

On the vertical-shear instability in astrophysical discs

A. J. Barker[★] and H. N. Latter

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

Accepted 2015 March 22. Received 2015 March 12; in original form 2015 January 30

ABSTRACT

We explore the linear stability of astrophysical discs exhibiting vertical shear, which arises when there is a radial variation in the temperature or entropy. Such discs are subject to a ‘vertical-shear instability’, which recent non-linear simulations have shown to drive hydrodynamic activity in the MRI-stable regions of protoplanetary discs. We first revisit locally isothermal discs using the quasi-global reduced model derived by Nelson et al. This analysis is then extended to global axisymmetric perturbations in a cylindrical domain. We also derive and study a reduced model describing discs with power-law radial entropy profiles (‘locally polytropic discs’), which are somewhat more realistic in that they possess physical (as opposed to numerical) surfaces. The fastest growing modes have very short wavelengths and are localized at the disc surfaces (if present), where the vertical shear is maximal. An additional class of modestly growing vertically global body modes is excited, corresponding to destabilized classical inertial waves (‘r modes’). We discuss the properties of both types of modes, and stress that those that grow fastest occur on the shortest available length-scales (determined either by the numerical grid or the physical viscous length). This ill-posedness makes simulations of the instability difficult to interpret. We end with some brief speculation on the non-linear saturation and resulting angular momentum transport.

Key words: accretion, accretion discs – hydrodynamics – instabilities – waves – planetary systems.

1 INTRODUCTION

Accretion through magnetorotational turbulence is only viable in sufficiently ionized regions of protoplanetary discs, namely at their inner and outer radii (Balbus & Hawley 1998; Armitage 2011). Between 1–20 au (the ‘dead zone’) non-ideal effects extinguish the MRI, and instead accretion may occur via magnetocentrifugally launched winds (e.g. Bai 2014; Lesur, Kunz & Fromang 2014). However, identifying additional hydrodynamic mechanisms for driving turbulence is essential, due to its potential impact on the dynamics of solids, and therefore for planet formation.

Though pure Keplerian shear flow is difficult to destabilize, several mechanisms have been proposed: subcritical baroclinic instability (Petersen, Stewart & Julien 2007; Lesur & Papaloizou 2010), convective instability (Ruden, Papaloizou & Lin 1988; Lesur & Ogilvie 2010) and gravitational instability (Toomre 1964; Lin & Pringle 1987), to name but a few. Another mechanism that has recently received attention is the ‘vertical-shear instability’ (hereafter VSI) which, as its name suggests, attacks rotating systems that exhibit vertical shear (Urpin & Brandenburg 1998; Urpin 2003).

Fundamentally, the VSI is a form of centrifugal instability and is a close cousin of the Goldreich–Schubert–Fricke instability, originally applied to stellar interiors (Goldreich & Schubert 1967; Fricke 1968). Observations of protostellar discs (Andrews & Williams 2005) and theoretical models of passively heated discs (Chiang & Goldreich 1997) suggest that they too should display destabilizing vertical shear, generated from radial variations in temperature or entropy. Recent numerical simulations indicate that the non-linear evolution of the VSI can produce hydrodynamic turbulence and modest levels of angular momentum transport (Nelson, Gressel & Umurhan 2013; Stoll & Kley 2014). It thus could be a potentially key player in the dynamics of protoplanetary disc dead zones.

The VSI was originally studied with a local (Boussinesq) approach by Urpin & Brandenburg (1998) and by Urpin (2003). More recently Nelson et al. (2013) described it with a quasi-global model that captures the full vertical structure of growing anelastic modes in radial geostrophic balance (assuming the background is locally isothermal). The VSI also appears in a further generalization of the vertically stratified shearing box, which dispenses with these two constraints (McNally & Pessah 2014). However, several properties of the linear VSI require further explanation, especially with respect to its global manifestation in more realistic disc models. This is a

[★] E-mail: adrianjohnbarker@gmail.com

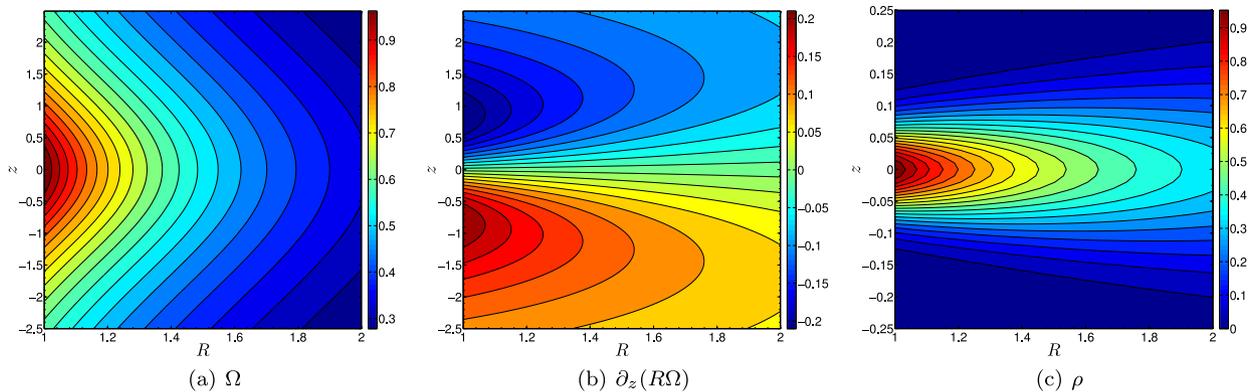


Figure 1. Basic state for the locally isothermal disc with $q = -1$, $p = -1.5$ and $c_0 = 0.05$. The left-hand panel shows a contour plot of Ω on the (R, z) plane. The middle panel is a similar contour plot, but this shows the magnitude of the vertical shear $\partial_z(R\Omega)$, which has a maximum at $|z| \sim 1$ (whereas the scaleheight at the inner radial boundary is 0.05). The right-hand panel shows the density ρ .

particularly important issue when trying to connect the linear theory to global simulations, and in interpreting their non-linear outcome. Our paper is devoted to exploring this aspect of the problem.

We perform linear stability analyses of astrophysical discs exhibiting global variations in temperature and entropy, and as a consequence vertical shear. We employ locally isothermal and polytropic models in both quasi-global and fully global 2D geometries, which revise and extend previous work.

In agreement with Nelson et al. (2013), we find that the VSI excites two classes of modes. The first class corresponds to classical free inertial waves (r modes) that are present in any astrophysical disc (Lubow & Pringle 1993; Korycansky & Pringle 1995; Kato 2001) but which have been destabilized by the vertical shear. These, referred to as ‘body modes’, grow at modest rates and typically exhibit longer wavelengths (though the radial wavelength of the waves is still short).

The second class corresponds to modes localized to the vertical surfaces of the disc where the vertical shear is maximal. These grow much faster and have very short wavelengths, making them difficult to resolve numerically. In fact, unless viscosity is included, the fastest growing modes possess arbitrarily small wavelengths, making their simulation problematic. Note that, though they have been termed ‘surface modes’, these are different to the classical surface gravity waves that appear in polytropic disc models, as they lie in a different frequency range; they are hence a form of localized low-frequency inertial wave. Strict isothermal models do not possess a physical vertical surface and hence do not support these surface modes. Polytropic disc models do, however, as should any realistic disc model that possesses a transition between an optically thick interior and an optically thin ‘corona’.

We begin by explaining why a radial variation in entropy or temperature generally leads to vertical shear in Section 2. There we also explain why such discs are likely to be unstable. After defining our basic disc models in Section 3, we analyse the resulting VSI in the locally isothermal disc in Sections 4 and 6 and the locally polytropic disc in Section 5. Finally, we will discuss the implications of our results in Section 7, where we also speculate on the non-linear evolution of the VSI and its efficiency at transporting angular momentum.

2 VERTICAL-SHEAR INSTABILITY

Discs with radial variations in temperature or entropy necessarily possess vertical shear. To see that this must be, consider the ‘thermal

wind equation’ (the azimuthal component of the vorticity equation for the axisymmetric basic state of the disc):

$$\partial_z(R\Omega^2) = -\mathbf{e}_\phi \cdot (\nabla\rho \times \nabla P) / \rho^2 \quad (1)$$

$$= \partial_R T \partial_z S - \partial_z T \partial_R S. \quad (2)$$

Here we have adopted cylindrical polar coordinates centred on the central object (R, ϕ, z) and ρ, P, S and T are the basic state density, pressure, specific entropy and temperature profiles, respectively. Equation (1) tells us that a radial variation in the background temperature or entropy generates a departure from cylindrical rotation through the baroclinic terms on the right-hand side. Thus the angular velocity $\Omega = \Omega(R, z)$, and consequently the disc exhibits a weak vertical shear. For illustration, we show the angular velocity and vertical shear for a disc with a radial variation in temperature in Fig. 1, and the vertical shear for a disc with a radial variation in entropy in Fig. 2 (both disc models and the notation adopted are defined in Section 3).

2.1 Physical picture

Vertical shear provides a source of free energy that can drive hydrodynamic instabilities. How might modes access this free energy? Consider a ring of fluid at a given location (A) within the disc with coordinates (R_A, z_A) , and hence specific angular momentum $h_A = R_A^2 \Omega(R_A, z_A)$. If we vertically displace this ring to a new position (B) with coordinates $(R_A, z_A + \Delta z)$, then its specific angular momentum will be conserved as long as viscosity can be neglected (i.e. we assume that $|\Delta z|$ is much larger than the viscous length). But if the angular momentum of fluid at the new location h_B is smaller (larger) than h_A , then the ring will be pushed outwards (inwards) by the centrifugal acceleration $(h_A^2 - h_B^2)/R_A^3$, leading to a dynamical instability. Given that $h_B^2 \approx h_A^2 + \Delta z \partial_z h^2$, instability occurs whenever $\partial_z h^2 < 0$ (or indeed > 0), i.e. if there is any vertical shear. Basically, this interchange of rings of fluid reduces the total energy of the configuration, leading to an instability that transports angular momentum in order to eliminate the vertical shear.¹ This

¹ Our illustrative perturbation is vertical for simplicity; any displacement lying within the angle between the rotation axis and a surface of constant angular momentum will do (as explained in Knobloch & Spruit 1982, for example).

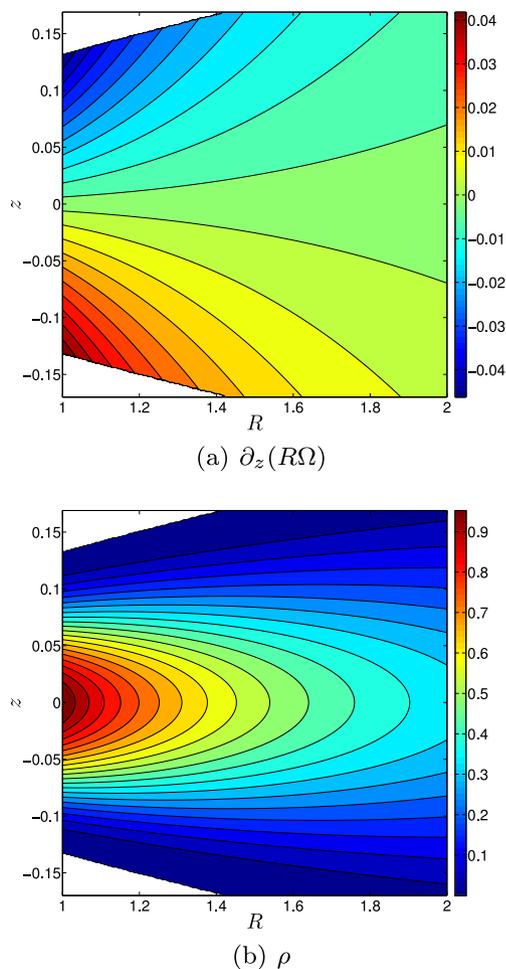


Figure 2. Basic state for the locally polytropic disc with $q_s = -1$, $p = -1.5$, $a_0 = 0.05$ and $\gamma = 1.4$ (therefore $H(R=1) \approx 0.13$). The top panel shows a contour plot of the vertical shear $\partial_z(R\Omega)$ on the (R, z) plane. For given R , the magnitude of the vertical shear has a maximum near the disc surface. The bottom panel shows the density ρ . Regions outside the disc surface where $\rho = 0$ are coloured white in both panels. (The angular velocity is not shown since the departure of constant Ω surfaces from the vertical is very small.)

is a modified form of Rayleigh’s argument for centrifugal instability. Though accretion discs are stable according to the classical Rayleigh criterion, any vertical shear permits its circumvention and hence the onset of instability.

This physical argument works for neutrally stratified discs, but must be altered when a stable vertical entropy stratification is present, as exhibited by protoplanetary disc dead zones. So we next introduce buoyancy and thermal diffusion. Buoyancy forces impede exchanges of the type described, and thus inhibit any adiabatic (dynamical) instability (cf. the Solberg–Høiland criterion). Instability is nevertheless possible if the buoyancy forces are eliminated, such as by sufficiently fast cooling or thermal diffusion. For this to work, displacements $|\Delta z|$ must then be much shorter than the thermal diffusion scale. The resulting instability is hence double-diffusive in character, possessing length-scales lying in the range bounded from below by the viscous length and above by the thermal diffusion length. Originally identified in the 1960s and applied to stellar interiors (Goldreich & Schubert 1967; Fricke 1968), only

much later was it recognized that such an instability could emerge in astrophysical discs (Urpin & Brandenburg 1998).²

2.2 Estimates from a local model

According to a local Boussinesq analysis, the growth rate of the VSI is

$$\sigma \approx |\partial_z(R\Omega)| \sim \epsilon |q| \Omega, \quad (3)$$

where $\epsilon = H/R$ is the disc aspect ratio and q is the exponent in the power law for temperature (or entropy), so that $T \propto R^q$ (Urpin & Brandenburg 1998; Urpin 2003; Nelson et al. 2013). For $|q| \sim 1$, the VSI will hence grow, and presumably saturate, relatively quickly on a time-scale not far from the dynamical time for thicker discs.

The VSI afflicts intermediate length-scales ℓ in the range

$$\ell_v \lesssim \ell \lesssim \ell_\chi, \quad (4)$$

where the viscous and thermal diffusion lengths are defined through

$$\ell_v = (\nu/\sigma)^{\frac{1}{2}}, \quad \ell_\chi = (\chi/N_z^2)^{\frac{1}{2}}. \quad (5)$$

Here ν is the kinematic viscosity, χ is the thermal diffusivity and $N_z > 0$ is the vertical buoyancy frequency. Length-scales above ℓ_χ are stabilized by buoyancy forces (as long as $N_z^2 > 0$); only when these are subdued by sufficiently rapid thermal diffusion is instability possible. On length-scales smaller than ℓ_v , viscosity neutralizes the excess angular momentum of a fluid element too quickly to allow the mode to grow. The fastest growing modes have vertical (k_z) and radial wavenumbers (k_R) that satisfy

$$\frac{k_z}{k_R} = \frac{\ell_R}{\ell_z} \lesssim \epsilon q, \quad (6)$$

so the radial wavelength of the mode (ℓ_R) is typically much shorter than the vertical wavelength (ℓ_z) (Urpin & Brandenburg 1998; Urpin 2003). Hence, being inertial waves, the group velocity points almost vertically, in accordance with the direction of transport outlined in the previous subsection. Note that modes grow at or near the fastest rate on short length-scales all the way to the viscous cut-off ℓ_v .

To give some sense of the numbers, the growth rate for the VSI is typically $\sigma \sim 0.3 \text{ yr}^{-1}$, if $\epsilon \sim 0.05$ at 1 au and $|q| \sim 1$. The microscopic kinematic viscosity at the mid-plane of a protoplanetary disc at 1 au may be estimated as $\nu \sim 2.5 \times 10^5 \text{ cm}^2 \text{ s}^{-1}$, yielding $\ell_v \sim 10^2 \text{ km}$, much shorter than the local disc thickness H (of the order of 10^7 km). The thermal diffusivity, on the other hand, is significantly larger, $\chi \sim 5 \times 10^{12} \text{ cm}^2 \text{ s}^{-1}$ (taking values typical for H_2 in a minimum mass solar nebula). Note that χ varies significantly with height in the disc (e.g. Bell & Lin 1994), but we ignore this complication when making simple estimates. It is unclear what value N_z should take in a dead zone in a protoplanetary disc, but it is probably much smaller than Ω . Therefore a (very crude) lower bound on the thermal diffusion length is $\ell_\chi \gtrsim 10^5 \text{ km}$. In all likelihood, however, $\ell_\chi \sim H$, not least because thermal diffusion becomes stronger as we approach the photosphere.

These estimates suggest the following ordering: $H \sim \ell_\chi \gg \ell_v$, with the VSI occurring in the wide gulf separating the viscous

² Urpin & Brandenburg (1998) coined the term ‘vertical-shear instability’. However, there is a good case for the retention of the name ‘Goldreich–Schubert–Fricke (GSF)’ instability, even if ‘VSI’ has the merit of clearly advertising the underlying physics.

and thermal diffusion lengths. We might expect the fastest growing modes to localize in regions of greatest vertical shear; in any realistic disc model, the magnitude of the vertical shear increases with distance from the mid-plane, and takes its maximum value at the disc surface.³ So we might expect the fastest growing modes to occur on very short length-scales (just above $\ell_v \sim 100$ km) located near the surface.

2.3 Non-linear evolution

We expect that the VSI will work to eradicate the destabilizing conditions from which it arose (the vertical shear) and ultimately return the system to a marginally stable, cylindrical rotation profile. Of course, resisting the VSI will be the driver of the vertical shear itself: the radiation field of the protostar. In the struggle between these two opponents, the system will probably reach a quasi-steady ‘balance’ in which the vertical shear is diminished (but not entirely removed) and some degree of hydrodynamical activity remains. The properties of this state will be determined by the relative efficiency of the VSI in wiping away the shear versus that of the protostar’s thermal driving. Presumably, if the VSI is inefficient then significant vertical shear persists and, as a consequence, significant turbulent motions.⁴ The locally isothermal simulations of Nelson et al. (2013) best describe this scenario, because the destabilizing gradients are fixed and cannot be modified by the VSI. On the other hand, if the thermal driving is weak, the VSI should eliminate the vertical shear, and subsequently its motions will settle down to a much lower level, because the system is near (if not at) marginal stability. This situation can be observed in simulations where the unstable equilibria possess long relaxation times (Nelson et al. 2013) or in cases where the disc is not thermally driven (Stoll & Kley 2014).

However it is generated, the properties of the final quasi-steady state are of key importance to dead zones, intervening potentially in the dynamics of solids and even in angular momentum transport. It certainly should be the focus of future numerical simulations. Note that this state will evolve slowly, on the long time-scale of the disc’s and protostar’s evolutionary track, and also presumably on the shorter time-scale of the protostar’s emission variability.

The VSI in protoplanetary discs is challenging to simulate adequately because the fastest growing modes have very short length-scales, many orders of magnitude shorter than both the disc thickness and (typically) the numerical grid scale. Global simulations are then not only ill-posed but exhibit a non-linear saturation whose characteristic length-scales are forced to be longer than is realistic. It may be that global simulations of the VSI greatly overestimate the amount of power in the largest length-scales. Indeed Stoll & Kley (2014) find no convergence in radial angular momentum transport (α) as resolution is increased, a result that emphasises the prominence of the smallest scales.

We can crudely estimate an upper limit for the resulting radial angular momentum transport if we assume that $\ell_R \sim |q|\epsilon\ell_z$ and

³ This is not the case for the locally isothermal thin-disc model discussed in Section 3.1.1, which has no surface and for which the magnitude of the vertical shear takes its maximum value at vertical infinity, where the density is negligible.

⁴ In this case, the fastest growing VSI-driven modes probably saturate through secondary shear instabilities – as is the case for fingering convection, another kind of double-diffusive instability (Brown, Garaud & Stellmach 2013). The resulting turbulence in the low-density regions near the disc surface may best be studied using local simulations, since they can access more realistic small scales.

if non-linear saturation occurs when the velocity amplitude is of the order $\ell_R\sigma$. Consequently, the turbulent viscosity scales as $\nu_T \sim \ell_R^2\sigma \sim |q|^3\epsilon^3\ell_z^2\Omega$. To make further progress we must estimate the size of ℓ_z . If the dominant scales at saturation are the largest ones, then $\ell_z \lesssim H$, which gives an upper bound on the α parameter:

$$\alpha \lesssim |q|^3\epsilon^3. \quad (7)$$

For a protoplanetary disc with $\epsilon \approx 0.05$, we get $\alpha \lesssim 10^{-4}$, which is broadly consistent with the results of current global simulations (Nelson et al. 2013; Stoll & Kley 2014).⁵ If, however, we assume instead that the dominant scale at saturation is much shorter (so that $\ell_z \ll H$), then this is a gross overestimate.

3 HYDRODYNAMIC EQUATIONS AND BASIC STATE PROFILES

In the previous section, we explained why discs with global radial variations in temperature or entropy⁶ are likely to be unstable to the VSI. We now list the equations and describe the simplified disc models that will be used to analyse the VSI. We begin with the equations of compressible hydrodynamics for an inviscid adiabatic fluid:

$$(\partial_t + \mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla P - \nabla \Phi, \quad (8)$$

$$(\partial_t + \mathbf{u} \cdot \nabla) \rho = -\rho \nabla \cdot \mathbf{u}, \quad (9)$$

$$(\partial_t + \mathbf{u} \cdot \nabla) S = 0, \quad (10)$$

where \mathbf{u} is the velocity. The ideal equation of state is $P = \mathcal{R}\rho T$ (where \mathcal{R} is the gas constant divided by the mean molecular weight).

If we adopt cylindrical polar coordinates (R, ϕ, z), the gravitational potential due to the central object is approximately that of a point mass

$$\Phi(R, z) = -\frac{GM}{\sqrt{R^2 + z^2}}, \quad (11)$$

$$\approx -\frac{GM}{R} \left[1 - \frac{z^2}{2R^2} \right] = \Phi_0 + \Phi_2 z^2, \quad (12)$$

where in some cases we expand for a thin disc ($|z| \ll R$; second line), and we define $\Phi_0 = -GM/R$ and $\Phi_2 = GM/(2R^3)$.

3.1 Basic state

The axisymmetric basic state of the differentially rotating disc has $\mathbf{u} = R\Omega(R, z)\mathbf{e}_\phi$, and satisfies the equations of radial and vertical force balance:

$$-R\Omega^2\mathbf{e}_R = -\frac{1}{\rho} \nabla P + \nabla \Phi. \quad (13)$$

Taking the curl then provides the thermal wind equation (equation 1). For simplicity, and to allow some analytical reduction and exploration, we restrict ourselves to studying two simple models (as in Nelson et al. 2013): the locally isothermal disc with a radial power law in temperature, and the locally polytropic disc with a

⁵ Note that many of the α calculations in Stoll & Kley (2014) are two dimensional and questions may be asked of angular momentum transport in this case (see arguments in Balbus 2000).

⁶ We do not set out to analyse the stability of local structures in the thermal properties of the disc, such as edges or pressure bumps, etc.

power-law entropy function. In both cases, we want to consider a disc with mid-plane density

$$\rho_m(R) = \rho_0 \left(\frac{R}{R_0} \right)^p, \quad (14)$$

where ρ_0 is the mid-plane density at a radius R_0 . It turns out that the value of p is unimportant for the VSI.

3.1.1 Locally isothermal disc with a radial power law in temperature

The first model that we will discuss is a locally isothermal disc with $P = c_s^2(R)\rho$ (where c_s is the isothermal sound speed), in which the temperature depends only on cylindrical radius as the power law

$$T(R) = T_0 \left(\frac{R}{R_0} \right)^q, \quad (15)$$

and similarly $c_s^2(R) = c_0^2 (R/R_0)^q$, where $c_0^2 = \mathcal{R}T_0$ is the square of the isothermal sound speed at a radius R_0 . The corresponding density profile that satisfies equation (13) is

$$\rho(R, z) = \rho_m(R) \exp \left(\frac{1}{c_s^2(R)} [\Phi_0(R) - \Phi(R, z)] \right). \quad (16)$$

Note that the disc has no surface and formally extends to infinity. The angular velocity that satisfies equation (13) is

$$\Omega(R, z) = \Omega_0(R) \left(1 + q + (p + q) \frac{H^2}{R^2} - \frac{qR}{\sqrt{R^2 + z^2}} \right)^{\frac{1}{2}}, \quad (17)$$

where $\Omega_0(R) = \sqrt{GM/R^3}$ and $H(R) = c_s(R)/\Omega_0(R) \propto (R/R_0)^{(q+3)/2}$ is the local disc scaleheight. The angular velocity therefore depends on z whenever $q \neq 0$. We set out to consider p and q so that these discs are stable according to the Solberg–Høiland criteria governing adiabatic axisymmetric perturbations (e.g. Tassoul 1978).

For illustration, in Fig. 1 we plot the angular velocity Ω , vertical shear $\partial_z(R\Omega)$ and density ρ on the (R, z) plane for a disc with $c_0 = 0.05$, $p = -1.5$ and $q = -1$. This shows that the angular velocity depends on z , and that the vertical shear increases monotonically with z until it reaches a maximum far away from the mid-plane, where the density is negligible.

3.1.2 Locally polytropic disc with a radial power law in entropy

The second model is a locally polytropic model in which $P = K_s(R)\rho^\gamma$ (where γ is the adiabatic index) with an entropy function

$$K_s(R) = P\rho^{-\gamma} = K_0 \left(\frac{R}{R_0} \right)^{q_s}, \quad (18)$$

with q_s and K_0 constants. Hence K_s is only a function of cylindrical radius, so that $S = S(R) \propto \ln K_s(R)$. This disc model is neutrally stratified in the vertical direction. The corresponding density profile which satisfies equation (13) is

$$\rho(R, z) = \rho_m(R) \left(1 + \frac{(1+m)}{K_s(R)} [\Phi_0(R) - \Phi(R, z)] \right)^m \quad (19)$$

$$\approx \rho_m(R) \left(1 - \frac{z^2}{H_0^2(R)} \right)^m, \quad (20)$$

where $m = 1/(\gamma - 1)$ is the polytropic index. For the last line, the potential has been expanded for a thin disc, and we have defined the

local disc thickness

$$H_0(R) = \sqrt{\frac{2(1+m)K_s(R)}{\Omega_0^2}}, \quad (21)$$

where Ω_0 was defined in Section 3.1.1. This disc model possesses a surface at which $\rho = 0$ when $z = H_0(R)$. The adiabatic sound speed is $a(R, z) = \sqrt{\gamma P(R, z)/\rho(R, z)}$, which becomes $a_m(R)$ at the mid-plane (taking the value a_0 at $R = R_0$), and we define $M(R, z) = R\Omega_0(R)/a(R, z)$ and $M_m(R) = R\Omega_0(R)/a_m(R)$ to be the Mach number and mid-plane Mach number of the flow. The angular velocity that satisfies equation (13) is

$$\Omega(R, z) = \Omega_0(R) \left(1 + q_s + \frac{p}{M_m^2} + \frac{q_s}{\gamma M^2} - \frac{q_s R}{\sqrt{R^2 + z^2}} \right)^{\frac{1}{2}}. \quad (22)$$

As earlier, the angular velocity depends on z whenever $q_s \neq 0$. Since this model is neutrally stratified in the vertical direction, it can become unstable to adiabatic disturbances for certain choices of p and q_s , whenever one of the Solberg–Høiland criteria are violated. However, we do not set out to study such instabilities in this work, and we instead focus on discs that would be stable to adiabatic perturbations.

For illustration, in Fig. 2 we plot the vertical shear $\partial_z(R\Omega)$ and density ρ on the (R, z) plane for a disc with $a_0 = 0.05$, $\gamma = 1.4$, $p = -1.5$ and $q_s = -1$. This shows that the angular velocity depends on z , and that the magnitude of the vertical shear increases monotonically with $|z|$ until it reaches a maximum just below the disc surface.

In the next three sections we work through the stability of these two equilibria. Because the full global analysis is challenging, we first treat their quasi-local approximations in a reduced model first outlined in Nelson et al. (2013). This helps tease out the most important features. Once this is done, we perform the full global analysis on the locally isothermal model only.

4 LINEAR STABILITY OF THE LOCALLY ISOTHERMAL DISC: REDUCED MODEL

In this section we revisit the reduced model of Nelson et al. (2013), which describes the dynamics of the VSI in locally isothermal discs with a radial power law in temperature. We begin by outlining the derivation of this model (a similar derivation is presented in detail in Appendix A for the locally polytropic disc) and go on to analyse its most important properties.

We define $\epsilon = H/R$ and consider a thin disc in which $\epsilon \ll 1$. We are interested in slow modes with frequencies and growth rates that are each $O(\epsilon\Omega)$, with vertical scales ℓ_z that are comparable with the thickness of the disc ($\ell_z \sim \epsilon R$) and radial scales ℓ_R that are much smaller ($\ell_R \sim \epsilon^2 R$), cf. equation (6). If we also consider vertical velocities that are mildly subsonic or transonic (by a factor ϵ) and radial velocities that are very subsonic (by a factor ϵ^2), then we require $O(\epsilon)$ density perturbations, for consistency. On such small radial scales, the curvature of the disc can be neglected and the geometry is locally Cartesian, similar to the classical shearing box (Goldreich & Lynden-Bell 1965; Umurhan & Regev 2004). In this limit, we are looking at low-frequency (inertial) dynamics that are anelastic and in approximate radial geostrophic balance.

The linearized reduced equations for the rescaled velocity perturbation (u, v, w) and the fractional density perturbation

$\Pi = \rho'/\bar{\rho}$ are (Nelson et al. 2013)

$$0 = 2v - \partial_x \Pi, \quad (23)$$

$$\partial_\tau v = -\frac{u}{2} - \frac{qzw}{2}, \quad (24)$$

$$\partial_\tau w = -\partial_z \Pi, \quad (25)$$

$$0 = \partial_x(\bar{\rho}u) + \partial_z(\bar{\rho}w). \quad (26)$$

Here τ is rescaled time, and x is a local radial variable. The background density is $\bar{\rho} = e^{-\frac{z^2}{2}}$, after appropriate normalization. The crucial term for the appearance of the VSI is the last one in equation (24), which arises from thermal wind balance if there is radial variation in temperature (see equation 1).

We seek solutions of this system of the form

$$\Pi = \text{Re} \left[\hat{\Pi}(z) e^{i(kx - \omega\tau)} \right], \quad (27)$$

and so on for other variables, where we subsequently drop the hats for clarity. This allows us to reduce the system to a single ODE that can be written most simply by defining new (complex) coordinates $\zeta = z\sqrt{1 + ikq}$, as

$$\frac{d^2 \Pi}{d\zeta^2} - \zeta \frac{d\Pi}{d\zeta} + \lambda \Pi = 0, \quad (28)$$

where we have set

$$\lambda = \frac{\omega^2 k^2}{1 + ikq}. \quad (29)$$

This is the well-known Hermite differential equation (the probabilist's version), which has solutions

$$\Pi = a_1 \text{He}_\lambda(\zeta) + a_2 {}_1F_1 \left(-\frac{\lambda}{2}, \frac{1}{2}, \frac{\zeta^2}{2} \right). \quad (30)$$

The first function is the Hermite function and the second function is a confluent hypergeometric function of the first kind, with a_1, a_2 arbitrary constants. Solutions that are polynomially bounded as $|\zeta| \rightarrow \infty$ (so that $|\rho'| \rightarrow 0$ as $|\zeta| \rightarrow \infty$) require $\lambda = n \in \mathbb{N}$ and $a_2 = 0$ (see Okazaki, Kato & Fukue 1987; Kato 2001). The regular solutions are therefore

$$\Pi \propto \text{He}_n(\zeta), \quad (31)$$

with He_n a Hermite polynomial of order n . The corresponding ω can be obtained from the dispersion relation equation (29). These solutions describe the vertical structure of modes in the low-frequency limit. Since they are (complex) polynomials in ζ , they describe global modes which are not localized in the vicinity of any particular location $\zeta \neq 0$.

4.1 Non-vertically shearing case, $q = 0$

Before treating the VSI we examine the case when $q = 0$ in order to make contact with the existing literature on wave modes in vertically stratified isothermal discs (namely Lubow & Pringle 1993). This then can assure us of the validity of the reduced model and the regime of its applicability.

Since λ is quantized, we obtain a discrete set of frequencies. When $q = 0$, these are real and

$$\omega = \pm \frac{\sqrt{n}}{k}, \quad (32)$$

which represents a pair of low-frequency inertial waves⁷ travelling in opposite directions. The full isothermal disc allows axisymmetric waves with frequencies that satisfy the following dispersion relation (Lubow & Pringle 1993):

$$(-\omega^2 + nc_s^2)(-\omega^2 + 1) - c_s^2 k^2 \omega^2 = 0. \quad (33)$$

There are two branches of solutions, which represent either high-frequency acoustic waves or low-frequency inertial waves. Note that there are no surface modes in the isothermal disc, since the model lacks a surface. The inertial waves have $\omega^2 \approx n/(k^2 + n)$; therefore equation (32) is a good approximation for modes with $k \gg \sqrt{n}$, as expected from the assumption $\ell_R/\ell_z \sim \epsilon$. When this is violated, physically incorrect solutions appear with frequencies larger than 1. The restriction on the size of the vertical quantum number n means that the model can never capture purely local modes (for which both n and k are large).

4.2 Vertically shearing case, $q \neq 0$

When $q \neq 0$ the disc is unstable to the VSI. The frequencies are now

$$\omega = \pm \frac{\sqrt{n}}{k} (1 + ikq)^{\frac{1}{2}}, \quad (34)$$

which are complex. The growth rate $\sigma = \text{Im}[\omega]$ may be computed from the positive root of

$$\sigma^2 = \frac{n}{2k^2} \left(\sqrt{1 + k^2 q^2} - 1 \right). \quad (35)$$

The unstable modes in this case are the classic inertial waves of the vertically stratified disc, destabilized by the vertical shear. They are all global ‘body modes’ that have no localization near to any $\zeta \neq 0$.

In the limit $|kq| \ll 1$,

$$\omega = \pm \frac{\sqrt{n}}{k} \left(1 + i \frac{kq}{2} + O(k^2 q^2) \right), \quad (36)$$

which possesses the same frequency as the classical modes, cementing the identification. They grow, however, at the rate $\sigma \approx \sqrt{n}|q|/2$. In the opposite limit $|kq| \gg 1$,

$$\omega = \pm \sqrt{\frac{nq}{2k}} (1 + i) + O\left(\sqrt{\frac{n}{k^3 q}}\right), \quad (37)$$

therefore the growth rate is $\sigma = \sqrt{n|q|/(2k)}$, which is no longer linear in the shear.

Note that the growth rate increases with polynomial order as \sqrt{n} , which may seem somewhat surprising but can be understood by considering the vertical shear profile. An isothermal thin disc lacks a surface, yet $|\partial_z(R\Omega)| \propto |qz/2|$, which increases without bound as $|z| \rightarrow \infty$. Since modes with large n extend over a greater vertical range, they can better tap into the energy associated with the larger shear at larger z . Hence it follows that the growth rate should increase with n (the square root comes from the fact that n has units of inverse length squared). It should be remembered, however, that these expressions apply only when the reduced model is valid, which requires $\sqrt{n} \ll k$.

Finally, note that the unstable modes here have a wave character, whereas those issuing from the local Boussinesq calculation grow monotonically (Urpin & Brandenburg 1998; Urpin 2003). Their

⁷ Note that these are sometimes referred to as r modes in the literature (and sometimes inappropriately as g modes even when buoyancy forces are absent).

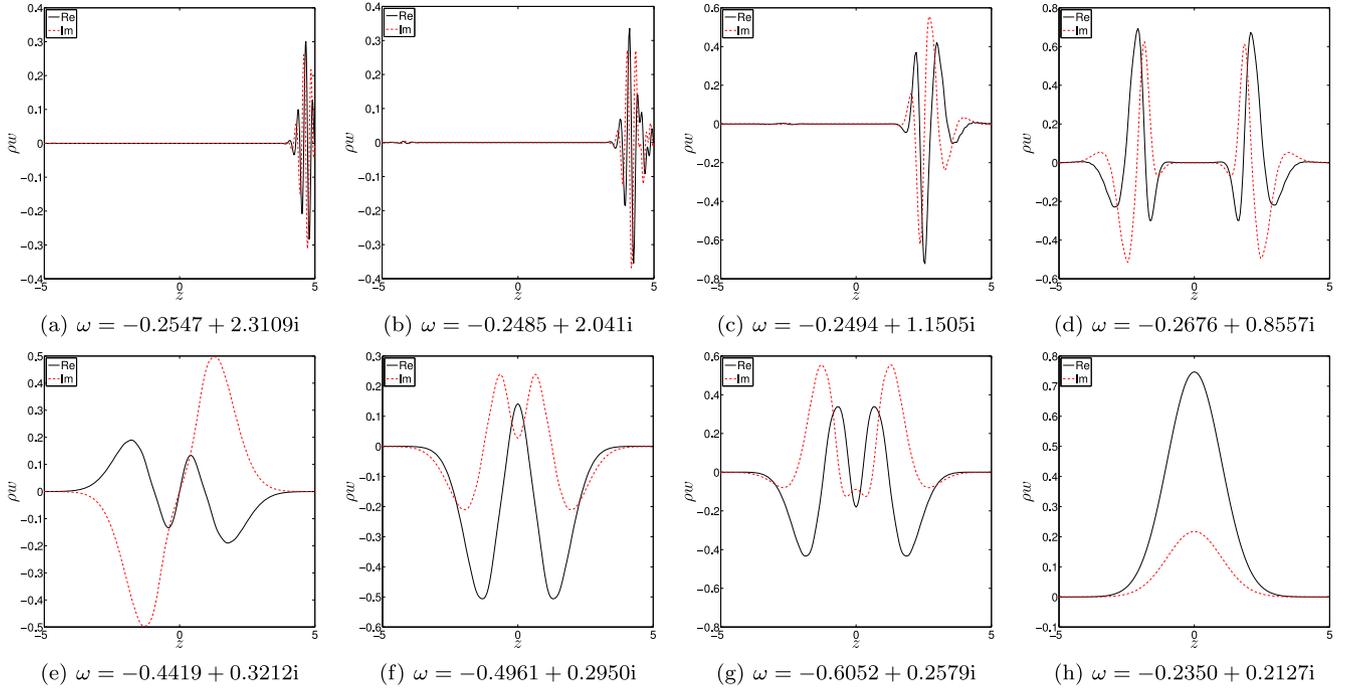


Figure 3. Illustration of the real and imaginary parts of the vertical momenta for a representative selection of the two types of modes in an isothermal disc with imposed numerical boundaries. Here $k = 10$, $q = -1$, $H = 5$ and $N = 300$. The figure labels indicate the complex frequencies of the modes. Surface modes are shown in the top row (the fastest growing mode is in the top-left panel), and body modes are shown in the bottom row. Similar surface modes localized at the bottom boundary are also obtained. The lowest frequency body mode is plotted in the bottom-right panel and is an $n = 1$ ($n = 0$ for w) inertial wave, i.e. the fundamental ‘corrugation mode’, which is well described by an $n = 1$ ($n = 0$ for w) Hermite polynomial in the absence of vertical boundaries, and its complex frequency is that predicted by equation (34).

scalings with wavenumber also differ. In the isothermal limit with $k/k_z \gg 1$ and $k \gg 1$, local Boussinesq modes grow at a rate $\sigma \sim \sqrt{k_z q/k}$, where k_z is the vertical wavenumber. The system, being scale free, only depends on the ratio of wavenumbers, k_z/k . In contrast, the quasi-global model separates out the vertical and radial scales, forcing the former to be near H . Consequently, it will always struggle to reproduce the local limit. Only if we force modes to localize at a fixed z (such as at a boundary) and then undertake a plane-wave analysis, can the reduced model produce the local Boussinesq results (Nelson et al. 2013).

4.3 Vertically shearing case with imposed boundaries: appearance of surface modes

To connect with recent numerical simulations of a locally isothermal disc (Nelson et al. 2013; Stoll & Kley 2014), we adopt an artificial boundary at a finite height in the vertical direction. In realistic discs there is surely some form of transition at the photosphere, but whether a numerical boundary adequately mimics this feature is unclear. We emphasize that within the confines of a strict isothermal model, a numerical boundary is an artificial addition, yet it has important effects.

We consider boundary conditions such that the vertical momentum $\rho_0 w = 0$ at $|z| = H$, where H is free parameter (other choices would lead to similar behaviour for the VSI). To illustrate the unstable modes in this case, we solve equations (23)–(26) numerically using a Chebyshev collocation method on $N + 1$ points of a Gauss–Lobatto grid. This results in a matrix eigenvalue problem that can be solved using a QZ method (Golub & van Loan 1996; Boyd 2001). Numerical convergence for the eigenvalues is verified by varying the resolution and comparing eigenvalues.

In the presence of an artificial boundary, a new class of ‘surface modes’ appears that localize near to the boundaries. These are in addition to the ‘body modes’ obtained previously. We show the vertical momenta for several examples of each type of mode in Fig. 3 for $k = 10$ and $q = -1$: the top panel shows a selection of typical surface modes, and the bottom panel shows typical body modes. The mode plotted in the top-right panel is one of the slowest growing surface modes, and its character is intermediate between the two classes of mode. The appearance of these two classes of modes was also found by Nelson et al. (2013) and McNally & Pessah (2014).

Mathematically, the emergence of surface modes can be understood by examining equation (30). In the presence of a boundary at a finite height, λ is complex and a_1 and a_2 are both non-zero, in general. New modes, localized near to $|z| = H$, appear due to the need to match boundary conditions at this location, rather than at infinity (this can be verified by plotting $\text{He}_\lambda(\zeta)$ for $\lambda \in \mathbb{C}$). Thus the new surface modes rely entirely on the boundary for their existence.

In Fig. 4 we show the dependence of the spectrum on disc height H . Here the numerical growth rates of the unstable modes are plotted in the complex frequency plane for three different heights, $H = 5, 6, 7$. The remaining parameters are $k = 10$ and $q = -1$. We also plot the $H = \infty$ case (equation 34), which is given by the solid red line in the figure. Roughly, surface modes correspond to the more ‘vertical’ segment of the spectrum, and the inertial waves to the more ‘horizontal’ segment.

Obviously, there is no convergence with increasing H , especially for the surface modes. As the height of the domain is increased, the maximum growth rate increases in direct proportion with H . This is what we would expect, since $\sigma_{\max} \propto \max|\partial_z(R\Omega)| = |q|H/2$ (even larger growth rates than shown in Fig. 4 are obtained for modes with

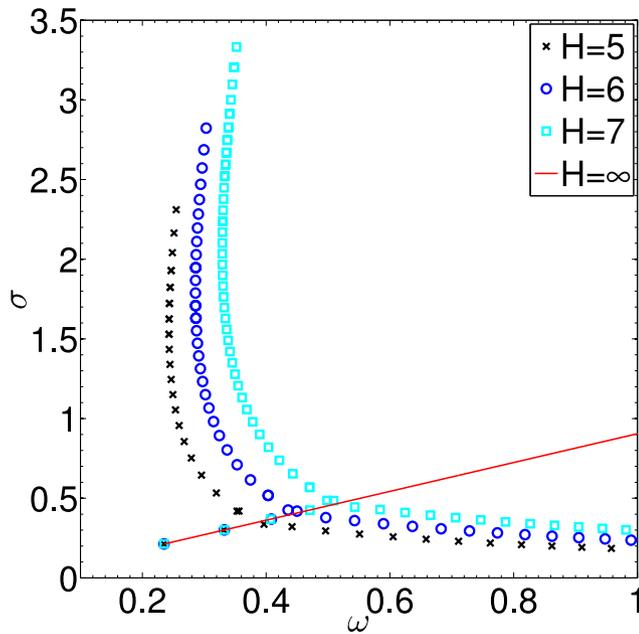


Figure 4. Growth rate of the unstable modes versus their (real) frequencies in an isothermal disc with artificial boundaries at $|z| = H$, illustrating the lack of convergence as H is varied for cases with $k = 10$ and $q = -1$ (converged results were obtained with $N = 300$). Note that the maximum growth rate in this case is approximately $|qH|/2$. We have also indicated the analytical prediction of equation (34) with the solid red line to guide the eye (however, it should be remembered that this represents a set of discrete modes and not a continuum). The branch that extends approximately vertically represents surface modes.

$k \gg 10$, which asymptotically attain this value). The presence of a boundary also strongly influences the inertial waves, only the lowest n of which are well described by equation (34). As H is increased, however, modes with increasingly larger n converge to the $H = \infty$ analytical prediction.

The appearance and lack of convergence of the surface modes is a special pathology of the isothermal model with imposed boundaries. It makes the interpretation of these surface modes, and their role in any ensuing turbulence, especially problematic. Are they merely numerical artefacts? In the next section we argue that they are, in fact, more than that and that it is the vertically isothermal model itself that is the problem.

5 LINEAR STABILITY OF THE LOCALLY POLYTROPIC DISC: REDUCED MODEL

In this section we analyse a reduced model of the VSI in discs with a radial power law in entropy. The vertical structure of this locally polytropic model is a good approximation for an optically thick disc, and is more realistic than the locally isothermal model because it possesses upper and lower surfaces. In addition, there is a well-defined maximum vertical shear rate, which occurs at the disc surface. Therefore this model is better defined mathematically and physically.

A formal derivation of the reduced model is relegated to Appendix A. The same scalings as in Section 4 are adopted, corresponding to anelastic radially geostrophic phenomena. In addition, thermal diffusion is likely to dominate the thermodynamics of such slow and short-scale modes. Consequently, we may neglect the en-

tropy perturbation entirely, as well as the stabilizing influence of stratification.

The resulting linearized equations for such perturbations are

$$0 = 2v - \partial_x h, \quad (38)$$

$$\partial_\tau v = -\frac{u}{2} - w \frac{q_s z}{2\gamma}, \quad (39)$$

$$\partial_\tau w = -\partial_z h, \quad (40)$$

$$0 = \partial_x u + \partial_z w - \frac{2mzw}{1-z^2}, \quad (41)$$

where the pseudo-enthalpy perturbation is $h = P'/(1-z^2)^m$ and P' is the (scaled) pressure perturbation. Rapid thermal diffusion means that the perturbations evolve isothermally, hence the equations are similar in form to equations (23)–(26).

We seek solutions of the form

$$u = \text{Re} [\tilde{u}(z)e^{i(kx - \omega\tau)}], \quad (42)$$

and so on for other variables. The resulting system of equations is equivalent to the single ODE

$$\frac{d^2 h}{dz^2} - z \left[\frac{2m}{1-z^2} + \frac{ikq_s}{\gamma} \right] \frac{dh}{dz} + \omega^2 k^2 h = 0. \quad (43)$$

Unfortunately, this cannot be solved in closed form analytically. However, we can attack the problem using a matched WKB approach, or numerically with the same method as Section 4.3: Chebyshev collocation on $N + 1$ points, followed by the eigen-solution of the resulting matrix equation using the QZ method. No explicit boundary conditions are imposed, but implicit regularity is assumed at $|z| = 1$ (this gives identical results to imposing a free-surface condition explicitly).

Finally, we note that the resulting system can be written in a scale-free manner by transforming our variables to hatted quantities, as follows:

$$v = k\hat{v}, \quad w = k\hat{w}, \quad \omega = \hat{\omega}/k, \quad q_s = 2\hat{q}\gamma/k. \quad (44)$$

A similar rescaling also applies to the isothermal model. Under this rescaling the governing equation depends solely on the two parameters \hat{q} and m . Each, discrete, growth rate then has the following behaviour:

$$\omega_n = \frac{1}{k} \hat{\omega}_n(kq_s, \gamma), \quad (45)$$

where the function $\hat{\omega}_n$ is determined numerically.

5.1 Non-vertically shearing case, $q_s = 0$

We first demonstrate that the reduced model correctly captures the inertial waves in a polytropic disc in the limit of large k , in order to gain confidence in the reduced model.

The full polytropic disc model without vertical shear ($q_s = 0$) possesses three different classes of neutrally stable axisymmetric modes: high-frequency acoustic modes, low-frequency inertial modes, and surface gravity modes of intermediate frequency (Korycansky & Pringle 1995; Ogilvie 1998). The reduced model is designed to capture only the low-frequency inertial modes.

In Fig. 5 we compare the frequencies of five inertial modes with vertical mode-numbers (vertical nodes) $n = 1, 2, 3, 4, 5$ and $q_s = 0$ for several k as predicted by the reduced model and the full polytropic model of Korycansky & Pringle (1995, solving their equations 4–8). The latter are represented by solid and the former by

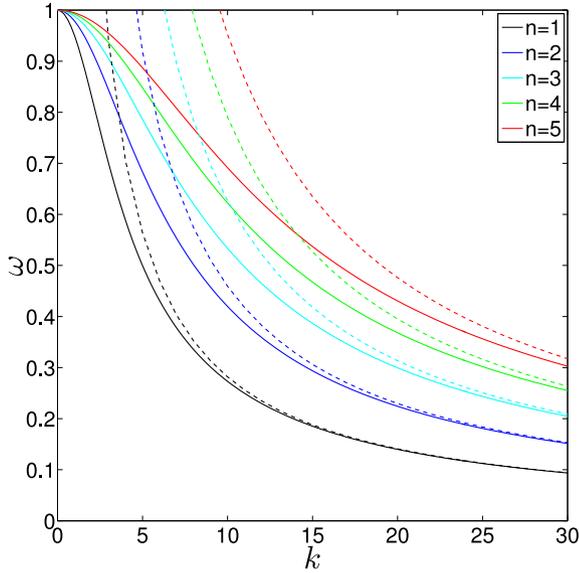


Figure 5. Comparison of the dispersion relation computed numerically from the locally polytropic reduced model (equations 38–41) with $q_s = 0$ (dashed lines) together with the numerically computed predictions for inertial modes in a polytropic disc (Korycansky & Pringle 1995, solid lines) with $n = 1, 2, 3, 4, 5$ vertical nodes adopting $\gamma = 1.4$. The reduced model is only valid when $k \gtrsim O(\sqrt{n})$.

dashed lines. Unlike the isothermal disc, these frequencies must generally be obtained numerically. There is general agreement for $k \gtrsim 20$, which illustrates that the reduced model correctly captures the frequencies of these inertial modes for sufficiently large k . However, only modes with $k \geq O(\sqrt{n})$ are correctly captured.

In the limit of large k , a WKB analysis shows that the frequencies take a simple form,

$$\omega = \pi(2n + m)/(4k), \quad (46)$$

where n is an integer. Though the formula is most accurate in the limit of large n , it does well across the whole range of frequencies. Details of this calculation can be found in Appendix B.

5.2 Vertically shearing case, $q_s \neq 0$

When $q_s \neq 0$, the disc is unstable to the VSI. As an illustrative example, we plot in Fig. 6 the complex frequencies of the unstable modes computed from the eigenvalue problem equations (38)–(41) when $k = 100$, $q_s = -1$ and $\gamma = 1.4$ for two different vertical resolutions $N = 200$ and 400. The unstable modes (for this k) are shown to be well resolved using $N = 200$.

Fig. 6 also shows clearly that a locally polytropic disc can be unstable to two different types of modes: (a) modestly growing inertial waves (‘body modes’; these occur on longer radial scales, as we will show in Fig. 9), and (b) rapidly growing short-wavelength surface modes (these only occur when $|kq_s| \gtrsim 30$, as we will also show in Fig. 9). The growth rates of the two classes of modes are labelled in Fig. 6. Examples for each type of mode are plotted in Fig. 7. Hence we find that the VSI in the locally polytropic disc, with well-defined physical surfaces, is qualitatively similar to the VSI in the isothermal disc with an imposed artificial surface at a finite height, as discussed in Section 4.3.

In Fig. 8, we plot the complex frequencies of the unstable modes when $q_s = -1$ and $\gamma = 1.4$ for several values of k . To accurately capture the fastest growing modes for larger k requires increasing

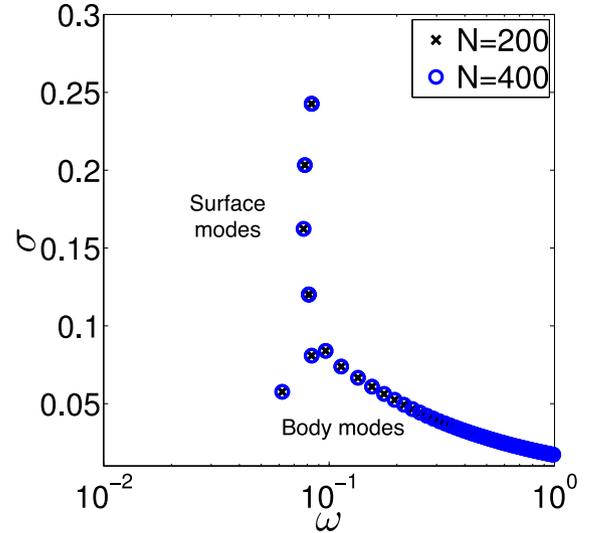


Figure 6. Growth rates of the unstable modes versus their (real) frequencies in a polytropic disc with $k = 100$, $q_s = -1$ and $\gamma = 1.4$ computed using $N = 200$ and 400 points. This shows that the modes are well captured using $N = 200$, and the distribution is similar to the isothermal case in Fig. 4.

the vertical resolution ($N = 400$ was required to obtain convergence when $k = 300$). Fig. 8 shows that the growth rate of the fastest growing mode increases, and its frequency decreases, as we consider smaller horizontal length-scales (larger k). For a given $|q_s|$, the number of rapidly growing surface modes increases with k . All unstable modes are body modes when $k = 10$, but as we increase k surface modes appear, and by $k = 300$ there is a large population of them. The maximum growth rate is on track to approach the maximum vertical shear rate $|\partial_z(R\Omega)| \sim |q_s|H/(2\gamma) \approx 0.357$ as $k \rightarrow \infty$, which we expect to provide an upper bound on the growth rate of the VSI.

In Fig. 9, we have exploited the rescaling of equation (44) to illustrate the general dependence of the scaled growth rate $\sigma/|q_s|$ of the VSI as a function of $k|q_s|$ for all unstable modes. We do not consider modes with real frequencies that are larger than unity – their omission results in a region where the VSI is absent in the bottom left of the figure. Pairs of surface modes appear for sufficiently large $k|q_s|$. The first, and fastest growing, pair of surface modes appear when $k|q_s| \gtrsim 30$, and corresponds to the $n = 2$ ‘breathing mode’ and the $n = 3$ inertial mode changing their character. The $n = 1$ fundamental ‘corrugation mode’ is the first to become unstable when $k|q_s| \gtrsim 3$, but its growth becomes weaker for larger $k|q_s|$. This is represented by the curve that crosses all others and ends up at the bottom right of the figure. As $k|q_s| \rightarrow \infty$, there are more unstable modes, and their scaled growth rates approach the scaled maximum vertical shear rate $1/(2\gamma) \approx 0.357$, as expected.

The fastest growing body mode is the $n = 1$ corrugation mode, and has a maximum growth rate of approximately $0.1|q_s|(1.4/\gamma)$ when $|kq_s| \lesssim 10$. The fastest growing mode is a surface mode, with a maximum growth rate that approaches the maximum vertical shear rate $|\partial_z(R\Omega)| \approx |q_s|/(2\gamma)$ on the smallest scales. (We have confirmed the dependence of the fastest growing surface mode on γ , though we omit this for brevity – this arises because the basic disc structure varies with γ .) Note that the full polytropic disc possesses a class of surface modes that are restored by gravity (Korycansky & Pringle 1995; Ogilvie 1998). However, these are unrelated to the

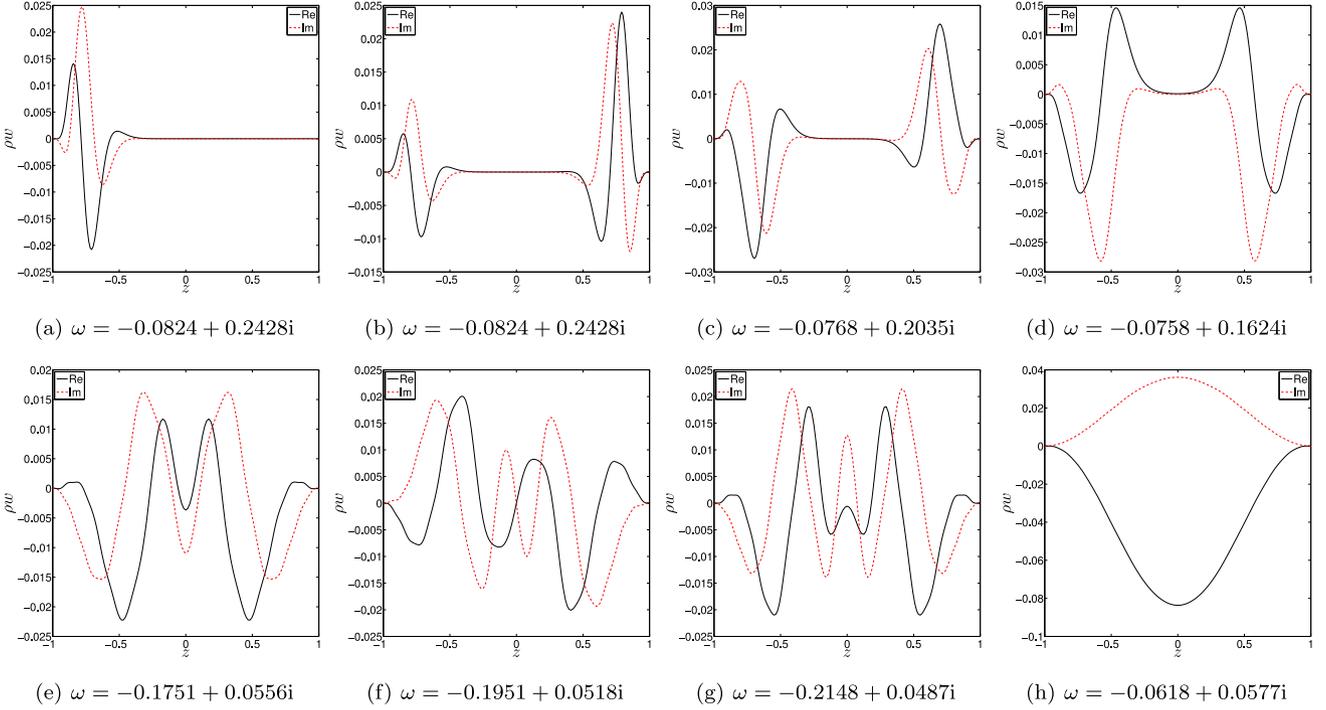


Figure 7. The real and imaginary parts of the vertical momenta for several representative examples of both types of modes in a polytropic disc with $k = 100$, $q_s = -1$ and $\gamma = 1.4$ using $N = 200$, where the figure labels indicate the complex frequencies of the modes. In the top row we show several examples of surface modes, with the fastest growing modes plotted in the two top-left panels. In the bottom row we show several examples of body modes. The lowest frequency mode is plotted in the bottom-right panel, and is an $n = 1$ ($n = 0$ for w) inertial wave, i.e. the fundamental ‘corrugation mode’. The other body modes are higher order body modes.

class of low-frequency surface modes that we have discussed in this section, and occur in a different frequency range.

Why do surface modes appear when $|kq_s| \gtrsim 30$? The vertical shear rate takes its maximum magnitude at the disc surface, so it makes sense that if localized modes were to appear, they should do so just below the disc surface. But when $|kq_s| \lesssim 30$, it is not possible for a mode (with an inertial wave character) to become sufficiently localized close to the disc surface (i.e. ℓ_z cannot be made sufficiently small). The criterion for the appearance of surface modes is therefore tied to the shortest vertical length-scale permitted.

Finally, in Appendix B we show that the body modes’ growth rate can be obtained analytically in the limit of small q_s :

$$\sigma \approx \frac{mq_s}{\pi\gamma(2n+m)} \log \left[\frac{1}{2} \pi(2n+m) \right], \quad (47)$$

where n is a (large) integer. The corresponding wave frequency is given by equation (46). It should be stressed that the growth rate σ being at subdominant order is only a rough estimate, because errors arising from the WKB method itself enter at the same order. Note that the growth rate is linear in the shear, but the larger n , the smaller σ , in contrast to the isothermal case, but in agreement with Figs 6 and 8. For general q_s an estimate for the surface modes’ maximum growth rate is $\sigma \approx q_s/(2\gamma)$, while its wave frequency scales as $\ln k/k$. Being leading order, this is a more robust estimate, which is in approximate agreement with our results.

In summary, the VSI in the locally polytropic model is very similar in character to the VSI in the locally isothermal model with artificially imposed vertical boundaries (Section 4.3). The presence of a surface, be it physically justified or a numerical convenience,

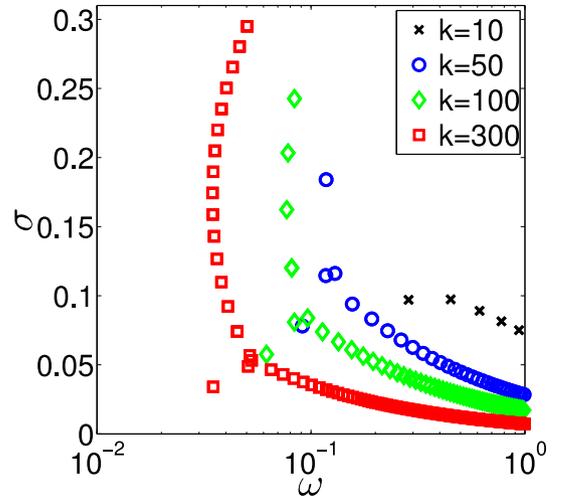


Figure 8. Growth rates of the unstable modes versus their (real) frequencies for several values of k in a polytropic disc with $q_s = -1$ and $\gamma = 1.4$, computed using up to $N = 400$ points (which was found to be sufficient for all k). This illustrates the dependence of the unstable modes on k .

provides a special location upon which modes can affix themselves and localize. The polytropic disc however yields solutions that are clearer to interpret because the location of the surface is specified (not an adjustable parameter). Consequently, there exists a well-defined maximum growth rate for the instability, set by the vertical shear rate at the disc surface.

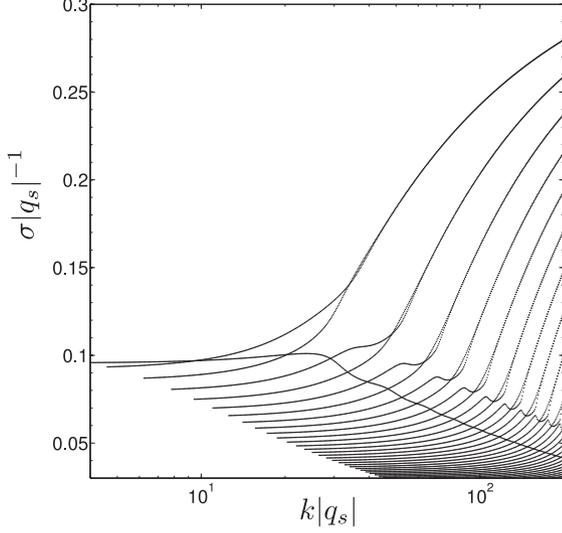


Figure 9. Normalized growth rates of all unstable modes as a function of the scaled radial wavenumber $k|q_s|$ in a polytropic disc with $\gamma = 1.4$, computed using $N = 200$. The absence of unstable modes in the lower left of the figure results from our elimination of modes with frequencies larger than 1. Surface modes appear in pairs for $k|q_s| \gtrsim 30$, with the fastest growing arising from the $n = 2$ ‘breathing mode’ and the $n = 3$ inertial mode. The $n = 1$ fundamental ‘corrugation mode’ lies along the line that ends up at the bottom right. The scaled growth rate of the fastest growing surface mode asymptotically approaches the maximum vertical shear rate (0.357) as $k|q_s| \rightarrow \infty$.

6 GLOBAL CALCULATIONS IN THE LOCALLY ISOTHERMAL DISC

We have so far analysed the VSI using reduced models of locally isothermal and polytropic discs. Here we present the first two-dimensional stability calculations of the VSI in a locally isothermal disc. One motivation for doing so is to verify the validity of the model analysed in Section 4, another is to reproduce the instability in a setup that more directly matches that of recent global simulations (Nelson et al. 2013; Stoll & Kley 2014). An issue with global stability calculations is that they are computationally demanding (primarily regarding their memory usage), so we are limited to studying the VSI with relatively low resolutions. This is primarily a problem for capturing the surface modes, since they occur on very short length-scales. However, the lowest order body modes have less vertical structure and are better captured in these calculations.

We consider axisymmetric perturbations $(u_R, u_\phi, u_z, \rho')$ to the global disc model of Section 3.1.1, assuming their time-dependence to be $\propto e^{-i\omega t}$. The equations governing their linear evolution are then

$$-i\omega u_R = 2R\Omega u_\phi - \frac{1}{\rho} \partial_R(c_s^2 \rho') + \frac{\rho'}{\rho^2} \partial_R(c_s^2 \rho) + f_R, \quad (48)$$

$$-i\omega u_\phi = -u_R \frac{1}{R} \partial_R(R^2 \Omega) - u_z \partial_z(R\Omega) + f_\phi, \quad (49)$$

$$-i\omega u_z = -\frac{1}{\rho} \partial_z(c_s^2 \rho') + \frac{\rho'}{\rho^2} \partial_z(c_s^2 \rho) + f_z, \quad (50)$$

$$-i\omega \rho' = -u_R \partial_R \rho - u_z \partial_z \rho - \frac{\rho}{R} \partial_R(R u_R) - \rho \partial_z u_z, \quad (51)$$

where $P' = c_s^2(R)\rho'$, and the background state quantities $c_s(R)$, $\Omega(R, z)$ and $\rho(R, z)$ are defined in Section 3.1.1. To regularize the solutions in some calculations we include a Navier–Stokes shear viscosity with constant kinematic viscosity ν , with the extra terms

$$\begin{aligned} \mathbf{f} = & \nu \rho \left[\nabla^2 \mathbf{u} + \frac{1}{3} \nabla(\nabla \cdot \mathbf{u}) \right] + \nu \rho' \nabla^2 \mathbf{U}_0 \\ & + \nu \nabla \rho \cdot \left[\nabla \mathbf{u} + (\nabla \mathbf{u})^\top - \frac{2}{3} (\nabla \cdot \mathbf{u}) \mathbf{I} \right] \\ & + \nu \nabla \rho' \cdot [\nabla \mathbf{U}_0 + (\nabla \mathbf{U}_0)^\top], \end{aligned} \quad (52)$$

where $\mathbf{U}_0 = R\Omega \mathbf{e}_\phi$. We adopt a 2D cylindrical domain with $R \in [R_0, R_1]$ and $z \in [-z_0, z_0]$, where z_0 is some multiple of $H_0 = \epsilon$ (this differs from the spherical wedge considered by Nelson et al. 2013 and Stoll & Kley 2014, but this difference is probably unimportant). Our units of length and time are chosen such that $R_0 = 1$ and $\Omega(R_0) = 1$. If $\mathbf{f} = 0$, boundary conditions are enforced such that

$$\rho u_R = 0 \quad \text{at } R = R_0 \quad \text{and } R = R_1, \quad (53)$$

$$\rho u_z = 0 \quad \text{at } z = -z_0 \quad \text{and } z = z_0, \quad (54)$$

otherwise we supplement these with stress-free conditions (no tangential stresses) on each boundary. This system is solved (after writing the equations in terms of momenta rather than velocities) numerically using a Chebyshev collocation method in both R and z , with $N_R + 1$ and $N_z + 1$ points on a Gauss–Lobatto grid, respectively. The resulting generalized eigenvalue problem involves matrices of size $L \times L$, where $L = 4(N_R + 1)(N_z + 1)$, and is solved using one of two methods: a QZ algorithm to obtain an approximation of the full spectrum, or an Arnoldi iterative method to obtain an approximation to a desired part of the spectrum. The QZ method is computationally expensive when $L \gtrsim 20\,000$, so our results are then supplemented by the Arnoldi method, once we determine which modes to focus on. In what follows we set $p = 0$, since preliminary investigation suggested this parameter to be unimportant.

The code has been tested in several ways. First, without viscosity and with a small radial domain (e.g. $R_1 = 1.001$), we have confirmed that our code accurately reproduces the inertial and acoustic modes of the vertically unbounded isothermal disc when $q = 0$ (Lubow & Pringle 1993) – except for modes with large n , for which the confining effect of the vertical boundaries modifies the solutions. Using a similarly small radial domain, we have verified that viscous damping produces the expected decay rate for a mode with a given short radial wavelength (νk^2 when $k \gg 1$). In all calculations, the discretization inevitably produces many junk modes, that involve oscillations on the grid scale. Where possible, we eliminate these modes by comparing eigenvalues for several different resolutions, and by inspecting the spatial structure of the eigenfunctions, discarding those that oscillate on the grid scale.

6.1 Inviscid calculations

We first illustrate the properties of the VSI in the absence of viscosity. As in the reduced model in Section 4, we obtain two classes of modes: modestly growing body modes, and rapidly growing surface modes localized near the numerical boundaries in z . As in Section 4.3, these modes would vanish if the boundaries could be taken to infinity. However, we do not discard these modes entirely as they still probably represent physical solutions in some sense (see Section 5).

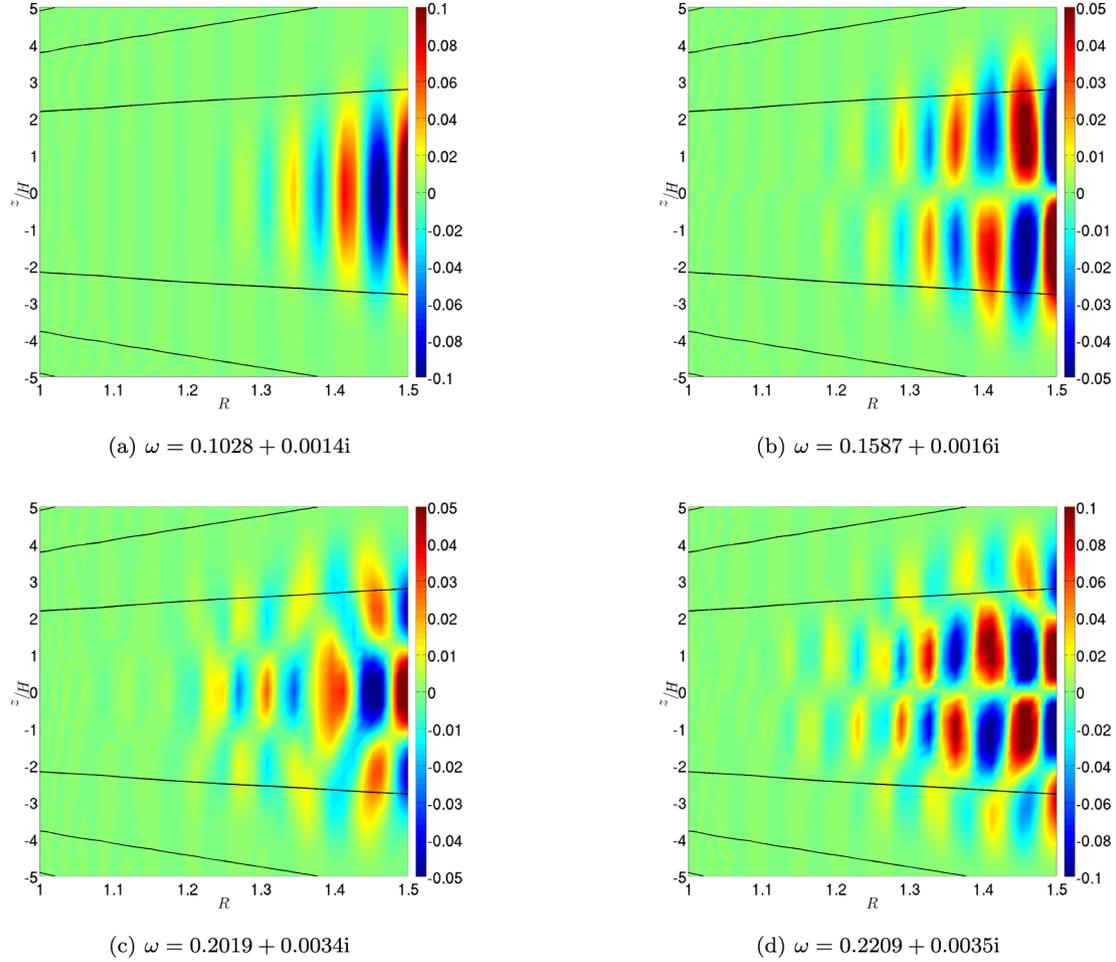


Figure 10. Illustration of the vertical momenta on the (R, z) plane for several body modes in the locally isothermal disc without viscosity with $\epsilon = 0.05$, $q = -1$, $z_0 = 5\epsilon$ and $R_1 = 1.5$, obtained using $N_R = N_z = 40$. These modes have vertical structures for u_z that are well described by Hermite polynomials with $n = 0, 1, 2$ and 3 . Many modes with similar vertical structures (but different radial structures) to each of these are also unstable. Also overplotted are three contours of constant density (solid black lines) with $\rho = 10^{-5}, 10^{-3}, 10^{-1}$ for reference. The amplitude in each panel is arbitrary.

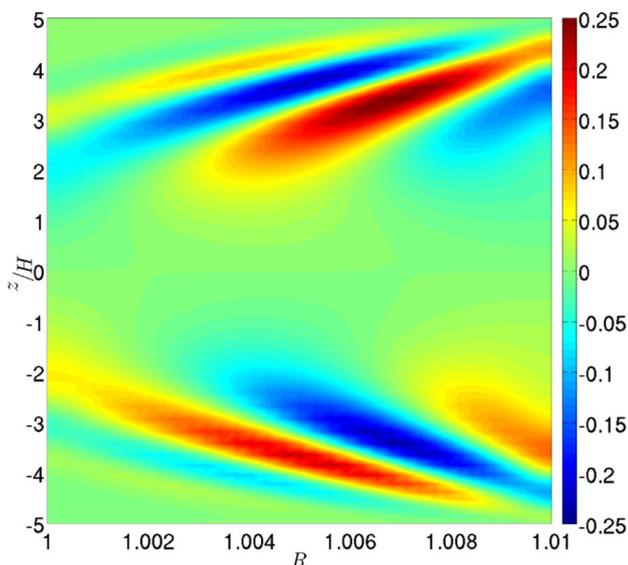
6.1.1 Body modes

As in Section 4.2, we obtain a set of body modes, essentially classical inertial waves that grow in the presence of vertical shear. We plot the vertical momenta for several examples with either even or odd symmetry in Fig. 10 in a domain with $R_1 = 1.5$ and $z_0 = 5\epsilon$. These modes have vertical structures with $n = 1, 2, 3$ and 4 ($n = 0, 1, 2, 3$ for ρw), which correspond with those obtained in the isothermal model without boundaries in Section 4.2 (and the bottom panels in Fig. 3) for various radial structures. The vertical shear is strongest at the inner radial boundary of the domain, whereas these modes are concentrated near the outer boundary. We might expect this to be the case because axisymmetric inertial waves are localized at and within their turning surfaces, defined by $\kappa = \omega$ (where $\kappa \approx \Omega$ is the epicyclic frequency). The best resolved modes in our calculations have low frequencies, and thus their turning surfaces lie near or beyond the outer boundary R_1 . Consequently such modes prefer the largest radii possible in the computational domain. Those plotted here were chosen to illustrate that the global model exhibits body modes with the same vertical structure as the reduced model in Section 4.2. The body modes are modestly growing, with a growth rate no larger than a third of the maximum vertical shear rate. These are the unstable modes with the longest wavelengths, and have a radial scale that is shorter than their vertical scale by a factor $O(\epsilon)$.

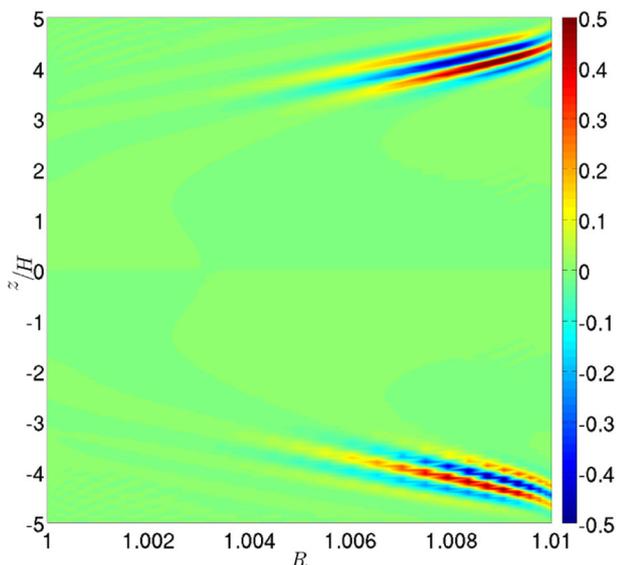
6.1.2 Surface modes

Just as in Section 4.3, we obtain a class of rapidly growing short-wavelength surface modes, which come in pairs with oppositely signed frequencies. For illustration, we show an example of a typical surface mode with and without viscosity in Fig. 11.

Without viscosity, the fastest growing mode always occurs on the smallest available length-scale. To demonstrate this, in the top panel of Fig. 12 we plot the growth rate and (real) frequencies of the fastest growing mode versus resolution N in a disc with $\epsilon = 0.05$ and $q = -1$, adopting a domain with $R_0 = 1.0025$ for several different resolutions $N = N_R = N_z \in [20, 30, 40, 50, 60]$. Such a small radial domain is chosen in order to better capture the rapidly growing surface modes. We also plot the vertical momenta for the fastest growing mode for several different resolutions in Fig. 13. As we increase the resolution, the fastest growing mode moves to lower frequency and exhibits increasingly shorter length-scales. In addition, its growth rate increases as we increase the resolution, gradually tending towards the maximum vertical shear rate (which is $|q|H_0\Omega_0/(2R_0) \approx 0.125$). This is because these modes become increasingly localized in the vicinity of the vertical boundary, where the vertical shear is maximal. Note that Fig. 12 does *not* indicate convergence as N is increased, because it is a *different* mode that is plotted (that is most unstable) for each N , and this mode always



(a) $N_R = N_z = 50, \nu = 10^{-8}, \omega = -0.0419 + 0.0681i$



(b) $N_R = 30, N_z = 120, \nu = 0, \omega = 0.0092 + 0.0607i$

Figure 11. Illustration of the vertical momentum on the (R, z) plane for a typical surface mode with and without viscosity, in an isothermal disc with $\epsilon = 0.05, q = -1, z_0 = 5\epsilon$ and $R_1 = 1.01$. The amplitude is arbitrary.

tracks the grid scale. This is demonstrated further in Fig. 13. Since these modes always occur on the grid scale, they are necessarily the most poorly resolved modes.

Another pathology of the isothermal model, that we first showed in Section 4.3, is that the maximum growth rate depends on the vertical domain size. To further demonstrate this, we illustrate the fastest growing mode for calculations with $z_0 = H\epsilon$, with $H \in [5, 6, 7, 8]$ in Fig. 14 in a radial domain with $R_0 = 1.0025$. We plot results using both a fixed resolution of $N_R = 20$ and $N_z = 60$ (blue circles and line), and for fixed $N_R = 20$ and $N_z/H = 12$ (black crosses and line). The growth rate continues to increase as we increase H . This is simply explained by the dependence of the vertical shear rate, which increases without bound as $O(z)$. However, the finite vertical

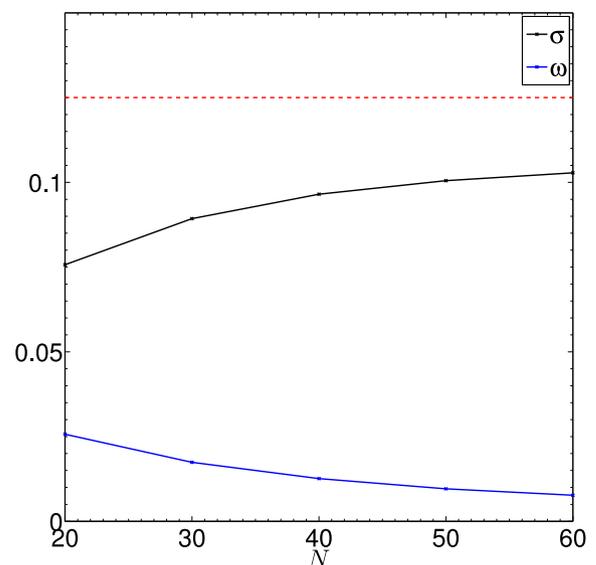


Figure 12. Growth rate and frequency of the fastest growing surface mode as a function of numerical resolution $N = N_R = N_z$ in an isothermal disc without viscosity, with $\epsilon = 0.05, q = -1, z_0 = 5\epsilon$ and $R_1 = 1.0025$. These results were computed using an Arnoldi method. The growth rate increases with resolution N , and the frequency correspondingly decreases, indicating a lack of convergence. The red dashed line is the maximum vertical shear rate.

resolution does not fully capture the fastest growing mode for each H (even with fixed N_z/H), therefore the dependence on H is slightly weaker than the linear extrapolation (red dashed line).

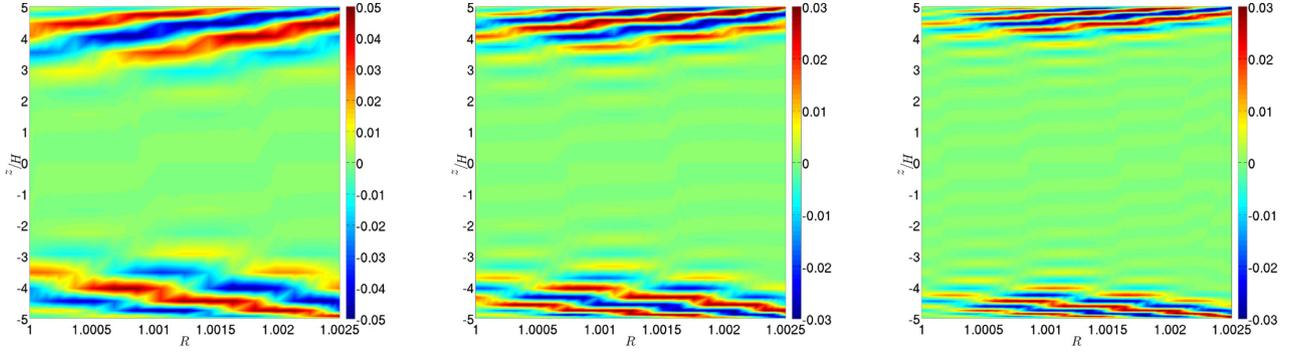
The maximum growth rate for the surface modes compares reasonably well with the numerical simulations of Nelson et al. (2013) for similar parameters (they obtain a growth rate of 0.125 whereas our maximum growth rate is approximately 0.1 for $N = 60$, for example – the difference is due to the inevitably smaller resolution in our case). However, growth rates are unconverged with respect to both (a) the numerical grid resolution and (b) the vertical domain size. The fastest growing mode occurs on the smallest available length-scale, which is always the grid scale. In addition, the fastest growing mode occurs at the boundaries in z . Since the boundary is artificially imposed and this choice sets the maximum growth rate, our results are strongly dependent on an arbitrary parameter H , which is far from ideal. These problems will also afflict non-linear simulations of the VSI in locally isothermal models.

6.2 Viscous calculations

Although the dependence on the vertical boundary cannot be removed, convergence with respect to resolution can be addressed via the inclusion of viscosity. This is necessary for the problem to be well-posed. As we will demonstrate here, even when we include viscosity, the dominant modes prefer the smallest available length-scales.

In Fig. 15, we plot the unstable modes on the complex frequency plane for converged modes in a domain with $z_0 = 5\epsilon$ in various radial domains with several values of ν . Note that these figures are busier than say Fig. 6 because they include modes of all resolvable vertical and radial quantum numbers; in Fig. 6, the radial wavenumber k is restricted to take only one value.

The left-hand panel shows unstable modes when $R_0 = 1.5$. Two classes of instability are obtained in this case: the VSI with $\omega \lesssim 0.4$ and the viscous overstability with $\omega \gtrsim 0.4$, which have been separated visually by the blue dot-dashed vertical line. The viscous



(a) $\omega = 0.0257 + 0.0757i$ ($N_R = N_z = 20$) (b) $\omega = 0.0174 + 0.0893i$ ($N_R = N_z = 30$) (c) $\omega = 0.0126 + 0.0965i$ ($N_R = N_z = 40$)

Figure 13. Illustration of the vertical momenta on the (R, z) plane for the fastest growing surface mode without viscosity for several different resolutions, in a locally isothermal disc with $\epsilon = 0.05$, $q = -1$, $z_0 = 5\epsilon$ and $R_1 = 1.0025$. These have increasing resolution from left-hand to right-hand panels, as indicated in the figure labels. This shows that the fastest growing surface modes occur on the smallest available length-scales; hence this problem is ill-posed because modes on the numerical grid scale will always be the fastest growing. Note the compressed radial scale, which was chosen to capture such rapidly growing surface modes – these modes are actually inclined by a small angle to the vertical, as we expect. The amplitude in each panel is arbitrary.

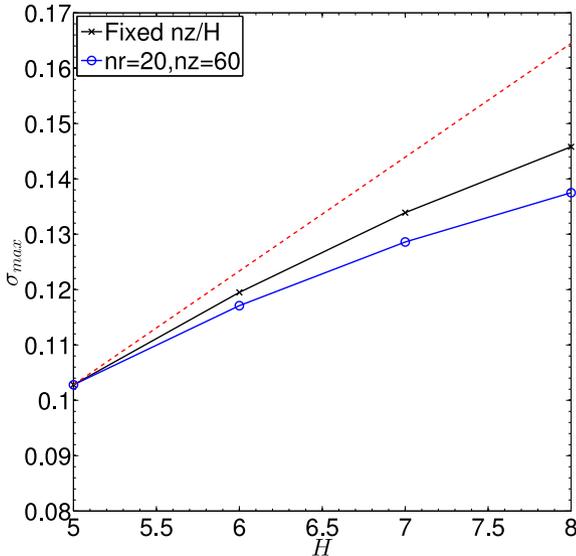


Figure 14. Growth rate of the fastest growing surface modes versus vertical domain size in a locally isothermal disc without viscosity for several values of $z_0 = H\epsilon$, with $\epsilon = 0.05$, $q = -1$ and $R_1 = 1.0025$. Results were computed using an Arnoldi method with fixed resolution $N_R = 20$, $N_z = 60$ (blue circles and line) and with $N_R = 20$ and a fixed value of $N_z/H = 12$ (black crosses and line). The growth rate of the fastest growing mode increases with H , indicating the lack of convergence as we increase the vertical size of our domain.

overstability (Kato 1978; Kley, Papaloizou & Lin 1993; Latter & Ogilvie 2006) preferentially excites modes with little vertical structure and long radial wavelengths $k^{-1} \gtrsim O(H_0)$, with frequencies comparable with the local rotational frequency and growth rates $O(\nu k^2)$. This can be distinguished from the VSI, which occurs only when $q \neq 0$, preferentially excites waves with short radial wavelengths, $\omega \ll 1$, and has growth rates bounded above by the maximum vertical shear rate. When $q = 0$, only the viscous overstability persists – this is modified by vertical shear when $q = -1$, but is found to remain in the same region of the complex plane. Given that our focus here is on the VSI, we will not discuss these modes any further.

The VSI is absent if the viscosity is too large ($\nu \gtrsim 10^{-6}$) since it preferentially excites (in this case) body modes with short radial

wavelengths which are damped by a weak viscosity. We note that a similar order of magnitude for ν is required to obtain the VSI in numerical simulations (Nelson et al. 2013). The fastest growing mode when $\nu = 10^{-6}$ is a body mode that has $n = 1$ (i.e. the fundamental ‘corrugation mode’) with a radial wavelength of approximately 0.05 (similar to the top-left panel of Fig. 10).

The right-hand panel in Fig. 15 plots the unstable modes in a smaller radial domain of $R_0 = 1.05$, for several viscosities. This demonstrates that as we decrease the viscosity, (a) there are more unstable modes and (b) the fastest growing mode has a larger growth rate, in tune with our expectations from the reduced model. The increase in the number of unstable modes is because the smaller the ν the shorter the viscous cut-off; consequently, there is a wider range of length-scales that are potentially unstable (cf. Fig. 8). When $\nu = 10^{-6}$ there is only one unstable mode in this case, which we omit for clarity.

We also observe that the fastest growing modes occur on the shortest available length-scales, since the mode frequency decreases with decreasing ν . However, we note that there are many unstable modes, with the fastest growth rate being somewhat smaller than the maximum magnitude of the vertical shear (0.125) as a result of viscous damping. Surface modes appear, and become dominant when $\nu \lesssim 10^{-7}$.

6.3 Summary

Our results in this section are in accord with the reduced model introduced by Nelson et al. (2013) and revisited in Section 4 but with imposed vertical boundaries. We obtain the same classes of modes and they have similar growth rates in both models (and compared with numerical simulations) in a finite vertical domain. We have highlighted that the VSI preferentially excites ultrashort-scale disturbances which occur on the smallest available length-scales, be they numerical or viscous. Adoption of explicit viscosity is therefore required to obtain results that are converged with respect to resolution.

True isothermal discs do not exhibit surface modes, as we have explained in Section 4. Only the presence of a rigid boundary permits those modes to exist, but such a model is poorly defined given the freedom regarding our choice of z_0 . Given the notable lack of convergence as the vertical domain size is varied, this indicates that

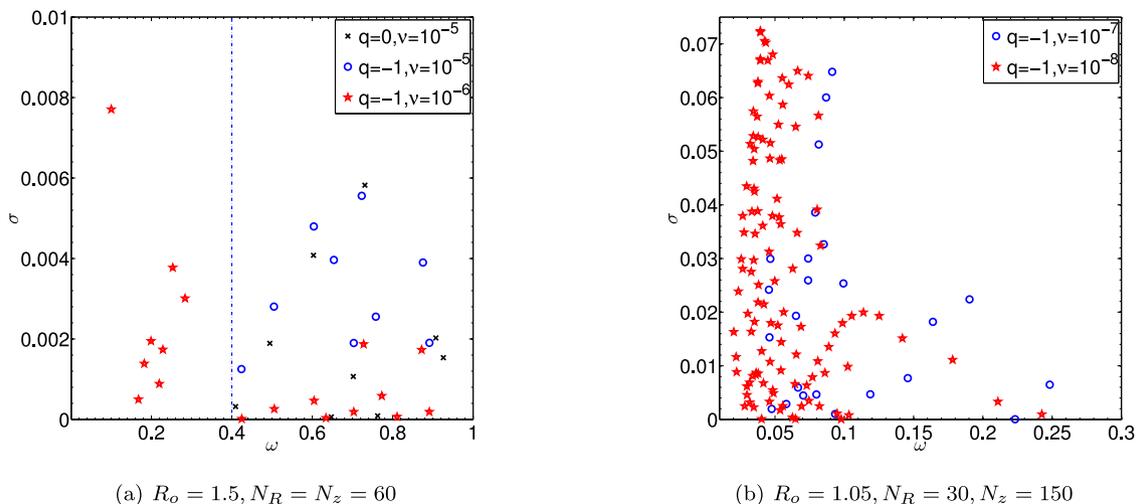


Figure 15. Growth rate of unstable modes versus their (real) frequencies in a locally isothermal disc with $\epsilon = 0.05$, $z_0 = 5\epsilon$, $R_1 = 1.5$ and 1.05 for several viscosities (resolution is listed in the figure labels). The dot-dashed line in the left-hand panel demarcates the low-frequency VSI modes from the viscously overstable modes. The VSI appears only when $\nu \lesssim 10^{-6}$. As the viscosity is decreased, there are more unstable modes, which occur on shorter length-scales, and their maximum growth rates increase. (We caution that some of the eigenvalues in the right-hand panel for $\nu = 10^{-8}$ in particular do not appear to be converged with the adopted resolution – however, we do not expect any of our conclusions to be adversely affected by this.)

the vertically isothermal model is not well suited for studying the linear properties of the VSI.

7 DISCUSSION AND CONCLUSIONS

We have analysed the linear stability of astrophysical discs with vertical shear, as a result of their radial variations in temperature or entropy. Such variations are expected to be present in real discs, as indicated by both observations and theory (e.g. Chiang & Goldreich 1997; Andrews & Williams 2005), and generally lead to vertical shear, thereby rendering the disc unstable to a hydrodynamic instability. Recent non-linear simulations of the resulting VSI have highlighted its potential to drive hydrodynamic activity in MRI-stable regions of protoplanetary discs (Arlt & Urpin 2004; Nelson et al. 2013; Stoll & Kley 2014). The aim of this work was to better understand the nature of the linear instability in two simple disc models: the locally isothermal disc with a radial power law in temperature (building on previous work by Nelson et al. 2013) as well as the locally polytropic disc with a radial power law in entropy.

In both models, there are two classes of unstable modes: modestly growing (vertically) global body modes, and rapidly growing ultrashort-scale surface modes.⁸ The latter only appear in discs with a vertical surface and, though this is not the case in strictly isothermal models, realistic discs should exhibit a density feature/transition upon which such modes can affix themselves. Ironically, artificially imposed boundaries present realistic behaviour even if this behaviour is a numerical artefact! The value of such models beyond the qualitative is unclear, however.

A separate issue is that surface modes preferentially occur on the smallest available length-scales. This necessitates the inclusion of viscosity to obtain convergence in any numerical simulation, otherwise the results will inevitably depend on numerical resolution.

⁸ The surface modes are somewhat analogous to the modes that appear in fingering convection, another type of double-diffusive instability (e.g. Brown et al. 2013).

It may be that the VSI saturates on length-scales much smaller than can be reached by non-linear global simulations, which would consequently pump power to artificially larger scales and hence misrepresent the ensuing turbulent state. This however will be very difficult to test numerically.

We have restricted our study to the locally isothermal and locally polytropic models, since studying the linear VSI in a more realistic model would require the two-dimensional disc structure to be computed numerically after accounting for the various sources of heating and cooling, which are uncertain. In any case, it is unlikely that there is much to be gained from doing this, owing to the similarity in its properties in both the locally isothermal disc (with an artificial boundary) and the locally polytropic disc.

As shown by recent work, the non-linear evolution of the VSI leads to wave activity or turbulence, which transports angular momentum vertically in order to eliminate the vertical shear (Arlt & Urpin 2004; Nelson et al. 2013; Stoll & Kley 2014). It may also transport angular momentum radially to enable the disc to accrete at modest levels. The amount of such hydrodynamical activity in reality depends on the battle between the external heating of the disc, and the efficiency of the VSI in eliminating the ensuing vertical shear. The disc must be externally heated sufficiently strongly or the VSI will eventually win out, leading to a very low level or no turbulence (as observed in Stoll & Kley 2014). That this process can occur on time-scales that are not very long may preclude the use of local computational models to determine the transport properties of the VSI. This is unfortunate, because global simulations of the VSI that are able to capture the fastest growing modes (even in the presence of viscosity) are computationally very challenging. Nevertheless, it would be worthwhile to determine the non-linear outcome of the VSI in more realistic disc models (continuing from Stoll & Kley 2014) to determine its longevity and transport properties.

An interesting byproduct of the VSI's saturation is its radial transport of angular momentum. If the VSI is sufficiently active in protoplanetary dead zones, it may provide the stresses necessary for these regions to accrete. The saturation of the VSI could be controlled by the smallest-scale surface modes, which may be ineffectual in this regard because of their small scales. On the other

hand, longer wavelength body modes carry greater quantities of angular momentum, but only if the saturated state endows them with sufficient power. If this is the case a crude upper limit on the α associated with this transport is of order $\epsilon^3 \sim 10^{-4}$ (which is roughly consistent with previous simulations). On the other hand, Stoll & Kley (2014) show that the measured α decreases with resolution, and is moreover unconverged, a result that suggests that it is in fact the smallest scales that are controlling the transport, not the largest, and that in real discs the value of α could be negligible. Obviously much more work is required to further test these ideas.

ACKNOWLEDGEMENTS

This research was partially supported by STFC through grants ST/J001570/1 and ST/L000636/1. AJB is supported by the Leverhulme Trust and Isaac Newton Trust through the award of an Early Career Fellowship. We would like to thank John Papaloizou for his helpful comments and for generously reading through an earlier version of the manuscript. HNL acknowledges useful discussions with Colin McNally and Martin Pessah. We also thank the referee for promptly reviewing the manuscript.

REFERENCES

- Andrews S. M., Williams J. P., 2005, *ApJ*, 631, 1134
 Arlt R., Urpin V., 2004, *A&A*, 426, 755
 Armitage P. J., 2011, *ARA&A*, 49, 195
 Bai X.-N., 2014, *ApJ*, 791, 137
 Balbus S. A., Hawley J. F., 1998, *Rev. Mod. Phys.*, 70, 1
 Balbus S. A., 2000, *ApJ*, 534, 420
 Bell K. R., Lin D. N. C., 1994, *ApJ*, 427, 987
 Boyd J. P., 2001, *Chebyshev and Fourier Spectral Methods*, 2nd edn. Dover Press, New York
 Brown J. M., Garaud P., Stellmach S., 2013, *ApJ*, 768, 34
 Chiang E. I., Goldreich P., 1997, *ApJ*, 490, 368
 Fricke K., 1968, *Z. Astrophys.*, 68, 317
 Goldreich P., Lynden-Bell D., 1965, *MNRAS*, 130, 125
 Goldreich P., Schubert G., 1967, *ApJ*, 150, 571
 Golub G. H., van Loan C. F., 1996, *Matrix Computations*. Johns Hopkins Univ. Press, Baltimore
 Kato S., 1978, *MNRAS*, 185, 629
 Kato S., 2001, *PASJ*, 53, 1
 Kley W., Papaloizou J. C. B., Lin D. N. C., 1993, *ApJ*, 409, 739
 Knobloch E., Spruit H. C., 1982, *A&A*, 113, 261
 Korycansky D. G., Pringle J. E., 1995, *MNRAS*, 272, 618
 Latter H. N., Ogilvie G. I., 2006, *MNRAS*, 372, 1829
 Lesur G., Ogilvie G. I., 2010, *MNRAS*, 404, L64
 Lesur G., Papaloizou J. C. B., 2010, *A&A*, 513, A60
 Lesur G., Kunz M. W., Fromang S., 2014, *A&A*, 566, A56
 Lin D. N. C., Pringle J. E., 1987, *MNRAS*, 225, 607
 Lubow S. H., Pringle J. E., 1993, *ApJ*, 409, 360
 McNally C. P., Pessah M. E., 2014, preprint ([arXiv:e-prints](https://arxiv.org/abs/1408.0001))
 Nelson R. P., Gressel O., Umurhan O. M., 2013, *MNRAS*, 435, 2610
 Ogilvie G. I., 1998, *MNRAS*, 297, 291
 Okazaki A. T., Kato S., Fukue J., 1987, *PASJ*, 39, 457
 Petersen M. R., Stewart G. R., Julien K., 2007, *ApJ*, 658, 1252
 Ruden S. P., Papaloizou J. C. B., Lin D. N. C., 1988, *ApJ*, 329, 739
 Stoll M. H. R., Kley W., 2014, *A&A*, 572, A77
 Tassoul J.-L., 1978, *Theory of Rotating Stars*. Princeton Univ. Press, Princeton, NJ
 Toomre A., 1964, *ApJ*, 139, 1217
 Umurhan O. M., Regev O., 2004, *A&A*, 427, 855
 Urpin V., 2003, *A&A*, 404, 397
 Urpin V., Brandenburg A., 1998, *MNRAS*, 294, 399

APPENDIX A: DERIVATION OF A REDUCED MODEL FOR THE VSI IN A LOCALLY POLYTROPIC DISC

We look for linear axisymmetric perturbations $(u_R, u_\phi, u_z, P', \rho', S')$ to equations (8)–(10), to which we also include thermal diffusion in the entropy equation of the form $\frac{1}{\rho T} \nabla \cdot [\chi \nabla T]$, where $\chi = \chi(\rho, T, \kappa) = \frac{16\sigma T^3}{3\kappa\rho}$ is the thermal conductivity. We will end up neglecting thermal diffusion when constructing the basic state, but will include it when considering the small-scale perturbations. The motivation is that we are considering slow perturbations with very short radial length-scales, on which thermal diffusion is very rapid, so the perturbations evolve approximately isothermally.

The system of linearized equations is

$$\partial_t u_R = 2R\Omega u_\phi - \frac{1}{\rho} \partial_R P' + \frac{\rho'}{\rho^2} \partial_R P, \quad (\text{A1})$$

$$\partial_t u_\phi = -u_R \frac{1}{R} \partial_R (R^2 \Omega) - u_z \partial_z (R\Omega), \quad (\text{A2})$$

$$\partial_t u_z = -\frac{1}{\rho} \partial_z P' + \frac{\rho'}{\rho^2} \partial_z P, \quad (\text{A3})$$

$$\partial_t \rho' = -u_R \partial_R \rho - u_z \partial_z \rho - \frac{\rho}{R} \partial_R (R u_R) - \rho \partial_z u_z, \quad (\text{A4})$$

$$\partial_t S' = -u_R \partial_R S - u_z \partial_z S + \frac{1}{\rho T} \nabla \cdot (\chi \nabla T'). \quad (\text{A5})$$

We adopt a non-dimensionalization such that our fiducial radius has $R_0 = 1$ and $\Omega_0 = 1$, and take $\epsilon = H_0/R_0 = H_0$ as our small parameter. To obtain a reduced model, we consider slow dynamics on vertical scales comparable with the disc thickness and radial scales that are much smaller. We also assume that the resulting velocities are comparable with the sound speed, with the exception of the radial velocity, which is assumed to be much slower. In particular, we define a slow time-scale $\tau = \epsilon t$ such that $\partial_t = \epsilon \partial_\tau$, along with new radial coordinate x and vertical coordinate z , such that

$$R - 1 = \epsilon^2 x, \quad \partial_R = \epsilon^{-2} \partial_x, \quad (\text{A6})$$

$$Z = \epsilon z, \quad \partial_Z = \epsilon^{-1} \partial_z. \quad (\text{A7})$$

This allows us to neglect curvature and to consider the geometry as locally Cartesian, similar to the classical shearing box (Goldreich & Lynden-Bell 1965; Umurhan & Regev 2004). We also define scaled velocity components (u, v, w) such that

$$u_R = \epsilon^2 u, \quad u_\phi = \epsilon v, \quad u_z = \epsilon w, \quad (\text{A8})$$

and appropriately scale the perturbed pressure, density and entropy for the approximation to be consistent as follows:

$$P' = \epsilon^3 \tilde{P}, \quad \rho' = \epsilon \tilde{\rho}, \quad S' = \epsilon \tilde{S}. \quad (\text{A9})$$

The background pressure is also rescaled as $P = \epsilon^2 \tilde{P}_0$. We also note that ∂_R acting on background quantities (e.g. ρ) is $O(1)$ and ∂_z acting on background quantities is $O(\epsilon^{-1})$. Under the above scaling assumptions (and neglecting thermal diffusion), the basic state in Section 3.1.2 can be written

$$\rho = \rho_0 \left[1 - \frac{z^2}{H_0^2} \right]^m, \quad (\text{A10})$$

$$P = P_0 \left[1 - \frac{z^2}{H_0^2} \right]^{m+1}, \quad (\text{A11})$$

$$R\Omega = R_0 \Omega_0 \left[1 + \frac{q_s z^2}{4\gamma} \epsilon^2 \right]^{\frac{1}{2}}, \quad (\text{A12})$$

where the disc thickness is

$$H_0 = \sqrt{2(1+m)\frac{P_0}{\rho_0}}. \quad (\text{A13})$$

Finally, we take $\chi = \tilde{\chi}\epsilon^\beta$, where β is an ordering parameter. We require $\beta > 2$ so that thermal diffusion can be neglected when constructing the basic state, since

$$\begin{aligned} -\rho T(\mathbf{u} \cdot \nabla S) &= \nabla \cdot (\chi \nabla T) \\ &= \epsilon^{-2+\beta} [\tilde{\chi} \partial_z^2 T + \partial_z \tilde{\chi} \partial_z T] \\ &\quad + \epsilon^\beta [\tilde{\chi} \partial_R^2 T + \partial_R \tilde{\chi} \partial_R T], \end{aligned} \quad (\text{A14})$$

where the left-hand side is $O(1)$. For the perturbations, we have

$$\begin{aligned} \nabla \cdot (\chi \nabla T') &= \epsilon^{-3+\beta} \tilde{\chi} \partial_x^2 \tilde{T} + \epsilon^{-1+\beta} \tilde{\chi} \partial_z^2 \tilde{T} \\ &\quad + \epsilon^{-1+\beta} \partial_R \tilde{\chi} \partial_x \tilde{T} + \epsilon^{-1+\beta} \partial_z \tilde{\chi} \partial_z \tilde{T}. \end{aligned} \quad (\text{A15})$$

The dominant term here is clearly the first one, and this dominates over all other terms in equation (A5) as long as $\beta < 3$, since the leading order term on the left-hand side of equation (A5) is $w \partial_z S$, which is $O(1)$. If $\beta < 3$, the influence of a stabilizing entropy gradient on the perturbations is eliminated by rapid thermal diffusion. We will therefore consider $\beta \in (2, 3)$. While this choice may seem contrived, what this means physically is that we are neglecting thermal diffusion for the basic state, but we are considering it to dominate the thermal evolution of the perturbations, which evolve isothermally.

Applying these scaling assumptions to equations (A1)–(A5) leads to the reduced model

$$0 = 2v - \partial_x h \quad (\text{A16})$$

$$\partial_t v = -\frac{u}{2} - w \frac{q_s z}{2\gamma} \quad (\text{A17})$$

$$\partial_t w = -\partial_z h \quad (\text{A18})$$

$$0 = \rho [\partial_x u + \partial_z w] + w \partial_z \rho \quad (\text{A19})$$

at leading order, where the pseudo-enthalpy perturbation is $h = \tilde{P}/\rho$. The perturbations are in radial geostrophic balance and are anelastic.

APPENDIX B: WKB ANALYSIS OF THE LOCALLY POLYTROPIC REDUCED MODEL

In this section we obtain approximate analytic solutions to equation (43) in the limit of large k . To ease the calculation, we first rescale ω and q_s , introducing the quantities $\hat{\omega} = k\omega$ and $\hat{q} = kq_s/(2\gamma)$. The limit of large k then becomes the limit of large $\hat{\omega}$ and \hat{q} . Next, equation (43) is transformed into Schrödinger form

$$\begin{aligned} \frac{d^2 H}{dz^2} + \left[\hat{\omega}^2 + m \frac{1 - (m-1)z^2}{(1-z^2)^2} + i\hat{q} \frac{1 - (1+2m)z^2}{1-z^2} + \hat{q}^2 z^2 \right] H \\ = 0, \end{aligned} \quad (\text{B1})$$

for the new dependent variable $H = (1-z^2)^{m/2} e^{-i\hat{q}z^2/2} h(z)$. Note that asymptotically equation (B1) is the harmonic oscillator equation for all z except for small regions near the boundaries at $z = \pm 1$.

Near the upper boundary $z = 1$, equation (B1) may be approximated by

$$\frac{d^2 H}{dz^2} + \left[\hat{\omega}^2 + \hat{q}^2 + \frac{m(2-m)}{4(1-z)^2} - \frac{im\hat{q}}{1-z} \right] H = 0, \quad (\text{B2})$$

which admits a solution in terms of generalized Laguerre functions and the Tricomi hypergeometric function. Only the former, however is regular at the boundary. The appropriate solution is hence

$$H \approx (1-z^2)^{m/2} e^{-\varpi(1-z)} L_{-\mu}^{m-1}[2\varpi(1-z)], \quad (\text{B3})$$

where $L_{-\mu}^{m-1}$ is the generalized Laguerre function, $\varpi = \sqrt{-\hat{\omega}^2 - \hat{q}^2}$, and

$$\mu = \frac{m}{2} \left(1 + \frac{i\hat{q}}{\varpi} \right). \quad (\text{B4})$$

In the limit of large ϖ , the solution takes the following asymptotic form far from the boundary:

$$H \sim \Gamma(m-\mu) \exp[\phi(z)] + \Gamma(\mu) \exp[-\phi(z)], \quad (\text{B5})$$

where Γ is the gamma function, and the complex phase function is

$$\phi(z) = \varpi(1-z) + \frac{im\hat{q}}{2\varpi} \ln[2\varpi(1-z)] + \frac{i\pi\mu}{2}. \quad (\text{B6})$$

This solution, in fact, should hold throughout the rest of the domain and so we impose boundary conditions (evenness or oddness) at the mid-plane. This yields the eigenvalue equation

$$\varpi + \frac{im\hat{q}}{2\varpi} \log(2\varpi) - \frac{1}{2} \log\left(\frac{\Gamma(\mu)}{\Gamma(m-\mu)}\right) + \frac{1}{2} i\pi(n+\mu) = 0, \quad (\text{B7})$$

where n is an integer.

Equation (B7) is transcendental in ϖ . However, in the non-vertically shearing case it can be solved easily and we obtain the very simple dispersion relation

$$\omega = \frac{\pi}{4k} (2n+m), \quad (\text{B8})$$

recalling that n is an integer, and m is the polytropic index.

For general \hat{q} , equation (B7) must be solved via a root-finding algorithm, usually an easier task than tackling the ODE itself. However, when $\hat{q} \sim 1$ (meaning $kq_s \ll 1$), we find that the wave frequencies of the unstable body modes are given by equation (B8) to leading order, while the growth rate is

$$\sigma \approx \frac{mq_s}{\pi\gamma(2n+m)} \log\left[\frac{1}{2}\pi(2n+m)\right]. \quad (\text{B9})$$

The growth rate is linear in the shear, but the larger n , the smaller σ , in contrast to the isothermal case. This therefore predicts the $n=1$ body mode to grow fastest at small k , as we have observed in Fig. 9 (though the growth rate is only correct to within an $O(1)$ factor, as might be expected for such small k).

Estimates for the surface mode frequencies can be obtained by equating the first and second terms in equation (B7), assuming that n is not too large. To leading order in large k , this balance yields

$$\omega \approx \frac{iq_s}{2\gamma} + \frac{m \ln k}{4k}. \quad (\text{B10})$$

Thus the growth rates of the surface modes are proportional to q_s , and their wave frequencies are typically a factor $\ln k/k$ smaller.