Mesoscale Gravity Waves and Midlatitude Weather:

A tribute to Fuqing Zhang

James H. Ruppert, Jr., a Steven E. Koch, b Xingchao Chen, c Yu Du, d,e Anton Seimon, f Y. Qiang Sun, g Junhong Wei, d,e and Lance F. Bosart h

a School of Meteorology, University of Oklahoma, Norman, OK
b Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ
c Department of Meteorology and Atmospheric Science and Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, PA
d School of Atmospheric Sciences, and Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, Sun Yat-sen University, Guangzhou, China
e Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China
f Center for Environmental Policy, Bard College, Annandale-on-Hudson, NY
f Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ
h Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, NY

Corresponding author: James H. Ruppert, Jr., jruppert@ou.edu

Early Online Release: This preliminary version has been accepted for publication in Bulletin of the American Meteorological Society, may be fully cited, and has been assigned DOI 10.1175/BAMS-D-20-0005.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

© 2021 American Meteorological Society
ABSTRACT

Over the course of his career, Fuqing Zhang drew vital new insights into the dynamics of meteorologically significant mesoscale gravity waves (MGWs), including their generation by unbalanced jet streaks, their interaction with fronts and organized precipitation, and their importance in midlatitude weather and predictability. Zhang was the first to deeply examine “spontaneous balance adjustment” – the process by which MGWs are continuously emitted as baroclinic growth drives the upper-level flow out of balance. Through his pioneering numerical model investigation of the large-amplitude MGW event of 4 January 1994, he additionally demonstrated the critical role of MGW–moist convection interaction in wave amplification.

Zhang’s curiosity-turned-passion in atmospheric science covered a vast range of topics and led to the birth of new branches of research in mesoscale meteorology and numerical weather prediction. Yet, it was his earliest studies into midlatitude MGWs and their significant impacts on hazardous weather that first inspired him. Such MGWs serve as the focus of this review, wherein we seek to pay tribute to his groundbreaking contributions, review our current understanding, and highlight critical open science issues. Chief among such issues is the nature of MGW amplification through feedback with moist convection, which continues to elude our understanding. The pressing nature of this subject is underscored by the continued failure of operational numerical forecast models to adequately predict most large-amplitude MGW events. Further research into such issues therefore presents a valuable opportunity to improve the understanding and forecasting of this high-impact weather phenomenon, and in turn to preserve the spirit of Zhang’s dedication to this subject.
Through unparalleled zeal, Fuqing Zhang made waves through every aspect of his work—from research to mentoring to community engagement. His insights leave a deep mark on our understanding of atmospheric dynamics across a range of phenomena, especially mesoscale gravity waves.

Gravity waves are ubiquitous entities in the global atmosphere, owing to their generation by topography, wind shear, latent heating, and flow imbalance associated with jets and fronts. While most small-scale gravity waves do not have a direct impact on midlatitude weather, coherent mesoscale gravity waves (MGW) capable of creating mesoscale bands of ice pellets, snow, and/or deep convection were found by Koch and Siedlarz (1999) to be present over the central United States roughly one-third of the time during the 45-day period of STORM-FEST from 1 February–15 March 1992. A subset of such MGWs grow to sufficiently large amplitude that they cause rapid surface pressure oscillations >5 hPa in less than an hour, with low-level winds capable of damaging structures, disrupting aviation, and driving rapid changes of water level in both coastal and inland water bodies (i.e., seiches and meteotsunamis; Bechle et al. 2016; Dusek et al. 2019). These large-amplitude MGWs wreak havoc on local weather forecasts, due especially to the rapid and unanticipated changes in winds, precipitation, and convective activity they cause (Uccellini 1975; Uccellini and Koch 1987; Ferretti et al. 1988; Schneider 1990; Bosart et al. 1998). Despite our growing understanding of how nascent MGWs form, their all-important amplifying mechanisms remain elusive (Plougonven and Zhang 2014).

The substantial impacts of MGWs are exemplified by the remarkable event of 4 January 1994, where a large-amplitude MGW wave of depression propagated along the Atlantic coast...
within a moderately intense coastal cyclone (Bosart et al. 1998). As the MGW propagated northeastward away from its parent cyclone center, it amplified dramatically and accelerated to 35–40 m s\(^{-1}\) across eastern New England (Fig. 1). As it passed, surface pressure falls exceeded 13 hPa (30 min\(^{-1}\)), with fall rates as steep as 1.3 hPa min\(^{-1}\), wind gusts as high as 29 m s\(^{-1}\), and rapid sea level changes at several coastal locations. The time series in Fig. 1 reveals the by-now familiar wind–pressure correlation commonly observed during MGW passage, the cause and implications for which are discussed below. Precipitation type and intensity changed abruptly before ceasing unexpectedly during the time of greatest pressure falls (Fig. 1).

Fuqing Zhang became captivated by large-amplitude MGW events from the start of his career, with the 4 January 1994 case serving as a central element of his Ph.D. dissertation (Zhang 2000). Facing an enigmatic meteorological phenomenon and the promise of a formidable intellectual challenge, MGWs would become one of his primary scientific fixations for the majority of his abbreviated career. In homage to Zhang’s unrivaled enthusiasm for this subject and for boosting early career researchers endeavoring into it, the intention of this review is to provide a reference on midlatitude large-amplitude MGWs – both to promote basic understanding for readers of all experience levels and summarize important progress and critical open science challenges. On balance with its share in progress to date, this review places heavy emphasis on the work of Zhang, his students, and his collaborators. This review should not be taken, however, to represent all of Zhang’s research, which spans a much wider scope. Furthermore, while MGWs of meteorological significance are found from the tropics (Mapes et al. 2003; Lane and Zhang 2011) to the arctic (Andrews et al. 1987), those tied to midlatitude cyclones and weather are emphasized here. Readers are also referred to the review by Plougonven and Zhang (2014) for a broader treatment of gravity wave theory and their nature and role in the climate of the middle–upper atmosphere.
Provided next is an introduction of key definitions and scales invoked in this review, followed by discussion of fundamental dynamical concepts and techniques for understanding and diagnosing MGW behavior. These discussions are framed primarily around the early observational and theoretical studies that first described large-amplitude MGW events and introduced hypotheses concerning their genesis, maintenance, and amplification. It is from this stage that Zhang picked up, launching into his prolific career. Our understanding of MGW genesis in connection with upper-tropospheric flow imbalance is now on solid footing owing in large part to his seminal work on this subject, which creatively leveraged both theory and high-resolution numerical models (Zhang et al. 2000, 2001; Zhang 2004). Through the keen design and application of numerical experiments, he additionally underscored the vital role of moist convection in both MGW amplification and in predictability more broadly. This review therefore follows a similar historical arc.

**Defining Mesoscale Gravity Waves**

A well-established archetype for the most significant MGW cases is the singular *wave of depression* (Fig. 1) (Brunk 1949; Tepper 1951; Ferguson 1967; Bosart and Cussen 1973; Uccellini and Koch 1987; Bosart and Seimon 1988; Koch and Siedlarz 1999; Trexler and Koch 2000; Zhang et al. 2001; Ruppert and Bosart 2014). A number of studies have also documented substantial sensible weather impacts caused by MGW *wave packets* (Eom 1975; Bosart and Sanders 1986; Ferretti et al. 1988; Koch et al. 1988; Schneider 1990; Du and Zhang 2019). It is common among large-amplitude MGW events for initially moderate-amplitude gravity wave packets to evolve into a solitary, large-amplitude wave (Trexler and Koch 1999; Bosart et al. 1998; Koch and Siedlarz 1999; Zhang et al. 2001). The causes for this evolution remain a source of continuing research.
“Large-amplitude” MGWs refer specifically to those with pressure perturbations of ≥5 hPa. Before MGWs grow to large amplitude, they form part of a substantially larger family of “moderate-amplitude” MGWs (perturbations of ≥0.2 hPa), which are nonetheless capable of triggering or modulating banded precipitation (Koch and Siedlarz 1999; Jacques et al. 2015, 2017). While the emphasis herein is on larger-amplitude events, the origins, behavioral traits, and environmental characteristics common to this broader class of MGWs are also considered.

MGWs form a subset of internal gravity waves, which are buoyancy oscillations with intrinsic frequencies bounded by the Brunt–Väisälä frequency and the Coriolis parameter $f$, or periods from minutes to the inertial period (e.g., ~17 h at 45ºN; Holton 1992). Inertia–gravity waves (IGWs) refer to the subset of waves at the lower-frequency end of this range, where rotational effects are important. While the mesoscale technically encompasses this entire range, the use of mesoscale in “MGW” usually refers specifically to buoyancy-dominated waves that strongly interact with precipitation, a nomenclature we invoke herein. Such waves are characterized by surface pressure perturbations that can reach ~15 hPa and horizontal wavelengths ranging from ~50–500 km, and hence partially overlap with Orlanski’s meso-β and meso-α subclasses (Orlanski 1975; Uccellini and Koch 1987).

**Identifying Mesoscale Gravity Waves: Pressure-Wind Relationships**

How do we know that a given disturbance in the precipitation, surface pressure, and wind fields is an MGW, as opposed to, e.g., a convectively generated cold pool or any other meso- to synoptic-scale feature? This can be established with a high degree of confidence by checking that the perturbations in surface pressure $p'$ and horizontal flow parallel to the direction of propagation $u'$ are highly correlated, and that the intrinsic wave phase velocity $c$ is consistent with that predicted by the wave impedance equation $c = p'/\rho u'$, where $\rho$ is air density (Gossard and Hooke 1975), or its more accurate nonlinear counterpart (Coleman and
Knupp 2010). To elaborate, an idealized MGW propagating horizontally in the lower troposphere is schematically depicted in Fig. 2a. This schematic has as its foundations several predecessor versions (Eom 1975; Bosart and Sanders 1986; Koch and Golus 1988; Ralph et al. 1993). This depiction assumes no tilt in vertical wave structure. The dynamics underlying the correlation between $p'$ and $u'$ can be understood as follows: for an eastward propagating MGW, as in Fig. 2a, low-level convergence is facilitated by relative easterly flow ($u' < 0$) in advance of westerly flow ($u' > 0$); this convergence lifts, deepens, and adiabatically cools the stable air, in turn increasing surface pressure. This lifted and cooled air, now finding itself negatively buoyant, promotes sinking motion and low-level divergence following the wave crest. Cloud and precipitation enhancement is promoted with the approach of MGW ridges, while the sinking motion trailing these ridges promotes drying and desiccation of clouds and precipitation. The time required for cloud and precipitation development, however, causes a slight lag between peak vertical motion and precipitation, as in Fig. 2a.

These relationships are exemplified by the MGW event of 4 January 1994, even though this event differs from the wave packet depicted in Fig. 2a, being instead characterized as a singular wave of depression. As the MGW propagates northeastward, the region of greatest pressure falls remains situated at the sharp back edge of radar reflectivity through Massachusetts and Rhode Island, which is where implied subsidence is strongest (Fig. 1, upper panel). The exceptional coherence of $p'$ and $u'$ (for which wind speed is assumed to be an adequate proxy) demonstrates their direct relationship during MGW passage (Fig. 1, lower panel). The strongest winds during trough passage (i.e., $p' < 0$) were northeasterly (i.e., $u' < 0$), highlighting positive correlation (Bosart et al. 1998).

The large-amplitude MGW that swept across the south–central U.S. on 7 March 2008 provides another clear example of these relationships (Fig. 3). In this event, an intense wave of depression tracked north of a cold front along the trailing edge of a widespread stratiform
rain shield, which stretched across Mississippi and Alabama (Ruppert and Bosart 2014). As the MGW trough passed, it caused rapid surface pressure falls as extreme as 6.7 hPa in 10 min and wind gusts exceeding 20 m s\(^{-1}\) across several states. Two examples of MGW passage are provided through surface meteograms in Fig. 3. The meteograms highlight that the strongest MGW-induced winds are east-northeasterly, i.e., \(u' < 0\), consistent with their correspondence with the wave trough, i.e., \(p' < 0\). Hence, these examples again highlight \(p'–u'\) correlation across this system. Furthermore, the signatures of vertical motion implied by suppressed precipitation in the region of decreasing pressure and enhanced precipitation in the ridge lend additional support to the interpretation as an MGW.

The literature abounds with further observed examples of both MGW waves of depression and wave trains, which primarily rely upon radar and satellite imagery and surface observations as clear evidence of MGW behavior (Uccellini 1975; Eom 1975; Bosart and Sanders 1986; Uccellini and Koch 1987; Bosart and Seimon 1988; Koch et al. 1988, 2008; Schneider 1990; Ramamurthy et al. 1993; Koch and O’Handley 1997; Koch and Siedlarz 1999; Zhang and Koch 2000; Zhang et al. 2001; Koch and Saleeby 2001; Jewett et al. 2003; Jacques et al. 2015, 2017; Zhang et al. 2020). High-resolution numerical modeling studies have lent further support to these interpretations through both controlled experiments and validation against observations (Tripoli and Cotton 1989; Fovell et al. 1992; Powers and Reed 1993; Jin et al. 1996; Zhang and Koch 2000; Koch et al. 2001, 2008; Zhang et al. 2001; Stephan and Alexander 2014, 2015).

The ability to identify and track MGW activity through \(p'\) and \(u'\) provides substantial impetus for increasing the density of surface station networks and increasing their temporal measurement frequencies. This need has been partially addressed by implementation of numerous state and local mesonetworks over the last two decades (Brock et al. 1995; Rauber et al. 2001; Horel et al. 2002; Jacques et al. 2015; Stephan et al. 2016; Brotzge et al. 2020).
Tremendous additional value is realized through the application of time-to-space conversion, a mesoanalysis technique pioneered by Fujita (1951, 1955, 1963) that converts measurements of high temporal frequency into spatial information, yielding a much more detailed horizontal analysis than that obtained from irregularly distributed stations (Bosart and Seimon 1988; Koch and Golus 1988; Koch and O’Handley 1997; Bosart et al. 1998; Koch and Saleeby 2001; Ruppert and Bosart 2014; Jacques et al. 2015).

Identifying Mesoscale Gravity Waves: Wave Ducting

Assuming that high $p’-u’$ correlation has been identified, detection of an atmospheric structure conducive to wave ducting further increases the likelihood that a given large-amplitude midlatitude disturbance in the surface pressure–wind field is an MGW. Under average conditions, positive static stability throughout the atmospheric column may enable tropospheric gravity waves to rapidly propagate into the stratosphere and dissipate, with minimal opportunity to influence the weather in the lower troposphere (e.g., Gossard and Hooke 1975). Under certain conditions, however, the environment can effectively trap or duct gravity waves in the lower troposphere by reflecting gravity wave energy back downward, thereby maximizing the opportunity for both wave amplification and interactions with weather (Lindzen and Tung 1976; Hooke 1986; Fovell et al. 2006).

While wave ducting is not strictly necessary for MGW propagation in the troposphere, studies of large-amplitude MGW events in midlatitudes lend strong support to the importance of a wave duct, the theory for which is described by Lindzen and Tung (1976). As depicted in Fig. 2a, the wave duct is composed of 1) a low-level stable layer, which acts as the primary propagation medium, and 2) an overlying reflecting layer characterized by negligible static stability, which reflects wave energy back downward. The reflecting layer in midlatitude situations is often promoted by having a critical level where the wave’s horizontal phase...
velocity $c$ matches the wind component in the direction of wave propagation $U$, and where the Richardson number $Ri < 0.25$. The ducting stable layer often takes the form of an inversion, while the reflecting layer is statically neutral and/or conditionally unstable (Fig. 2b) (Marks 1975; Uccellini and Koch 1987). This vertical structure contrasts with that in the tropics, where such inversions are rare, yet partial reflection of convectively forced or low frequency gravity waves, such as diurnal gravity waves, can occur at the tropopause (Schmidt and Cotton 1990; Mapes 1993; Mapes et al. 2003; Lane and Zhang 2011; Ruppert and Zhang 2019; Ruppert et al. 2020).

According to wave ducting theory, the depth of the duct determines the vertical scale of the MGW, and in turn, its horizontal phase velocity (Lindzen and Tung 1976). In the example in Fig. 2, one-half of the vertical wavelength is contained between the ground and the height of the critical level. Sounding information can therefore be directly used to estimate the ducted phase speed, which can in turn be directly compared with a disturbance’s observed propagation. Past studies of MGW events have conducted this exercise and noted a strong match, lending evidence to confirming MGW behavior and the importance of wave ducting in such events (Bosart and Sanders 1986; Bosart and Seimon 1988; Ralph et al. 1993; Koch and O’Handley 1997; Zhang and Koch 2000; Zhang et al. 2003a; Ruppert and Bosart 2014; Du and Zhang 2019). Several studies, however, describe other environmental conditions that can support wave ducting beyond these specific criteria (Wang and Lin 1999; Fovell et al. 2006; Adams-Selin and Johnson 2013).

Many studies, including by Zhang et al. (2001), have indicated that the conditions for wave ducting are only partially met, implying some degree of leakage of energy to the stratosphere. Since MGWs are in many such cases observed to maintain amplitude or even amplify, something else must either provide this energy input or further limit vertical energy propagation. As Zhang underscored through his research, latent heating and cooling appear to
be key to this conundrum (Zhang 2000; Zhang et al. 2001, 2003a; Du and Zhang 2019). The role of both wave ducting, even if imperfect, and latent heating processes in MGW behavior has come to be referred to as *ducted wave–CISK*, where wave–CISK (Conditional Instability of the Second Kind) refers to wave amplification through the feedback with latent heating processes (Lindzen 1974; Raymond 1975; Powers and Reed 1993). We return to this issue later in relation to mechanisms for MGW maintenance and amplification.

**The East Coast “Snow Bomb” MGW Event of 4 January 1994**

The major MGW event of 4 January 1994 that would ultimately serve as the subject of Zhang’s Ph.D. dissertation was first documented by Bosart et al. (1998). The observational analysis conducted by Bosart et al. revealed the dramatic sensible weather effects of a major MGW wave of depression and its relation to an inland band of very heavy snow, aptly named a “Snow Bomb”, which was characterized by peak hourly snowfall rates of 10–15 cm h⁻¹ (Fig. 4b). The path of the MGW corresponds conspicuously to the southeastern boundary of the Snow Bomb, implying the potential role of the MGW in altering the overall hydrometeorological impacts of this storm. Neither the MGW nor the Snow Bomb were well forecast by operational prediction models of the time.

While the MGW was first detected as a moderate-amplitude wave packet poleward of a surface frontal boundary, the primary large-amplitude wave of depression ultimately grew from this packet as it advanced northeastward. Amplification occurred as the MGW encountered an increasingly deeper and stronger wave duct in the cold air damming region east of the Appalachians and in the vicinity of strong forcing for ascent associated with the upper-level flow pattern. Based on these observations, Bosart et al. (1998) hypothesized that MGW genesis was caused by “unbalanced flow,” and that amplification was triggered by perturbation of the wave duct by vigorous latent heat release.
An infrared satellite image from the time of MGW maturity depicts a cold cloud shield with a very sharp trailing edge, where implied MGW-induced sinking motion promotes cloud desiccation (Fig. 4a). Cloud banding implies two regions of MGW-induced sinking motion (dashed lines), consistent with the appearance of two troughs in sea level pressure (Fig. 4c; cf. also Bosart et al. 1998). Precipitation ceased in connection with the dominant, leading trough, however (Fig. 1). A caveat of this assessment of vertical MGW structure, based on radar and satellite depictions alone, is that it neglects the potential influence of wind shear and other factors on vertical structure (e.g., Koch and Golus 1988).

Captivated by this remarkable MGW event, Zhang seized the opportunity to advance our understanding through the creative application of theory, novel analysis techniques, and high-resolution numerical modeling to this case. As the central component of his Ph.D. dissertation (Zhang 2000), he used modeling to reproduce the event (Fig. 4d,e), which in turn enabled him to examine hypotheses put forth by Bosart et al. (1998). As documented in the seminal paper by Zhang et al. (2001), he invoked wavelet analysis to unambiguously track and analyze the evolution of the MGW within the context of its changing local environment – the first time this technique had ever been applied in such a manner. Wavelet analysis possesses the ability to provide localized time-frequency information, unlike traditional spectral analysis, thus permitting study of the temporal and vertical evolution of waves with specified scales. Through a set of model sensitivity tests in which the phase change of water substance was disallowed for varying periods of time, Zhang also provided convincing evidence that convective latent heating was crucial to MGW amplification through the wave–CISK feedback (discussed later) but did not directly trigger the waves. Rather, through novel quantitative techniques described shortly, he provided the most convincing evidence yet for the emerging argument that MGW genesis is triggered by synoptic-scale flow imbalance.
Synoptic-scale Setting of Large-amplitude MGW Events: Implications for Geostrophic Adjustment

By the time of the 4 January 1994 event, the number of case studies of large-amplitude MGW events had grown to a sufficiently large number to accommodate a robust synthesis of the salient, recurring synoptic-scale environmental patterns. Uccellini and Koch (1987) took this opportunity and conducted a review of documented cases up to that time in order to examine various hypotheses on MGW behavior. Variability among cases notwithstanding, their review suggested a clear synoptic-scale flow paradigm for MGW occurrence, which ultimately provided the greatest support for the role of upper-level flow imbalance in MGW genesis via geostrophic adjustment. Zhang’s subsequent work both during and following his Ph.D. reflected his desire for a deeper, more quantitatively supported understanding of how this imbalance arises within this pattern and how it in turn excites MGWs. Prior to delving into those exploits, it is worth first discussing this flow paradigm and understanding how it lends evidence to the role of flow imbalance in MGW genesis.

The flow paradigm associated with MGW occurrence is schematically illustrated in Fig. 5, which draws on the review by Uccellini and Koch (1987) and incorporates further insights of Zhang et al. (2001) based on their numerical modeling study of the 1994 Snow Bomb event. This figure conveys the progression of the baroclinic lifecycle in connection with the low-level frontal occlusion process. MGW genesis is favored as a jet streak propagates away from the upper-level trough axis and toward the inflection axis between the trough and downstream ridge, especially when 1) the horizontal wavelength between this trough and ridge shortens with time, and 2) the jet streak separates from its geostrophic component at the base of the trough and propagates toward the ridge. These developments are favored by the maturing
low-level cyclone, and especially the latent heating accompanying convection beneath the inflection axis, where warm, moist air is lifted by the occluding frontal system.

The increasing separation between the geostrophic wind maximum and the actual jet streak is the manifestation of intensifying upper-level flow imbalance, which can be understood as follows. An air parcel near the inflection axis is located in the exit region of the geostrophic wind maximum (black arrow in Fig. 5) where, according to quasi-geostrophic theory, a thermally indirect ageostrophic circulation directed toward higher geopotential heights should occur to convert the excess kinetic energy of the jet to potential energy (Holton 1992). However, the parcel is simultaneously located in the entrance region of the actual jet streak (green arrow in Fig. 5). As a result, a thermally direct ageostrophic circulation instead occurs at the inflection axis, in direct contradiction with the predictions of quasi-geostrophic theory. Thus, the flow is unbalanced, and both MGW emission and large time-tendencies in divergence may result. Flow curvature plays a central role here by promoting parcel acceleration from the trough to the downstream ridge. This acceleration, and hence the flow imbalance, therefore intensifies as the wavelength between the trough and ridge shortens, all else being equal.

Based on detailed analysis of the 1994 Snow Bomb event, Zhang et al. (2001) proposed an update to this paradigm to link the development of flow imbalance with the tropopause folding process, which is common among especially strong baroclinic systems. As incorporated in Fig. 5 (lower-right), flow imbalance is greatest in the region of strongest ascending motion within the tropopause fold, which is part of the aforementioned thermally direct ageostrophic circulation coupled to the entrance region of the jet streak. Hence, MGWs may first be detected immediately downstream of this region (Fig. 5). Once excited, the MGWs propagate both horizontally and vertically toward the downstream ridge axis as small- to moderate-amplitude waves. Provided the conditions are suitable for wave ducting and that these waves perturb the low-level stable layer, they may become ducted. Since the locus of
flow imbalance and MGW emission is often on the cold side of a surface warm/stationary front, this synoptic-scale configuration is indeed usually favorable for wave ducting, provided the free troposphere above the stable layer contains a critical level and is weakly stratified (Figs. 2 and 5).

This MGW genesis paradigm is akin to and has often been referred to as geostrophic adjustment, which originates from the classical studies of Rossby (1938), Cahn (1945), and Blumen (1972). These studies described geostrophic adjustment as the process by which the atmosphere removes geostrophic imbalance between the mass and momentum fields by emitting IGWs. A refinement that Zhang strongly advocated in the context of realistic (i.e., curved) flows, however, is to refer to this process as the more generalized balance adjustment, since the balance relevant to strongly curved flows is the higher-order gradient wind balance, rather than strict geostrophic balance (Zhang et al. 2001; Zhang 2004).

**Quantifying Flow Imbalance**

While subjective analysis of the synoptic-scale flow configuration helps in identifying areas of potential flow imbalance and MGW genesis, Zhang made a seminal advancement by providing detailed new guidance as to how to both interpret and quantitatively assess the presence of upper-level flow imbalance (Zhang 2000; Zhang et al. 2000, 2001). He systematically appraised and evaluated several key approaches for diagnosing flow imbalance, including the cross-stream Rossby number, the nonlinear balance equation, potential vorticity inversion, the psi vector, and the generalized omega equation.

Based on these assessments, Zhang determined that the greatest indication of flow imbalance is provided by the cross-stream Rossby number and the residual of the non-linear balance equation (Zhang et al. 2000). As a measure of geostrophic imbalance, the Lagrangian Rossby number is the ratio of parcel acceleration to the Coriolis force. This term can be well
approximated by the cross-stream Rossby number, which is the ratio of the ageostrophic wind normal to the flow to the total wind (Koch and Dorian 1988).

The nonlinear balance equation provides a more appealing and accurate approach for shorter time scales, however, since it directly incorporates the higher-order balance valid for strongly curved flows (Uccellini and Koch 1987; Koch and Dorian 1988; Koch and O’Handley 1997; Zhang et al. 2000). This equation is specifically the residual of the four largest terms in the total divergence tendency equation, after ignoring terms representative of situations smaller in scale than synoptic. It can be written as

\[ \Delta \text{NBE} = 2J(u, v) - \beta u + f\zeta - \nabla^2 \phi, \]

where \( J(u, v) \) is the Jacobian of horizontal flow, \( \beta \) is the latitudinal variation of the Coriolis parameter, \( \zeta \) is vertical relative vorticity, and \( \phi \) is geopotential. Large-magnitude \( \Delta \text{NBE} \) highlights where the flow is unbalanced and large temporal changes in divergence can be expected. When the horizontal wavelength between the upstream trough and the downstream ridge dramatically shortens with time, for example, this term grows due especially to growth of the Jacobian (first) and Laplacian (fourth) terms in this equation.

Leveraging his numerical modeling study of the 1994 Snow Bomb MGW event, Zhang directly applied each of these quantitative approaches to diagnose the measure of flow imbalance in that event, which represented many firsts, and in turn yielded the clearest evidence yet for the importance of flow imbalance in MGW genesis (Zhang 2000; Zhang et al. 2000, 2001). Determined to further lift the veil on MGW genesis, however, he subsequently brought these diagnostic techniques into more idealized numerical model configurations to examine MGW emission through flow imbalance more deeply, in the context of baroclinic development of the jet–front system (Zhang 2004). That matter is the subject of the next section.
Spontaneous Balance Adjustment

The first numerical demonstration of IGW genesis within an evolving baroclinic jet–front system was made by O'Sullivan and Dunkerton (1995) based on a 3D hemispheric hydrostatic primitive equation model (Fig. 6). IGWs with intrinsic frequencies between $f$ and $2f$ were identified principally in the upper-tropospheric jet streak exit regions during the stage of baroclinic wave development when parcel accelerations were largest. However, the basic nature and propagation of these waves were at odds with observed MGWs: since the smallest horizontal grid spacing used in their hemispheric model was approximately 50 km, the simulated gravity waves possessed overly long horizontal wavelengths of 600–1000 km.

Zhang later exploited the leaps in computing power to push this experiment to much finer resolution, in addition to leveraging the new insights into MGW genesis that arose through his Ph.D. work (Zhang 2000, 2004). He conducted a high-resolution idealized baroclinic wave experiment using a multiply nested mesoscale numerical model, with grid spacing as fine as 3.3 km, which successfully produced long-lived vertically propagating MGWs with horizontal wavelengths of ~100–200 km originating from the exit region of the upper-tropospheric jet streak (Fig. 7). Drawing on his Ph.D. work, he further demonstrated that the timing and location of MGW generation are well predicted by $\Delta\text{NBE}$.

Zhang’s seminal 2004 study went far beyond a proof-of-concept. By emphasizing that the complex, curved flows in nature often adhere to a high-order state of balance while strongly departing from geostrophic balance, he advocated for a move away from geostrophic adjustment and the adoption of balance adjustment as its generalization. To support this argument, Zhang invoked potential vorticity inversion (Davis and Emanuel 1991) to produce a precisely balanced, yet baroclinically unstable, initial condition characterized by curved flow. His experiment thus demonstrated how the progression of the baroclinic life cycle leads to the
continuous development of pockets of ΔNBE, and hence to spontaneous, continuous MGW emission. In Zhang’s subsequent work, the name of the paradigm was adapted to spontaneous balance adjustment to account for this continuous nature (Wang and Zhang 2010; Wei and Zhang 2014).

Five distinct gravity wave modes can in fact be identified in Zhang’s (2004) experiment (annotated as WP1–WP5 in Fig. 7). This discovery prompted a flurry of new research to better understand the distinct characteristics, sources, and behavior of these modes, much of it led by Zhang, his students, and his collaborators. Plougonven and Snyder (2005, 2007) demonstrated the influence of the background sheared flow on the character and propagation on these modes. Wang and Zhang (2007) found that the intrinsic frequencies of the wave packets in the jet-exit region (i.e., WP1 and WP2) tend to increase with the growth rates of the baroclinic waves, underscoring the strong sensitivity to the timescale of the evolution of the balanced flow (Reeder and Griffiths 1996). Ray-tracing analysis conducted by Lin and Zhang (2008) and Wei and Zhang (2015) suggests that WP3 and WP4, however, are likely forced by the surface frontal system (Griffiths and Reeder 1996). Many similarities between the idealized simulations and MGW case studies in terms of wave characteristics and wave front orientation relative to the background flow lend credence to the idealized model frameworks (Lane et al. 2004; Koch et al. 2005).

More broadly, Plougonven and Zhang (2007) extended the work of Zhang (2004) by formalizing spontaneous balance adjustment theory. Using scale analysis, they directly linked ΔNBE to the right-hand side of an analytical wave equation linearized around a balanced background flow. They further demonstrated through this approach how the imbalance can be attributed to the distinct forcing by vorticity, divergence, and thermodynamic sources.

Further progress in understanding spontaneous balance adjustment was achieved through model experiments using an idealized vortex-dipole jet. As arguably the simplest
analogue to a jet streak possible, this framework is valuable for more deeply understanding gravity wave emission. Examples of this framework are provided from the work of Snyder et al. (2007) and Wang et al. (2009) in Fig. 8. Zhang played an important role in these studies, as a coauthor of Snyder et al. (2007) and as the Ph.D. advisor of Wang (Wang 2008), respectively. Using a Boussinesq primitive equation model, Snyder et al. (2007) was the first study to isolate and examine spontaneous MGW emission via balance adjustment in the vortex-dipole jet framework. In their experiment, IGW packets spontaneously form in the jet exit region, which are characterized by low intrinsic frequency between \( f \) and \( 2f \), phase lines that are nearly perpendicular to the jet, and shrinkage of wavelengths toward the periphery of the jet due to flow deformation (Fig. 8a) (McIntyre 2009; Wang and Zhang 2010). Wang et al. (2009) extended this experiment to a fully nonhydrostatic mesoscale model and further highlighted the upward- and downward-propagating IGWs that are excited by an upper-level jet streak (Fig. 8b). To better understand the source mechanisms of these waves, Wang and Zhang (2010) next developed a linear numerical model based on the heuristic semi-analytical derivation of Plougonven and Zhang (2007). Their model successfully replicated the salient characteristics of the gravity waves (i.e., location, phase, and amplitude) from the fully nonlinear model. By isolating the dynamic sources of flow imbalance, they found that the vorticity component was the leading contributor to gravity wave emission in their model, though they cautioned that this result is likely flow-dependent.

The idealized vortex-dipole framework has also been invoked to better understand how the nature of the forcing and environmental wind shear affect the scale and propagation characteristics of gravity waves (Snyder et al. 2007; Wang et al. 2009, 2010). By forcing a linear model with a prescribed, stationary gravity wave forcing corresponding to a dipole-jet, Wang et al. (2010) demonstrated that wave attributes such as wavelength and phase are strongly constrained by the environmental wind shear and flow deformation pattern. This influence of
the sheared flow is referred to as wave capture (Badulin and Shirma 1993; Bühler and McIntyre 2005).

Progressing toward a more complete picture of MGW behavior in nature, subsequent studies, including by Zhang and his students (e.g., Wei 2015; Sun 2017), incorporated moisture and diabatic forcing into the idealized baroclinic jet–front model framework (Mirzaei et al. 2014; Wei and Zhang 2014, 2015). Wei and Zhang (2014) performed a convection-allowing model experiment using this framework to examine the sensitivity of MGW behavior to convective instability. In a set of otherwise identical simulations, they varied the initial moisture content progressively from 0–100% of a reference relative humidity state, as depicted in Fig. 9a–f. This experiment highlights the extreme sensitivity of both MGW amplitude and wavelength to moist convection. The fully dry experiment (Fig. 9a) simulates gravity wave modes that are consistent with previously discussed dry simulations, reproducing all five distinct MGW modes identifiable in Zhang (2004; annotated in Fig. 7). With just 20% initial moisture included (Fig. 9b), however, the development of moist convection causes a clear, considerable amplification of MGWs, especially of those wave packets situated in the ridge and downstream northwesterly flow (i.e., WP5). Further increases of moisture lead to progressively more vigorous moist convection and hence greater feedback with MGW emission, MGW amplitude, and the pace of the baroclinic life cycle. With greater moisture, the baroclinic life cycle progresses more rapidly, and hence MGWs are emitted sooner (Fig. 9f). Wei and Zhang (2014) identified six distinct wave packets (see caption for Fig. 9). Wei and Zhang (2015) then invoked two-dimensional spectral decomposition and four-dimensional time-dependent ray-tracing techniques to demonstrate that the distinctions between the wave packets of the dry and moist simulations were partly caused by the direct influence of moist convection on the generation and propagation of the wave modes.
To quantify the upscale influence of these gravity waves in the idealized jet–front model framework, Wei et al. (2016) further documented the resolved gravity wave spectral characteristics in the dry and moist models and estimated the associated wave-induced momentum fluxes. This was the first time that the relationships between gravity wave momentum flux and phase velocity were presented in detail based on an ensemble of high-resolution idealized moist baroclinic jet–front systems. This effort highlighted various key input parameters for improving the parameterization of non-orographic gravity waves, which is still an area of substantial uncertainty and oversimplification (e.g., Haynes 2005; Richter et al. 2010; Wei et al. 2019; Plougonven et al. 2020) and is also poorly constrained by observational tools (Wu and Zhang 2004; Alexander et al. 2010; Zhang et al. 2013, 2015).

**Wave Amplification: The Role of Moist Convection**

Beyond deepening our understanding of spontaneous balance adjustment and its role in MGW genesis, Zhang’s research also more firmly established the key role of moist convection in MGWs. His numerical modeling study of the 1994 Snow Bomb MGW event demonstrated that, while convection did not directly trigger MGW genesis, it was essential to wave amplification and maintenance in that event (Zhang 2000; Zhang et al. 2001). Examined in this section are the various conceptual frameworks and hypotheses concerning the role of convection in large-amplitude MGW events, highlighting the contributions by Zhang toward this still-unresolved problem.

Moist convection is a well-established mechanism for exciting high-frequency gravity waves—as a localized buoyancy source, a perturbation of stratified layers, and a flow obstacle (Pierce and Coroniti 1966; Clark et al. 1986; Bretherton 1988; Hauf and Clark 1989; Fovell et al. 1992; Lin et al. 1998; Pandya and Alexander 1999; Lane et al. 2001; Beres et al. 2002; Beres 2004). In some circumstances, such waves can serve as a mechanism for convective upscale...
growth and propagation through their remote environmental influence (Mapes 1993; McAnelly et al. 1997; Fovell 2002; Fovell et al. 2006; Tulich and Mapes 2008; Adams-Selin and Johnson 2013; Stephan et al. 2016, 2020; Parsons et al. 2019; Adams-Selin 2020). While convective generation is unlikely to play a role in the origin of most large-amplitude MGW events, evidence strongly implicates the role of moist convection as a mechanism for wave maintenance and/or amplification in the large-amplitude and long-duration events. Since wave ducting is very often imperfect, mechanisms such as the feedback with convection are required to explain the maintenance of MGWs that at times endure for many hours. Furthermore, wave ducting cannot explain the dramatic amplification of MGWs into ~10-hPa disturbances, as in many documented events.

The most commonly invoked hypothesis to explain MGW maintenance and amplification through convection–MGW interaction is ducted wave–CISK (Powers and Reed 1993; Bosart et al. 1998; Koch et al. 2001; Zhang et al. 2001; Du and Zhang 2019). This feedback is schematically depicted in Fig. 10a. This mechanism conceptually operates as follows: for a gravity wave propagating in a wave duct, the gravity wave supplies the divergence and convergence patterns that regulate moisture and convection, while the gravity wave is, in turn, supplied energy by the divergence of convective mass fluxes. Thus, convection localized to the ascending branches of the gravity wave may provide a source of energy for waves, thereby compensating for energy leakage in an inefficient wave duct, or potentially causing wave amplification. According to wave-CISK theory, the strongest upward motion within the convection occurs around the critical level (Fig. 10a) (Zhang et al. 2001; Du and Zhang 2019). This is fundamentally distinct from pure linear ducting theory, wherein the gravity wave signal decays to zero at the critical level (Fig. 2a; Lindzen and Tung 1976). Furthermore, wave ducting was not incorporated in the original wave–CISK theory (Lindzen...
1974; Raymond 1975). Nonetheless, by incorporating these two frameworks together, they have become in effect a unified conceptual paradigm, referred to as *ducted* wave–CISK.

Zhang et al. (2001) argued that ducted wave–CISK was the primary wave amplification mechanism in the 1994 Snow Bomb event. In their numerical model experiment, a “Fake-Dry” simulation revealed that the dominant MGWs weakened in the absence of latent heating, which was set to zero. Although Powers and Reed (1993) found similar results using the “Fake-Dry” technique, they had interpreted the role of convection as a mechanism of MGW genesis. Zhang, in contrast, emphasized that there are likely multiple feedbacks at play here. He noted that in the absence of latent heating, the intensity of the cyclone is weaker, and the wavelength between the upper-level trough and ridge is longer, which results in weaker parcel divergence tendency and flow imbalance (also see Pokrandt et al. 1996). In other words, a likely key mechanism for MGW genesis is suppressed through the removal of latent heating (Fig. 9; Wei and Zhang 2014). Zhang et al. (2001) applied several powerful approaches to separate the distinct scales of convective feedback, including wavelet analysis and a detailed examination of the roles of microphysical processes. By doing so, they confirmed that while MGW–convection interaction did not play a role in MGW genesis, it was vital to MGW maintenance and amplification, with the likely role of wave–CISK implied by large-magnitude vertical motion centered around the critical level.

While wave–CISK offers a valuable conceptual framework for understanding convection-wave interactions, as a strict, linear framework for making quantitative predictions of wave behavior, it has important deficiencies. Theoretical wave CISK-driven modes display an instability that increases monotonically with wavenumber, which is inconsistent with reality, unless shear or time-lagged updrafts are artificially imposed. Further, this theory invokes the *quasi-equilibrium* assumption traditionally applied for the treatment of tropical convection on the large scale, which is only valid if the life cycle of convective clouds is
negligibly small compared to the wave-induced convergence forcing; this assumption has
necessitated an assumed phase lagging in the model, which may be called into question given
observed MGW and squall line characteristics. Finally, in an idealized convection-allowing
modeling study, Lane and Zhang (2011) found a large discrepancy between the speed of
convection and MGWs – evidence against the role of wave–CISK. Ducted wave–CISK has,
onetheless, provided important impetus for conceptually examining various ways convection
and gravity waves interact (e.g., Zhang et al. 2001; Ruppert and Bosart 2014; Du and Zhang
2019).

Another pattern that emerges from a review of many large-amplitude MGW events is
the strong parallel between MGWs – especially singular waves of depression – and squall line
wake low–mesohigh couplets (Bosart and Seimon 1988; Koch et al. 1988; Koch and Siedlarz
1999; Coleman and Knupp 2009, 2011; Ruppert and Bosart 2014). This conceptual framework
is schematically depicted in Fig. 10b. When a squall line is organized according to the canonical
leading convective line–trailing stratiform structure, its dominant surface pressure features are
usually 1) the surface mesohigh, which develops within and immediately trailing the intense
line of convection; 2) the wake low, which forms at the back edge of the trailing stratiform
region; and 3) the nonhydrostatic presquall low (Fujita 1955; Johnson and Hamilton 1988;
Johnson 2001). When the wake low and mesohigh are fully developed, they are separated by
a strong pressure gradient between the leading convective line and trailing stratiform region
(Johnson and Hamilton 1988). While below-cloud evaporation of rainfall and hydrometeor
loading are the long-established drivers of mesohigh formation (Fujita 1959; Sanders and
Emanuel 1977; Fovell and Ogura 1988; Nicholls et al. 1988), an understanding of wake low
formation has remained more elusive (Johnson 2001). Highlighting the role of evaporative
cooling aloft for driving low-level subsidence in the stratiform region, Zipser (1977) proposed
that the wake low could result from subsidence warming exceeding this latent cooling and
overshooting its level of neutral buoyancy (Johnson 2001). In support of this idea, Gallus (1996) found that evaporatively-forced sinking motion causes wake low formation only when the evaporative cooling ceases, because when this occurs, the downward inertia of the subsiding air drives adiabatic warming and lowering of surface pressure. Such overshooting implies the development of excess buoyancy, which may in turn trigger a gravity wave response (e.g., Schmidt and Cotton 1990).

As a striking parallel, a number of MGW investigators have emphasized the importance of the shutdown of evaporation at the trailing edge of the stratiform anvil for MGW amplification, which is facilitated by the dry conveyor belt in the context of the larger-scale cyclone (Bosart and Seimon 1988; Koch et al. 1988; Zhang et al. 2001; Jewett et al. 2003; Marsham et al. 2010). Zhang et al. (2001) emphasized the importance of such drying associated with a cold front aloft (Hobbs et al. 1990) in the 1994 Snow Bomb event. As this front and associated advection of mid-tropospheric dry air caught up with weak-amplitude incipient gravity waves, rapid MGW amplification and scale contraction ensued, after which time the front and wave were indistinguishable. This front and its role in relation to the MGW and precipitation are incorporated in Fig. 10b.

Of interest is the fact that the pressure couplet defined by the wake low and mesohigh largely exhibits positive $p'-u'$ correlation as is characteristic of MGWs, as noted earlier (Koch and Siedlarz 1999; Coleman and Knupp 2009). The interpretation of the wake low–mesohigh couplet as a gravity wave is further supported by the idealized linear modeling study of Haertel and Johnson (2000), which demonstrated that an imposed stratiform-like forcing characterized by latent cooling alone can excite a gravity wave response consistent with a mesohigh–wake low couplet. Notably, they found the amplitude of this gravity wave response to maximize when the motion of the cooling matches the intrinsic gravity wave phase velocity (~10–20 m s$^{-1}$). This body of work on both squall line pressure couplets and large-amplitude MGW events...
implies that strong MGWs in the form of waves of depression are dynamically synonymous to squall line wake lows in terms of convection–circulation feedback. Sudden large-amplitude pressure drops observed in typhoons undergoing extratropical transition have also been attributed to similar microphysical/gravity wave arguments (Matsumoto and Okamura 1985; Fudeyasu et al. 2007).

In addition to MGW–convection interaction, a nonlinear feedback process unrelated to moist physics has also been proposed as a candidate mechanism for maintaining long-lived MGWs. Since MGWs are dispersive waves (i.e., a wave packet composed of the superposition of waves with distinct phase velocities), the waves will be flattened over time unless the wave dispersion is balanced by nonlinear wave steepening. When this balance occurs, a long-lasting, large amplitude, solitary wave can form (Lin and Goff 1988), which is typical of high-amplitude MGW cases. Rottman et al. (1992) and Rottman and Einaudi (1993) showed that a critical level located just below the tropopause can act as a nearly perfect wave reflector for waves displaying horizontal length scales of order 100 km, resulting in gravity wave overreflection and rapid amplification. The potential role of such nonlinear processes in the maintenance and amplification of observed long-lived MGW events remains unresolved, particularly in the presence of deep moist convection.

Zhang’s work played a key role in quantitatively demonstrating the importance of moist convection in MGW maintenance and amplification, which in turn helped to stimulate new research into the dynamics of MGW–convection interaction. Furthermore, it was through this work that Zhang became entranced by the fundamental energetic role of moist convection, not only in MGW amplification, but in the upscale growth of errors in convection-allowing models, and hence in weather prediction more broadly.

Mesoscale Weather Prediction
Despite the major progress made in operational numerical forecast systems, Zhang was fascinated by the challenges of accurately predicting the hazardous weather associated with large-amplitude MGW events. The model frameworks he devised and invoked to simulate the high-impact weather effects of MGWs provided excellent testing platforms, in turn, to explore the underlying causes for these challenges. Alongside his efforts to examine MGWs, this drive to more deeply understand predictability would ultimately become a central, and eventually the dominant, component of his research. Discussed here are his most salient contributions to our understanding of predictability.

Regarding the practical predictability of high-impact weather events, the convection-allowing model experiment of Zhang et al. (2003b) emphasized that precipitation forecasts are highly sensitive to model resolution, physics schemes, and the initial conditions generated by data assimilation. This study, conducted in the context of the surprise snowstorm of 24–25 January 2000, highlights the importance of developing high-density observational networks, high-performance computing facilities, advanced data assimilation techniques, and advanced modeling systems. Zhang et al. (2003b) additionally showed that, even with a perfect model and nearly perfect initial conditions, the prediction of precipitation can degrade dramatically within 1-2 day forecast lead time, raising the question of the intrinsic predictability limit for such weather events (Lorenz 1969). Further analysis demonstrated that moist processes are the key drivers of the initial rapid small-scale error growth, which may inherently limit the intrinsic predictability of winter snowstorms at the mesoscale (Zhang et al. 2003b; Hohenegger and Schär 2007; Sun and Zhang 2016).

By analyzing the underlying nature of error growth within the convection-allowing model experiments of Zhang et al. (2003b), Zhang et al. (2007) proposed a novel three-stage error growth model. This model is depicted in Fig. 11 and can be described as follows: (1) the initial convective growth stage, which begins with convective instability followed by rapid
error saturation at small scales ($O(1 \ h)$), (2) the intermediate adjustment stage, during which small-scale errors project to the balanced fields with a time scale of $O(10 \ h)$, and (3) the large-scale growth stage, where error grows with large-scale baroclinic instability. The error growth dynamics are dominated by convective and baroclinic instabilities at the small (stage 1) and large scales (stage 3), respectively. Zhang et al. (2007) also hypothesized several mechanisms that potentially underlie the upscale error transition at stage 2, which include gravity waves produced by convective latent heating followed by geostrophic adjustment, cold pools at low levels, and instability of the low-level cold front itself.

While the dependence of error growth processes on dynamics at different scales makes it difficult to derive a uniform quantitative mathematical description, the above conceptual three-stage error growth model provides a fundamental framework for understanding the mesoscale predictability of midlatitude moist baroclinic waves. Follow-up studies have added support to this conceptual model and developed more comprehensive models based on it (e.g., Selz and Craig 2015). Extending Zhang’s three-stage model, Baumgart et al. (2019) added a fourth stage to describe the evolution of the errors up to planetary scale. Bierdel et al. (2018) further investigated the error evolution in stage 2 and showed that geostrophic adjustment may be responsible for upscale error growth through mesoscale processes.

In a fully turbulent flow, errors are transferred between eddies of different scales through an inverse cascade, with the transfer rate determined by the background kinetic energy spectrum (Lorenz 1969; Rotunno and Snyder 2008). While the homogeneous turbulence view of error growth is highly idealized, this error growth behavior is found to be similar to realistic, full physics simulations (Durran and Gingrich 2014; Leung et al. 2020; Sun and Zhang 2020). A defining aspect of the kinetic energy spectrum around the mesoscale is its -5/3 slope (Nastrom and Gage 1985). Recent studies highlight the key role of IGWs in this portion of the spectrum (Callies et al. 2014; Sun et al. 2017), implying that IGWs may play a key role in the
transfer of error between scales. Further work is necessary to examine this question and the intrinsic predictability limit of large-amplitude gravity waves, however.

While Zhang’s work on predictability was initially motivated by his attempts to understand and accurately predict large-amplitude MGW events, predictability became a central focus of his career. Indeed, in one of his last lead-author publications, he presented an outlook for the future potential of midlatitude weather predictability (Zhang et al. 2019b). His work on this subject has helped shape both subsequent research and the design of NWP products. The prediction of large-amplitude MGW events, and high-impact mesoscale weather events in general, remains a critical challenge. His dedication to improving and advancing data assimilation, especially to better utilize valuable satellite radiance data, have helped pave the way towards addressing this challenge (Zhang et al. 2019a; Hartman et al. 2021).

The Road Forward

Before Zhang’s premature passing in 2019, his outsized research contributions over just a 20-year period helped to pin down the underlying nature and dynamics of MGW genesis in association with upper-level flow imbalance, the influence of environmental flow on MGW propagation, and the role of moist convection in MGW maintenance and amplification. He drew tremendous inspiration and insight from the observational studies of large-amplitude MGW events that came before his time, beginning his career with a numerical modeling study of one such event. In addition to providing new techniques to quantify upper-level flow imbalance, the theoretical and modeling work he conducted, together with many of his students and other collaborators, greatly strengthened our grasp of the emission of MGWs from this imbalance.

Zhang’s legacy of contributions paves the way for new research yet to be done. Consistent with his career-long dedication to both a deep theoretical understanding and real-
world forecast challenges, this subject presents valuable opportunities for new research on both fronts. His work underscored the importance of moist convection for MGW maintenance and amplification; yet the precise nature of this interaction as it manifests in observed large-amplitude events remains elusive. In particular, we lack a theoretical understanding of how initially weak MGW wave packets consolidate and dramatically amplify into singular waves of depression as they couple with organized precipitation. Although this dramatic evolution is common to many large-amplitude cases, these events are relatively rare (several per year) compared to the flow configurations that excite small-amplitude MGWs. Thus, the special ingredients that promote such dramatic MGW amplification – potentially related to the specific distribution of moisture and convective instability or convective organization – remain to be identified. High-resolution numerical modeling case studies of observed large-amplitude MGW events, which are strikingly scant since Zhang’s Ph.D. work, would likely help guide new theoretical insights into this evolution.

Furthermore, given the substantial impacts of large-amplitude MGWs on the midlatitude sensible weather, efforts to develop, test, and apply a new real-time monitoring and prediction system for MGWs would be highly valuable to operational weather forecasting. Such an undertaking would directly align with the NOAA/NWS Weather-Ready Nation program (Uccellini and Ten Hoeve 2019), which is intended to improve public warning services and enhance community response and resiliency to hazardous weather impacts. Such effort would also help preserve the spirit of Zhang’s dedication to addressing real-world forecast challenges.

Acknowledgments
The authors were inspired to write this review by the important role that Fuqing Zhang played in their careers and understanding of the MGW phenomenon. Zhang was a doctoral student of SEK at North Carolina State University (NCSU). Through a shared passion for studying MGWs, he became a mentee of LFB and a close colleague to AS. As students of Zhang, either formally or in effect, the careers of JHR, XC, YD, QS, and JW have all been greatly shaped for the better by Zhang. We acknowledge Stephen Saleeby (Colorado State University), Zhang’s graduate school office mate at NCSU, for valuable discussions. We thank Dan Lindsey (NOAA) for kindly providing infrared satellite imagery for the 4 January 1994 event. We thank Todd Lane and two anonymous reviewers for their valuable comments that helped improve the manuscript. JHR acknowledges support from the National Science Foundation (NSF) grant AGS-1712290 and the Center for Advanced Data Assimilation and Predictability Techniques at Pennsylvania State University; XC is supported by the Office of Science of DOE Biological and Environmental Research as part of the Regional and Global Modeling and Analysis program; YD is supported by the National Natural Science Foundation of China (grant 42075006) and Guangzhou Science and Technology Plan Projects (grant 202002030346); YS is supported by the National Oceanic and Atmospheric Administration (grant NA18OAR4320123); JW is supported by the National Natural Science Foundation of China (grant 42075005) and Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies (2020B1212060025); and LFB is supported by NSF grant AGS-1854886. JHR conceived the original plan for this manuscript, JHR and SEK led its drafting, and all other coauthors otherwise contributed equally.


34


Accepted for publication in Bulletin of the American Meteorological Society. DOI 10.1175/BAMS-D-20-0005.1.
Uccellini, L. W., 1975: A Case Study of Apparent Gravity Wave Initiation of Severe
4093(1975)103<0497:ACSOAG>2.0.CO;2.

——, and S. E. Koch, 1987: The Synoptic Setting and Possible Energy Sources for

——, and J. E. Ten Hoeve, 2019: Evolving the National Weather Service to build a

Wang, S., 2008: Gravity Waves from Vortex Dipoles and Jets, PhD dissertation, Texas A&M
University. 189 pp. Available from https://hdl.handle.net/1969.1/ETD-
TAMU-2875.

——, and F. Zhang, 2007: Sensitivity of Mesoscale Gravity Waves to the Baroclinicity of

——, and ——, 2010: Source of Gravity Waves within a Vortex-Dipole Jet Revealed by a

——, ——, and C. Snyder, 2009: Generation and Propagation of Inertia–Gravity Waves from

——, ——, and C. C. Epifanio, 2010: Forced gravity wave response near the jet exit region

Wang, T.-A., and Y.-L. Lin, 1999: Wave Ducting in a Stratified Shear Flow over a Two-

Pennsylvania State University, 183 pp. Available from
https://etda.libraries.psu.edu/files/final_submissions/10596.


——, ——, and J. H. Richter, 2016: An Analysis of Gravity Wave Spectral Characteristics in
15-0316.1.

Gravity Waves with Unbalanced Large-Scale Flows: Pseudomomentum-Flux
18-0337.1.

with satellite observations and a mesoscale model. *J. Geophys. Res. Atmos.*, **109**, n/a-n/a,

Zhang, F., 2000: Unbalanced dynamics and topography in the generation of mesoscale
Fuqing Zhang was born to a humble home in China, on a tiny island in the Yangtze River. After earning his bachelor’s and master’s degrees in atmospheric science from Nanjing University in 1991 and 1994, he traveled to the U.S. to earn his Ph.D. from North Carolina State University, graduating in 2000. Following a postdoctoral fellowship at NCAR, he took a faculty position at Texas A&M University in 2001 and moved to Penn State University in 2008. A mere 19 years after earning his Ph.D., he was promoted to Distinguished Professor at Penn State, the same year of his premature passing.

Fuqing was inspired by observational scientists and theoreticians alike. He took up the mantle on the geostrophic adjustment problem first laid out by Carl-Gustaf Rossby, his direct academic ancestor, and the matter of upscale error growth and its role in predictability, as first pioneered by Ed Lorenz. With a profound drive to address real-world forecast challenges, he conducted all of his research with a determination to find practical insights and applications.

While Fuqing left lasting marks across multiple branches of atmospheric science, no topic held such sway over his pursuits as did gravity waves. His work on this subject was groundbreaking. In his doctoral research, he shed new light on the role of moist-convective processes in gravity wave amplification and detailed new techniques for unambiguously quantifying flow imbalance, which in turn provided a novel approach for ensuring a balanced initial condition in forecast models. From this foundation, he later went on to develop “spontaneous balance adjustment” theory, which extended Rossby’s geostrophic adjustment problem to complex, three-dimensional flows and illuminated the nature of inertia–gravity wave genesis in the real world.
Complementary to his work on gravity waves was his career-long pursuit to better understand predictability, from fundamental theory to direct forecast application. Extending Lorenz’s original ideas, he demonstrated how small-amplitude initial condition errors fundamentally limit the predictability of severe weather events, highlighting the leading role of moist physics in error growth. His multistage conceptual model of upscale error growth tied to moist processes has now been widely adopted internationally.

Upon moving to Penn State, Fuqing founded and led the Center for Advanced Data Assimilation and Predictability Techniques. To improve the forecasting of hurricanes and severe weather, he developed the PSU WRF-EnKF analysis and prediction system through this center and helped pioneer the assimilation of high-resolution Doppler radar and satellite radiance measurements. Run in real time since 2008, this experimental system has continually outperformed official hurricane intensity predictions, in turn helping to spur the periodic adoption of novel assimilation techniques in operation models. The insights and new techniques arising from his work leave an indelible mark on our understanding of weather and its prediction, both within academia and at operational centers around the world.

Undoubtedly, one of his greatest contributions to our field is the boundless energy and enthusiasm that he devoted to engaging, mentoring, and inspiring students and young scientists throughout his short career. Those who knew him will remember his welcoming, warm spirit, and his penchant for humor and friendly competition, which enlivened every meeting he attended.

List of Awards and Achievements

Having authored more than 200 peer-reviewed publications, Fuqing ranked first on the list of the most impactful scientists during 2011–2015 in the category “Meteorology and

Accepted for publication in Bulletin of the American Meteorological Society. DOI 10.1175/BAMS-D-20-0005.1.
Atmospheric Science” based on ISI Web of Science data analytics by the Chinese Academy of Sciences. He received numerous awards over his career in recognition of his research and its many scientific and societal impacts. In 2018 he was awarded the prestigious Pennsylvania State University Faculty Scholar Medal, before being promoted to Distinguished Professor in 2019. He received two major awards from the AMS: the 2009 Clarence Leroy Meisinger Award, for his “outstanding contributions to mesoscale dynamics, predictability and ensemble data assimilation,” as well as the 2015 Banner I. Miller Award, for “valuable insights into incorporating real-time airborne Doppler radar measurements via ensemble data assimilation, leading to improvements in forecasts of tropical cyclone track and intensity.” He was also awarded the prestigious Pennsylvania State University Faculty Scholar Medal in 2018 and the Joanne Simpson Medal from AGU in 2019. He was an elected fellow of both the AMS and AGU and was selected as the 2015 Rossby Fellow in the International Meteorological Institute of Sweden.
Figure 1. Mesoanalysis for the large-amplitude mesoscale gravity wave (MGW) of 4 January 1994 in the Northeast U.S. (upper) Composite base reflectivity (shaded, $\Delta = 10$ dBZ, beginning 15 dBZ) from the three WSR-88D radar sites (ENX, BOX, and OKX) denoted by squares, and manual analysis of 20-minute pressure change (solid for pressure falls, dashed for rises; contoured at 0 and ±0.5, 1, 2, 4, and 8 hPa). The vector denotes the approximate MGW horizontal phase velocity ($c$). (lower) Time series of maximum (red) and minimum (blue) 30-s wind speed (m s$^{-1}$) at Boston (BOS), MA, and perturbation pressure $p'$ (black; hPa) at Bedford (BED), MA. Station locations are denoted by stars in the map. $p'$ axis is reversed for comparison with wind speed. While $p'$ has been detrended, the inset depicts a longer time series of surface pressure at BED without detrending. Adapted from Bosart et al. (1998).
Figure 2. (a) Schematic depiction of a ducted linear MGW train propagating toward the right in a wave duct and (b) the corresponding Skew T–log p diagram. (a) The wave duct traps gravity wave energy in the lower troposphere, which is made up of 1) the ducting layer of strong static stability (curves denote isentropes) and 2) an overlying reflecting layer characterized by negligible static stability and containing a wave critical level, where the gravity wave’s phase velocity $c$ (thick arrow) matches the flow component in the direction of propagation $U$ and the Richardson number $R_i < 0.25$. Clouds and precipitation are enhanced in and slightly lagging the region of greatest gravity wave-induced boundary layer convergence and lifting (thin arrows). The anomalies in surface pressure $p'$ and wind speed parallel to the direction of propagation $u'$ are perfectly correlated. Modified from Ralph et al. (1993). (b) corresponding Skew T–log p diagram, where red and blue curves denote temperature and dew point temperature and dotted lines denote a characteristic dry adiabat and pseudoadiabat.
Figure 3. Mesoanalysis for the large-amplitude MGW of 7 March 2008. (upper) Manual surface mesoanalysis for 0900 UTC, with mosaic base reflectivity (shaded; dBZ), sea-level pressure (black contours; Δ= 2 hPa; first two digits excluded), temperature (blue contours; °C), fronts, and 5-min cloud–ground lightning strikes (thinned for clarity; red dots). Red dashed line denotes the MGW trough axis, and the 200-km length scale indicates its approximate horizontal wavelength. Surface station data are plotted in standard convention. (lower) Meteograms for 5-min ASOS stations GWO and JAN (locations annotated in the mesoanalysis), including \( p' \) (black; hPa) and \( u' \) (gray; m s\(^{-1}\)) with time-means subtracted, wind gusts (red; m s\(^{-1}\)), precipitation (green; mm), and wind barbs denoting horizontal flow (thinned for clarity). Adapted from Ruppert and Bosart (2014).
Figure 4. Overview of the large-amplitude MGW event of 4 January 1994. (a) Satellite infrared brightness temperature at 1418 UTC (GOES-7, band 8; 11.2 µm). Dashed (solid) lines denote inferred MGW-induced desiccation (enhancement) of cloud through sinking motion (lifting). (b, d) Isochrones of primary MGW surface trough in (b) observations and (d) a model simulation. Area outlined with dotted line in (d) denotes accumulated precipitation ≥ 20 mm from 0000–1800 UTC. (c, e) Sea level pressure (hPa), surface winds, and MGW troughs (dashed lines) with inferred frontal positions at 1200 UTC in (c) observations with manual analysis and (e) model simulation. Panels (b, c) adapted from Bosart et al. (1998) and (d, e) adapted from Zhang et al. (2003a). Wind barbs are in standard convention here and in subsequent figures.
Figure 5. Conceptual model of the synoptic-scale environment favoring MGW genesis (dark-shaded region) and subsequent MGW activity (light-shaded region) in relation to upper-level flow imbalance, based on Uccellini and Koch (1987) and Zhang et al. (2001). Shown are scenes both prior to (left) and during (right) the time of most active MGW genesis in relation to the occlusion process of a baroclinic wave. (upper) 300-hPa geopotential height field (gray curves and “L”) and jet streaks (arrows; black for geostrophic and green for total), with the inflection axis (dotted line), and surface cyclone (black “L”) and fronts. (lower) Cross sections corresponding to the A–B line in upper-panels, with total vertical–horizontal flow (green arrows), isentropes (gray curves), and the dynamic tropopause (“DT”; black curves). In the lower-right panel, MGWs are excited with phase velocity $c$ (black arrow) in the region of the folding tropopause, with an upward component above genesis level and downward component below.
Figure 6. The first numerical simulation of inertia–gravity wave (IGW) genesis in connection with the dry baroclinic life cycle, by O'Sullivan and Dunkerton (1995). The general behaviors of (a) large-scale patterns (geopotential height and wind at 503 hPa) and (b) IGW signals (horizontal divergence at 130 hPa; $\Delta = 10^{-5}$ s$^{-1}$; thick contours for positive values). Adapted from Figs. 2 and 3 of O'Sullivan and Dunkerton (1995).
Figure 7. MGW emission in Zhang’s (2004) dry idealized baroclinic jet–front system experiment, showing the layered upper-level flow pattern at 120 h. Depicted are 13-km pressure (blue contours; Δ = 2 hPa) and horizontal flow (maximum of 25 m s$^{-1}$), 8-km jet streaks (shaded above 40 m s$^{-1}$, Δ = 5 m s$^{-1}$), and MGW signals via 13-km horizontal divergence (red lines, solid for positive; Δ = 2×10$^{-6}$ s$^{-1}$). The green rectangle highlights a set of pronounced MGW packets that appear in the jet streak exit region. Multiple distinct MGW modes are highlighted in red text (i.e., WP1–WP5), which have been the subject of subsequent articles (Plougonven and Snyder 2005, 2007; Wang and Zhang 2007; Lin and Zhang 2008; Wei and Zhang 2014, 2015). Adapted from Zhang (2004).
Figure 8. IGWs emitted by idealized vortex-dipole jet streaks situated (a) at the surface and (b) aloft. (a; left) Horizontal plan view of vertical velocity (shaded; red is positive) at 125 m and potential temperature (thin contours) at 62.5 m. (a; right) Cross section according to the thick line in (a, left) with vertical velocity and section-parallel horizontal velocity (contours). Adapted from Snyder et al. (2007). (b) Cross section of divergence (shaded/contoured; red is positive) and horizontal velocity (black; m s\(^{-1}\)). Adapted from Wang et al. (2009).
Figure 9. Sensitivity MGW emission and dynamics to convective instability in the idealized baroclinic jet–front system, showing six convection-allowing idealized baroclinic wave simulations, which differ only in initial moisture content: including (a) a dry experiment, (b) a weak moist experiment, (c, d) two moderate moist experiments, and (e, f) two strong moist experiments. Snapshots are at selected times when the tests are at similar stages of the baroclinic lifecycle. MGW packets are depicted via 12-km horizontal divergence (blue lines, positive; red lines, negative; Δ = 2×10^{-6} s^{-1}; range within ±1.2×10^{-5} s^{-1}; zero-value omitted). The other contours denote 1-km temperature (yellow; Δ = 5 K), 8-km horizontal wind (black; contour levels at 40, 45, 50, 55 m s^{-1}), and the dynamic tropopause (turquoise; 1.5 PVU surface at 7 km). Five gravity wave modes (i.e., WP1–WP5) are annotated in (a) and (b). WP6, a novel mode only present with moisture included, is identified most clearly at the earlier stage of the corresponding moist baroclinic wave life cycle (not annotated here). Adapted from Wei and Zhang (2014, 2015).
Figure 10. Schematic depiction of MGW–convection interaction according to two hypotheses. (a) Ducted wave–CISK, with an MGW wave train coupled to deep convection. Encircled arrows denote latent heating- and cooling-induced vertical motion (solid for positive, upward). Other features are as in Fig. 2. (b) A singular MGW wave of depression, akin to a squall line wake low, with a pronounced stratiform anvil precipitation region (dash-dotted line denotes 0°C isotherm). Surface trough is amplified by strong sinking motion at the trailing edge of the anvil, which is promoted by transient melting and evaporative cooling in combination with drying due to a descending rear-inflow jet and synoptic-scale dry conveyor belt behind a cold front aloft.
Figure 11. Evolution of the domain-integrated difference total energy ($E_\Delta$, m$^2$ s$^{-2}$) at three different characteristics scales (S: smaller scale $L < 200$ km; M: intermediate scale $200 < L < 1000$ km; and L: larger scale $L > 1000$ km) and the three-stage framework proposed by Zhang et al. (2007). Dotted lines denote simulations that exclude latent heating. Adapted from Fig. 7 of Zhang et al (2007).
Figure SB1. Photo of Fuqing Zhang, courtesy of Robert Houze.