The Genesis of African Easterly Waves by Upstream Development

MICHAEL DIAZ AND ANANTHA AIYYER
North Carolina State University, Raleigh, North Carolina

(Manuscript received 13 December 2012, in final form 27 June 2013)

ABSTRACT

A genesis mechanism for African easterly waves (AEWs) is proposed. In the same manner that new troughs and ridges in the midlatitudes form downstream of existing ones through a mechanism known as downstream development, it is proposed that new AEWs can be generated upstream of existing AEWs. A local eddy kinetic energy budget of the AEW that ultimately became Hurricane Alberto (2000) demonstrates that upstream development explains its genesis more convincingly than previous theories of AEW genesis. The energetics and ageostrophic secondary circulation of a composite AEW are consistent with a new AEW forming as a result of this mechanism. Some strengths and weaknesses of upstream development as a paradigm for AEW genesis are discussed with respect to other potential mechanisms.

1. Introduction

a. African easterly wave genesis

Despite continued advances in our understanding of African easterly waves (AEWs), a satisfactory theory for their genesis has remained elusive. Early studies of AEWs place their origin as far east as central and eastern Africa. Carlson (1969) speculated that they originate from squall lines initiated over elevated terrain in eastern Africa, and Frank (1970) suggested that they may originate from the interaction of the upper-tropospheric easterly jet with the mountains of East Africa. Beginning in the 1970s, the consensus reached was that AEWs form because of an instability of the African easterly jet (AEJ). Burpee (1972) confirmed that the AEJ is associated with a reversal in the meridional gradient of potential vorticity (PV) and therefore satisfies the Charney–Stern necessary condition for hydrodynamic instability (Charney and Stern 1962). Many numerical modeling studies have since demonstrated that the structure of the most unstable normal mode of idealized representations of the AEJ resembles that of AEWs (e.g., Thorncroft and Hoskins 1994a, and references therein). It is also well documented that AEWs follow two distinct tracks: one north of the AEJ and one south of the AEJ (e.g., Reed et al. 1977). To account for this observation, it has been suggested that northern-track AEWs result primarily from baroclinic instability and that southern-track AEWs result from barotropic instability and interaction with deep convection (e.g., C.-B. Chang 1993; Pytharoulis and Thorncroft 1999; Chen 2006).

Recently, explanations for AEW genesis have veered away from normal-mode instability theory. One alternative that has gained wide support is that AEW genesis is better explained by the interaction of moist convection with the AEJ. Based on an analysis of a numerical simulation of AEWs, Hsieh and Cook (2005) conclude that their generation relates more closely to condensational heating within the intertropical convergence zone (ITCZ) than to hydrodynamic instability of the AEJ. In a follow-up study, they demonstrate that this heating can quickly produce a strong PV gradient reversal that in turn supports the growth of AEWs, as convective overturning circulations induce barotropic and baroclinic instabilities (Hsieh and Cook 2008). By contrast, other studies have emphasized the role of condensational heating in producing finite-amplitude perturbations that can themselves grow into AEWs. For example, in a case study of the AEW that ultimately led to Hurricane Alberto (2000), Berry and Thorncroft (2005) hypothesize that it originated from a mesoscale convective system (MCS) initiated over the high terrain of Darfur in western Sudan. They reason that the large disturbance to the wind field induced by the MCS perturbs the unstable basic state and leads to the development of an AEW. However, in a case study of the same AEW, Lin et al.
(2005) come to a slightly different conclusion. They suggest that it originated from an orographically forced wave in the lee of the Ethiopian highlands that subsequently merged with a mesoscale convective vortex (MCV). Nevertheless, despite differences in the exact mechanism, each of these studies agrees that convection is important in the genesis of AEWs. This idea has been supported in many subsequent observational (e.g., Mekonnen et al. 2006; Kiladis et al. 2006; Berry et al. 2007; Leroux et al. 2010; Ventrice and Thorncroft 2013) and modeling studies (e.g., Hsieh and Cook 2007; Thorncroft et al. 2008; Cornforth et al. 2009).

As more and more studies have emphasized the importance of convection in initiating AEWs, recent work has questioned whether normal-mode instability theory is even an appropriate framework to explain AEW genesis. Hall et al. (2006) demonstrate in an idealized numerical model that the AEJ is rendered neutral or slightly stable with relatively modest surface damping. They argue that the slow growth rate of the most unstable normal mode of the AEJ cannot account for the observed amplitudes of AEWs, because individual AEWs do not spend enough time within the unstable region of the AEJ. They even suggest that AEWs may be a manifestation of the slowest decaying normal mode. Based on their results, they hypothesize that AEWs require large finite-amplitude precursors for their genesis. In a follow-up study using the same modeling framework, Thorncroft et al. (2008) demonstrate that convective precursors can initiate AEWs on a neutrally stable AEJ. They prescribe circular regions of diabatic heating and cooling to mimic convection and examine the ensuing downstream response. They find that structures closely resembling AEWs emerge within a few days after the diabatic forcing is applied. They coin this genesis paradigm the “triggering hypothesis” and identify a triggering region near the Darfur mountains that is most sensitive to convective triggers. To further refine this hypothesis, Leroux and Hall (2009) find that the subsequent response to this heating is very sensitive to the state of the AEJ, with some configurations leading to a strong response and others to a weaker one. The list of “triggers” has even been extended to include disturbances along the midlatitude storm track (Leroux et al. 2011). Regardless of the triggering mechanism, the general theme of these studies is that, although mature AEWs grow by extracting energy from the AEJ through mixed baroclinic and barotropic energy conversion, normal mode instability does not explain their initiation; it merely determines their scale and structure.

However, the triggering hypothesis as presented by current research has several shortcomings. To our knowledge, no observational study has conclusively demonstrated that an initial convective precursor directly triggered an AEW where no AEW had existed before. Convection that appears to precede an AEW could just as easily be the result of a preexisting AEW that went undetected. Additionally, with widespread convection common almost every day over tropical Africa, convection preceding an AEW may often be merely coincidental. This cause-and-effect relationship is particularly difficult to separate because of the lack of weather observations in eastern Africa and because of the difficulty of identifying weaker AEWs. It is also unclear how deep convection, whose period is strongly diurnal, could lead to disturbances with a 3–5-day period. If AEWs are primarily a triggered phenomenon and the triggers are independent of the AEWs, then it is not obvious why they do not exhibit the 1-day periodicity of convection or the aperiodicity of tropical cyclones, which clearly require the combination of a finite-amplitude precursor and a favorable ambient environment to support their continued growth (e.g., Gray 1968).

Additionally, although Thorncroft et al. (2008) clearly demonstrate that upstream diabatic forcing can trigger AEWs in an idealized model, the initial diabatically forced vortex that forms in their simulation is unrealistically large compared with observed vortices generated by convection alone. The 1080-km-wide region of prescribed diabatic forcing in their experiment produces a vortex with a diameter of maximum winds of 1600 km within 1 day [see Fig. 4 in Thorncroft et al. (2008)]. Since the diabatic forcing is meant to represent an MCS, it follows that this vortex is an MCV. However, it is significantly larger than observed vortices, which tend to be about 150–300 km in diameter (Bartels and Maddox 1991). Assuming the diameter of maximum wind represents a half wavelength, the modeled MCV would have a full wavelength of 3200 km. This wavelength is similar to that of the most unstable normal mode of the AEJ in their model setup [3500 km in Hall et al. (2006)]. Thus, this experiment does not separate the scale of MCVs from that of AEWs. For a true cause-and-effect relationship, it must be established how an initial MCV can reach the scale of an AEW within such a short time frame before an AEW is present. Furthermore, despite the large perturbation, the initial 650-hPa AEW trough that is directly attributable to the circulation of the MCV is very weak (about 1 m s⁻¹ at jet level). By contrast, the simulation produces many subsequent AEW troughs that are much stronger than the first, despite the fact that their circulations cannot be directly attributed to MCVs. Therefore, most AEWs in this simulation did not directly originate from a finite-amplitude precursor. This behavior is inconsistent with the triggering hypothesis in
its most literal sense, that is, that each AEW requires a unique convective trigger. Alternatively, one could assume the viewpoint that the role of the convective trigger is to force the most unstable (or least stable) normal mode. This hypothesis, which is also clearly stated in Thorncroft et al. (2008), is more consistent with their results. However, if this is the correct interpretation, then it would be difficult observationally to trace an individual AEW back to its convective precursor, which may have preceded it by many wave periods.

The continued generation of AEWs in the simulation of Thorncroft et al. (2008) after the initial perturbation is one of the more intriguing aspects of their experiment. What causes these additional AEWs to become increasingly strong despite the basic state being stable to small perturbations? With a neutrally stable basic state, their experimental design would seem to rule out normal mode instability. However, except for the initial AEW, it also rules out upstream convective triggers. And, with no topography, it rules out orographic forcing. This result motivates us to explore an alternative genesis mechanism for AEWs. Recognizing that they exhibit the dispersion characteristics of Rossby waves (Diaz and Aiyyer 2013), it is plausible to expect them to form because of a mechanism analogous to downstream development of Rossby waves on midlatitude westerly jets.

b. Upstream and downstream development

The concept of downstream development has had a long history. Early theoretical studies attributed it to energy dispersion of barotropic Rossby waves (Rossby 1949; Yeh 1949). These studies found that disturbance energy can propagate downstream much faster than the disturbance itself because of the rapid geostrophic adjustments between the mass and velocity fields. They also showed that this energy travels at the Rossby wave group velocity. Hovmöller (1949) clearly confirmed the prevalence of downstream energy dispersion in the real atmosphere in his now well-known trough–ridge diagram.

The advent of numerical modeling allowed for the study of energy dispersion and downstream development in a more controlled setting. Simmons and Hoskins (1979) examined energy dispersion for unstable baroclinic waves as an initial-value problem in a numerical model. They found that new disturbances developed both upstream and downstream of the initial disturbance and that the growth rate of these new disturbances is substantially faster than that of the most unstable normal mode of their model’s basic state. Their growth is not linked to a single normal mode, but rather to the constructive and destructive interference of the large spectrum of normal modes required to localize the initial disturbance. From their results, they suggested that the main trigger for baroclinic instability may not be small, random perturbations, but rather energy dispersion from neighboring systems. Although downstream development clearly dominates the evolution of large-scale Rossby waves on midlatitude westerly jets, Thorncroft and Hoskins (1990) proposed that upstream development may explain the development of small-scale waves that develop along frontal boundaries. They attribute these waves to neutral Rossby waves that disperse along a surface baroclinic zone. Despite being neutral, these waves exhibit large growth rates. However, very few if any observational studies have focused on the upstream development of baroclinic Rossby waves.

To quantify downstream development in a case study of cyclogenesis, Orlanski and Katsfey (1991) derived a local eddy kinetic energy (EKE) budget that explicitly contains a term to represent energy dispersion. In this energy budget, they partitioned the pressure work term into a baroclinic conversion term and a geopotential flux convergence term. Downstream development through energy dispersion is represented by the convergence of geopotential fluxes. In a case study of midlatitude cyclogenesis, they found that the disturbance grew initially by baroclinic conversion but ceased growing once energy exported by geopotential fluxes exceeded baroclinic growth. This energy was then used to grow a new disturbance downstream. Building on observational studies (e.g., E. K. M. Chang 1993) and idealized modeling (e.g., Orlanski and Chang 1993), Orlanski and Sheldon (1995) provide a thorough overview of the use of a local energetics budget and how it relates to energy dispersion and downstream development.

c. Objectives

The goal of the present study is to propose upstream development through energy dispersion as a genesis mechanism for AEWs. The results of Diaz and Aiyyer (2013) suggest that this idea has merit. Using the same approach as Orlanski and Katzfey (1991) applied to composite AEWs, it was found that, in their growing phase, EKE growth due to convergence of geopotential fluxes exceeds that of baroclinic and barotropic growth, whereas in their decaying phase, EKE decay due to divergence of geopotential fluxes exceeds baroclinic and barotropic growth. For AEWs, the energy dispersion is eastward and thus new disturbances form upstream of existing ones. Westward advection by the AEJ counteracts the relatively fast eastward energy dispersion and leads to an approximately 3 m s⁻¹ upstream (eastward) group velocity (Diaz and Aiyyer 2013). One limitation of the study of Diaz and Aiyyer (2013) is that, because it is based on a composite of AEWs in many different stages of life, it cannot definitively conclude that energy dispersion is a genesis mechanism. In this study, we will
attempt to extend the results of Diaz and Aiyyer (2013) to a case of AEW genesis.

This study is organized as follows: Section 2 describes the local EKE budget; section 3 presents a case study of the AEW that ultimately led to Hurricane Alberto (2000); section 4 generalizes some of the results of section 3 using a composite AEW; section 5 outlines a conceptual model of upstream development, places it within the context of existing genesis theories, and discusses which aspects of AEW genesis it can and cannot explain; and section 6 recaps the major findings of this study.

2. Local EKE budget

To study the upstream development of AEWs, we adopt the local EKE budget proposed by Orlanski and Katzfey (1991). This approach offers a quantitative tool to compare the relative importance of energy dispersion to baroclinic and barotropic growth for individual AEWs. Here we provide a brief overview of the equations and their application. For a more complete discussion, refer to Orlanski and Katzfey (1991) and Orlanski and Sheldon (1995).

The equation for the time tendency of EKE in pressure coordinates as derived by Orlanski and Katzfey (1991) is as follows:

\[
\frac{\partial \left( K_e \right)}{\partial t} + \mathbf{V}_m \cdot \nabla K_e + \mathbf{v} \cdot \nabla_3 K_e \\
= - \left( \mathbf{v} \cdot \nabla \phi \right) - \left[ \mathbf{v} \cdot \mathbf{V}_m \cdot \mathbf{V}_m \right] + \left[ \mathbf{v} \cdot \left( \mathbf{V}_e \cdot \mathbf{V}_e \right) \right] - \text{diss}_e + \mathbf{v} \cdot \mathbf{F}_o ,
\]

(1)

where \( K_e \) is the EKE, \( \mathbf{V}_m \) is the time-averaged velocity, \( \mathbf{v} \) is the perturbation velocity, \( \phi \) is the perturbation geopotential, diss, refers to dissipative forcing, and \( \mathbf{F}_o \) denotes the forcing that maintains the time-averaged circulation. The terms on the left-hand side of Eq. (1) are the local EKE tendency, advection of EKE by the mean wind, and advection of EKE by the eddies, respectively. The first term on the rhs is work done by the pressure field. The second and third terms are Reynolds stress terms. They denote the exchange of kinetic energy between the mean flow and the eddies and are often referred to as barotropic conversion. The fourth term is dissipation of EKE by the eddies. The last term is the impact of the steady-state forcing on the eddies and is generally found to be small.

Orlanski and Katzfey (1991) showed that the pressure work term is useful in diagnosing downstream development in midlatitude baroclinic waves. This term, when vertically integrated, is approximated as follows:

\[- \mathbf{v} \cdot \nabla \phi = \nabla \cdot (\mathbf{v} \phi) - \omega \alpha ,\]

(2)

where \( \omega \) is the vertical velocity in pressure coordinates and \( \alpha \) is the specific density. The second term on the rhs of Eq. (2) is the baroclinic term, which describes the conversion of eddy potential to eddy kinetic energy. The first term on the rhs of Eq. (2) is geopotential flux convergence. It is central to our discussion because of its relationship with energy dispersion and group velocity. Over a sufficiently large domain, it averages to zero and thus makes no contribution to the global energetics; it merely redistributes EKE from one place to another.

The term \( \mathbf{v} \phi \) represents an energy flux. To determine the direction of this flux, the large nondivergent part associated with geostrophic balance can be removed to yield the following:

\[\mathbf{v} \phi = \left( \mathbf{v} - \frac{\mathbf{k}}{f_0} \times \nabla \phi \right) \phi ,\]

(3)

where \( f_0 \) is the Coriolis parameter at a reference latitude. This term is referred to as the ageostrophic geopotential flux (AGF). For small-amplitude quasigeostrophic waves, it represents an energy flux that when averaged over a full wavelength points in the direction of the group velocity.

It is important to make a distinction between a local EKE budget and a volume-averaged EKE budget, such as that used by previous studies of the energetics of AEWs (e.g., Norquist et al. 1977; Hsieh and Cook 2008; Cornforth et al. 2009). With a volume-averaged EKE budget, one cannot unambiguously distinguish between EKE growth resulting from the intensification of existing AEWs and that resulting from the genesis of new AEWs. Additionally, whereas a volume-averaged EKE budget diagnoses energy exchanges only between AEWs and the mean state, the local EKE budget also diagnoses EKE exchanges between individual AEWs. Thus, with a local EKE budget, we are in a better position to understand cause and effect with regard to AEW genesis.

3. Case study

To demonstrate that upstream development through energy dispersion can effectively generate AEWs, we present a case study of the genesis and early life of the AEW that ultimately led to Hurricane Alberto (2000). We chose this particular AEW because its genesis has been attributed to a convective precursor by multiple case studies (e.g., Hill and Lin 2003; Berry and Thorncroft 2005; Lin et al. 2005; Ventrice and Thorncroft 2013). As mentioned in the introduction, discrepancy remains
regarding exactly what generated this AEW. Berry and Thorncroft (2005) hypothesize that an MCS perturbing the low-level temperature gradient initiated it on 30 July over the Darfur highlands. On the other hand, Lin et al. (2005) hypothesize that the interaction between an MCV and an orographically generated vortex in the lee of the Ethiopian highlands led to its genesis on 28 July. By contrast, a detailed quantitative analysis using the local-energetics approach suggests that its genesis is better explained by energy dispersion from a preceding AEW on 26 July. As such, the two MCSs previously cited as the cause of this AEW were more likely an effect of it.

a. Data and methods

Our analysis uses the gridded Interim European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-Interim) (Dee et al. 2011). These data have a grid spacing of $1.5^\circ \times 1.5^\circ$ and are available every 6 h. Both Berry and Thorncroft (2005) and Lin et al. (2005) base much of their studies on the ECMWF operational analysis, because the ERA-Interim did not become available until 2006. With its more sophisticated data assimilation system and better access to data, the ERA-Interim offers substantial improvement over the operational ECMWF analysis. Thus, our earlier detection of this AEW in comparison with previous studies may at least partly result from the improved dataset.

The mean state for calculating Eq. (1) is constructed using the July–August 2000 time average for each meteorological variable. For the perturbation quantities, we apply a 2-day low-pass filter to remove the strong diurnal cycle and semidiurnal tides. The budget terms are vertically averaged from near the surface (i.e., the lowest available pressure level above the ground) to 200 hPa. With most of the disturbance amplitude concentrated below 500 hPa, the interpretation of our results is insensitive to moving the upper boundary of averaging between 500 and 100 hPa. The reference latitude used to calculate the AGF [Eq. (3)] is $15^\circ$N.

One significant problem that arises when attempting to balance the local EKE budget is that we cannot assume that the dissipation by the eddies is small [fourth term on rhs of Eq. (1)]. Although this assumption was valid for the case study of Orlanski and Katzfey (1991), which featured a disturbance over the ocean with most of its EKE at upper levels, our case study features a shallow disturbance over land embedded within a deep, convective boundary layer. As we cannot calculate eddy dissipation directly from the reanalysis, it contributes to a large negative residual that is of the same order of magnitude as the other terms in the budget. Although a large residual is not optimal, its behavior is consistent with friction; it is almost always negative and it correlates well with regions of strong surface wind. To help verify that the budget is calculated correctly and to find the approximate order of magnitude of frictional EKE dissipation, we tested the calculations in a numerical simulation of idealized AEWs with a simple boundary layer parameterization in which all of the EKE budget terms could be accounted for (not shown). The results suggest that frictional EKE dissipation does become the same order of magnitude as the EKE growth terms in regions where surface winds are strong. An additional source of error in balancing Eq. (1) is the nonphysical process of assimilating model output with raw observational data. Finally, some of the residual can be attributed to calculating it using coarse (6 hourly) temporal resolution. This factor would become especially problematic when the EKE field is rapidly evolving or where the advection terms become large.

b. Overview

Energy dispersion tends to organize waves into coherent wave packets. To reveal the wave-packet nature of the AEWs in the time period surrounding our case study, we show a Hovmöller plot of the square of the 2–6-day filtered 850-hPa meridional wind ($m^2 s^{-2}$) averaged from 5° to 18°N for a period in 2000. Lines are drawn along the group velocity of several prominent wave packets. The dashed line marks a wave packet north of the 5°–18°N latitude band that is better shown in Fig. 3. “A” marks the genesis location of the pre-Alberto AEW, and “B” marks the genesis location of the preceding AEW.
respectively, and lines are drawn along the group velocity of several prominent wave packets.

The time period surrounding our case study features multiple upstream-propagating wave packets that have an apparent life span of about 10–20 days (Fig. 1). The genesis of AEW A and AEW B is linked to a wave packet that first appears on 21 July near 25°E. The first AEW in this wave packet is in turn linked to a preceding wave packet first visible on 16 July near 30°E. This continuity suggests that wave energy can propagate upstream in one coherent wave packet and then initiate a second wave packet. During its life span, AEW A also becomes involved in a wave packet near 15°E and near 15°W, which is near the West African coast. Both periods were associated with increases in its 850-hPa meridional wind signature. These observations suggest that upstream energy dispersion may be responsible for some of the intensity fluctuations of AEWs.

Figure 2 shows the track of AEW A in relationship with the 925-hPa potential temperature and 700-hPa potential vorticity averaged for July–August 2000. The track of the northern vortex is plotted with dots and the track of the southern vortex with crosses. One feature essential for AEW growth is the meridional PV gradient reversal (Fig. 2). This reversal is caused by the combination of PV destruction by dry convection over the Sahara and PV production by deep, moist convection in the ITCZ near 10°N (Thorncroft and Blackburn 1999; Dickinson and Molinari 2000). According to hydrodynamic instability theory (e.g., Hoskins et al. 1985), disturbances can amplify through interactions between the northward-directed surface potential temperature gradient and the southward-directed midlevel PV gradient (i.e., baroclinic instability) or through interactions between the positive and negative meridional PV gradients at midlevels (i.e., barotropic instability). The former interaction explains the existence of northern-track AEWs, while the latter interaction explains the existence of southern-track AEWs (e.g., Pytharoulis and Thorncroft 1999). Consistent with this explanation, the northern vortex for AEW A remains within the region of reversed temperature gradient, and the southern vortex tracks south of or directly along the zonally elongated PV maximum. In agreement with previous studies (e.g., Berry and Thorncroft 2005), the northern and southern vortices traverse Africa as one coherent feature. According to the Rossby wave dispersion relationship (Rossby 1949), eastward energy dispersion is possible in either the positive meridional PV gradient south of 10°N or the positive meridional temperature gradient north of 10°N. Note that surface potential temperature gradients are equivalent to surface PV gradients (Bretherton 1966). The origin of the southern vortex within the strongest positive PV gradient is at least consistent with a genesis due to energy dispersion.

Although we consider 26 July to be the genesis date of AEW A, it is difficult to unambiguously specify an exact genesis date and location. A careful analysis suggests some continuity between AEW A and a weak perturbation over Saudi Arabia beginning on 23 July. Because the averaging latitude band of Fig. 1 is too far south and west to reveal this relationship, we show a zoomed-in Hovmöller plot of the same field as Fig. 1 averaged from 22° to 28°N (Fig. 3). As suggested by the line drawn along the group velocity, an upstream-propagating wave packet appears on 20 July and lasts until 30 July. For reference, the line drawn in Fig. 3 is also drawn in Fig. 1 as a dashed line, even though the corresponding waves do not explicitly appear in Fig. 1. Though it is clear that weak wave activity occurs upstream of AEW A and B, in order to focus our case study, we do not include an EKE budget analysis of this wave activity. Our reasoning is
that the wave activity before 26 July appears to also result from an upstream-propagating wave packet and thus represents essentially the same genesis mechanism. Also, these perturbations weakened significantly as they approached the Red Sea on 25 July (Fig. 3). Thus, we consider the events that unfolded on 26 July to be the primary instigator of the genesis of AEW A. Nevertheless, we cannot exclude the possibility that weak perturbations originating in Saudi Arabia had a non-negligible contribution to the genesis of AEW A. It is worth noting that the location of the wave packet over Saudi Arabia corresponds with a climatological maximum in 2.5–5-day meridional wind variance (Albignat and Reed 1980).

c. Energy budget

To quantify AEW A’s growth due to various energetics processes, we present a local EKE budget of its early life from 26 to 30 July (Figs. 4–8). Figures 4a–8a show geopotential flux convergence [first term on rhs of Eq. (2)] and ageostrophic geopotential flux vectors on pressure surfaces [Eq. (3)]. Figures 4b–8b show baroclinic conversion [second term on rhs of Eq. (2)] and 850-hPa perturbation wind. Figures 4c–8c show the barotropic term [second and third terms on rhs of Eq. (1)] and 650-hPa perturbation wind. Figures 4d–8d show the residual and surface wind perturbation. Figures 4e–8e show the advective terms [second and third terms on lhs of Eq. (1)] and full 650-hPa wind. Figures 4f–8f show the nonadvective tendencies [sum of first, second, and third terms on rhs of Eq. (1)] and full 850-hPa wind. The winds shown in each plot are meant to facilitate understanding of the physical processes underlying the budget terms. The 650-hPa winds correspond with the typical height of the AEJ and hence the level at which the eddies extract EKE from the zonal shear of the AEJ. The 850-hPa winds are located within the low-level baroclinic zone and hence are important for baroclinic processes. The surface winds are shown to highlight the general relationship between the residual and regions where the surface wind stress is large. For convenience, the EKE maxima linked to AEW A and AEW B are labeled in Figs. 4c–8c with “A” and “B,” respectively.

The genesis of AEW A can be traced to its interaction with AEW B on 26 July. At 0000 UTC 26 July, AEW B is growing through both barotropic and baroclinic conversion (Figs. 4b,c). Based on the southwest-to-northeast tilt of the 650-hPa trough south of the mean position of the AEJ, we can deduce that AEW B is extracting EKE from the horizontal shear of the AEJ (Fig. 4c). The strongest baroclinic conversion is focused on the surface temperature gradient along 17°N (Fig. 4b). This baroclinically generated EKE is being exported upstream by the AGF, which is convergent upon a region near the Red Sea (Fig. 4a). This upstream transfer of EKE will begin the process leading to the genesis of AEW A. On 27 July, EKE rapidly increases within the northerly flow of AEW A (Fig. 5). From Figs. 5a and 5b, we see that EKE generated by baroclinic conversion within the southerly flow of AEW B is being transferred upstream by the AGF into the northerly flow of AEW A. Because the convergence of AGF is the dominant growth term, we attribute the initial growth of AEW A to energy dispersion emanating from AEW B. Although barotropic and baroclinic conversion are insignificant at this time for AEW A, the large perturbation to the meridional wind extending from 5° to 20°N is disturbing the regions of reversed meridional PV and temperature where the necessary criterion for hydrodynamic instability is met (Fig. 2). If these gradient reversals can support its continued growth, then we should expect this disturbance to soon amplify by extracting EKE from the basic-state kinetic and potential energy. However, as demonstrated by the large residual, not all of the EKE conversion is being realized (Fig. 5d). In fact, the residual is of the same order of magnitude as the total nonadvective EKE growth (Fig. 5f). Because the residual is almost exclusively negative and corresponds to regions where the surface wind is strong within the deep Saharan boundary layer, it is consistent with EKE spindown through friction. Indeed, if frictional EKE dissipation is as large as Fig. 5f suggests, then it would strongly stabilize the AEJ to perturbations growing from small amplitudes. This observation agrees with

In contrast with AEW A, EKE growth for AEW B diminishes on 27 July. Although it has the characteristic boomerang shape that one would expect from an eddy extracting EKE from the zonal shear of the AEJ (Fig. 5c), its overall growth is near zero as weak growth through baroclinic and barotropic conversion is compensated by decay through the divergence of AGF and friction (Figs. 5d,f). Thus, AEW A is growing at the expense of AEW B.

On 28 July, AEW A begins to grow through a combination of baroclinic and barotropic conversion (Figs. 6b,c). The strongest baroclinic conversion is concentrated on the northern vortex, where northerly flow has led to a substantial warm perturbation (Fig. 6b). The strong cyclonic surface winds and frictional EKE dissipation suggest that Ekman pumping is forcing localized convergence and ascent within this warm anomaly (Fig. 6d). On the northern edge of the EKE maximum, the barotropic term has become large with $\nu^2 \partial V/\partial y$ contributing to nearly all of the barotropic growth (Fig. 6c). As
evident by the 850-hPa wind vectors in Fig. 6f, rather than extracting EKE from the zonally oriented AEJ (i.e., $u'u'\partial U/\partial y$). AEW A is growing along the confluence line between the monsoon southerlies and the dry northerlies. In contrast with the previous day, the convergence of AGF emanating from AEW B has become the smallest EKE growth term (Fig. 6a). However, to continue the upstream flux of EKE, the AGF from the northerlies of AEW A are beginning to converge upstream. Their convergence will accelerate southerly flow on the eastern side of AEW A and complete its cyclonic circulation. Once again, friction consumes much of the EKE growth for AEW A (Fig. 6d). Thus, despite robust EKE growth through baroclinic and barotropic conversion, AEW A is actually nearing a steady state where EKE growth is balanced by EKE dissipation. This balance may place a strong constraint on the intensity of a surface-intensified vortex such as a northern-track AEW, because the production of eddy available potential energy through advection [Eq. (4.2) in Orlanski and Katzfey (1991)] and dissipation of EKE through friction are both proportional to the low-level wind speed. Nevertheless, one must be cautious ascribing physical processes to the residual, especially since reanalysis datasets are not necessarily dynamically consistent. It should be noted that the largest residual is directly over Khartoum, Sudan, where an upper-air sounding is assimilated into the reanalysis. At least some of the residual can be attributed to the data assimilation system “correcting” for this observation.

Meanwhile, as AEW A has been rapidly strengthening, AEW B has been slowly weakening (Fig. 6d). The life cycle for AEW B so far bears some similarity with that of the extratropical cyclone analyzed by Orlanski.
and Katzfey (1991); its growth through baroclinic and barotropic conversion terminates once AGF begin to export EKE upstream as fast as it is generated. However, in contrast with the upper-level cyclone analyzed by Orlanski and Katzfey (1991), which decayed primarily through divergent energy fluxes, frictional EKE dissipation plays an equally large role in the weakening of AEW B.

After reaching a peak intensity near 1200 UTC July 28, AEW A begins to slowly weaken on July 29 as frictional EKE dissipation exceeds EKE growth (Fig. 7). The northern vortex and southern vortex make nearly equal contributions to baroclinic conversion (Fig. 7b). According to satellite imagery and 925-hPa temperature (not shown), the former is associated with temperature advection and the latter with moist convection. Much of this EKE is being transferred upstream by the AGF to the southerly flow of AEW A (Fig. 7a). Advection begins to transport AEW A more rapidly westward as it is picked up by the AEJ (Fig. 7e). In contrast, advection on the previous day was transporting it southward (Fig. 6e). Figure 1 shows this westward acceleration as an increase in zonal phase speed on 29 July. Although not shown in Fig. 7, it is important to note that AEW B restrengthens substantially on 29 July.

Beginning on 30 July, AEW A rapidly intensifies at the expense of AEW B (Fig. 8). Cold advection in the southerly flow of AEW B leads to a large region of vigorous baroclinic conversion (Fig. 8b). However, divergent AGF more than compensate for this EKE growth (Fig. 8a). These fluxes converge upstream, where they accelerate the northerly flow of AEW A. Figure 1 reveals this energy transfer as an upstream-propagating wave packet. This time frame coincides with the
initiation of a large MCS (Berry and Thorncroft 2005). Its contribution to EKE is evident by the concentrated region of baroclinic conversion centered at 8°N, 22°E. Although the reanalysis correctly places the convection, it appears to have been initiated slightly before the real event. Nevertheless, the enhanced energy dispersion event (i.e., the wave packet in Fig. 1) began in the reanalysis on 29 July and clearly preceded the convection. Therefore, it is plausible that strong dynamical forcing associated with energy dispersion from AEW B initiated this MCS rather than the MCS initiating AEW A. We will elaborate on this possibility in more detail in sections 3d and 3e. Additionally, with its 650-hPa trough tilting upshear, AEW A is finally extracting EKE from the zonal shear of the AEJ (Fig. 8c). Thus, barotropic conversion from the horizontal shear of the AEJ does not become important to the growth of AEW A until well after its inception.

d. Physical mechanism

The physical processes leading to energy dispersion and the genesis of AEW A can be understood with more transparency using the concept of an ageostrophic secondary circulation. Although one can achieve an analogous explanation using PV-advection arguments (e.g., Hoskins et al. 1985), the secondary circulation concept relates more directly to the terms in the local EKE budget. To reveal the lower branch of this secondary circulation, we plot divergence, geopotential, and ageostrophic flow at 900 hPa (Figs. 9a,c). To reveal its rising and sinking branches and the dynamical forcing maintaining it, we plot 800-hPa vertical velocity, 850-hPa temperature, and 850-hPa wind (Figs. 9b,d). All of the quantities except for temperature are perturbations from the mean state. The rising and sinking branches are labeled with an “R” and
an “S,” respectively, and the geopotential high and low with an “H” and an “L,” respectively. To demonstrate that the forcing that drives this circulation can be described using quasigeostrophic (QG) arguments, we plot 850-hPa Q vectors and their divergence. Following Kiladis et al. (2006), we substitute the geostrophic wind used in the standard Q-vector equation with the perturbation wind. For reference, Q-vector convergence implies QG forcing for ascent, Q-vector divergence implies QG forcing for descent, and the Q vector itself points in the general direction of the low-level ageostrophic flow. For brevity, we focus on two times: genesis at 0000 UTC 27 July and reintensification at 0600 UTC 30 July.

From Figs. 9a and 9b, we can deduce an ageostrophic secondary circulation directed from west to east across the geopotential ridge. Its sinking branch coincides with cold advection and low-level divergence, and its rising branch with warm advection and low-level convergence. In its lower branch, eastward-directed ageostrophic flow within a westerly current is consistent with gradient wind balance, which requires supergeostrophic flow within a ridge. The pattern of vertical motion and ageostrophic flow closely resembles that which can be attributed to QG forcing (Fig. 10a). Positive baroclinic conversion demonstrates that this secondary circulation converts potential energy to kinetic energy (Fig. 5b). However,
the resulting horizontal wind accelerations are asymmetric about the ridge. With its sinking branch centered on the region of QG forcing west of the geopotential ridge axis, the ageostrophic flow within the ridge is predominantly eastward (Fig. 9c). According to Eq. (1), this eastward flow down the gradient of geopotential generates EKE. It preferentially accelerates the northerly flow on the up-stream (eastern) side of the ridge rather than the southerly flow on the downstream (western) side. This asymmetry explains why Fig. 5a shows a strong upstream flux of energy directed away from the region of baroclinic conversion.

The energetics contributions of the rising and sinking branches also exhibit asymmetry. From the sign of the baroclinic term (Fig. 5b), we know that the sinking branch coincides with cold air. By contrast, the rising branch is not yet associated with a significant temperature perturbation, because the accelerating northerly flow has not had enough time to produce a warm anomaly by advection. Hence, at this time, baroclinic conversion is occurring only in the sinking branch of the circulation. This baroclinically generated EKE is being fluxed upstream to accelerate the northerlies of AEW A. As the warm advection in these northerlies leads to a warm perturbation, baroclinic conversion will begin in the rising branch (Fig. 6b). The rising branch will then induce another ageostrophic secondary circulation across the geopotential low of AEW A (Fig. 7a). In essence, energy is transferred upstream in a series of ageostrophic secondary circulations. These circulations link together successive AEWs and provide a mechanism whereby existing AEWs can extract EKE from the basic-state potential energy and transfer it upstream to initiate new AEWs. Although the preceding discussion focused on the evolution of the northerly and southerly flow within the EKE maxima, one can apply similar reasoning to the effect of convergence and divergence on vorticity stretching and compression within the rising and sinking branches of the secondary circulations (Simmons and Hoskins 1979). For example, low-level convergence and ascent east of the geopotential ridge stretches vorticity and induces surface cyclogenesis.

A second energy-dispersion event unfolds during AEW A’s reintensification phase on 30 July (Figs. 9c,d). On 29 July, AEW B rapidly intensified. Cold advection
within its southerly flow contributed to strong baroclinic conversion and intensified the ageostrophic secondary circulation, linking AEW B and AEW A through the geopotential high (Figs. 8, 9c,d, 10b). The eastward-directed ageostrophic flow down the gradient of geopotential intensifies the northerlies of AEW A. Because these events began just before the MCS on 30 July, the intensification of this secondary circulation may have been important in initiating the MCS that Berry and Thorncroft (2005) hypothesize triggered AEW A.

e. Observations

1) SURFACE

Because the ERA-Interim is based heavily on model output in data-sparse regions such as northeastern Africa, we will examine more direct observations to help validate the genesis time and location of AEW A. For this purpose, we plot time series of temperature, dew-point, and sea level pressure for three stations located near the site of surface cyclogenesis, namely, Abu Hamed, Atbara, and Khartoum, Sudan (Fig. 11). For reference, these stations are plotted in Fig. 12. To better observe the AEW signal, we subtract the July–August 2000 hourly mean from each variable to remove the diurnal cycle. All plotted values are anomalies from this mean. For missing data, we linearly interpolate from the surrounding times. We then apply a 1–2–1 smoother to remove the small-scale wiggles. Vertical lines denote the passage of troughs and ridges in sea level pressure.

The sea level pressure time series exhibits a 5.5-day period AEW at all three stations. The passage of the ridge axis near 0000 UTC 26 July and the trough axis near 1800 UTC 28 July closely aligns with the reanalysis. The sea level pressure perturbation at Khartoum reached $-3.4$ hPa at 1800 UTC 28 July. For comparison, 19 out of 33 of the AEWs observed by Carlson (1969) to cross the West African coast in 1968 had sea level pressure perturbations between $-2.0$ and $-4.0$ hPa. According to the EKE budget and the analysis of the secondary circulation, the falling pressures on 27–28 July resulted from in situ cyclogenesis rather than from an approaching disturbance. As the surface low intensified, northerly winds advected hot dry air from the north. This temperature advection is clearly evident at Abu Hamed and Khartoum. Consistent with a warm-core surface low, Khartoum reported a $+3.8$-K warm anomaly coincident with the lowest sea level pressure. The temperature anomalies are smaller at Abu Hamed because it is located north of the strongest temperature gradient and north of the track of AEW A. Strangely, the temperature at Atbara does not follow the expected pattern but rather drops on the afternoon of 27 July and reaches a minimum when the surface low passes directly overhead. There is no obvious meteorological explanation for this drop, since the station reported no clouds or precipitation at the time. It is possible, though highly speculative, that strong ascent forced by Ekman pumping directly in the center of the developing surface low altered the convective–radiative equilibrium in the boundary layer by deepening the mixing layer. Assuming surface heating makes the low-level lapse rates superadiabatic, deeper mixing would cool the surface temperature without a corresponding increase in surface hydrostatic pressure, because surface cooling...
would be balanced by warming aloft. Alternatively, the temperature drop may have resulted from other microscale processes or even measurement errors. In addition to the warm advection, the southward advection of dry air shows up as a sharp drop in dewpoints. This drop, which occurs later at southern sites and earlier at northern ones, suggests a southward-moving moisture front. In the strong southerly flow behind AEW A on 29 July, dewpoints rose sharply, with Abu Hamed reporting a nearly +11-K warm anomaly.

Previous observational studies suggest that events similar to AEW A are a common occurrence in northern Sudan. Albignat and Reed (1980) find evidence of an AEW genesis maximum in the same location where AEW A appears to form. Using radiosonde data for a 28-day period in 1974, they observe a maximum in the 2.5–5-day-period meridional wind at 850 hPa in northern Sudan [Fig. 3 in Albignat and Reed (1980)]. These radiosonde data reveal that this maximum corresponds with a region of cyclonic shear at 850 hPa and an enhanced southward flux of momentum at 700 hPa.

2) PRECIPITATION

To complete our case study, we will examine how AEW A interacted with moist convection using Tropical Rainfall Measuring Mission (TRMM) 3B42–estimated rainfall rates gridded at 0.25° × 0.25°. This dataset merges TRMM rainfall estimates with additional satellite observations and rain gauge data. For reference, surface pressure anomalies and total surface winds from the ERA-Interim are overlaid. Once again, we remove periods of less than 2 days from the wind and pressure field using a low-pass filter. The locations of the surface observations are plotted by letters (‘‘K’’ for Khartoum, ‘‘H’’ for Abu Hamed, and ‘‘A’’ for Atbara), and the southern vortex is marked with a cross.

At 0000 UTC 28 July, surface cyclogenesis is ongoing in a dry region in northeastern Sudan (17°N, 34°E) (Fig. 12a). Throughout the day, convection increases to the south and east of the surface low (Fig. 12b). Based on the variations in surface dewpoints (Fig. 11), it is plausible that AEW A modulates convection through its effect on moisture advection. With the southward-directed moisture gradient, the dry northerly flow suppresses convection, while the moist southerly flow enhances it. The small region of convection centered near 8°N, 28°E appears more directly associated with AEW A, because it was initiated near the center of the southern vortex (Fig. 12b). After nightfall, the convection is organized into two MCSs, one in the lee of the Ethiopian highlands and a second in association with a squall line that initially formed near the southern vortex but has since propagated southwestward (Fig. 12c). The former is the MCS Lin et al. (2005) hypothesize generated an MCV that subsequently merged with an orographically generated vortex to produce this AEW. By contrast, surface observations and reanalysis data strongly support AEW A existing prior to this convection and that its origin is better explained by energy dispersion. It is possible that the moist southerly flow behind AEW A provided a favorable environment for this MCS to propagate off the Ethiopian highlands. If events similar to AEW A are common, then they may...
cause the 2–6-day spectral peak in the convection observed in this region (Mekonnen et al. 2006).

By 0000 UTC 30 July, the surface low has weakened from its previous day’s intensity (Fig. 12d). However, as noted by Lin et al. (2005), it is visible in infrared satellite imagery as a swirl of low clouds. This time is approximately 6 h before an intense MCS erupts from within this swirl of clouds. Because this MCS was initiated near the diurnal convection minimum (about 0700 local time), it seems likely that it was triggered by strong dynamical forcing. From the preceding energetics analysis, we can deduce this forcing. As discussed in section 3d, the surface low of AEW A is located directly beneath the rising branch of an ageostrophic secondary circulation with QG forcing for ascent (Figs. 9c,d, 10b). Intensifying baroclinic conversion in the sinking branch located in the southerlies of AEW B strengthened the secondary circulation and provided strong dynamical forcing for ascent upstream. Convective initiation occurred within this region of ascent and low-level convergence. Thus, it appears that the same dynamical forcing that intensified AEW A also triggered the

![Fig. 12. TRMM 3B42–estimated rainfall rate (shaded, mm day$^{-1}$) overlaid with ERA-Interim total surface winds (vectors, m s$^{-1}$) and surface pressure perturbations (contours, hPa). A 2-day low-pass filter is applied to the reanalysis fields. The location of the surface observations in Fig. 11 are shown by letters, and the position of the southern vortex is annotated with a cross. Contours for surface pressure perturbations are drawn every 1 hPa with the zero contour omitted.](image-url)
convection. As evidence of this possibility, the downstream surface ridge, which was under the sinking branch of the secondary circulation, intensified simultaneously. If this reasoning is correct, then it would look at least superficially as if convection triggered this AEW, when in reality, both the convection and the intensification of the AEW were linked to the same dynamical forcing.

4. Composite secondary circulation

As a step toward generalizing these results, we will compare them with a composite AEW early in its life cycle. Although we cannot unambiguously determine cause and effect from a composite, it is important to determine if ageostrophic secondary circulations of the kind seen in section 3d are a general feature of AEWs. For the composite, we use the lagged-regression dataset constructed by Diaz and Aiyyer (2013) to analyze the energetics of composite AEWs. It is calculated by regressing relevant meteorological fields against the time series of 2–10-day meridional wind at 10°N, 15°W using the ERA-Interim from 1990 to 2010 for July–September. Note that this composite is in West Africa. We chose this region so that it is directly comparable with the energy budget of Diaz and Aiyyer (2013) and because the statistical signal for AEWs is much stronger over West Africa. If energy dispersion is an intrinsic property of AEWs, then we should see its signature in both East and West Africa.

Figure 13a shows divergence, geopotential, and ageostrophic flow at 900 hPa, and Fig. 13b shows...
800-hPa pressure-coordinate vertical velocity, 850-hPa temperature, and 850-hPa wind. All quantities except for temperature are perturbations from the mean. This figure was made to match Fig. 9 as closely as possible for comparison. The ageostrophic secondary circulation is very similar to that of the case study (Fig. 9); the rising branch is collocated with warm advection, the sinking with cold advection, and the lower is supergeostrophic, consistent with gradient wind balance. The eastward ageostrophic flow down the geopotential gradient increases EKE and converges on a region of low-level cyclogenesis.

The vertical distribution of ageostrophic geopotential fluxes yields insight into the east-to-west evolution of the vertical structure of AEWs. At 18°N, where AEWs tend to be more baroclinic, the ageostrophic geopotential fluxes are directed upstream (eastward) below 800 hPa and downstream (westward) above 800 hPa (Fig. 14a). Because the direction of energy dispersion of Rossby waves depends on the sign of the meridional PV gradient (e.g., Hoskins et al. 1985), this pattern is a manifestation of the meridional PV gradient reversing sign with height (Fig. 2). Idealized modeling studies of midlatitude baroclinic waves reveal a similar vertical arrangement of energy dispersion and ageostrophic geopotential fluxes [Fig. 4 in Simmons and Hoskins (1979) and Fig. 6 in Orlanski and Chang (1993)]. As a consequence, upstream development of AEWs near this latitude is favored below 800 hPa. This preference may explain why composite studies (e.g., Kiladis et al. 2006; Fyfe 1999) and idealized modeling studies (e.g., Hall et al. 2006; Thorncroft et al. 2008) show that AEWs centered near 15°N develop first at low levels and then become strongest at midlevels. In essence, the lower portion of an AEW packet disperses upstream, while the upper portion disperses downstream. Over time, midlevel wave activity becomes concentrated to the west and low-level wave activity to the east. The downstream dispersion at midlevels appears less effective in generating new AEWs, because it is weaker and generally directed away from the baroclinic zone.

In contrast with the pattern at 18°N, ageostrophic geopotential flux vectors point upstream at all heights at 8°N.
8°N (Fig. 14b). This arrangement indicates eastward energy dispersion throughout the lower to middle troposphere and is consistent with the mean PV gradient being positive at all heights south of about 10°N. This observation may explain why southern-track AEWs develop first at midlevels and show little change in the height of their maximum amplitude as they progress westward [our Fig. 14b; Fig. 6 of Kiladis et al. (2006)].

5. Discussion

A detailed local energetics analysis of the pre-Alberto AEW strongly suggests that energy dispersion from a preceding AEW was the direct cause of its genesis on 26 July and played a significant role in its subsequent reintensification on 30 July. Drawing on our case study, an EKE budget of composite AEWs (Diaz and Aiyer 2013), and previous studies of upstream and downstream development (e.g., Orlanski and Sheldon 1995), we propose the following sequence of events to explain the genesis of an AEW by upstream development and its subsequent life cycle:

(i) The convergence of energy fluxes from a preexisting AEW generates a wind perturbation upstream. The initial EKE source is provided by convergent energy fluxes from the preexisting AEW.

(ii) The new perturbation disturbs the gradients of basic-state PV and potential temperature and begins to produce anomalies of temperature and PV.

(iii) Once the new perturbation reaches sufficient amplitude, baroclinic and barotropic conversion become the dominant energy source. The EKE of the disturbance grows at the expense of the basic-state potential and kinetic energy. During this stage of development, the PV gradient reversals and coupling with convection are essential to continued growth. If its environment does not support continued EKE growth, then the disturbance will quickly decay through frictional dissipation.

(iv) As the new AEW intensifies, it begins to flux its own energy upstream. If EKE growth through baroclinic and barotropic conversion is overcome by EKE exportation by divergent energy fluxes and EKE dissipation through friction, then the AEW will begin to weaken. As it decays, it may generate a new AEW upstream.

In light of the finding of Hall et al. (2006) that friction stabilizes the AEJ to small perturbations, upstream development offers an attractive explanation for the genesis of AEWs. In contrast with normal-mode instability theory, the initial growth rate of AEWs is not tied to that of the most unstable normal mode, but rather to the constructive and destructive interference of the larger spectrum of normal modes, which can be neutral or even decaying. Theoretical studies of baroclinic waves show that the growth rate due to this mechanism can greatly exceed that of the most unstable normal mode (e.g., Simmons and Hoskins 1979). From an energetics perspective, the new AEWs do not grow by directly extracting energy from the basic state; their initial energy source is provided by the convergence of energy fluxes from a preexisting AEW. This process merely redistributes EKE, because flux convergence and EKE growth in one location requires flux divergence and EKE decay in another. Thus, the growth of new AEWs is canceled by the decay of existing AEWs.

As pointed out by Hall et al. (2006) and Thorncroft et al. (2008), one of the failures of normal-mode instability theory in explaining the genesis of AEWs is that the growth rate of the most unstable normal mode is too slow to explain their observed amplitudes; the unstable region of the AEJ is simply too short for individual AEWs to undergo significant amplification before they leave the AEJ. Given these circumstances, one possible way to account for AEW genesis is to require large finite-amplitude precursors. This requirement would be most applicable if all of the wave energy propagates away from its source region at the westward phase velocity. However, if some of the wave energy propagates at the eastward group velocity, then a second possibility emerges; successive AEWs can “share” energy by passing it upstream and extend the slow growth across multiple AEWs. This scenario likely plays out in the simulations of Thorncroft et al. (2008). After the initial weak AEW is triggered, the subsequent AEWs become increasing strong upstream. The upstream AEWs grow through a combination of energy extraction from the basic state and energy dispersion from their predecessors. We can even see a similar process for our case study in Fig. 1. From about 22 to 28 July, two weaker AEWs precede AEW A. These two AEWs are in turn linked to another wave packet that appears on 16 July near 30°E. The upstream flux of energy within these wave packets leads to successively stronger AEWs that culminate in the pre-Alberto AEW.

Several aspects of the behavior of AEWs are consistent with their being generated by upstream development. For example, upstream development is a periodic genesis mechanism. Once AEW activity develops, new AEWs would be generated with nearly the same 3–5-day period and 2500–4000-km wavelength of
the existing AEWs. By contrast, it is unclear how a triggering mechanism alone could account for this regular period and wavelength. Upstream development also explains why AEWs tend to travel in groups or "wave trains" (Pytharoulis and Thorncroft 1999; Carlson 1996; our Fig. 1). These wave trains are best described as dispersive wave packets (Diaz and Aiyyer 2013). The evolution of the vertical structure of AEWs is also consistent with a genesis through upstream development. Regression analyses (Kiladis et al. 2006; Fyfe 1999) and dry numerical simulations (Hall et al. 2006; Thorncroft et al. 2008) show that AEWs at 15°N develop first at low levels and then become strongest at midlevels. This evolution is consistent with their initial growth being supported by an upstream dispersing surface Rossby wave. By contrast, AEWs south of the PV gradient reversal at 10°N develop first at midlevels (Kiladis et al. 2006). This observation is consistent with upstream energy dispersion within the strong positive midlevel meridional PV gradient.

One weakness of upstream development as a general genesis mechanism for AEWs is that it does not explain the origin of the first AEW. After all, a genesis mechanism that assumes AEWs already exist is incomplete. If the real AEJ is as stable to small perturbations as the numerical simulations of Hall et al. (2006) suggest, then a finite-amplitude disturbance would be necessary to initiate AEW activity (e.g., Thorncroft et al. 2008). However, if the initial disturbance disperses a significant amount of its energy upstream, then additional triggers may not be necessary. This scenario would be compatible with the hypothesis of Thorncroft et al. (2008, p. 3600) that "the prescribed heating not only caused the initial westward-moving trough but also forced the leading linear normal mode." Alternatively, it is probable that certain instantaneous configurations of the AEJ are more unstable to small perturbations than is the mean state (Leroux and Hall 2009) or that nonlinear processes may be important for the genesis of AEWs (Thorncroft and Hoskins 1994b) or that other factors not represented by this idealized model destabilize the AEJ–AEW system. Under these circumstances, the initial disturbance is best explained through hydrodynamic instability. In either case, once initiated, AEW activity would be at least temporarily self-sustaining as new disturbances continue to form upstream through energy dispersion.

Another important aspect of AEW genesis left unexplained by upstream development is the intermittency of AEW activity. If upstream energy dispersion is an inherent property AEWs, then one might expect it to produce continuous trains of AEWs. To reconcile this discrepancy, it must be considered within the context of existing theories of AEW genesis and amplification. After their initial growth, continued amplification of these upstream AEWs will be highly sensitive to the structure of the AEJ and its associated PV gradient reversals (e.g., Leroux and Hall 2009; Hsieh and Cook 2008) and their ability to couple with convection (e.g., Berry and Thorncroft 2012). For example, although the AEW in our case study grew initially through energy dispersion from the preceding AEW, its subsequent growth resulted from baroclinic and barotropic conversion. Indeed, if the upstream environment is unsuitable for their continued growth, then friction will quickly spin down these neutral disturbances. A complete description of AEW genesis must include the role of finite-amplitude triggers, hydrodynamic instability, and energy dispersion. It is probable that each of these mechanisms explains some aspects of AEW genesis.

6. Conclusions

Upstream development is a viable mechanism to explain the genesis of AEWs. As it does not rely on either normal-mode instability of the AEJ or finite-amplitude precursors, it offers an alternate to both modal instability theory and the triggering hypothesis, although it does not exclude either possibility. Its signature is apparent in statistical composites of AEWs (Diaz and Aiyyer 2013), and it can explain the genesis of the pre-Alberto AEW. In this genesis paradigm, the role of convection and hydrodynamic instability is to amplify AEWs after they have formed, rather than directly lead to their genesis. Nevertheless, as we have presented only one case study, future work is needed to establish the generality of these results.

Acknowledgments. This work was supported by NSF Grant ATM-0847323. We are grateful to George Kiladis, Walter Robinson, Gary Lackmann, and an anonymous reviewer for their constructive criticism, which greatly helped to improve this manuscript.

REFERENCES


