Chapter 16

Extratropical Cyclones: A Century of Research on Meteorology’s Centerpiece

DAVID M. SCHULTZ, a LANCE F. BOSART, b BRIAN A. COLLE, c HUW C. DAVIES, d
CHRISTOPHER DEARDEN, e DANIEL KEYSER, b OLIVIA MARTIUS, f PAUL J. ROEBBER, g
W. JAMES STEENBURGH, h HANS VOLKERT, i ANDREW C. WINTERS b

a Centre for Atmospheric Science, School of Earth and Environmental Sciences, University of Manchester, Manchester, United Kingdom
b Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, New York
c School of Marine and Atmospheric Sciences, Stony Brook University, State University of New York, Stony Brook, New York
d Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
e Centre of Excellence for Modelling the Atmosphere and Climate, School of Earth and Environment, University of Leeds, Leeds, United Kingdom
f Oeschger Centre for Climate Change Research, Institute of Geography, University of Bern, Bern, Switzerland
g Atmospheric Science Group, Department of Mathematical Sciences, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin
b Department of Atmospheric Sciences, University of Utah, Salt Lake City, Utah
i Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

ABSTRACT

The year 1919 was important in meteorology, not only because it was the year that the American Meteorological Society was founded, but also for two other reasons. One of the foundational papers in extratropical cyclone structure by Jakob Bjerknes was published in 1919, leading to what is now known as the Norwegian cyclone model. Also that year, a series of meetings was held that led to the formation of organizations that promoted the international collaboration and scientific exchange required for extratropical cyclone research, which by necessity involves spatial scales spanning national borders. This chapter describes the history of scientific inquiry into the structure, evolution, and dynamics of extratropical cyclones, their constituent fronts, and their attendant jet streams and storm tracks. We refer to these phenomena collectively as the centerpiece of meteorology because of their central role in fostering meteorological research during this century. This extremely productive period in extratropical cyclone research has been possible because of 1) the need to address practical challenges of poor forecasts that had large socioeconomic consequences, 2) the intermingling of theory, observations, and diagnosis (including dynamical modeling) to provide improved physical understanding and conceptual models, and 3) strong international cooperation.

1. The continua of the atmosphere and history

The atmosphere and history can both be viewed from a common perspective. Both are continua with a multitude of processes acting simultaneously and at a variety of time and space scales. To make sense of either the atmosphere or history, we humans have the habit of defining categories to provide focus—be they atmospheric scales, physical processes, theory, and observations, or historically defined separations between epochs (e.g., discovery of America, First World War, Treaty of Versailles, end of Second World War, and atomic era).

Within this atmospheric continuum, we focus on extratropical cyclones, low pressure systems that are frequently

Corresponding author: Prof. David M. Schultz, davis.schultz@ manchester.ac.uk

DOI: 10.1175/AMSMONOGRAPHS-D-18-0015.1
© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
born of and evolve with the jet stream, producing in some midlatitude locations as much as 85%–90% of the annual precipitation (Hawcroft et al. 2012) and as many as 80% of extreme precipitation events (Pfahl and Werml 2012). Although extratropical anticyclones are the counterpart to extratropical cyclones, for the purposes of this chapter, we focus only on the cyclonic sibling.

Within this historical continuum, our focus for this chapter is nominally 1919–2018. In addition to the founding of the American Meteorological Society (AMS), 1919 was important to this chapter for two other reasons. The first reason was the publication of the first widely accepted conceptual model for the structure of the extratropical cyclone by the Bergen School of Meteorology in Norway (Bjerknes 1919). Understanding extratropical cyclones—their dynamics, structure, and evolution—was the big advance that came from the Bergen School meteorologists, which makes this chapter extra pertinent to the AMS 100th anniversary. The energy and enthusiasm coming from the Bergen School was ignited by the leadership of Vilhelm Bjerknes and his colleagues in Norway following World War I, constituting a dramatic paradigm shift within the meteorological community and providing the foundation for the rise of modern synoptic meteorology (e.g., Friedman 1989, 1999; Jewell 1981, 2017). For synoptic meteorology, the development of what we now call the Norwegian cyclone model and accompanying polar-front theory proposed by Bjerknes (1919), and further developed in Bjerknes and Solberg (1921, 1922) and Bjerknes (1930), provided a common framework and language by which researchers and forecasters could communicate. Although this model had its roots in earlier research by Vilhelm Bjerknes and German scientists (e.g., Volkert 1999), it was its blending of theoretical and practical research, as well as its focus on operational forecasting, that made it so influential. Much of the terminology introduced in the cyclone model is still in use today (e.g., cold front, warm sector, occlusions, polar front), and, as we will see later, some ideas that were introduced at that time were lost and rediscovered (e.g., seclusion, bent-back front). Later, applying physical principles to polar-front theory allowed quantitative analysis and testing of the mechanisms for cyclogenesis, culminating in the discovery of baroclinic instability (Charney 1947; Eady 1949). These reasons are why we refer to extratropical cyclones as the centerpiece of meteorology.

Despite its immense utility as a conceptual model for routine synoptic analysis, polar-front theory was adopted slowly in the United States. The early development of the Norwegian cyclone model was covered extensively in Monthly Weather Review, which was published by the U.S. Weather Bureau at that time. Specifically, Monthly Weather Review was one of the two journals that printed Bjerknes (1919) and it also reported on American Anne Louise Beck’s yearlong fellowship at the American Anne Louise Beck’s yearlong fellowship at the Bergen School (Beck 1922). Despite these efforts by early career scientists to sell the Norwegian cyclone model to American forecasters (e.g., Meisinger 1920; Beck 1922), the management at the U.S. Weather Bureau resisted (e.g., Bornstein 1981; Namias 1981, 1983; Newton and Rodebush 1999; Fleming 2016, 52–59). For example, Monthly Weather Review Editor Alfred Henry (Henry 1922b,c) reviewed Bjerknes and Solberg (1921, 1922), arguing that the Norwegian cyclone model was not necessarily applicable to weather systems in the United States because of their different geographies and the much larger number of surface observing stations needed in the United States to achieve data densities rivaling that of Norway (Henry 1922a,b). Following the arrival of Carl-Gustaf Rossby to the United States in 1926, the ascent to leadership of the Bureau by Bergen-trained Francis Reichelderfer in 1938, and the subsequent birth of meteorology programs at U.S. universities during World War II helmed by Bergen-trained academics, polar-front theory established stronger roots within the U.S. meteorological community (Bornstein 1981; Namias 1981, 1983; Newton and Rodebusch 1999).

Similarly, resistance occurred in Europe. The United Kingdom also faced similar challenges to adoption of the Bergen School methods (e.g., Douglas 1952; Sutcliffe 1982; Ashford 1992). In Meteorologische Zeitschrift, the leading German-language meteorological research journal in Europe, Ficker (1923) compiled an in-depth critical review of the Bergen school publications before 1922. He lauded the introduction of a compact analysis scheme with clear and memorable diagrams, as well as the short and characteristic names for the relevant phenomena, but he strongly disagreed that a radical new theory had been presented.

1 The reason why the 1919 paper was published simultaneously in two different journals is a bit of a mystery. Because Jakob was young, it is likely that Vilhelm chose the options for the journals. Vilhelm was more aware of the need to get the preliminary findings published quickly. Geofysiske Publikationer was brand new and aimed to reach both sides of a scientific world split by the postwar environment. Still, the new journal was as yet unproven in its ability to serve as a vehicle for groundbreaking research. Vilhelm probably saw Monthly Weather Review as the most reliable venue because its publication was relatively unaffected by the war and probably the least provocative to Germans and Austrians. Vilhelm had previously turned to Monthly Weather Review in line with his past connections with Cleveland Abbe, as well as his connections with the Carnegie Institution in Washington, DC (R. M. Friedman 2018, personal communication).
The second reason why 1919 is important to this chapter was the creation of a new system of international cooperation through a series of meetings in Brussels, Belgium, in July 1919, where international bodies such as the International Association of Meteorology came into formal existence (Ismail-Zadeh and Beer 2009; Ismail-Zadeh 2016). International cooperation is a key theme that runs through this chapter. Members of the Bergen School and its disciples came from various countries, traveled to various countries to found meteorology programs, collaborated internationally on their research, and collected data during international field programs (e.g., Bjerknes 1935; Bjerknes and Palmén 1937). Although Bergen School meteorologists were effective at pursuing international cooperation (Fig. 16-1), there were a few bumps along the way. One bump was the signing of the Treaty of Versailles on 28 June 1919, bringing to a close World War I. One of its immediate consequences for international research cooperation occurred at a 28 July 1919 meeting in Brussels (Ismail-Zadeh 2016) during which the International Research Council (IRC; later renamed ICSU) was founded containing, for example, the International Union of Geodesy and Geophysics (IUGG), which in turn was composed of six sections (later, associations), among them the International Association of Meteorology [IAM; later IAMAP and now the International Association of Meteorology and Atmospheric Physics/Sciences (IAMAS)]. The treaty also meant that the Central Powers were explicitly excluded from membership in any of the bodies mentioned above, a glaring example of how international cooperation was not always such a positive experience. Nevertheless, these nongovernmental international organizations and learned societies (e.g., AMS) in some ways resemble the global and synoptic scales in the social networks akin to those in the atmospheric continuum (Volkert 2017). In addition, individual scientists and their employers (e.g., universities, governmental laboratories, national hydrometeorological services) often obtain energy, inspiration, and motivation from such nonprofit

![Fig. 16-1. International milieu at the Bergen School of Meteorology, two years after the founding of AMS. Shown are the participants at the Eighth Meeting of the International Commission for the Scientific Investigation of the Upper Air on 25 July 1921 in Bergen, as discussed by Volkert (2017). The photograph is provided through the courtesy of the University of Bergen Library.](http://journals.ametsoc.org/mono/article-pdf/doi/10.1175/AMSMONOGRAPHS-D-18-0015.1/4941507/amsmonographs-d-18-0015_1.pdf)
networks on these different scales. The progress reported in all the chapters of this monograph should be viewed within the context of these important cooperative structures.

During the 100 years since 1919, extratropical cyclone research was the centerpiece for the international atmospheric science community, not least because it combined basic research efforts in dynamical meteorology with applied forecasting endeavors using synoptic-scale data analyses and later numerical weather prediction (NWP) techniques. The progress achieved during the past century is traced throughout this chapter in a series of sections by an ensemble of authors and their personal perspectives. For a comparison with previous syntheses, we refer to the AMS-sponsored volumes *Compendium of Meteorology* (Malone 1951), *Extratropical Cyclones: The Erik Palmén Memorial Volume* (Newton and Holopainen 1990), and *The Life Cycles of Extratropical Cyclones* (Shapiro and Grönås 1999).

The *Compendium of Meteorology* was written at the middle of the twentieth century for “taking stock of the present position of meteorology . . . as we are on the threshold of an exciting era of meteorological history” (Malone 1951, p. v). Five chapters summarized the state of science on extratropical cyclones at that time. Bjerknæs (1951) reviewed the then-current state of polar-front theory and exemplified its relevance through a juxtaposition with the life cycle of the storm over North America during 7–10 November 1948. Palmén (1951) presented three-dimensional manual analyses from observational data including fronts, providing evidence for “the role of extratropical disturbances as links in the general atmospheric circulation as cells for the meridional exchange of air masses” (p. 599). The problem of cyclone development in early efforts of numerical forecasting was also referred to by Eady (1951) and Charney (1951). Finally, Fultz (1951) reviewed his own and previous efforts to obtain, among other things, cyclonic eddies in rotating tank experiments and frontal movement in a stratified environment. These chapters highlighted the need for closer correspondence between theory and observations, with Palmén (1951, pp. 618, 619) concluding, “Meteorologists are still in disagreement about many fundamental aspects of the cyclone problem.” and “If the complexity of the cyclone problem is considered, it does not seem likely that any satisfactory theoretical solution can be achieved in the near future.”

During the 1970s and early 1980s, the promise of operational NWP faced a severe challenge. Operational forecast systems frequently failed to predict rapidly developing cyclones (Sanders and Gyakum 1980; Bosart 1981; Gyakum 1983a,b; Anthes et al. 1983). Reed and Albright (1986) described an especially egregious forecast of explosive cyclogenesis over the eastern Pacific by the Limited-Area Fine Mesh Model (LFM), which completely missed the storm development and resulted in a 55-hPa central pressure error. These failures sparked a fertile period of cyclone research in the 1970s, 1980s, and 1990s that included major field programs such as Cyclonic Extratropical Storms (CYCLES; Hobbs et al. 1980), Genesis of Atlantic Lows Experiment (GALE; Dirks et al. 1988), Experiment on Rapidly Intensifying Cyclones over the Atlantic (ERICA; Hadlock and Kreitzberg 1988), Alaskan Storm Program (Douglas et al. 1991), and Fronts and Atlantic Storm-Track Experiment (FASTEX; Joly et al. 1997, 1999). These field programs revealed the structure and evolution of cyclones, as well as their attendant fronts and precipitation. Concurrently, advances in computer infrastructure, model resolution, and model physics led to idealized and real-data simulations capable of resolving these structures. These improvements in models and computer hardware also allowed operational forecasting of the intensification rate of explosive cyclones to improve considerably during this time. The groundwork was laid for a fresh perspective on frontal-cyclone evolution. The seminal nature of this body of research becomes evident from the prominent celebrations of Erik Palmén resulting in *Extratropical Cyclones: The Erik Palmén Memorial Volume* (Newton and Holopainen 1990) and of the 75th anniversary of Bjerknæs (1919) resulting in *The Life Cycles of Extratropical Cyclones* (Shapiro and Grönås 1999).

This chapter advances the narrative in the 20 years since Shapiro and Grönås (1999) while bringing a 100-year perspective to the topic. We are influenced by the conceptual model for scientific inquiry introduced by Shapiro et al. (1999) (Fig. 16-2), which embodies the evolution of research on cyclones during the 100 years that have elapsed since the introduction of polar-front theory. Shapiro et al.’s (1999) model involves theoretical, diagnostic (including dynamical modeling), and observational approaches, swirling cyclonically and then ascending to produce improved physical understanding and conceptual models. The following sections honor this mixing process through the organization of the remainder of this chapter.

Section 2 (written by Roebber and Bosart) describes the ubiquitousness of extratropical cyclones in atmospheric processes, as well as how the depiction of extratropical cyclones have changed over the past century, using East Coast cyclones as an example. Section 3 (written by Davies) presents an overview of theories of cyclone development including the divergence hypotheses of Dines and Sutcliffe, frontal-wave instability, baroclinic instability, quasigeostrophic and semigeostrophic theories, potential-vorticity thinking, and deterministic
2. Extratropical cyclones—The Forrest Gump of the atmosphere

In the popular feature film *Forrest Gump*, the titular character says “Life is like a box of chocolates. You never know what you are going to get.” During the film, which covers the period from the mid-1940s through the early 1980s, Forrest Gump encounters a wide variety of American popular culture icons ranging from Elvis Presley to three U.S. Presidents (Kennedy, Johnson, and Nixon) and experiences—and sometimes influences—notable events such as the Vietnam War, the opening of diplomatic relations with China, the Watergate scandal, and the early days of Apple Computer. Similarly, one can randomly select one cyclone event or another and find that each one is different, owing to the complex interplay of baroclinic and diabatic processes in their development. Likewise, as detailed by Lorenz (1967; discussed in section 4 of this chapter), the instability of the general circulation to baroclinic disturbances necessitates their ubiquity and inevitability, just as Forrest Gump appears everywhere, influencing a half-century of American life.

A succinct and direct definition of an extratropical cyclone, proffered by Fred Sanders and which he attributed to Jule Charney, is that a cyclone is a process and not a thing. By that, Sanders and Charney are referencing the formation and growth of transient baroclinic eddies through dynamic and thermodynamic processes, whose surface manifestation as a pressure minimum is what we recognize as a cyclone. Cyclones were perhaps initially recognized as pressure minima when the first crude synoptic analyses were able to be constructed, which in real time occurred following the introduction of the telegraph and corresponding synoptic observing systems (Kutzbach 1979). The collection of these surface observations led to the production of surface synoptic weather maps (e.g., Reed 1977). Petterssen (1969) presented several examples of early cyclone models resulting from analysis of surface synoptic maps: the 1861 opposing currents model of Master Mariner Jimman, Fitzroy’s 1863 model of cyclonic whirls, the 1883 cyclone weather distribution model of Abercromby, and Shaw’s (1911) cyclone model (Fig. 16-3). It was the Bergen...
School, however, that advanced understanding of these systems by setting forth these observations in the form of a four-dimensional picture that is the now-famous frontal cyclone model (Bjerknes 1919; Bjerknes and Solberg 1922; Fig. 16-4). Eliassen (1999) and Volkert (1999) present further details of advances in European understanding.

As one example of a region with high-impact extratropical cyclones that ties the sections in this chapter together, we consider northeastern U.S. snowstorms (also colloquially known as northeasters or nor'easters). The high population density combined with lots of meteorologists living in this region and the occasional big snowstorm was an excellent recipe for a “perfect storm” of meteorological awareness and weather lore (Kocin and Uccellini 2004) that goes back to the nineteenth century, as evidenced by the legendary East Coast blizzards of 1888 and 1899 (Kocin 1983; Kocin et al. 1988). Characteristic northeastern U.S. storm tracks parallel to the Atlantic coast and from the Ohio Valley northeastward down the St. Lawrence River Valley were described in an atlas prepared by Bowie and Weightman (1914). Austin (1941) and Petterssen (1941) provided illustrative examples of typical northeastern U.S. cyclones. Miller (1946) documented two types of East Coast cyclones, which he termed type A and type B. Type-A cyclones typically originated along a frontal boundary near the coast, whereas type-B coastal secondary cyclones formed in conjunction with the death of a primary cyclone west of the Appalachians. Type-B cyclones represented a greater forecast challenge because of uncertainties associated with the forecast location and timing of secondary cyclone development, a challenge that remains today. A famous example of a type-A cyclone was the New York City blizzard of 26–27 December 1947 (Uccellini et al. 2008). Snowfall amounts of about 67 cm in less than 24 h were reported in New York City with higher amounts in the suburbs (Weather Bureau 1948). This storm brought New York City to a standstill.

Although the synoptic-scale location and structures of these cyclones were critical to getting the forecast correct, northeasters also produce important mesoscale structures that could cause large changes in hazardous weather over short distances, further frustrating forecasters. Spar (1956) showed an example of a type-A cyclone that contained embedded areas of high winds near the surface warm front that could be associated with downward momentum mixing and discrete warm-front propagation. Bosart et al. (1972) and Bosart (1975) first documented the existence of mesoscale coastal fronts ahead of Atlantic coastal cyclones. He showed that coastal fronts served as a locus of surface frontogenesis and cyclonic vorticity generation and that northeastward-propagating coastal cyclones tended to track along a preexisting coastal front. Coastal fronts served as boundaries between frozen and unfrozen precipitation with the heaviest precipitation falling along and on the cold side of the boundary. The impact of enhanced
diabatic heating due to precipitation along and toward the cold side of coastal fronts impacted the cyclogenesis process through enhanced low-level convergence and cyclonic vorticity generation (e.g., Keshishian and Bosart 1987). Tracton (1973) and Ellenton and Danard (1979) showed that unrepresented diabatic heating and the associated low-level convergence and cyclonic vorticity generation in NWP models could be a source of significant model forecast error in northeastern U.S. cyclones, a finding that could also be linked to coastal-frontogenesis processes. Furthermore, stratified air masses on the cold side of coastal fronts proved to be effective in providing wave ducts for the passage of long-lived, large-amplitude mesoscale inertia–gravity waves (e.g., Bosart and Sanders 1986; Uccellini and Koch 1987; Bosart and Seimon 1988; Bosart et al. 1998). An excellent example of a long-lived, large-amplitude mesoscale inertia–gravity wave and “snow bomb” associated with a strong Atlantic coastal cyclone occurred on 4 January 1994 (Bosart et al. 1998) (Fig. 16-5).

The catastrophic failure of then-operational forecast models to predict the infamous Presidents’ Day coastal storm of 19 February 1979 (Bosart 1981; Bosart and Lin 1984; Uccellini 1990; Uccellini et al. 1984, 1985) had a major impact on operational NWP. Bosart (1981) showed that the then-NMC (predecessor to NCEP) operational forecast model known as LFM-II had nary a clue about the intensity and location of the eventual Presidents’ Day storm. A strong coastal front that was associated with the storm enabled it to hug the coast and intensify rapidly in an environment favorable for strong latent heating, low-level convergence, and cyclonic vorticity generation (Bosart 1981). The then-operational LFM-II had no parameterization for latent heat flux as was evident from a comparison of the observed and predicted coastal planetary boundary layer structure (Fig. 22 in Bosart 1981). The absence of assimilation of significant-level sounding data into the NMC operational forecast system at that time likely further contributed to the deficient operational forecasts of the storm (Bosart 1981). The forecast debacle that was the Presidents’ Day storm in the Washington, D.C., area was a watershed moment that helped to usher in significant advances to the then-NMC operational forecasting enterprise in subsequent years. Another important NMC operational model forecast failure occurred in conjunction with an early-season coastal storm that occurred on 4 October 1987. This storm dumped more than 50 cm of snow on portions of interior eastern New York and western New England and was investigated by Bosart and Sanders (1991). They showed that the forecast failure could likely be linked to an improperly analyzed low-level wind field and vertically integrated moisture field.
The Presidents’ Day storm coupled with the publication of the first comprehensive climatology of “bomb” cyclones by Sanders and Gyakum (1980) opened the floodgates to further studies of now famous Atlantic coast storms such as the Megalopolitan storm (Sanders and Bosart 1985a,b), the Queen Elizabeth II storm (Gyakum 1983a,b; Uccellini 1986), the eastern Ohio Valley bomb cyclone of 25–26 January 1978 (e.g., Hakim et al. 1995), the “perfect storms” of late October and early November 1991 (e.g., Cordeira and Bosart 2010, 2011), and the 13–14 March 1993 Superstorm (e.g., Uccellini et al. 1995; Bosart et al. 1996; Dickinson et al. 1997). The importance of upstream precursor disturbances on western Atlantic Ocean cyclogenesis cases was also identified (e.g., Sanders 1986a, 1987; Lackmann et al. 1997; Cordeira and Bosart 2010). Results from field programs such as GALE in 1986 (Dirks et al. 1988) and ERICA in 1988–89 (Hadlock and Kreitzberg 1988) solidified the importance of previously neglected diabatic heating processes during intense oceanic cyclogenesis and illustrated the importance of upstream precursors to downstream cyclogenesis.

Statistical analyses and climatologies of explosively deepening western North Atlantic cyclones motivated by these field experiments established the existence of a skewed distribution of explosively deepening extratropical cyclones toward the rapid deepening end (e.g., Roebber 1984, 1989). Further numerical investigations of explosively deepening extratropical cyclones by Roebber and Schumann (2011, p. 2778) has revealed “that the strongest maritime storms are the result of the baroclinic dynamics of the relative few being preferentially enhanced through feedback with the available moisture. Strong baroclinic forcing, in the absence of this moisture availability and resultant latent heating, does not produce the skewed rapid deepening tail behavior.” These results indicate that very rapidly deepening intense oceanic extratropical cyclones are the result of a fundamentally distinct pattern of behavior characteristic of maritime cyclones as compared with continental cyclones and that this behavior is the result of process interactions (i.e., baroclinic dynamics and latent heat release). These results further indicate that the combination of diabatic forcing associated with latent heat release in a highly baroclinic environment can account for the skew on the right side of the cyclone intensity distribution, pointing the way toward future research on rapidly intensifying oceanic cyclones and associated atmospheric predictability studies.

Using an example of a northeaster, one measure of how much cyclone knowledge and its graphical representation has advanced in 100 years is to compare the idealized depictions of cyclones (Figs. 16-3–16-4) with a modern depiction of a real extratropical cyclone from gridded model analyses (Fig. 16-6). A strong, sub 965-hPa cyclone lay off the east coast of North America at 1200 UTC 4 January 2018 (Fig. 16-6a). This cyclone easily met the Sanders and Gyakum (1980) condition for a bomb cyclone, with rapid intensification occurring between the favored equatorward entrance region of the jet streak to the north and the poleward exit region of the jet streak to the south. The cyclone was located near the thermal ridge in the 1000–500-hPa thickness field with strong warm-air advection to the north and east and strong cold-air advection to the south and west. The strong sea level pressure gradient on the southwestern side of the storm was associated with exceptionally strong surface westerly winds estimated to have exceeded 40 m s\(^{-1}\). The cruise ship Norwegian Breakaway was caught in these strong winds, with resulting injuries to passengers and crew and considerable damage to the vessel (http://newyork.cbslocal.com/2018/01/05/cruise-through-storm/).

The 4 January 2018 storm can be illustrated in a modern dynamical perspective through a dynamical-tropopause view (Fig. 16-6b), an analysis of upper-level potential vorticity (PV), and upper-level divergent irrotational wind outflow (Fig. 16-6c), representing the underlying physical processes in the extratropical cyclone in a way that the conceptual models in Figs. 16-3–16-4 cannot. A classic signature of an explosively deepening extratropical cyclone is a PV hook as evidenced by potential temperature values less than 310 K approaching the cyclone center (Fig. 16-6b) and accompanying layer-mean 925–850-hPa relative vorticity along the bent-back front as the cyclone approaches its occluded stage. Good agreement exists between the location of the bent-back 925–850-hPa vorticity in Fig. 16-6b with the 600–400-hPa layer-mean ascent in Fig. 16-6c. Diabatically generated outflow from the deep ascent in the northern semicircle of the storm is manifested by a starburst pattern in which negative PV advection by the irrotational wind acts to strengthen the PV gradient from the southwestern to northeastern side of the storm with an associated tightening of the horizontal PV gradient and a strengthening of the downstream jet to over 100 m s\(^{-1}\) (not shown).

With this background and perspective on extratropical cyclones, we turn to their dynamics and the theoretical frameworks during the past century that have helped advance our understanding of the development of cyclones.

3. Theories of cyclones and cyclogenesis

The dominating presence of cyclones and anticyclones within the atmosphere’s chaotic extratropical flow prompts fundamental theoretical questions related to their raison d’etre, ubiquity, variety, and characteristic
space–time scales. Not surprisingly then, the quest to understand the day-to-day development of synoptic-scale flow and to formulate perceptive theories for extratropical cyclogenesis has been one of meteorology’s long-standing objectives. Indeed, Margules in his parting contribution to meteorology identified extratropical cyclogenesis as one of the discipline’s grand challenges and avowed, “I consider it unlikely that observations alone will suffice to provide a useful model of cyclogenesis. An individual equipped with sufficient knowledge of the observations and endowed with imagination and abundant patience may attain this goal” (Margules 1906, p. 497).

The response to this grand challenge has been chronicled in several studies overviewing theories of cyclogenesis (e.g., Hoskins 1990; Reed 1990; Pierrehumbert and Swanson 1995; Davies 1997; Thorpe 2002). In this section, a digest is provided of the iconic theories that have been advanced from around the time of the AMS’s founding with consideration being given to each theory’s essence, emergence, and explanatory power.

The period around 1919 was a propitious time to address the Margulesian challenge. The disputations of the mid-1800s between protagonists favoring James Pollard Espy’s thermal versus William Redfield’s mechanical conception of cyclones and cyclogenesis had long since abated (e.g., Kutzbach 1979), quasi-real-time surface synoptic datasets were accruing from the newly established but sparsely spaced observational networks, limited upper-air soundings were becoming available, and the key classical laws of physics pertinent for atmospheric flow had been established (e.g., Abbe 1901; Bjerknes 1904). Furthermore, case-study analyses were beginning to tease out inchoate characteristics of a cyclone’s low-level features from the seeming morass of mildly related surface observations. More trenchantly at this time, two nascent hypotheses for cyclogenesis were being advanced. Thus, like Robert Frost’s traveler, the meteorological community was confronted in 1919 with

---

5 \times 10^{-2} \text{ hPa s}^{-1}, \text{with only negative values shown}, \text{and 300−200-hPa layer-averaged irrotational wind (vectors, starting at 3 m s}^{-1}; \text{length scale at lower-left corner}). (b) Dynamic tropopause potential temperature (K; color shading) and wind barbs (pennant, full-barb, and half-barb denote 25, 5, 2.5 m s}^{-1}, \text{respectively}; \text{along with 925−850-hPa layer-mean cyclonic relative vorticity (solid lines every 0.5 \times 10^{-4} \text{ s}^{-1}). (c) Sea level pressure (solid lines every 4 hPa), 1000−500-hPa thickness (dashed lines every 6 dam, with a changeover from blue dashed lines to red dashed lines between 540 and 546 dam), precipitable water (mm; color shading), and 250-hPa wind speeds (m s}^{-1}; \text{gray shading}). The figure was provided by H. Archambault.
“two paths diverging…,” and the theme of divergence was also to permeate the subsequent history of cyclogenesis.

a. Two nascent hypotheses

The central theme of the first of the nascent hypotheses was indeed horizontal divergence. The hypothesis is encapsulated in the following statement: “...a cyclone is produced by the withdrawal laterally of the air at a height of from 8 to 10 kilometres” (Dines 1912, p. 46). The features identified by Dines were the result of a prodigious feat of inspired analysis conducted with the available meager data (Fig. 16-7). It revealed the distinctive structure of mature cyclones near the tropopause with a cold central core located beneath a lowered tropopause that was itself surmounted by a warm core in the lower stratosphere.

The hypothesis correctly eschewed the inference, asserted by some, that the surface low had a stratospheric cause, but rather pointed to tropopause-level divergence as the mediator of the overall vertical structure. However, the hypothesis neither established a determining process for the divergence nor accounted for the earlier perceptive observational detection by Ley (1879) and Bigelow (1902) that a growing cyclone’s center of low pressure tilted upstream with increasing height in the lower troposphere.

Furthering this hypothesis was hampered by two factors. First, there was a lack of adequate upper-air data to shed light on the space–time development of the cyclone’s vertical structure. Notwithstanding, Ficker (1920) provided a prescient illustration of a surface low pressure center developing as a major flow feature (i.e., an upper-level trough) advanced toward a secondary feature (i.e., a surface trough). Observations acquired in the subsequent decades revealed an empirical link between certain recurring upper-airflow patterns such as the delta-shaped jet exit region with surface cyclogenesis, and suggestive, but incomplete, arguments were advanced to account for this linkage by Scherhag (1934) as discussed by Volkert (2016) and Namias and Clapp (1949). The second major limiting factor was that this nascent theory’s emphasis on horizontal divergence highlighted an Achilles’s heel of atmospheric dynamics that was to bedevil progress for decades. Margules (1904) had deduced that its accurate computation with the available data would be challenging, and Jeffreys (1919) noted that geostrophic flow implied weak horizontal divergence, thwarting attempts at direct calculation of the divergence.

The other nascent hypothesis was that associated with the Bergen School under the leadership of Vilhelm Bjerknes. The Bergen School’s contribution can be viewed as comprising two components related respectively to the morphology of surface weather patterns and to the occurrence of cyclogenesis. First, the Bergen School came to conceive synoptic-scale atmospheric flow as being dominated by an elongated sloping frontal boundary separating air masses of different temperature, and the interface itself was depicted as deforming into alternate cold- and warm-frontal segments (sections 5 and 6). This portrayal of surface weather patterns was an amalgam of a reconstituted synthesis of earlier studies and a brilliant conceptualization of the extant surface observational data. Its crisp depiction of cold and warm fronts remains (with some refinements) a staple ingredient of synoptic analysis charts to this day.

The second component arose from the Bergen School’s observation that the frontal interface was the seat for wave undulations that subsequently evolved to form a train of cyclones (Bjerknes and Solberg 1922; Fig. 16-8). They hypothesized that these undulations were attributable to the instability of the sloping frontal interface, and an attempt was made to determine the stability of a basic state comprising a uniformly sloping interface separating two homogeneous incompressible fluids of different uniform densities and velocities. This setting replicated that already proposed by Margules (1906), and the hypothesis would yield striking explanatory power provided the most unstable perturbations of the interface were to correspond to the characteristic
space–time scale of observed frontal-wave cyclones. However, numerous studies, based upon variants of the Margulesian front, conducted first by the Bergen School (Bjerknes and Godske 1936) and subsequently by many others have not yielded fully persuasive support for the hypothesis. Thus, like the hypothesis for upper-level driven cyclogenesis, the Bergen School’s hypothesis of frontal instability lacked firm theoretical underpinning.

b. Two substantive theories

By the middle of the twentieth century, two substantive theories emerged that were to exert an enduring influence upon studies of cyclogenesis. A hallmark of both theories was their distinctive approach to divergence. One theory focused explicitly on estimating the divergent component of the flow, whereas the other avoided its direct consideration. Key to both approaches were 1) a refined interpretation of divergence, as embodied in the term quasigeostrophy coined by Durst and Sutcliffe (1938, p. 240), “...departures of the wind velocity from the geostrophic value...are generally small...(so that the whole motion can be described as quasigeostrophic) but they are of fundamental dynamical significance” and 2) the realization that a simplified version of the equation for the vertical component of the vorticity was appropriate for synoptic-scale flow (Rossby 1940).

The first theory (Sutcliffe 1938, 1947) set out to diagnose the weaker ageostrophic (or divergent) flow component from a knowledge of the geostrophic component itself. It proved possible to infer qualitatively (using conventional geopotential and thermal charts) the sign of the difference between upper- and lower-level horizontal divergence, and thereby identify preferred regions for cyclogenesis (and anticyclogenesis) along with the direction of translation of pressure systems (Fig. 16-9). Sutcliffe (1947, p. 383) concluded with seeming diffidence that, “Since the arguments and deductions are susceptible both to physical interpretation and to practical test, they may have some acceptable virtue.”

This theory amplified Dines’s hypothesis, provided a tool for estimating flow development (i.e., the evolution of weather patterns), and was readily applicable. The theory also helped fuse synoptic and dynamic meteorology. Its virtue is attested by the fact that meteorological terminology soon became replete with terms such as diffluent and confluent troughs, left exit of the jet stream, and thermal steering that referred to certain developmental patterns (Fig. 16-9).

The second theory, baroclinic instability (Charney 1947; Eady 1949), resulted from an examination of the stability of a steady uniform baroclinic shear flow in the extratropics. Eady (1949, p. 33) concluded that “small disturbances of simple states of steady baroclinic largescale atmospheric motion...are almost invariably unstable,” and that, in the f-plane limit, the most unstable perturbation possessed a spatial scale and growth rate akin to that of larger-scale cyclones. In effect, although a latitudinal temperature gradient can be balanced by a commensurate zonal flow, wave perturbations of that balanced state can feed from the associated available potential energy. A subsequent simulation with a simple numerical model indicated that growth of the disturbance to finite amplitude resulted in cyclogenesis and frontogenesis (Phillips 1956, 141–142): “The wave begins as a warm low, and...the final stages look very much like those of an occluded cyclone....Definite indications of something similar to cold and warm fronts are to be seen in the 1000-mb [hPa] contours.” This theory views fronts as emerging during cyclogenesis and therefore differs radically from the Bergen School concept of fronts being the source of cyclogenesis.

Together these two theories helped establish meteorology as a scholarly scientific discipline in the broader scientific community. They also encapsulated in embryonic form the diagnostic and predictive components of the so-called quasigeostrophic set of equations, whose formal derivation soon followed. The first theory was

FIG. 16-8. A train of frontal-wave cyclones (Bjerknes and Solberg 1922, their Fig. 9). The figure is provided through the courtesy of the Norwegian Geophysical Society.
generalized to yield the diagnostic component of the quasigeostrophic set, the so-called $\omega$ equation (Fjortoft 1955). In addition to its deployment for forecasting (e.g., Sutcliffe and Forsdyke 1950; Petterssen 1955), this equation was used to detect the occurrence of cyclogenesis linked to an upper-level trough advancing toward a surface baroclinic zone (Petterssen 1956, p. 335), to classify different types of cyclogenesis (Petterssen and Smebye 1971), and to undertake case study analyses of, for example, events of explosive maritime cyclogenesis. Contemporaneous with these early studies, the contribution of kinematically estimated upper- and lower-level divergence to the three-dimensional development of, and the link between, cyclogenesis and frontogenesis was being elicited in a stream of perceptive diagnostic studies (e.g., Newton 1954, 1956; Newton and Palmén 1963).

Baroclinic instability theory was followed by the formal derivation of the predictive component of the quasigeostrophic set (Charney 1948; Eliassen 1949). This single and self-contained equation states that there is a quasigeostrophic form of the potential vorticity that is conserved following the flow. It is a radical simplification of the primitive equations and refers only to the geostrophic flow (thereby circumventing direct consideration of the divergent component). It has provided a fruitful test bed for pursuing studies of baroclinic instability and cyclogenesis because it is amenable both to numerical solution and to mathematical analysis.

Three-dimensional numerical simulations conducted with this equation (e.g., Mudrick 1974), its semi-geostrophic counterpart (e.g., Hoskins 1975; Hoskins and West 1979; Heckley and Hoskins 1982; Davies et al. 1991; Snyder et al. 1991; Wernli et al. 1998), and the primitive equations (e.g., Mudrick 1974; Simmons and Hoskins 1978, 1979; Snyder et al. 1991; Thorncroft et al. 1993; Rotunno et al. 1994; Wernli et al. 1998; Schultz and Zhang 2007) 1) confirmed that the nonlinear phase of baroclinic instability replicates cyclogenesis with accompanying cold- and warm-frontal accoutrements, 2) showed that a wide panoply of cyclone types and fronts can result from the ambient flow possessing jet-like features or lateral shear, 3) calibrated the modifying role of cloud diabatic heating, and 4) demonstrated that a localized upper-tropospheric anomaly can effectively trigger surface cyclogenesis. Mathematical analysis of the equation (Charney and Stern 1962; Pedlosky 1964) established general instability criteria for two-dimensional basic states, and thereby helped both guide and interpret the

---

**Fig. 16-9.** Some classical developmental patterns: (a) Thickness contours for a diffluent thermal ridge [labeled "(a)"], a confluent thermal ridge [labeled "(b)"], a diffluent thermal trough [labeled "(c)"], and a confluent thermal trough [labeled "(d)"]. (b) A thermal jet complex. (c) The development of a warm-sector depression. In (a) and (b), the symbols A and C respectively refer to preferred regions for anticyclogenesis and cyclogenesis. In (c), the cyclonic development due to the thermal development process at C favors the continuous distortion of the pattern. The panels are from Sutcliffe and Forsdyke (1950, their Figs. 22, 24, and 23a); © Royal Meteorological Society.
results of exploratory studies. Likewise, the concept of baroclinic instability is central to the theories for the atmosphere’s general circulation (Held 2019).

The compatibility of the two substantive theories discussed above is illustrated in Fig. 16-10. It shows features of cyclogenesis derived from a variety of approaches: three-stage cyclone formation accompanying strong vorticity advection aloft based upon \(v\)-equation considerations (top row), synoptic syntheses of flow in the lower half of the troposphere in three stages (middle row) and a three-dimensional schematic (left panel in bottom row), and surface and tropopause-level patterns resulting from a semigeostrophic nonlinear simulation of baroclinic instability of a jet flow in the Eady configuration (middle and right panels in bottom row).

c. Two paradigm-changing frameworks

In the second half of the twentieth century, two theoretical advances resulted in new paradigms for studying synoptic-scale flow development and cyclogenesis. These paradigms are the potential vorticity perspective and deterministic chaos. The former regards the space–time development of the interior PV and the surface potential temperature to be key to understanding balanced...
flow, and that knowledge of the instantaneous distributions of these variables “...is sufficient to deduce, diagnostically, all the other dynamical fields, such as winds, temperatures, geopotential heights, static stabilities, and vertical velocities” (Hoskins et al. 1985, p. 877).

In its mature form, the PV perspective is a coalescence, generalization, and exploitation of several aspects of atmospheric dynamics, namely depiction of the flow on isentropic surfaces (Shaw 1930; Rossby et al. 1937; Namias 1939), exploitation of the Lagrangian conservation property of PV under adiabatic and frictionless conditions (Rossby 1940; Ertel 1942), extension of the quasigeostrophic concepts of partition and inversion (Charney 1963) to higher forms of balanced flow (Davis and Emanuel 1991), and detection and quantification of diabatic changes following air-parcel trajectories (Whitaker et al. 1988; Uccellini 1990; Wernli and Davies 1997).

For cyclogenesis, the PV perspective focuses attention on the dominant time-evolving, coherent flow features of PV in the interior (i.e., wave and vortex-like features in the upper troposphere and lower stratosphere, cloud-modified regions of the troposphere) and of potential temperature at the surface (i.e., frontal undulations and cutoff cold and warm pools). In this framework, the archetypical upper-level induced surface cyclogenesis can be viewed as an upper-level localized PV anomaly instigating and sustaining, via its far-field effect, a perturbation on an underlying lower-level front. More generally, a suitably located and isolated PV anomaly (generated by adiabatic or diabatic processes) can trigger disturbances on a surface front or upper-level jet. Such vortex–wave interaction bears comparison with aspects of upstream and downstream development, extratropical transition, Rossby wave breaking, diabatic Rossby waves, and also the train of surface frontal-wave cyclones akin to that portrayed by the Bergen School (Fig. 16-8).

Likewise, classical baroclinic instability can be viewed as a wave–wave interaction involving a PV wave near the tropopause and a potential temperature wave on the surface. For the classical Eady configuration, the interaction is between potential temperature waves on the upper and lower bounding surfaces (Davies and Bishop 1994). In both settings, maximum instantaneous growth prevails when the upper and lower waves are in quadrature before they transit to a shape-preserving (i.e., normal-mode) structure. The latter state prevails when the two waves remain stationary relative to one another under the influence of their differing upper- and surface-level ambient flow fields. One import of this result is that the fastest-growing normal mode is not the optimum perturbation for maximizing transient growth, illustrated elegantly by Farrell’s (1982) example of rapid nonmodal growth. More circumspectly, consideration of nonmodal perturbations introduces questions related to the nature of the growth, namely where (e.g., global, regional), when (i.e., over what time span), and of what (i.e., selection of a suitable metric).

In addition, the perspective invites consideration of other aspects of cyclogenesis. For example, tracing the origin of the high-PV air that surmounts a surface cyclone by computing backward trajectories can shed light on subtle dynamics of cyclone formation by highlighting the contribution and differing source regions of the high-PV air (e.g., Rossa et al. 2000) and demonstrating that forecast error growth can be associated with the misrepresentation of these differing airstreams (e.g., Davies and Didone 2013).

The second paradigm-changing concept referred to above is that of deterministic chaos. Edward Lorenz, the principal architect of this concept, showed that deterministic flow systems that exhibit nonperiodicity are unstable, and he went on to note in his breakthrough study, “When our results...are applied to the atmosphere, which is ostensibly nonperiodic, they indicate that prediction of sufficiently distant future is impossible by any method, unless the present conditions are known exactly” (Lorenz 1963, p. 141).

Large-scale atmospheric flow is indeed an exemplar of an intrinsically chaotic system. Consonant with this observation, NWP simulations demonstrate a sensitive response to small differences in the initial state so that with time the trajectories of these simulations diverge in phase space. This is an apologia, par excellence, for the failure of single deterministic forecasts, and a prompter for applying an ensemble approach to NWP (section 7b).

The import of Lorenz’s result for cyclogenesis studies is manifold. For example, on the time scale of days, uncertainty in the specification of an NWP’s initial state could in principle result in the under- or overdevelopment—or even the simulated nondevelopment or unrealized development—of a cyclogenesis event. For example, Fig. 16-11 illustrates the sensitivity to the specification of the initial conditions exhibited by the 42-h operational ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) for the major European Cyclone Lothar (Wernli et al. 2002) in December 1999. Only 13 of the 50 (26%) ensemble members produced a cyclone with an intensity equal to or greater than that observed. Such depictions provide a practical measure of the predictability of such storms, and the subsequent challenge is to decipher what, if any, small variations of the atmosphere’s initial flow state can significantly promote or
inhibit an event’s subsequent occurrence. Again, on the subseasonal time scale, a sector’s flow can be dominated by a particular weather regime (i.e., characterized for example by the occurrence of a series of transient cyclones or a sequence of collocated blocking anticyclones), prompting questions related to the predictability of weather regimes. The challenge is to determine and understand the nature of the linkage between individual weather events and the sustained forcing factors (e.g., sea surface temperature anomalies, stratospheric flow state), and whether this linkage is associated with predictability—or unpredictability—islands in the troposphere’s chaotic flow.

Lorenz’s concept has patently lifted cyclogenesis studies to a new realm, and this paradigm-changing effect has been mirrored in other scientific fields. The citation accompanying Lorenz’s award of the prestigious Kyoto Prize states that deterministic chaos “has profoundly influenced a wide range of basic sciences and brought about one of the most dramatic changes in mankind’s view of nature since Sir Isaac Newton.”

Each of the iconic theories discussed in this section sought to establish the basic dynamics governing cyclogenesis, and with the passage of time the tropopause-level jet stream and its associated across-stream temperature gradient, emerged as key factors. In the next section, attention shifts to discussing the influence of these factors upon the geographic distribution of the birth, growth, and decay of extratropical cyclones, as well as their dependence upon and subtle contribution to the jet stream.

4. Where do extratropical cyclones occur? Jet streams and storm tracks

Climatologies show that cyclogenesis tends to occur in specific geographic locations (Fig. 16-12). Specifically, maxima of cyclogenesis occur across the North Atlantic
Ocean and North Pacific Ocean in the Northern Hemisphere winter (Fig. 16-12a) and across the Southern Ocean and east of Australia and New Zealand in the Southern Hemisphere winter (Fig. 16-12b). Why maxima in cyclogenesis occur over the oceans is the principal topic of this section.

Understanding the locations and conditions for cyclogenesis requires a gaze upward to the upper troposphere and the jet stream. Storm tracks are preferred areas of the jet stream that control the genesis, movement, and lysis of synoptic-scale pressure systems, and they are critical to midlatitude dynamics in several ways (e.g., Chang et al. 2002).

First, cyclones and storm tracks are an essential part of the atmospheric general circulation (e.g., Held 2019). A large fraction of the meridional energy and momentum transport in the midlatitude atmosphere occurs within the storm tracks (Fig. 16-13b), and the storm tracks thereby sustain the eddy-driven (or polar) jet streams. Starr (1948), in his famous essay on the general circulation, considered the role of anticyclones and cyclones in the poleward transfer of absolute angular momentum. He noted that the distribution and shapes of individual time-mean subtropical anticyclones over the oceans facilitate the poleward transfer of absolute angular momentum from the easterly trade winds. He also remarked that typical midlatitude cyclones as studied by Bjerknes et al. (1933) served to facilitate the downward transport of absolute angular momentum from upper levels because rising air ahead of cyclones was closer to Earth’s axis of rotation than descending air behind cyclones. Lorenz (1967) provided a now-famous first quantitative analysis of Earth’s general circulation in a World Meteorological Organization monograph. He stressed that, because the general circulation would be unstable to small-scale baroclinic disturbances, the observed circulation would have to contain mature cyclones and anticyclones, in agreement with the results from Bjerknes (1937). Newton (1970) further quantified the role of extratropical cyclones in Earth’s general circulation. He calculated that the kinetic energy produced during the extratropical transition of Hurricane Hazel in 1954 (Palmén 1958) was $19 \times 10^{13}$ W or about 25% of the kinetic energy production in the entire extratropical region. This result led Newton (1970, p. 148) to conclude that “only 4 or 5 active disturbances would suffice to account for the total (kinetic energy) generation, in harmony with the conclusion...that a few disturbances could accomplish the required meridional and vertical heat exchange.”

Second, the location and temporal variability of the storm tracks determines the midlatitude mean climate (Namias 1950), as well as the frequency and intensity of weather and climate extremes. On interannual time scales, latitudinal shifts or the zonal extension and contraction of the storm tracks result in regional precipitation and temperature anomalies in the area of the storm tracks and farther downstream. Examples are the effects of the Atlantic storm-track variability on Mediterranean precipitation (e.g., Zappa et al. 2015) or the

![Fig. 16-12. Winter climatologies of (a) Northern Hemisphere [December-February (DJF)] cyclogenesis and (b) Southern Hemisphere [June-August (JJA)] cyclogenesis for 1958-2001. The units are number of events per 10^4 km^2. The field has been calculated on a 3° × 3° latitude-longitude grid and is not plotted in regions where the topography exceeds 1800 m. The figure is adapted from Wernli and Schwierz (2006, their Figs. 6a and 7a).](image-url)
changes in the Pacific storm track during strong El Niño events and associated precipitation anomalies over North America (e.g., Andrade and Sellers 1988; Chang et al. 2002) and South America (e.g., Grimm et al. 1998).

Third, storm tracks are teleconnection agents. They translate Rossby wave forcing (e.g., from tropical convection, stratospheric-temperature anomalies, and sea-ice anomalies) to regional impacts in areas remote from the original forcing. The role of the storm tracks extends beyond the mere transfer of a disturbance, however. The storm tracks can amplify the low-frequency Rossby waves in the jet stream via eddy feedbacks on the background flow (e.g., Held et al. 1989; Hartmann 2007).

As a consequence of these three reasons, a detailed understanding of storm-track dynamics and proper representation in numerical models is essential for capturing the midlatitude dynamical response to external forcings, understanding internal variability or forecasting for seasons and beyond.

a. Global occurrence

The existence of storm tracks has historically been recognized by meteorologists since before the twentieth century (e.g., Kropotkin 1893; Van Bebber 1891; Van Bebber and Köppen 1895; Chang et al. 2002 provide an overview). In the middle of the twentieth century, Northern Hemisphere storm tracks based on surface weather charts were compiled by Klein (1951, 1957, 1958) and Petterssen (1956, 266–276). With the emergence of gridded analysis datasets by the end of the century, new and more comprehensive views of the storm tracks became possible.

Specifically, two complementary diagnostic methods have been used to identify storm tracks from these gridded meteorological fields. Early computational studies identified storm tracks from time-filtered fields in the Northern Hemisphere (Fig. 16-13a; e.g., Blackmon 1976; Lau and Wallace 1979) and the Southern Hemisphere (Fig. 16-13a; e.g., Trenberth 1991; Berbery and Vera 1996). This approach identifies the storm tracks from variability maxima in meteorological fields (e.g., relative vorticity, height, wind) associated with the passage of synoptic-scale eddies. These methods are still frequently used as they link to the energy and momentum budgets, are computationally inexpensive, and are easy to apply. Alternatively, synoptic-scale eddies can be tracked using manual tracking (e.g., Klein 1957), lagged correlations (e.g., Wallace et al. 1988), or automated feature-tracking algorithms (e.g., Hodges 1995; Fig. 16-13a contours), providing information on the entire storm life cycle from genesis to lysis and hence a Lagrangian perspective of the storm tracks (e.g., Hoskins and Hodges 2002, 2005; Wernli and Schwierz 2006).
The Northern Hemisphere possesses two main storm tracks over the North Atlantic and North Pacific Ocean basins (Fig. 16-13a), comparable in magnitude. The Southern Hemisphere possesses one storm track spiraling across the South Atlantic and south Indian Oceans turning poleward over the western Pacific (Fig. 16-13a). A second subtropical storm track at lower latitudes extends from southern Australia across the Pacific with a southerly tilt over the eastern Pacific. The maximum in number of storms is located over the South Atlantic and Indian Oceans.

The storm tracks in each hemisphere generally reach their maximum in eddy kinetic energy during the winter season when the equator-to-pole temperature gradients are strongest (Chang et al. 2002). An interesting exception is the North Pacific storm track. In midwinter, eddy kinetic energy decreases slightly over the Pacific storm track (Nakamura 1992), a local minimum referred to as the midwinter suppression. A possible explanation for the midwinter suppression is the faster progression of eddies across the baroclinic zone in winter due to a stronger background flow, reducing baroclinic amplification (Chang 2001) and resulting in shorter lifetimes of the cyclones (e.g., Schemm and Schneider 2018). Along similar lines, vertical trapping of baroclinic eddies resulting in reduced vertical interaction has also been suggested (Nakamura and Sampe 2002). Another explanation is variability in the number of cyclones that reach the Pacific storm track from upstream (Penny et al. 2010). Another midwinter storm-track suppression mechanism is provided by Schemm and Schneider (2018). They find that the number of cyclones in the North Pacific storm track remains high in the Pacific in the midwinter but the mean eddy kinetic energy per cyclone is reduced (Schemm and Schneider 2018). In contrast, Southern Hemisphere storm-track intensity variations between seasons are small (e.g., Hoskins and Hodges 2005). In the summer hemispheres, the storm track shifts poleward (e.g., Hoskins and Hodges 2005; Wernli and Schwierz 2006) and the upper-level jets shift with the storm tracks (e.g., Koch et al. 2006).

Maxima in cyclogenesis also occur downstream of major mountain ranges such as the Rocky Mountains and Alps in the Northern Hemisphere (Fig. 16-13a) and the Andes and the Antarctic Peninsula in the Southern Hemisphere (Fig. 16-13b). Cyclogenesis in the lee of the Rocky Mountains was first studied by Newton (1956), building upon earlier work by Hess and Wagner (1948). Newton’s (1956) time-dependent three-dimensional analysis enabled him to interpret a lee cyclone on 17–18 November 1948 in terms of dynamical principles by connecting the cyclonic vorticity advection aloft along the 300-hPa jet stream to the ascent and upper-level divergence above the developing lee cyclone. He linked his results to Petterssen’s (1955) finding that the “cyclone development at sea level occurs where and when an area of positive vorticity advection in the upper troposphere becomes superimposed on a frontal zone in the lower troposphere” (Newton 1956, 528–529). Newton further showed how the period of rapid surface lee cyclogenesis was associated with maximum 500-hPa ascent beneath the jet. In what was a landmark finding for that time, he showed that the maximum ascent at 500 hPa was superimposed over the maximum surface downslope flow, indicative of the importance that lower-tropospheric vertical stretching and the associated horizontal stretching and cyclonic relative vorticity growth played in the lee-cyclogenesis process. Furthermore, Newton (1956) showed that differential lateral friction over sloping terrain east of the Rockies was as important as dynamically induced lower-tropospheric vertical stretching in the production of cyclonic vorticity during lee cyclogenesis.

Sanders (1988), linking back to Petterssen (1955), noted that surface cyclogenesis is primarily a response to the approach of a preexisting trough at upper levels. Accordingly, Sanders (1988) investigated the origin of preexisting disturbances over the Northern Hemisphere. His analysis was based on the behavior of 500-hPa troughs as identified by the evolution and configuration of the 552-dam geopotential height contour from twice-daily upper-level maps for a 9-yr period. In an indication of the importance of major mountain barriers over the Northern Hemisphere, Sanders (1988) found that the two primary centers where trough births exceeded trough deaths were located over and downstream of the Rocky Mountains and the Tibetan Plateau whereas a weak maximum of trough deaths over trough births was found about 1000 km upstream of the Rocky Mountains and the Tibetan Plateau. Thus, the maxima of lee cyclogenesis appear to be connected, at least in part, to the formation of mobile short-wave troughs in the jet stream.

b. The dynamics of storm tracks

The release of potential energy by upward and poleward transport of warm air through baroclinic instability is the fundamental mechanism behind the formation and growth of transient baroclinic eddies that compose the storm track and whose surface manifestation includes cyclones (section 3). Baroclinicity is a measure for the growth potential of baroclinic eddies and is proportional to the meridional temperature gradient and inversely proportional to the effective static stability taking into account the effects of latent heat release (e.g., Charney
Latent heating is asymmetrically linked to the vertical winds with heating occurring only in ascent. Because of latent heating, the effective stability is reduced relative to the dry static stability. For example, at 50° latitude in both hemispheres, effective stability is about 60% of the dry static stability (O’Gorman 2011), an indication that latent heating affects the dynamics of individual eddies. The jet is also maintained against surface friction by momentum fluxes (e.g., Lau and Holopainen 1984; Chang et al. 2002; Hartmann 2007; Shaw et al. 2016). The baroclinic eddies converge momentum into the upper-level jet during the final nonlinear stage of their life cycle (e.g., Thorncroft et al. 1992). The eddy momentum fluxes are not constant over time and depend on the location of the jet. A positive feedback exists because the meridional location of the jet affects the shape of the high-frequency eddies. The shape of these eddies determines the direction of the associated momentum fluxes, which in turn affects the meridional position of the jet stream (e.g., Gerber and Vallis 2007, 2009; Rivière 2009; Barnes and Hartmann 2010). Cyclonically breaking waves are favored with a more equatorward jet stream, and the momentum fluxes associated with these cyclonically breaking waves keep the jet in its equatorward position. The opposite is true for a poleward-shifted jet and anticyclonic wave breaking. Thus, this feedback results in the persistence of the meridional jet and storm-track position on medium-range to subseasonal time scales.

These maxima in the momentum fluxes are located downstream of the maxima in the heat fluxes. A simple interpretation of this spatial relationship is that it is a direct representation of an idealized baroclinic life cycle propagating eastward. The idealized life cycle of a baroclinic wave is characterized by strong low-level poleward temperature fluxes during the early stage of the life cycle and upper-level momentum fluxes into the jet during the final stage of the life cycle (e.g., Thorncroft et al. 1993). However, this simple explanation falls short of the complexity of real-life storm tracks where baroclinic eddies are embedded in coherent wave packets that consist of several eddies [review by Wirth et al. (2018)]. The wave packets propagate with an eastward group velocity that exceeds the eastward phase velocity of individual eddies, and there is downstream transfer of energy from one eddy to the next eddy within the wave packets (e.g., Simmons and Hoskins 1979; Orlanski and Katzfey 1991; Chang 1993; Orlanski and Chang 1993), a process called downstream development.

In addition to the dry dynamics discussed above, diabatic processes, and particularly latent heating, shape both cyclone and storm-track dynamics. Latent heating in the midlatitudes is strongest in baroclinic eddies (Sutcliffe 1951) and hence within the storm tracks. More specifically, latent heating occurs in the warm conveyor belts of extratropical cyclones (e.g., Harrold 1973; Carlson 1980; Browning 1990; Browning and Roberts 1996; Wernli 1997; Wernli and Davies 1997; Joos and Wernli 2012; Pfahl et al. 2014). Such latent heating affects the structure of cyclones (e.g., Danard 1964) through low-level diabatic PV production (e.g., Reed et al. 1992; Davis et al. 1993), resulting in a moderate to strong correlation between cyclone intensification rate and the strength of warm conveyor belts, as measured by the number and mass of the warm conveyor belt trajectories associated with the cyclone at low levels during its strongest intensification (Binder et al. 2016).

Latent heating is also part of the answer to the question posed by Hoskins and Valdes (1990): Why do storm tracks exist? Baroclinic eddies feed on baroclinicity and, by transporting heat northward during their life cycle, they act to destroy the baroclinicity. As a consequence, the next eddy would be expected to form in a different location where the baroclinicity is still high, arguing against the formation of a coherent storm track. So, which processes contribute to the self-maintenance of the storm track? Hoskins and Valdes (1990) found that thermal forcing, predominantly via latent heating associated with the baroclinic eddies, is the most important factor in maintaining the baroclinicity and hence the storm tracks. Sensible heat fluxes restore most of the baroclinicity near the surface (e.g., Hotta and Nakamura 2011), whereas latent heating dominates in the free troposphere (e.g., Papritz and Spengler 2015). Then, vorticity fluxes associated with the baroclinic eddies promote convergent flow in the entrance region of the storm tracks (Hoskins et al. 1983) that strengthens the temperature gradient and thereby counters the effects of the temperature fluxes by the eddies (Hoskins and Valdes 1990). In addition, energy fluxes by stationary planetary-scale waves increase the baroclinicity in the storm-track entrance region (e.g., Lee and Mak 1996; Kaspi and Schneider 2013). Last, the low-level flow induced by the eddies exerts wind stresses on the oceans that help maintain the warm boundary currents and thereby baroclinicity (Hoskins and Valdes 1990).

Diabatic processes also influence storm-track variability. There are distinct differences between the eastern and the western North Atlantic. Over the western Atlantic, the maxima in sensible and latent heating remain anchored to areas of strong sea surface temperature gradients, whereas in the eastern Atlantic the areas of maximum latent heat release shift meridionally in tandem with the storm track and hence help to maintain the anomalous storm-track positions (Woollings et al. 2016).
The interdependency between the creation of baroclinicity by diabatic processes and destruction of baroclinicity by the release of baroclinic instability on subseasonal time scales may explain oscillations of storm-track intensity on these time scales (e.g., Ambaum and Novak 2014; Novak et al. 2017). Besides latent heating, other diabatic processes (e.g., cloud radiative processes) affect the storm tracks, as well (e.g., Shaw et al. 2016).

Storm tracks extend longitudinally beyond the maximum surface baroclinicity as a result of downstream development and there is no obvious end to this downstream extension. So which factors control the downstream extent of the storm tracks? First, increased surface roughness and drag over the downstream continents results in energy dissipation (Chang and Orlanski 1993). However, zonally confined storm tracks form continents (Kaspi and Schneider 2011); therefore, other processes must be involved. Indeed, stationary planetary-scale waves destroy the baroclinicity downstream of storm tracks ( Hoskins and Valdes 1990; Kaspi and Schneider 2011, 2013). These stationary planetary-scale waves arise from orography and warm ocean currents ( Held et al. 2002). The Atlantic storm track’s extent and southwest–northeast tilt are strongly influenced by the geometry and major orography of North America (e.g., Brayshaw et al. 2009; Gerber and Vallis 2009) and by Atlantic SST gradients ( Brayshaw et al. 2011). In addition, the weaker background flow in the storm-track exit areas gives rise to Rossby wave breaking and thereby the termination of baroclinic-wave packets ( Swanson et al. 1997).

Having considered the large-scale aspects of how the jet stream affects extratropical cyclones, we now transition to scales smaller than the cyclone, to investigate how the dynamics and kinematics of the cyclone itself create structures called fronts that regulate the distribution of heat, moisture, winds, and precipitation within extratropical cyclones.

5. Fronts and frontogenesis

Characteristic features of extratropical cyclones are the baroclinic zones, or fronts. Fronts are characterized by vertically sloping transition zones in the thermal and wind fields ( Keyser 1986). The study of fronts, and the process by which they form (i.e., frontogenesis), was energized by the Bergen School in the wake of World War I. Later, dovetailing observational, theoretical, and diagnostic research encapsulated in Fig. 16-2 has resulted in substantial growth in the dynamical understanding of frontogenesis, as well as in the systematic refinement of conceptual models of fronts. This section documents and discusses major advances in understanding fronts and frontogenesis during the past 100 years, with a focus on the synergy between observational, theoretical, and diagnostic frontal research.

a. Observations of fronts

In their development of polar-front theory, the Bergen School astutely integrated sparse quantitative and visual observations to construct a conceptual model for the three-dimensional thermal structure of a midlatitude cyclone ( Bjerknes 1919; Bjerknes and Solberg 1921, 1922). The polar front constituted a substantial component of the Norwegian cyclone model and was hypothesized to encircle the globe and to separate polar air masses at high latitudes from tropical air masses at low latitudes within the Northern Hemisphere. The temperature contrast associated with the polar front subsequently represented the energy source for cyclogenesis and the concomitant development of a frontal wave ( section 3b).

The evolution of the frontal wave within the Norwegian cyclone model featured distinct warm- and cold-frontal boundaries that were positioned at the leading edge of advancing warm and cold currents of air, respectively, within the circulation of the cyclone ( Fig. 16-4; Bjerknes and Solberg 1921, 1922). The vertical structure of warm and cold fronts was characterized by across-front gradients in vertical motion and precipitation, as well as zero-order, tropospheric-deep discontinuities in temperature and alongfront wind that sloped over the colder air. During the latter stages of cyclogenesis, the Norwegian cyclone model depicted advancing cold air behind the cold front catching up to the warm front to produce an occluded front ( Fig. 16-8). Both warm-type and cold-type occluded fronts were proposed as complementary descriptions of the vertical temperature structure associated with an occluded front, with the prevailing type governed by the temperature of the air mass behind the cold front relative to the temperature of the air mass ahead of the warm front.

The introduction of routine upper-air observations during the 1930s ushered in an era of revision to polar-front theory. In particular, detailed analyses of the vertical structure of fronts consistently demonstrated that fronts were characterized by sloping transition zones in the thermal and wind fields, rather than the zero-order discontinuities proposed by the Bergen School (e.g., Bjerknes and Palmén 1937; Palmén and Newton 1948).

The rising tide of observations challenging polar-front theory fed the discontent of Fred Sanders and Richard Reed, who lamented “the nearly blind acceptance by
many meteorologists” of polar-front theory during the midtwentieth century (Reed 2003, p. 3). Of particular interest to Sanders and Reed was the notion that fronts may not be tropospheric-deep entities, as implied by polar-front theory. To this end, Sanders (1955) analyzed surface and upper-air observations during the development of a strong surface-based frontal zone over the south-central United States (Fig. 16-14). Consistent with previous analyses, Sanders identified a frontal zone that featured an intense temperature contrast near the surface, strong cyclonic relative vorticity, and enhanced static stability. A novel aspect of the Sanders (1955) analysis, however, was that the frontal zone was confined exclusively within the lower troposphere. In contrast, Reed and Sanders (1953), Newton (1954), and Reed (1955) identified zones of intense thermal contrast and cyclonic wind shear that were confined solely within the middle and upper troposphere (Fig. 16-15). The observation of frontal structures in the middle and upper troposphere laid the foundation for the concept of upper-level

**FIG. 16-14.** (a) Surface observations at 0330 UTC 18 Apr 1953, with sea level pressure contoured with thin solid lines every 6 hPa and the boundaries of the surface frontal zone contoured by the thick solid lines. (b) Cross section along A–A', as indicated in (a), at 0300 UTC 18 Apr 1953, with potential temperature contoured by thin solid lines every 5 K, the horizontal wind component normal to the cross section contoured by dashed lines every 5 m s⁻¹ (with positive values representing flow into the cross section), and the boundaries of the frontal zone contoured by thick black lines. The figure and caption are adapted from Sanders (1955, his Figs. 2 and 9).

**FIG. 16-15.** (a) Observed 500-hPa temperature, dewpoint, and wind at 0300 UTC 15 Dec 1953, with geopotential height [thin solid lines every 200 ft (61 m)], temperature (dashed lines every 4°C), and the boundaries of the frontal zone (thick red lines). (b) Cross section along B–B', as indicated in (a), of geostrophic wind speed normal to the cross section (thin solid lines every 20 m s⁻¹), potential temperature (dashed lines every 10 K), the tropopause (thick solid line), and the jet core (indicated by the red “J”). The figure and caption are adapted from Reed (1955, his Figs. 7 and 13).
frontogenesis, a process by which a wedge of stratospheric air is extruded into the middle troposphere to produce a tropopause fold (e.g., Keyser and Shapiro 1986, 454–458).

In contrast to surface frontogenesis, which Sanders primarily attributed to horizontal deformation, upper-level frontogenesis resulted from across-front gradients in vertical motion that positioned the most intense subsidence on the warm side of the developing frontal zone (e.g., Reed and Sanders 1953; Reed 1955; Bosart 1970). This description of upper-level frontogenesis countered the conventional wisdom that the tropopause was a material surface separating stratospheric and tropospheric air, because concomitant tropopause folding represented a process that was conducive to stratosphere–troposphere exchange (e.g., Danielsen 1964, 1968; Shapiro 1978, 1980). Considered together, the analyses by Sanders, Reed, and Newton established the notion that surface and upper-level fronts were distinct structural and dynamical entities. Consequently, their analyses represented profound breaks from polar-front theory and served as benchmarks against which future theoretical and diagnostic analyses of fronts would be compared.

Advances in observational capabilities during the latter half of the twentieth century spurred further revisions to polar-front theory. For example, the advent of satellite technology provided greater detail on the distribution of clouds and precipitation within midlatitude cyclones. Carlson (1980) was among the first to synthesize satellite observations through the construction of a conceptual model that expanded upon the Norwegian cyclone model and included the three-dimensional movement of airstreams within a mature, steady-state cyclone (Fig. 16-16). Although providing a common language for describing the airstreams in midlatitude cyclones, further refinements of Carlson’s (1980) model would occur over future years with the advent of air-parcel trajectory calculations (e.g., Whitaker et al. 1988; Kuo et al. 1992; Mass and Schultz 1993; Schultz and Mass 1993; Reed et al. 1994; Wernli and Davies 1997; Wernli 1997; Schultz 2001; Schemm et al. 2013; Schemm and Wernli 2014; Slater et al. 2015, 2017).

Observations from case studies and intensive field campaigns also demonstrated that the evolution and distribution of fronts within midlatitude cyclones did not always adhere to the model conceptualized by the Bergen School. These observations illuminated some of the synoptic-scale and mesoscale frontal structures that differed from those incorporated in the original polar-front theory (Table 16-1). Observations of occluded cyclones have also suggested that warm-type and cold-type occlusions are more accurately governed by the static stability rather than the temperature of the air mass behind the cold front relative to the air mass ahead of the warm front (Stoelinga et al. 2002). One result of this alternative perspective on occluded fronts is that cold-type occlusions would rarely be observed (e.g., Schultz and Mass 1993; Schultz and Vaughan 2011; Schultz et al. 2014).

During the last quarter of the twentieth century, the modification of fronts and their associated narrow precipitation bands by topography at coastlines and mountains became a special focus of observational investigation in the flourishing field of mesoscale meteorology. Examples of investigations from three continents include 1) the Cyclonic Extratropical Storms project (CYCLES) that studied fronts along the west coast of North America (e.g.,

![Fig. 16-16. Schematic composite of the three-dimensional airflow through a midlatitude cyclone. Heavy solid streamlines depict the warm conveyor belt, dashed lines represent the cold conveyor belt (drawn dotted where it lies beneath the warm conveyor belt or dry airstream), and the dot-dashed line represents flow originating at midlevels within the tropics. Thin solid streamlines pertain to dry air that originates at upper levels west of the trough. Thin solid lines denote the heights of the airstreams (hPa) and are approximately normal to the direction of the respective air motion (isobars are omitted for the cold conveyor belt where it lies beneath the warm conveyor belt or beneath the jet streamflow). Scallop marks the regions of dense clouds at the upper and middle levels, stippling indicates sustained precipitation, and streaks denote thin cirrus. Small dots with tails mark the edge of the low-level stratus. The major upper-tropospheric jet streams are labeled “Jet” and “Dry Tongue Jet.” The limiting streamline for the warm conveyor belt is labeled “LSW.” Warm and cold fronts are identified by the thick red and blue lines, respectively, and coincide with the boundaries between airstreams. The figure and caption are adapted from Carlson (1980, his Fig. 9).](http://journals.ametsoc.org/doi/abs/10.1175/AMSMONOGRAPHS-D-18-0015.1?journalCode=mono)
TABLE 1. Examples of observed frontal structures that differ from the Norwegian cyclone model.

<table>
<thead>
<tr>
<th>Frontal structure</th>
<th>Selected citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Katafronts and anafronts</td>
<td>Bergeron (1937), Sansom (1951), and Browning (1990, 1999)</td>
</tr>
<tr>
<td>Backdoor fronts</td>
<td>Carr (1951) and Bosart et al. (1973)</td>
</tr>
<tr>
<td>Coastal fronts</td>
<td>Bosart et al. (1972, 2008) and Bosart (1975, 1981)</td>
</tr>
<tr>
<td>Prefrontal troughs and wind-shift lines</td>
<td>Schultz (2005)</td>
</tr>
</tbody>
</table>

Hobbs et al. 1980, 2) the British–French campaign FRONT85 that studied Atlantic fronts landfalling on western Europe (Thorpe and Clough 1991), 3) a 5-yr program called Fronts and Orography centered in southern Germany and neighboring countries (e.g., Volkert et al. 1991; Egger and Hoinka 1992), and 4) the Cold Fronts Research Programme that studied fronts over the Southern Ocean impinging on southeastern Australia (e.g., Ryan et al. 1985). These investigations provided close-up looks into the three-dimensional structure of precipitation and moisture within frontal zones and, thus, research datasets for prototypical simulations of frontal dynamics. In particular, the latter two investigations helped to quantify the often frontogenetical forcing of mountain masses caused by low-level blocking of the airflow in the vicinity of the European Alps and Australian Alps, respectively.

b. Theory of fronts

As observations further revealed the characteristics of frontal zones, theoretical studies sought to reproduce and interpret their development within idealized frameworks. The conceptualization of fronts as transition zones coincided with the advent of baroclinic instability theory (e.g., Charney 1947; Eady 1949) and quasigeostrophic theory (section 3b). An important shift represented by these theories was that intense fronts were not a necessary precursor to cyclogenesis, but rather that intense fronts developed as a consequence of cyclogenesis. This shift placed emphasis on the role of horizontal deformation in subsequent theoretical studies of frontogenesis.

In a quasigeostrophic framework, frontogenesis is driven by geostrophic deformation that acts to intensify the horizontal temperature gradient. This process is subsequently accompanied by the development of an across-front ageostrophic circulation that arises to preserve thermal wind balance (e.g., Hoskins et al. 1978). Studies employing two-dimensional quasigeostrophic prognostic models were successful in producing frontal zones with some fidelity (e.g., Stone 1966; Williams 1968, 1972; Williams and Plotkin 1968). However, quasigeostrophic solutions featured a number of deficiencies relative to observations. Namely, frontogenesis occurred too slowly at the surface, the frontal zone did not exhibit a vertical tilt, the frontal zone featured areas of both cyclonic and anticyclonic relative vorticity, and the frontal zone exhibited static instability.

The deficiencies of quasigeostrophic solutions are understood by recognizing that fronts are synoptic scale in length but mesoscale in width. Consequently, whereas the alongfront wind is approximately geostrophic for straight fronts, the across-front wind can be substantially ageostrophic. In what would become a pioneering contribution to semigeostrophic theory (Hoskins 1975), Sawyer (1956) modified the quasigeostrophic solution for the across-front ageostrophic circulation to include across-front ageostrophic advections of temperature and alongfront wind and vertical advections of perturbation temperature and alongfront wind. However, Sawyer’s solution was limited in that it only considered the frontogenetical effect of geostrophic confluence. Eliassen (1962) expanded upon Sawyer’s work to include the frontogenetical effects of geostrophic horizontal shear and differential diabatic heating in his solution for the across-front ageostrophic circulation, diagnosed from what would later be termed the Sawyer–Eliassen equation. The across-front ageostrophic circulations diagnosed from the Sawyer–Eliassen equation in regions of geostrophic confluence and horizontal shear (Fig. 16-17) represented a significant theoretical advance in the attempt to better understand the dynamics of frontogenesis and to reproduce the characteristics of observed fronts.

Two-dimensional semigeostrophic prognostic models, which included across-front ageostrophic and vertical advections of temperature and alongfront wind, demonstrated a greater ability than their quasigeostrophic counterparts to reproduce observed surface and upper-level fronts under adiabatic and frictionless conditions (e.g., Hoskins 1971, 1972; Hoskins and Bretherton 1972). In particular, the semigeostrophic models identified frontogenesis as a two-step process, in which geostrophic deformation strengthens the horizontal temperature gradient and subsequently accompanied by the development of an across-front ageostrophic circulation that arises to preserve thermal wind balance (e.g., Hoskins et al. 1978). Studies employing two-dimensional quasigeostrophic prognostic models were successful in producing frontal zones with some fidelity (e.g., Stone 1966; Williams 1968, 1972; Williams and Plotkin 1968). However, quasigeostrophic solutions featured a number of deficiencies relative to observations. Namely, frontogenesis occurred too slowly at the surface, the frontal zone did not exhibit a vertical tilt, the frontal zone featured areas of both cyclonic and anticyclonic relative vorticity, and the frontal zone exhibited static instability.

The deficiencies of quasigeostrophic solutions are understood by recognizing that fronts are synoptic scale in length but mesoscale in width. Consequently, whereas the alongfront wind is approximately geostrophic for straight fronts, the across-front wind can be substantially ageostrophic. In what would become a pioneering contribution to semigeostrophic theory (Hoskins 1975), Sawyer (1956) modified the quasigeostrophic solution for the across-front ageostrophic circulation to include across-front ageostrophic advections of temperature and alongfront wind and vertical advections of perturbation temperature and alongfront wind. However, Sawyer’s solution was limited in that it only considered the frontogenetical effect of geostrophic confluence. Eliassen (1962) expanded upon Sawyer’s work to include the frontogenetical effects of geostrophic horizontal shear and differential diabatic heating in his solution for the across-front ageostrophic circulation, diagnosed from what would later be termed the Sawyer–Eliassen equation. The across-front ageostrophic circulations diagnosed from the Sawyer–Eliassen equation in regions of geostrophic confluence and horizontal shear (Fig. 16-17) represented a significant theoretical advance in the attempt to better understand the dynamics of frontogenesis and to reproduce the characteristics of observed fronts.

Two-dimensional semigeostrophic prognostic models, which included across-front ageostrophic and vertical advections of temperature and alongfront wind, demonstrated a greater ability than their quasigeostrophic counterparts to reproduce observed surface and upper-level fronts under adiabatic and frictionless conditions (e.g., Hoskins 1971, 1972; Hoskins and Bretherton 1972). In particular, the semigeostrophic models identified frontogenesis as a two-step process, in which geostrophic deformation strengthens the horizontal temperature gradient and subsequently accompanied by the development of an across-front ageostrophic circulation that arises to preserve thermal wind balance (e.g., Hoskins et al. 1978). Studies employing two-dimensional quasigeostrophic prognostic models were successful in producing frontal zones with some fidelity (e.g., Stone 1966; Williams 1968, 1972; Williams and Plotkin 1968). However, quasigeostrophic solutions featured a number of deficiencies relative to observations. Namely, frontogenesis occurred too slowly at the surface, the frontal zone did not exhibit a vertical tilt, the frontal zone featured areas of both cyclonic and anticyclonic relative vorticity, and the frontal zone exhibited static instability.

The deficiencies of quasigeostrophic solutions are understood by recognizing that fronts are synoptic scale in length but mesoscale in width. Consequently, whereas the alongfront wind is approximately geostrophic for straight fronts, the across-front wind can be substantially ageostrophic. In what would become a pioneering contribution to semigeostrophic theory (Hoskins 1975), Sawyer (1956) modified the quasigeostrophic solution for the across-front ageostrophic circulation to include across-front ageostrophic advections of temperature and alongfront wind and vertical advections of perturbation temperature and alongfront wind. However, Sawyer’s solution was limited in that it only considered the frontogenetical effect of geostrophic confluence. Eliassen (1962) expanded upon Sawyer’s work to include the frontogenetical effects of geostrophic horizontal shear and differential diabatic heating in his solution for the across-front ageostrophic circulation, diagnosed from what would later be termed the Sawyer–Eliassen equation. The across-front ageostrophic circulations diagnosed from the Sawyer–Eliassen equation in regions of geostrophic confluence and horizontal shear (Fig. 16-17) represented a significant theoretical advance in the attempt to better understand the dynamics of frontogenesis and to reproduce the characteristics of observed fronts.

Two-dimensional semigeostrophic prognostic models, which included across-front ageostrophic and vertical advections of temperature and alongfront wind, demonstrated a greater ability than their quasigeostrophic counterparts to reproduce observed surface and upper-level fronts under adiabatic and frictionless conditions (e.g., Hoskins 1971, 1972; Hoskins and Bretherton 1972). In particular, the semigeostrophic models identified frontogenesis as a two-step process, in which geostrophic deformation strengthens the horizontal temperature gradient and subsequently accompanied by the development of an across-front ageostrophic circulation that arises to preserve thermal wind balance (e.g., Hoskins et al. 1978). Studies employing two-dimensional quasigeostrophic prognostic models were successful in producing frontal zones with some fidelity (e.g., Stone 1966; Williams 1968, 1972; Williams and Plotkin 1968). However, quasigeostrophic solutions featured a number of deficiencies relative to observations. Namely, frontogenesis occurred too slowly at the surface, the frontal zone did not exhibit a vertical tilt, the frontal zone featured areas of both cyclonic and anticyclonic relative vorticity, and the frontal zone exhibited static instability.

The deficiencies of quasigeostrophic solutions are understood by recognizing that fronts are synoptic scale in length but mesoscale in width. Consequently, whereas the alongfront wind is approximately geostrophic for straight fronts, the across-front wind can be substantially ageostrophic. In what would become a pioneering contribution to semigeostrophic theory (Hoskins 1975), Sawyer (1956) modified the quasigeostrophic solution for the across-front ageostrophic circulation to include across-front ageostrophic advections of temperature and alongfront wind and vertical advections of perturbation temperature and alongfront wind. However, Sawyer’s solution was limited in that it only considered the frontogenetical effect of geostrophic confluence. Eliassen (1962) expanded upon Sawyer’s work to include the frontogenetical effects of geostrophic horizontal shear and differential diabatic heating in his solution for the across-front ageostrophic circulation, diagnosed from what would later be termed the Sawyer–Eliassen equation. The across-front ageostrophic circulations diagnosed from the Sawyer–Eliassen equation in regions of geostrophic confluence and horizontal shear (Fig. 16-17) represented a significant theoretical advance in the attempt to better understand the dynamics of frontogenesis and to reproduce the characteristics of observed fronts.
and induces an across-front ageostrophic circulation. This circulation further strengthens the horizontal temperature gradient, resulting in a contraction of the width of the frontal zone at the surface, and accounts for the vertical tilt of the frontal zone (Fig. 16-18). Two-dimensional semigeostrophic models forced by geostrophic confluence (Fig. 16-19) and horizontal shear, as well as their primitive equation counterparts, also affirmed the role of subsidence during upper-level frontogenesis and the concomitant production of a tropopause fold (e.g., Hoskins 1972; Keyser and Pecnick 1985; Reeder and Keyser 1988).

Despite the success of two-dimensional semigeostrophic models in reproducing aspects of the observed structure of fronts, idealized simulations of midlatitude cyclones using three-dimensional primitive equation models revealed that the semigeostrophic equations inaccurately represented the structure of fronts relative to the primitive equations for cases in which the ratio of the ageostrophic relative vorticity to the Coriolis parameter was large (e.g., Snyder et al. 1991; Rotunno et al. 1994). In response to the deficiencies of semigeostrophic theory, Muraki et al. (1999) derived a first-order correction to quasigeostrophic theory that extended the conceptual simplicity of quasigeostrophic theory to higher orders of Rossby number. The subsequent application of this first-order correction resulted in frontal structure that aligned more favorably with that simulated in primitive equation models (Rotunno et al. 2000). Three-dimensional primitive equation models also reproduced canonical surface and upper-level frontal structures observed within midlatitude cyclones. In particular, Davies et al. (1991) and Thorncroft et al. (1993) showed that the character of the background barotropic across-jet shear differentiated between cyclones that developed following the Norwegian cyclone model and the more-recent Shapiro–Keyser cyclone model (Shapiro and Keyser 1990; section 6d). The degree of along-jet shear in the form of confluence or diffluence was also shown to differentiate between the two models (e.g., Schultz et al. 1998; Schultz and Zhang 2007).

The addition of diabatic and frictional processes into idealized modeling frameworks further reconciled
idealized simulations of frontal structure with observations. For instance, a number of idealized studies illuminated the influence of condensational heating and differential surface heating on frontogenesis (e.g., Szeto et al. 1988a,b; Huang and Emanuel 1991; Koch et al. 1995; Szeto and Stewart 1997) and on the modulation of the structure and intensity of across-front ageostrophic circulations (e.g., Baldwin et al. 1984; Hsie et al. 1984; Mak and Bannon 1984; Thorpe and Emanuel 1985). Furthermore, surface fluxes, friction, and turbulent mixing within the planetary boundary layer were found to influence the structure of fronts within idealized simulations (e.g., Keyser and Anthes 1982; Cooper et al. 1992; Hines and Mechoso 1993; Rotunno et al. 1998; Tory and Reeder 2005; Reeder and Tory 2005; Schultz and Roebber 2008; Sinclair and Keyser 2015).

Last, the paradigm of PV thinking has provided a contemporary theoretical framework from which to examine surface and upper-level fronts (e.g., Hoskins et al. 1988a,b; Huang and Emanuel 1991; Koch et al. 1995; Szeto and Stewart 1997) and on the modulation of the structure and intensity of across-front ageostrophic circulations (e.g., Baldwin et al. 1984; Hsie et al. 1984; Mak and Bannon 1984; Thorpe and Emanuel 1985). Furthermore, surface fluxes, friction, and turbulent mixing within the planetary boundary layer were found to influence the structure of fronts within idealized simulations (e.g., Keyser and Anthes 1982; Cooper et al. 1992; Hines and Mechoso 1993; Rotunno et al. 1998; Tory and Reeder 2005; Reeder and Tory 2005; Schultz and Roebber 2008; Sinclair and Keyser 2015).

Last, the paradigm of PV thinking has provided a contemporary theoretical framework from which to examine surface and upper-level fronts (e.g., Hoskins et al. 1988a,b; Huang and Emanuel 1991; Koch et al. 1995; Szeto and Stewart 1997) and on the modulation of the structure and intensity of across-front ageostrophic circulations (e.g., Baldwin et al. 1984; Hsie et al. 1984; Mak and Bannon 1984; Thorpe and Emanuel 1985).

zones of enhanced potential temperature gradients on the dynamic tropopause (e.g., Morgan and Nielsen-Gammon 1998) and may precede the development of coherent tropopause disturbances (e.g., Pyle et al. 2004; Cavallo and Hakim 2010). In the PV framework, the development of upper-level fronts may be alternatively described in terms of PV frontogenesis (Davies and Rossa 1998), which corresponds to increases in the magnitude of the PV gradient on an isentropic surface, and foldogenesis (Wandishin et al. 2000), which corresponds to increases in the slope of the dynamic tropopause.

c. Diagnosis of fronts

Diagnostic studies of fronts have provided a bridge between observations and theory by leveraging a suite of quantitative tools to investigate the structure and dynamics of fronts. The two-dimensional Petterssen frontogenesis equation (Petterssen 1936, 1956, 200–202) served as a seminal breakthrough by providing a quantitative basis for diagnosing frontogenesis. In the context of this equation, frontogenesis is defined as the Lagrangian rate of change of the magnitude of the horizontal temperature gradient and is forced by horizontal convergence and deformation in the absence of vertical motion and diabatic effects. Reed and Sanders (1953), Newton (1954), and Sanders (1955) were among the first to calculate the Lagrangian rate of change of the across-front temperature gradient in their respective diagnoses of upper-level and surface fronts by applying a related
form of the Petterssen frontogenesis equation (Miller 1948; discussed further in Schultz 2015).

Applications of the Sawyer–Eliassen equation to idealized and analyzed cases have further illuminated the dynamics of frontogenesis and across-front ageostrophic circulations. Todsen (1964) provided the first known application of the Sawyer–Eliassen equation to an observed front and quantified the influence of latent heat release in strengthening the across-front ageostrophic circulation. An advance in conceptual understanding of upper-level frontogenesis resulted from Shapiro’s (1981) application of the Sawyer–Eliassen equation. In particular, Shapiro demonstrated that alongfront cold-air advection in the presence of ageostrophic horizontal shear shifted the across-front ageostrophic circulation relative to the upper-level jet axis so as to force subsidence on the warm side of the developing upper-level front. Termed the Shapiro effect by Rotunno et al. (1994), this shift highlighted the role of differential subsidence during upper-level frontogenesis originally discussed by Reed and Sanders (1953) and became a substantial topic of interest in subsequent diagnostic examinations of upper-level fronts (e.g., Newton and Trevisan 1984; Keyser and Pecnick 1985; Rotunno et al. 1994; Schultz and Doswell 1999; Schultz and Sanders 2002; Lang and Martin 2010, 2013a; Schultz 2013). Applications of the Sawyer–Eliassen equation have also highlighted the influence of uncoupled (Fig. 16-20) and coupled (Fig. 16-21) upper- and lower-tropospheric across-front ageostrophic circulations on convective initiation, as well as on cyclogenesis and poleward moisture transport (e.g., Shapiro 1982; Uccellini et al. 1985; Hakim and Keyser 2001; Winters and Martin 2014).

Despite the diagnostic utility of the Sawyer–Eliassen equation, its rigorous application is restricted to across-front ageostrophic circulations in straight fronts. The Q vector (e.g., Hoskins et al. 1978; Hoskins and Pedder 1980) is not subject to this restriction, and thus its introduction provided an important tool for diagnosing three-dimensional ageostrophic circulations in straight and curved fronts. The diagnostic power of the Q vector becomes apparent in a framework where the Q vector is partitioned into across- and along-isotherm components (e.g., Keyser et al. 1992). Within this framework, the across-isotherm component of the Q vector reduces to the geostrophic form of the two-dimensional Petterssen frontogenesis equation, whereas the along-isotherm component of the Q vector diagnoses changes in the orientation of the temperature gradient. The latter component, in particular, provided insight into the wrap-up process associated with the occlusion of midlatitude cyclones (e.g., Martin 1999, 2006).

The psi vector (Keyser et al. 1989) provided a tool complementary to the Q vector for diagnosing three-dimensional ageostrophic circulations in straight and curved fronts. Specifically, the psi vector represents the irrotational part of the three-dimensional ageostrophic circulation, and its application has demonstrated considerable explanatory power in the context of upper-level frontogenesis by allowing the separation of the irrotational ageostrophic circulation into across- and alongfront components. A key result from Keyser et al.’s (1989) application of the psi vector was the notion that subsidence in the vicinity of developing upper-level fronts featured both across-front and alongfront components. The alongfront component of subsidence occurring in conjunction with upper-level frontogenesis has received additional consideration by Mudrick (1974) and Martin (2014). Building on the results of Mudrick (1974), Martin (2014) demonstrated that, within regions of geostrophic cold-air advection in the presence of cyclonic shear, the contribution to frontal tilting associated with alongfront subsidence induced by negative shear-vorticity advection by the thermal wind dominates the contribution associated with across-front subsidence induced by geostrophic frontogenesis.

Last, the application of PV inversion (e.g., Davis and Emanuel 1991) has provided insight into the dynamics of frontogenesis (e.g., Morgan 1999; Korner and Martin 2000), as well as into the dynamics of across-front ageostrophic circulations in the vicinity of upper-level fronts (e.g., Winters and Martin 2016, 2017). Furthermore, diabatically generated lower-tropospheric PV anomalies near fronts have been linked to enhanced alongfront moisture transport within the warm conveyor belt of midlatitude cyclones (e.g., Lackmann and Gyakum 1999; Lackmann 2002; Reeves and Lackmann 2004; Brennan et al. 2008; Joos and Wernli 2012; Lackmann 2013). This enhanced alongfront moisture transport can foster a positive feedback whereby the lower-tropospheric frontal structure can be strengthened in response to additional latent heat release.

d. Summary

Ignited by the advent of polar-front theory in the wake of World War I, scientific knowledge regarding fronts and frontogenesis has been characterized by a powerful synergy of observational, theoretical, and diagnostic research. This research has spurred revisions to polar-front theory to account for the variety of frontal structures and dynamics within the midlatitude atmosphere. The next section discusses conceptual models and addresses their utility in revealing classifications of the variety of midlatitude cyclones.
and their associated frontal structures and precipitation systems.

6. Conceptual models of cyclone and frontal evolution

One of the ways that meteorologists make sense of the variety of observed weather systems is through the construction of conceptual models, idealized schematics that represent common characteristics of weather systems. With the understanding that comes from cyclone and frontal dynamics, this section explores how the synthesis of these schematics has led to greater insight into the structure and dynamics of cyclones and their attendant fronts.

a. Norwegian cyclone model

As summarized in Bjerknes and Solberg (1922), the birthplace of the frontal cyclone is a nearly straight boundary, or polar front, separating cold easterly flow from warm westerly flow (Fig. 16-22a). This boundary bulges toward the cold air at the location of the incipient
low center, forming a frontal wave (Fig. 16-22b), which amplifies into an open-wave cyclone (Fig. 16-22c). As cold air moves cyclonically around the low center, the warm sector narrows (Fig. 16-22d), and eventually the cold front overtakes the warm front south of the low center, cutting off a pocket of warm-sector air, known as the warm-core seclusion (Fig. 16-22e). Eventually the warm sector disappears entirely, and the cyclone becomes occluded (Fig. 16-22f). Gradually, the occluded boundary dissipates and the cyclone becomes a symmetrical vortex of cold air (Fig. 16-22g), followed by death (Fig. 16-22h).

Modern textbooks for meteorologists and non-meteorologists still use elements of the Norwegian cyclone model, which is sometimes condensed into a four-stage conceptual model consisting of the initial frontal wave, open-wave cyclone, narrowing warm sector, and frontal occlusion, with the seclusion omitted (e.g., Schultz and Vaughan 2011). Bergeron (1959, p. 457) suggests that the seclusion was based on a hypothesis that was “better than the data by which it was achieved,” although modern observations and modeling confirm seclusion development during intense extratropical cyclogenesis through processes not envisioned by the Bergen School (e.g., Shapiro and Keyser 1990; Kuo et al. 1992; Galarneau et al. 2013). The Norwegian cyclone model also suggested explanations for cyclone development (section 3) and frontal precipitation processes (section 5).

b. Bergen School contributions through the mid-twentieth century

Bergen School meteorologists continued to refine knowledge of frontal cyclones after publication of the original Norwegian cyclone model (e.g., Bergeron 1959). By the middle of the twentieth century, these refinements included the following: 1) awareness that fronts are better regarded as discontinuities in temperature gradient rather than temperature (section 5a); 2) identification of frontolysis along the cold front near the low center during the open-wave phase, a predecessor to what is referred to today as the frontal fracture; 3) recognition of the three-dimensional structure of cyclones, including the role of upper-level waves; and 4) knowledge
FIG. 16-22. Life cycle of the ideal cyclone: (a) initial phase, (b) incipient cyclone and frontal wave, (c) amplification of the warm wave (open-wave cyclone), (d) narrowing in of the warm tongue/sector, (e) warm-core seclusion, (f) occluded cyclone, (g) cold-air vortex, and (h) death. The figure is from Bjerknes and Solberg (1922, their Fig. 2) and is provided through the courtesy of the Norwegian Geophysical Society.
of the potential for a secondary surface trough to develop in the polar airstream behind the low during cases of extreme cyclogenesis, with a back-bent occlusion coincident with the trough near the low center (e.g., Bergeron 1937; Godske et al. 1957, their chapters 14 and 15).

Godske et al. (1957) provided a revised conceptual model of a strong occluded cyclone at maximum intensity (Fig. 16-23) based largely on work by Bergen School meteorologists Tor Bergeron, Jacob Bjerknes, and Erik Palmén (Bjerknes 1930; Bergeron 1937, 1959). They illustrated the occlusion as warm type and included an upper cold front, which is coincident with a tongue of warm air aloft, sometimes called a trowal (i.e., trough of warm air aloft; Crocker et al. 1947; Godson 1951; Penner 1955; Galloway 1958, 1960; Martin 1999). The upper cold front may have a stronger temperature contrast than the surface occluded front and demarcates an important transition in cloud and precipitation. The secondary trough and back-bent occlusion extend into the polar airstream behind the low center, with the latter identified with cold-front symbols. In Norway in the 1960s, meteorologists were trained to watch for strong winds [termed the sting jet by Browning (2004)] associated with this “poisonous tail” of the back-bent occlusion (Grønnås 1995), also known as the bent-back occlusion, retrograde occlusion, back-bent (or bent-back) front (Bjerknes 1930; Bergeron 1937). Research on the origin and mechanisms of strong winds along the bent-back front in Shapiro–Keyser cyclones has been an active topic for debate over the past 15 years (e.g., Browning 2004; Clark et al. 2005; Gray et al. 2011; Schultz and Sienkiewicz 2013; Smart and Browning 2014; Slater et al. 2015, 2017; Coronel et al. 2016; Schultz and Browning 2017; Volonté et al. 2018).

c. Beyond the Bergen School

Surface and upper-air observations, satellite remote sensing, ground-based remote sensing, numerical modeling, and intensive field programs have transformed our understanding of the life cycle of extratropical cyclones since the middle of the twentieth century. In particular, modern observational and numerical modeling capabilities show that

- fronts are often a consequence of cyclogenesis rather than the cause, with frontal zones better regarded as regions of active frontogenesis rather than semipermanent phenomena (Phillips 1956; Reed 1990),
- upper-level and surface-based fronts are not necessarily structurally continuous through the troposphere and respond to different dynamical processes (Keyser 1986; Reed 1990; Shapiro and Keyser 1990),
- cyclogenesis is better viewed as a consequence of baroclinic instability and the interaction of upper-level, surface, and diabatically generated PV anomalies rather than frontal instabilities (e.g., Charney 1947; Eady 1949; Hoskins et al. 1985; Davis and Emanuel 1991), and
- pathways for extratropical cyclone development include not only cyclogenesis along a preexisting frontal boundary but also cyclogenesis in polar airstreams (e.g., Reed 1979) and the extratropical transition of tropical cyclones (e.g., Evans et al. 2017).
Coming from the need to better predict poorly forecasted explosive cyclones, the 1980s and 1990s were a fruitful time for extratropical cyclone research. An outcome of this period of extensive research, Shapiro and Keyser (1990) synthesized knowledge from field-program observations and numerical modeling into a new four-stage conceptual model of a marine extratropical frontal cyclone (Fig. 16-24). Their model begins with incipient cyclogenesis along a continuous and broad frontal zone (stage I). During the early stages of cyclogenesis, a fracturing of the previously continuous frontal zone occurs, along with contraction of the now discontinuous warm- and cold-frontal temperature gradients (stage II). The warm front then develops westward into the northern airstream behind the low, where Shapiro and Keyser (1990) refer to it as a bent-back warm front, and the warm sector narrows, leading to a pronounced frontal T-bone (stage III). Last, a warm-core seclusion forms as the cold air and the bent-back warm front encircle the low center (stage IV). The name bent-back warm front often leads to confusion (Schultz et al. 1998, p. 1770). For simplicity, and to avoid confusion with other frontal archetypes, we recommend bent-back front be applied for this feature.

The Shapiro–Keyser model differs from the Norwegian cyclone model in several ways, but perhaps the most distinctive is that it does not include the process of occlusion. Instead, the warm and cold fronts become aligned perpendicular to each other (i.e., the frontal T-bone), and only in the late stages of cyclogenesis is there some narrowing of the warm sector. Synoptic analysis illustrates, however, that extratropical cyclones may exhibit frontal structures and life cycles that may resemble the Norwegian cyclone model, the Shapiro–Keyser model, or other alternatives (Schultz et al. 1998; Catto 2016). This broad spectrum reflects the diversity of dynamical factors and physical processes contributing to cyclone evolution including variations in the large-scale flow (e.g., Simmons and Hoskins 1978; Hoskins and West 1979; Davies et al. 1991; Thorncroft et al. 1993; Schultz et al. 1998; Wernli et al. 1998; Schultz and Zhang 2007), surface characteristics (e.g., Hines and Mechoso 1993; Thompson 1995; Rotunno et al. 1998), diabatic heating (e.g., Nuss and Anthes 1987; Terpstra et al. 2015), and orographic effects (e.g., Pichler and Steinacker 1987; Hobbs et al. 1990, 1996; Tibaldi et al. 1990; Steenburgh and Mass 1994; McTaggart-Cowan et al. 2010a,b; West and Steenburgh 2010). As a result, there are well-documented cases of occlusions forming and lengthening as the cold front overtakes the warm front as depicted by the Norwegian cyclone model (e.g., Schultz and Mass 1993; Market and Moore 1998; Martin 1998, 1999), occlusions forming through alternative processes (e.g., Palmén 1951; Anderson et al. 1969; Reed 1979; Hobbs et al. 1990, 1996; Neiman and Wakimoto 1999), and cyclones that instead develop a frontal T-bone (e.g., Neiman and Shapiro 1993). How can these contrasting paradigms be reconciled?

Schultz and Vaughan (2011) proposed that the key physical process operating in all of these paradigms is the wrap-up of the thermal wave by differential rotation and deformation. They argued that, in many cyclones, the cold front undeniably catches up to the warm front, but that this catch up is not an explanation for occlusion. Instead, they defined the occlusion process as “the separation of warm-sector air from the low center through the wrap-up of the thermal wave around the cyclone” (Schultz and Vaughan 2011, p. 446). The cold front overtaking the warm front is a consequence of differential rotation and deformation thinning the warm sector and drawing the two fronts together (Martin 1999). Differential rotation and deformation also act to elongate the warm tongue and extend the length of the occlusion, explaining why in some cases the occluded front is much longer than can be explained by the merger of the cold and warm fronts, as illustrated by the highly wrapped-up occluded fronts in cyclones described by Reed et al. (1994) and Reed and Albright (1997). Although the Shapiro–Keyser model omits the occluded front, the
separation of the low center from the warm sector, development of the intervening warm front, and formation of their back-bent warm front are consistent with the wrapping up of the thermal wave. Thus, the wrap-up of the thermal wave through differential rotation and deformation serves as a framework for understanding frontal cyclone evolution in a variety of contexts.

e. Precipitation structure and rainbands

The precipitation structure of cyclones was a key component of the Norwegian cyclone model, including the formation of precipitation as warm air ascends the wedge of cold air ahead of the warm front and the generation of a narrow band of precipitation as the cold front intrudes into the warm sector (Fig. 16-4). It was not until the development of weather radars, and their subsequent incorporation within observation networks, that progress was made in understanding rainfall patterns associated with extratropical cyclones. By the 1970s, the mesoscale structure of such precipitation features began to be revealed (e.g., Browning and Harrold 1970; Harrold and Austin 1974; Browning 1974; Houze et al. 1976). The term rainband was first introduced by Houze et al. (1976), referring to elongated mesoscale areas of precipitation that favor certain locations relative to the fronts themselves. Based on the range of observations collected during the CYCLES project, Houze et al. (1976) introduced a general classification scheme identifying six types of common rainbands—warm frontal, warm sector, narrow cold frontal, wide cold frontal, wavelike, and postfrontal. This list was later refined by Hobbs (1978), Matejka et al. (1980, their Fig. 1), and Houze and Hobbs (1982) which separated warm-frontal bands according to their position relative to the surface warm front, and also added the surge band in the vicinity of upper-level cold fronts. The current classification, presented in Houze (2014) and illustrated in Fig. 16-25, introduced the concept of the occlusion band found in the northwest quadrant (e.g., Sanders and Bosart 1985a,b; Sanders 1986b; Martin 1998; Novak et al. 2004, 2006, 2008, 2009, 2010; Rauber et al. 2014).

In addition to radar observations, the CYCLES project also provided valuable in-cloud aircraft measurements that stimulated interest in the role of microphysics and hydrometeor transport. In the early 1980s, a number of idealized modeling studies were designed to complement the in situ observations and elucidate the influence of microphysical processes on frontal rainbands. These studies revealed the importance of the ice phase in particular, demonstrating a link between ice crystal growth and surface precipitation flux in both warm-frontal rainbands (Rutledge and Hobbs 1983) and narrow cold-frontal rainbands (Rutledge and Hobbs 1984). Similar conclusions were reached by Cox (1988), who performed idealized two-dimensional simulations of
both warm-frontal and narrow cold-frontal rainbands for comparison with field observations. Simulations including only liquid-phase processes could not accurately model the surface precipitation flux and neither could they produce realistic distributions of latent heat release.

Research into the role of precipitation phase in the evolution of frontal rainbands grew over the next decade, motivated by Clough and Franks (1991) who suggested that frontal downdrafts could be enhanced by sublimating snow. This idea was later supported by modeling studies (e.g., Parker and Thorpe 1995; Marecal and Lemaître 1995) and also by Clough et al. (2000) using observational data from the FASTEX field campaign (Joly et al. 1997, 1999). Numerical simulation of a FASTEX winter case study by Forbes and Clark (2003) also demonstrated how the rate of sublimation-induced cooling beneath slantwise ascending frontal updrafts can influence the development of postfrontal rainbands. Indeed, along with ice crystal size and habit (or shape), sublimation rate is an important factor in determining the density of snow, and hence snowfall depths in winter storms (e.g., Roebber et al. 2003; Stark et al. 2013).

The role of diabatic effects in precipitation banding has been the subject of further investigation in recent years. Idealized baroclinic-wave simulations have shown that latent heating and cooling associated with microphysical processes can perturb vertical velocity across the warm conveyor belt, leading to the creation of multiple precipitation bands (e.g., Norris et al. 2014). Observations of cool-season European cyclones also suggest the possibility of a link between precipitation banding, diabatic heating, and finescale wind structure below 800 hPa on the equatorward side of intense storms (Vaughan et al. 2015). Such results serve as a reminder of the importance of high-quality observations for the validation of numerical models, ultimately to enable a deeper understanding of the morphology of high-impact weather embedded within low pressure systems.

In the quest to extend our knowledge and our ability to predict cyclones on smaller and smaller scales with increased accuracy, we highlight the need for high-quality observations of cloud microphysical processes to challenge NWP models. How we arrived at this point and the more recent history of NWP focused specifically on extratropical cyclones is discussed in the next section.

7. Prediction

The failed prediction of major extratropical cyclones has been a catalyst for research programs and improvements in our understanding throughout time, again highlighting why extratropical cyclones are the centerpiece of meteorology. One of the first events hindcasted using NWP by computer was of the 5 January 1949 cyclone over central North America (Charney et al. 1950). Later, Leary (1971) documented systematic underpredictions of oceanic cyclones and overpredictions of Rocky Mountain lee cyclones in the NMC primitive equation model, but it was not until later that decade when a catalyst occurred that energized the research community. The infamous 19 February 1979 Presidents’ Day storm along the east coast of North America (e.g., Bosart 1981) was severely underpredicted by the LFM-II Model. The motivation for the definition and study of rapidly developing cyclones was in part due to their poor performance in the operational models at the time (Sanders and Gyakum 1980). With this definition and recognition, an explosion (pun intended) in research on rapidly developing cyclones occurred. The National Science Foundation, Office of Naval Research, and other funding bodies invested heavily in extratropical cyclone research, including field programs, climatologies, theory, and numerical modeling. We have already seen the outcomes of much of this work in other sections, but in this section, we focus on NWP, with the specific goal to discuss some of the NWP advances and predictability of extratropical cyclones, to highlight some of the forecast challenges, and to propose some ideas for future directions to improve cyclone predictability.

a. NWP advances and systematic errors

Accurate operational forecasts of extratropical cyclones require accurate numerical guidance. Following on from Leary (1971), Charles and Colle (2009a) gathered some validation statistics over the eastern United States to show how cyclone displacement errors have evolved over the decades (Fig. 16.26). During the 1978/79 cool season, the LFM-II displacement errors over the continental United States and surrounding oceans ranged from about 300 to 440 km from hours 24 to 48 (Silverberg and Bosart 1982). By the late 1980s, cyclone position errors over the western Atlantic had improved by about 30%. By the 2002–07 cool seasons, the displacement errors in the North American Meso-scale Forecast System (NAM) and Global Forecast System (GFS) had improved by another 30%–40%, which suggests that cyclone position forecasts had continued to improve since these earlier studies, albeit at a modest rate.

Despite this overall improvement, the predictability of extratropical cyclones can still vary greatly from case to case. At issue is whether the forecast errors are due to errors in the initial conditions or errors in the physical
processes represented within the model (e.g., moist convection). In the 1980s, forecast busts were more common than now, even for the short-term (0–3-day) forecasts (e.g., Reed and Albright 1986; Reed et al. 1988; Bosart and Sanders 1991). Even in 2001–02, landfalling cyclones on the U.S. West Coast with large errors (200–400 km and >10 hPa) were happening even in 24–48 h forecasts, related to errors in the initial conditions over the North Pacific Ocean (McMurdie and Mass 2004). These low-predictability cases were sensitive to flow regime, with storms tracking from the southwest having the largest sensitivity to initial conditions (McMurdie and Ancell 2014). In another example, the 25 January 2000 East Coast cyclone was another bust in which initial condition errors were important (Zhang et al. 2002). However, Zhang et al. (2003) showed that the predictability for this event was limited (near its intrinsic limit) because even adding small random white noise to the initial temperature resulted in large forecast differences by 30 h. This rapid upscale error growth was the result of moist convective processes within the baroclinic wave (Zhang et al. 2007). In contrast, Durran et al. (2013) showed for two extratropical cyclones that the initial errors were concentrated in some of the longer wavelengths (100–1000 km), and not from an upscale growth process. The variety of these results suggest that a study on the error growth characteristics in a larger sample of cyclones might shed light on this behavior.

Despite this uncertainty into why the errors are happening, the good news is that the frequency of large forecast busts has diminished. A few decades ago, only a few deterministic operational models were run at fairly coarse resolution with limited physics and primitive data assimilation (e.g., LFM, NGM, and AVN). Currently, large ensembles are generated for global models and are combined with advanced data-assimilation approaches such as four-dimensional data assimilation and ensemble Kalman filter (EnKF). As a result, forecasters now have access to over 100 ensemble members for a particular cyclone that are run at higher resolution with more sophisticated physics, so the chances of all the ensemble members completely missing a storm are much less.

The bad news is that there are still systematic errors for extratropical cyclone forecasts in many of these operational models and ensembles. The deterministic models have had systematic underprediction bias in predicting the intensity (e.g., central pressure) of these storms over the decades. When the LFM was operational back in 1972 at 190.5-km grid spacing, extratropical cyclones in the Pacific and Atlantic were too shallow by 6–10 hPa at 48 h (Silberberg and Bosart 1982). Mullen (1994) showed that there was a systematic underprediction error in the global model (AVN) initial analysis for cyclones for 1 November 1989–31 January 1990, and that the model underestimated the deepening rates. Uccellini et al. (2008) also found that 4-day cyclone forecasts from the NOAA/NCEP Ocean Prediction Center were frequently underforecast, especially for more intense storms. More recently, Korfe and Colle (2018) showed that major operational modeling systems (Canadian, NCEP, and ECMWF) still underpredict relatively deep cyclones in the medium range, particularly near the Gulf Stream. The models all had a slow long-track bias that was significant from 24 to 90 h, and they had a left-of-track bias from 120 to 144 h. The ECMWF ensemble errors have been decreasing from 2007 to 2014 at all lead times from 0 to 6 days, but only at short lead times at CMC and not as much at NCEP.

b. Use of ensembles

With limited computer power, early NWP focused on improving model resolution and model physics. As computer power increased, running a number of forecasts to produce an ensemble was able to be realized. Ensembles embrace the uncertainty in predictions that Lorenz identified (section 3c), although others previously had enunciated such concerns (e.g., Lewis 2005). Since then, numerous studies have identified the benefits of using ensembles for both the short- and medium-range forecasts of extratropical cyclones (e.g., Froude et al. 2007; Park et al. 2008; Johnson and Swinbank 2009; Charles and Colle 2009b). For example, Froude et al. (2007) verified extratropical cyclone tracks in the 0–7-day forecasts from ECMWF and NCEP ensemble prediction systems between January and April 2005. The ECMWF ensemble consisted of 50 perturbed members.

Fig. 16-26. Extratropical cyclone displacement errors (km) vs forecast hour for the LFM-II (Silberberg and Bosart 1982; 1978/79 cool season for CONUS and oceans), the NGM and AVN (Smith and Mullen 1993; 1987/88 and 1989/90 cool seasons for the Atlantic), and the NAM and GFS (2002–07 cool seasons) (Atlantic). The figure is from Charles and Colle (2009a, their Fig. 16).
with a spectral resolution of T255L40, whereas the NCEP ensemble consisted of 10 perturbed members with a resolution of T126L28. The ECMWF ensemble was slightly more accurate than the NCEP ensemble for cyclone intensity in the Northern Hemisphere, whereas the NCEP ensemble was significantly more accurate for cyclones in the Southern Hemisphere. In another example, Froude (2011) compared nine ensemble prediction systems from TIGGE in 2008 for both the Northern and Southern Hemispheres. For about one-half of the models, the cyclone intensity and position errors were 10%–20% larger in the Southern Hemisphere than in the Northern Hemisphere, but errors in other models by other centers (e.g., ECMWF and Met Office) were more comparable, with some coherent biases in most of the models. More than one-half of the models were too weak with the cyclones in both hemispheres (Figs. 16-27a,b), and most models had a slow bias (Figs. 16-27c,d).

More recently, Korfe and Colle (2018) validated the ECMWF, Canadian (CMC), and NCEP ensembles over the eastern United States and western Atlantic for the 2007–15 cool seasons. For lead times less than 72 h, the NCEP and ECMWF ensembles had comparable mean absolute errors in cyclone intensity and track, whereas the CMC errors were larger (Fig. 16-28). For 4–6-day forecasts, the ECMWF had 12–18 and 24–30 h more accuracy for cyclone intensity than NCEP and CMC, respectively. The ECMWF also had greater probabilistic skill for intensity and track than CMC and NCEP. Korfe and Colle (2018) showed that the 90-member multimodel ensemble from all three centers (NCEP + CMC + ECMWF) had more probabilistic skill than
any single ensemble, thus illustrating the importance of adding model diversity. For example, Korfe and Colle (2018) showed that, for the 3–6-day forecasts from 2007 to 2015, cyclones fell outside of the envelope for the ECMWF ensemble 5.6%, 5.2%, and 4.1% of the cases for cyclone intensity, along-, and cross-track positions, respectively. For the NCEP ensemble, these values were 13.7%, 10.6%, and 11.0%, respectively. Using a multimodel ensemble (90-member NCEP + CMC + ECMWF), however, reduces the percentage of cases outside the envelope of the 90-member ensemble: 1.9%, 1.8%, and 1.0% of cases, respectively. How many of these outside-the-envelope cases are near their intrinsic predictability limit is not known, which is an area of potentially important future research.

One existing challenge of cyclone verification research is that the feature-tracking algorithms have large uncertainties and often cannot track weak cyclones in many members. Therefore, the true accuracy or skill of the ensemble is often not being assessed. Zheng et al. (2017) developed a scenario-based method, which includes all ensemble members by using an empirical orthogonal function (EOF) and fuzzy clustering methodology. The EOF analysis at the verification time is used to determine the dominant patterns of variations in ensemble sea level pressure forecasts. The principal components (PCs) corresponding to the leading two EOF patterns are used as a base to perform fuzzy clustering on the ensemble sea level pressure forecasts over the verification region. Each ensemble member is assigned a weight that identifies its relative strength of membership to each of the five clusters depending on its distance from the cluster mean in the PC phase space. An ensemble member is assigned to the cluster with the largest weight (Zheng et al. 2017), and an ensemble mean cluster is also determined for those members closest to the mean. Once the clusters are obtained, spatial plots can be made to demonstrate the synoptic clusters associated with each cluster using, for example, a “spaghetti” plot of a particular contour.

To illustrate this approach, consider the 90-member (NCEP + CMC + ECMWF) 6-day ensemble forecast initialized at 1200 UTC 21 January 2015. The forecast mean cyclone position was about 200 km to the southwest of the analyzed cyclone, and the largest spread was mainly to the west of the ensemble mean cyclone (Fig. 16-29a). A spaghetti plot of the 996-hPa contours from the ensemble also illustrates the spread of the cyclone position, which appears to cluster by ensemble system, with the ECMWF ensemble members to the west relative to the NCEP ensemble members (Fig. 16-29b).
Figures 16-29c and 16-29d show the leading two EOF patterns for this 6-day sea level pressure forecast, which explains 42.9% and 28.7% of the variance over the verification region, respectively. The first EOF (EOF1) has a maximum located about 400 km west of the ensemble mean position of the surface cyclone (Fig. 16-29c). This pattern represents a deeper storm with a westward shift and a weaker storm with an eastward shift relative to the ensemble-mean cyclone at 6 days. Meanwhile, the dipole pattern with EOF2 (Fig. 16-29d) is an asymmetric dipole pattern, with a positive pattern representing the deepening and northeastward shift of the cyclone and a negative pattern representing the weakening and southwestward shift of the cyclone. Figure 16-30 shows the ensembles in the PC1 and PC2 phase space and the clusters, including an ensemble mean cluster and the verifying analysis. This example highlights how the ensemble systems tend to cluster together, which explains why all three ensembles
together verify the best on average (Korfe and Colle 2018).

c. Physical processes

There has been no detailed investigation for why the model forecasts have improved over the decades, which would require systematically varying model resolution, data assimilation approaches and observations, and model physics. However, smaller cyclone errors are likely linked to increased operational model resolution as grid spacings on global models have decreased from about 200 km in the early 1970s to about 80 km in the early 1990s to 20–30 km in the early 2000s. Increasing resolution has allowed models to better resolve important physical processes, such as low-level temperature gradients (e.g., along the coastlines, SST boundaries), orographic effects (e.g., flow blocking, lee cyclogenesis), and diabatic effects (e.g., condensation, surface fluxes, latent heating). For example, as highlighted in section 2, the importance of diabatic heating on these storms has been well documented.

Systematic errors from dry dynamical forcing are likely relatively small as grid spacings approach 10–20 km, but latent heating biases are likely still prevalent because most operational global models today still run with a convective parameterization. Thus, smaller-scale embedded convection, such as that associated with the warm conveyor belt (section 6), may be important to correctly predict cyclone intensity.

There is also interest in how these extratropical cyclones may change during the next 100 years (Colle et al. 2015); however, global climate models typically underestimate the intensity of extratropical cyclones in the North Atlantic because of their relatively coarse resolution (100–300-km horizontal grid spacing) (Chang et al. 2013; Colle et al. 2013; Zappa et al. 2013; Seiler and Zwiers 2016; Seiler et al. 2018). Colle et al. (2013) found that those climate models that best predicted cyclone tracks and intensities had the higher model resolution. Jung et al. (2006) and Champion et al. (2011) also found that the extratropical cyclone intensity increases with increasing horizontal resolution. Thus, one must be careful when using relatively coarse climate models to understand future cyclone changes given these systematic errors.

However, as climate-model scales steadily converge toward weather-model scales, many of the same issues faced by the weather community also exist, which will continue to foster collaboration between both modeling communities. For example, Willison et al. (2013, 2015) showed that latent heat release increases the intensity of extratropical cyclones in the future model projections as grid spacings are decreased from 120- to 20-km grid spacing. Zhang and Colle (2018) showed that most of the strong underprediction bias in a climate model can be removed if the grid spacings are decreased to around 20 km, such that latent heating associated with precipitation can be better resolved. As
with other studies, Michaelis et al. (2017) found using regional climate model grid spacings down to 4 km that the total number of strong storms in the North Atlantic storm track may decrease during the twenty-first century because of a weaker low-level temperature gradient. However, both Michaelis et al. (2017) and Colle et al. (2013) found increased occurrence of cyclones in the future along the East Coast. Zhang and Colle (2018) hypothesized that a decreased low-level temperature gradient may be compensated by additional latent heating within the entrance of the storm track.

d. Mesoscale challenges

As NWP improves, there will be fewer large forecast bust cases but still events with large predictability challenges, because relatively small changes in the cyclone position can lead to significant changes in the axis of heavy precipitation that have large societal impacts. A good example is the 26–27 January 2015 East Coast cyclone in which even short-term 24-h forecast uncertainties in the western edge of a sharp precipitation gradient caused major issues for the New York City region. For example, Fig. 16-31a shows the regional radar at 0600 UTC 27 January 2015, and Fig. 16-31b shows the location of the 25.4-mm (1-in.) storm-total threshold from the NCEP ensemble (Greybush et al. 2017). Those members with a more eastern cyclone position (about 100 km east of observed) had the heavy snow more over Long Island, whereas those members farther to the west of the observed had the heaviest precipitation to the west of New York City. This uncertainty and the sharp western gradient in the precipitation were evident in many other ensembles (Greybush et al. 2017), which complicates matters for the forecaster because there are many different potential impacts for the New York City area.

There is a wide spectrum of important mesoscale phenomena associated with these storms that cause forecast challenges (e.g., precipitation bands, gravity waves, severe convective storms, freezing level issues, cyclone interaction with terrain, orographic precipitation). Currently, operational convective-allowing models are run deterministically at 3–4-km grid spacing, which helps with the prediction of these phenomena, but there are few high-resolution ensembles at this grid spacing. Multimodel convective-allowing models up to 90 members have been run for various projects, such as the NOAA Hazardous Weather Testbed Spring Experiment (Clark et al. 2018), but, at the time of this writing, only the High-Resolution Rapid Refresh Ensemble (HRRRE) is run operationally over the contiguous United States (CONUS) using 20 members at 3-km grid spacing and a lagged-ensemble approach (S. Benjamin 2018, personal communication).

e. Opportunities

Despite improvements in NWP predictions, systematic errors still lead to the loss of probabilistic skill. Historically, much of the model performance has been quantified using basic-state variables (e.g., temperature, wind, precipitation), for standard metrics (e.g., 500-hPa anomaly correlations, root-mean-square errors), and averaged over a relatively large geographic region or for select points. This approach helps quantify how good or bad the model may be in general, but it does not help target the origin of the errors in order to help improve the model. Another way to perform verification is to calculate the errors around an object of interest, such as a convective line, snowband, hurricane, and extratropical cyclone. In the future, better use of object-oriented verification is needed.

A number of process-oriented metric and diagnostic approaches have been applied recently to models, but mainly to climate models. Physical processes have been evaluated for the Madden–Julian oscillation (e.g., Kim et al. 2009, 2014), east Pacific warm pool variability (e.g., Maloney et al. 2014), tropical cyclones (e.g., Kim 2017), and extratropical cyclones (e.g., Booth et al. 2018). For an extratropical cyclone, understanding the horizontal temperature gradients, surface fluxes, vertical stability, moisture budget around the storm, and diabatic heating profiles are important for model developers who need to understand how changes in the physical parameterizations (e.g., microphysics, surface layer physics, convective schemes) impact the parameters leading to any model biases. This effort also requires the community to obtain and archive important nonconventional quantities for model development, such as surface fluxes, planetary boundary layer height, and heating profile estimates.

An operational convective-allowing model ensemble is needed that can be run out 2–3 days to predict mesoscale phenomena associated with extratropical cyclones that can produce large gradients in snowfall or precipitation amount over relatively short areas, causing particular problems along populated coast lines. To reduce underdispersion, these ensembles need both physics diversity (e.g., microphysical, surface layer physics, convective schemes) and initial-condition diversity. Last, it needs to be determined whether the upscale growth of errors from some of these mesoscale phenomena are leading to some intrinsic predictability limits or still uncertainties in the regional-or-larger scale. More tools are needed for the forecaster and other users to better utilize ensembles for these cyclones and associated weather impacts. Above, a fuzzy clustering approach was highlighted to break down the ensemble into different possible scenarios. Object-based tracking will allow various
mesoscale features (e.g., snowbands) within these storms to be tracked and communicated probabilistically. More advanced postprocessing approaches, such as machine learning and statistical approaches (e.g., Bayesian model averaging), can be applied to better calibrate the ensemble for these storms. More ensemble graphics are needed operationally besides mean, spread, and basic probabilities, with a focus on the feature or hazard in question.

8. The past, present, and future

Over the past 100 years, extratropical cyclone research, as described in this chapter, has made remarkable strides.
Can we determine the key ingredients that were conducive to that progress in the past? Are there indications of what would make the line of progress sustainable, analogous to persistence in forecasting? What can be seen as prerequisites for further progress in the future?

Going back to **section 1**, we believe that the characteristics that have made this period so successful are based on the triad of 1) practical challenges of addressing poor forecasts that had large socioeconomic consequences; 2) the intermingling of theory, observations, and diagnosis as depicted in **Fig. 16-2**; and 3) strong international cooperation.

As the first note of the triad, poor forecasts of cyclones sinking Norwegian fishing vessels motivated Vilhelm Bjerknes to develop the observing system. Synoptic analysis of the data from this network of stations led to the development of the Norwegian cyclone model. Forecasts of cyclogenesis were also among those first forecasts performed by computer. And, the large interest in cyclogenesis in the 1980s and 1990s was due to the poor NWP forecasts of rapidly developing cyclones.

This societal need led to the second note of the triad: the fruitful application of **Fig. 16-2**. The century began with the collection of finescale weather observations that led to the birth of the Norwegian cyclone conceptual model and its continual refinement through the 1950s to the present. The further collection and analysis of routine observations, through specialized field programs to provide targeted data collection, as well as the development of new observing tools of radar and satellite, also shed light on the structures and processes within cyclones. Theories for cyclones at 1919 were incomplete and being debated, but frameworks of divergence, baroclinic instability, quasigeostrophy, and potential vorticity have been developed that have largely led to the “cyclone problem” (e.g., Palmén 1951, 618–619) being solved. But the crucial test of these theories was how they compared with observations. The first attempt at calculating the weather forecast was attempted shortly after the start of our century (Richardson 1922); but, since the first forecast by computer at midcentury (Charney et al. 1950), remarkable progress on NWP has occurred, driven in part by theoretical advances in modeling and architecture, improved observations, their incorporation into the initial conditions of models through improved methods of data assimilation, and accounting for chaos. All these theories, observations, and diagnosis through numerical modeling have led to improved understanding of relevant physical processes.

In part, the success in effective application of **Fig. 16-2** depends upon the character of individual researchers and their willingness to cross Valleys of Death, whether it be the Valley of Death between observation and theory, the Valley of Death between operations and research (e.g., National Research Council 2000), or the Valley of Death between observations and modeling. Rossby (1934, p. 32) famously noted, “The principal task of any meteorological institution of education and research must be to bridge the gap between the mathematician and the practical man, that is, to make the weather man realize the value of a modest theoretical education and to induce the theoretical man to take an occasional glance at the weather map.” To many of us coauthors who have had success in our meteorological institutions, we dedicate this chapter to our advisors and mentors who prepared us for the journey, taught us not to be afraid, and gave us the confidence to cross the valleys.

The third culminating note of the triad, a decisive ingredient for the success achieved during the past 100 years, has been the international voluntary cooperation (section 1; Volkert 2017). The international associations for sharing science that emerged at the end of World War I and the AMS—at first a national organization, but which later became one of the leading professional societies for atmospheric science worldwide—both originated in 1919. Many of the Bergen School meteorologists were Norwegian, but they also came from other European countries, the United Kingdom, and the United States for training. These apostles for science traveled the globe, many settling elsewhere, to help advance their methods and to lead the development of NWP in the United States, United Kingdom, and Europe. International field research programs (e.g., Global Atmospheric Research Program, Alpine Experiment, FASTEX, THORPEX) were tasked with improved understanding of extratropical cyclones and their fronts. The formation of ECMWF in 1975 and its more than 40 years of operation were critical to supporting international cooperation on NWP [e.g., chapter 20 in Woods (2006)]. International conferences continue to serve as a focal point for fruitful scientific discussion, including the Cyclone Workshop (e.g., Gyakum et al. 1999), soon to celebrate its 19th incarnation.

Given this impressive progress over the last century, what are the current trends?

- The development of reanalysis datasets consisting of gridded meteorological data using a consistent analysis and modeling system allows the possibility of many years of meteorological analyses. Indeed, such progress has already led to many advances. We expect more will come.
- In part to address the large amounts of data available through reanalyses, automated approaches for detection
and analysis of these datasets will become even more developed in the near future. Machine-learning and artificial-intelligence approaches to further interrogate the data will become more prevalent.

- The practice of sharing datasets and code communally to allow wider access and analysis of these datasets is growing and will continue to grow, driven in part by national regulations and the expectations of funding agencies.
- As detailed in sections 6e and 7, cyclone structure and predictability on the mesoscale and microscale can be very sensitive to cloud-microphysical processes. To what extent do the details of the cloud microphysics matter to the evolution and predictability of cyclones? Are any systematic effects missing or misrepresented in the current generation of NWP models? More collection of relevant datasets, as well as theoretical developments, to better understand these processes will be required for further progress.
- Last, as the world warms because of nearly 200 years of anthropogenic greenhouse gas input, climate change will have a profound influence on regional climates. Their atmospheric and oceanic responses to changes in a warmer climate—including the potential loss of Arctic sea ice and melting of permafrost—will change extratropical weather systems. Recent research is showing conflicting results as to the magnitude of this effect for the jet stream, but further investigations should reduce this uncertainty.

Beyond these current trends, what is the outlook for the next century? Certainly, 100 years is well beyond the deterministic predictability limit. Therefore, the provision of a forecast for the next century is too daring to be made in any detail. However, in combination with the recently observed changes to the use of observations in data assimilation (e.g., Davies 2005, especially pages 374–375), some trends are suggested, along which future research agendas may develop:

- Traditional boundaries of research areas will be less clear cut or will disappear altogether. Extratropical cyclones may continue to be the Forrest Gump of meteorology, but they will not likely remain its centerpiece. Instead, extratropical cyclone research is likely to be assimilated into predictability research and directly linked to data assimilation and ensemble prediction.
- The midlatitudes of both hemispheres will continue to be areas of high scientific interest (in addition to the tropics between and the polar regions beyond), yet the embedded cyclones alone may lose their special status as core research objects.
- Dedicated field experiments will continue to serve as catalysts for progress, especially if new technology is applied, be it on airborne platforms (manned and unmanned aircraft) or satellite missions with active sensors (radar and lidar). The North Atlantic Waveguide and Downstream Impact (NAWDEX) campaign of 2016 may serve as a recent example (Schäfler et al. 2018).
- Near-global coverage of line-of-sight motion vectors should be available soon. [The ESA satellite Atmospheric Dynamics Mission (ADM)-Aeolus launched in August of 2018; http://www.esa.int/Our_Activities/Observing_the_Earth/Aeolus/Overview2]. This mission is likely to open new horizons for both data-assimilation techniques and a series of systematic case studies.
- Extratropical cyclones and their composited storm tracks will continue to be of great interest for studies of regional reanalyses (e.g., Buizza et al. 2018) on the way toward seamless prediction of weather and climate (e.g., Palmer et al. 2008; Hoskins 2013).

 Altogether, the evolution of extratropical cyclone research over a full century (i.e., the current lifetime of the AMS and the Norwegian cyclone model) carries some analogies with their life cycles driven by the ceaseless wind (Dutton 1976).

- The main features of cyclones and research about them are slowly evolving and exhibit some inherent predictability. So, extratropical cyclones stayed on the research agenda during the entire century.
- A multitude of disturbances of much smaller scale are embedded, making every depression and its life cycle distinct. Equally, inventions and technological developments of, for example, computers and satellite sensors, transformed the tools to study cyclones and disseminate results.

Thus, hitherto unforeseeable pieces of technology may redefine extratropical cyclone research considerably, but the impact of actual weather systems in the extratropical belts around Earth will continue to remind researchers and the general public alike of their special relevance for the atmospheric sciences as a whole.

Over 50 years ago, Bjerknes (1964, p. 314), the author of the Norwegian cyclone model, remarked at the

---

5 A counterexample from the past of just such a visionary outlook, and a well-documented one, is the May 1957 speech by Lloyd Berkner about the coming era of satellite meteorology and operational NWP (Droessler et al. 2000).
inaugural award ceremony for the AMS’s Harald Sverdrup Gold Medal:

“But yet I would give highest recommendation to the less narrow and more basic field of meteorology, which was the concern of the founders of our science, and which still is our first duty to society: weather forecasting. All too frequently, students, and professors too, shy away from the subject of weather forecasting and go shy one of the nice little research specialties which are more nerve racking, and which do not force you to show the public how often you are wrong. But, fortunately, the weather forecaster will soon be better off. Electronic automation has already relieved him of much of the overwhelming load of data handling, and now also presents him with electronically computed forecast maps.”

As of today, many researchers and students are heavily committed to improving forecasting of extratropical cyclones, taking the risk of making errors, but also making efforts to quantify the inherent uncertainties. In the future, the international patchwork of nation states—a globally interlinked society—still has to provide the basis for both the global atmospheric observation system and the education of the next generations of researchers. Thus, we must maintain the high standards of our discipline and, it is hoped, extend them, as has happened during the past 100 years.

Acknowledgments. The authors gratefully acknowledge the support of the NOAA/Next Generation Global Prediction System through Grants NA16NWS4680018 (Roeber), NA15NWS4680003 (Colle), and NA17NWS4680004 (Colle); the National Science Foundation through Grants AGS-155960 (Bosart and Keyser), AGS-1656406 (Bosart), and AGS Postdoctoral Research Fellowship AGS-1624316 (Winters); the U.K. Natural Environment Research Council through Grants NE/I005234/1, NE/I026545/1, and NE/N003918/1 to the University of Manchester (Schultz); the German Research Foundation DFG via the collaborative research center Waves to Weather (Volkert); and the ongoing support of the National Science Foundation and the National Weather Service CSTAR program (Steenburgh). We thank Heather Archambault for providing Fig. 16-6, Robert Marc Friedman for his comments about Bjerknes (1919), and two anonymous reviewers for their comments.

REFERENCES


Wallace, J. M., G. H. Lim, and M. L. Blackmon, 1988: Relationship between cyclone tracks, anticyclone tracks and baroclinic


Weather Bureau, 1948: The snowstorm of December 26–27, 1947


