Extended Sawyer–Eliassen Equation for Frontal Circulations in the Presence of Small Viscous Moist Symmetric Stability

QIN XU

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma/NOAA, Norman, Oklahoma

(Manuscript received 3 June 1987, in final form 10 April 1989)

ABSTRACT

The Sawyer–Eliassen (S–E) equation for frontal circulations forced by a geostrophic stretching deformation is extended to include the effects of both negative moist potential vorticity (MPV) and eddy viscosity. Since the moist (precipitation) region depends on the vertical motion and thus needs to be solved together with the frontal circulation, the extended S–E equation is a nonlinear, elliptic, partial differential equation of sixth order. When MPV is positive and viscosity is negligible, this equation degenerates into the conventional S–E equation. The existence, uniqueness and stability of the solutions of the extended S–E equation in the presence of negative MPV (but still stable to viscous moist symmetric perturbations) are examined both analytically and numerically.

1. Introduction

The Sawyer–Eliassen (S–E) equation has been widely used in many studies of frontal circulation (Hoskins 1982). It is well known that the second order differential operator of the S–E equation is elliptic (hyperbolic) and the problem of solving for frontal circulations from the S–E equation is well-posed (ill-posed) if the geostrophic potential vorticity is positive (negative), in which case the quasi-two-dimensional geostrophic flow is stable (unstable) to viscous symmetric perturbations. When the above criterion is applied to a saturated region, then we need to consider the moist (equivalent) potential vorticity (MPV) instead of dry potential vorticity. Since zero or negative MPV is often observed in regions of frontal precipitation, the associated moist frontal circulations can not be properly solved from the S–E equation.

By using the S–E equation, frontal circulations and frontogenesis in the presence of small positive MPV have been studied by Emanuel (1985, referred as E85 hereafter) and Thorpe and Emanuel (1985, referred as TE85 hereafter). Their results have shown that latent heating can readily be concentrated on the mesoscale for a moderately small positive MPV, with the moist ascent more intense near and on the warm side of the region of maximum geostrophic frontogenesis. On the other hand, the study of conditional symmetric instability (CSI) of Xu (1986b, referred as X86b hereafter) has shown that inviscid CSI occurs as MPV becomes negative and the moist ascents in the most unstable inviscid CSI modes are infinitely narrow. These previous studies suggest that as MPV decreases to zero the moist ascent in a forced inviscid frontal circulation tends to become infinitely narrow and intense. In this case, as mentioned above, the S–E equation becomes invalid and the singularity in the solution for the frontal circulation is clearly due to the inviscid assumption. As shown by X86b, viscous CSI occurs only when negative MPV becomes low enough and the moist ascents in the most unstable viscous CSI modes are narrow but of finite width. This suggests that the S–E equation might be extended to include the effect of eddy viscosity and that the extended equation might remain valid before the onset of viscous CSI. In this paper we show theoretically that the above speculation is true. In section 2 we discuss the singularity of the inviscid, moist, frontal circulation at the asymptotic limit of zero MPV and introduce the concept of a viscous internal boundary layer (IBL) that is between the moist ascent and dry descent. In section 3 we derive the extended S–E equation based on IBL scale analysis in the presence of zero or negative MPV. The existence, uniqueness and stability of the solutions for the extended S–E equation are examined in sections 4 and 5, respectively. Numerical solutions of the extended S–E equation are given in section 6. A summary and conclusions follow in section 7.

2. Inviscid frontal circulation and singularity in the presence of zero MPV

Consider a unidirectional (along y-coordinate) geostrophic flow described by the following three parameters:

© 1989 American Meteorological Society
\[ F^2 = f(f + \eta), \quad S^2 = (g/\Theta_0)\partial_x \theta_g = f \partial_z v_g, \]
\[ N_d^2 = (g/\Theta_0)\partial_\theta \theta_g, \]  
(2.1)

where \( \eta = \partial_x v_g; v_g \) and \( \theta_g \) are the basic geostrophic wind and potential temperature satisfying the thermal wind relationship; \( \Theta_0 \) is the constant reference temperature; the Coriolis parameter \( f \) and acceleration of gravity \( g \) are assumed to be constant. Here \( z = [1 - (p/p_0)^k]^c \Theta_0/g \) is the pseudohight (Hoskins and Bretherton 1972).

Assume that the geostrophic flow (2.1) is superimposed on the following frontogenetic stretching deformation flow:
\[ u'_x = -\alpha x, \quad v'_x = \alpha y, \]
\[ \varphi = -\alpha^2(x^2 + y^2)/2 + f \alpha xy. \]  
(2.2)

According to Hoskins (1982), the semigeostrophic equations for a two-dimensional frontal ageostrophic circulation \( (u, w) \) in the physical space \( (x, z) \) are
\[ F^2 u + S^2 w = -f(\partial_t + D_g + \alpha) v_g, \]  
(2.3a)
\[ S^2 u + N_d^2 w = -(g/\Theta_0)(\partial_t + D_g) \theta_g + S_h, \]  
(2.3b)
\[ \partial_x u + \partial_z w = 0, \]  
(2.3c)
where \( D_g = u'_x \partial_x - \alpha x \partial_x \) is the geostrophic advection and \( S_h \) is the latent heating. The source term \( S_h \) can be related to the vertical motion through the following idealized expression
\[ S_h = (N_d^2 - N^2) w, \]  
(2.4)

where
\[ N^2 = \begin{cases} 
N_w^2 = (\Gamma_w/\Gamma_d)(g/\Theta_0)\partial_x \theta_w, & \text{when saturated and } w > 0 \\
N_d^2, & \text{otherwise}.
\end{cases} \]  
(2.5)

The terms \( \Gamma_w \) and \( \Gamma_d \) are the moist and dry adiabatic lapse rates, respectively, and \( \theta_w \) is the wet bulb potential temperature. Using (2.4)–(2.5) we can rewrite (2.3b) into a compact form in which \( S_h \) is dropped and \( N_d^2 \) is replaced by \( N^2 \). The S–E equation for the ageostrophic circulation is obtained from (2.3a–c) by using the thermal wind relationship and eliminating the tendency terms of \( (v_g, \theta_g) \). If the potential vorticity \( q = N^2 F^2 - S^4 \) is positive, then the S–E equation is elliptic and the problem of solving for the frontal circulation from the S–E equation with an appropriate boundary condition is well posed. Here \( N^2 \) is defined in (2.5), so that in the moist (saturated and \( w > 0 \)) region \( q \) is the moist potential vorticity (MPV) \( q_w = N_w^2 F^2 - S^4 \). In the dry region \( q \) is the dry potential vorticity \( q_d = N_d^2 F^2 - S^4 \).

Equations (2.3)–(2.5) have been used in E85 and TE85 to study frontal circulation and frontogenesis in the presence of small positive MPV. As mentioned earlier, their results have shown that the moist ascent of the inviscid frontal circulation can readily be concentrated on the mesoscale even for a moderately small positive MPV. As MPV decreases toward the negative, the basic geostrophic flow tends to become unstable and slantwise convection. In this case, as shown in X86b, the unstable inviscid CSI modes always tend to have infinitely narrow moist ascents. The results of E85, TE85 and X86b suggest that as MPV decreases toward zero the inviscid solution of the frontal circulation from the S–E equation will become singular with the moist ascent concentrated into an infinitely thin slantwise sheet. The rapid or even discontinuous variation of the ageostrophic motion in the vicinity of this thin singular sheet suggests that there should be a viscous internal boundary layer (IBL) between the moist ascent and dry descent. Within such an IBL the effect of eddy viscosity on the ageostrophic component cannot be neglected.

A viscous extension of the S–E equation has been proposed by Thorpe and Nash (1984) using CISK-type parametrizations of convective heating and PBL physics. Since the latent heating was parameterized in their extended S–E equation, fine structure in the frontal circulation and associated IBL were ignored. Apart from this, they did not consider the cloud mixing effect, which could increase the eddy viscosity dramatically, especially in the vertical. Based on their scale analysis, the eddy viscous effect on the ageostrophic wind was negligible, but the eddy viscous effect on the geostrophic wind was retained as a forcing term on the right hand side of the extended S–E equation.

As far as the effect of eddy viscosity is concerned, the scale analysis and extended S–E equation proposed by Thorpe and Nash (1984) were similar to the analysis of Shapiro (1981). The differential operator for the ageostrophic streamfunction in these previous extended S–E equations was exactly as in the conventional S–E equation. If the heating is related to upward motion locally rather than parameterized in association with the vertical motion at the top of PBL, then the extended S–E equation of Thorpe and Nash (1984) becomes essentially the same as the S–E equation used in E85 or TE85 and the solution will be singular at the limit of vanishing MPV. In order to eliminate the singularity and obtain a uniformly valid solution for the frontal circulation in the presence of zero or negative MPV, we need to consider the viscous regime of ageostrophic motion and make an extension of the conventional S–E equation. In the next section we derive such an extended S–E equation based on IBL scale analysis.

3. IBL scale analysis and extended S–E equation

To filter the inertial-gravity waves, we assume that there is no ageostrophic component in the geopotential (or pressure) field. In other words, the total geopotential is exactly geostrophic and given by \( \varphi + \varphi' \), where \( \varphi \) is...
associated with the thermal wind field \((v_g, \theta_g)\) in (2.1) and \(\phi_\tau\) is defined in (2.2). Consequently, the ageostrophic components are

\[
(u, v, w, \theta).
\]

(3.1)

We assume that these ageostrophic components are functions of \((t, x, z)\). The total flow field consists of (2.1), (2.2) and (3.1). Substituting the total field into the viscous Boussinesq equations gives

\[
(D_t - D)u - \alpha u - f v = 0,
\]

\[
(D_t - D)f u + F^2 u + S^2 w = -f[\partial_x D_g + \alpha - D]v_g = A,
\]

\[
(D_t - D)w - (g/\theta_0)\partial_x \theta = 0,
\]

\[
(D_t - D)(g/\theta_0)\theta + S^2 u + N^2 w = - (g/\theta_0)\{\partial_x D_g - D\} \theta_g = B,
\]

\[
\partial_z u + \partial_z w = 0,
\]

(3.2)

where \(N^2\) is given in (2.5), \(D_t = \partial_t + D_x + u \partial_x + w \partial_z\) is the total advection, \(D_g\) is defined in (2.3), and \(D = \nu_1 \partial_x^2 + \nu_2 \partial_z^2\). Here the coefficients of eddy viscosity \((\nu_1, \nu_2)\) are assumed constant and the Prandtl number is 1. Outside the IBL the scale analysis of (3.2) recovers the semigeostrophic approximation which leads to (2.3) and the conventional S–E equation, so we only need to examine the IBL.

We choose \((\alpha^{-1}, L, H) = (10^5 \text{ s}, 10^3 \text{ km}, 10 \text{ km})\) as the scales for the \((t, x, z)\) variations of the thermal wind \((v_g, \theta_g)\). The time scale \(\tau\) for the ageostrophic wind in the IBL can be assumed as slow as \(\alpha^{-1}\) if the basic flow is stable to viscous CSI and the moist bands are quasi-stationary relative to the frontal system. In this paper we are only interested in those quasi-stationary rainbands, so we assume that \(\tau \ll \alpha^{-1}\) may only occur if the basic flow is unstable to viscous CSI. The \((x, z)\) dimensions of the IBL, denoted by \((l, h)\), depend on eddy viscosity and the slope of the interface, or, loosely, the slope of the moist ascent (see Fig. 1). Note that the moist ascent tends to develop along the most unstable or least stable direction, which is between the slope \(k_o = -f^2/S^2\) of the absolute momentum surface and the slope \(k_g = -S^2/N^2\) of the moist isentropic surface (Xu and Clark 1985). When MPV is either positively or negatively small, \(k_o \approx k_g\) and Fig. 1 suggests that for the interior IBL (away from the upper and lower boundaries)

\[
\beta = h/l \approx k_o = f^2/S_0^2 \approx f/N_0
\]

ranges from \(O(1))\) to \(O(10^{-2})\). (3.3)

where \((f^2, S_0^2, N_0)\) are the scales of \((F^2, S^2, N_w)\). Here (3.3) covers a wide range of \(S_0^2\) and \(N_0\) in various situations of small MPV. The scale for \((l, h) = (1, \beta)l\) is determined by the following condition:

\[
O(D/f) = \max(E_1, E_2) = 1,
\]

(3.4)

where \(D\) is defined in (3.2), \(E_1 = \nu_1/(f^2)\) and \(E_2 = \nu_2/\nu_1\) are the Ekman numbers for the horizontal and vertical diffusions, respectively. With \(\nu_2 = 10^3 \text{ m}^2 \text{ s}^{-1}\) and \(\nu_1 = 10^5 \text{ m}^2 \text{ s}^{-1}\), (3.3)–(3.4) estimate \((l, h) \approx (1, \beta)30 \text{ km}\), where \(1 \gg \beta \gg 10^{-2}\) according to (3.3). When both the moist stratification and thermal wind shear are very weak, the moist ascent tends to be nearly upright (\(\beta \approx k_0 \approx 1\)) and \(h\) may exceed the total depth \(H\) of the frontal circulation. In this case \(h\) should be considered as a length scale for the variation of the ageostrophic flow along the vertical direction in the interior IBL, rather than the depth of the IBL. In the corner region (i.e., the region where the IBL meets the upper or lower PBL) \(E_1 \approx E_2 \approx O(1)\), so \(h / l \approx O(\sqrt{\nu_2 / \nu_1}) \approx O(10^{-2})\) is independent of the slope \(k_o\) (or \(k_0\) if \(S_0^2 < 0\)) of the moist ascent. In order to perform a scale analysis which is valid for both the interior IBL and corner regions, we have to treat \(\beta\) and \(k_0\) as two independent parameters.

The scaling for the ageostrophic and geostrophic components can be introduced as follows:

\[
(t, x, z) \leftrightarrow (t/\tau, x/l, z/h),
\]

\[
(u, v, w, \theta) \leftrightarrow \left[ \frac{u}{u_0}, \frac{v}{u_0}, \frac{w}{\beta u_0}, \frac{(g/\theta_0)\theta}{fu_0/\beta} \right],
\]

(3.5)

\[
(t, x, z) \leftrightarrow (t \alpha, x/L, z/H),
\]

\[
(v_g, \theta_g) \leftrightarrow \left[ v_g/V_0, (g/\theta_0)\theta_g/(fu_0/k_0) \right],
\]

(3.6)

where \(V_0 (10^3 \text{ m}^2 \text{ s}^{-1})\) and \(u_0 (1 \text{ m}^2 \text{ s}^{-1})\) are the scales for \(v_g\) and \((u, v)\), respectively. In (3.6) we assume that the geostrophic components \((v_g, \theta_g)\) are smooth functions in the IBL, specifically, their nondimensional derivatives with respect to \((t, x, z)\) remain \(O(1)\). This assumption may be true initially, but will become invalid as the geostrophic flow is disturbed by the banded ageostrophic circulation and develops a banded substructure. The banded substructure in the geostrophic flow field, for example denoted by \((\tilde{v}_g, \tilde{\theta}_g)\), cannot reach the full intensity of the total geostrophic flow.
because if $O(\hat{v}_g, \hat{\theta}_g) \approx O(v_g, \theta_g)$ the viscous terms $D(\hat{v}_g, \hat{\theta}_g)$ would be much larger than all other terms and lead to a rapid decay of $(\hat{v}_g, \hat{\theta}_g)$. Thus, $(\hat{v}_g, \hat{\theta}_g)$ should be separated from $(v_g, \theta_g)$ and scaled by $u_0$. Consequently, $D(\hat{v}_g, \hat{\theta}_g)$ and $D_k(\hat{v}_g, \hat{\theta}_g)$ are in the same order of magnitude as $D(v, \theta)$ and $D_k(v, \theta)$, respectively, while all the other $(\hat{v}_g, \hat{\theta}_g)$ terms are negligible in comparison with the $(v_g, \theta_g)$ terms. For neatness, we prefer not to write the $(\hat{v}_g, \hat{\theta}_g)$ terms explicitly in the following nondimensional equations, but rather keep the results of the above analysis in mind for later discussion.

Since the restoring force (mainly negative buoyancy) in the dry region is strong, the IBL is expected to develop favorably on the moist side of the interface. Thus we focus our analysis on the moist IBL. Substituting (3.5)–(3.6) into (3.2) with $N^2 = N_w^2$ gives

$$(\tilde{D}_i - \tilde{D})u - (\alpha/f)u - v = 0,$$

$$(\tilde{D}_i - \tilde{D})v + (F^2/f^2)u + \gamma(S^2/S_0^2)w$$

$$= -[\text{Ro}(\partial_i + D_k) + \alpha/f - \epsilon D](v_g/\epsilon_0),$$

$$\beta^2(\tilde{D}_i - \tilde{D})w - \theta = 0,$$

$$(\tilde{D}_i - \tilde{D})\theta + \gamma(S^2/S_0^2)u + \gamma^2(N_w^2/N_0^2)w$$

$$= -[\text{Ro}(\partial_i + D_k) - \epsilon D](\theta_g/\epsilon_0),$$

$$\partial_x u + \partial_z w = 0,$$  \hspace{1cm} (3.7)

where

$$\tilde{D}_i = (f \tau)^{-1} \partial_i + \tilde{D}_k + \tilde{Ro}(u \partial_x + w \partial_z),$$

$$\tilde{D}_k = -(\alpha/f) \tilde{x} \partial_z,$$

$$\text{Ro} = u_0/(fL) \approx 0.3,$$

$$\tilde{D} = E_1 \partial_z^2 + E_2 \partial_z^2,$$

$$\text{Ro} = V_0/(fL) \approx \alpha/f \approx 0.1,$$

$$D_k = -\chi \partial_x,$$

$$\epsilon_0 = u_0/V_0 \approx 0.1,$$

$$\varphi = \beta/k_0,$$

$$\epsilon = \nu_2/(fH^2) \approx 10^{-3},$$

$$D = (v_1/v_2)(H/L) \partial_x^2 + \partial_z^2.$$

As mentioned earlier, if the geostrophic flow is stable to viscous CSI and the moist bands are quasi-stationary to the frontal system, then $(f \tau)^{-1} \approx \alpha/f \ll 1$ and $(f \tau)^{-1} \partial_i$ (the first term in $\tilde{D}_i$) is negligible. The value of $D_k$ (the second term in $\tilde{D}$) is not generally small, however, unless the banded substructure attenuates rapidly as $\tilde{x} > O(\alpha/\alpha)$. It is necessary to point out that the nonlinear ageostrophic advective term (the third term in $\tilde{D}_i$) is actually smaller than $O(\text{Ro}) \approx 0.3$, because of a cancellation between $u \partial_x$ and $w \partial_z$ in the interior IBL. This cancellation is implied by the following two facts: (i) the ageostrophic wind in the interior IBL is largely parallel to the moist updraft, or approximately along the interface; (ii) the gradients of the ageostrophic fields are nearly perpendicular to the interface. In the corner region where the IBL meets PBL, both $w$ and $u$ are small and so is the nonlinear ageostrophic advection. Thus, the nonlinear ageostrophic advection can be neglected although $\text{Ro} \approx 0.3$ is not very small.

For the interior IBL, (3.3) suggests $\beta \approx k_0 \approx O(1)$ and $\gamma \approx O(1)$. In the corner region $\beta \approx O(\sqrt{v_2/v_1}) \approx O(10^{-2})$ and $\gamma \approx O(10^{-2}) - O(1)$. Thus, the parameters $\beta$ and $\gamma$ may vary in a wide range from $O(1)$ to $O(10^{-2})$ and all the terms associated with $\beta$ and $\gamma$ in (3.7) need to be retained for generality. Besides, even though $(\epsilon/\epsilon_0)D(v_g, \theta_g)$ appears very small, we still need to retain the geostrophic viscous terms because the earlier mentioned implicit viscous terms $D(\hat{v}_g, \hat{\theta}_g)$ have the same order of magnitude as the ageostrophic viscous terms $\tilde{D}(v, \theta)$.

The nondimensional equations for the dry IBL have the same form as (3.7) except that $N_w^2$ should be replaced by $N_d^2$. Since $N_d^2/N_0^2 \gg O(1)$ and the term $\gamma^2(N_w^2/N_0^2)w$ should not exceed $O(1)$, it follows that $\beta^2w = \gamma^2 k_0^2 w < \gamma^2 w < 1, \theta < 1$, and more terms can be neglected for the dry IBL. Thus, the general form for the (both moist and dry) IBL equations should have the same form as the moist IBL equations except that $N_w^2$ is replaced by $N_d^2$.

Based on the above analysis we may neglect those uniformly small terms in (3.7) and obtain the following (dimensional) equations:

$$(\delta \partial_i + D_k - D)u - \varphi \tilde{w} = 0,$$  \hspace{1cm} (3.8a)

$$(\delta \partial_i + D_k - D)\tilde{w} + (\phi \theta_0) \partial_i = 0,$$  \hspace{1cm} (3.8b)

$$(\delta \partial_i + D_k - D)(\delta \theta_0 \theta_0) + s^2 u + N_0^2 w = B,$$  \hspace{1cm} (3.8d)

$$\delta \partial_i u + \partial_i w = 0,$$  \hspace{1cm} (3.8e)

where $(A, B)$ are defined in (3.2), $\delta$ is a trace index, and $\delta = 0$ (or 1) if the geostrophic flow is stable (unstable) to viscous CSI. In this paper we only consider the stable case ($\delta = 0$). Outside the IBL, especially in the remote dry region, we have $D_k/f \ll 1$ and $D/f \ll 1$, which implies that $\varphi \tilde{w}$ in (3.8a) and $\theta_0$ in (3.8c) are also very small. In this case the ageostrophic terms associated with $(v, \theta)$ in (3.8b) and (3.8d) can be neglected and (3.8a–e) degenerate into (2.3a–c). Clearly, (3.8a–e) are uniformly valid, both inside and outside the IBL.

Substituting $(D_k - D)(\delta \partial_i(3.8c) - \partial_i(3.8a))$ into $\partial_i(3.8d) - \partial_i(3.8b)$ with $\delta = 0$ gives the following equation for the streamfunction $\psi$ ($u = -\partial_z \psi$ and $w = \partial_x \psi$):

$$\partial_z [F^2 + (D - D_k)^2] \partial_x - \partial_y S^2 \partial_x - \partial_y S^2 \partial_x$$

$$+ \partial_z [N^2 + (D - D_k)^2] \partial_x \psi = 2Q,$$  \hspace{1cm} (3.9)

where $Q = f \partial(v_g, u_g')/\partial(x, z) = \alpha S^2$. We call (3.9) the extended S–E equation. When the basic flow (2.1) is dry and stable, the differential operator in (3.9) is
strongly elliptic and the ageostrophic circulation is spatially smooth, so that $D - D_\gamma$ can be neglected and (3.9) reduces to the conventional S–E equation. For a moist frontal circulation, if

"the banded substructure attenuates rapidly

as \( x > O(f/\alpha) \),"  \( (3.10) \)

then, as mentioned earlier, \( \tilde{D}_\gamma \) will be small in (3.7) and \( D_\gamma \) can be dropped from (3.9). In the following sections, we assume that the above condition (3.10) is satisfied and \( D_\gamma \) can be neglected. This assumption will be checked against the numerical solutions in section 6.

4. Existence of solution

Under the above condition (3.10), \( D_\gamma \) can be neglected and (3.9) reduces to

\[
L_0 \psi = [\partial_x (F^2 + D^2) \partial_x - \partial_z S^2 \partial_x - \partial_x S^2 \partial_z
+ \partial_y (N^2 + D^2) \partial_y] \psi = 2Q. \tag{4.1}
\]

The boundary conditions are

\[
\psi = \partial_x \psi = D \partial_x \psi = 0 \quad \text{on} \quad z = 0, H
\]

\[
\psi = \partial_y \psi = D \partial_y \psi = 0 \quad \text{on} \quad x = \pm L, \tag{4.2}
\]

which describe nonslip rigid boundaries, i.e., \( u = v = 0 \) on \( z = 0, H \), and \( u = v = \theta = 0 \) on \( x = \pm L \).

The interface conditions are

\[ \psi, \nabla \psi, D\psi, D\nabla \psi, D^2 \partial_x \psi, (D^2 + N^2) \partial_y \psi \text{ continuous.} \tag{4.3} \]

The first four conditions in (4.3) are due to the continuity of \( (u, w, v, \theta) \) across the interface. The last two conditions are due to the continuity of the forcing \( (A, B) \) in (3.8). Note from (2.5) that \( N^2 \) is not continuous across the interface.

To examine the existence of the solutions for the system (4.1)–(4.3), we need to define a complete functional space under the constraints of (4.2)–(4.3). Specifically, by \( J^0(\Omega) \) we denote the space of all smooth two-dimensional fields \( \psi \) satisfying (4.2)–(4.3), where \( \Omega \) denotes the domain in (4.2). On \( J^0(\Omega) \) we define the norm

\[
\| \psi \| = \left\{ \int \int |D \nabla \psi|^2 dx \, dz \right\}^{1/2} = \left\{ \int \int |Dv|^2 dx \, dz \right\}^{1/2}, \tag{4.4}
\]

where \( v = (u, w) \) and the integration is over \( \Omega \). The completion of \( J^0(\Omega) \) in the norm (4.4) is a subspace of the Hilbert space (specifically, the Sobolev space) of two-dimensional functions on \( \Omega \), which we denote by \( W^0(\Omega) \). After the above preparation, we can prove the following lemmas and theorems.

Lemma 1. The boundary value problem of (4.1)–(4.3) is equivalent to the variational principle \( \delta \Phi(\psi) = 0 \) for \( \psi \in W^0(\Omega) \) under constraint (2.5), where

\[
\Phi(\psi) = \Phi_0(\psi) + 4 \int \int Q \psi dx \, dz, \tag{4.5}
\]

\[
\Phi_0(\psi) = \int \int \{ (D \partial_x \psi)^2 + (D \partial_y \psi)^2 + F^2 (\partial_x \psi)^2
+ N^2 (\partial_x \psi)^2 - 2S^2 \partial_x \psi \partial_y \psi \} dx \, dz
= \int \int \{ |Dv|^2 + v \cdot \nabla \psi \} dx \, dz \tag{4.6}
\]

\[ \Pi \] is the stability tensor whose components are \( \Pi_{ij} = F_i, \Pi_{12} = \Pi_{21} = S^2, \) and \( \Pi_{12} = N^2 \).

Proof. By using integration by parts with the boundary and interface conditions (4.2)–(4.3) we can show that the variation of \( \Phi \) with respect to a change of \( \delta \psi = \psi' - \psi \) equals

\[
\delta \Phi(\psi) = 2 \int \int \{ L_0 \psi - 2Q \} \delta \psi dx \, dz
+ \Phi_0(\delta \psi, N^2(\psi)) + \int \int \int_{\delta \Omega} (N_x^2 - N_z^2) w^2 \, dx \, dz, \tag{4.7}
\]

where \( \Phi_0(\delta \psi, N^2(\psi)) \) is as in (4.6) except that \( \psi \) is replaced by \( \delta \psi \) explicitly but \( N^2 \) is still associated implicitly with \( w = \partial_x \psi \) (rather than \( dw = \partial_x \psi / \partial x \)) under the constraint (2.5); the last term in (4.7) is the variation due to the dependence of \( N^2 \) on \( w' = \partial_x \psi' = w + \delta \psi, \delta \Omega_0 = \Omega_0 - \delta \Omega_0, \delta \Omega_0 \) and \( \delta \Omega_0 \) represent the moist regions corresponding to \( w' \) and \( w \), respectively.

As \( \delta \psi \to 0 \), measure \( (\delta \Omega_0) \to 0 \) and \( \delta \Omega_0 \) shrinks onto the interface—zero contour of \( w \). For a small \( \delta \psi \), \( \delta \Omega_0 \) is the narrow region (either positive or negative depending on whether \( \delta \psi \) is in \( \Omega_0 \) or in \( \Omega_0 \)) between the zero contour of \( w \) and zero contour of \( w' \). Since \( w' \) is continuous up to the third order differentiation, \( w' \approx O(\delta \psi) \approx O(\delta \psi) \) in \( \delta \Omega_0 \) and the last term in (4.7) is \( O((\delta \psi)^3) \). This can be rigorously shown as follows. We choose curvilinear orthogonal coordinates \( (\xi, \tau) \) in \( \delta \Omega_0 \) such that \( \xi = 0 \) gives the zero contour of \( w' \), i.e., one boundary of \( \delta \Omega_0 \). The other boundary of \( \delta \Omega_0 \) is the zero contour of \( w \) on which we denote by \( \xi = \xi_0(\tau) \) or \( \xi_0 \). Expressing \( w' \) and \( w \) in these new coordinates \( (\xi, \tau) \), we obtain \( w'(0, \tau) = 0 \) and \( w(\xi_0, \tau) = w'(\xi_0, \tau) \)

\[
- \delta w(\xi_0, \tau) = \xi_0 [\partial \xi / \partial \xi] \xi(0, \tau) + O(\xi_0^2) - O(\xi_0^2). \]

It follows that \( 0 \leq |\xi| \leq |\xi_0| \) \( \approx O(\delta \psi) \approx O(\delta \psi) \) and \( w' = w'(\xi, \tau) = \xi [\partial \xi / \partial \xi] \xi(0, \tau) + O(\xi^2) \approx O(\delta \psi) \) in \( \delta \Omega_0 \). Since measure \( (\delta \Omega_0) \approx O(\delta \psi) \), the last term of (4.7) is \( O((\delta \psi)^3) \).

Note that \( \Phi_0(\delta \psi, N^2(\psi)) \) is \( O((\delta \psi)^2) \), so the last two terms in (4.7) are higher order small and can be
neglected. Thus, \( \delta \Phi(\psi) = 0 \) on \( O(\delta \psi) \) if \( \psi \) is a solution of (4.1)-(4.3) under constraint (2.5). Furthermore, since \( \delta \Phi \) is an arbitrary function, the function \( \psi \in W^0(\Omega) \) giving \( \delta \Phi(\psi) = 0 \) on \( O(\delta \psi) \) under constraint (2.5) satisfies (4.1) almost everywhere and is a (generalized) solution of (4.1)-(4.3). (When the forcing \( 2Q \) is continuous, the generalized solution is also a classic solution.) QED.

**Lemma 2.** The functional \( \Phi_0(\psi) \) defined in (4.6) is positive definite in \( W^0 \) if \( \Pi \) is positive definite or if the eddy viscosity is strong enough (when \( \Pi \) is not positive definite). A sufficient estimation for the latter is

\[
E^2 = \pi^4 \left( \frac{\nu_1}{L^2} + \frac{\nu_2}{H^2} \right)^2 > \lambda_0, \quad (4.8)
\]

where \( \lambda_0 = \min \{ \lambda_{\min}(x, z) : (x, z) \in \Omega \} \) and \( \lambda_{\min}(x, z) \) is the minimum eigenvalue of \( \Pi = \Pi(x, z) \).

**Proof.** Obviously, \( \Phi_0(\psi) \) will be positive definite if \( \Pi \) is positive definite, so we only need to consider the case in which \( \Pi \) is not positive definite, i.e., \( \lambda_0 > 0 \). By using the inequality (4.4) of Xu (1987), integration by parts, and interface conditions (4.2)-(4.3), and the Cauchy–Schwarz inequality, we obtain

\[
\int \int u^2 \, dx \, dz \leq K_1 \int \int (\partial_x u)^2 \, dx \, dz
\]

\[
= -K_1 \int \int u (\partial_x^2 u) \, dx \, dz, \quad (4.9a)
\]

\[
\int \int u^2 \, dx \, dz \leq K_2 \int \int (\partial_x u)^2 \, dx \, dz
\]

\[
= -K_2 \int \int u (\partial_x^2 u) \, dx \, dz, \quad (4.9b)
\]

\[
E \int \int u^2 \, dx \, dz \leq - \int \int u (Du) \, dx \, dz
\]

\[
\leq \left( \int \int u^2 \, dx \, dz \right)^{1/2} \left( \int \int (Du)^2 \, dx \, dz \right)^{1/2} \quad (4.9c)
\]

\[
E^2 \int \int u^2 \, dx \, dz \leq \int \int (Du)^2 \, dx \, dz, \quad (4.9d)
\]

where \( K_1 = (L/\pi)^2, K_2 = (H/\pi)^2, \) and \( E \) is defined in (4.8). Here (4.9c) is obtained from (4.9a) + (4.9b). Similarly, (4.9d) can be applied to \( w \) and thus, to \( v \):

\[
E^2 \int \int |v|^2 \, dx \, dz \leq \int \int |Dv|^2 \, dx \, dz = \| \psi \|^2. \quad (4.10)
\]

Substituting (4.10) into \( \Phi_0(\psi) \) gives

\[
\Phi_0(\psi) \geq \| \psi \|^2 - \lambda_0 \int \int |v|^2 \, dx \, dz
\]

\[
\geq \| \psi \|^2 (1 - \lambda_0 / E^2),
\]

which shows that \( \Phi_0(\psi) \) is positive definite in \( W^0 \) when (4.8) is satisfied. QED.

**Theorem 1.** At least one solution exists for (4.1)-(4.3) if \( \Phi_0(\psi) \) is positive definite, or, sufficiently, if (4.8) is satisfied.

**Proof.** Use Lemmas 1-2 and note that the \( W^0(\Omega) \) is compactly embedded in the space of square integrable functions and, thus, \( \Phi(\psi) \) has at least one minimum in \( W^0(\Omega) \) if \( \Phi_0(\psi) \) is positive definite. QED.

5. Local uniqueness and stability

Although the global uniqueness of the solution of (4.1)-(4.3) is suggested by numerical solutions in section 6, a rigorous proof seems difficult because \( N^2 \) depends on \( w \) and the operator \( L_0 \) in (4.1) is nonlinear. Nevertheless, we can prove the following local uniqueness (i.e., isolation) of the solution.

**Theorem 2.** Under the condition (4.8) a solution \( \psi = \psi_0(\cdot)-(4.3) \) is uniquely local in a small ball of small enough radius centered at \( \psi_0 \) in the functional space \( W^0 \).

**Proof.** To prove the theorem, we only need to show that \( \Phi(\psi_0) \) is a local minimum of \( \Phi(\psi) \). Consider \( \delta \psi = \psi - \psi_0 \) and \( \delta \Phi(\psi) = \Phi(\psi) - \Phi(\psi_0) \), we can rewrite (4.7) as

\[
\Phi(\psi) = \Phi(\psi_0) + \Phi_0(\delta \psi, N^2(\psi_0)) + \int \int_{\Omega_0} (N^2 - N_{\delta \psi}^2) w^2 \, dx \, dz, \quad (5.1)
\]

where \( w_0 = \partial \psi_0 / \partial x \) and the fact that \( L_0 \psi_0 - 2Q = 0 \) is used. The distribution of \( N^2(\psi_0) \) in \( \Phi_0(\delta \psi, N^2(\psi_0)) \) is associated with \( \psi_0 \) and independent of \( \delta \psi \), so \( \Phi_0(\delta \psi, N^2(\psi_0)) \) is exactly quadratic with respect to \( \delta \psi \). This is different from \( \Phi_0(\psi) \), in which \( N^2 = N^2(\psi) \) depends on \( \psi \) under constraint (2.5). The proof of Lemma 2, however, is independent of the dependence of \( N^2(\psi) \) on \( \psi \), so Lemma 2 also applies to \( \Phi_0(\delta \psi, N^2(\psi_0)) \). Thus, \( \Phi_0(\delta \psi, N^2(\psi_0)) \) is quadratic and positive definite, the last term in (5.1) is \( O((\delta \psi)^3) \), and \( \Phi(\psi_0) \) is a local minimum of \( \Phi(\psi) \). QED.

In view of the fact that the moist region \( \Omega_0 \) is given in association with the solution \( \psi = \psi_0 \) in Theorem 2, the condition (4.8) can be improved to

\[
E^2 = \pi^4 (1/d_1^2 + 1/d_2^2)^2 > \lambda_0
\]

\[
= \max \{ S^4/F^2 - N_{\psi_0}^2(x, z) : (x, z) \in \Omega_0 \}, \quad (5.2)
\]

where \( d_1 \) and \( d_2 \) are the maximum and minimum diameters of the convex hull of \( \Omega_0 \) in the viscous-normalized space \( (x/\sqrt{\nu_1}, z/\sqrt{\nu_2}) \). The detailed proof of (5.2) is given in the Appendix.

**Corollary 2.** A solution \( \psi = \psi_0 \) of (4.1)-(4.3) is locally unique if (4.8) or (5.2) is satisfied.

**Theorem 3.** A solution \( \psi = \psi_0 \) of (4.1)-(4.3) is stable to viscous CSI if (4.8) or (5.2) is satisfied.

**Proof.** We superimpose a small perturbation \( \psi' \) on \( \psi_0 \) and substitute the total fields \( (u, v, w, \theta) \) obtained from the composed \( \psi = \psi_0 + \psi' \) into (3.8a-d) with \( \delta \)
= 1 and \( D_x = 0 \). Following the manipulations which lead to (3.9) and using the fact that \( \psi_0 \) satisfies (4.1), the above equations can be combined to give the following equation:

\[
\{ \partial_z (F^2 + (D - \partial_z)^2) \partial_z^2 - \partial_x S^2 \partial_x - \partial_z S^2 \partial_x \}
\quad + \partial_z \{ N^2 + (D - \partial_z^2) \partial_x \} \psi' = 0, \tag{5.3}
\]

where \( N^2 = N^2(\psi_0) \) is associated with \( w_0 = \partial \psi_0 / \partial x \). Here, like the last term in (4.7) or (5.1), the perturbation due to the dependence of \( N^2 \) on \( w \) is higher order small and has been neglected in (5.3).

Assuming \( \psi(x, z, t) = Re \{ \hat{\psi}(x, z) e^{i\sigma} \} \), where \( \sigma \) and \( \hat{\psi}(x, z) \) are complex, we may replace \( \partial_z \), with \( \partial_x \) in (5.3). To prove the theorem we only need to show \( Re(\sigma) < 0 \). By using integration by parts with the boundary and interface conditions (4.2)–(4.3), we obtain

\[
0 = \int\int \hat{\psi}^*(4.5) \, dx \, dz
\]

\[
= \Phi_0(\hat{\psi}, N^2(\psi_0)) + 2 \sigma \Phi_1(\hat{\psi}) + \sigma^2 \Phi_2(\hat{\psi})
\]

or

\[
\sigma = -\Phi_1(\hat{\psi})/\Phi_2(\hat{\psi}) = \{ [\Phi_1(\psi) + \Phi_1(\psi)]^2
\quad - \Phi_0(\psi, N^2(\psi_0))/\Phi_2(\psi) \}^{1/2}, \tag{5.4}
\]

where \((\cdot)^*\) represents the complex conjugate of \((\cdot)\),

\[
\Phi_0(\psi, N^2(\psi_0)) = \int\int \{ |D\hat{\psi}|^2 + \hat{\psi}^* \cdot \Pi(\psi_0) \hat{\psi} \} \, dx \, dz
\]

\[
\Phi_1(\psi) = \int\int \{ \nu_1 |\partial_x \hat{\psi}|^2 + \nu_2 |\partial_z \hat{\psi}|^2 \} \, dx \, dz
\]

\[
\Phi_2(\psi) = \int\int |\hat{\psi}|^2 \, dx \, dz.
\]

We can show that \( \Phi_0(\psi, N^2(\psi_0)) \) is positive definite under condition (4.8) or (5.2), which is similar to the positiveness of \( \Phi_0(\psi, N^2(\psi_0)) \) in the proof of Theorem 2 or Corollary 2. Since \( \Phi_1(\psi) \) and \( \Phi_2(\psi) \) are also positive definite, (5.4) indicates \( Re(\sigma) < 0 \). QED.

In the proofs of Theorems 2 and 3, the local uniqueness and stability of the solution \( \psi_0 \) is essentially ensured by the definite positiveness of \( \Phi_0(\psi, N^2(\psi_0)) \). Since the positiveness of \( \Phi_0(\psi, N^2(\psi_0)) \) can be easily checked numerically, the following corollary is useful.

**Corollary 3.** A solution \( \psi = \psi_0(\psi, N^2(\psi_0)) \) is unique locally and stable to viscous CSI if \( \Phi_0(\psi, N^2(\psi_0)) \) is positive definite in \( W^0 \).

The reverse statement for the stability problem in Corollary 3 is that viscous CSI instability occurs when \( \Phi_0(\psi, N^2(\psi_0)) \) is not positive definite, while the marginal instability occurs at \( \min \Phi_0(\psi, N^2(\psi_0)) = 0 \). Clearly, this reverse statement can be proven to be true by using (5.4) and showing \( Re(\sigma) \geq 0 \) as in the proof of Theorem 3. In the absence of forcing and moist processes, \( \min \Phi_0(\psi) = 0 \) recovers the variational formulation of Emanuel (1979) for marginal dry viscous symmetric stability.

### 6. Numerical solutions

Numerically, the generalized solution is obtained in a truncated functional space \( \tilde{W}^0 \) (a subspace of \( W^0 \)) constructed by the two-dimensional cubic-spline basis functions on rectangle finite elements. Each element is associated with \( (2 \times 3)^2 = 36 \) basis functions:

\[
B^{\beta}_{ij}(x, z) = B_{i}^{\alpha}(x) B_{j}^{\beta}(z), \quad i, j = 1, 2
\]

and \( \alpha, \beta = 0, 1, 2 \), where \( B_{i}^{\alpha}(x) \) and \( B_{j}^{\beta}(z) \) are one-dimensional basis functions. The basis function \( B_{i}^{\alpha}(x) \) (with fixed \( i \) and \( \alpha \)) and its derivatives of order \( \alpha' = 0, 1, 2 \) at the nodes \( x = x_i \) (\( i' = 1, 2 \)) satisfy

\[
\partial_{x_i} \partial_{x_{i'}} B_{i}^{\alpha}(x_i') = \delta_{ii'} \delta_{aa'}, \tag{6.1}
\]

where \( \delta_{ii'} \) and \( \delta_{aa'} \) are Kronecker deltas. To construct \( B_{i}^{\alpha}(x) \) over a line element \( \Delta x = x_2 - x_1 \), we divide \( \Delta x \) equally into three segments and thus obtain 12 constraints for \( B_{i}^{\alpha}(x) \) : 6 of them are given by (6.1) at the two nodes \( x = x_1 \) and \( x = x_2 \) and the remaining 6 constraints are the continuity of \( \partial_{x_i} \partial_{x_{i'}} B_{i}^{\alpha}(x) \) (\( \alpha' = 0, 1, 2 \)) at the two internal knots \( x = x_1 + \Delta x/3 \) and \( x = x_1 + 2\Delta x/3 \) between the segments. Note that each cubic polynomial contains 4 parameters over each interval of \( \Delta x/3 \) and the total number of parameters are \( 4 \times 3 = 12 \), so \( B_{i}^{\alpha}(x) \) can be determined uniquely by the cubic-spline Hermite interpolation over \( \Delta x \). The situation for \( B_{i}^{\beta}(z) \) is similar and the detailed formulations can be found in Yuan (1985). The two-dimensional basis functions are obtained from \( B_{i}^{\alpha}(x) B_{j}^{\beta}(z) \) over a rectangle element, while the element is divided into 9 subelements.

The expression for \( \psi \in \tilde{W}^0 \) is

\[
\psi(x, z) = \psi^{\beta}_{ij} B^{\beta}_{ij}(x, z), \tag{6.2}
\]

where the summation convention (implied by double indices) is used and \( \psi^{\beta}_{ij} \) is the value of \( \partial_x \partial_x \partial_z \partial_z \psi \) at the node \( (x, z) = (x_i, z_j) \). Each node has 9 components of \( \psi^{\beta}_{ij} \) corresponding to \( \alpha, \beta = 0, 1, 2 \). The basic state parameter \( N^2 \) equals either \( N_u^2 \) or \( N_f^2 \) within each subelement, so that a slantwise interface between the moist and dry regions is approximated by a staircase boundary between the moist and dry subelements (see Fig. 2a). If the moist subelements are prescribed, then we can substitute (6.2) into \( \Phi(\psi) \), set the variation \( \delta \Phi(\psi^{\beta}_{ij}) = 0 \) with respect to all \( \psi^{\beta}_{ij} \) (except for those with given boundary values), and obtain a linear algebraic system for \( \psi^{\beta}_{ij} \). Once the \( \psi^{\beta}_{ij} \) are calculated, (6.2) gives the truncated generalized solution of the extended S–E equation. This solution, however, may not satisfy condition (2.5). If not, then adjust the moist region according to the vertical velocity field averaged over each subelement and iterate the above procedure until condition (2.5) is satisfied.
According to Corollary 3, the numerical solution will be stable to viscous CSI if the functional \( \Phi(\psi^\phi) \) is positive definite in the truncated functional space \( \tilde{\mathcal{W}}^0 \), or, equivalently, if the matrix associated with \( \psi^\phi \) in the above linear algebraic system is positive definite. The latter can be easily checked when the system is solved by Gaussian elimination. In this way, all the solutions in Figs. 2–4 are found to be stable to viscous CSI.

To reduce the effect of lateral boundaries, the computation domain is chosen to be wide enough: \( L = 10L_R \), where \( L_R = HN_1/f = 100H \) is the Rossby radius of deformation and \( N_1 = 10^{-2} \text{ s}^{-1} \) is the scale of \( N_d \). To increase the resolution in the central region of the domain where the frontogenetic forcing is large and banded substructure occurs, the domain is divided non-uniformly in the horizontal direction by 16 columns of elements (48 columns of subelements). In the vertical direction are 8 equally divided layers of elements (24 layers of subelements), so the total number of subelements is \( 9 \times 16 \times 8 = 48 \times 24 \). Figure 2a shows the subelements in the central region \([ -L_R, L_R ]\)
which is only one-fifth of the computation domain. The open (half width) subelements on each side of Fig. 2a are outside the central region, beyond which are 5.5 more columns of subelements (not shown).

The following idealized frontogenetic forcing [similar to (19) of E85] is used:

$$2Q = 2\alpha S^2 = 2Q_0 e^{-\xi/L_b}$$  \hfill (6.3)

where $2Q_0 = 2\alpha S_0^2$ (constant) is the forcing amplitude, $\xi = x + z/k_0$, $k_0 = f^2/S_0^2$, and $S_0^2$ are defined in (3.3), and $L_b$ is the attenuation length of the forcing. (The forcing attenuates by a factor of $e^{x} = 23.1$ over $L_b$ away from the region of maximum forcing.) To represent a concentrated forcing, such as the one used in E85, we may choose $L_b/L_R = 0.14$. For a widely distributed forcing, we choose $L_b/L_R = 1.57$ as a representative parameter setting.

Based on the form of $2Q = 2\alpha S^2$ specified in (6.3) and the thermal wind relationship, the basic state parameters in (2.1) should be functions of space and have the following forms:

$$S^2 = S_0^2 e^{-\xi/L_b}$$

$$F^2 = f^2 \{1 + e^{-\xi/L_b} - e^{-\xi/L_b} \}$$

$$N^2 = N_0^2 \{1 + Ri^{-1}e^{-\xi/L_b} \}$$  \hfill (6.4)

where $Ri = N_0^2 f^2/S_0^4$ is the scale Richardson number. The moist stratification is independent of (6.4). We assume that $N_0^2 = constant$ and the moist region is confined within the region of $|x| \leq L_b$ and $z \geq 0.125H$ (see Figs. 2a and 3a).

Figures 2a, b show the moist region and streamfunction of the numerical solution obtained with the following parameter setting: $\nu_1 = 10^5 \text{ m}^2\text{s}^{-1}$, $\nu_2 = 10^2 \text{ m}^2\text{s}^{-1}$, $f = 10^{-8} \text{s}^{-1}$, $N_0^2 = 10^{-4} \text{s}^{-2}$, $L_b = 30 \text{ km}$, $RI = 1.3$ (or $S_0^2 = 0.707 \times 10^{-13} \text{s}^{-2}$), $N_0^2 = 0.0$ (or $q_w = -S^4$), $2Q_0 = 10^{-11} \text{s}^{-2}$ (or $\alpha = 0.707 \times 10^{-13} \text{s}^{-2}$), $H = 100 \text{ km}$ and $L_b = 140 \text{ km}$ (or $L_b/L_R = 0.14$). In Fig. 2a the moist subelements are plotted with dark boundary lines. If the vertical velocity averaged over a subelement is positive, then a plus sign is plotted at the subelement center. As shown, all the moist subelements match the plus signs and the constraint (2.5) is well satisfied. The dashed line in Fig. 2 indicates the axis of maximum forcing, i.e., the line of $\xi = x + z/k_0 = 0$. The streamline intervals in Fig. 2b are $0.002 \times Q_0 H^2/f^2 = 200 \text{ m}^2\text{s}^{-1}$, so the ageostrophic circulation is weak ($u \leq 0.5 \text{ m s}^{-1}$ and $w \leq 1 \text{ cm s}^{-1}$). The moist region is similar to Fig. 2, except for $L_b = 1570 \text{ km}$ (or $L_b/L_R = 1.57$). In this case the forcing is widely distributed and the volume-integrated forcing is much stronger than in Fig. 2. The negative MPV ($q_w = -S^4$) is also widely distributed. Note that the streamline intervals in Fig. 3 are $800 \text{ m}^3\text{s}^{-1}$. The circulation is moderately intense and evolves into a multiply banded substructure in response to the widespread forcing and negative MPV. The ageostrophic wind ($u \leq 3 \text{ m s}^{-1}$ and $w \leq 8 \text{ cm s}^{-1}$) in Fig. 3 does not exceed the assumed order of magnitude in the scale analysis of section 3. The second band in Fig. 3 is very weak and the overall banded substructure (in either Fig. 3 or Fig. 2) attenuates rapidly away from the region of maximum forcing, so the condition (3.10) is satisfied and the model is self-consistent, at least for the solutions in Figs. 2 and 3.

The moist ascent in either Fig. 2 or Fig. 3 is steeper than the axis of maximum forcing and tilts up into the warm air with a slope between that of the absolute momentum surface and moist isentropic surface (not shown), so the moist inertial-buoyancy energy of the basic state is released efficiently. This feature is commonly observed in all other solutions of different parameter settings (see Xu 1989). When the forcing is horizontally concentrated the frontal circulation generally contains only one moist band. The band becomes narrow and intense as the viscous CSI stability decreases. If the forcing is widespread and MPV is negative (but still stable to viscous CSI), however, then a multibanded substructure will occur. The bands become intense and narrow as MPV becomes more negative and/or the eddy viscosity becomes smaller (until the solution becomes marginally unstable to viscous CSI). Detailed discussion and a physical explanation for the above findings are given in Xu 1989).

The above and many other (not shown) numerical solutions indicate that the extended S–E equation (4.1) can be generally valid (in terms of self-consistency) until the solution becomes unstable to viscous CSI. An example is given in Fig. 4 in which the solution is almost unstable (but still stable) to viscous CSI instability. The parameter setting for the solution in Fig. 4 is as in Fig. 3, except that $N_0^2 = max \{ -10^{-8} \text{s}^{-2}, -(S^4 + 10^{-13} \text{s}^{-4})/f^2 \}$ and the moist region is extended beyond $x = L_R$. Note that $N_0^2 = -(S^4 - q_w)/f^2$, so the moist parameter setting in Fig. 4 is that either $N_0^2$ is constant $= -10^{-8} \text{s}^{-2}$ or $q_w = constant = -10^{-13} \text{s}^{-2}$ whichever gives a higher value for $N_0^2$ (or $q_w$). Specifically, the MPV in Fig. 4 is more (less) negative than in Fig. 3 in the moist region on the warm side far away from (close to) the axis of maximum forcing, i.e., the region $\xi = x + z/k_0 \geq 0$ and $\xi = x + z/k_0 < 402 \text{ km}$. The minimum MPV ($q_w = -10^{-13} \text{s}^{-2}$) in Fig. 4 is uniformly distributed in the moist region between $0 < \xi = x + z/k_0 < 402 \text{ km}$, while in Fig. 3 the MPV ($q_w = -S^4$) reaches the minimum $(-5 \times 10^{-13} \text{s}^{-2})$ at the lower left corner of the moist region where the moist region intersects the maximum forcing line. Consequently, in comparison with Fig. 3, the first moist band in Fig. 4 is slightly wider and weaker, its induced dry subsidence on the warm (right) side is weaker and narrower, but the second moist band becomes slightly stronger. The ageostrophic wind ($u \leq 3 \text{ m s}^{-1}$ and $w \leq 5 \text{ cm s}^{-1}$) in Fig. 4 does not exceed the assumed order of magnitude in the scale analysis of section 3. The second band in Fig. 4 is still weak (al-
Fig. 3. As Fig. 2, except that $L_0 = 1570$ km and the contour intervals are $800 \text{ m}^2 \text{ s}^{-1}$.

though slightly stronger than the second band in Fig. 3) and the overall banded substructure attenuates rapidly away from the region of maximum forcing, so the condition (3.10) is satisfied and the extended S–E equation (4.1) remains valid. Here we need to point out that (4.1) may become invalid for some special cases. For example, if the forcing is widespread and both the negative MPV and dry potential vorticity are uniformly low, then the second and/or third band will become moderately strong and the $D_e$ terms can no longer be neglected in (3.8) or (3.9). In this case, it seems that a valid (self-consistent) solution for the moist frontal circulation should be obtained from the extended S–E equation (3.9). This problem needs further investigation.

The convergence rates of iterations for different parameter settings have also been examined. Generally, the more stable the moist geostrophic flow, the faster the iterations converge. When the parameter setting is very close to marginal viscous CSI instability, the iterations converge very slowly and usually take 20–30 steps. As the parameter setting moves toward significant CSI instability, the iterations do not seem to converge. For a given parameter setting, if the iterations converge, they always converge to the same solution even when started with very different initial guesses. In other
words, all the numerical solutions obtained so far (most of them not shown) are unique not only locally but also globally as long as they are stable to viscous CSI.

The solutions presented in this paper (Figs. 2–4) are all stable to viscous CSI and the bands are rather wide and located on the warm side of the region of maximum forcing. If the axis of maximum forcing is considered as a cold front, then the single band in Fig. 2 may resemble a cold frontal rainband while the multibands in Fig. 3 or Fig. 4 may resemble warm-sector rainbands (or warm-frontal rainbands in a partially occluded frontal system). Here it is interesting to note that as MPV is below the onset point of viscous CSI in the region immediately above the axis (dashed line) of maximum forcing in Fig. 2, the main ascent in Fig. 2 is found (not shown) to break up into fine and intense multibands in the region of instability, which seems to resemble some of the features in the numerical simulations of multiple cold-frontal rainbands carried out by Knight and Hobbs (1988) in conditions unstable to CSI. But, in this case, our solution can be obtained only with a given distribution of \( N^2 \) which largely but not exactly satisfies the constraint (2.5). Besides, the solution is slightly unstable to viscous CSI, so the result
is incomplete unless the contribution from the weakly unstable CSI mode is included or proven to be small. (Presumably the weakly unstable mode will not grow to a large amplitude.) This problem is still under investigation.

7. Summary and conclusions

In this paper the S–E equation for frontal circulations is extended to include the effects of both negative MPV and eddy viscosity. This extension is justified by IBL scale analysis that shows that the effect of eddy viscosity to ageostrophic motion is important in the presence of zero and negative MPV. On the other hand, the inviscid ageostrophic circulation obtained from the conventional S–E equation becomes singular with the moist ascent concentrated into an infinitely thin slantwise sheet as the MPV decreases towards zero.

It is also rigorously shown that a solution exists for the extended S–E equation when the negative MPV is above the onset point of viscous CSI while the latter condition is ensured by the definite positiveness of the variational formulation of the homogeneous part (left hand side) of the extended S–E equation. Under the same condition the solution is proven to be at least locally unique and stable. Numerical solutions are obtained for the extended S–E equation with a cubic-spline finite-element method and the results suggest that the solutions are unique and stable not only locally but also globally under the above condition. Numerical solutions are also obtained for broad ranges of parameter settings (many of them not shown) and it is generally found that the intensity, structure and scale of the bands are controlled by the competition between the frontogenetic forcing, negative MPV and eddy viscosity. Detailed discussions of these findings are reported in Xu (1989).

Our results suggest that the extended S–E equation may be useful for diagnosing CSI-related frontal rainbands. Pure CSI theory has been useful for explaining and assessing some fast-growing bands. But, the fast-growing CSI bands can not last long after their initial development, unless the moisture supply is maintained by the large-scale frontal lifting or by rapid remoistening and recycling of the dry subsiding air at the lower levels (Xu 1986a, 1986c). Thus, the pure CSI concept is often hard-pressed to explain some observed long-lasting rainbands. If the band formation is assisted by frontogenetic forcing, however, then longer persistence can often be predicted. In view of the fact that frontal rainbands often coexist with frontogenetic forcing and small or negative MPV, we may expect that the extended Sawyer–Eliassen equation will find wide use.

Finally, we need to remind readers that in the derivation of the extended S–E equation [either (3.9) or (4.1)] we have filtered the inertial–gravity waves, assumed that the frontal circulation and its embedded substructure are quasi-stationary, and, consequently, ignored those propagating frontal rainbands, such as fast moving squall lines. These are the limitations of our extended S–E equation.

Acknowledgments. I am grateful to Professor D. K. Lilly for his encouragement and support and to Dr. Si Yuan for his help in developing the numerical model of finite element method. The comments and criticisms from the anonymous reviewers have been helpful in many aspects. Thanks also extended to Dr. Bill McCaul for proofreading. The financial support was provided by the NOAA contract NA85-RAH05046 and NSF Grant ATM-8701522 at CIMMS, University of Oklahoma.

APPENDIX

Proof of (5.2)

We separate $N_s^2$ in the moist region $\Omega_0$ into two parts: $S^2/F^2$ and $N_s^2 = N_s^2 - S^2/F^2$. Obviously, $N_s^2 \geq -\Lambda_0$ and $\Lambda_0$ is defined in (5.2). The separation for $\Pi_w$ is $\Pi_w = \Pi_S + \Pi_\lambda$ where

$$\Pi_S = \begin{pmatrix} F^2 & S^2 \\ S^2 & S^2/F^2 \end{pmatrix}, \quad \Pi_\lambda = \begin{pmatrix} 0 & 0 \\ 0 & N_s^2 \end{pmatrix}.$$

Note that $\Pi_S$ is positive semidefinite and $\Pi = \Pi_d$ is positive definite in the dry region $\Omega - \Omega_0$, so

$$\Phi_0(\psi, N^2(\psi_0)) > \int_{\Omega_0} \{ |Dv|^2 + v \cdot \Pi v \} \, dx \, dz$$

$$= \int_{\Omega_0} \{ |Dv|^2 + v \cdot \Pi_S v + N_s^2 w^2 \} \, dx \, dz$$

$$\geq \int_{\Omega_0} \{ |Dv|^2 - \Lambda_0 w^2 \} \, dx \, dz$$

$$> \int_{\Omega_0} \{ (Dw)^2 - \Lambda_0 w^2 \} \, dx \, dz$$

$$= \sqrt{\nu_1 \nu_2} \int_{\Omega_0} \{ \nabla_0^2 w \}^2 - \Lambda_0 w^2 \} \, dx_0 \, dz_0, \quad (A.1)$$

where $\nabla_0^2$ is the Laplacian in the viscous-normalized space $(x_0, z_0) = (x/\sqrt{\nu_1}, z/\sqrt{\nu_2})$. By rotating the coordinates $(x_0, z_0)$ to $(x'_0, z'_0)$ such that the $z'_0$-coordinate is along the direction of the minimum diameter $d_2$ of the convex hull of $\Omega_0$ and using the property that $w = 0$ on $\partial \Omega_0$ (i.e., the boundary of $\Omega_0$), we may follow the same steps in (4.9a)–(4.9d) and obtain

$$\int_{\Omega_0} (\nabla_0^2 w)^2 \, dx_0 \, dz_0 = \int_{\Omega_0} (\nabla_0^2 w)^2 \, dx_0 \, dz'_0$$

$$\geq \tilde{E}^2 \int_{\Omega_0} w^2 \, dx_0 \, dz_0 = \tilde{E}^2 \int_{\Omega_0} w^2 \, dx_0 \, dz_0, \quad (A.2)$$

where $\nabla_0^2$ is the Laplacian in $(x'_0, z'_0)$ and $\tilde{E}^2$ is defined.
in (5.2). Substituting (A.2) into (A.1), we see immediately that \( \Phi_0(\psi, N^2(\psi_0)) \) is positive definite under the condition (5.2). QED.

REFERENCES


