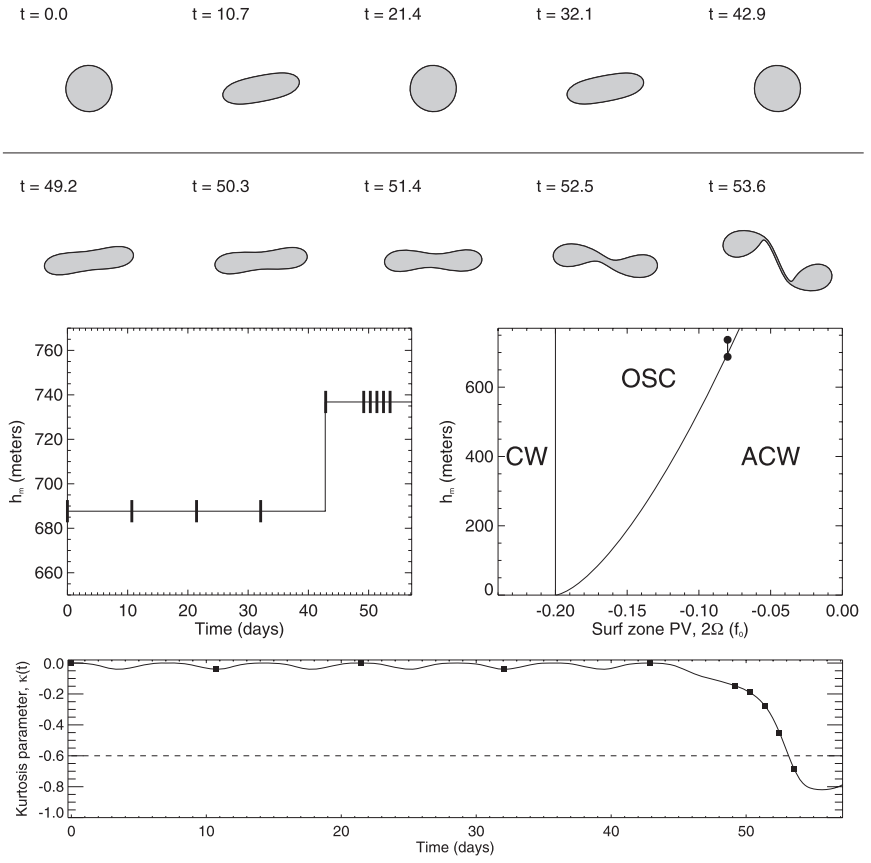

FIG. 5. (top) Snapshots of vortex evolution at a selection of times during the SSW prototype experiment with $\mathcal{B} = 0$, $k = 2$, and $\lambda = 1.162$. (middle left) Forcing amplitude as a function of time in dimensional units, with tick marks corresponding to times at which snapshots are shown. (middle right) Location of the experiment parameters in dimensional parameter space. (bottom) Evolution of $\mathcal{K}(t)$ as a function of time, with box symbols marking times at which snapshots are shown. The dashed line marks the threshold $\mathcal{K}(t) = -0.6$ at which vortex splitting is diagnosed.

regime (as predicted by both the Kida and weakly nonlinear theories), and (Ω, M_2) in the OSC regime.

The top panels of Fig. 5 show the evolution of the vortex patch over the duration of the experiment, with the evolution of M as a function of time, together with the times of the snapshots, indicated in the lower-left panel. When $M = M_1$, disturbances on the vortex edge propagate anticlockwise, in keeping with the Kida and weakly nonlinear theoretical predictions. Furthermore, the evolution is stable, with no shedding or filamentation of PV away from the main vortex core. Six periods are completed by $t = 42.9$ days, as evidenced by the behavior of the $\mathcal{K}(t)$ diagnostic in the bottom panel. At $t = 42.9$ days, when the topographic forcing height increases to $M = M_2$, the vortex enters the OSC regime. The effect of the change is not evident for another 7 days. By day 49, however, the amplitude of the disturbance on the vortex exceeds the maximum amplitude observed up to that point. The vortex then begins to rotate clockwise and simultaneously becomes pinched near its centroid. The pinch becomes increasingly pronounced over the following 2 days, and by day 53.6 the vortex has split.

In principle the onset of the vortex split in the experiment described above might have been caused by an infinitesimal change either in M or Ω. Note, however, that the periods of the vortex motion become infinite close to the transition between the ACW and OSC regimes. The onset of the split will therefore take longer as the magnitude of the change in M or Ω is reduced. Nevertheless, it has been demonstrated above that a relatively modest change in M is sufficient to cause a split on the physically relevant time scale of a few days.

6. Conclusions

The nonlinear evolution of the stratospheric polar vortex under the influence of a topographic forcing has

been considered in a single-layer *f*-plane model. It has been demonstrated that most aspects of the nonlinear behavior exhibited by the model can be understood using either of two distinct analytical reductions, each valid in a different region of parameter space. The important feature of the resulting amplitude Eqs. (22) and (34) is that each exhibits a bifurcation in vortex Rossby wave amplitude across a well-defined curve in parameter space. It has been shown that it is just such a bifurcation in wave amplitude that causes vortex splits in the fully nonlinear model, and we propose an amplitude bifurcation of exactly this type as the fundamental dynamical mechanism underlying vortex splitting SSWs.

The insights from the nonlinear analyses allowed the construction of a ''prototype vortex split SSW'' (see section 5) during which, following a long period of stable and relatively low-amplitude oscillation, a small (in principle, infinitesimal) change in forcing amplitude leads to a spontaneous vortex split. The small change necessary to cause the split could equally have arisen from a small change in the underlying stratospheric circulation, controlled in our model by the parameter Ω, as might be associated with, for example, the seasonal cycle. The key point illustrated by the prototype SSW is that it is not necessary, notwithstanding the efforts of numerous authors, to invoke an anomalous tropospheric ''event,'' which then leads to anomalous vertical propagation of Rossby waves, in order to explain the occurrence of the SSW. Rather, a vortex splitting SSW can arise as a spontaneous bifurcation of the forced vortex system as the system's parameters evolve smoothly in time. It is to be emphasized that linear theory (e.g., Tung and Lindzen 1979b) gives no clue as to the existence of the bifurcation.

Plumb (1981b) has previously obtained the forced nonlinear oscillator Eq. (34) in connection with SSWs in a β-channel model. While in many respects Plumb's work has strongly influenced the present study, there are two key differences in interpretation that merit comment.

- First, Plumb (1981b) makes reference to ''wave–wave interaction'' where the interaction is between a ''stationary topographic wave'' and a freely propagating Rossby wave. Plumb's terminology has been influential; for example, Hio and Yoden (2007) recently discussed wave–wave interaction in a topographically forced single-layer shallow water model. The present work has deliberately avoided this terminology and any associated interpretation because of the fact that in any quasigeostrophic system the influence of the topography is equivalent to that of a time-independent effective ''forcing'' velocity field $\mathbf{u}_M$, which is determined by the distribution of the topography $h(\mathbf{x})$. Although $\mathbf{u}_M$ might exhibit a wavelike pattern—for instance, if $h(\mathbf{x})$ is chosen to be spatially periodic—it is not a wave in any meaningful sense. This is most obvious in the Kida vortex limit of broad topographic forcing discussed above, for which $\mathbf{u}_M$ becomes a uniform strain flow. Our view is therefore that it is meaningful to think in terms of only a single wave, the freely propagating Rossby wave, which oscillates in amplitude under the influence of a stationary forcing velocity field $\mathbf{u}_M$.
- Second, although Plumb was aware that the system governed by Eq. (34) exhibits a bifurcation in oscillation amplitude [see in particular section 3 of Plumb (1981a)], Plumb chose instead to identify SSWs with linear instability of the steady solutions of Eq. (34), rather than the bifurcation itself. Our view is that Plumb's conceptual model is flawed for the same reason we believe the linear theory explanation of SSWs to be flawed, because the instability growth rates in Plumb's theory [see Eq. (3.8) of Plumb 1981b] depend smoothly on the model parameters, and therefore one is once again required to invoke an abrupt change in the parameter values (such as the magnitude of the topographic forcing) to obtain a rapid onset SSW. Our hope is that the unambiguous demonstration given in section 5 that the bifurcation governed here by Eq. (43) directly leads to a rapid onset SSW makes a compelling case for the new paradigm.

The application of the nonlinear theory to vortex displacement SSWs, which have some complicating features due to their baroclinic vertical structure (e.g., Manney et al. 1999; Matthewman et al. 2009), is the subject of Part II (Esler and Matthewman 2011).

Acknowledgments. We thank three anonymous referees for their useful comments that helped improve the readability of this paper. NJM acknowledges funding from a UK NERC studentship and JGE acknowledges support from NERC Grant NE/G003122/1. This paper is dedicated to the memory of Chris Bell, who contributed many helpful suggestions during the early stages of this work.

APPENDIX

Details of the Weakly Nonlinear Analysis

a. Expansion of the integral expression determining the velocity field

In $\mathbb{R}^2$ it is well known that the solution of Eq. (11), expressing the streamfunction in terms of the PV, can be given in Green's function form (e.g., Esler 2004) as

$$\psi_R(\mathbf{x}) = -\frac{1}{2\pi}\int\int_{\mathbb{R}^2} K_0(\mathcal{B}|\mathbf{x} - \tilde{\mathbf{x}}|)q(\tilde{\mathbf{x}})\, d^2\tilde{\mathbf{x}}. \quad \text{(A1)}$$

For the special case where $q(\mathbf{x})$ is a uniform patch of constant unit vorticity in an area $\mathcal{D}$ as here, the curl of the vector field $\psi_R\mathbf{k}$ in Eq. (A1) can be taken, and Green's theorem in the plane used to convert the area integral to a contour integral around the patch boundary $\partial\mathcal{D}$:

$$\mathbf{u}_R(\mathbf{x}) = \frac{1}{2\pi}\oint_{\partial\mathcal{D}} K_0(\mathcal{B}|\mathbf{x} - \tilde{\mathbf{x}}|)q(\tilde{\mathbf{x}})\, d\tilde{\mathbf{x}}. \quad \text{(A2)}$$

Equation (A2) gives the self-induced velocity $\mathbf{u}_R$ of the vortex patch in terms of the location of its boundary $\partial\mathcal{D}$. It is assumed in all that follows that the location of the boundary can be expressed in polar form as $r = 1 + \epsilon\eta(\theta)$, where ε is a small parameter set equal to $M^{1/3}$ in section 4d above.

In fact, to expand Eq. (12) in powers of ε, only $(\mathbf{u}_R \cdot \mathbf{n})_c[\theta]$ need be evaluated. Here $\mathbf{n}$ is the vector perpendicular to the patch edge defined by Eq. (13), and the subscript c denotes evaluation at the patch boundary. To proceed note first that the position vectors $\mathbf{x}$ and $\tilde{\mathbf{x}}$ in Eq. (A2) can be expanded as

$$\mathbf{x}(\theta) = [1 + \epsilon\eta(\theta)]\hat{\mathbf{r}}(\theta),$$
$$\tilde{\mathbf{x}}(\phi) = [1 + \epsilon\eta(\phi)]\hat{\mathbf{r}}(\phi),$$

so that $d\tilde{\mathbf{x}} = \{\epsilon\eta_\phi(\phi)\hat{\mathbf{r}}(\phi) + [1 + \epsilon\eta(\phi)]\hat{\boldsymbol{\phi}}(\phi)\}\, d\phi$ and $(\mathbf{u}_R \cdot \mathbf{n})_c$ can be written

$$(\mathbf{u}_R \cdot \mathbf{n})_c(\theta) = \frac{1}{2\pi}\int_0^{2\pi} K_0(\mathcal{B}|\mathbf{x}(\theta) - \tilde{\mathbf{x}}(\phi)|)\left(\left\{\epsilon\eta_\phi(\phi) - \epsilon\frac{\eta_\theta(\theta)[1 + \epsilon\eta(\phi)]}{1 + \epsilon\eta(\theta)}\right\}\cos(\theta - \phi)\right.$$
$$\left. + \left[1 + \epsilon\eta(\phi) + \epsilon^2\frac{\eta_\theta(\theta)}{1 + \epsilon\eta(\theta)}\eta_\phi(\phi)\right]\sin(\theta - \phi)\right) d\phi, \quad \text{(A3)}$$

where the identities

$$\hat{\mathbf{r}}(\theta) \cdot \hat{\mathbf{r}}(\phi) = \hat{\boldsymbol{\phi}}(\theta) \cdot \hat{\boldsymbol{\phi}}(\phi) = \cos(\theta - \phi) \quad \text{and}$$
$$\hat{\mathbf{r}}(\theta) \cdot \hat{\boldsymbol{\phi}}(\phi) = -\hat{\mathbf{r}}(\phi) \cdot \hat{\boldsymbol{\phi}}(\theta) = \sin(\theta - \phi)$$

have been used. The Bessel function terms in Eq. (A3) can be expanded using the (easily verified) Taylor series expansion, valid for any smooth function f, constant ν, and vectors $\mathbf{a}$ and $\mathbf{b}$:

$$f(\nu|\mathbf{a} + \epsilon\mathbf{b}|) = \left(1 + \left\{\epsilon\frac{(\mathbf{a}\cdot\mathbf{b})}{(\mathbf{a}\cdot\mathbf{a})} + \frac{\epsilon^2}{2}\left[1 - \epsilon\frac{(\mathbf{a}\cdot\mathbf{b})}{(\mathbf{a}\cdot\mathbf{a})}\right]\frac{(\mathbf{b}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{a}) - (\mathbf{a}\cdot\mathbf{b})^2}{(\mathbf{a}\cdot\mathbf{a})^2}\right\}\nu\frac{d}{d\nu}\right.$$
$$\left. + \left[\frac{\epsilon^2}{2}\frac{(\mathbf{a}\cdot\mathbf{b})^2}{(\mathbf{a}\cdot\mathbf{a})^2} + \frac{\epsilon^3}{2}\frac{(\mathbf{b}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{a})(\mathbf{a}\cdot\mathbf{b}) - (\mathbf{a}\cdot\mathbf{b})^3}{(\mathbf{a}\cdot\mathbf{a})^3}\right]\nu^2\frac{d^2}{d\nu^2} + \frac{\epsilon^3}{6}\frac{(\mathbf{a}\cdot\mathbf{b})^3}{(\mathbf{a}\cdot\mathbf{a})^3}\nu^3\frac{d^3}{d\nu^3}\right)f(\nu|\mathbf{a}|) + O(\epsilon^4). \quad \text{(A4)}$$

Using Eq. (A4) with $f \equiv K_0$, $\nu = \mathcal{B}$, $\mathbf{a} = \hat{\mathbf{r}}(\theta) - \hat{\mathbf{r}}(\phi)$, and $\mathbf{b} = \eta(\theta)\hat{\mathbf{r}}(\theta) - \eta(\phi)\hat{\mathbf{r}}(\phi)$, and noting that

$$\mathbf{a}\cdot\mathbf{a} = 2[1 - \cos(\theta - \phi)],$$
$$\mathbf{a}\cdot\mathbf{b} = [\eta(\theta) + \eta(\phi)][1 - \cos(\theta - \phi)],$$
$$\mathbf{b}\cdot\mathbf{b} = [\eta(\theta) - \eta(\phi)]^2 + 2\eta(\theta)\eta(\phi)[1 - \cos(\theta - \phi)],$$

the Bessel function expression in Eq. (A3) can be expanded in ϵ as

$$K_0(\mathcal{B}|\mathbf{x}(\theta) - \tilde{\mathbf{x}}(\phi)|) = \left\{1 + \epsilon\frac{\eta(\theta) + \eta(\phi)}{2}\mathcal{B}\frac{d}{d\mathcal{B}} + \epsilon^2\frac{[\eta(\theta) + \eta(\phi)]^2}{8}\left[\frac{1 + \cos(\theta - \phi)}{1 - \cos(\theta - \phi)}\right]\mathcal{B}\frac{d}{d\mathcal{B}}\right.$$
$$+ \epsilon^2\frac{[\eta(\theta) + \eta(\phi)]^2}{8}\mathcal{B}^2\frac{d^2}{d\mathcal{B}^2} + \epsilon^3\frac{[\eta(\theta) + \eta(\phi)][\eta(\theta) - \eta(\phi)]^2}{16}\left[\frac{1 + \cos(\theta - \phi)}{1 - \cos(\theta - \phi)}\right]$$
$$\left.\times\left(\mathcal{B}^2\frac{d^2}{d\mathcal{B}^2} - \mathcal{B}\frac{d}{d\mathcal{B}}\right) + \epsilon^3\frac{[\eta(\theta) + \eta(\phi)]^3}{48}\mathcal{B}^3\frac{d^3}{d\mathcal{B}^3}\right\}K_0\left[\mathcal{B}\sqrt{2 - 2\cos(\theta - \phi)}\right] + O(\epsilon^4). \quad \text{(A5)}$$

The integral in Eq. (A3) can now be expanded in powers of ϵ according to Eq. (27), thus defining linear, quadratic, and cubic functionals of η. These can now be obtained as

$$(\mathbf{u}_R\cdot\mathbf{n})_c^l[\eta]=\frac{1}{2\pi}\int_0^{2\pi}[\eta_\theta(\theta+\tilde\psi)-\eta_\theta(\theta)\cos\tilde\psi]K_0\left(\mathcal{B}\sqrt{2-2\cos\tilde\psi}\right)d\tilde\psi, \tag{A6}$$

$$(\mathbf{u}_R\cdot\mathbf{n})_c^q[\eta]=\frac{1}{2\pi}\int_0^{2\pi}\left\{[\eta_\theta(\theta+\tilde\psi)-\eta_\theta(\theta)\cos\tilde\psi]\left[\eta(\theta+\tilde\psi)-\eta(\theta)+\frac{\eta(\theta+\tilde\psi)+\eta(\theta)}{2}\mathcal{B}\frac{d}{d\mathcal{B}}\right]-\eta_\theta(\theta+\tilde\psi)\eta_\theta(\theta)\sin\tilde\psi\right\}K_0\left(\mathcal{B}\sqrt{2-2\cos\tilde\psi}\right)d\tilde\psi, \tag{A7}$$

$$\begin{aligned}(\mathbf{u}_R\cdot\mathbf{n})_c^c[\eta]=\frac{1}{2\pi}\int_0^{2\pi}\Bigg([\eta_\theta(\theta+\tilde\psi)-\eta_\theta(\theta)\cos\tilde\psi]\Bigg\{&\eta(\theta)[\eta(\theta)-\eta(\theta+\tilde\psi)]\\&+\frac{\eta(\theta+\tilde\psi)^2-3\eta(\theta)^2}{4}\mathcal{B}\frac{d}{d\mathcal{B}}+\frac{[\eta(\theta+\tilde\psi)+\eta(\theta)]^2}{8}\left(\mathcal{B}^2\frac{d^2}{d\mathcal{B}^2}+\mathcal{B}\frac{d}{d\mathcal{B}}\right)+\frac{[\eta(\theta+\tilde\psi)-\eta(\theta)]^2}{4(1-\cos\tilde\psi)}\mathcal{B}\frac{d}{d\mathcal{B}}\Bigg\}\\&+\eta_\theta(\theta)\eta_\theta(\theta+\tilde\psi)\sin\tilde\psi\left\{\eta(\theta)-\frac{\eta(\theta+\tilde\psi)+\eta(\theta)}{2}\mathcal{B}\frac{d}{d\mathcal{B}}\right\}\Bigg)K_0\left(\mathcal{B}\sqrt{2-2\cos\tilde\psi}\right)d\tilde\psi.\end{aligned} \tag{A8}$$

Note that the relation

$$\sin\tilde\psi\mathcal{B}\frac{d}{d\mathcal{B}}K_0\left(\mathcal{B}\sqrt{2-2\cos\tilde\psi}\right)=(1-\cos\tilde\psi)\frac{d}{d\tilde\psi}K_0\left(\mathcal{B}\sqrt{2-2\cos\tilde\psi}\right),$$

together with integration by parts, has been used to simplify Eqs. (A6)–(A8) where possible.

The expressions (A6)–(A8) hold for any $\eta(\theta)$ and thus facilitate the expansion of $(\mathbf{u}_R\cdot\mathbf{n})_c$ in powers of ϵ for any boundary disturbance profile. The linear functional $(\mathbf{u}_R\cdot\mathbf{n})_c^l$ naturally appears in the integro-differential Eq. (14) describing the linear problem. The remaining quadratic and cubic functionals are required for the weakly nonlinear theory.

b. *Evaluation of the quadratic and cubic velocity functionals for wavelike perturbations*

Next, the aim is to evaluate the quadratic and cubic velocity functionals given by Eqs. (A7) and (A8) for the case where

$$\eta=\eta_1+\epsilon\eta_2+\epsilon^2\eta_3+O(\epsilon^3),$$

and η_1 is given by the wavelike solution to the leading-order equation in the weakly nonlinear expansion for the near-resonant problem:

$$\eta_1(\theta,T)=\mathrm{Re}A(T)e^{ik\theta}.$$

To proceed we adopt the notation $P_k=I_k(\mathcal{B})K_k(\mathcal{B})$ for the Bessel function product as introduced by Penfold et al. (2007) (see main text). The P_k appear directly in the Fourier expansion [Eq. (16)] and can be manipulated using the following two (independent) differentiation identities:

$$P_k'=\frac{\mathcal{B}}{2k}(P_{k+1}-P_{k-1}), \tag{A9}$$

$$\mathcal{B}(P_{k+1}'+P_k')=2kP_k-2(k+1)P_{k+1}, \tag{A10}$$

each of which helps facilitate the manipulation of our results into the form given below. Note that derivative terms can be eliminated (where necessary) between the expressions to give a recurrence relation. Penfold et al. (2007) also show that for all $k\geq 1$ the P_k can be shown to be positive, bounded, monotonically decreasing functions on $\mathcal{B}\geq 0$.

The quadratic functional (A7) can be evaluated for $\eta=\eta_1$ by replacing the K_0 Bessel function term with its Fourier series [Eq. (16)]. All of the terms in the integral reduce to elementary Fourier integrals and after some simplification the following result is obtained:

$$(\mathbf{u}_R\cdot\mathbf{n})_c^q[\eta_1]=\mathrm{Re}\frac{1}{4}ikA(T)^2e^{2ik\theta}[2(P_{2k}-P_k+P_1)+\mathcal{B}(2P_k'+P_{2k}'-P_1')]. \tag{A11}$$

Examination of the kinematic equation at next order in the weakly nonlinear expansion (29) reveals that

$$\eta_2 = \mathrm{Re}\beta_1|A|^2 + \beta_2 A(T)^2 e^{2ik\theta},$$

for constants β_1, β_2 (see discussion in the main text where their values are given). Nonzero η_2 means that the quadratic functional (A7) also contributes terms $O(\epsilon^3)$, which also reduce to Fourier integrals that on evaluation give

$$\begin{aligned}(\mathbf{u}_R\cdot\mathbf{n})_c^q[\eta_1,\eta_2] &= \frac{1}{4}\beta_2 ikA|A|^2 e^{ik\theta}[4P_k - 4P_{2k} + 2P_1 + \mathcal{B}(2P'_k + P'_{2k} - P'_1)]\\ &+ \frac{1}{4}\beta_2 ikA(T)^3 e^{3ik\theta}[6P_{3k} - 4P_{2k} - 2P_k + 6P_1 + 3\mathcal{B}(P'_k + P'_{2k} + P'_{3k} - P'_1)]\\ &+ \frac{1}{4}\beta_1 ikA|A|^2 e^{ik\theta}[\mathcal{B}(P'_k - P'_1)].\end{aligned} \tag{A12}$$

Finally, evaluation of the cubic functional $(\mathbf{u}_R\cdot\mathbf{n})_c^c[\eta_1]$ is required. A complicating factor is the term involving the factor $(1-\cos\tilde{\psi})^{-1}$ appearing in the integral, which does not allow for reduction to standard Fourier integrals. Instead, it is evaluated using the following identities (these results have been derived by transforming the integral to the unit circle in the complex plane, and have been verified numerically):

$$\int_0^{2\pi}\frac{[\cos\beta - \cos(k\tilde{\psi}+\beta)]^2}{1-\cos\tilde{\psi}}\cos m\tilde{\psi}\,d\tilde{\psi} = \pi\begin{cases}2(k-m) - m\cos 2\beta & 0\le m\le k\\ -(2k-m)\cos 2\beta & k<m\le 2k\\ 0 & m>2k,\end{cases}$$

$$\int_0^{2\pi}\frac{\sin(k\tilde{\psi}+\beta)[\cos\beta - \cos(k\tilde{\psi}+\beta)]^2}{1-\cos\tilde{\psi}}\cos m\tilde{\psi}\,d\tilde{\psi} = \frac{\pi}{2}\begin{cases}2k\sin\beta & 0\le m\le k\\ (4k-2m)\sin\beta + (k-m)\sin 3\beta & k<m\le 2k\\ (m-3k)\sin 3\beta & 2k<m\le 3k\\ 0 & m>3k.\end{cases}$$

Using the above results, at third order, the cubic functional can be evaluated to obtain

$$\begin{aligned}(\mathbf{u}_R\cdot\mathbf{n})^c[\eta_1] &= \mathrm{Re}\frac{1}{4}ikA|A|^2 e^{ik\theta}[P_k - P_{2k} - P_1 + \frac{\mathcal{B}^2}{4}(2P_{2k} - P_{2k+1} - P_{2k-1})\\ &+ \frac{\mathcal{B}}{4}(P'_1 - 6P'_k - P'_{2k}) - \frac{3}{2}(\mathcal{B}^2+1)P_1 + 2(\mathcal{B}^2+k^2)P_k - \frac{1}{2}(\mathcal{B}^2+4k^2)P_{2k}]\\ &+ \frac{1}{4}ikA(T)^3 e^{3ik\theta}[P_k - P_{2k} - P_1 + \frac{3\mathcal{B}^2}{4}(2P_{2k} - P_{2k+1} - P_{2k-1}) + \frac{\mathcal{B}}{4}(3P'_1 - 7P'_k + P'_{2k} + P'_{3k})\\ &-\frac{1}{2}(\mathcal{B}^2+1)P_1 + \frac{3}{2}(\mathcal{B}^2+k^2)P_k - \frac{3}{2}(\mathcal{B}^2+4k^2)P_{2k} + \frac{1}{2}(\mathcal{B}^2+9k^2)P_{3k}].\end{aligned}$$

The results above can be used as described in the main text to obtain the nonlinear frequency correction $\omega_2(\mathcal{B};k)$ given by Eq. (36).

c. *Limiting forms of the nonlinear frequency correction ω_2*

As discussed above, the coefficient $\omega_2(\mathcal{B};k)$ determines the magnitude of the nonlinear frequency correction for Rossby waves on the vortex, and its magnitude is important in determining the location of the bifurcation in parameter space. The exact expression (36) has been carefully checked numerically for several values of $\mathcal{B}$ and k and also contains two separate known limits that merit discussion.

The first relevant limit is the barotropic limit $\mathcal{B}\to 0$ in which the governing equations reduce to the 2D Euler equations. The weakly nonlinear analysis of the motion of a vortex patch governed by the 2D Euler equations has been studied previously by Su (1979). The results

$$\lim_{\mathcal{B}\to 0} P_k = \frac{1}{2k},\quad \lim_{\mathcal{B}\to 0} P'_k = O(\mathcal{B})$$

[Abramowitz and Stegun 1964, Eqs. (9.6.7)–(9.6.9)] can be used to show that

$$\lim_{\mathcal{B}\to 0} \omega_2(\mathcal{B};k) = -\frac{k(k-1)}{4},$$

in agreement with Su [1979; see Su's Eq. (16), which gives an equivalent result for a scaled angular speed $\lambda_2 = \pi\omega_2/k$].

The second limit is the "straight contour" limit, in which the wavelength of the Rossby waves is comparable to the Rossby radius but much smaller than the vortex radius. The problem of weakly nonlinear Rossby wave propagation on a straight (zonal) PV jump has been previously examined by Esler (2004) in the context of modulational or Benjamin–Feir instability. The results

$$\lim_{k\to\infty} P_k\left(\frac{k}{\gamma}\right) = \frac{\gamma}{2k(1+\gamma^2)^{1/2}}, \quad \lim_{k\to\infty} P'_k\left(\frac{k}{\gamma}\right) = -\frac{\gamma^2}{2k^2(1+\gamma^2)^{3/2}},$$
$$\lim_{k\to\infty} P_1\left(\frac{k}{\gamma}\right) = \frac{\gamma}{2k}, \qquad \lim_{k\to\infty} P'_1\left(\frac{k}{\gamma}\right) = -\frac{\gamma^2}{2k^2}$$

[see Abramowitz and Stegun 1964, their Eqs. (9.7.7)–(9.7.10) and (9.7.5)–(9.7.6)] can be used to access this limit. Specifically, the joint limit $k, \mathcal{B}\to\infty$ is required, with the ratio $k/B = k_x$ held constant. In this limit

$$\lim_{\substack{k,\mathcal{B}\to\infty \\ \mathcal{B}=k/k_x}} k^{-2}\omega_2(\mathcal{B};k) = -\frac{1}{16k_x}[4(1+k_x^2)^{1/2} - (1+4k_x^2)^{1/2} - 3],$$

recovering the corresponding result in Esler [2004, see Eq. (30) therein]. Here k_x is a nondimensional wavenumber, scaled on L_R^{-1}, and rescaling also explains the k^2 prefactor necessary in the limit, because in Esler (2004) the wave amplitude η is scaled on the inverse of the dimensional wavenumber ($k_x L_R = k^{-1}R$) as opposed to the vortex radius R used here. Note that only the first three terms in expression (36) contribute to the straight contour limit.

The fact that expression (36) for $\omega_2(\mathcal{B};k)$ contains the known results for the "barotropic" and straight contour in distinguished limits, and agrees with numerical results away from these limits, indicates that despite its somewhat opaque appearance and lengthy derivation [Eq. (36)] is likely to be correct.

REFERENCES

Abramowitz, M., and I. A. Stegun, 1964: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, 1046 pp.

Chao, W. C., 1985: Sudden stratospheric warmings as catastrophes. *J. Atmos. Sci.,* **42,** 1631–1646.

Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *J. Climate,* **20,** 449–469.

——, and Coauthors, 2007: A new look at stratospheric sudden warmings. Part II: Evaluation of numerical model simulations. *J. Climate,* **20,** 470–488.

Chen, P., 1996: The influences of zonal flow on wave breaking and tropical–extratropical interaction in the lower stratosphere. *J. Atmos. Sci.,* **53,** 2379–2392.

Dritschel, D. G., 1986: The nonlinear evolution of rotating configurations of uniform vorticity. *J. Fluid Mech.,* **172,** 157–182.

——, 1988: Contour surgery: A topological reconnection scheme for extended integrations using contour dynamics. *J. Comput. Phys.,* **77,** 240–266.

——, 1990: The stability of elliptical vortices in an external straining flow. *J. Fluid Mech.,* **210,** 223–261.

——, and R. Saravanan, 1994: Three-dimensional quasi-geostrophic contour dynamics, with an application to stratospheric vortex dynamics. *Quart. J. Roy. Meteor. Soc.,* **120,** 1267–1297.

Esler, J. G., 2004: Benjamin–Feir instability of Rossby waves on a jet. *Quart. J. Roy. Meteor. Soc.,* **130,** 1611–1630.

——, and R. K. Scott, 2005: Excitation of transient Rossby waves on the stratospheric polar vortex and the barotropic sudden warming. *J. Atmos. Sci.,* **62,** 3661–3682.

——, and N. J. Matthewman, 2011: Stratospheric sudden warmings as self-tuning resonances. Part II: Vortex displacement events. *J. Atmos. Sci.,* **68,** 2505–2523.

——, L. M. Polvani, and R. K. Scott, 2006: The Antarctic sudden stratospheric warming of 2002: A self-tuned resonance? *Geophys. Res. Lett.,* **33,** L12804, doi:10.1029/2006GL026034.

Hio, Y., and S. Yoden, 2007: A parameter sweep experiment on quasiperiodic variations of a polar vortex due to wave–wave interaction in a spherical barotropic model. *J. Atmos. Sci.,* **64,** 4069–4083.

Holton, J. R., and C. Mass, 1976: Stratospheric vacillation cycles. *J. Atmos. Sci.,* **33,** 2218–2225.

Kida, S., 1981: Motion of an elliptic vortex in a uniform shear flow. *J. Phys. Soc. Japan,* **50,** 3517–3520.

Limpasuvan, V., D. W. J. Thompson, and D. L. Hartmann, 2004: The life cycle of the Northern Hemisphere sudden stratospheric warmings. *J. Climate,* **17,** 2584–2596.

Love, A. E. H., 1893: On the stability of certain vortex motions. *Proc. London Math. Soc.,* **23,** 18–42.

Manney, G. L., J. D. Farrara, and C. R. Mechoso, 1994: Simulations of the February 1979 stratospheric sudden warming: Model comparisons and three-dimensional evolution. *Mon. Wea. Rev.,* **122,** 1115–1140.

——, W. A. Lahoz, R. Swinbank, A. O'Neill, P. M. Connew, and R. W. Zurek, 1999: Simulation of the December 1998 stratospheric major warming. *Geophys. Res. Lett.,* **26,** 2733–2736.

Matsuno, T., 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.,* **28,** 1479–1494.

Matthewman, N. J., J. G. Esler, A. J. Charlton-Perez, and L. M. Polvani, 2009: A new look at stratospheric sudden warmings. Part III: Polar vortex evolution and vertical structure. *J. Climate,* **22,** 1566–1585.

Melander, M. V., N. J. Zabusky, and A. S. Styczek, 1986: A moment model for vortex interactions of the two-dimensional Euler equations. Part 1. Computational validation of a Hamiltonian elliptical representation. *J. Fluid Mech.,* **167,** 95–115.

Mitchell, T. B., and L. F. Rossi, 2008: The evolution of Kirchhoff elliptic vortices. *Phys. Fluids,* **20,** 054103, doi:10.1063/1.2912991.

Nayfeh, A. H., and D. T. Mook, 1979: *Nonlinear Oscillations*. Wiley, 704 pp.

O'Neill, A., 2003: Stratospheric sudden warmings. *Encyclopedia of Atmospheric Sciences,* J. R. Holton, J. A. Pyle, and J. A. Curry, Eds., Elsevier, 1342–1353.

Penfold, R., J.-M. Vanden-Broeck, and S. Grandison, 2007: Monotonicity of some modified Bessel function products. *Integr. Transforms Spec. Funct.,* **18,** 139–144.

Plumb, R. A., 1981a: Forced waves in a baroclinic shear flow. Part 2: Damped and undamped response to weak near-resonant forcing. *J. Atmos. Sci.,* **38,** 1856–1869.

——, 1981b: Instability of the distorted polar night vortex: A theory of stratospheric warmings. *J. Atmos. Sci.,* **38,** 2514–2531.

Polvani, L. M., and R. A. Plumb, 1992: Rossby wave breaking, microbreaking, filamentation, and secondary vortex formation: The dynamics of a perturbed vortex. *J. Atmos. Sci.,* **49,** 462–476.

——, and R. Saravanan, 2000: The three-dimensional structure of breaking Rossby waves in the polar wintertime stratosphere. *J. Atmos. Sci.,* **57,** 3663–3685.

Rong, P.-P., and D. W. Waugh, 2004: Vacillations in a shallow-water model of the stratosphere. *J. Atmos. Sci.,* **61,** 1174–1185.

Saffman, P. G., 1992: *Vortex Dynamics*. Cambridge University Press, 311 pp.

Scott, R. K., and L. M. Polvani, 2006: Internal variability of the winter stratosphere. Part I: Time-independent forcing. *J. Atmos. Sci.,* **63,** 2758–2776.

Smith, A. K., 1989: An investigation of resonant waves in a numerical model of an observed sudden stratospheric warming. *J. Atmos. Sci.,* **47,** 3038–3054.

Su, C. H., 1979: Motion of fluid with constant vorticity in a singly-connected region. *Phys. Fluids,* **22,** 2032, doi:10.1063/1.862502.

Swanson, K. L., 2000: Stationary wave accumulation and the generation of low-frequency variability on zonally varying flows. *J. Atmos. Sci.,* **57,** 2262–2280.

Tung, K. K., and R. S. Lindzen, 1979a: A theory of stationary long waves. Part I: A simple theory of blocking. *Mon. Wea. Rev.,* **107,** 714–734.

——, and ——, 1979b: A theory of stationary long waves. Part II: Resonant Rossby waves in the presence of realistic vertical shears. *Mon. Wea. Rev.,* **107,** 735–750.

Watson, G. N., 1944: *A Treatise on the Theory of Bessel Functions*. 2nd ed. Cambridge University Press, 804 pp.

Waugh, D. W., 1997: Elliptical diagnostics of stratospheric polar vortices. *Quart. J. Roy. Meteor. Soc.,* **123,** 1725–1748.

——, and D. G. Dritschel, 1999: The dependence of Rossby wave breaking on the vertical structure of the polar vortex. *J. Atmos. Sci.,* **56,** 2359–2375.

Yoden, S., 1987: Bifurcation properties of a stratospheric vacillation model. *J. Atmos. Sci.,* **44,** 1723–1733.