Secondary frontal waves in the North Atlantic region: A dynamical perspective of current ideas

By D. J. PARKER*

University of Reading, UK

(Received 22 July 1996; revised 10 April 1997)

SUMMARY

The current understanding of the dynamics of secondary cyclogenesis on frontal systems in the North Atlantic region is reviewed. These secondary cyclones are often poorly forecast, and in extreme cases can have damaging consequences, but because of their small scales and potential for rapid growth detailed observations and analyses are limited: in this paper recent work is described and compared. It is argued that, on the synoptic scale, the development of 'primary' baroclinic waves is well understood, but on the subscales of these systems, where the secondaries develop, there is a wider variety of mechanisms which may cause or modify cyclone growth. This variety may be reflected in a wider range of resulting systems.

Various dynamical models have been postulated: secondary frontal waves are seen as the instability of a low-level potential-vorticity strip or warm band at a front, with possible finite amplitude triggering by upper-level features. Latent heating is thought to provide a suitable unstable potential-vorticity strip but the degree to which the latent heating contributes to the instability itself is a matter of dissent. Secondary systems are often shallow, so boundary-layer processes are important, but the influence of these is not well understood: laboratory experiments have suggested that the boundary layer will act to suppress instability. In general, frontogenetic strain flow, while acting to intensify the basic-state front, suppresses wave development—in contrast, recent work suggests that a frontolitic strain flow may be important in enhancing the development of one class of waves.

The FASTEX experiment, an international field study of North Atlantic secondary cyclogenesis, has recently attempted to observe these systems in detail. It is hoped that the data which have been obtained will help to resolve some of the outstanding issues in this field.

KEYWORDS: Cyclone Front Frontal wave

1. INTRODUCTION

Although interest in frontal waves, as instabilities of midlatitude frontal zones, goes back to the times of the Bergen school, there has been a resurgence of work since the late 1980s. Despite understanding of midlatitude frontal cyclogenesis development having originated with the study of North Atlantic systems, these cyclones are not yet fully understood on all scales, and some of the issues raised by the early researchers (e.g. Bjerknes and Sollberg 1922) have yet to be resolved. Recent severe storms to have affected the British Isles, such as the ‘October Storm’ of 1987 (see the special issue of Weather, Vol. 43, No. 3, for a description of this storm and its effects), the ‘Burns Day Storm’ of 1990 (McCallum 1990) and the ‘Braer Storm’ of January 1993, have reaffirmed the importance of frontal-wave development.

The idea of secondary frontal waves stems from a duality in the understanding of frontal instability, which again dates to the times of the Bergen school. In midlatitudes, cyclonic systems develop as instabilities of the meridional temperature gradient, or 'polar front', as typified by the Charney and Eady models of baroclinic instability. Within these cyclones, or baroclinic waves, the action of the shearing and deformation of the large-scale state is to form locally intense fronts which will reach small length-scales in finite time (as shown by Hoskins and Bretherton (1972); also Hoskins and West (1979) and Davies et al. (1991)). These intense fronts associated with the developed baroclinic waves may themselves undergo instability, forming secondary cyclones, for example, on the trailing cold front (Fig. 1). The secondary systems occur on the mesoscale (with typical wavelengths of order 1000 km), can have extremely large growth rates (developing over one or two days) and are notoriously hard to forecast; a significant example being the

* Present affiliation: The Environment Centre, University of Leeds, Leeds, LS2 9JT, UK.
'October Storm' of 1987. The recent studies of frontal waves focus on these secondary instabilities.

Bjerknes and Solberg (1922) described a model of cyclone families, in which successive frontal cyclones develop on the 'polar front': each cyclone thus resembles a secondary development on the trailing cold front of its forerunner (Fig. 2). However, implicit in this early model is the idea that these cyclones share the same dynamics and belong to the same class of systems. Subsequent ideas have suggested that the dynamics of the secondaries may be markedly different from those of their 'parent' waves. In particular, it is now thought that a wide range of dynamical ingredients can make up the secondary waves, and that these waves are not necessarily developmental, perhaps 50% of identified waves failing to deepen their pressure significantly.

Understanding of primary cyclogenesis may be said to be comprehensive. Waves develop as a result of baroclinic instability, deriving their energy from the potential energy of the larger-scale state of the atmosphere, either as a wave train ('type A') or in a transient manner ('type B'; Petterssen and Smebye 1971). The essential understanding of baroclinic instability stems from the two-dimensional Charney (1947) and Eady (1949) models, which may be interpreted as instabilities resulting from the interaction between Rossby wave trains (Hoskins et al. 1985). Modification of these simple models to include physical processes such as latent-heat release (e.g. Emanuel et al. 1987) and surface friction have allowed the role of these processes in baroclinic instability to be assessed. Modern numerical techniques have enabled high-resolution 'lifecycle' simulations to be performed (Simmons and Hoskins 1978; Thornicroft et al. 1993): the agreement between detailed observations, lifecycles and idealized models means that the process of primary cyclogenesis is well documented.

On the smaller scales at which the secondary frontal waves exist, the dynamics are less simple. Theoretical models have shown (as will be discussed in section 3) that various other processes are important to wave growth on the mesoscale: these include the shear at the frontal zone (contributing a barotropic component to the instability), the large-scale strain field in the environment of the front, effects of latent heating in clouds, boundary-layer processes, and the influence of a local strip of maximum potential vorticity. Theoretical work in these areas has gone a long way towards understanding the effects of these various
SECONDARY FRONTAL WAVES

Figure 2. Formation of a secondary cyclone as a wave on the cold front of the 'mother cyclone'. Implicit in this early model is the idea that the secondary is dynamically similar to the primary. (From Bjerknes and Solberg (1922).)

Processes, but in contrast with the study of baroclinic waves, the link with observations has yet to provide generality in our understanding. To rephrase this point, there is, as yet, no general statement of the importance and role of each of the various dynamical processes contributing to secondary frontal-wave development. It is likely that there is a broad spectrum of systems, with differing scales and structures, according to the relative importance of instability, triggering, moisture, mixing and background strain. Without a good observation set these problems cannot be addressed and this has been the goal of the Fronts and Atlantic Storm Track EXperiment (FASTEX) which took place in the first two months of 1997.

Detailed observations of cyclogenesis have recently taken place during ERICA (Experiment on Rapidly Intensifying Cyclones over the Atlantic) in 1988/1989, studying rapidly intensifying cyclones over the west Atlantic, close to the east coast of the USA. Although these systems develop rapidly on the narrow baroclinic zone between the Atlantic and the North American continent, with small horizontal scales, they are regarded as representing 'primary' cyclogenesis, as they occur at the entrance to the Atlantic storm track. In any case, the planetary-scale flow feeding these systems, with westerlies from the continent flowing over the warm Gulf Stream water, provides different conditions for cyclone growth from the mature fronts which approach northern Europe. Other mesoscale cyclones in the North Atlantic include 'polar lows' (e.g. Noldeng and Rasmussen 1992), which develop as cold polar air flows over high-latitude oceans and are often associated with ice boundaries. Again, these phenomena are here regarded to be different from the secondary frontal cyclones which commonly affect the British Isles and western Europe. A ‘typical’ secondary frontal wave observed during the third Intensive Observing Period (IOP 3) of the field experiment FRONTS 92 will be described in section 2.
Various practical issues are at stake in this research. The mechanisms by which intense or ‘explosive’ growth rates occur must be found and, perhaps more importantly for forecasting, the reasons why many systems are not developmental while a few have the potential to grow explosively must be isolated. Allied with this is a need to appreciate the structure of the mature frontal wave: most analyses tend to plot secondary developments as a scaled-down version of a primary frontal cyclone (e.g. Fig. 1) but this pattern need not be universally applicable. In the case of the primary frontal cyclone, knowledge of the nature of the features which generally accompany such systems is important to forecasting, and forecasters will need a similar understanding of the substructure of secondary systems in order to interpret model forecasts in terms of significant weather. If, as is likely, secondary frontal waves can exist on all scales, new concepts will be required.

This paper reviews, from a dynamical perspective, the current ideas regarding secondary frontal waves over the Atlantic. Section 2 discusses some recent observations and analyses, and will give a picture of the ‘typical’ frontal wave (although it must be accepted that there may be no universally applicable structure for all instabilities to a midlatitude front). Section 3 describes the range of theoretical models that have been used to explain frontal waves and, in particular, section 3(c) addresses the question of the role of moisture in wave development. The concluding section synthesizes these ideas.

2. RECENT OBSERVATIONAL STUDIES

Observations of secondary frontal waves have been hampered by the same factors which make them difficult to forecast—their small scales and their rapid growth rates—although retrospective studies of significant events such as the ‘October Storm’ of 1987 (Hoskins and Berrisford 1988; Shutts 1990; also Browning and Roberts 1994) have proved revealing. Despite such difficulties, recent study of the European Centre for Medium-Range Weather Forecasts (ECMWF) analyses by Ayrault et al. (1995; hereafter ALJL) and an observational programme carried out by the Joint Centre for Mesoscale Meteorology (JCMM), FRONTS 92, have greatly improved our understanding of the climatology and the structure of these phenomena. Here, the FRONTS 92 results will be used to illustrate typical features of North Atlantic frontal waves. The field experiment FASTEX is expected to yield a great deal more information about the process of secondary cyclogenesis (Thorpe and Shapiro 1996).

ALJL used ECMWF 1984–94 cold-season analyses for the North Atlantic region, at spectral resolution T106 for the early part of this period and T213 after September 1991. These resolutions imply that before 1991 the horizontal scale of resolved features was only around 400 km, but statistical analysis of the results suggests that there was no qualitative change in the behaviour of analysed frontal waves due to the change in resolution. This study started by identifying four weather regimes for the large-scale flow, within which higher frequency (12–36 hour) disturbances would be located by an objective method: these regimes concurred, in the main part, with regimes diagnosed elsewhere in the analysis of primary baroclinic waves. ALJL make the point that the high-frequency signal used to locate frontal waves may embrace many systems, such as polar lows or Mediterranean cyclones, and this regime-dependent analysis is intended to differentiate between the qualitatively different phenomena. Within the individual weather regimes, the high-frequency signal is maximized at the exit regions of the storm tracks, confirming that the high frequencies represent systems which coincide with the decay of primary baroclinic waves (objective filtering removes structures related to filamentation of the primary cyclones from the analysis). Also, this observation indicates that the rapidly growing frontal waves of the east Atlantic are of a different nature to the systems which have been studied in
the western Atlantic, off the east coast of the USA (such as those observed in ERICA). ALJL's study focusses on the 'zonal' weather regime, in which westerlies extend across the Atlantic and maximum high-frequency activity occurs in the region affecting France and the British Isles. In this regime, the two principal groups of high-frequency disturbances are labelled 'type 1' and 'type 2' and have distinct characteristics; although there are some problems inherent in the averaging involved with this work, these two types of secondary cyclone offer a generality to the study of frontal waves: ALJL describe them in detail.

Both types have a characteristic scale of 1000 km (Fig. 3). Type 1 cyclones exhibit a prominent cold front, which represents a local intensification of the parent frontal zone. In contrast, type 2 cyclones have a stronger warm front. Both types show a relatively weak surface-pressure signal of order 3 mb depression, and a temperature wave of amplitude 2.5 K: ALJL suggest that the clustering technique may be responsible for ameliorating the amplitude of the resulting structures, and the FRONTS 92 IOP 3 case, which in some ways resembles the type 2 system, has a much stronger temperature wave (see later). Both types exhibit anomalies in the relative-humidity field (Fig. 3): in type 1 cyclones this feature is evident well before the reference time at which the secondary system was objectively analysed, while in type 2 it is weak at -6 h and intensifies rapidly as it spreads over the warm front. The warm anomaly of type 1 is colocated with the relative humidity maximum while for type 2 the relative humidity at 700 mb is shifted ahead of the 850 mb warm maximum. In this way, the type 2 again shows a structure akin to that of a system dominated by a classical warm front.

Classical models of baroclinic instability (e.g. Eady 1949; Petterssen and Smebye 1971) involve an interaction between upper- and lower-tropospheric features, or precursors. Potential-vorticity analysis in conjunction with ALJL's studies shows high-frequency behaviour related to the objectively analysed cyclones, and it is shown that, despite having weak pressure signals, the systems do have an upper-level (300 mb) component on a similar scale. In type 1 cases the upper-level anomaly can be traced back 18 h but type 2 cyclones have little signal at -6 h, in low-level equivalent potential temperature (θe) or 300 mb potential vorticity. Once again, the lack of clear precursors may possibly be due to the averaging techniques used in ALJL's work, but the implication of these results is that the type 1 system involves an interaction between dynamical features that is commonly observed in primary cyclogenesis, while for type 2 systems other mechanisms may dominate the development.

Various recent studies have emphasized the role of the large-scale strain flow in the suppression of cyclone growth on fronts which may otherwise be dynamically unstable. In view of this, ALJL analysed the strain in the frontal environment of the type 1 and type 2 systems. In terms of the large-scale flow, type 1 cyclones appear to develop in an environmental deformation field that is weakly frontogenetic (Fig. 4). These systems form on a well-defined cold front that is reinforced by the secondary wave's cold front. In the case of type 2 cyclones, the background flow is frontolytic and the background baroclinicity is weaker than in type 1. This interesting observation pertains directly to the discussions of Bishop (1993a,b) and Bishop and Thorpe (1994a,b) on the role of deformation frontogenesis in suppressing frontal-wave growth: if, as is thought to be the case, frontogenesis suppresses frontal waves, does frontolysis enhance them? Recent results of Parker (1998) show that the action of frontolysis on barotropic instability is to allow the possibility of a short period of rapid wave growth, and this will be discussed in section 3(b).

To reiterate, the type 1 secondary cyclone, dominated by its cold front, is associated with frontogenetic deformation while the type 2, with a strong warm front, grows in a frontolytic environment. In contrast, Davies et al. (1991) showed that for baroclinic
Figure 3. The structures of the 'type 1' and 'type 2' cyclones identified by Ayrault et al. (1995). Solid contours denote 850 mb equivalent potential temperature (every 3 K), dashed contours are 1000 mb height, and regions of 700 mb relative humidity greater than 60% are shaded.
waves, the dominance of the cold or the warm front in the mature system was dependent on the strength and sign of the barotropic shear of the basic state. While the shear flow may be viewed kinematically as the sum of a rotation and a deformation flow, the link between these results is not clear, and may require a more precise study of the evolution of individual cases.

ALJL performed energy calculations showing that both types of cyclone have dominant baroclinic sources at lower levels. Type 1 also shows a positive barotropic contribution but type 2 cases lose energy to this term. Type 1 cyclones also gain energy from upper-level baroclinicity; this is consistent with the presence of clear upper-level precursors for type 1.

In summary, ALJL suggest that the type 1 case may be described by familiar concepts of wave development: the cyclone appears to occur as a mixed barotropic/baroclinic instability of a mature front, gaining energy both from the kinetic and the potential energy of the basic front, and involving the interaction with an upper-level feature. These kinds of mechanisms are to be discussed in section 3. In contrast, type 2 systems seem to be shallow and baroclinic, with little evidence of upper-level precursors, and develop on a relatively weak low-level front. The apparent lack of precursors offers few clues as to the mechanism of rapid evolution, but ALJL suggested that type 2 cyclones represent waves which are largely dominated by the background frontolytic deformation field. Although Parker (1998) confirms that frontolysis can indeed enhance growth in the simple case of purely barotropic instability of a vortex strip, the role of deformation in type 2 development requires further theoretical investigation.

ALJL’s work offers a generality to the study of frontal cyclones; a field where detailed observations and analyses of individual cases have been rare. FRONTS 92 was an observational programme which made use of the UK Meteorological Office (UKMO) Meteorological Research Flight C130 aircraft, employing dropsondes and airborne radar in conjunction with ground-based observing systems to observe secondary frontal waves. Coincident with the FRONTS 92 experiment, a subjective climatology of frontal waves in the North Atlantic sector was performed by T. D. Hewson (JCM) in the three-month period of March–May 1992. The results confirmed that these systems are relatively common, an
average of one wave per day being observed, with around 50% deepening in pressure to some degree. Cold-front waves were found to be the most common and col waves (developing in the ‘col’ between two high pressure systems, as in Fig. 1) the most developmental. Many features of the IOP 3 case (Hewson 1993; Browning et al. 1995) concur with the analysed ‘type 2’ disturbance of ALJL. IOP 3 was the most successfully observed FRONTS 92 case. The wave developed rapidly, south-west of the British Isles, growing by 12 mb in the period 1200 UTC 27 April to 0000 UTC 28 April 1992. It was during this period that intense observations were made by the aircraft. The horizontal scale of the system was around 500 km and it was one of a succession of disturbances to a trailing cold front over the Atlantic. The system could be traced back for over 2 days (Fig. 5), when genesis appeared to occur as a col wave, lying in the west Atlantic between high-pressure systems over the Azores and eastern Canada on 25 April 1992 (Fig. 5), in a region of very intense sea surface temperature (SST) gradients (around 4 K(100 km)−1). There is evidence from the UKMO Limited Area Model (LAM) output that the interaction with an upper-level trough caused the growth of the wave on 27/28 April as it approached the British Isles. Figure 6(a) (from Browning et al. (1995)) shows the Meteosat visible imagery for 1730 UTC 27 April 1992, indicating the frontal convection (running SW-NE) and the ‘cloud head’ which developed to the north of it, sketched in Fig. 6(b). Qualitatively, these features are typical of secondary frontal-wave developments.
Significant features of this wave, as observed in the dropsonde data, include a pronounced warm front (as in ALIL's type 2) and two significant cold fronts, as discussed by Browning et al. (1995) (Fig. 6(c)), large vertical wind shears outside the warm sector and a temperature range of order 10 K through the wave. As with ALIL's type 2 cyclones, the relative-humidity anomaly of IOP 3 seemed to lie ahead of the warm sector. A rainband, which persisted for at least 15 h, lay parallel to the warm front and ahead of it, and seemed to be riding up the warm-frontal surface.

The warm-frontal structure as derived from the dropsonde data is shown in Fig. 7 (from Browning et al. (1995)). The rainband was associated with a ‘kink’ in the frontal surface, where \( \theta_e \) was constant, or even decreasing, with height and this may have been associated with upper front CF1 overrunning from the west. Similar kinks were observed by Locatelli et al. (1994) for shorter-lived cold-frontal rainbands. Since it is reasonable to hope that IOP 3 is one of a generic type of frontal instability (such as ALIL’s type 2), the environment of this warm-frontal rainband may have a general occurrence—the interaction between this type of kink on the frontal zone and the rainband deserves attention. Locatelli et al. (1994) ascribe the rainbands of their case to forcing by a trapped gravity wave between the frontal surface and the ground. The existence, nature and role of gravity waves at synoptic fronts are not entirely clear at present: numerical studies such as those of Snyder et al. (1993) have shown, with parametrized diffusion, that gravity waves may produce a low-level maximum in upward velocity ahead of a front, which is stationary with respect to the front. The effect of latent-heat sources and sinks on these waves is not explored but might be expected to reinforce their structure: in particular, evaporative cooling of precipitation may interact to give a positive feedback on the intense and narrow downdraughts found beneath the frontal kink. However, other self-consistent arguments could be invoked to explain the coupling between the frontal kink and the rainband: in IOP 3 the rainband persisted for at least 15 h and lay above a maximum in the along-front wind (‘cold conveyor belt’) associated with warm advection; it is likely that balanced processes had a bearing on the propagation over this time-scale. Parker and Thorpe (1995) show how a precipitation band in a geostrophically balanced model tends to lie above the warm low-level jet to the flank of the low-level potential-vorticity maximum in a frontal system. In this model, the frontal kink would be an effect of the diabatic heating and cooling through which \( \theta_e \) becomes constant with height. A feedback on the jet at low levels then occurs through the local intensification of the horizontal temperature gradient, giving a greater vertical shear above the jet through thermal-wind balance. Parker and Thorpe (1995) describe a propagation mechanism for such a feature: a drawback in using the trapped gravity-wave model to describe the IOP 3 warm-frontal rainband is that in contrast with the observations of Locatelli et al. (1994), the IOP 3 rainband propagates relative to the front. It might further be possible to relate this rainband to conditional symmetric instability (CSI; Bennetts and Hoskins 1979), which occurs when the equivalent potential vorticity becomes negative, but Browning et al. (1995) point out that accurate diagnosis of the equivalent potential vorticity is not simple, even from data of relatively high resolution. It is difficult to be specific about the origins of the warm-frontal rainband of IOP 3 and, at present, its persistence and interaction with the frontal surface are not explained.

Browning et al. (1995) studied the multiple cold fronts of IOP 3 in greater detail (Fig. 6). The leading cold front (CF1) was more strongly reflected in the Meteosat infrared (IR) imagery, the second cold front (CF2) exhibiting shallow convection up to about 1.5 km. CF1 and CF2 propagated more rapidly than the system: as CF1 overran the warm sector its surface feature weakened, probably due to boundary-layer mixing processes, to leave CF1 as an upper cold front, which subsequently rode over the warm-frontal zone. Both CF1 and CF2 were associated with marked dry intrusions, visible in the water-vapour imagery (not
Figure 6. Detailed observations taken from IOP 3 of FRONTS 92, around 1800 UTC on 27 April 1992. (a) Meteosat visible image for 1730 UTC, showing the well defined clearance of cloud immediately behind cold front 1 (CF1). The inner box indicates the area covered in (b) and (c). (b) The surface and cloud analysis for 1800 UTC (pressure at 2 mb intervals). (c) Distribution of wet-bulb potential temperature, $\theta_w$, at 950 mb for 1800 UTC, derived from a subjective analysis of the dropsondes. Isopleths are labelled with large numbers (°C). The small numbers show the $\theta_w$ values for individual soundings displaced according to the system velocity. The tip of the warm sector/low centre was determined by a ship observation. (From Browning et al. (1995).)
shown) and the dropsonde data (Fig. 8). There was clearly a rich structure on the subscales of this system.

Overall, IOP 3 of FRONTS 92 may be regarded as illustrating a ‘typical’ frontal wave for this geographical region. However, whether the substructure, in the form of the fronts and upper-level features, has a general relevance will only be ascertained from study of other cases. Browning and Golding (1995) showed how a dry intrusion in a vigorous cyclone over the British Isles was involved in the generation of a tornadic squall line, and indicated that the high-potential-vorticity air originated in the stratosphere rather than being formed at low levels by latent-heat release. Again, aspects of these observations need greater theoretical understanding: the origin of the dry intrusions and the resulting banded structure which forms as they overrun the warm sector are not yet fully explained, but seem to be critical in determining the weather and frontal substructure associated with frontal cyclones.

ERICA took high-resolution observations of primary cyclogenesis off the east coast of North America, during the period December 1988 to February 1989. Although this study pertained to primary cyclogenesis, occurring where cold continental air meets the warm boundary-layer air over the Gulf Stream, the small-scale features observed within the cyclones are worth discussing at this stage. Neiman et al. (1993) describe remarkably detailed observations of the mesoscale structure of an explosively developing ERICA system (60 mb in 24 h). Their analysis shows frontal widths of 2 km and frontal vorticity of up to $20 \times 10^{-4} \text{s}^{-1}$ at the convectively active fronts, and the authors compare these features with thunderstorm outflows (or ‘cold pools’) which are often described as gravity currents. These fronts exhibit shallow (below 750 mb) smaller-scale vortices down to the 10 km horizontal scale, which are clearly recognizable in a signature of deep convection. Although the context of these systems is different from that of the cyclones of interest to
European forecasters, it does seem that secondary frontal cyclones can exist down to the smallest scales currently resolved by observations or analyses.

FASTEX was a major international field experiment (Thorpe and Shapiro 1996) in which several research aircraft, based at Shannon Airport, took observations of a number of frontal-wave cyclones in the east Atlantic. In conjunction with this operation, upstream flights took observations of the precursors of likely systems in the west Atlantic. Increased upper-air observations over the eastern seaboard of North America and periods of intense upper-air soundings over western Europe, in conjunction with specially stationed ships in mid-Atlantic, provided other high-resolution observations of the environment of the cyclones. It is hoped that the analysis of this data set will yield major advances in our knowledge of the development and substructure of frontal waves.

3. Theoretical models

(a) The front

It is inevitable that a prerequisite of understanding frontal waves is an understanding of the front itself. Study of midlatitude fronts remains an area of continuing research; high-resolution observations of several features within the north-east Atlantic region were made
Figure 8. A cross-section through the cold fronts of IOP 3 of FRONTS 92, with isolines of relative humidity indicating the multiple dry intrusions associated with the system. The surface cold fronts are labelled CF1 and CF2 and the positions of the dropsondes are indicated by the arrows on the top of the figure. (From Browning et al. (1995).)

during the field experiment FRONTS 87 (Thorpe and Clough 1991). Although traditionally associated with a temperature contrast, such fronts are often better defined by wet-bulb potential temperature, $\theta_w$, which is a better air-mass tracer in regions where latent heating occurs. Hewson (1997) discusses this point with regard to fronts in the UKMO LAM data. Typically, these fronts are also regions of horizontal wind shear: more significantly from a dynamical perspective this may be associated with a band of positive potential-vorticity anomaly at low levels. The reader is referred to Thorpe and Clough (1991) for more details of the structures deduced from FRONTS 87 observations.

It can reasonably be argued that there is as yet no ideal model of a mature front, and in discussing frontal-instability models it is important to keep in mind the limitations of the basic-state model being used. Semi-geostrophic theory describes the process of balanced frontogenesis with great success (Lagouvardos et al. (1993) showed this using FRONTS 87 data), but implicit in this kind of model is the move to an unstable frontal region, at which point turbulent mixing must be assumed to take place (Hoskins and Bretherton 1972). Some studies have addressed the stability of a steady semi-geostrophic deformation front for which frontogenesis has ceased, but it is not clear how this approach could be applied to fronts which develop in a baroclinic wave from shearing rather than deformation of the temperature field. The gravity-current model of a front is highly successful at describing smaller-scale intense fronts where mixing through Kelvin–Helmholtz instability is intrinsic to the system, but there is no simple extension of this theory to rotating, stratified fluids (the instability of rotating gravity currents in the laboratory will be described later). Early
dynamical descriptions of a front were based on the idea of 'air-masses', but observations do indicate that a front is in fact a transition zone rather than a discontinuity, and may propagate at a speed somewhat different from that of the local wind (see review by Smith and Reeder (1988)). These latter properties are better represented in semi-geostrophic models. However, studies based on two-layer frontal models (e.g. Orlanski 1968; Sinton and Heise 1993), because of their relative dynamical simplicity, have helped to illuminate the mechanisms and energetics of frontal instability.

(b) Dry instabilities

The gravity-current model of a front is useful at the smaller frontal scales and has been proposed as a mechanism of instability or 'frontal fracture' (Hobbs and Persson 1982). However, the scales of instabilities to gravity currents, which are 1–2 km in the atmosphere, and the importance of rotational effects at synoptic fronts means that the gravity-current model is unlikely to be useful for an explanation of the larger frontal waves discussed here (scales of 500 km, lifetimes of days). Snyder (1995) found Kelvin–Helmholtz instabilities, on scales of kilometres, to be the only short-wave features on an intense deformation front: Kelvin–Helmholtz instability is an intrinsic feature of gravity-current behaviour, which involves a dynamic equilibrium between the drag on the gravity current due to turbulent mixing, and the hydrostatic pressure gradient across the front (Benjamin 1968). The propagation of this kind of front is robust to the wind profile (Parker 1996), although when intense moist convection is possible, the interaction of the convection with the low-level gravity current may produce characteristic mesocyclones on scales of tens to hundreds of kilometres (e.g. Skamarock et al. (1994), and references therein).

To summarize, the instability of a front on the gravity-current scale consists of Kelvin–Helmholtz instability: the front coexists with this kind of instability, even in complex ambient flow (although convection may interact with a front to produce convective-scale cyclones). To understand the instability of mature atmospheric fronts, the effect of planetary rotation must be taken into account. Study of the limits of balanced theory (e.g. Snyder et al. 1993) have indicated that balance may remain a good tool even in the late stages of frontal development; since this review focusses on secondary cyclones, by definition forming on fronts which have developed in a large-scale balanced flow, the gravity-current perspective is not pursued at greater length here.

Griffiths and Linden (1981) described two-layer laboratory experiments in which buoyant fluid was released in the centre of a rotating tank of water, and used the instability of the resulting density interface, or front, which rapidly approaches geostrophic balance, as a model for geophysical frontal waves. They found a mixed barotropic/baroclinic instability, deriving energy both from the kinetic and the potential energy of the front. The instability was characterized by a relative depth scale $\delta_0 = h_0/H$, where $h_0$ and $H$ are the depth of the buoyant fluid and the total depth of the ambient fluid, respectively, and a Richardson number, effectively the square of the ratio of the Rossby radius of the flow to the initial radius of the buoyant fluid. In general, the waves grew principally through barotropic instability for large Richardson number and small $\delta_0$, that is, flows which are both shallow and have a horizontal scale which is small compared with the Rossby radius. For small Richardson number and large $\delta_0$, waves were baroclinically driven. In these experiments marked differences were found between flows at the free surface, where friction is small, and dense flows on the rigid lower boundary of the tank. It seems that the presence of friction at the lower boundary stabilizes some modes. Since secondary frontal waves are often shallow features, this suggests that boundary-layer influences are likely to be critical to the behaviour of frontal waves in the atmosphere, especially when they reach land. The
influence of the boundary layer is one which has been neglected in this field, and should be addressed more fully.

Orlanski (1968) studied the stability of a front separating air masses of differing densities, as a function of the Rossby number and Richardson number, respectively,

\[ Ro = \frac{U_2 - U_1}{2f} \]
\[ Ri = \frac{gH(\rho_1 - \rho_2)}{\bar{\rho}(U_2 - U_1)^2} \]

with \( U_1, U_2 \) and \( \rho_1, \rho_2 \) the velocities and densities of the two layers, \( H \) the total channel depth, \( f \) the Coriolis parameter, \( \bar{\rho} \) a mean density and \( k \) the along-front wave number. \( R_i \) is essentially a measure of the ratio of the potential energy of the basic state to its kinetic energy. Although this is a highly idealized basic state, with the frontal structure contained in a singularity, so that it cannot admit, for example, frontal propagation, nor give a description of the potential-vorticity evolution in a cyclone, it allows an appreciation of some essential mechanisms of frontal instability. For weak buoyancy contrast between the air-masses, the basic-state front is almost vertical, and \( R_i \) is small. In this regime waves exhibit flow which is quasi-horizontal and essentially represents shear instability. At intermediate \( R_i \) and high \( R_o \), the instability resembles Kelvin–Helmholtz waves, while for lower \( R_o \) a barotropic instability is possible. Of particular interest in the study of secondary cyclones is the fact that at higher \( R_i > 2 \), baroclinic instabilities are possible at intermediate along-front wave numbers and values of \( R_o \) which are not small. Orlanski’s work unified the results of previous studies and, apart from classifying the possible wave modes in terms of their energetics, made the important point that in this model there are unstable modes possible at all values of the along-front wave number.

Moore and Peltier (1987) published a description of the first recent attempts to understand secondary development as the instability of a semi-geostrophic frontal zone. In essence, they examined the stability of a balanced deformation front (as constructed by Hoskins and Bretherton (1972), but for which deformation has ceased) to three-dimensional perturbations, using the primitive equations. The solutions showed a principal unstable mode of typical wavelength 1000 km, trapped near the lower boundary and deriving its energy in the most part from the baroclinicity of the basic state. As noted by Schär and Davies (1990), these modes are not balanced, in that they could not derive from the balanced equations: Moore and Peltier’s (1987) modes must be nonbalanced because, by being concentrated close to one boundary, they violate the Charney–Stern theorem. The idea of 1000 km scale waves which derive solely from nonbalanced dynamics has remained a source of unease regarding these results in the research community. Notably, recent work by Snyder (1995), which examined the stability of steady fronts with uniform potential vorticity, identical to those of Moore and Peltier (1987), by applying a white-noise initial perturbation in a primitive-equation numerical model, found that the only nonbalanced modes which resulted were Kelvin–Helmholtz instabilities, on a scale of order 1 km: Moore and Peltier’s (1987) ‘cyclone-scale’ modes were not seen. At this stage it should be recalled that the basic-state front used by Moore and Peltier (1987) and Snyder (1995) is an idealized, steady one, in which there are no potential-vorticity gradients, and dry dynamics are assumed for the instability: the absence of instabilities at intermediate wavelengths in Snyder’s study does not preclude them from the general problem.

Although it is now widely realized that normal-mode analysis can fail to represent even the small amplitude fastest-growing structures, calculation of normal modes, as time-
independent structures, is still a revealing exercise. Schär and Davies (1990) and Joly and Thorpe (1990) both presented work which developed normal-mode instabilities of idealized frontal regions, characteristic of those embedded in primary baroclinic waves.

For dry, normal-mode instability of balanced (such as quasi-geostrophic) systems, the Charney–Stern theorem states that it is a necessary condition that the basic state potential-vorticity gradient and the equivalent potential-vorticity gradients which represent the boundary potential-temperature gradients (Bretherton 1966) have regions of opposite sign. Schär and Davies (1990) and Joly and Thorpe (1990) allowed shallow, balanced instabilities in a frontal model by introducing such potential-vorticity-gradient sign changes into their models by assuming, respectively, a surface warm band ahead of the cold front and a strip of low-level positive potential-vorticity anomaly due to diabatic heating (such potential-vorticity strips in baroclinic lifecycles appear as a result of numerical diffusion). It can be argued that, from the Bretherton analogy, these models have dynamical equivalence. At this stage it should perhaps be observed that in flows where moist processes are significant, the Charney–Stern theorem need not be relevant. Snyder and Lindzen (1991) showed how the potential-vorticity anomalies induced by diabatic heating in an infinite fluid (not containing potential-vorticity gradient reversals) could act in a similar manner to anomalies induced by advection in a dry flow. Joly and Thorpe (1991) illustrated in another simple model, with three-dimensional instabilities to a front with no potential-vorticity anomaly but retaining diabatic heating terms, how the Charney–Stern theorem may again become redundant. There are good reasons why latent heating will become progressively more important as the horizontal scale of systems decreases, and this will be discussed in section 3(c).

Schär and Davies (1990) showed how their warm band undergoes a mixed barotropic/baroclinic instability. A narrower warm band produces faster growth of the instability at a shorter wavelength. Similarly, a more pronounced temperature maximum produces faster growth. The instability is restricted to low levels. In the nonlinear regime, initializing with the normal mode, saturation occurs, leading to reduction in the growth rate from exponential. In other words, although the small-amplitude normal mode can grow exponentially, when nonlinear processes become important the growth rate falls: this fall in the growth rate is associated with the warm band breaking up into isolated warm anomalies, while the cold front develops a wave-like cold front/warm front structure (Fig. 9). The doubling time of the instability is around 1.5 days. Schär and Davies (1990) suggested that the saturation of their modes in the nonlinear regime is an explanation of the failure of many observed features to develop deep pressure features, and they commented that in practice, an upper-level trigger may be required in order to produce significant development.

Joly and Thorpe (1990) also used a dry semi-geostrophic model to demonstrate the instability of a positive low-level potential-vorticity anomaly. They showed how the available energy for the instability depends on the shape of the anomaly, with a narrower potential-vorticity strip having more available kinetic energy, suggesting that the barotropic conversion will dominate an instability. As in Schär and Davies (1990), a narrow anomaly produces shorter, faster-growing waves, dominated in this case by the barotropic energy conversion. A vertically deeper anomaly likewise produces faster growth. Joly and Thorpe (1990) also used a moist-modified Eady wave, as computed by Emanuel et al. (1987), as a steady basic state for the dry semi-geostrophic instability of the low-level potential vorticity. It is interesting to note that the modes found by Joly and Thorpe (1990) show vertical velocity and potential-temperature perturbations to be out of phase at low levels, as seems to be the case in ALIII’s type 2 (but not type 1) cyclone and IOP 3 of Fronts 92.

Subsequent work by Malardel et al. (1993) extended Joly and Thorpe’s (1990) results by running nonlinear simulations in a primitive-equation model. As in Schär and Davies
Figure 9. An example of the nonlinear evolution obtained in the numerical simulations of Schar and Davies's (1990) warm band frontal instability, at Rossby number 0.95. (a) Evolution of the surface temperature field: negative isotherms are dashed and the prefrontal warm band is stippled. (b) Surface pressure (contoured): the stippled region indicates areas where the vorticity exceeds the Coriolis parameter.

(1990) they found saturation of the modes. Again, they noted that the growth is restricted so that the instability is often insufficient to produce intense pressure changes in the developing lows, contrary to surface observations of many such systems. In terms of energetics, the barotropic conversion is initially dominant but subsequently it decreases and the baroclinic conversion dominates; at this stage the barotropic conversion becomes negative and is a sink of energy and the cause of equilibration. Only the waves which go through this two-stage (barotropic then baroclinic) process deepen in pressure significantly. In order for this to occur, a suitable initial perturbation is required, vertically deep enough to extract energy from the ambient baroclinicity, and sufficient ambient baroclinicity must be present. This is broadly in agreement with Schar and Davies's (1990) ideas that since shallow modes saturate before they can develop significantly, a baroclinic development of relatively deep vertical extent must subsequently occur if the system is to intensify.

At this stage it is worth noting that, in an intuitive sense, a barotropic instability, which does not involve vertical motion, would not be expected to produce deep pressure anomalies: without the vertical motion, there will be no vortex stretching to produce the intense vortices which could be seen in a baroclinic wave. Since low pressure is generally associated with high vorticity in geostrophically balanced systems, barotropic instability
is not likely to produce cyclones with deep central pressures. Indeed, some observed systems do not deepen their central pressure significantly over much of their evolution. This raises the question of how best to quantify the intensity of frontal waves; whether in terms of pressure anomaly, maximum wind speed, potential-vorticity anomaly or some other measure. In part, this returns to the problem of determining a typical structure for a frontal wave, in order to determine the features which best classify these phenomena.

An intriguing point that is made by Malardel et al. (1993) is the qualitative similarity between mature instabilities to baroclinic zones of different horizontal scales (Fig. 10). This idea might suggest that it is reasonable to consider secondary cyclones, in their fine structure, as smaller versions of their parent systems. However, it must be recalled that these simulations of Malardel et al. (1993) neglect moist dynamics and other physical processes which might be thought to have greater influence on smaller cyclones.

Simple consideration of the barotropic or baroclinic instability of a given front neglects the role of the larger-scale environment in influencing the development of growing waves. Dritschel et al. (1991) showed that for the case of two-dimensional vortex fila-
ment instability, the linear wave growth was suppressed by an ambient strain of magnitude greater than \( \frac{4}{1} \) of the vorticity of the vortex strip: this result was taken as an indication of why such instability may be uncommon in observations of such vortex filaments in two-dimensional turbulence. In the atmospheric context, Bishop (1993a) explored analytical, non-modal solutions (i.e., linear solutions which need not preserve their structure with time) to the development of instabilities on a time dependent basic state of a front undergoing deformation frontogenesis (Fig. 4), and these methods were used by Bishop and Thorpe (1994a,b) to investigate the influence of frontogenesis on the growth of frontal cyclones. Bishop and Thorpe (1994a) showed how frontogenetic deformation acting on the low-level vorticity strip at a front favours the growth of instabilities which are initially of short wavelength: the deformation field acts to increase the wavelength of the instability (towards that of maximum normal-mode growth) whilst increasing the potential growth rate by making the frontal vorticity strip narrower. However, Bishop and Thorpe (1994b) describe how a sufficiently large deformation flow will suppress nonlinear wave development, as the waves are literally compressed in the across-front direction.

The role of environmental deformation in suppressing or allowing frontal wave growth has been investigated for several analysed cases by Renfrew et al. (1997). For the cases documented, rapid amplification was coincident with decreasing, or small, frontogenetic environmental deformation of the frontal vorticity strip, whereas a strong deformation rate appeared to inhibit the growth of two frontal waves. Bishop (1993a) comments on the fact that frontolytic deformation can increase the growth rate of instabilities, and the processes involved in the nonlinear regime of a frontolytic flow are certainly pertinent to the understanding of ALJL's type 2 cyclones, which seem to develop in this kind of large-scale environment. Parker (1998) extended Dritschel et al.'s (1991) work to the case of negative, or frontolytic, deformation and showed how this allows a short period of very large wave growth for suitable initial perturbations. As in the frontogenetic case, the development is a combination of the kinematic distortion of the wave by the deformation flow, which amplifies the wave in frontolytic flow and suppresses it in frontogenetic flow, and the 'barotropic' instability of the vortex strip itself. Extension of Parker's (1998) results to more physically realistic fronts, admitting baroclinic instability, would be useful.

Bishop's work highlights the way in which the structure of a growing wave may change with time as it develops; it is now well established in the theory of cyclogenesis that 'non-modal' linear disturbances, which change their form as they grow, may experience much more rapid development than normal-mode structures (Farrell 1984). Joly (1995) explored these ideas in the context of secondary frontal waves and showed how transient growth rates could be three times greater than those of the normal modes. Another possible consequence of the role of deformation in modifying the structure of growing waves is that this implies the existence of a broader spectrum of relevant wavelengths than might be suggested from normal-mode analysis applied to steady fronts.

Thornicroft and Hoskins (1990) describe dry baroclinic wave lifecycle simulations, on the sphere, in which cyclogenesis was observed to occur on the trailing cold front of a primary baroclinic wave as a result of the interaction with upper-level cyclonic vorticity. This kind of finite-amplitude interaction is a way in which the instability mechanisms of normal-mode solutions act transiently to give explosive growth, or a manner in which slowly developing shallow waves may be boosted. The wave observed during IOP 3 of FRONTS 92 was one of a series of waves on a long cold front but the only one to achieve deep growth in pressure, and there is evidence that this was involved with an upper-level trough. Such finite-amplitude interactions may also be pertinent to the description of ALJL's type 1 cyclone, in which precursors to the development were evident at upper and lower levels for some time before the period of growth. Finite-amplitude studies have a further
advantage in that they can show the Lagrangian aspects of the flow, as small-amplitude modes may fail to do. This aspect is important in studying the origin of the elements of an instability (for example the warm conveyor belt or a stratospheric intrusion). However, as for non-modal studies, it can be argued that finite-amplitude models lack the generality of normal-mode studies and, while providing a more precise description of an instability, may fail to illuminate its dynamical processes.

Although upper- and lower-level interactions are a cornerstone in the understanding of primary baroclinic instabilities, the short wavelengths of many observed frontal waves mean that the Rossby scale height of disturbances, defined as,

$$H_R = \frac{fL}{N},$$

with $L$ a horizontal length-scale and $N$ the Brunt–Väisälä frequency, will be small. Thus, the height of influence of a given dynamical anomaly of, say, potential vorticity is small for short waves. Thinking in terms of canonical Eady wave dynamics, this kind of argument is an explanation why short waves may not grow from baroclinic instability: when $H_R$ is small the upper and lower Rossby wave trains cannot interact. However, other canonical forms of baroclinic instability exist, such as the Charney model, where there is no rigid lid but a gradient of potential vorticity (a ‘$\beta$-plane’), Green’s (1960) model which has a potential-vorticity gradient and rigid lid (allowing short-wave instability) or Snyder and Lindzen’s (1991) model where an equivalent potential-vorticity gradient is provided by diabatic sources; in addition, Joly and Thorpe (1990) show how a potential-vorticity strip which is vertically deep and narrow undergoes a principally barotropic instability, having quasi-horizontal motion and gaining energy from the kinetic energy of the basic state. Given the range of baroclinic, diabatic and barotropic instabilities possible, it is perhaps unwise to focus attention solely on upper/lower-level interactions as the source of rapid growth on the smaller scales, even from the perspective of transient development.

In summary, frontal instability over the North Atlantic has been seen as a growing disturbance on a low-level potential-vorticity strip, which is a consequence of the frontogenesis processes (in conjunction with latent heating and friction) which are inherent in primary baroclinic wave development. The secondary instability may be enhanced or triggered by transient interactions with upper levels. In general, systems seem to correspond to a mixed baroclinic/barotropic instability, feeding on both the potential energy and the kinetic energy of the basic state: for the simple semi-geostrophic deformation front, for which large-scale forcing becomes negligible, there seems to be no dry short-wave instability except for Kelvin–Helmholtz waves (Snyder 1995). The general energetics of a frontal-wave disturbance on a potential-vorticity strip depend on the cross-sectional aspect ratio of the strip (Joly and Thorpe 1990) and the wavelength and structure of fastest-growing normal modes depends on the width (narrower strip gives shorter, faster-growing waves). Deformation frontogenesis is seen to modify the development, as it deforms the developing disturbance and alters the width and stability of the basic front; frontogenetic flow tends to suppress waves but the role of frontal flow seems to be to allow a short period of very rapid growth. The general role of deformation is, through kinematically altering the wavelength of any disturbance, to suggest that waves may exist on a broader spectrum of wavelengths than would be suggested by normal mode or non-modal analysis: stability studies for simplified basic-state fronts show instability to be possible at all wave numbers (Orlanski 1968). Moisture in the instability has been seen as the source for the basic potential-vorticity strip, through the latent heating at the basic front, and many analyses have been of a dry disturbance to this potential-vorticity distribution. The structure of such dry disturbances, as incipient waves and through their ‘lifecycles’, has been documented,
and in the broadest sense these models are in concord with the limited observations of secondary frontal waves. The next subsection will describe what is understood about the role of moisture in the disturbances themselves.

(c) Moist instability

Malardel et al. (1993) state briefly that the inclusion of moist dynamics in a frontal-wave instability has little effect. In contrast, Ferris (1989) describes how frontal instability in a three-dimensional mesoscale model can be dominated by the convective latent heating, and both Hoskins and Berrisford (1988) and Shutts (1990) suggest that latent heating was a critical factor in the intense development of the ‘October Storm’ of 1987. Certainly, there are many outstanding aspects of the moist dynamics of secondary cyclogenesis which remain to be explored.

Satellite images of frontal waves show that moist convection is taking place (e.g. Moore and Peltier 1987). Polar lows (shallow intense cyclones that develop over high-latitude oceans) are thought to have analogous dynamics to tropical cyclones, dominated by latent heating (e.g. Nordeng and Rasmussen 1992) and though these are not specifically the same phenomena as the frontal waves of interest here, they exist on similar horizontal scales and it is likely that they share some features, or that there may be a spectrum of such cyclones, over which latent heating plays a differing role.

Starting from the perspective of the primary systems on which frontal waves grow, it is known that moist processes can significantly modify baroclinic instability (e.g. Emanuel et al. 1987; Craig and Cho 1988; Snyder and Lindzen 1991; Parker and Thorpe 1995), especially when included in a ‘conditional’ or ‘moist up, dry down’ manner, increasing growth rates and narrowing updraught scales. Although the growth of the dry modes of Malardel et al. (1993) saturates, and large pressure perturbations are not obtained, it would be reasonable to expect that suitable moist parametrizations may have a similar qualitative influence on this instability. Thorpe and Emanuel (1985) demonstrated that a purely moist analogue of a baroclinic Eady wave may be constructed, with $\theta_c$ taking the role of $\theta$. In this case, the horizontal wavelength of fastest growth scales as $\sqrt{q_c/q}$, where $q_c$ and $q$ are the equivalent and dry potential vorticities, respectively. In regions of convection, where $q_c$ is observed to be small, moist baroclinic instability tends to favour short horizontal scales. The implication here is that the short, secondary frontal waves of interest will be influenced by moist processes to a greater degree than the primary disturbances, because they project more strongly on to the growing ‘moist’ modes. Using similar arguments it is possible to construct a moist analogue of the Rossby scale height,

$$H_m = H_R \sqrt{\frac{q}{q_c}},$$

where $H_R$ is defined in (3). In cases of small $q_c$, where latent heating is strong, $H_m$ may become large, so that small-scale features may interact more strongly in the vertical than would be possible if the dynamics were dry. In this way, finite-amplitude triggering mechanisms such as suggested by Thorncroft and Hoskins (1990) may remain potent on scales well beyond the short-wave cut off for dry baroclinic instability. At the extremely short scales of frontal instability, where unbalanced dynamics dominate, deep convection can be of first importance in its interaction with gravity-current dynamics, as is known from the study of squall lines (e.g. Skamarock et al. 1994). As a general rule, then, it seems that the shorter the horizontal scale of a system, the greater the potential for interaction with moist dynamics.
In qualitatively assessing the role of moisture in a developing frontal wave it is important to have an understanding of the velocity and thermodynamic fields associated with a perturbation to the basic potential-vorticity strip, since this determines, interactively, the diabatic forcing of the instability. This interaction between the flow and the associated moist processes is a subject of much current discussion, which has perhaps spilled over from the theory of large-scale tropical disturbances (e.g. Emanuel et al. 1994). At mid-latitude fronts there is evidence (e.g. Hobbs and Persson 1982) that, as at squall lines, the finite-amplitude upward displacement at the low-level front acts to ‘trigger’ a band of intense convection. This is further justified by the diagnostic analyses of FRONTS 87 data performed by Lagouvardos et al. (1993) who show that a convective parametrization based on such assumptions is needed for a good representation of the cross-frontal flow. Contrary arguments by Thorpe and Emanuel (1985) would seem to indicate that the ascent region in the vicinity of a midlatitude front exists in a state close to moist slantwise neutrality, implying that parametrizations based on this assumption should be used. This controversy is not resolved, and in the context of frontal waves it is likely that the scale of the wave is important: features of the scale of cloud cells, 10 km say, may be ‘triggered’ phenomena (such as the rainbands of Locatelli et al. (1994)) but on synoptic scales moist equilibrium is likely to be more pertinent.

Since frontal waves exist at the mesoscale, the question of how moist processes interact with a wave must take into account the lack of general understanding of the basic couplings, and this should be borne in mind when generalizing results. It is insufficient to consider only one parametrization of the moist effects. At a further level, addressing the issue of forecasting of frontal waves, the position within the theoretical hierarchy of the moist parametrizations used in operational numerical weather-prediction (NWP) models must be addressed.

Joly and Thorpe (1991) performed a numerical study of normal modes on a time-evolving basic state (so that the spatial structure obtained is constant but not necessarily exponentially growing), in which moist processes were allowed to occur in both the basic-state front and the small-amplitude perturbation. To summarize their findings for waves on moist Eady-wave basic states (such as those of Emanuel et al. 1987), short-wave instability (along-front wavelength of around 1000 km) was possible for slowly evolving but intense fronts, in which case mean e-folding times over a 24-hour period could be less than half a day. Once again, as shown by Bishop and Thorpe (1994b) for deformation frontogenesis and observed by AlJL in the ECMWF analyses, it seems that strong frontogenesis, in this case shearing deformation, is detrimental to frontal-wave growth.

Joly and Thorpe (1991) found the existence of the short wave, three-dimensional modes to depend crucially on the presence of the diabatic heating. These modes were periodic in the along-front direction (unfortunately the spatial structures were not shown) and it was commented that asymmetric structures, with narrower updraught regions, would be likely to experience faster growth. Study of such structures and the mechanism of interaction between diabatic heating and the wave would be valuable.

Much work on the effect of moist processes on cyclogenesis has focussed on ‘primary’ instabilities, for example the intense synoptic-scale systems observed to form off the east coast of the USA. Davis et al. (1993) studied model simulations of three extratropical cyclones; one over the continental USA and two over the west Atlantic. They found that although the potential-vorticity anomalies attributable to the diabatic source may account for up to one third of the anomalous potential vorticity of the system (although there was much variation in the significance of latent heating between the cases), the fact that the anomalies are of relatively short horizontal scale means that they do not have a great vertical penetration, and hence have little influence on the baroclinic deepening. While this is a
result relating to primary development, for secondary frontal cyclones the point applies regardless of the presence of moist processes, because it is partly their short horizontal scale which defines these phenomena, as noted earlier.

Since the modelled cyclones of Davis et al. (1993) showed reasonable agreement with surface rainfall observations, there is some confidence that the integrated diabatic sources were well represented in these studies. However, the authors do discuss how different convection schemes, in modifying the vertical distribution of heating and consequently the height of the potential-vorticity anomalies, can provide noticeably different degrees of surface deepening.

Valuable insight into moist dynamics was gained from the results of the ERICA project, and some of the conclusions, which pertain to the general dynamics of frontal instability in a moist environment, are discussed here. It is hoped that the observations obtained in FASTEX will allow more specific conclusions to be made regarding the secondary systems which are the focus of this paper. Reed et al. (1993) show a difference due to a convective parametrization scheme in a sensitivity study on simulations of the ERICA IOP 5 storm. The principal results of this work are that the moist processes are crucial to the explosive deepening of the storm: the absence of moisture gives a storm which propagates more slowly (as follows from arguments of Parker and Thorpe (1995)) and is shallower, by 19 mb, than the control simulation (a ratio of deepening of 2.58 between control and dry simulations). However, as found by Davis et al. (1993), the surface temperature features were not significantly modified by the absence of moisture. Removal of the surface fluxes from the simulation produced a system which was likewise slower and was shallower (by 6 mb) than the control; this is linked to the reduction in latent heating. Removal of the surface fluxes also intensified the surface thermal pattern: the fluxes act to warm the cold air and may cool the warm flow (some downward heat fluxes are seen in the control simulation).

Reed et al. (1993) make an interesting comparison of various other studies of explosive cyclogenesis, calculating for each case the ratio of the control pressure drop and that computed without surface fluxes, $\frac{\Delta p \text{ (control)}}{\Delta p \text{ (no fluxes)}}$. This ratio can be negative or positive, for different storms, and Reed et al. (1993) ascribe this to the presence or absence of negative surface heat fluxes in advance of the cyclone. In the ERICA IOP 5 case, there are positive fluxes in advance of the system and these fluxes enhance the deepening (through subsequent latent-heat release and a maintenance of the thermal advection). Neiman and Shapiro (1993) attempted to quantify the surface fluxes in a rapidly deepening system off the east coast of the USA and found that the occurrence of negative fluxes to the east of the system coincided with the end of cyclogenesis. Again, the precise mechanism of this phenomenon is not explained (nor, at a more direct level, is the coupling with latent heating), but as the magnitude of the positive boundary-layer heating was diagnosed to reach orders of 5 K h$^{-1}$ (albeit in a system where winter continental air passes over warm Gulf Stream water) the phenomenon would seem to be extremely important. In other cases where negative fluxes occur ahead of the cyclone, they act to weaken the system.

Balasubramanian and Yau (1994) used a 2-layer model to simulate explosive moist cyclogenesis on a baroclinic zone of around 3000 km extent (i.e. with a view to understanding primary cyclogenesis), using a parametrization of slantwise convection. The model did not show great differences from a dry simulation until it reached a stage at which nonlinear advection produced a 'bent-back warm front'. At this time, a surge of cold advection occurred at the mid-level (in the 2-layer model). This cold advection corresponded to a fall in the upper-level height which produced an explosive deepening in the surface pressure. Balasubramanian and Yau (1994) note that it is the frontogenesis process which rapidly enhances the cold advection, giving an upper response: upper/lower-level interactions are
crucial to this model, as are interactions between the latent heating at the front and the cyclone. This is an example of the scale interactions induced by moist dynamics, as discussed by, for example, Joly and Thorpe (1989). Again, moist processes tend to reduce the horizontal scale of the fields of vertical motion and hence of the diabatically produced potential-vorticity fields. Inversion of these potential-vorticity fields, however, still tends to be dominated by the dry Rossby radius, so that narrow bands of convection can have a broad influence, as in Balasubramanian and Yau's (1994) model. This phenomenon of the interaction of large and small scales is a recurring feature of the study of secondary frontal waves, since they exist on a scale between that of the parent synoptic cyclone and its fronts.

The above studies were performed with a view to describing moist influences on primary, synoptic-scale cyclogenesis. Although much insight has been gained into these systems, there is still a lack of understanding of the interaction with latent heating in the mechanisms of rapid growth. This, it appears, can have a variable degree of influence from case to case. Dynamical arguments suggest that, as a general rule, as horizontal scales become shorter the potential for a disturbance to interact with moist dynamics increases: secondary frontal waves are likely to be more strongly influenced by intense latent heating, should it occur. Since these processes are currently poorly understood, this remains an important avenue of future research. In particular, it may be hoped that the FASTEX data will allow a more precise picture of the role of moist processes in secondary cyclogenesis to be obtained.

4. Conclusions and Outlook

The earliest ideas of frontal-wave development regarded secondary systems within cyclone families to be incipient versions of the primary systems. For example, Bjerknes and Solberg (1922) showed families of frontal cyclones in which successive systems appeared on the trailing cold front of their predecessor. Since then, studies have indicated that the secondary systems are qualitatively different from the 'parent' cyclones. While the primary waves represent baroclinic instability on the synoptic scale, the secondaries may occur as a result of a wider range of growth mechanisms, including modulation by strain, moisture and friction and, it seems, may be manifest in a broader range of physical types.

Theoretical work has increased in recent years. Idealized frontal-instability models of Joly and Thorpe (1990) and Schär and Davies (1990) have exposed the simplest mechanisms of secondary cyclone growth. While instability occurs on a baroclinic zone (as for the primary systems but on shorter horizontal scale), the waves may extract energy not only from the baroclinicity but also through barotropic instability (taking energy from the wind shear across a front). In the absence of upper-level interaction, 'dry' low-level waves require a potential-vorticity strip, or equivalent warm band, at a front in order to be unstable. Such features may arise in nature as a result of latent heating at the front. Transient, non-modal waves, or waves with strong moist dynamics, may grow on more general frontal states. Thornicroft and Hoskins (1990) described a finite-amplitude mechanism for cyclone growth, whereby, as an integral part of the parent system's evolution, an upper-level potential-vorticity feature advects over the surface front and triggers secondary development. This mechanism links directly to the cyclogenesis mechanism of primary baroclinic instability and conceptually describes the mechanism of growth as being that of a mesoscale baroclinic wave, with upper/lower-level interaction. It seems that this kind of mechanism is important in practice: whether it dominates over 'instability' (meaning growth from arbitrarily small disturbances) is not yet clear.
Saturation of shallow modes in the nonlinear regime has been proposed as an explanation for the observed lack of development in many incipient waves. It has been suggested that only those waves which interact baroclinically with an upper-level feature reach large amplitude. Ambient strain is also proposed as a limiting factor on the growth of systems: frontogenetic deformation and shear, while acting to make the front more intense and therefore potentially more unstable, tend to inhibit physically the roll-up of cyclones by compressing them in the cross-front direction. However, in the arguments relating to the rate at which secondary waves evolve, it is not entirely clear if there is a well-defined measure of the development: some ‘intense’ storms may be observed to undergo little deepening of surface pressure.

Although much work has gone into the study of moist and turbulent diffusive processes in west Atlantic primary cyclones, revealing the crucial importance of these in explosive growth, the detailed interactions with the dynamics often remain opaque. Such work has not yet been undertaken for the frontal waves of the east Atlantic. Limited studies of recent frontal-wave systems have indicated variety in their structure and evolution (e.g. ALJL’s type 1 and type 2) and the fact that this review openly excludes discussion of related systems such as polar lows is perhaps a confirmation that, at the scales of secondary frontal waves, there is a wide range of physical processes that are of first order in the dynamics. It is to be hoped that the data which have been obtained during FASTEX will enable us to pin down more precisely, over a wider sample of systems, the dominant mechanisms responsible for the frontal waves of interest.

Finally, perhaps the overriding question remaining in the study of secondary frontal waves is the one of scale. It is not clear whether frontal waves exist uniformly throughout the spectrum of scales, nor is it entirely clear how the dynamics and structure of a wave will vary with scale. At the synoptic scale, the behaviour of the primary waves is well documented: these systems are baroclinic instabilities of the midlatitude meridional temperature and potential-vorticity gradient, somewhat modified by latent heating and friction, and their substructure is familiar (e.g. Davies et al. 1991; Thorncroft et al. 1993). At the other extreme, of small, kilometre scales, fronts undergo Kelvin–Helmholtz instability, in many ways behaving like gravity currents and, if moist convection is intense, develop substructures linked to the convection. Between these two extremes of horizontal scale lie the frontal waves under consideration here: these are thought to involve various physical processes in conjunction. Typically, forecasters plot such waves as small examples of the primary, ‘Norwegian’, cyclones: studies such as that of Browning et al. (1995) have shown that they are likely to be more complex, and as yet no generic structure is known—it is not clear whether such a generic structure exists. Perhaps these ideas may be summarized by the two linked questions:

- Is there a strong scale selection for secondary waves?
- What is the substructure of secondary waves?

To answer the first of these requires a greater dynamical understanding of the mechanisms behind wave growth, and observational grounding for this. If there are dominant mechanisms producing secondary development, then simple models for the structure of systems may be obtained. If, however, the spectrum of possible instabilities is broad, dependent on the relative importance of baroclinicity, horizontal shear, moist dynamics, large-scale strain and boundary-layer dynamics to differing degrees in different systems, we may have to acknowledge a spectrum of substructures. Again, it is to be hoped that the FASTEX data will enable us to make progress in this work. The coming years may be a productive time in the study of secondary frontal cyclones.
ACKNOWLEDGEMENTS

I would like to thank Alan Thorpe, Tim Hewson, Franck Ayrault and Ian Renfrew for their many interesting comments on this review. During this work I have been supported by Natural Environment Research Council grant GR3/9036. Figure 9 is reproduced by kind permission of the American Meteorological Society.

REFERENCES


Charney, J. G. 1947 The dynamics of long waves in a baroclinic westerly current. Q. Meteorol., 4, 135–163


Eady, E. T. 1949 Long waves and cyclone waves. Tellus, 1, 33–52


