The spontaneous generation of inertia–gravity waves during frontogenesis forced by large strain: theory

Callum J. Shakespeare[1] **and J. R. Taylor**[1],†

[1]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

Density fronts are common features of ocean and atmosphere boundary layers. Field observations and numerical simulations have shown that the sharpening of frontal gradients, or frontogenesis, can spontaneously generate inertia–gravity waves (IGWs). Although significant progress has been made in describing frontogenesis using approximations such as quasi-geostrophy (Stone, *J. Atmos. Sci.*, vol. 23, 1966, pp. 455–565, Williams & Plotkin *J. Atmos. Sci.*, vol. 25, 1968, pp. 201–206) semi-geostrophy (Hoskins, *Annu. Rev. Fluid Mech.*, vol. 14, 1982, pp. 131–151), these models omit waves. Here, we further develop the analytical model of Shakespeare & Taylor (*J. Fluid Mech.*, vol. 736, 2013, pp. 366–413) to describe the spontaneous emission of IGWs from an initially geostrophically balanced front subjected to a time-varying horizontal strain. The model uses the idealised configuration of an infinitely long, straight front and uniform potential vorticity (PV) fluid, with a uniform imposed convergent strain across the front, similar to Hoskins & Bretherton (*J. Atmos. Sci.*, vol. 29, 1972, pp. 11–37). Inertia–gravity waves are generated via two distinct mechanisms: acceleration of the large-scale flow and frontal collapse. Wave emission via frontal collapse is predicted to be exponentially small for small values of strain but significant for larger strains. Time-varying strain can also generate finite-amplitude waves by accelerating the cross-front flow and disrupting geostrophic balance. In both cases waves are trapped by the oncoming strain flow and can only propagate away from the frontal zone when the strain field weakens sufficiently, leading to wave emission that is strongly localised in both time and space.

Key words: atmospheric flows, ocean processes, waves in rotating fluids

---

1. Introduction

Inertia–gravity waves (IGWs) are of fundamental importance in both the atmosphere and ocean due to their ability to transfer significant amounts of energy, momentum and other tracer properties. For example, in the ocean IGWs provide a mechanism for energy loss from large-scale flows (e.g. Williams, Haine & Read 2008) and can drive significant vertical mixing (e.g. Polzin 2010). In the atmosphere, these waves

† Email address for correspondence: J.R.Taylor@damtp.cam.ac.uk

have the ability to initiate and modulate convection (Zhang 2004). Substantial sources of IGWs include flow over topography and moist convection in the atmosphere (Plougonven & Zhang 2014), and forcing by tides and surface wind stresses in the ocean (Wunsch & Ferrari 2004). Recent work suggests that fronts, or regions of large horizontal density gradient, and their associated geostrophic jets, are a further important source of IGWs (e.g. Alford, Shcherbina & Gregg 2013; Plougonven & Zhang 2014). Despite the importance of IGWs, their generation mechanisms at fronts are still not fully understood. There are two broad mechanisms of IGW generation at fronts that are widely discussed in the literature: geostrophic adjustment (e.g. Rossby 1938; Blumen 1972; Ou 1984; Tandon & Garrett 1994; Blumen 2000; Shakespeare & Taylor 2013) and spontaneous emission (e.g. Ford, McIntyre & Norton 2000; Viudez & Dritschel 2006; Plougonven & Zhang 2007; Danioux *et al.* 2012). In geostrophic adjustment, the system is instantaneously displaced from geostrophic balance and then adjusts back towards this state, releasing energy in the form of IGWs in the process (e.g. Blumen 1972). On the other hand, spontaneous emission involves the generation of IGWs from an initially 'balanced' (i.e. long time scale, slow manifold) flow. Geostrophic adjustment relies on an imposed external influence or 'initial condition', whereas spontaneous emission is a fundamental transient feature of the dynamical equations. The process of spontaneous emission is therefore expected to be ubiquitous in the ocean and atmosphere, and is arguably more important in terms of the global wavefield and associated tracer fluxes (Plougonven & Zhang 2014). Unfortunately, spontaneous wave emission is also far more difficult to quantify since, by definition, it involves the breakdown of the balanced approximations (e.g. geostrophy, quasi-geostrophy, semi-geostrophy) that form the basis of our understanding of most geophysical flows (Vanneste 2013). Zhang (2004), among others, proposed that the mechanism of spontaneous emission can be considered as a generalisation of geostrophic adjustment, called 'balance adjustment'. In this paradigm, departures from a suitably balanced state are treated as internally forced imbalances to which the system adjusts via the emission of waves, as per the classical geostrophic adjustment scenario. As such, the two mechanisms of wave generation at fronts, geostrophic adjustment and spontaneous emission, may be closely related – a topic investigated further in the present work.

Baroclinic instability is a major source of spontaneous IGWs in both the atmosphere and the ocean (Vanneste 2013), which exhibit large-scale baroclinic waves and eddy fields. Sharp horizontal buoyancy gradients (fronts) and strain flows are ubiquitous features of these fields. A convergent strain can act to amplify relatively weak buoyancy gradients through a process known as frontogenesis (e.g. Hoskins & Bretherton 1972). This strain-driven sharpening of the buoyancy gradients ultimately causes a breakdown of geostrophic balance (Juckes 1994) and the generation of IGWs as the frontal scale collapses (Snyder, Skamarock & Rotunno 1993). In addition, the eddies are themselves naturally time-dependent according to a 'baroclinic life-cycle' and thus cause the surrounding strain field to accelerate and decelerate with time. Eddy fields consequently exhibit highly variable strains in time and space, and this variability can itself generate IGWs even in the absence of a sharp frontal gradient (Snyder *et al.* 1993; Viudez & Dritschel 2006; Vanneste 2013). Williams *et al.* (2008) measured the generation of IGWs during a baroclinic life-cycle in a rotating two-layer annulus experiment at small (∼0.1) Rossby numbers. They found that the baroclinic eddies leak ∼1 % of their energy to IGWs each inertial cycle. Alford *et al.* (2013) obtained a result with a similar order of magnitude for the energy loss via wave emission from an actively strained, mesoscale front observed in the ocean mixed

layer. Collectively, these studies suggest that wave generation at strained fronts within eddy fields may provide a significant contribution to the ocean energy budget.

Given the complexity of fully three-dimensional baroclinic instability discussed above, it has proven useful to use two-dimensional (2D) models to examine small-scale flow evolution within a large-scale eddy field. Hoskins & Bretherton (1972, hereafter, HB) introduced a 2D, rigid-lid model to study the effect of convergent strain within a baroclinic wave on the evolution of a smaller-scale buoyancy gradient (a front). They derived an analytical solution for the secondary flow in the limit of small strain, using the semi-geostrophic equations. The semi-geostrophic approximation requires that the cross-front acceleration is much slower than the inertial acceleration, i.e. $\mathscr{D}u \ll fv$ where $\mathscr{D}$ represents the material derivative. Time derivatives are therefore neglected in the semi-geostrophic limit, and thus IGWs are filtered from the solution. While the semi-geostrophic approximation is valid (and IGW generation negligible) for many flows, it breaks down for unbalanced initial conditions and strongly strained flows where the horizontal strain rate α is of similar order to the inertial frequency f; one example is at submesoscale fronts in the ocean mixed layer (Shcherbina *et al.* 2013).

Among its various applications, the HB solution is a convenient balanced state for the study of spontaneous wave emission. Ley & Peltier (1978) computed the neglected 'unbalanced' ageostrophic flow arising from the model and used it to force a correction to the HB solution. This correction was found to take the form of a packet of IGWs, which produced a pressure minimum at a fixed distance ahead of the front. Ley & Peltier proposed that such wave emission provides a mechanism for the formation of squall lines in the warm sector ahead of an advancing cold front – such squall lines are a common feature of atmospheric weather systems (e.g. Karan *et al.* 2010). Garner (1989) and Snyder *et al.* (1993) studied the difference between the HB solution and numerical primitive equation models, and found it to be dominated by a field of largely stationary IGWs. These waves were attributed to three mechanisms: generation via the geostrophic adjustment of initial conditions of semi-geostrophic balance (mechanism 1), implying the existence of some more balanced primitive equation solution (Garner 1989), generation via a time-dependent strain field (mechanism 2) and generation via frontal collapse (mechanism 3). Griffiths & Reeder (1996) and Reeder & Griffiths (1996) obtained similar results from their numerical model of upper-level frontogenesis and, in particular, found that wave generation becomes more pronounced as frontogenesis varies rapidly. Despite these efforts, a theoretical model for spontaneous wave generation from this idealised 2D system via the above mechanisms (2, 3) is still lacking.

Shakespeare & Taylor (2013, hereafter, ST13) examined frontogenesis in a quasi-2D, rigid-lid model of a frontal system, as initially introduced by HB and discussed above. A key feature of their model was the introduction of the 'generalised momentum coordinate', which allowed the unification of previous work by Blumen & Wu (1995), Blumen (2000) and others, describing geostrophic adjustment and wave emission at unbalanced fronts, with that of HB describing the evolution of balanced fronts subjected to a convergent strain flow. The generalised model formulated in ST13 is able to simultaneously represent strain-driven frontal collapse and the generation and propagation of IGWs. Unlike HB, no assumption was made as to the smallness of the cross-front acceleration. However, ST13 focused on the breakdown of the HB model due to unbalanced initial conditions at small strains, $\delta = \alpha/f \sim 0.1$, and the attendant wave generation via geostrophic adjustment (mechanism 1). In that limit, the generalised model solution is composed of an IGW field driving oscillations

about a time-varying mean state given by the HB solution. The strain flow traps the generated IGWs in the frontal zone and drives the decay of the wave amplitude with time.

Here, we will consider the breakdown of semi-geostrophy associated with larger strain fields, but purely balanced initial conditions. Since semi-geostrophic balance breaks down for large strain, the requirement of balanced initial conditions will necessitate the expansion of the ST13 model to incorporate a time-dependent strain field, such that the model can be initialised with zero strain in a geostrophically balanced state. The analytical solution presented in ST13 also required an *ad hoc* linearisation of the horizontal momentum equations, which was shown to be valid *a posteriori* for the cases considered therein. Here, we will more rigorously consider the system via a scaling analysis of the governing equations, assuming balanced initial conditions, and determine the parameter values for which the neglect of the nonlinear terms is valid. We find that the neglected nonlinear terms are small compared to the leading-order retained terms wherever the product of the Rossby number and non-dimensional strain, $Ro\,\delta$, is sufficiently small. Two subsets of the generalised model will be investigated: first, the limit of small strain δ, corresponding to the semi-geostrophic solution of HB, which does not permit IGWs; and secondly, the limit of small Rossby number Ro, which permits large strain and accurately describes second-order effects such as wave generation. However, in contrast to ST13, here we will only consider 'spontaneously generated' waves, which we define in the present context as waves generated at an initially balanced front subject to an imposed convergent strain. Such waves will arise both through time-variation of the strain field, even for an arbitrarily weak buoyancy gradient (mechanism 2), as well as for a constant strain in the limit of frontal collapse (mechanism 3). The spatial structure of the generated wavefield in each case is strongly influenced by the effect of the large-scale strain flow on wave propagation, as argued by Plougonven & Snyder (2005).

The layout of the paper is as follows. In § 2 we introduce the governing equations following ST13, generalised here to include a time-dependent strain field. A scaling analysis is performed (§ 2.1) to identify the parameter values for which the nonlinear terms in the governing equations may be neglected. In § 3, we consider the special case of zero potential vorticity (PV) to show that time-variation in the large-scale strain field drives inertial oscillations about the geostrophically balanced state. Then, in § 4, the more general uniform PV flow is solved for constant strain using a Green's function method. We demonstrate the tendency of the flow to develop a strong stationary wavefield as the front collapses, and separate the flow into a wave and secondary circulation component. In § 5, we consider the most general case of uniform PV and a time-dependent strain field, using the results from previous sections. Lastly, in § 6 we discuss the implications of our results in terms of frontal evolution and energy loss from balanced flows.

2. Model equations

Following ST13, we consider a uniform PV flow trapped between rigid lids and subject to a strain field $(U, V, 0) = (-\alpha x, \alpha y, 0)$ in the Cartesian (x, y, z) directions. The action of the strain drives a secondary flow (u, v, w). The model is quasi-2D in the sense that along-front (y) gradients are neglected. ST13 assumed that the strain α was constant in time – here, we permit the strain to vary in time to allow the model to describe a greater range of geophysical situations. The key feature of the model is

the use of the generalised momentum coordinate, defined (for a time-dependent strain) by the transformation

$$X = e^{\beta(t)} \left(x + \frac{v}{f} \right), \quad Z = z, \quad T = t, \quad (2.1a\text{–}c)$$

where f is the inertial frequency and, following HB, $\beta(t)$ is the time-integrated strain,

$$\beta(t) = \int_0^t \alpha(t') \, dt', \quad (2.2)$$

which is non-dimensional. We will write the strain parameters as explicit functions of time (i.e. $\alpha = \alpha(t)$, $\beta = \beta(t)$) where appropriate throughout this work to prevent confusion with the constant-strain case. The coordinate X defined in (2.1) is conserved for any strain $\alpha(T)$; that is,

$$\frac{DX}{Dt} = 0. \quad (2.3)$$

All derivatives and other properties of the coordinate transformation are listed in table 1 of ST13, where $\exp \delta T$ must be replaced by $\exp \beta(T)$ for a time-dependent strain field – the expressions are otherwise identical. Following ST13, the incompressible, inviscid, Boussinesq, rotating fluid equations, when transformed to generalised momentum coordinates, become

$$\frac{\partial u}{\partial T} + Ro\, w\, \frac{\partial u}{\partial Z} - v - \delta(T)\, u + v_g = 0, \quad (2.4a)$$

$$\frac{\partial v}{\partial T} + Ro\, w\, \frac{\partial v}{\partial Z} + u + \delta(T)\, v = 0, \quad (2.4b)$$

$$\frac{\partial b}{\partial T} + Ro\, w\, \frac{\partial b}{\partial Z} = 0, \quad (2.4c)$$

plus continuity (see table 1 of ST13) and hydrostatic balance. The geostrophic velocity v_g is defined by

$$v_g = \frac{1}{\rho_0 f} \frac{\partial p}{\partial x}, \quad (2.5)$$

for reference density ρ_0 and perturbation pressure p. Equation (2.5) and hydrostatic balance give rise to the thermal wind equation, which in momentum coordinates becomes

$$\frac{\partial v_g}{\partial Z} - Ro\, e^{\beta(T)} \left(\frac{\partial b}{\partial X} + \frac{\partial v}{\partial X} \frac{\partial v_g}{\partial Z} - \frac{\partial v}{\partial Z} \frac{\partial v_g}{\partial X} \right) = 0. \quad (2.6)$$

Conservation of PV, q, may be derived from (2.4) and continuity as

$$\frac{\partial q}{\partial T} + Ro\, w\, \frac{\partial q}{\partial Z} = 0, \quad q = \frac{\partial b}{\partial Z} \left(1 - e^{\beta(T)} Ro \frac{\partial v}{\partial X} \right)^{-1}. \quad (2.7)$$

The assumption of uniform PV then implies that

$$\frac{\partial b}{\partial Z} - q_0 \left(1 - e^{\beta(T)} Ro \frac{\partial v}{\partial X} \right) = 0, \quad (2.8)$$

Buoyancy scale		ΔB
Horizontal velocity scale		$\sqrt{\Delta B H}$
Vertical velocity scale		$\sqrt{(\Delta B H^3)/L^2}$
Time scale		$1/f$
Rossby number	Ro	$(\sqrt{\Delta B H}/fL)$
Froude number	F	$\sqrt{(\Delta B/N^2 H)}$
Deformation ratio (or large-scale Rossby number)	$\delta(T)$	$(\alpha(T))/f$
Burger number (Ro/F)	Bu	(NH/fL)

TABLE 1. The non-dimensional parameters and variable scales employed herein. The fundamental physical scales are the inertial frequency f, the horizontal buoyancy step across the front ΔB, the buoyancy frequency N, the strain α, the initial horizontal width L and the height H.

where the uniform PV, q_0, can be written in terms of the Froude number, $q_0 = F^{-2}$ (see table 1). The equations have been non-dimensionalised using the same scales as ST13, which are also listed in table 1 for reference. The physical scales are the initial width of the frontal zone L, the height H, the buoyancy transition across the front ΔB, the background stratification N^2, the inertial frequency f and the strain $\alpha(T)$. Following ST13, the time evolution of the hydrostatic system ($L/H \gg 1$) is thus entirely controlled by the three independent non-dimensional parameters; the Rossby number $Ro = \sqrt{(\Delta B H)}/(fL)$, the non-dimensional strain $\delta(T) = \alpha(T)/f$ and the Burger number $Bu = (NH)/(fL)$ (also see table 1).

Of the three independent non-dimensional parameters (Ro, Bu, δ), the buoyancy difference ΔB across the front only appears in the Rossby number. Here, Ro thus provides a non-dimensional measure of the size of the buoyancy step $b_0(X)$ imposed on the rigid lids. Indeed, the parameter Ro is equivalent to the 'nonlinearity parameter' governing the relative surface height displacement in shallow-water step-adjustment problems (e.g. Kuo & Polvani 1997). We observe that while the convergent strain will act to increase horizontal gradients (i.e. by reducing the frontal width with time), it will not alter the net magnitude of buoyancy transition ΔB across the front. A small Ro therefore implies an initially weak front, but does not preclude the formation of a sharp/strong front with time as the frontal width contracts. While the frontal width L remains constant in generalised momentum coordinates, the frontal width in Eulerian coordinates (i.e. the actual width) is $L_E = L J^{-1}$, where J^{-1} is the inverse Jacobian of the coordinate transformation:

$$J^{-1} = \frac{\partial x}{\partial X} = \left(e^{-\beta(T)} - Ro \frac{\partial v}{\partial X} \right). \qquad (2.9)$$

The physical/frontal Rossby number at a given time is therefore $Ro_f = Ro\,J$, which grows with time, and becomes infinite in the limit of discontinuity formation. For balanced initial conditions, the along-front velocity v will be non-zero initially, implying that the frontal Rossby number Ro_f will exceed Ro even at time zero.

The background stratification N^2 only appears in the Burger number Bu, and thus Bu is the non-dimensional scale that governs the speed of wave propagation and adjustment in the system or, equivalently, the background PV (dimensionally, $q_0 = N^2$

with our definition of PV; see table 1 of ST13). The non-dimensional strain $\delta(T)$ governs the magnitude of the imposed large-scale convergent flow field. In fact, $\delta(T)$ may be interpreted as a large-scale Rossby number Ro_L characterising the large-scale strain flow:

$$Ro_L = \left| \frac{1}{f} \frac{\partial U}{\partial x} \right| = \alpha(T)/f = \delta(T). \tag{2.10}$$

A small strain δ thus implies that the large-scale phenomenon that is responsible for generating the strain field (such as a baroclinic eddy field or wave) is characterised by a small Rossby number.

Perhaps the major advantage of generalised momentum coordinates is the simplicity of the buoyancy (2.4c) and PV conservation (2.8) equations. Indeed, these two equations, plus the rigid-lid boundary condition $w = 0$ and $b = b_0(X)$ (for arbitrary b_0) at $Z = 0, 1$, are sufficient to linearly relate the perturbation along-front velocity v, the cross-front streamfunction ψ and the buoyancy field b. The xz-plane velocity components may be determined from ψ in the usual way, via

$$u = \frac{\partial \psi}{\partial z} \quad \text{and} \quad w = -\frac{\partial \psi}{\partial x}. \tag{2.11a,b}$$

It will prove convenient to express the relations between the field variables v, ψ and b in terms of the integral of the along-front velocity,

$$\phi(X, Z, T) = \int_0^Z v(X, Z', T)\, \mathrm{d}Z', \tag{2.12}$$

such that

$$v(X, Z, T) = \frac{\partial}{\partial Z} \phi(X, Z, T), \tag{2.13a}$$

$$b(X, Z, T) = b_0(X) + F^{-2}Z - Ro\, F^{-2} \mathrm{e}^{\beta(T)} \frac{\partial}{\partial X} \phi(X, Z, T), \tag{2.13b}$$

$$\psi(X, Z, T) = \left(-\delta(T) - \frac{\partial}{\partial T} \right) \phi(X, Z, T). \tag{2.13c}$$

The relations (2.13b), (2.13c) have been modified here (cf. (2.27) and (2.30) of ST13) to incorporate a time-dependent strain. Note that (2.13) depend only on buoyancy and PV conservation, and thus apply independent of any dynamical balance assumption. The entire time evolution of the system is thus known if $\phi(X, Z, T)$ can be determined subject to boundary conditions of $\phi(X, 0, T) = \phi(X, 1, T) = 0$ to enforce $w = -\partial_x \psi = 0$ on the rigid lids. Applying the horizontal momentum (2.4a), (2.4b), thermal wind (2.6) and PV conservation (2.8) equations, in addition to relations (2.13a) and (2.13b), one may write the evolution of ϕ as

$$\left[\left(\frac{\partial^2}{\partial T^2} + 1 - \delta(T)^2 + \delta'(T) \right) \frac{\partial^2}{\partial Z^2} + Bu^2 \mathrm{e}^{2\beta(T)} \frac{\partial^2}{\partial X^2} \right] \phi(X, Z, T) = \mathscr{F} + \mathscr{N}. \tag{2.14}$$

The function $\mathscr{F}$ on the right-hand side of (2.14) is the linear forcing defined by

$$\mathscr{F} = Ro\, b_0'(X)\, \mathrm{e}^{\beta(T)}, \tag{2.15}$$

where $b_0'(X) = \partial_X b \mid_{Z=0,1}$ (from (2.13b)) is the imposed buoyancy gradient at the rigid lids that forces motion in the interior as it is amplified by the convergent strain field. The function $\mathcal{N}$ on the right-hand side of (2.14) is the sum of the nonlinear terms:

$$\mathcal{N} = Ro \frac{\partial}{\partial Z} \left[w \frac{\partial u}{\partial Z} + \left(\delta - \frac{\partial}{\partial T} \right) \left(w \frac{\partial v}{\partial Z} \right) \right] - Ro\, e^{\beta(T)} \left[\frac{\partial v}{\partial X} \frac{\partial v_a}{\partial Z} - \frac{\partial v}{\partial Z} \frac{\partial v_a}{\partial X} \right], \quad (2.16)$$

where $v_a \equiv v - v_g$ is the ageostrophic velocity. Similar to ST13, we will neglect the explicit nonlinear terms. In the next section, we perform a scaling analysis to determine under what conditions the neglect of $\mathcal{N}$ in (2.14) is valid.

2.1. *Scaling analysis*

Unlike ST13, here we will only consider initial conditions that are appropriately 'balanced' such that the time-variation in the solution is controlled by the strain forcing, implying that the time derivative ∂_T 'scales with' the strain δ – here denoted as $\partial_T \sim \delta$ – rather than the natural wave response of the system. Applying this result and relations (2.13), we can write scales for the velocity field and streamfunction as

$$v \sim \Phi, \quad \psi \sim \delta\, \Phi, \quad (2.17a,b)$$

assuming that $\phi \sim \Phi$. In determining further scales we will neglect any nonlinear terms, and then employ the linearised scales to determine under what conditions the neglect is valid. Using (2.17), scales for the velocities $u = \partial_z \psi$ and $w = -\partial_x \psi$ may be written as

$$u \sim \delta\, \Phi, \quad w \sim \delta\, e^{\beta(T)}\, \Phi, \quad (2.18a,b)$$

since $\partial_x \sim e^{\beta(T)}$. Substitution of (2.18) into the *x*-momentum equation (2.4a) yields a scale for the ageostrophic velocity,

$$v_a \sim \delta^2\, \Phi, \quad (2.19)$$

neglecting the contribution of the nonlinear advection term. Using these results, it may be shown that the nonlinear terms $\mathcal{N}$ (2.16) scale as

$$\mathcal{N} \sim Ro\, \delta^2\, e^{\beta(T)}\, \Phi^2, \quad (2.20)$$

the linear forcing $\mathcal{F}$ (2.15) as

$$\mathcal{F} \sim Ro\, e^{\beta(T)} \quad (2.21)$$

and the left-hand side (LHS) of (2.14) as

$$\text{LHS} \sim \Phi + \delta^2\, \Phi + Bu^2\, e^{2\beta(T)}\, \Phi. \quad (2.22)$$

Neglecting $\mathcal{N}$, the magnitude of ϕ must be determined by a balance between the linear forcing (2.21) and the LHS (2.22):

$$\Phi \sim \frac{Ro\, e^{\beta(T)}}{1 + \delta^2 + Bu^2\, e^{2\beta(T)}}. \quad (2.23)$$

We can now determine under what conditions the neglect of the nonlinear terms is valid. For a first estimate of model validity, we require that the summed magnitude

of the linear terms (LHS) greatly exceeds that of the neglected nonlinear terms; that is,

$$(1 + \delta^2 + Bu^2\, e^{2\beta(T)})\, \Phi \gg Ro\, \delta^2\, e^{\beta(T)}\, \Phi^2, \tag{2.24}$$

or, substituting (2.23),

$$\mathcal{E} = \frac{Ro^2\, \delta^2\, e^{2\beta(T)}}{\left(1 + \delta^2 + Bu^2\, e^{2\beta(T)}\right)^2} \ll 1, \tag{2.25}$$

where $\mathcal{E}$ is a measure of the relative fractional error in neglecting the nonlinear terms. Thus, for the neglect of the nonlinear terms to be valid at time zero, we must have

$$Ro^2\, \delta^2 \ll \left(1 + \delta^2 + Bu^2\right)^2. \tag{2.26}$$

In other words, the product of Ro and δ must be sufficiently small for the model to be valid at time zero. The relative error $\mathcal{E}$ will be less for larger Bu. Equation (2.26) represents the weakest possible limit on model validity. If the leading-order terms on the left-hand side of (2.14) balance the forcing such that the scaling in (2.26) holds, then the large-scale circulation predicted by the model should be accurate, but smaller-amplitude features such as higher-mode waves may not be well described by the model. On the other hand, if the leading-order terms on the left-hand side of (2.14) happen to cancel, the neglected terms could play a more significant role, and the model solution may not be valid. A more stringent constraint on model validity can be obtained by requiring that the magnitude of the smallest linear term (i.e. in (2.22)) exceeds that of the largest neglected nonlinear term:

$$\delta^2\, \Phi \gg Ro\, \delta^2\, e^{\beta(T)}\, \Phi^2, \tag{2.27}$$

or, substituting the scale for ϕ (2.23),

$$\frac{Ro^2\, e^{2\beta(T)}}{1 + \delta^2 + Bu^2\, e^{2\beta(T)}} \ll 1. \tag{2.28}$$

Thus, for the neglect of the nonlinear terms to be valid at time zero in this more stringent limit, we must have

$$Ro^2 \ll 1 + \delta^2 + Bu^2. \tag{2.29}$$

In other words, the Rossby number Ro must be sufficiently small for validity at time zero, where the term 'validity' in this more stringent limit implies that both the large-scale circulation and smaller-scale features will be well described by the model.

We have derived constraints on the parameter values for the validity of the neglect of the nonlinear terms at time zero, assuming suitably balanced initial conditions. We now consider the validity at later times. First, note that the coordinate system (and model) breaks down when the inverse Jacobian of the momentum coordinate transformation (2.9) vanishes. The vanishing of J^{-1} implies the formation of a discontinuity in velocity and buoyancy fields. The inverse Jacobian vanishes when ϕ grows sufficiently large; that is, from (2.9),

$$\Phi \sim Ro^{-1}\, e^{-\beta(T)}, \tag{2.30}$$

using the scales derived above. We observe that the constraint on model validity (2.27) in the more stringent limit discussed above may be rewritten as

$$\Phi \ll Ro^{-1} e^{-\beta(T)}. \tag{2.31}$$

Thus, if (2.29) holds at $T = 0$, $\Phi \ll Ro^{-1}$, then the model is valid except near the time or location of discontinuity formation, when $e^{\beta(T)} \gg 1$. Similarly, the weak constraint on model validity (2.24) may be rewritten as

$$\Phi \ll Ro^{-1} e^{-\beta(T)}(1 + \delta^{-2}(1 + Bu^2)). \tag{2.32}$$

Thus, given that the initial condition (2.26) is satisfied, the model is valid except near the time or location of discontinuity formation. Further, if δ is small, then the neglect of the nonlinear terms will be valid even in the limit of discontinuity formation.

2.2. *The generalised model*

Given the above scaling analysis, the neglect of the nonlinear terms in (2.14) is valid – in the sense that these terms are small compared to leading-order linear terms – as long as the product $Ro\,\delta$ is sufficiently small, as defined by (2.26). As such, we call (2.14) with $\mathcal{N} \equiv 0$ the 'generalised model'. For constant strain, (2.14) with $\mathcal{N} \equiv 0$ is equivalent to (2.34) of ST13. A vital feature of the generalised model is that it permits the propagation of IGWs (as a result of the explicit time derivative in (2.14)). The weak constraint on model validity (2.26) ensures that the large-scale secondary circulation associated with frontogenesis is well described, but second-order effects, such as wave generation, may not be. The relative size of neglected effects may be estimated from (2.25). If we instead consider the limit of small Rossby number, defined by (2.29), both first- and second-order flow will be well described by the generalised model – neglected effects are strictly third order. The physical interpretation of this small Rossby limit is that of an initially weak front, somewhat analogous to the small-step limit of shallow-water free-surface height adjustment problems. The front is only weak at early times and will sharpen with time as the frontal width contracts (e.g. the frontal Rossby number is $Ro_f = Ro\,J$, as discussed previously), although the above scaling arguments suggest that the generalised model may break down where $O(Ro_f) = 1$. Perhaps the most important feature of this small Rossby limit is that it permits an accurate analytical description of wave generation occurring at a strained front, since the included time derivative and strain terms in (2.14) that are associated with wave generation are guaranteed to exceed neglected terms in this limit.

One important subset of the generalised model is the limit of weak strain, $\delta^2 \ll 1$. In this limit, the generalised model equation (2.14) may be further simplified by the neglect of the $O(\delta^2)$ terms on the left-hand side, yielding

$$\left[\frac{\partial^2}{\partial Z^2} + Bu^2\, e^{2\beta(T)} \frac{\partial^2}{\partial X^2} \right] \phi = \mathcal{F}, \tag{2.33}$$

where $\mathcal{F}$ is given by (2.15), as previously. This derivation assumes that the time-variation in the strain is also small, $\delta'(T) \ll 1$. Equation (2.33) is identical to the 'HB model' of frontogenesis (Hoskins & Bretherton 1972). The neglect of the $O(\delta^2)$ terms (including the time derivative) in (2.14) filters IGWs from the solution. As δ is small, the relative error $\mathcal{E}$ associated with the neglected nonlinear terms (2.25) will be

small. Given this reduction of our model to the HB model in the limit of small strain, our model can be described as a generalisation of the HB model permitting large strain rates and IGWs. The physical interpretation of the weak-strain limit is that the large-scale flow is strongly affected by rotation, or $Ro_L^2 \ll 1$, where Ro_L is the Rossby number of the large-scale flow. While we model the (dimensional) large-scale flow as $(-\alpha x, \alpha y, 0)$ in the region of the front, this is only an approximation to a flow that would have significant spatial and temporal variability. For instance, in the atmosphere the 'large-scale flow' might be associated with locally convergent regions of a global-scale baroclinic wave (e.g. Hoskins 1982). In the ocean, the 'large-scale flow' could be associated with convergence within a mesoscale eddy field (e.g. Thomas, Tandon & Mahadevan 2008). In each case, δ is the Rossby number describing the dynamics of that spatially and temporally varying larger-scale flow field. For a global baroclinic wave in the atmosphere or a mesoscale ocean eddy field, $Ro_L \sim 0.1$ and this $\delta^2 \ll 1$ limit is valid.

Given the influence of the Rossby number Ro on the validity of the generalised model, it is useful to consider the effect of the Rossby number on the dynamics. As already observed, a larger Rossby number Ro implies a stronger initial front and consequently more rapid discontinuity formation and model breakdown. Furthermore, the Rossby number appears in the governing equation (2.14) in two places: (i) in the momentum coordinate transformation, $X = e^{\beta(T)}(x + Ro\,v)$; and (ii) in the forcing $\mathscr{F}$ on the right-hand side (defined by (2.15)). As shown in ST13, the appearance of the Rossby number in the coordinate transformation is related to the tilting/slumping of the front during frontogenesis; the position of the front on the boundary will be $O(Ro)$ by the time of model breakdown. A small Rossby number thus implies that the front remains relatively upright during frontogenesis due to rotational effects dominating the gravity-driven tendency for frontal slumping. The Rossby number in the forcing term $\mathscr{F}$ acts to scale the amplitude of the flow response to the applied strain. As expected, stronger fronts lead to a larger amplitude secondary circulation.

Before proceeding further and solving the generalised model in specific cases, we can gain significant insight into the dynamics from direct examination of the PDE (2.14) with $\mathscr{N} = 0$. Using the fact that ϕ must vanish on the boundaries at $Z = 0, 1$, we can write ϕ as a sum of vertical sine modes:

$$\phi(X, Z, T) = \sum_{n=1}^{\infty} \widehat{\phi}(X, n, T)\,\sin n\pi Z. \tag{2.34}$$

Substituting (2.34) into (2.14), we obtain

$$\left[\frac{\partial^2}{\partial T^2} + 1 - \delta(T)^2 + \delta'(T) - \left(\frac{Bu}{n\pi}\right)^2 e^{2\beta(T)}\frac{\partial^2}{\partial X^2}\right]\widehat{\phi}(X, n, T) = Ro\,A_n\,e^{\beta(T)}\,b_0'(X), \tag{2.35}$$

with A_n defined as

$$A_n = \frac{2\,(-1 + (-1)^n)}{n^3\pi^3}. \tag{2.36}$$

Equation (2.35) is hyperbolic (wave-like) for all n so long as $\delta \neq 0$ and $Bu \neq 0$. While we will not attempt to solve (2.35) analytically at this point, we can readily write down the equations for the characteristics:

$$\mathscr{X}_\pm = e^{-\beta(T)}\left(\mathscr{X}_0 \pm \frac{Bu}{n\pi}\int_0^T e^{\beta(T')}\,\mathrm{d}T'\right), \tag{2.37}$$

defined here in terms of the regular momentum coordinate, $\mathscr{X} = x + Ro\,v = e^{-\beta(T)}\,X$, which equals the Eulerian coordinate x at mid-depth (since $v = 0$ at $Z = 0.5$). The region between the plus ($\mathscr{X}_+$) and minus ($\mathscr{X}_-$) characteristics is the region of influence of $\mathscr{X}_0$, and indicates the area over which amplitude/energy associated with vertical mode n, and initially at location $\mathscr{X}_0$, has spread by time T. In the case of a constant strain δ, the characteristics from (2.37) become

$$\mathscr{X}_{\pm}^{\delta=\text{const.}} = e^{-\delta T}\,\mathscr{X}_0 \pm \frac{Bu}{n\pi\,\delta}(1 - e^{-\delta T}). \tag{2.38}$$

Thus, at constant strain, the region of influence of a given $\mathscr{X}_0$ is

$$\mathscr{X}_0\,e^{-\delta T} - \frac{Bu}{n\pi\delta} \leqslant \mathscr{X} \leqslant \mathscr{X}_0\,e^{-\delta T} + \frac{Bu}{n\pi\delta}, \tag{2.39}$$

at time T. If we now consider the limit of infinite time, $T \to \infty$, then the region of influence for $\mathscr{X}_0$ becomes

$$-\frac{Bu}{n\pi\delta} \leqslant \mathscr{X} \leqslant \frac{Bu}{n\pi\delta}, \tag{2.40}$$

assuming that $\delta > 0$. Since (2.40) applies for any $\mathscr{X}_0$, it implies that the solution ϕ is identically zero for $|\mathscr{X}| > Bu/(n\pi\delta)$ in the limit $T \to \infty$ for an appropriate choice of boundary conditions (see appendix B for the detailed derivation). This 'confinement' result will prove important in formulating the general solution to (2.14) in §4.

The fact that at large enough time the frontal circulation is entirely confined within the region $|\mathscr{X}| \leqslant Bu/(n\pi\delta)$ is a powerful result. It gives a fundamental scale for the width of frontal circulation at large time. This width decreases with increasing strain, δ, implying that for large strains we can expect significantly larger vertical velocities. The edge of the confinement region is equivalent to the points at which the group speed of the fastest IGWs of vertical mode n, $c_g = Bu/(n\pi)$, equals the strain flow speed at mid-depth, $\delta\mathscr{X}$. In other words, $\mathscr{X} = \pm Bu/(n\pi\delta)$ are the points at which outward-propagating waves are expected to stagnate in the strain flow. Thus, the convergent strain is effectively trapping the circulation within the region of possible wave propagation. By contrast, in the infinite time limit of the HB model, which does not include waves, there is no equivalent confinement effect, and the amplitude of the solution is finite everywhere. We will discuss these differences between the generalised and HB models in more detail when we formulate the general solution in §4.

In the following sections, we examine in detail the dynamics of frontogenesis and wave generation using the generalised model derived above. We focus on parameter values corresponding to the small Rossby limit (2.29), where the scaling analysis indicates the generalised model will be most accurate.

3. Inertial oscillations in zero-PV flow

It is useful to first examine the limit of zero PV, where the equations greatly simplify. The assumption of zero PV requires that the background stratification N^2 vanish and thus $Bu = (NH)/(fL) \to 0$. This limit allows the isolation of spontaneous wave generation via acceleration of the large-scale flow from other mechanisms of generation that occur simultaneously at non-zero PV. The zero-PV limit is also of

practical interest, since the PV in ocean and atmosphere boundary layers is often close to zero. With $Bu \rightarrow 0$, (2.14) with $\mathcal{N} \equiv 0$ becomes simply

$$\left[\frac{\partial^2}{\partial T^2} + 1 - \delta(T)^2 + \delta'(T) \right] \frac{\partial v}{\partial Z} = Ro \, b_0'(X) \, e^{\beta(T)}, \tag{3.1}$$

written here for convenience in terms of the along-front shear. The geostrophic component of this shear may be computed by substitution of the buoyancy (2.13b) into the thermal wind equation (2.6):

$$\frac{\partial v_g}{\partial Z} = Ro \, b_0'(X) \, e^{\beta(T)}. \tag{3.2}$$

Equation (3.2) represents the 'directly forced' part of the flow and is equivalent to the solution of the HB model (i.e. (2.33)). Using (3.1) and (3.2), we can write down an equation for the time evolution of the ageostrophic part of the flow:

$$\zeta''(T) + (1 - \delta(T)^2 + \delta'(T)) \, \zeta(T) = -2 \, \delta'(T) \, e^{\beta(T)}, \tag{3.3}$$

where

$$\frac{\partial v_a}{\partial Z} = Ro \, b_0'(X) \, \zeta(T). \tag{3.4}$$

Consistent with the assumptions made in the model derivation (§ 2.1), we choose initial conditions of geostrophic balance with $v = v_g$ and $\partial_T v = 0$, or $\zeta(0) = \zeta'(0) = 0$. We therefore require a strain field with $\delta(0) = \delta'(0) = 0$ such that there is no forcing to the ageostrophic flow at time zero (i.e. the right-hand side of (3.3) vanishes). However, as the strain is turned on and $\delta'(T)$ becomes non-zero, ageostrophic flow is inevitably forced. As a simple example, consider applying a 'pulse of strain' to the system

$$\delta(T) = \delta_0 \sin^2 \left(\frac{\pi T}{\tau} \right) H(\tau - T), \tag{3.5}$$

where H is the Heaviside function. Such a strain profile could represent a region of convergence interior to an eddy field, where the eddies evolve on a characteristic time scale τ. Figure 1 shows the time-dependence of the solution for $\delta_0 = 0.2$ and $\tau = 8\pi$ (4 days) in terms of the ageostrophic part $\zeta(T)$, the geostrophic part $\exp \beta(T)$ and the full flow $\exp \beta(T) + \zeta(T)$. The strain field from (3.5) is shown in (a). The time-variation in the strain forces ageostrophic flow (waves), which amplifies over time as the strain acts and the integrated strain $\beta(T)$ in (3.3) increases. The waves become exactly inertial and obtain a constant magnitude as the strain vanishes ($T = \tau$). In other words, for $T > \tau$, the net flow is composed of non-decaying inertial oscillations about the state of geostrophic balance – this end-state behaviour is identical to the behaviour of the Blumen (2000) zero-PV model of frontogenesis. The key difference is that the Blumen (2000) model assumed unbalanced initial conditions, whereas here we have demonstrated a mechanism for the generation of inertial waves from purely balanced initial conditions.

The spatial and temporal separability of the solution (3.2), (3.4) implies that the generation of the inertial waves in this zero-PV model is independent of the sharpness of the front; waves would be generated even with a linear buoyancy gradient (i.e. constant $b_0'(X)$). As such, the model provides an example of wave generation due to the acceleration (or transience) of the large-scale flow (e.g. Viudez
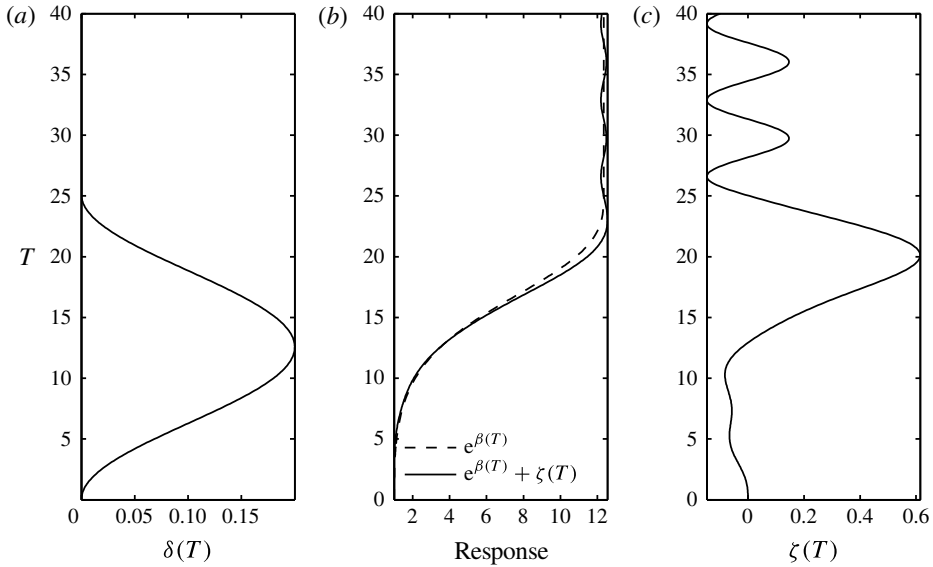
FIGURE 1. The geostrophic and ageostrophic responses to a pulse of strain in zero-PV flow. The strain $\delta(T)$ (*a*) is defined by (3.5) with $\delta_0 = 0.2$ and $\tau = 8\pi$. The time-variation of the strain drives ageostrophic flow $\zeta(T)$ in the form of an oscillation about the geostrophic solution, $\exp \beta(T)$.

2007; Vanneste 2013), rather than generation via frontal collapse. Note that wave generation via this transience mechanism requires the imposed strain field to vary on a sufficiently fast time scale. Therefore, this mechanism by itself does not provide a complete model for spontaneous wave generation from a slowly varying balanced flow, since such a model would also require a description of the processes giving rise to the rapid time-variation in the strain.

4. Constant strain

In the previous section we isolated the process of spontaneous IGW generation by acceleration of the large-scale flow. In this section we wish to similarly isolate the process of spontaneous IGW generation via frontal sharpening, and as such will require δ to be constant here before returning to the more general case of time-varying strain in §5.

The presence of a non-zero background stratification (non-zero PV) significantly complicates the situation compared with the zero-PV case discussed above. As the strain field amplifies frontal gradients, the system is now able to adjust via the emission of IGWs. Following Zhang (2004), the problem can be interpreted as one of continuous adjustment (on wave time scale $1/f$) to the applied strain (on time scale $1/\alpha$) and, unlike the zero-PV case, geostrophic balance is no longer necessarily a sensible balanced state for the flow. Indeed, a balanced state may not even exist owing to the breakdown of time-scale separation as $\delta = \alpha/f$ tends to 1 (e.g. McIntyre 2009). We instead consider the flow field as being composed of two parts; a secondary circulation, as in HB – that is, a directly forced response to the applied strain – and a wavefield. In this section we have two main objectives. First, we seek to determine a generalised secondary circulation (GSC) that extends the HB secondary circulation

to finite strain; secondly, we seek to quantify the size of the wavefield in comparison to this 'quasi-balanced' secondary circulation.

The evolution of the frontal system in the generalised model is fully described by the PDE (2.14) with constant strain δ and $\mathcal{N} \equiv 0$. We seek a forced solution to the PDE of the form

$$\phi(X, Z, T) = \int_{-\infty}^{\infty} \phi_I((X - X_0)\, e^{-\delta T}, Z)\, b_0'(X_0)\, dX_0, \tag{4.1}$$

such that time-dependence in the solution only arises through the strain-driven contraction of the horizontal coordinate. This form of solution eliminates the propagating waves associated with initial conditions that were studied in ST13. Equation (4.1) may be written more intuitively in terms of the regular momentum coordinate (e.g. as in Blumen 2000),

$$\mathcal{X} = X\, e^{-\delta T} = x + Ro\, v, \tag{4.2}$$

as

$$\phi(\mathcal{X}, Z, T) = \int_{-\infty}^{\infty} \phi_I(\mathcal{X} - \mathcal{X}_0, Z)\, \frac{\partial}{\partial \mathcal{X}_0} b_0\left(\mathcal{X}_0\, e^{\delta T}\right)\, d\mathcal{X}_0. \tag{4.3}$$

The function $\phi_I(\mathcal{X}, Z)$ in (4.3) is the time-independent impulse response, or Green's function, for the problem. We show in appendix A that the Green's function satisfies

$$\left[\frac{\partial^2}{\partial Z^2}\left(\delta^2\, \mathcal{X}^2\, \frac{\partial^2}{\partial \mathcal{X}^2} + \delta^2\, \mathcal{X}\, \frac{\partial}{\partial \mathcal{X}} + 1 - \delta^2\right) + Bu^2\, \frac{\partial^2}{\partial \mathcal{X}^2}\right] \phi_I(\mathcal{X}, Z) = Ro\, \underline{\delta}(\mathcal{X}), \tag{4.4}$$

where $\underline{\delta}$ denotes the Dirac delta function. Given that this impulse response is independent of time, time dependence only arises in the solution via the strain-driven contraction of the boundary buoyancy gradient. As indicated by (4.3), the solution at a given time is obtained by a convolution of the impulse response function with the boundary gradient at that time; time has thus been reduced to a parameter in the solution, as in the classical HB model. Note that the solution to the HB model of frontogenesis may be obtained from a similar convolution, where the HB impulse response $\phi_{I,HB}$ is defined by (4.4) with the $O(\delta^2)$ terms neglected:

$$\left[\frac{\partial^2}{\partial Z^2} + Bu^2\, \frac{\partial^2}{\partial \mathcal{X}^2}\right] \phi_{I,HB}(\mathcal{X}, Z) = Ro\, \underline{\delta}(\mathcal{X}). \tag{4.5}$$

In other words, (4.5) is equivalent to the classical Sawyer–Elliassen equation with the buoyancy gradient forcing set to a delta function. The generalised model equation (4.4), including the $O(\delta^2)$ terms, can thus be thought of as a large-strain generalisation of the Sawyer–Elliassen equation.

Apart from generating the full solution via (4.3), the impulse response functions defined by (4.4) also provide information about the long-time state of the system. As time progresses, any typical initially smooth boundary profile $b_0(\mathcal{X})$ will be squeezed towards a step, $b_0(\mathcal{X}\, e^{\delta T}) \to H(\mathcal{X})$, where H is the Heaviside function. The gradient of the boundary profile therefore approaches a delta function, and thus the general solution from (4.3) will collapse towards the impulse response function. Specifically,

$$\lim_{T \to \infty} \phi(\mathscr{X}, Z, T) = \lim_{T \to \infty} \int_{-\infty}^{\infty} \phi_I(\mathscr{X} - \mathscr{X}_0, Z) \frac{\partial}{\partial \mathscr{X}_0} b_0 \left(\mathscr{X}_0 \, e^{\delta T} \right) \, d\mathscr{X}_0$$

$$= \int_{-\infty}^{\infty} \phi_I(\mathscr{X} - \mathscr{X}_0, Z) \, \underline{\delta}(\mathscr{X}_0) \, d\mathscr{X}_0$$

$$= \phi_I(\mathscr{X}, Z). \tag{4.6}$$

We will employ this result in § 4.1 below to predict generic features of fronts at large time from the impulse response functions.

4.1. *The impulse response solution*

Here, we derive and physically interpret the impulse response $\phi_I(\mathscr{X}, Z)$ defined by (4.4). The solution proceeds most readily if we consider ϕ_I as sum of vertical Fourier modes (similar to (2.34)) such that the homogeneous boundary conditions, $\phi_I(\mathscr{X}, 0) = \phi_I(\mathscr{X}, 1) = 0$, are satisfied. The PDE (4.4) in ϕ_I is then reduced to an ODE in $\widehat{\phi}_I$ for each vertical mode,

$$\left[(\mathscr{X}^2 - \mathscr{X}_s(n)^2) \frac{\partial^2}{\partial \mathscr{X}^2} + \mathscr{X} \frac{\partial}{\partial \mathscr{X}} + \sigma^2 \right] \widehat{\phi}_I(\mathscr{X}, n) = \frac{A_n \, Ro}{\delta^2} \underline{\delta}(\mathscr{X}), \tag{4.7}$$

where A_n is defined as per (2.36),

$$\sigma = \frac{\sqrt{1 - \delta^2}}{\delta}, \quad \mathscr{X}_s(n) = \frac{Bu}{n\pi\delta}, \tag{4.8a,b}$$

and we assume that $\delta < 1$. It is further convenient to change the horizontal coordinate $\mathscr{X}$ to

$$\epsilon(\mathscr{X}) = \sigma \, \arctan \frac{\mathscr{X}}{\sqrt{\mathscr{X}_s(n)^2 - \mathscr{X}^2}}, \tag{4.9}$$

which transforms (4.7) to a very simple form,

$$\left[\frac{\partial^2}{\partial \epsilon^2} - 1 \right] \widehat{\phi}_I(\mathscr{X}(\epsilon), n) = \frac{n\pi A_n \, Ro}{Bu\sqrt{1 - \delta^2}} \underline{\delta}(\epsilon). \tag{4.10}$$

The boundary conditions on ϕ_I are that it must vanish infinitely far from the front, or $\widehat{\phi}_I \to 0$ as $\mathscr{X} \to \pm\infty$. In appendix B, we use the characteristics of the generalised model PDE (2.35) to show that, with these boundary conditions, the nth vertical mode of $\phi(\mathscr{X}, Z, T)$ vanishes in the region $|\mathscr{X}| > Bu/(n\pi\delta)$ in the limit $T \to \infty$. Equation (4.6) implies that the same result must apply to the impulse response $\widehat{\phi}_I$. The boundary conditions on $\widehat{\phi}_I$ may also be verified by solving (4.7) in Fourier space (see appendix A, (A 3)) and taking the inverse Fourier transform numerically. The resulting $\widehat{\phi}_I$ is zero for $|\mathscr{X}| > Bu/(n\pi\delta)$.

The physical reason for the sharp cutoff in amplitude at $\pm\mathscr{X}_s(n)$ is associated with the dynamics of IGWs in this strained, hydrostatic flow. The waves are responsible for the transfer of energy away from the source at the origin and can only propagate where their outward group velocity exceeds the inward strain flow speed (e.g. Jones 1969). Since $\mathscr{X} = \mathscr{X}_s(n)$ is the location at which the strain flow magnitude first exceeds the maximum wave group speed for each mode n, a wavepacket of vertical mode n can only propagate out to this point. We call the set of $\mathscr{X}_s(n)$ the 'stagnation

points' for each mode (hence the subscript 's'; this nomenclature was also used in ST13). Further, the group speed of IGWs in a hydrostatic flow asymptotes to the maximum group speed $Bu/(n\pi)$ in the limit of large horizontal wavenumber. The convergent strain will tend to increase the wavenumbers of a wavepacket propagating in the strain flow, $k \to \infty$, and thus the group speed of all wavenumber components will asymptote towards $Bu/(n\pi)$ at large time. All wavenumbers will thus be concentrated at the stagnation point at large time. In other words, the fact that the large wavenumbers of the flow are non-dispersive implies that there is no spreading mechanism to counteract the squeezing imposed by the strain field, leading to a sharp cutoff in amplitude at the stagnation points. If non-hydrostatic effects are included in the model, resulting in dispersion at high horizontal wavenumbers, then the sharp cutoff no longer occurs, and the impulse response tends smoothly to zero beyond the stagnation points. These non-hydrostatic dynamics will be examined in detail in a upcoming paper.

It is interesting to note that the problem in ϵ coordinates (4.10) is mathematically identical to that of calculating the geostrophically adjusted velocity for an initial step displacement in the free surface height in rotating shallow water, as studied by Gill (1976). Gill's equation (5.5) may be written as

$$\left(\frac{\partial^2}{\partial \epsilon^2} - 1\right) v(\epsilon) = v_0''(\epsilon) - \eta_0'(\epsilon), \tag{4.11}$$

where v is the steady state along-front flow, ϵ is the horizontal coordinate, and v_0 and η_0 are the initial velocity and height fields, respectively. With $v_0 = 0$ and η_0 equal to a unit step, (4.11) is identical to (4.10), up to scaling by a constant. While Gill studied this problem in an infinite domain, the boundary conditions introduced above imply that the analogous shallow-water flow is trapped between no-slip channel walls at $\epsilon = \pm \sigma \pi/2$. The presence of a strain field in the frontogenesis problem may thus be thought of as constraining the flow between walls; as the strain tends to zero, we have $\sigma \to \infty$ and so the 'walls' vanish.

Following Gill, it is straightforward to write down the exact solution to (4.10):

$$\widehat{\phi}_I(\mathscr{X}, n) = \begin{cases} \dfrac{n\pi A_n\, Ro}{2\, Bu\sqrt{1-\delta^2}} \left[\mathrm{e}^{-|\epsilon|} - \mathrm{e}^{-(\pi \sigma)/2} \dfrac{\cosh \epsilon}{\cosh \dfrac{\pi \sigma}{2}} \right] & |\mathscr{X}| < \mathscr{X}_s(n) \\[20pt] 0 & |\mathscr{X}| \geqslant \mathscr{X}_s(n). \end{cases} \tag{4.12}$$

The solution (4.12) is composed of two terms; the first, $\exp(-|\epsilon|)$, is the directly forced part (or particular solution) resulting from the applied impulse forcing and is the solution – vanishing at infinity in ϵ coordinates – that one would obtain in the absence of the constraint $|\epsilon| \leqslant \sigma \pi/2$ (e.g. see (5.7) of Gill). However, with these 'walls' this part of the solution does not independently satisfy the boundary conditions. A second term, $\cosh \epsilon$, corresponding to the wave (or homogeneous) solutions, must be introduced to satisfy the boundary conditions. We can understand the origin of this $\cosh \epsilon$ term by consideration of the infinite-domain shallow-water problem where, as shown by Gill, the adjustment of the initial height displacement generates Poincaré waves that propagate unimpeded away from the adjustment region. However, with the introduction of channel walls, the generated waves will instead reflect from the channel walls and set up a standing wavefield in the channel. We can interpret the $\cosh \epsilon$ part of the solution as corresponding to the time-averaged part of
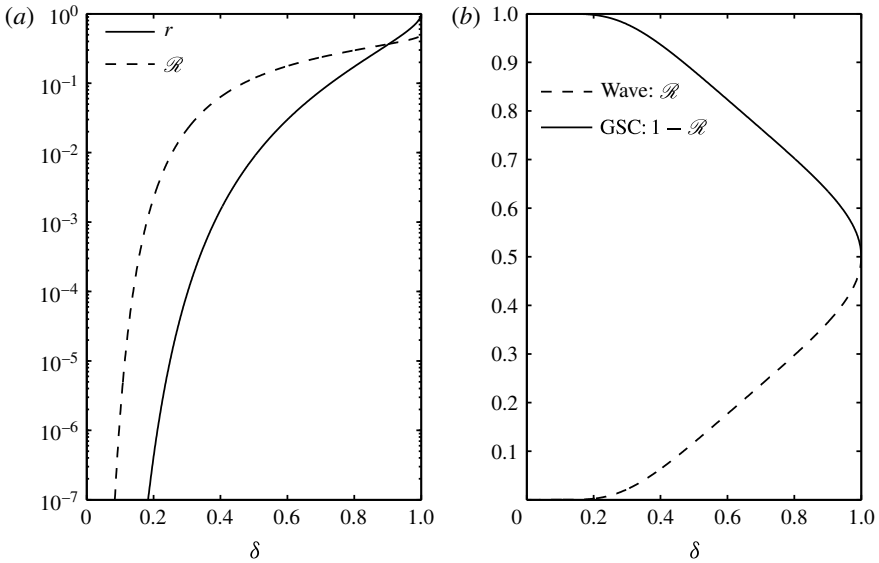
FIGURE 2. (*a*) The relative amplitude r (4.13) of the 'wave' term in the generalised model solution (4.12), and the relative strength of wave generation $\mathscr{R}$ (4.20), as a function of the strain δ. (*b*) The fraction of amplitude from an impulse forcing going into the wave, $\mathscr{R}$, and GSC, $1 - \mathscr{R}$, components of the flow for a given value of strain.

this standing wavefield. The amplitude of this 'wave' term relative to the particular solution $(\exp(-|\epsilon|))$ is

$$r = \frac{e^{-(\pi\sigma)/2}}{\cosh \dfrac{\pi \sigma}{2}}. \tag{4.13}$$

Equation (4.13) is a function only of the strain δ and is plotted in figure 2(*a*). The figure shows that the 'wave' term is exponentially smaller than the particular solution for strains less than about $\delta = 0.2$, but becomes of comparable size for larger values of strain. This exponential smallness is an important feature, which we will revisit below.

We now consider the problem of explicitly dividing the impulse response solution (4.12) into secondary circulation and wave components. The solution (4.12) is defined in terms of the function ϕ which, while mathematically useful, is difficult to interpret physically. However, using the relation between the streamfunction and ϕ (2.13*c*), and the definition of the impulse response solution (4.3), it may be shown that the streamfunction impulse response is defined by

$$\psi_I(\mathscr{X}, Z) = \delta \left(\mathscr{X} \frac{\partial \phi_I}{\partial \mathscr{X}} - \phi_I \right). \tag{4.14}$$

Substituting ϕ_I from the solution (4.12) into yields

$$\widehat{\psi_I}(\mathscr{X}, n) = \frac{-\delta\, n\pi A_n\, Ro}{2\, Bu\sqrt{1-\delta^2}} \left[\underbrace{\left(e^{-|\epsilon|} - r\,\cosh\epsilon\right)\left(1 + |\mathscr{X}|\,\epsilon'(\mathscr{X})\right)}_{\text{GSC}} \right.$$

$$\left. + \underbrace{r\left(\mathscr{X}\,\epsilon'(\mathscr{X})\,\sinh\epsilon + |\mathscr{X}|\,\epsilon'(\mathscr{X})\,\cosh\epsilon\right)}_{\text{WAVE}} \right]. \qquad (4.15)$$

Now consider dividing $\widehat{\psi_I}(\mathscr{X}, n)$ into a GSC, accounting for the effects of finite strain, and a wave part. We require that the GSC part is finite everywhere in the domain and limits to the HB secondary circulation impulse response in the limit of small strain. We require that the wave component of the streamfunction be proportional to r (4.20), as in the general solution for ϕ_I (4.12), such that it is exponentially small as $\delta \to 0$. The streamfunction impulse response in (4.15) has been divided into two components satisfying these requirements. Indeed, it may be shown that at small δ,

$$\widehat{\psi_{I,GSC}}(\mathscr{X}, n) = \widehat{\psi_{I,HB}}(\mathscr{X}, n) + O(\delta^3), \qquad (4.16)$$

plus higher-order terms, where

$$\widehat{\psi_{I,HB}}(\mathscr{X}, n) = \frac{-\delta\, n\pi A_n\, Ro}{2\, Bu} e^{-(n\pi)/Bu|\mathscr{X}|}\left(1 + \frac{n\pi}{Bu}|\mathscr{X}|\right) \qquad (4.17)$$

is the HB model impulse response (as may be determined from substituting the solution to (4.5) into (4.14)).

The exact separation of the circulation into GSC and wave components (4.15) is somewhat arbitrary, and the distinction between the components becomes more difficult to interpret at large values of δ. Nonetheless, the separation performed above has two important properties. First, in the limit of small strain, the GSC component limits to the secondary circulation from the HB model as indicated in (4.16), and as will be shown in §4.2, the wave component is well described by stationary mode one IGWs. Secondly, if the strain forcing turns off (i.e. a time-dependent strain, see §5), the propagating part of the resulting flow is entirely contained within the wave component. In light of these observations, the term 'wave' is invoked here to qualitatively describe the dynamical response of the system, even though the distinction between the wave and GSC parts of the flow may not be formally valid (particularly for large values of the strain).

The streamfunction impulse responses from the HB model, and the GSC and wave components from the generalised model, are shown in figure 3 for two values of strain δ. First, considering the small-strain case, $\delta = 0.2$, we observe that differences between the GSC impulse response and the HB impulse response are relatively small, as should be expected. The major difference is that the GSC response is confined within the region $|\mathscr{X}| < Bu/(\pi\delta)$, whereas the HB response is non-zero for all $\mathscr{X}$. This confinement of the secondary circulation in the generalised model is due to the effect on strain on wave propagation; that is, waves can only propagate where their outward group speed exceeds the incoming strain flow speed. By contrast, waves in the HB model are assumed to be 'free' and can thus propagate out to $\pm\infty$. As seen in figure 3(a), the confinement of the GSC leads to an increased amplitude in the centre of the frontal zone (near $\mathscr{X} = 0$) and a reduction on the flanks, compared to the HB secondary circulation.
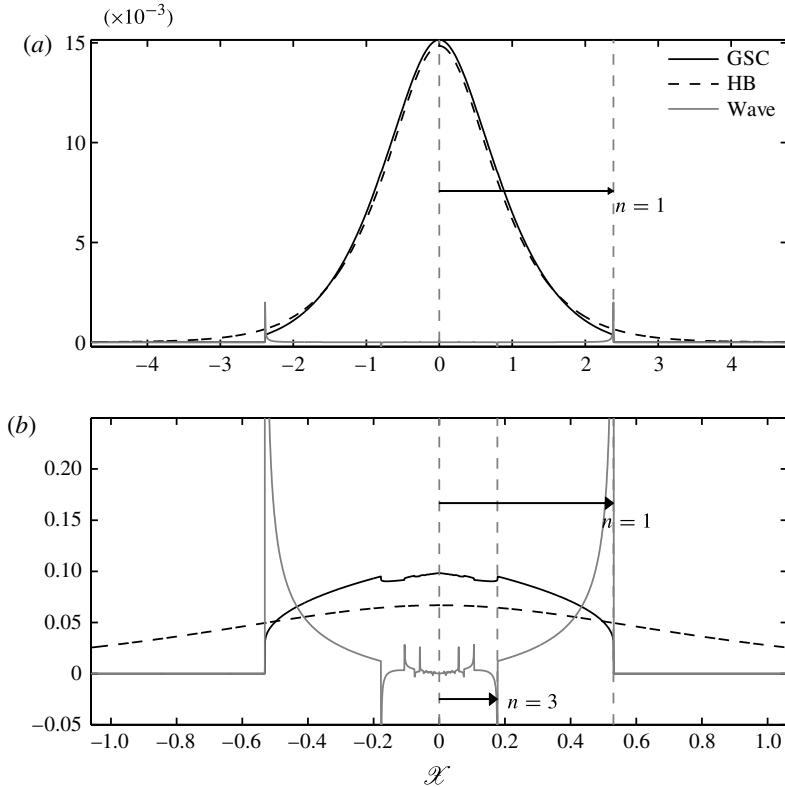
FIGURE 3. The streamfunction impulse response ψ_I evaluated at $Z = 0.5$ for the GSC and wave components of the general solution (4.15), and the HB solution ((4.17), HB), for strains of (a) $\delta = 0.2$ and (b) $\delta = 0.9$. The impulse forcing is located at $\mathscr{X} = 0$, and the spread of energy away from this point is indicated by arrows for the first two non-zero modes $n = 1$ and $n = 3$. Energy can only spread over the region $\mathscr{X} \leqslant Bu/(n\pi\delta)$ for each mode. The parameter values used are $Bu = 1.5$ and $Ro = 0.6$, consistent with other examples considered later in this paper.

Now consider the wave part of the streamfunction, $\psi_{I,W}$, as defined in (4.15). Since $\psi_{I,W}$ scales with r, which is exponentially small for small strains (see figure 2), we expect $\psi_{I,W}$ to be relatively small in comparison to the GSC for $\delta = 0.2$. Figure 3 shows that $\psi_{I,W}$ is close to zero everywhere except near the stagnation points, $\mathscr{X}_s = \pm Bu/(n\pi\delta)$, where it becomes infinite. As with the GSC response, the wave response is identically zero for $\mathscr{X} > \mathscr{X}_s$. This structure of $\psi_{I,W}$ is consistent with the idea that IGWs generated by the squeezing of the front are confined within $\mathscr{X} < \mathscr{X}_s$, and stagnate at $\mathscr{X} = \mathscr{X}_s$, and thus wave energy will accumulate at $\mathscr{X} = \mathscr{X}_s$. The infinities in the wave impulse response at the stagnation points require careful interpretation. We observe that the full time-dependent wave streamfunction ψ_W, which may be reconstructed from $\psi_{I,W}$ via

$$\psi_W(\mathscr{X}, Z, T) = \int_{-\infty}^{\infty} \psi_{I,W}(\mathscr{X} - \mathscr{X}_0, Z) \frac{\partial}{\partial \mathscr{X}_0} b_0(\mathscr{X}_0 e^{\delta T}) \, d\mathscr{X}_0, \qquad (4.18)$$

will remain finite for all time for any smooth initial buoyancy profile b_0. However, in the absence of the formation of a discontinuity at the front or viscous effects, the

infinities in the impulse response imply that the magnitude of ψ_W at the stagnation points will continuously increase with time. In other words, the waves stagnating at $\pm\mathscr{X}_s$ will continuously amplify with time, consistent with the wave capture paradigm of Buhler & McIntyre (2005).

Figure 3(b) shows the streamfunction impulse responses for a larger value of strain, $\delta = 0.9$. The difference between the HB and GSCs is now of order one for all $\mathscr{X}$, due to the strong confinement of energy in the generalised model greatly intensifying the secondary circulation and localising it near the front. The wave component $\psi_{I,W}$ is an order-one contribution to the overall streamfunction, and greatly exceeds the GSC near the first vertical mode stagnation points. The higher vertical mode signals ($n = 3, 5$ etc.) are also visible as large-amplitude spikes near each of their stagnation points, $\mathscr{X}_s = \pm Bu/(n\pi\delta)$, indicating the accumulation of wave energy from each mode at those locations.

The integral of the net streamfunction impulse responses (ψ_I) is equal for both the HB and generalised models, and is controlled by the magnitude of the impulse forcing in the original differential equation (4.7),

$$\int_{-\mathscr{X}_s}^{\mathscr{X}_s} \psi_I(X, n) \, \mathrm{d}\mathscr{X} = -2\delta \int_{-\mathscr{X}_s}^{\mathscr{X}_s} \phi_I(X, n) \, \mathrm{d}\mathscr{X} = -2\delta A_n Ro, \qquad (4.19)$$

by substituting (4.14), followed by (4.12). For the generalised model, this 'impulse of amplitude' is split between the GSC and wave streamfunctions. As shown in figure 3, the relative size of the wave component varies significantly with strain δ. To quantify this variation, we compute the $\mathscr{X}$-integrated wave streamfunction, normalised by the total streamfunction amplitude from (4.19),

$$\mathscr{R} = \frac{\displaystyle\int_{-\mathscr{X}_s}^{\mathscr{X}_s} \psi_{I,W}(X, n) \, \mathrm{d}\mathscr{X}}{-2\delta A_n Ro} = \frac{\mathrm{e}^{-(\pi\sigma)/2} + \sigma}{\mathrm{e}^{-(\pi\sigma)/2} + \mathrm{e}^{(\pi\sigma)/2}} = \frac{r}{2}(1 + \sigma \, \mathrm{e}^{(\pi\sigma)/2}). \qquad (4.20)$$

The ratio $\mathscr{R}$ can be interpreted as the relative strength of wave – as opposed to secondary circulation – generation associated with an element of horizontal buoyancy gradient acted on by convergent strain δ. The ratio is a function of δ only and is plotted in figure 2 (2a; log scale, 2b; linear scale). Unsurprisingly, given the definition of the wave streamfunction in (4.15), the ratio $\mathscr{R}$ is exponentially small for small strains, but finite for larger strains. Indeed, for $\delta \to 0$ (4.20) implies that $\mathscr{R} \to 1/\delta \exp -\pi/(2\delta)$, meaning that virtually all the energy supplied by the impulse forcing goes into the secondary circulation, consistent with the Hoskins & Bretherton (1972) paradigm. In contrast, for a forcing frequency approaching the inertial, $\delta \to 1$, (4.20) implies that $\mathscr{R} \to 1/2$, such that the impulse forcing is evenly split between the wave and GSC streamfunctions. We will see in §5 that the scale given by (4.20) provides a useful upper limit on the amplitude of propagating frontogenesis waves generated by a time-dependent strain field acting on a front – this is despite the separation of the wave and GSC streamfunctions in (4.15) being somewhat arbitrary.

The exponential smallness of spontaneous wave generation for small strain (i.e. small large-scale Rossby number Ro_L) implied by (4.20) has previously been predicted in 'toy models' (see Vanneste 2008, and references therein). The sharp cutoff in wave generation at approximately $\delta = 0.2$ in figure 2 has important consequences for the production of waves in eddy fields, which exhibit substantial spatially varying strain fields. Figure 2 suggests that we should expect wave generation to be localised

to a number of distinct regions of relatively large strain, with exponentially smaller (i.e. negligible) generation outside of these regions. Such behaviour is consistent with numerical studies of spontaneous generation in eddy fields (e.g. Danioux *et al.* 2012). Equation (4.20) may also provide a theoretical foundation for gravity-wave parametrisations in numerical models, which often depend on the local strain (Plougonven & Zhang 2014, and references therein).

Based on the impulse responses displayed in figure 3, we can infer some features of the flow at large time during frontogenesis forced by a constant strain – or more generally around any sufficiently sharp, strained front. The GSC and frontal zone will have a fundamental length scale, or width, of $2\mathscr{X}_s(1) = 2Bu/(\pi\delta)$ corresponding to the stagnation point for the fastest-propagating mode one wave. In dimensional units, the width is $2NH/(\pi\alpha)$. Stationary gravity waves associated with each vertical mode will appear at or near the stagnation points $\pm Bu/(n\pi\delta)$. The wave amplitudes will grow as frontogenesis proceeds and wave energy accumulates at these locations. The highest-amplitude wave will be associated with the first vertical mode appearing at the stagnation points on the edges of the frontal region. Referring to figure 3, this mode will drive a thin band of intense upwelling (since $w = -J\,\partial_X\psi$) ahead of the front on the warm side (*vice versa* on the cool side). These mode one waves can be expected to be visible even for relatively small values of strain, since the GSC response tends to zero at these locations. Waves associated with the higher vertical modes overlay the frontal zone and will, for small strains, tend to add relatively weak 'fine-scale structure' to the GSC-dominated fields in this region.

4.2. *Full solution*

Equation (4.3) demonstrates how the full time-dependent solution $\phi(\mathscr{X}, Z, T)$ may be generated by summation of the impulse responses of infinitesimal elements $\mathrm{d}\mathscr{X}_0$ of buoyancy gradient at that time. The partial time derivative $\partial_T\phi(\mathscr{X}, Z, T)$ may be similarly generated from its impulse response:

$$\partial_T\phi(\mathscr{X}, Z, T) = -\delta \int_{-\infty}^{\infty} (\mathscr{X} - \mathscr{X}_0) \frac{\partial\phi_I(\mathscr{X} - \mathscr{X}_0, Z)}{\partial\mathscr{X}} \frac{\partial}{\partial\mathscr{X}_0} b_0\left(\mathscr{X}_0\,\mathrm{e}^{\delta T}\right) \mathrm{d}\mathscr{X}_0. \quad (4.21)$$

Once $\phi(\mathscr{X}, Z, T)$ and $\partial_T\phi(\mathscr{X}, Z, T)$ have been obtained, the along-front velocity, buoyancy and streamfunction fields at time T can be computed by expressing (2.13) in coordinate $\mathscr{X}$:

$$v(\mathscr{X}, Z, T) = \frac{\partial}{\partial Z}\phi(\mathscr{X}, Z, T), \quad (4.22a)$$

$$b(\mathscr{X}, Z, T) = b_0\left(\mathscr{X}\,\mathrm{e}^{\delta T}\right) + F^{-2}Z - Ro\,F^{-2}\frac{\partial}{\partial\mathscr{X}}\phi(\mathscr{X}, Z, T), \quad (4.22b)$$

$$\psi(\mathscr{X}, Z, T) = -\delta\,\phi(\mathscr{X}, Z, T) - \partial_T\phi(\mathscr{X}, Z, T). \quad (4.22c)$$

Further, we can define the time-dependent GSC as

$$\psi_{GSC}(\mathscr{X}, Z, T) = \int_{-\infty}^{\infty} \psi_{I,GSC}(\mathscr{X} - \mathscr{X}_0, Z) \frac{\partial}{\partial\mathscr{X}_0} b_0\left(\mathscr{X}_0\,\mathrm{e}^{\delta T}\right) \mathrm{d}\mathscr{X}_0, \quad (4.23)$$

where $\psi_{I,GSC}$ is the GSC impulse response defined in (4.15). The wave streamfunction may then be computed as the difference, $\psi_W = \psi - \psi_{GSC}$. As noted previously, this
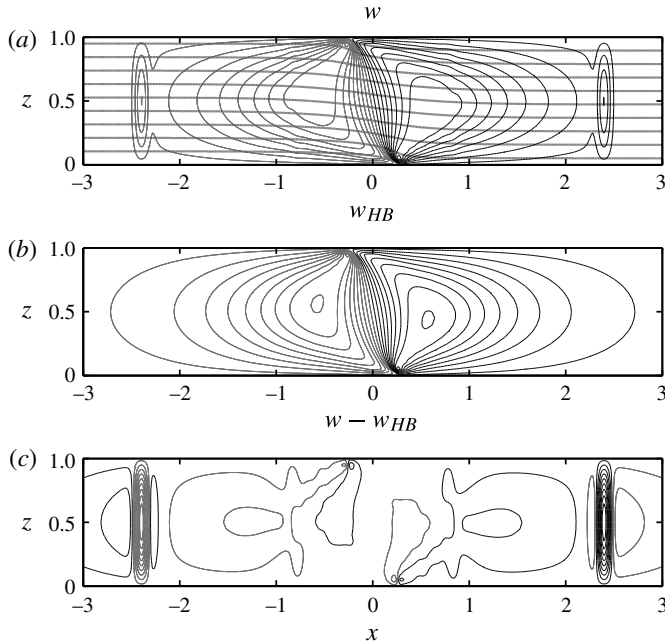
FIGURE 4. The vertical velocity field from the generalised model w, the Hoskins and Bretherton model w_{HB} and the difference between the two, evaluated just prior to the critical time for parameter values of $Ro = 0.6$, $Bu = 1.5$, and $\delta = 0.2$. Black denotes positive and grey denotes negative velocities. Contours of buoyancy are overlaid on the top plot (thick grey lines).

separation of wave and GSC streamfunctions is not unique. We will now examine the behaviour of these time-dependent fields in detail.

First, it is useful to consider an explicit example of the differences between the generalised and HB models in a parameter regime, δ and Ro both small, where both are expected to be valid. In figure 4 we display the vertical velocity fields at late time arising from the generalised model, w, the HB model, w_{HB}, and the difference between the two, $w - w_{HB}$, for parameter values of $\delta = 0.2$ and $Ro = 0.6$, and a buoyancy profile of $b_0(X) = 1/2 \, \text{erf}(X/\sqrt{2})$. Since all time dependence in the model arises through the straining of the boundary buoyancy profile as per (4.3), initial conditions on the model fields cannot be explicitly set and are instead determined implicitly from relations (2.13) using the field ϕ from (4.3) with $T = 0$. The HB model predicts a single thermally direct overturning cell with upwelling on the warmer (right-hand) side, and downwelling on the cooler side. The generalised model velocity field is broadly similar, dominated by an analogous large-scale overturning cell, but contains a number of additional features. The most obvious addition is the formation of distinct lines of intensified vertical flow on the periphery of the frontal region (at $\mathscr{X}_s(1) = \pm Bu/(\pi\delta) = \pm 2.4$), associated with the mode one gravity wave. Similar bands of up/downwelling were observed by Garner (1989, e.g. see their figure 5) and Snyder *et al.* (1993, e.g. see their figure 1) in their numerical simulations, although in those cases it is difficult to determine whether the feature is arising due to waves generated by frontogenesis (as here) or waves arising due to the initial conditions (e.g. as in ST13), since both types of waves stagnate at $\pm\mathscr{X}_s(1)$. One key difference
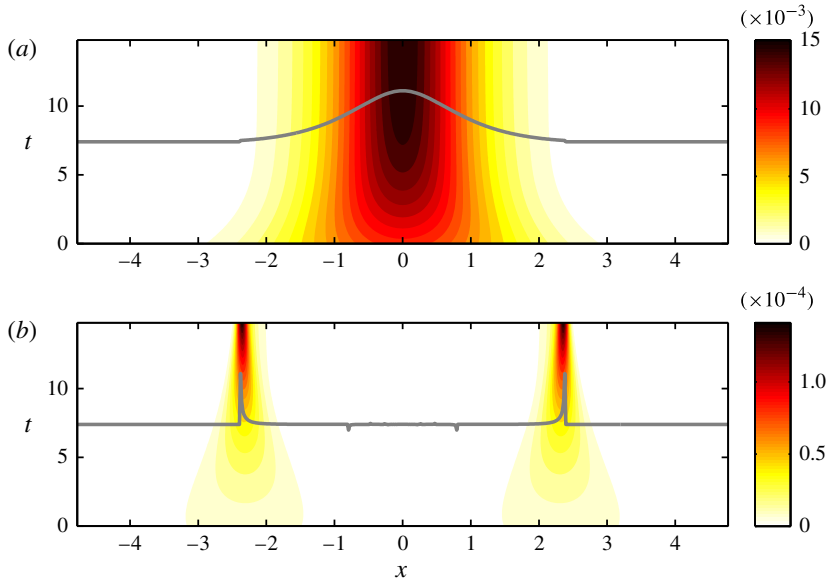
FIGURE 5. The GSC (ψ_{GSC}, a) and wave streamfunction (ψ_W, b), for parameter values of $Ro = 0.6$, $Bu = 1.5$, and $\delta = 0.2$. The GSC and wave impulse response functions from figure 3(a) have been overlaid in grey.

between the wave types is that the amplitude of the waves associated with initial conditions decay with time due to the action of the strain field (see ST13), whereas the amplitude of the waves associated with frontogenesis amplify with time as the front sharpens. Thus at sufficiently late time, or for a sufficiently sharp front, the frontogenesis waves are expected to dominate.

Figure 4 also shows additional fine structure associated with higher vertical modes in the vertical velocity field. For example, note the slight 'bumps' in the contours near $x \simeq \pm 0.8$ associated with the third vertical mode. There are also sizeable differences between the HB and generalised models at the location of the front ($x \simeq \pm 0.3$) on the upper and lower boundaries; this 'updraft–downdraft couplet' feature was observed by Snyder *et al.* (1993, their figure 2 and footnote 2) when comparing the HB prediction with the output of their numerical model. The feature appears due to the higher along-front velocity magnitudes (v) near the origin in the generalised model (as a result of the strain-imposed limit on wave propagation ensuring that more energy remains in the frontal zone; see figure 3), which cause the front to slump/slant further as represented by the momentum coordinate transformation, $x = \mathscr{X} - Ro\,v$. Consequently, there is an outward shift of the frontal position (and the associated extremum of the vertical velocity) in the generalised model compared with the HB model, giving rise to the couplet feature observed in the difference field.

In figure 5, we consider the time evolution of the system (same parameter values) via a Hovmöller plot of the GSC and wave streamfunctions at mid-depth. The GSC at mid-depth is initially squeezed inwards and amplified by the strain, but by $t \sim 10$ it approaches a steady state consistent with the GSC impulse response function (i.e. from figure 3a, overlaid in grey). Thus, as argued in the previous section, the impulse response provides a snapshot of the long-time state of the system. The wave streamfunction in figure 5 shows the accumulation of wave energy into distinct packets

at the stagnation points, consistent with the wave impulse response (overlaid in grey). However, in contrast to the GSC, the convergent strain then acts to continually amplify these wavepackets with time. The waves do not propagate owing to the trapping effect of the strain field, and are essentially fixed in space at the stagnation points. Note that the wave streamfunction in figure 5 is two orders of magnitude smaller than the GSC. A consequence of this is that distinct wave features (e.g. as in figure 4) are only visible at late time, once wave energy has accumulated at the stagnation points.

One of the most important applications of the generalised model is to flows that have order-one strains (i.e. order-one large-scale Rossby numbers Ro_L), such as frontogenesis in a submesoscale eddy field in the ocean mixed layer. As an example, consider a case with parameter values of $\delta = 0.9$, but $Ro = 0.6$ and $Bu = 1.5$ as before. The initial conditions are again implicitly defined through (4.3) and relations (2.13). Figure 6 shows the time evolution of the wave and GSC streamfunctions, with buoyancy contours overlaid in black at each time step. In contrast to the previous small-strain example, here the wave and GSC streamfunctions are of the same order of magnitude. Initially, both streamfunctions are characterised by a single-cell thermally direct overturning ($t = 0$), which rapidly intensifies as the frontal scale contracts ($t = 1.3$). The GSC remains as a single cell and continues to intensify as time proceeds. However, the wave streamfunction splits into two distinct wavepackets ($t = 2$), which then intensify with time. As in the previous example, the waves do not propagate and are essentially fixed in space at the stagnation points, $\mathscr{X}_s(1) = \pm 0.53$. For a more realistic flow with a temporally or spatially varying strain field (see § 5), these generated waves tend to be released and disperse rather than remaining fixed at the stagnation points and amplifying indefinitely. The structure of the wave and GSC streamfunctions at $t = 2.7$ (i.e. at long time, $T \gg \delta^{-1}$) is consistent with the impulse responses for the same parameter values shown in figure 3(*b*).

Between $t = 0$ and $t = 1.3$ in figure 6, the strain rapidly amplifies the horizontal buoyancy gradients, but this amplification is counterbalanced by a flattening of the isopycnals in the frontal zone. It was observed in ST13 that such flattening is indicative of adjustment and wave generation processes, giving credence to the 'balance adjustment' theory of Zhang (2004). Viewed from this perspective, the system adjusts to the strain-generated imbalance in the frontal zone by generating inertia–gravity wavepackets ($t = 1.3, 2$). The amplitude of these waves then increases with time as their horizontal scale contracts.

The adjustment process can only occur in the interior, since buoyancy conservation requires that $b = b_0(X) + F^{-2}Z$ on the boundaries at all time, leading to the formation of a sharp front on the boundary (see $t = 2.7$ in figure 6). A scale for the depth of the frontal feature can be determined based on the impulse solutions above. Taking the GSC ($\exp{-|\epsilon|}$) part of the impulse response solution (4.12), we observe that the amplitudes of the vertical modes in this solution decay exponentially with mode number n. Such exponential decay implies that the lowest vertical modes dominate the solution. The solution amplitude (and thus degree of adjustment) near to the boundary, $Z \ll 1/(n\pi)$, must therefore be small, and the front will be sharp in this region. For small $\mathscr{X}$, where the front is located, (4.12) yields a scale depth for the front of

$$h \sim \frac{1}{n\pi} \sim \frac{\sqrt{1 - \delta^2}}{Bu}, \tag{4.24}$$

for $\delta < 1$. Thus, for order-one strains we expect frontal features to be concentrated in a very shallow boundary layer near the surface, as figure 6 exemplifies. Strong adjustment and wave generation is expected outside the boundary layer.
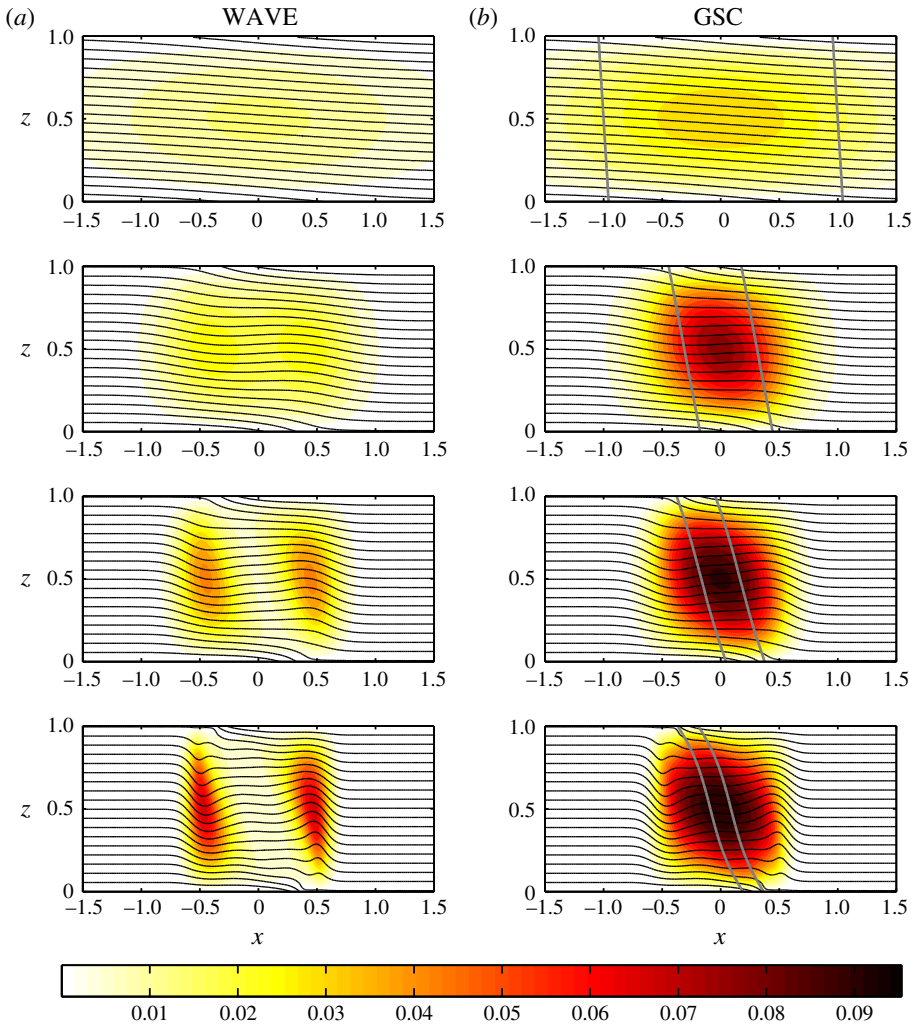
FIGURE 6. The time evolution of the wave ψ_W (a) and GSC ψ_{GSC} (b) streamfunctions for parameter values of $Ro = 0.6$, $Bu = 1.5$ and $\delta = 0.9$, at times of $t = 0$, 1.3, 2 and 2.7 (just prior to the critical time), with time increasing down the page. Buoyancy contours are overlaid in black in each case. Grey contours enclose the region $|X| \leqslant 1$, demonstrating the convergent action of the strain field, and the simultaneous slumping of the front.

5. Time-dependent strain

Here, we return to the more general case of a time-dependent strain field, with the objective of describing the response of an initially weak front, as defined by (2.29), that is initially in geostrophic balance, to an imposed strain field. We will employ the pulse-like strain profile from (3.5) to exemplify the situation of a strain field that is smoothly switched on, and acts for a finite period of time τ, during which it obtains a maximum value of δ_0, before being smoothly switched off. Such a strain field is expected to be ubiquitous in geophysical flows, particularly in strong eddying regions in the ocean mixed layer. The temporal variation of the strain in the model can be thought of as either (a) representing the lifetime of the eddies responsible for forming
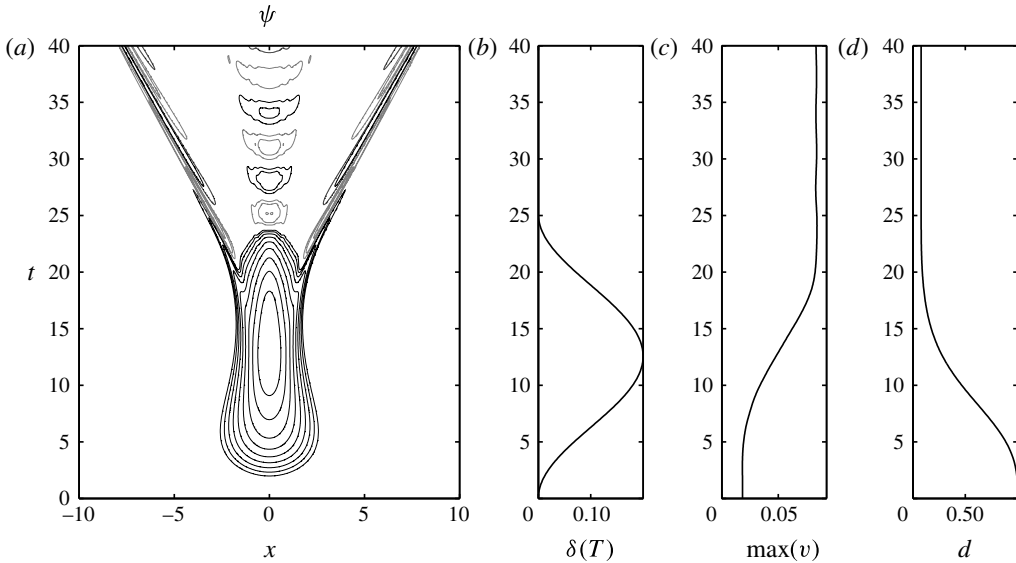
FIGURE 7. The solution of the time-dependent model (2.14) with $\mathcal{N} \equiv 0$, computed numerically, for parameter values of $Ro = 0.1$ and $Bu = 1$. The imposed strain field (*b*) is defined in (3.5) with $\delta_0 = 0.2$ and $\tau = 8\pi$. The boundary profile of buoyancy is $b_0(X) = 1/2 \, \mathrm{erf}(X/\sqrt{2})$. (*a*) Hovmöller of the streamfunction evaluated at mid-depth. The contour spacing is logarithmic from 1 to 100 % of the maximum value, with grey contours indicating negative (clockwise) overturning. (*c*) The maximum value of the along-front velocity v. (*d*) The frontal width $d = J^{-1}$.

the convergent strain field or (*b*) the time taken for a patch of fluid (comprising the front) to advect through a convergent region within a large-scale eddy field. The latter case invokes a temporal variation in the model strain to represent a spatial variation in the physical strain. The numerical solutions presented below proceed by Fourier transforming the PDE (2.14) with $\mathcal{N} \equiv 0$ (noting that the right-hand side forcing term vanishes sufficiently far from the front) and solving the resultant time ODE for each mode, then reconstructing the solution from a summation of the modes.

Consider an explicit example with parameter values representative of a front in a mesoscale eddy field in the ocean mixed layer; $Ro = 0.1$ (initially weak front), $Bu = 1$ (a length scale of the order of the Rossby radius), $\delta_0 = 0.2$ (the approximate Rossby number for a mesoscale eddy) and $\tau = 8\pi$ (the time scale for the eddy of $2\pi/\delta \sim \tau = 4$ days). Figure 7 displays the solution of our model for the above parameter values. The pulse of strain (figure 7*b*) pushes the system out of the initial geostrophic balance and drives a thermally direct secondary circulation (similar to the HB model), which forms the dominant feature of the Hovmöller plot of the streamfunction at mid-depth (figure 7*a*). Note that the contours in this plot are logarithmically spaced from 1 to 100 % of the maximum value. The time-varying strain also acts to smoothly sharpen the front, as demonstrated by the time series of frontal width shown in figure 7(*d*). Consistent with this frontal sharpening, the along-front velocity magnitude (figure 7*c*) increases to maintain the front close to geostrophic balance. The streamfunction (figure 7*a*) exhibits wave generation on the flanks of the secondary circulation (i.e. associated with the first vertical mode) as the front sharpens, similar to the case of constant strain (§ 4). These frontogenesis

waves are initially trapped by the strain field but begin to propagate as the strain weakens. The time-variation in the strain field also drives the generation of waves. These transience waves are visible as a near-inertial oscillation in the frontal zone for $T > \tau$, and slowly propagate outwards with time. Thus, for $T > \tau$ the system consists of near-inertial oscillations about a state of geostrophic balance, plus a propagating wavefield. The behaviour of the front for $T > \tau$ is thus identical to the large-time behaviour predicted by the model of ST13 for an unstrained flow. In that case unbalanced initial conditions were responsible for wave generation, whereas here the dual mechanisms of frontogenesis and acceleration of the large-scale strain flow themselves generate IGWs from purely balanced initial conditions.

While the impulse response solution presented in § 4 is no longer valid for a time-dependent strain field, many of the properties of that solution still apply. The characteristics of the sine-transformed generalised model equation (2.35) were derived in § 2.2 for both time-varying (2.37) and constant (2.38) strains. As noted in that section, the region between the plus ($\mathscr{X}_+$) and minus ($\mathscr{X}_-$) characteristics indicates the area over which amplitude/energy associated with vertical mode n, and initially at location $\mathscr{X}_0$, has spread by time T. The highest horizontal wavenumbers ($k \to \infty$) propagate along the lines $\mathscr{X}_\pm$, while the lowest wavenumbers ($k \to 0$) essentially remain fixed along the centre line, $\mathscr{X} = \mathscr{X}_0 \mathrm{e}^{-\beta(T)}$. Comparing (2.37) with (2.38), we expect the behaviour of time-dependent strain flows to be similar to the constant-strain case, except that the stagnation points will be time-dependent. In both cases, a pair of positive (or negative) characteristics initially separated by a distance $\Delta \mathscr{X}_0$ will be squeezed together by the convergent strain field, to a separation distance of $\Delta \mathscr{X}_0 \mathrm{e}^{-\beta(T)}$ by time T. This convergence implies that waves will be concentrated in a small region of width $\Delta \mathscr{X}_W = \mathrm{e}^{-\beta(T)}$ around the stagnation points as time progresses, leading to the formation and amplification of wavepackets in these narrow regions as seen in § 4. The relative amplitude $\mathscr{R}_A$ of the wave streamfunction compared to the GSC may be estimated using the ratio $\mathscr{R}$ derived previously (4.20) for the net integrated amplitudes, and the relative width of the regions in which the streamfunctions are concentrated,

$$\mathscr{R}_A = \mathscr{R} \frac{\Delta \mathscr{X}_{GSC}}{2 \Delta \mathscr{X}_W} = \mathscr{R} \frac{Bu \, \mathrm{e}^{\beta(T)}}{n\pi \, \delta} \approx \frac{Bu \, \mathrm{e}^{\beta(T)}}{n\pi \, \delta^2} \mathrm{e}^{-\pi/(2\delta)} \quad \text{for } \delta \leqslant 0.2, \qquad (5.1)$$

where δ is an appropriate mean (or maximum) value of the time-dependent strain. These results will prove useful in interpreting the numerical solution introduced in figure 7.

In figure 8, we examine the generation and properties of the IGWs in detail by dividing the flow into the secondary circulation and wave components. GSC (left-hand panel) is computed as defined in § 4, but evaluated at each time T using the value of the strain at that time; that is,

$$\psi_{GSC}(X, Z, T) = \int_{-\infty}^{\infty} \psi_{I,GSC}((X - X_0) \, \mathrm{e}^{-\beta(T)}, Z) \, b_0'(X_0) \, \mathrm{d}X_0. \qquad (5.2)$$

This part of the flow is directly forced, and only exists during the time $T < \tau$ where the strain is non-zero (see (4.15)). Given the form of the time-dependence in (5.2) it is clear that – as with the general constant-strain solution in § 4 – propagating waves will not be present in the GSC part of the flow. Defining the wave part of flow (subscript W, right-hand panel in figure 8) as the difference between the GSC and the
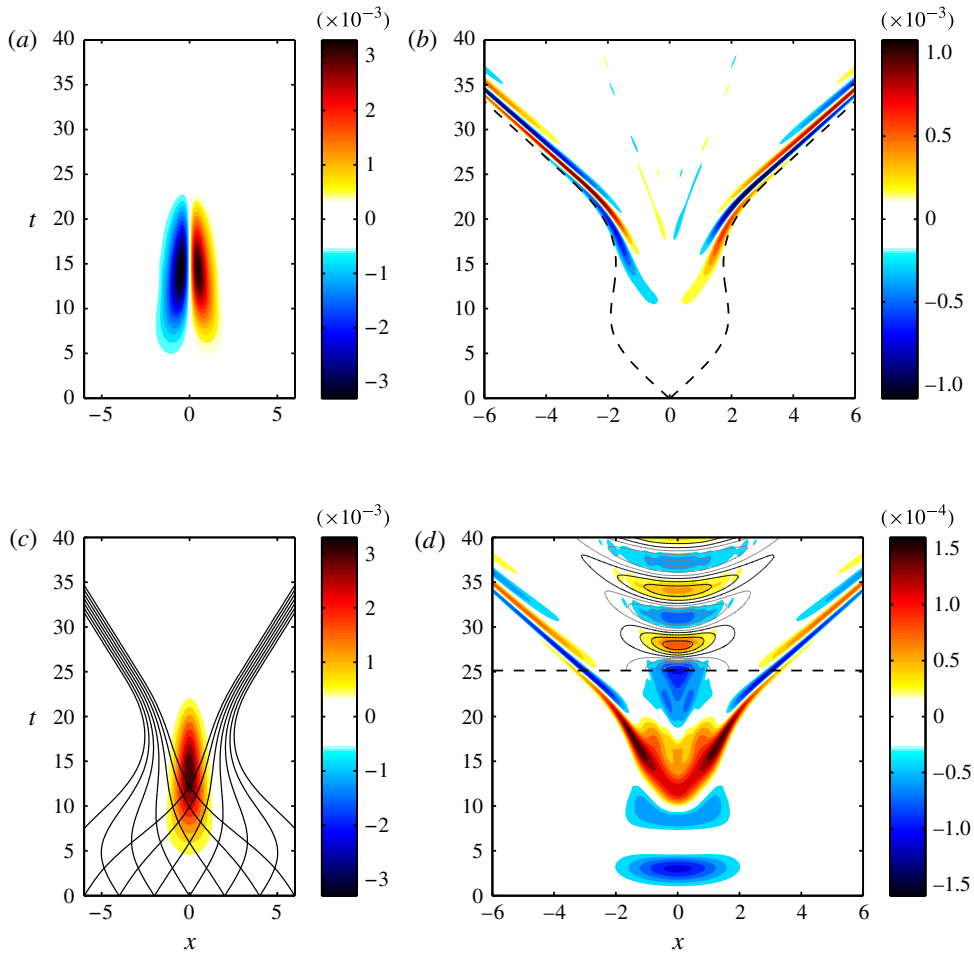
FIGURE 8. The solution of the time-dependent model (2.14) with $\mathcal{N} \equiv 0$, computed numerically, for parameter values of $Ro = 0.1$ and $Bu = 1$, as in figure 7. Hovmöller plots of the GSC and wave components of the vertical velocity (a,b) and streamfunction (c,d) are displayed, evaluated at mid-depth $(z = 0.5)$. The stagnation lines for the first vertical mode are shown on the (b) (black dashed lines; from (2.37) with $\mathcal{X}_0 = 0$). The family of characteristics $\mathcal{X}_+$ and $\mathcal{X}_-$ defined by (2.37) for the first vertical mode are shown on the (c) (solid black lines). Streamfunction contours for the unstrained adjustment of the maximum GSC for $T > \tau = 8\pi$ (shown by a dashed black line) are overlaid on the (c), with black denoting positive (anticlockwise) and grey negative (clockwise) overturning.

full flow, $\psi_W = \psi - \psi_{GSC}$, ensures that all propagating signals are contained within this component. In other words, while ψ_W is not a unique separation of the wavefield, it is an upper limit. As expected, the wavefield in figure 8 persists when the strain field is switched off. Hovmöller plots of the streamfunction (c,d) and vertical velocity (a,b) fields are displayed in the figure, evaluated at mid-depth. The wave streamfunction is at least an order of a magnitude weaker than the secondary circulation (see the colour bar scales).

We will first consider the frontogenesis waves. The amplitude of the frontogenesis wave streamfunction may be estimated by substitution of the parameter values

into (5.1), yielding a relative amplitude $\mathscr{R}_A = 0.04$ in agreement with figure 8. The maximum distance from the origin of the frontogenesis wavepacket with time is plotted in figure 8(*a*) as dashed black lines. These 'stagnation lines' for the first vertical mode are given by (2.37) with $\mathscr{X}_0 = 0$ and $n = 1$. Note that these stagnation lines converge between roughly $t = 10$ and $t = 20$ in the figure. During this time, wave energy cannot disperse and instead accumulates along these lines (as seen in §4), leading to the formation of a wavepacket. This behaviour is reinforced by the family of characteristics $\mathscr{X}_+$ and $\mathscr{X}_-$ for the first vertical mode (defined by (2.37)), shown in figure 8(*c*) as solid black lines. The convergence of the characteristics between $t = 10$ and $t = 20$ indicates the confinement of energy initially spread over a broad region ($\Delta\mathscr{X} \sim 1$) at $t = 0$ into a region of width $e^{-\beta(t)}$ by time t, leading to the formation of a (relatively) high-amplitude wavepacket. Correspondingly, an initial wavenumber of k_0 becomes $k = k_0 e^{\beta(t)}$ after a time t. By time τ, the integrated strain is $\beta \sim 2.5$, implying an order-of-magnitude amplification ($k/k_0 = e^\beta \sim 12$) of the initial wavenumbers. For $T > \tau$, the wavepacket is composed of a narrow band of high-wavenumber, high-frequency waves, which disperse weakly with time as the packet propagates outwards (at or near to the maximum speed of Bu/π). Specifically, the wavepacket is composed of only the first vertical mode $n = 1$ and a spread of horizontal mode numbers around $k = e^{\beta(t)}$. Applying the hydrostatic IGW dispersion relation (e.g. ST13), the dominant frequency of a first vertical mode wavepacket subjected to integrated strain β is

$$\omega = \sqrt{1 + \left(\frac{Bu\, e^\beta}{\pi}\right)^2}. \tag{5.3}$$

For the current example, the dominant frequency is thus $\omega \sim 4$, or $4f$ in dimensional units. The result is clearly highly sensitive to the integrated strain β – a smaller β would lead to a wavepacket consisting of a broader band of lower-wavenumber, lower-frequency, lower-amplitude waves that disperses more strongly. For example, an integrated strain of half the present value leads to a dominant frequency of $\omega \sim 1.5$, closer to the inertial frequency.

In contrast to the frontogenesis waves, the transience waves in figure 8 are dominated by low-wavenumber, near-inertial signals. The transience waves are generated as a result of the time-dependent strain changing the quasi-steady secondary circulation (GSC) that can be supported at the front. For instance, consider impulsively switching off the strain field at some time T_0. In the absence of strain, for $T > T_0$, the system possesses a steady state corresponding to geostrophic balance with zero secondary circulation. Thus, switching off the strain will trigger the geostrophic adjustment of the GSC towards this steady state. Such geostrophic adjustment of a now unstrained flow will be associated with the generation of freely propagating IGWs – or transience waves. For a smoothly switched off strain, as in figure 8, the same adjustment process gives rise to transience wave generation, but the wave amplitude will be smaller than in the impulsive case and the waves will be (partially) trapped until the strain vanishes. In either case, the wavenumber spectrum of the waves is controlled by the GSC streamfunction that is undergoing adjustment, implying lower wavenumbers, frequencies and group velocities compared with the frontogenesis waves.

The generation and propagation of waves due to geostrophic adjustment at a front was discussed in ST13. In that case, the initial condition considered was an unbalanced buoyancy gradient (or mass imbalance). Here, the 'initial condition' is the

state of the flow when the strain is switched off, and corresponds to an unbalanced streamfunction (or momentum imbalance). However, the dynamics of the adjustment process are unchanged and we can use the results of ST13 to understand the transience wave generation seen in figure 8. By approximating the smooth switch-off (over time scale $\tau/2$) of the strain field supporting a frontal secondary circulation of ψ_{GSC} at $\tau/2$ as an impulsive switch-off, we can derive an approximate response streamfunction $\psi_{T,W}$ for the transience wave generation for $T > \tau$. Following ST13, the Fourier mode amplitude of the response is

$$\widehat{\psi_{W,T}}(k, n, T) = \widehat{\psi_0} \, \cos\left(\sqrt{1 + \left(\frac{Bu\,k}{n\pi}\right)^2}\,(T - \tau)\right), \tag{5.4}$$

where k is the wavenumber in momentum coordinates. The contours of the (approximate) transience wave streamfunction $\psi_{W,T}$, computed from (5.4), are overlaid on the wave streamfunction plot in figure 8. The initial streamfunction ψ_0 has been taken as the GSC at $T = \tau/2$; that is, $\psi_0 = \psi_{GSC}(X, Z, \tau/2)$. The displayed contours of the approximate transience wave streamfunction closely match the full wave response for $T > \tau$.

Transience waves are also generated by the switching on of the strain field via an analogous adjustment mechanism to that discussed above. The waves are visible in the Hovmöller plot of ψ_W in figure 8(d) as near-inertial oscillations for $T < \tau/2$. In the limit of an impulsive switching on of the strain field at $T = 0$, these transience waves are identical to the waves associated with initial conditions that were studied in ST13.

6. Discussion

In §2, we introduced the governing equations for a quasi-2D, hydrostatic, Boussinesq, uniform PV fluid, trapped between rigid lids with an initial buoyancy gradient, $b'_0(X)$. A time-dependent strain field then amplifies this gradient and drives a flow in response. This configuration provides an idealised model to examine frontogenesis and the associated spontaneous wave generation occurring in eddy fields in the atmosphere and ocean. We derived a generalised model that is valid – in the sense that neglected terms are small compared to leading-order retained terms – if the product $Ro\,\delta$ is sufficiently small (as defined by (2.26)) and for appropriately balanced initial conditions. One subset of the generalised model is the weak-strain limit studied by Hoskins & Bretherton (1972), $\alpha^2 \ll f^2$, where α is the strain rate and f is the inertial frequency, in which the along-front velocity remains geostrophically balanced for all time and no wave motions exist. We call this weak-strain limit the 'HB model'. A second subset of the generalised model is the weak-front limit, as defined by (2.29), where the initial buoyancy gradient on the boundary must be small, but the strain α can be order f. The key feature of the weak-front limit is that the generalised model yields an accurate description of both IGW generation and the large-scale secondary circulation. The weak-front limit does not neglect any terms in the rotating fluid equations which are retained in the HB model, but instead retains additional terms relating to wave propagation. The additional terms vanish for sufficiently weak strains, and thus the weak-front limit reduces to the HB model. The time evolution of the system in the generalised and HB models are completely described by the PDEs in field ϕ, (2.14) with $\mathscr{N} \equiv 0$ and (2.33), respectively. The

along-front velocity, buoyancy and cross-front streamfunction may then be generated from ϕ for each limit via (2.13).

The generalised model was solved analytically in §4 for the case of a steady strain field, using a Green's function method. While the model solution breaks down for a sufficiently sharp front (as does the HB model), we are able to make some qualitative statements about the differences between the two models. First, a finite strain field has the effect of confining the frontal circulation within a region of $2NH/(\pi\alpha)$ approximately the deformation axis – we define a secondary circulation (GSC) that accounts for this effect and generalises the HB secondary circulation to finite strain. Secondly, a steady strain field drives the formation of stationary IGWs as the front sharpens. The first vertical mode tends to appear on the periphery of the frontal zone (at a distance of $\pm NH/(\pi\alpha)$ from the centre of the front). For small Ro, the front remains largely vertical and this mode is visible a distinct band of strong vertical flow ahead of the front. As such, this wave feature provides a potential dynamical model for the formation of squall lines ahead of cold fronts in the atmosphere. Ley & Peltier (1978) previously obtained a similar result from an *ad hoc* fast-time-scale correction to the solution of Hoskins & Bretherton (1972). Stationary waves associated with the higher vertical modes tend to overlay the frontal zone and add fine structure to the fields in this region. We emphasise that the formation of these stationary waves is independent of the initial conditions, suggesting that they should be a ubiquitous feature of fronts, at least for small Rossby numbers. Similar fine structure and squall line type features have been observed in numerical models of frontogenesis such as Garner (1989) and Snyder *et al.* (1993).

Since the generalised model only accurately describes second-order flow features in the limit of small Rossby number (2.29), the wave features described above are not expected to be quantitatively valid for larger Rossby numbers. However, even neglecting the wavefield, the effect of the strain is important in modifying the first-order secondary circulation for all Rossby numbers. Specifically, the GSC contains an additional dynamical/physical correction relative to the HB secondary circulation; namely, that spread of energy and momentum during frontogenesis is associated with a finite group velocity, which limits the outward spread of the frontal circulation into the oncoming strain flow. As such, this confinement effect is expected to be a ubiquitous feature of fronts for parameter values where the generalised model is applicable (2.26), which includes the weak-strain HB limit.

Spontaneous IGW emission at a front occurs via two mechanisms: large-scale flow acceleration (transience waves) and frontal sharpening (frontogenesis waves). The flow acceleration mechanism was isolated in the limit of zero-PV flow (§3) where inertial waves are generated via a time-varying strain flow independent of the frontal sharpness. In §4, it was demonstrated that stationary IGWs evolve naturally from a front subject to constant strain, as the frontal scale contracts. In the more general situation of an initially weak, balanced front subjected to a time-dependent strain field (as in §5), both mechanisms are active. With a time-varying strain flow, the frontogenesis IGWs are initially confined by the convergent strain field and wave energy is forced to accumulate at locations set by the vertical mode number, leading to the formation of high-amplitude wavepackets as the frontal scale shrinks. As the strain weakens, these wavepackets begin to propagate away from the front. The time-dependent strain also generates transience waves. These waves arise since a change in the strain magnitude implies a change in the magnitude of the secondary circulation that can be sustained at a front, thus giving rise to an adjustment process. The transience waves tend to exhibit lower wavenumbers and frequencies than

frontogenesis waves, although the propagation of the waves away from the frontal zone can still only occur once the strain field weakens sufficiently. In general, it is likely to be difficult to determine which wave generation mechanism is responsible for individual features in observed flow fields, particularly as waves triggered by initial conditions may also be present.

Using the analytical solutions developed in §4 we were able to compute estimates of the relative amplitude of the frontogenesis wave streamfunction compared to the secondary circulation (the 'mean flow'). A scaling for the relative wave amplitude was given in (5.1). In terms of the physical scales of the problem, this reduces to

$$\mathcal{R}_A = \frac{N H f}{\pi \, \alpha^2 \, L_F} \, \mathrm{e}^{-(\pi f / 2\alpha)}. \tag{6.1}$$

The parameter L_F in (6.1) is the minimum frontal width, which in our model is given by $L_F = L \, \mathrm{e}^{-\beta(t)}$. A key feature of (6.1) is that it implies exponentially small wave generation for strain rates α smaller than $\sim 0.1f$–$0.2f$, but finite generation above this threshold. Equation (6.1) may thus provide a theoretical basis for parametrisations of gravity-wave emission from frontal zones in numerical models of the atmosphere, which at present are based on largely arbitrary thresholds (Plougonven & Zhang 2014).

Our results may have important implications for energy loss from balanced flows via IGW emission. Danioux *et al.* (2012) demonstrate that wave generation in an eddy field is highly intermittent in both space and time. In particular, wavepackets tend to be emitted in discrete 'generation events' from thin filamentary density structures associated with high-strain regions. Our model provides a possible explanation for this intermittency. In §4, we demonstrated that wave generation via frontal sharpening is exponentially small for small strain (see (4.20) and figure 2) but becomes first order for larger strain, and consequently we expect a localisation of wave emission to small regions of high strain. Further, as noted above, waves generated via convergent straining across a front remain trapped in the vicinity of the front while the strain remains large and then begin to propagate as distinct wavepackets as the strain weakens – for example, by advection of the frontal feature into a region of weaker strain in an eddy field. Consequently, we expect a tendency for the emission of wavepackets to be highly localised in time as well as space. The localisation in time due to trapping by the strain field emphasises that the effect of the large-scale strain on the propagation of waves is vitally important in determining the structure of the observed wavefield (e.g. Plougonven & Snyder 2005).

The present model provides a simple, idealised theoretical framework from which to study IGW generation at fronts in more complex situations. It should be noted, however, that other factors – including spatially variable mean and eddy flows, vertical variations in shear and stratification, along-front variations, and many other features of more realistic flows, all of which are neglected here – can significantly influence the propagation and generation of waves. Furthermore, the generalised model developed herein is only stringently valid for small Rossby numbers, whereas many fronts in the ocean and atmosphere have order-one Rossby numbers. Despite this, the qualitative similarity of the wavefields in our analytical solutions to those observed in prior numerical models of the same system at order-one Rossby numbers (e.g. Snyder *et al.* 1993) suggest that the mechanism of spontaneous wave generation may be well captured by the generalised model. The effects of larger Rossby numbers and the other factors noted above will be examined in a future study directly comparing the model predictions developed here to the observed frontogenesis and IGW generation in a fully nonlinear numerical model.

Appendix A. Green's function solution

Here we show that, as stated in the text, the convolution given by (4.1) – with the impulse response defined by (4.4) – is the forced solution to the generalised model (PDE (2.14) with $\mathcal{N} \equiv 0$) when the strain δ is constant. We begin by taking a Fourier transform in X of (2.14) with δ constant to obtain

$$\left[\frac{\partial^2}{\partial Z^2} \left(\frac{\partial^2}{\partial T^2} + 1 - \delta^2 \right) - k^2 Bu^2 e^{2\beta(T)} \right] \widehat{\phi}(k, Z, T) = -\iota k \, Ro \, e^{\delta T} \, \widehat{b}_0, \qquad \text{(A 1)}$$

where k is the horizontal wavenumber in generalised momentum coordinates and a hat denotes the Fourier transform. The forced solution is assumed to have the form given by (4.1), the Fourier transform of which is

$$\widehat{\phi}(k, Z, T) = -\iota k \, \widehat{b}_0 \, e^{\delta T} \, \widehat{\phi}_I(k \, e^{\delta T}, Z). \qquad \text{(A 2)}$$

Substituting (A 2) into (A 1) and simplifying yields

$$\left[\frac{\partial^2}{\partial Z^2} \left(1 + 3\delta^2 \, \mathcal{K} \frac{\partial}{\partial \mathcal{K}} + \delta^2 \, \mathcal{K}^2 \frac{\partial^2}{\partial \mathcal{K}^2} \right) - \mathcal{K}^2 Bu^2 \right] \widehat{\phi}_I(\mathcal{K}, Z) = Ro, \qquad \text{(A 3)}$$

where $\mathcal{K} = k \, e^{\delta T}$ is the wavenumber in regular momentum coordinates (i.e. $\mathcal{X} = X \, e^{-\delta T}$). At this point, we make use of the Fourier identities

$$\mathscr{F}^{-1} \left[\mathcal{K}^n \widehat{f}(\mathcal{K}) \right] = \iota^n \frac{\partial^n}{\partial \mathcal{X}^n} f(\mathcal{X}), \qquad \mathscr{F}^{-1} \left[\frac{\partial^n}{\partial \mathcal{K}^n} \widehat{f}(\mathcal{K}) \right] = \iota^n \, \mathcal{X}^n f(\mathcal{X}), \qquad \text{(A 4a,b)}$$

where $\mathscr{F}^{-1}[\,]$ denotes the inverse Fourier transform. Inverse Fourier transforming (A 3) with respect to $\mathcal{K}$, and applying the identities (A 4), yields

$$\left[\frac{\partial^2}{\partial Z^2} \left(\delta^2 \, \mathcal{X}^2 \frac{\partial^2}{\partial \mathcal{X}^2} + \delta^2 \, \mathcal{X} \frac{\partial}{\partial \mathcal{X}} + 1 - \delta^2 \right) + Bu^2 \frac{\partial^2}{\partial \mathcal{X}^2} \right] \phi_I(\mathcal{X}, Z) = Ro \, \underline{\delta}(\mathcal{X}).$$
$$\text{(A 5)}$$

Equation (A 5) is identical to (4.4) in the main text. Thus, (4.3) with the impulse response defined by (4.4) is the forced solution to the generalised model at constant strain.

Note that the general solution of the generalised model (PDE (2.14) with $\mathcal{N} \equiv 0$) at constant strain is the sum of the forced solution derived above, plus the wave solutions associated with initial conditions, which were studied in ST13 (see (4.23) of ST13).

Appendix B. Boundary conditions on the Green's function solution

Here we show, using the characteristics of the constant-strain hydrostatic generalised model (2.38), that: (i) the frontal circulation is identically zero for $|\mathcal{X}| > Bu/(n\pi\delta)$ in the limit of infinite time; and (ii) the appropriate boundary conditions on the Green's function derived in §4 are $\widehat{\phi}_I(|\mathcal{X}| > Bu/(n\pi\delta), n) = 0$.

Consider the region of dependence of a point $\mathscr{X}_1$ just outside of the confinement region for the nth vertical mode (defined as $|\mathscr{X}| < Bu/(n\pi\delta)$). That is, let

$$\mathscr{X}_1 = \frac{Bu}{n\pi\delta} + \Delta\mathscr{X} \tag{B 1}$$

for $\Delta\mathscr{X}$ a small positive constant. The region of dependence of the point $\mathscr{X}_1$ at time T_1 is bordered by the positive ($\mathscr{X}_+$) and negative ($\mathscr{X}_-$) characteristics (from (2.38)) that intersect at this point. Thus, the region of dependence at time T_1 is $\mathscr{X}_0 \leqslant \mathscr{X} \leqslant \mathscr{X}_2$, where

$$\mathscr{X}_0 = \frac{Bu}{n\pi\delta} + \Delta\mathscr{X}\, e^{\delta T_1} \quad \text{and} \quad \mathscr{X}_2 = \frac{Bu}{n\pi\delta}\left(2e^{\delta T_1} - 1\right) + \Delta\mathscr{X}\, e^{\delta T_1}. \tag{B 2a,b}$$

Now consider the limit $T_1 \to \infty$ and define

$$\mathscr{X}_0^\infty = \lim_{T_1 \to \infty} \mathscr{X}_0 = \lim_{T_1 \to \infty}\left[\frac{Bu}{n\pi\delta} + \Delta\mathscr{X}\, e^{\delta T_1}\right]. \tag{B 3}$$

If we apply boundary conditions to ensure that ϕ (and its vertical sine transform $\widehat{\phi}$) vanish infinitely far from the front such that

$$\lim_{|\mathscr{X}| \to \infty} \widehat{\phi}(\mathscr{X}, n, T) = 0 \tag{B 4}$$

for all times, it follows that

$$\widehat{\phi}(\mathscr{X} > \mathscr{X}_0^\infty, n, T) = 0. \tag{B 5}$$

In other words, the boundary condition ensures that $\widehat{\phi}$ is everywhere zero in the region of dependence of the point $\mathscr{X}_1$: $\widehat{\phi}(\mathscr{X}_0 \leqslant \mathscr{X} \leqslant \mathscr{X}_2, n, 0) = 0$. Further, we choose a boundary buoyancy gradient profile $b_0'(X = \mathscr{X}\, e^{\delta T})$ such that $b_0'(X > \mathscr{X}_0^\infty\, e^{\delta T}) = 0$. The forcing to the model PDE (2.35), $Ro\, A_n\, e^{\beta(T)} b_0'(X)$, is therefore everywhere zero in the region of dependence. Thus, as $T_1 \to \infty$, since the initial condition and forcing are zero in the region of dependence, it follows that

$$\widehat{\phi}\left(\mathscr{X}_1 = \frac{Bu}{n\pi\delta} + \Delta\mathscr{X},\ n,\ T_1 \to \infty\right) = 0. \tag{B 6}$$

Since this result applies for any $\Delta\mathscr{X}$, it follows that

$$\lim_{T \to \infty} \widehat{\phi}(\mathscr{X}, n, T) = 0 \quad \text{for } \mathscr{X} > \frac{Bu}{n\pi\delta}. \tag{B 7}$$

A similar argument may be made to show that

$$\lim_{T \to \infty} \widehat{\phi}(\mathscr{X}, n, T) = 0 \quad \text{for } \mathscr{X} < -\frac{Bu}{n\pi\delta}. \tag{B 8}$$

Thus we conclude that the solution to (2.35) is identically zero for $|\mathscr{X}| > Bu/(n\pi\delta)$ in the limit $T \to \infty$, and hence the frontal circulation is entirely confined in the region $|\mathscr{X}| \leqslant Bu/(n\pi\delta)$. Further, since the nth vertical mode Green's function $\widehat{\phi}_I(\mathscr{X}, n)$ is the infinite time limit of $\widehat{\phi}(\mathscr{X}, n, T)$, as per (4.6), the appropriate boundary conditions for the Green's function are $\widehat{\phi}_I(|\mathscr{X}| > Bu/(n\pi\delta), n) = 0$.

REFERENCES

ALFORD, M. H., SHCHERBINA, A. Y. & GREGG, M. C. 2013 Observations of near-inertial internal gravity waves radiating from a frontal jet. *J. Phys. Oceanogr.* **43**, 1225–1239.

BLUMEN, W. 1972 Geostrophic adjustment. *Rev. Geophys. Space Phys.* **10**, 485–528.

BLUMEN, W. 2000 Inertial oscillations and frontogenesis in a zero potential vorticity model. *J. Phys. Oceanogr.* **30**, 31–39.

BLUMEN, W. & WU, R. 1995 Geostrophic adjustment: frontogenesis and energy conversion. *J. Phys. Oceanogr.* **25**, 428–438.

BUHLER, O. & MCINTYRE, M. E. 2005 Wave capture and wave-vortex duality. *J. Fluid Mech.* **534**, 67–95.

DANIOUX, E., VANNESTE, J., KLEIN, P. & SASAKI, H. 2012 Spontaneous inertia–gravity-wave generation by surface-intensifed turbulence. *J. Fluid Mech.* **699**, 153–1732.

FORD, R., MCINTYRE, M. E. & NORTON, W. A. 2000 Balance and the slow quasimanifold: some explicit results. *J. Atmos. Sci.* **57**, 1236–1254.

GARNER, S. T. 1989 Fully Lagrangian numerical solutions of unbalanced frontogensis and frontal collapse. *J. Atmos. Sci.* **46**, 717–739.

GILL, A. E. 1976 Adjustment under gravity in a rotating channel. *J. Fluid Mech.* **77**, 603–621.

GRIFFITHS, M. & REEDER, M. J. 1996 Stratospheric inertia–gravity waves generated in a numerical model of frontogenesis. I: model solutions. *Q. J. R. Meteorol. Soc.* **122**, 1153–1174.

HOSKINS, B. J. 1982 The mathematical theory of frontogenesis. *Annu. Rev. Fluid Mech.* **14**, 131–151.

HOSKINS, B. J. & BRETHERTON, F. P. 1972 Atmospheric frontogenesis models: mathematical formulation and solution. *J. Atmos. Sci.* **29**, 11–37.

JONES, W. L. 1969 Ray tracing for internal gravity waves. *J. Geophys. Res.* **74** (8), 2028–2033.

JUCKES, M. 1994 Quasigeostrophic dynamics of the tropopause. *J. Atmos. Sci.* **51**, 2756–2768.

KARAN, H., FITZPATRICK, P. J., HILL, C. M., LI, Y., XIAO, Q. & LIM, E. 2010 The formation of multiple squall lines and the impacts of WSR-88D radial winds in a WRF simulation. *J. Atmos. Sci.* **25**, 242–262.

KUO, A. C. & POLVANI, L. M. 1997 Time-dependent fully nonlinear geostrophic adjustment. *J. Phys. Oceanogr.* **27**, 1614–1634.

LEY, B. E. & PELTIER, W. R. 1978 Wave generation and frontal collapse. *J. Atmos. Sci.* **35**, 3–17.

MCINTYRE, M. E. 2009 Spontaneous imbalance and hybrid vortex-gravity structures. *J. Atmos. Sci.* **66**, 1315–1326.

OU, H. W. 1984 Geostrophic adjustment: a mechanism for frontogenesis. *J. Phys. Oceanogr.* **14**, 994–1000.

PLOUGONVEN, R. & SNYDER, C. 2005 Gravity waves excited by jets: propagation versus generation. *Geophys. Res. Lett.* **32**, L18802.

PLOUGONVEN, R. & ZHANG, F. 2007 On the forcing of inertia–gravity waves by synoptic-scale flows. *J. Atmos. Sci.* **64**, 1737–1742.

PLOUGONVEN, R. & ZHANG, F. 2014 Internal gravity waves from jets and fronts. *Rev. Geophys.* **52**, 33–76.

POLZIN, K. L. 2010 Mesoscale eddy-internal wave coupling. Part II: energetics and results from PolyMode. *J. Phys. Oceanogr.* **40**, 789–801.

REEDER, M. J. & GRIFFITHS, M. 1996 Stratospheric inertia–gravity waves generated in a numerical model of frontogenesis. II: wave sources, generation mechanisms and momentum fluxes. *Q. J. R. Meteorol. Soc.* **122**, 1175–1195.

ROSSBY, C. G. 1938 On the mutual adjustment of pressure and velocity distributions in certain simple current systems, II. *J. Mar. Res.* **1**, 239–263.

SHAKESPEARE, C. J. & TAYLOR, J. R. 2013 A generalised mathematical model of geostrophic adjustment and frontogenesis: uniform potential vorticity. *J. Fluid Mech.* **736**, 366–413.

SHCHERBINA, A. Y., D'ASARO, E. A., LEE, C. M., KLYMAK, J. M., MOLEMAKER, M. J. & MCWILLIAMS, J. C. 2013 Statistics of vertical vorticity, divergence, and strain in a developed submesoscale turbulence field. *Geophys. Res. Lett.* **40**, 4706–4711.

SNYDER, C., SKAMAROCK, W. & ROTUNNO, R. 1993 Frontal dynamics near and following frontal collapse. *J. Atmos. Sci.* **50**, 3194–3211.

STONE, P. H. 1966 Frontogenesis by horizontal wind deformation fields. *J. Atmos. Sci.* **23**, 455–565.

TANDON, A. & GARRETT, C. 1994 Mixed layer restratification due to a horizontal density gradient. *J. Phys. Oceanogr.* **24** (6), 1419–1424.

THOMAS, L. N., TANDON, A. & MAHADEVAN, A. 2008 Submesoscale processes and dynamics. In *Ocean Modelling in an Eddying Regime*, Geophysical Monograph Series, vol. 177. American Geophysical Union.

VANNESTE, J. 2008 Exponential smallness of inertia–gravity wave generation at small Rossby number. *J. Atmos. Sci.* **65**, 1622–1637.

VANNESTE, J. 2013 Balance and spontaneous wave generation in geophysical flows. *Annu. Rev. Fluid Mech.* **45**, 147–172.

VIUDEZ, A. 2007 The origin of the stationary frontal wavepacket spontaneously generated in rotating stratified vortex dipoles. *J. Fluid Mech.* **593**, 359–383.

VIUDEZ, A. & DRITSCHEL, D. G. 2006 Spontaneous generation of inertia–gravity wave packets by balanced geophysical flows. *J. Fluid Mech.* **553**, 107–117.

WILLIAMS, P. D., HAINE, T. W. N. & READ, P. L. 2008 Inertia–gravity waves emitted from balanced flow: observations, properties, and consequences. *J. Atmos. Sci.* **65**, 3543–3556.

WILLIAMS, R. T. & PLOTKIN, J. 1968 Quasi-geostrophic frontogenesis. *J. Atmos. Sci.* **25**, 201–206.

WUNSCH, C. & FERRARI, R. 2004 Vertical mixing, energy and the general circulation of the oceans. *Annu. Rev. Fluid Mech.* **36**, 281–314.

ZHANG, F. 2004 Generation of mesoscale gravity waves in upper-tropospheric jet-front systems. *J. Atmos. Sci.* **61**, 440–457.