Understanding Atmospheric Catastrophes

Winston C. Chao

NASA/Goddard Space Flight Center, Greenbelt, MD 20771, USA
Winston.c.chao@nasa.gov
Book Chapter, "Recent Progress in Atmospheric Sciences with Applications to the Asia-Pacific Region", Co-Editors: K.-N. Liou and M.-D. Chou. The World Scientific Pub. Co., to appear. (In press)

Abstract

The atmosphere, as in other parts of nature, is full of phenomena that involve rapid transitions from one (quasi-) equilibrium state to another—i.e., catastrophes. These (quasi-) equilibria are the multiple solutions of the same dynamical system. Unlocking the mystery behind a catastrophe reveals not only the physical mechanism responsible for the transition, but also how the (quasi-) equilibria before and after the transition are maintained. Each catastrophe is different, but they do have some common traits. Understanding these common traits is the first step in studying these catastrophes. In this article, three examples are reviewed to illustrate how atmospheric catastrophes can be studied.

1. Introduction

Changes in the physical world are of two basic types: gradual and sudden. Gradual changes are usually the norm; sudden changes are less frequent, larger in magnitude, and easily recognizable. A physical system may experience gradual change over a long period of time, and then "all of a sudden," a rapid change may occur. For example, at a typical tropical island station in the Caribbean, the surface air temperature gradually rises after sunrise and continues to rise till late afternoon, but then suddenly drops due to a thunderstorm. Often a sudden change occurs without any sudden or fast external stimulus such that a slow gradual change in a system gives way to a sudden change. Water suddenly starting to boil in a heated pot is an example. At other times a sudden change is triggered by external events, but the amount of energy provided by the trigger is usually much less than that associated with the rapid change in the physical system initiated by the trigger. An example is the triggering of cumulus convection by air flowing over hills.

When the dependence of the equilibrium state of a dynamical system on an external parameter is nonlinear, multiple equilibrium states may occur. (The equilibrium state may be stationary (fixed point), time-periodic (limit cycle), or time aperiodic (quasi-equilibrium, including strange attractor)). Fig. 1 illustrates such a conceptual picture. When the externally adjustable parameter ε is $\varepsilon_1<\varepsilon<\varepsilon_2$ two (quasi-) equilibria exist. If initially the system is on the lower branch when ε is raised passing ε_2, the system jumps from the lower branch to the upper branch. The speed of the transition and the degree of overshooting before the system settles down on the new solution branch varies from system to system. Regardless, a transition, such as this, from one (quasi-) equilibrium solution to another when an external parameter passes a critical value is called a catastrophe. Alternative to an external parameter passing a critical value, a catastrophe can be triggered by an external disturbance providing enough energy for the system to overcome the energy hump, which is of course larger when the external parameter is further away from the critical value of ε_2. During the jump, the system may overshoot the new equilibrium and bounce back and oscillate about the new equilibrium before settling down. The overshooting may or may not occur, depending on the transient response of the system and the frictional force it experiences. A simple example to further illustrate a catastrophe is as follows (Fig. 2). A particle is held in a potential well as the structure of the well changes, the well eventually disappears, and the particle falls into a neighboring well. This fall takes place suddenly; it is called a spontaneous catastrophe. The fall can also take place without the well the particle resides in disappearing, if the particle is pushed over the top of the well (energy barrier). This is called a triggered catastrophe.

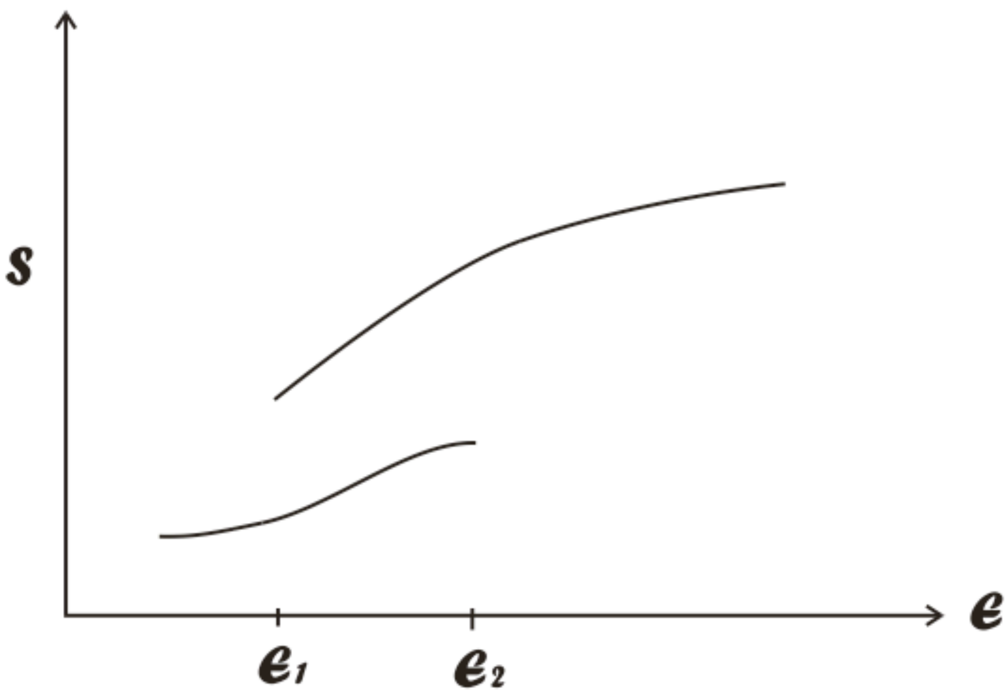


Fig. 1 Schematic diagram showing the state of a system as a function of an external parameter ε. The system has two states between ε_1 and ε_2.

Fig. 2 Schematic diagram showing the state of a particle in a potential well. As an external parameter changes the shape of the potential well changes such that the particle changes its position suddenly.

The concept of catastrophe (Thom 1972, Poston and Stewart 1978, Saunders 1980, Iooss and Joseph 1980) exists in many disciplines and has been discovered by many researchers independently. Thus, it is not surprising that catastrophes have also been called by many other names. In many disciplines a catastrophe is called a subcritical instability (e.g., Drazin 1992). The name hysteresis is also used when the emphasis is on the occurrence of multiple (quasi-) equilibria instead of the suddenness of the transitions. In engineering, it is called a structural instability. In the physical sciences, it is often called a critical phenomenon due to the fact that a parameter must exceed a critical value for the catastrophe to occur; or it may be called a nonlinear instability due to the fact that, unlike in a linear instability, the growth rate is not constant. The term nonlinear instability includes subcritical and supercritical instabilities (Drazin 1992). The forcing acting on a system to pull it away from an equilibrium that has just disappeared (in the case of a spontaneous catastrophe) increases as the system moves further away from the equilibrium. This is followed by a reduction of the forcing as the system approaches the other equilibrium. Thus the jump is under positive and then negative acceleration after the system overshoots the new equilibrium. Such accelerated growth lasts, of course, only for a finite time period and eventually will have to stop due to the finite energy supply of the basic state; i.e., when the system reaches the other quasi-equilibrium. The name explosive instability (Sturrock 1966) is sometimes used because the forcing pushing the system away from the initial (quasi-) equilibrium increases with time initially. Since it can be triggered by a finite-amplitude initial disturbance (a trigger), it is also called finite-amplitude instability. In using the term finite-amplitude instability, one should note that triggering is merely an alternative way to initiate the instability and is not necessary when an external parameter is changed to exceed a critical value. Nonetheless, in many cases triggering is the more likely way an instability gets started due to the frequent occurrence of perturbations in the atmosphere. Before a triggered catastrophe occurs, the system is stable and this stability is called metastability.

The term “catastrophe”, for a while, was considered an oversold concept and it incurred some connotation of a fraud, when it was overly publicized in the seventies. It has since regained its respectability after the initial commotion died down. It is now accepted as a standard mathematical term (Arnold 1981) without the implication of something negative happening necessarily. We use it instead of its equivalent terms—explosive instability, subcritical instability, finite-amplitude instability, critical phenomenon, and structural instability—for brevity.

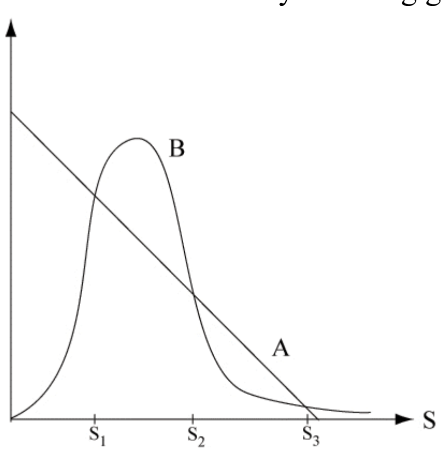
In attempting to identify if a particular phenomenon is a catastrophe, one strives to ascertain if the common traits of catastrophes can be found. These common traits are bimodality (or multiple equilibria with at least two stable equilibria and an unstable one in between), suddenness of occurrence and spontaneous or triggered onset. Each catastrophe has its own cause for multiple equilibria, feedback processes in the transition, and triggering mechanisms (if they exist). The majority of the catastrophes can be easily identified through observation. One looks first for suddenness of the change. Suddenness means the duration of the change is short relative to the less eventful periods before and after the event. Bimodality--the next thing one looks for--means the states before and after the change are clearly different: i.e., the change is sizable relative to the changes that take place before and after the transition. A phenomenon termed “onset” or "genesis" is very likely a catastrophe. Identification of the triggers of a catastrophe is not always very easy since triggers do not always exist. Moreover, events concurrent with a catastrophe may be a result of that catastrophe rather than the triggers of it. Thus identifying triggers when they are not obvious can only be reliably accomplished through theoretical and modeling work, as in the case of a complete understanding of each catastrophe.

As far as understanding catastrophe is concerned, the terms in the governing equation of a system that exhibits catastrophic behavior can be grouped (at least conceptually) into two competing sets (or two forcings), A and B. Thus,

$$\partial S/\partial t = A - B \quad (1)$$

Fig. 3 is an example of Eq. (1). Initially, the state is at the stable equilibrium represented by point S_1 in Fig. 3[1]. Point S_2 is an unstable (quasi-) equilibrium state and Point S_3 is another stable equilibrium state. As one or more parameters change such that curve A moves right ward while keeping its slope and/or the peak of curve B diminishes to such a degree that the point S_1 (quasi-) equilibrium can no longer be sustained, the state moves rapidly to (quasi-) equilibrium point S_3. The movement is propelled by the difference between A and B in a "free fall" (more about this shortly). Fig. 3 was used by Held (1983) to explain the topographically induced Rossby wave instability of Charney and DeVore (1979), which is a catastrophe. It is obvious that the existence of multiple quasi-equilibria alone is not sufficient for a catastrophe to occur. The relative movement of A and B induced by changes in external parameters or a trigger is necessary for a catastrophe to occur.

Fig. 3 is a useful conceptual figure; however, it needs some clarification. According to Eq. (1) when the state reaches S_3, $\partial S/\partial t=0$ and the state should stop changing and no overshooting should occur. But overshooting is often observed. The correct description is that A and B are functions of not only S but also $\partial S/\partial t$; what is depicted in Fig 3 are A and B when $\partial S/\partial t=0$--i.e., steady state forcings. Held (1983) pointed out that the curve B he used represents the drag exerted on the zonal flow by the *steady* forced waves generated by topography in the presence of dissipation. Thus the "free fall" is driven by a forcing greater than what appears in Fig. 3 as A-B.



[1] Since we have used S to denote a gross state, points 1 and 3 may be quasi-equilibria, rather than equilibria, in a multi-dimensional system.

Fig. 3. A and B as functions of S. The equilibrium states S_1 and S_3 are stable, but not the one at S_2.

Moreover, within the domain of the S-variable where one of the forcings varies highly nonlinearly, the other forcings usually does not vary highly nonlinearly and can thus be represented by a linear or a quasi-linear line; but there can be exceptions.

Other examples of catastrophes are abundant. A simple example is the buckling found in flipping a wall switch or an earthquake. Another example is an explosion of any type. Other atmospheric examples are also abundant. In this article, three atmospheric catastrophes: tropical cyclogenesis, stratospheric warming, and monsoon onset--chosen based on the author's research interest--are reviewed in some detail, to illustrate how one can study atmospheric catastrophes. Our understanding of these phenomena is still not complete; we will try to point out areas that require further study. Many more atmospheric catastrophes are enumerated below with varying degrees of commentary. In studying these catastrophes, one strives to identify the forces or forcings within that explain the associated multiple equilibria to arrive at a schematic diagram like Fig. 3. Also, one must support diagrams such as this by theoretical arguments and/or numerical experiments.

2. Tropical cyclogenesis

Tropical cyclogenesis is the transition from a cloud cluster to a tropical cyclone (TC). Prior to the transition, the cloud cluster can last for days without much change in its characteristics. Thus, one can claim that the cloud cluster is in a quasi-equilibrium state. After the transition, the tropical cyclone can also last for days without changing its basic characteristics. Hence, one can identify the tropical cyclone as a quasi-equilibrium state as well.

Tropical cyclogenesis typically takes only about two or three days. Relative to the duration of either the cloud cluster or the TC this transition period is very short. Tropical cyclogenesis occurs when the cloud cluster (quasi-equilibrium) state can no longer be sustained (a spontaneous catastrophe) or when a trigger acts on a cloud cluster (a triggered catastrophe). A spontaneous tropical cyclogenesis must be associated with a certain condition being met; in a triggered tropical cyclogenesis this condition is very nearly being met. Although the precise details of this condition are still unknown, it is generally associated with what is already known: SST higher than 26.5°C, low background vertical wind shear, and a sufficiently high Coriolis parameter (Chap. 15 of Palmen and Newton 1969). When the condition is met, all cloud clusters can turn into TCs. It is well-known that a series of TCs often occur concurrently. Through these identified characteristics: two quasi-equilibrium states and the rapid transition between the two, etc., tropical cyclogenesis clearly can be identified as a catastrophe.

It is heuristic to construct a schematic diagram similar to Fig. 3 in understanding the catastrophic nature of tropical cyclogenesis. One can use the 3-d mass-weighted average temperature in the core region (e.g., within a 30 km radius) of a disturbance minus that of the environment--i.e., the degree of warming of the core region--as the state variable S. Curve A represents the diabatic heating--which is mostly cumulus heating--in the core region. Curve B is the adiabatic cooling due to upward motion in the same region. Assuming the thermal balance is mainly between A and B, A-B is zero at both quasi-equilibria: the cloud cluster and the TC. Fig. 4a shows A-B at the moment the cloud cluster loses its equilibrium status. A-B is zero at both the cloud cluster quasi-equilibrium and the TC quasi-equilibria. As shown in Fig. 4a at the moment the cloud cluster loses it equilibrium status, A-B does not cross the abscissa at the cloud cluster quasi-equilibrium S_1 but only touches it tangentially. A-B crosses zero at the TC quasi-equilibrium S_3, which is a stable quasi-equilibrium, since a perturbation away from this quasi-equilibrium will be reduced by A-B to zero. The positive value of A-B on the left side of the TC quasi-equilibrium S_3 ensures that the state starting from the cloud cluster status will move to the TC status.

Curve B can be roughly represented by a straight line through the origin. This is because a disturbance with a higher core temperature has a stronger meridional circulation, which implies stronger adiabatic cooling in the core region. As a result of these considerations, A and B at the moment of the cloud cluster's losing its quasi-equilibrium status can be represented as in Fig. 4b. Before that moment, the picture is as depicted in Fig. 4c. In this figure, the cloud cluster corresponds to S_1 and the TC S_3. S_2 is an unstable quasi-equilibrium. As the boundary conditions change (e.g., the SST increases), curve A moves upward and S_1 and S_2 merge and then both disappear--i.e., the cloud cluster can no longer be maintained. The system rapidly moves toward the TC state in a "free fall" pulled by a forcing greater than A-B. There can be overshooting.

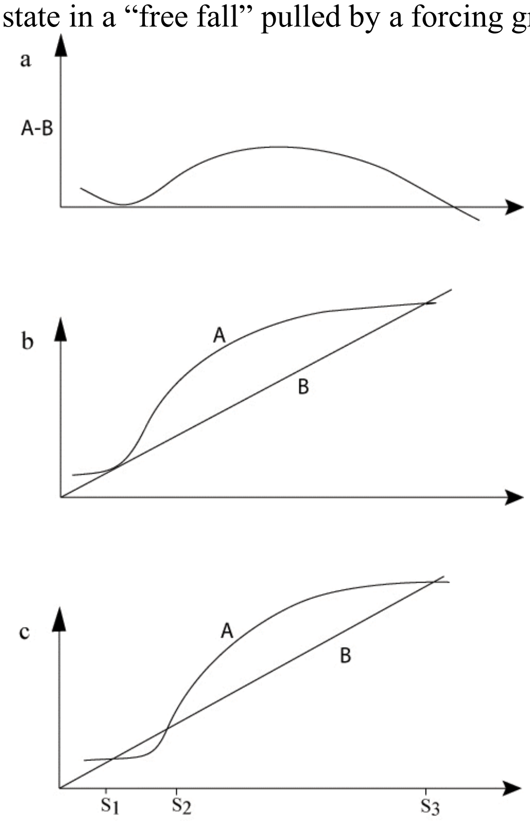


Fig. 4. Schematic diagram showing the strength of the forcings on a tropical disturbance. The abscissa is the state of the system, S_1 denotes the cloud cluster state and S_3 the TC. Curve A is the adiabatic heating and line B is the cooing due to adiabatic upward motion in the core region. See the text for more detail.

The remaining task is to explain the shape of curve A. As S increases, curve A increases because a higher S causes a stronger meridional circulation, which brings in more moisture and causes more evaporation through the associated stronger surface winds, thus resulting in more convective heating. The rate of increase of A is initially modest, but becomes greater as a result of the increase in wind speed in the boundary layer and the corresponding increase in evaporation at the surface as the vortex spins up. A increases at a reduced rate, as the state gets close to S_3, as a result of the higher temperature in the core region, which reduces vertical instability and hinders further increase in convective activity. Moreover, with increased convergence at low levels and increased divergence at high levels, inertial stability provides further inhibition of meridional circulation. Also, as the boundary-layer air approaches saturation, evaporation rate cannot keep increasing, thus limiting the further intensification of convective activity.

The foregoing interpretation of tropical cyclogenesis as a catastrophe was presented in Chao et al. (2003) and needs to be supported with modeling effort.

3. Stratospheric Sudden Warming

In the polar region of the stratosphere, the temperature in winter can rise by more than 40°K in as little as a week followed by a substantial drop, which is just as fast. The corresponding zonal mean zonal wind, consistent with the thermal wind relationship, can turn easterly in a major warming event. This phenomenon is called stratospheric sudden warming (SSW). It may occur more than once during a winter; the last episode, called the final warming, is often the strongest and precedes the spring flow pattern.

SSW is a catastrophe, due to the interaction between the zonal flow and planetary waves. Chao (1985) proposed a mechanism explaining SSW, as depicted in Fig. 3. This figure depicts the "forcings," or attractions, acting on the zonally-averaged zonal wind [u] (the abscissa). The effect of a stationary baroclinic Rossby wave of a zonal wave number on the zonal mean flow is represented by Curve B; a positive value denotes westward acceleration. This acceleration occurs when the Rossby waves (i.e., the eddies) carry an upward westward momentum flux that is then deposited in the stratosphere due to radiative damping in a quasi-equilibrium situation. However, the transient effect of wave-zonal wind interaction (Section 12.4 of Holton 2004)[2] is prominent during the jump. Curve B has a peak at the zonal mean zonal wind speed corresponding to the resonance speed (i.e., $c=u-\beta/k^2=0$ for the stationary Rossby wave, where c is the phase speed of the Rossby wave, u the zonally averaged zonal wind, $\beta= \partial f/\partial y$; when the phase speed of the Rossby wave matches the speed of the mountain, which is zero. At resonance, the momentum flux due to planetary-scale topography transported upward by the Rossby wave reaches its maximum.

The forcing on [u] due to the meridional gradient of the radiative-convective equilibrium temperature is depicted is depicted by line A. Line A pulls [u] toward the state of radiative equilibrium (a positive A represents eastward acceleration), which is

[2] Related to A and B being functions of $\partial S/\partial t$, as mentioned in the Introduction.

represented by the point where line A intersects the x-axis, S_3. In mid-winter, this point has a very high zonal mean zonal wind, and it is the state that exists when there is no topographical forcing (i.e., when Curve B is zero) and the high zonal mean zonal wind reflects, through the thermal wind relationship, the large meridional gradient of the radiative equilibrium temperature. In mid-winter the state is at S_3. As winter progresses toward spring, the zonal mean zonal wind corresponding to the radiative equilibrium meridional temperature gradient becomes smaller, and thus line A moves toward the left in Fig. 3 while retaining its slope. Eventually S_2 and S_3 merge and both then disappear. At this time, the difference between line A and curve B reflects a westward acceleration of the zonally-averaged zonal wind. This difference becomes large as [u] diminishes and approaches the resonance speed. In the meantime, the amplitude of the Rossby wave increases due to the resonance. Thus, [u] races toward the other equilibrium state, S_1. Chao's (1985) model experiments showed S_1 to be a quasi-equilibrium state rather than an equilibrium state, in fact, an oscillatory state in the simplified model used. Therefore, Fig. 3 can only serve as a conceptual aid not as an actual depiction of the catastrophe. Due to the tremendous difference between curve B and line A at resonance, overshooting can occur; thus [u] can pass S_1 and, in some extreme cases, become negative. That is why the westerly wind [u] weakens fast and may even turn easterly; correspondingly, the temperature in the polar region rises rapidly. However, the warmer polar temperature exists only momentarily, since the state of the system soon bounces back from the overshooting. Such bouncing back can be as fast as the overshooting itself; hence, sudden warming is followed immediately by sudden cooling.

The description above is based on a model that has only one zonal wave number (see Chao (1985) for details.) When all zonal wave numbers are included, how the description should be modified still remains to be worked out. Moreover, what is described above explains the final warming well; warmings earlier in the season remain to be explained.

4. Abrupt latitudinal movement of the ITCZ (monsoon onset)

The intertropical convergence zone (ITCZ) is the location where surface winds converge and rise into the upward branch of the Hadley/Walker circulation. The latent heat released in the ITCZ drives the Hadley/Walker circulation. In other words the ITCZ is a part of the Hadley/Walker circulation. The latitudinal location and the intensity of the ITCZ impact the tropical surface wind distribution, which, in turn, impact the air-sea interaction. The latter is known to play a crucial role in El Nino. The convection within the ITCZ in the Indian Ocean and western Pacific has an intraseasonal oscillation, which is called the Madden-Julian oscillation (MJO). The MJO is strongest when the ITCZ is over the equator. Thus the annual cycle of the ITCZ latitudinal movement determines the seasonal variation of the MJO intensity.

Because of its importance the latitudinal location of the ITCZ has attracted the attention of many prominent researchers, such as Charney (1971). Charney's theory of the latitudinal location of the ITCZ was built on his CISK theory (conditional instability of the second kind, Charney and Eliassen 1964), which he used to explain tropical cyclogenesis. Chaney's ITCZ theory is a zonal mean version of his CISK idea.

According to CISK, synoptic-scale convection relies on frictionally-induced boundary-layer convergence (i.e,, the Ekman pumping), and thus favors higher latitudes; i.e., the poles. Charney also invoked the idea that the moisture supply is higher over the equatorial tropics than over the polar regions. The compromise of these two factors gives the ITCZ a latitudinal location close to, but not at, the equator.

The Charney ITCZ theory was quite influential. Many people adopted or embellished it (see the appendix of Chao and Chen 2004). However, Charney's ITCZ theory and its offshoots turned out to be in conflict with the aqua-planet (AP) experiments of Sumi (1991), Kirtman and Schneider (1996), Chao (2000) and Chao and Chen (2001a, 2004). These researchers used AP models with uniform sea surface temperature (SST) and solar angle--which eliminated the second factor in Charney's theory--and obtained ITCZ over or near the equator (at about 14º N or S) instead of over the polar region, as the first factor of Charney's ITCZ theory would predict. What is relevant to our topic is that the region adjacent to the equator (e.g., from 4° to 8°) is not, or rarely, accessible by the AP ITCZ, which gives a clear indication of multiple equilibria.

The following is the author's explanation (from Chao and Chen 2004) for the AP model results. An AP model with uniform SST and solar angle would have uniform time-averaged precipitation if the earth's rotation rate, Ω, were set to zero, since in the absence of Ω, one location on the globe is indistinguishable from another. When the earth's true rotation rate is used, convection finds preferred latitudes. To explain these preferred latitudes, one should look at the cause of convection, which is that vertical instability in the presence of the Coriolis parameter turns negative; in other words the squared frequency of the inertial gravity wave, $\sigma^2 = f^2 + \alpha^2 N^2 + |F|$, turns negative, convection occurs (Eq. 8.4.23 of Gill 1982). Here, N^2 is the vertical stability, $|F|$ is the stabilization due to friction, and α is the ratio of the vertical scale to horizontal scale of the wave or convective cell. The fact that f^2 is added to a term proportional to N^2 in the definition of σ^2 indicates that f^2 has an equivalent effect on convection as N^2; see Chao and Chen (2001a) for an explanation of this equivalence. When α is large, as in the case of individual clouds with small horizontal scales, f^2 can be ignored. However, when we consider synoptic-scale convective systems in the ITCZ, f^2 is not negligible. The equivalence of f to N makes the equator an attractor for the ITCZ. Thus when N^2 is globally uniform, σ^2 is a minimum at the equator. This implies that convection, or the ITCZ, should occur at the equator--this being a first effect of f on the ITCZ. In an AP model convection must occur somewhere given the destabilizing effects of radiative cooling and the surface sensible and latent heat fluxes. Convection, or more generally the ITCZ, occurs at the latitude where the atmosphere is most unstable--i.e., $\partial\sigma^2/\partial\phi = 0$, where ϕ is latitude. This means that the ITCZ occurs at the latitude where $-\partial f^2/\partial\phi$ is balanced by $\partial(\alpha^2 N^2)/\partial\phi$, if $|F|$ is ignored. N^2 is reduced by the boundary evaporation, which is enhanced by the tangential wind component in the boundary layer, which, in turn, is induced by the Coriolis parameter—which is a second effect of f. Thus the poles are additional attractors of the ITCZ.

$\partial(\alpha^2 N^2)/\partial\phi$ is *the latitudinal gradient of the f-modified surface wind-evaporation feedback mechanism (the role of α^2 is yet to be explored)*. Unfortunately, it remains a very difficult challenge, if not an impossibility, to obtain an analytical expression for $\partial(\alpha^2 N^2)/\partial\phi$. For the stability of the individual clouds, N^2 is equal to $g\partial \ln\theta/\partial z$, where θ is the potential temperature (or equivalent potential temperature for a saturated atmosphere); but for synoptic-scale cloud systems (or as in GCMs, where individual clouds are not simulated and cumulus parameterization is used to represent an ensemble of clouds), N^2 should be the vertical stability for cumulus ensembles rather than for individual clouds. Thus, N^2 has no known simple tractable analytical expression. Therefore, one has to seek other means to push the investigation forward, as we will soon discuss.

The gradient $\partial f^2/\partial\phi$ can be identified as the forcing due to the ITCZ attractor at the equator, and $\partial(\alpha^2 N^2)/\partial\phi$ as the forcing due to the other attractors at the poles. The former gradient is equal to $8\Omega^2 \sin\phi \cos\phi$, and is represented by curve A in Fig. 5a; the latter gradient is represented by curve B in Fig. 5a. Since curve A has a known analytic form and does not depend on the model design, it is easily understood. Curve B, however, has no known analytic form; it has been constructed through numerical experimental results and theoretical arguments. Curve B depends on the way N^2 is affected by the model design and, in particular, by the design of the model physics. Since curve B represents the attraction on convective systems due to the attractor at the poles through the second effect of f, curve B is zero at the poles--the center of the attractor. Also, curve B depends on the gradient of f--remember that curve B is the gradient of $\alpha^2 N^2$, and $\alpha^2 N^2$ is affected by f. In other words, curve B depends on β. This gives Curve B a maximum at the equator. However, since the convective system is fairly sizeable, when its center is close to the equator a part of the system is in the other hemisphere; therefore, it also experiences the attraction by the other pole. Thus, Curve B has to be zero at the equator, where the attractions due to both poles cancel. This is shown in Fig. 5b with the dashed curve being the attraction due to the second effect of f if the size of the convective system is not accounted for. The solid curve, curve B--being the running average of the dashed curve with an averaging window the size of a tropical synoptic system—represents the net attraction experienced by the convective system.

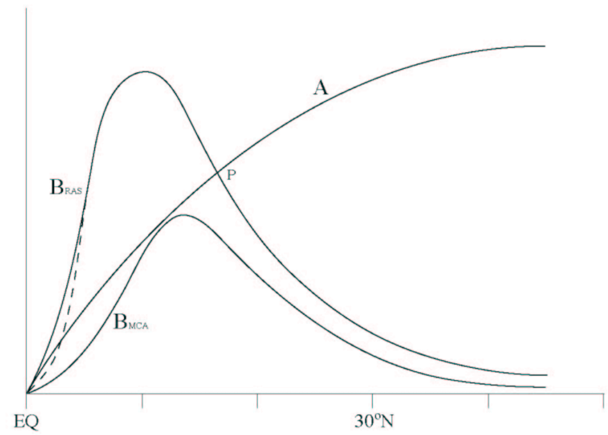


Fig. 5a Schematic diagram showing the strength of the two types of attraction acting on the ITCZ. The difference between the solid and dashed curve Bras lies in their slopes at the equator.

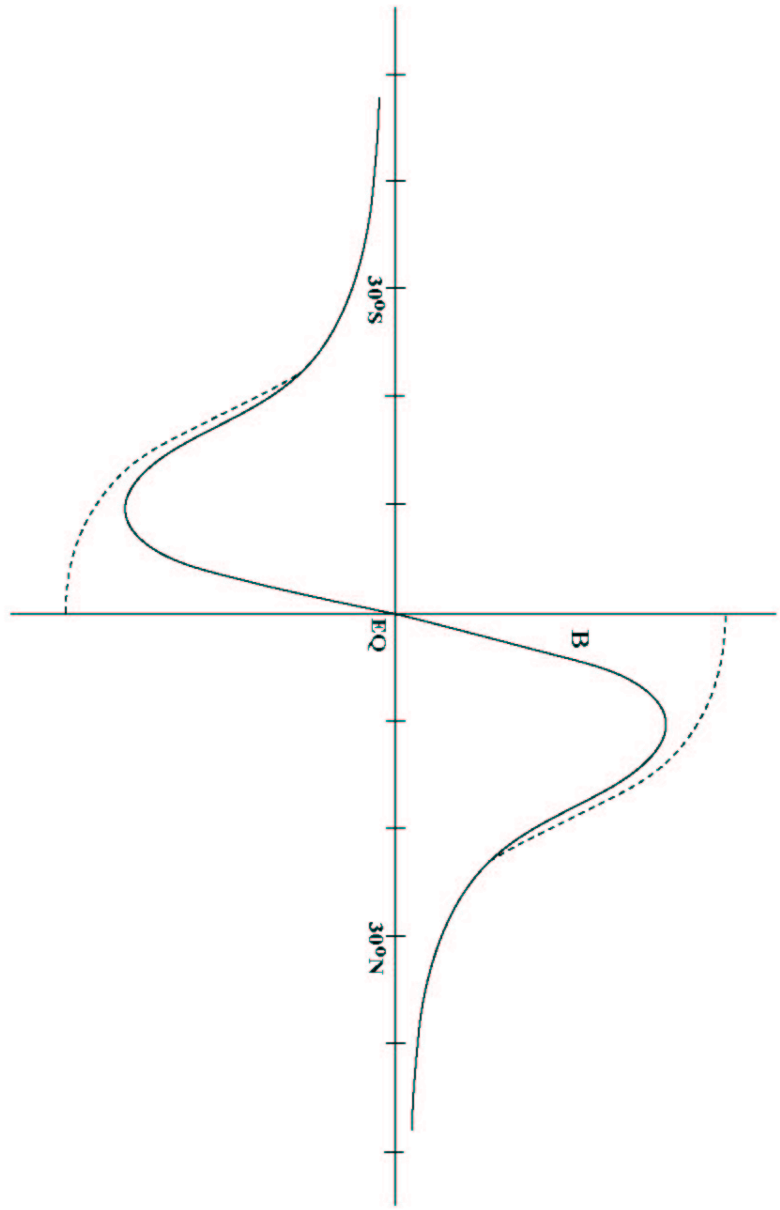


Fig.5b. Schematic diagram showing curve B (the solid line) as the latitudinal running mean of the magnitude of the attraction toward the poles

The dependence of Curve B on the cumulus scheme is illustrated in Fig. 5a. Bras and Bmca represent Curve B when the relaxed Arakawa and Schubert scheme (RAS, Moorthi and Suarez 1992) and Manabe's moist convective adjustment (MCA) scheme are used, respectively. As for the relative height of Bras and Bmca, let's first recall that curve B represents the latitudinal gradient of f-modified surface heat fluxes. When f is larger the surface winds that converge toward the center of a synoptic-scale convective system (which is a constituent of the ITCZ) develops a greater tangential wind component, which adds to the wind speed and thus enhances the surface heat fluxes. Convective cells simulated with MCA are usually smaller than those simulated with RAS and have faster surface wind speed toward the center of convective system which does not allow enough time for the surface wind to be fully modified by the Coriolis parameter to allow the tangential wind component to fully develop. Therefore, curve B under MCA is smaller than that under RAS. See Chao and Chen (2004) for more detailed discussion. If a condition is imposed on the RAS cumulus parameterization scheme such that the boundary layer relative humidity must exceed a critical in order for the scheme to operate, then this critical value is increased from 90% to 95%, curve B will change from Bras to Bmca. At 90%, the criterion does not impose much restriction on RAS, but at 95% the restriction is strong enough to make RAS behave more like MCA. MCA requires both neighboring levels of the model to be saturated for it to operate, and thus is quite restrictive. In Fig. 5a as Curve B rises from Bmca to Bras the location of the intercept between Curve A and Curve B--i.e., the location of the ITCZ--remains at the equator until the slope of Curve B at the equator exceeds that of Curve A at the equator. Then the equator is no longer a stable equilibrium location, and the ITCZ jumps to the latitude of the other intercept, P. This jump is very fast because the ITCZ is pulled in a "free fall" by the large difference between the two curves. On the other hand, when Curve B decreases from Bras to Bmca, point P--or the ITCZ--moves gradually toward the equator until the peak of Curve B gets below Curve A. At this moment P disappears and the state jumps toward the equator. But in this case the difference between the two curves is much smaller than in the case of the ITCZ jumping away from the equator, which involves a rising Curve B. Thus, the move of the ITCZ back to equator is not as abrupt.

The above deductions are borne out in experiments shown in Fig. 6a, where the boundary layer relative humidity criterion is held at 90% for the first 200 days and is changed linearly in time to 95% over the next 100 days and then kept unchanged for the reminder of the experiment. As expected, at the beginning of the experiment shown in Fig. 6a the ITCZ is away from the equator and then moves to the equator gradually as the RH criterion is increased. The fact that the simulated ITCZ is asymmetric with respect to the equator in the first phase of the experiment remains to be explained. Fig. 6b shows the results of an identical experiment except that the values of 90% and 95% are switched. In this experiment the move away from the equator is abrupt, characteristic of a catastrophe. Also, in another experiment, the wind speed used in computing the surface sensible and latent fluxes is fixed at 5 m/s everywhere. Because Curve B has been eliminated, the ITCZ is located at the equator. A further experiment shows that radiation can have impact on the ITCZ location through its effect on N^2 and in turn on Curve B. Different radiation packages lead to different amounts of forcing for the cumulus convection, resulting in differing convective systems of different intensity. This further means different magnitudes of surface winds and thus different heights for Curve B.

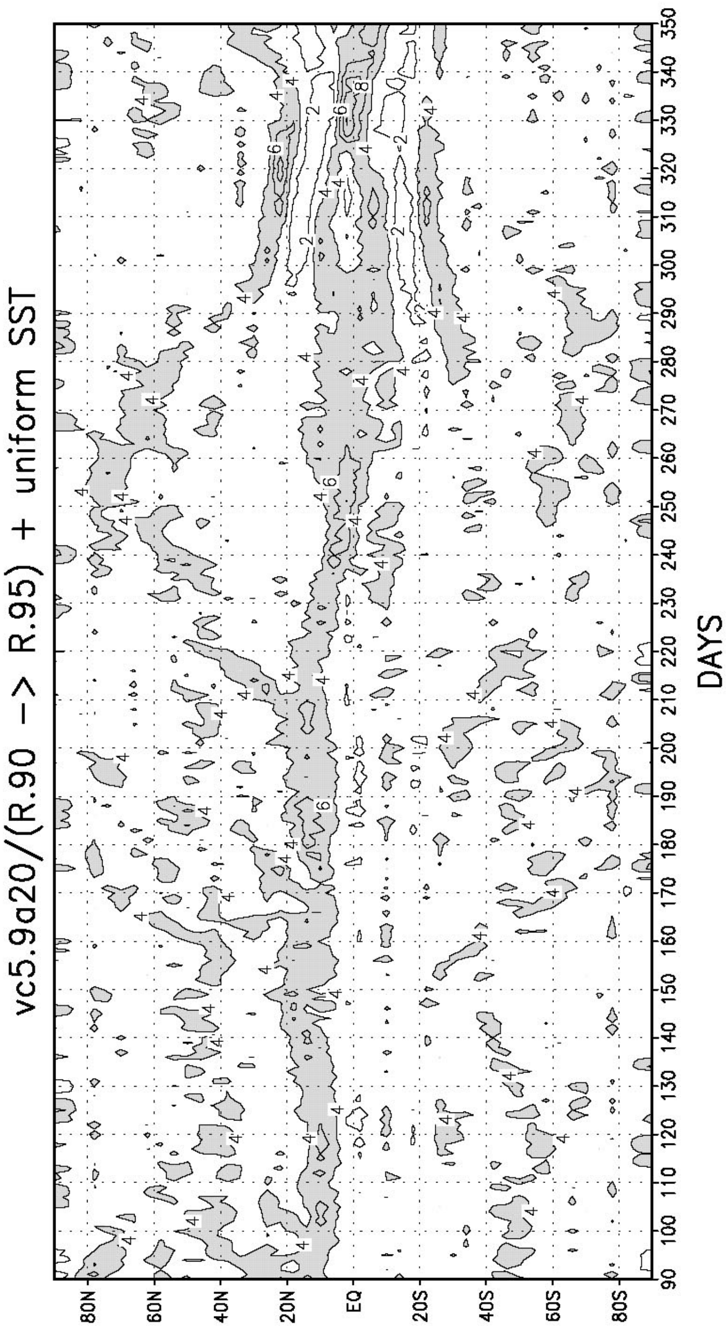


Fig. 5a Zonally averaged precipitation rate (mm/day) in an AP experiment with globally and temporally uniform SST (29°C) and solar angle using RAS. The critical boundary layer humidity is set at 90% in the first 200 days and is changed to 95% linearly in the next 100 days.

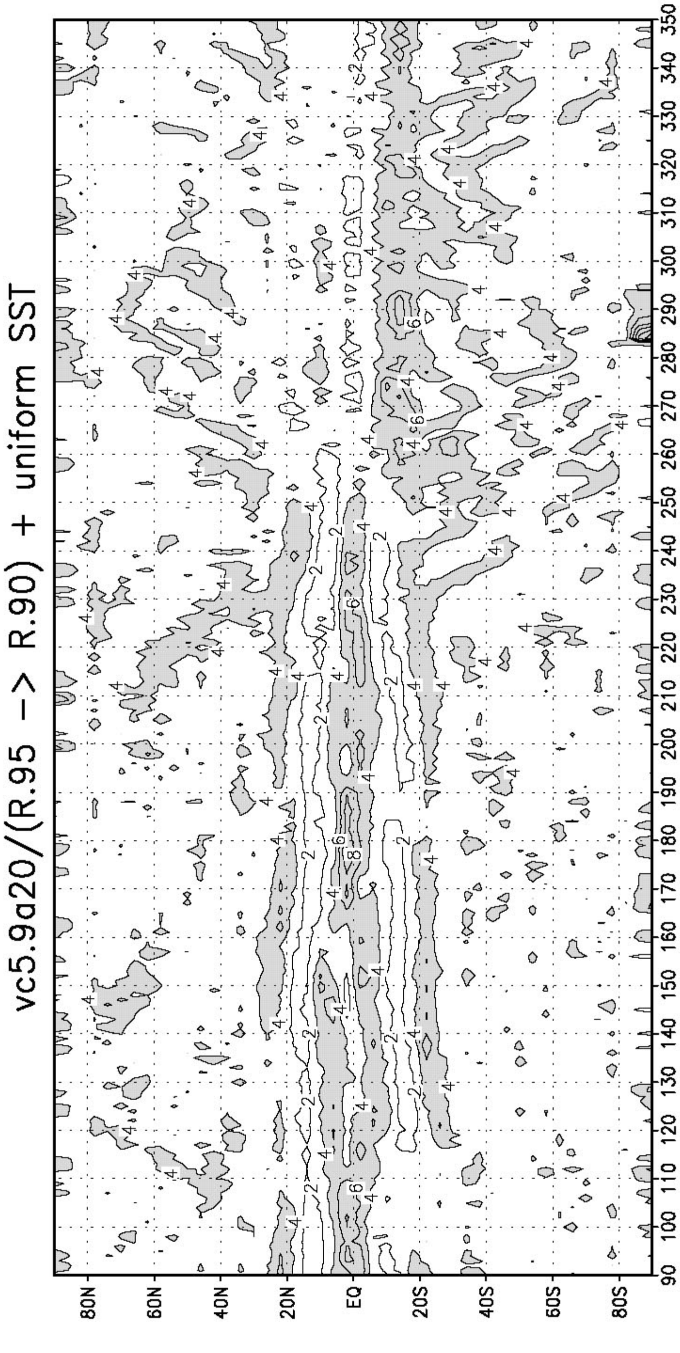


Fig. 5b Same as Fig. 5a except that the values of 90% and 95% are switched.

The above interpretation of the abrupt ITCZ latitudinal movement may be expanded to explain monsoon onset. Chao and Chen (2001b) argued that a monsoon is no more than an ITCZ sufficiently removed from the equator during its annual cycle of latitudinal movement. This movement is not completely smooth; there can be a sudden large shift away from the equator, which can be interpreted as a monsoon onset (Chao 2000, Chao and Chen 2001a). This concept also lead to the argument that a monsoon is an ITCZ after it suddenly shifts away from the equator and it does not have to rely on land-sea thermal contrast for its existence (Chao and Chen 2001b.)

The location of the ITCZ is obviously also influenced by land-sea distribution of the globe and by ocean-atmosphere interaction, which makes the subject more complex and interesting than it would be otherwise. The reader is referred to Philander et al. (1996) and Xie and Saito (2001) about these additional factors.

5. Other examples of atmospheric catastrophes

In this section we will mention more atmospheric catastrophes, with varying degrees of commentary. These catastrophes either are already well-known as catastrophes or can be easily recognized as such. The underlying dynamics of some of them are not very clear; therefore, many of them are good research topics.

Polar ice cap instability in energy-balance climate models

This most well-known example of a catastrophe in climatology is found in energy-balance climate models. Budydo (1969) and Sellers (1969) demonstrated, in a one-dimensional model, a catastrophic transition from a globe that has ice in the polar regions only to an ice-covered globe by lowering the solar constant by a few percent. A description of this catastrophe can be found in textbooks on climatology, such as Hartmann (1994).

Abrupt climate changes

Ocean-atmosphere interaction gives rise to additional catastrophes. The onset of an ice age, a long-term drought, and abrupt changes in ocean general circulation are some examples.

Barotropic and baroclinic instabilities

The textbook explanation for these instabilities presents only a linear growth rate depiction based on linear analyses, which is of limited use. The linear analysis approach starts from a basic (quasi-) equilibrium state, such as S_1, S_2 and S_3, in Fig. 3, adds perturbation and determines if the perturbation can grow and, if so, its growth rate. If a state such as S_2 in Fig. 3 is unstable, the perturbation will then grow at a constant growth rate, and can be of either sign. The obvious limitations of linear analysis are that the unstable equilibrium is not physically realizable and that the analysis is valid only for a very short period of time, before nonlinear effects become sizeable. Therefore, to get a more complete picture of these instabilities, one should turn to the concept of catastrophe. In studying these instabilities as catastrophes, one needs to ask what the equilibria are before and after these instabilities and what the forcings sustaining the equilibria are, in order to arrive at a complete picture of the whole life span of the phenomenon. In a catastrophe, the growth rate is initially zero, when the system loses its equilibrium, and then increases and finally decreases over the lifespan of the catastrophe. The growth can be in one direction only. The trigger in a triggered catastrophe has to be in the same direction as the growth of the catastrophe.

Middle-latitude explosive cyclogenesis

This is a good research topic. Linear baroclinic instability studies (Section 6.e of Hoskins et al. 1985) are not adequate to explain cyclogenesis, due to its nonlinear nature and the convective heating involved.

Jumping of Mei-yu front

The Mei-yu front in China jumps in its movement northward. This is a good research topic.

Boundary layer instability

There can be catastrophes in the boundary layer (e. g., Randall and Suarez 1983).

Blocking onset

The topographically induced Rossby wave instability has been used to explain blocking onset (Charney and DeVore 1979).

Onset of cumulus convection

This is the most obvious, but not necessarily the most understood, atmospheric catastrophe.

Low-order models

Lorenz's celebrated equations, exhibiting sudden changes in the flow characteristics of the system he studied, are very well-researched (e.g., Sparrow 1986, Lorenz 1993).

ENSO

ENSO has been studied intensely. The suggestion that ENSO can be triggered by the MJO (Yu and Rienecker 1998) implies the possibility that the onset of ENSO is a catastrophe.

Tornado-genesis

This is an excellent research topic, since so little is known about its dynamics (Church 1993, Mak 2001).

Annulus experiments

The various flow regimes in annulus experiments show a considerable amount of overlap in the regime diagram (Lorenz 1967, Holton 2004) and the transition among them is often quite abrupt.

Transition to turbulence.

Routes to turbulence have many catastrophes.

Martian polar warming

The stratospheric sudden warming mechanism set forth in Section 3 has been used to explain Martian polar warming (Barnes and Hollingsworth 1987).

Onset of Martian dust storms

Dust storms lasting months can engulf nearly the whole Martian atmosphere within weeks of their onset. Their growth is explosive and certainly fits the description of a catastrophe. There have been attempts to explain this growth as a catastrophe, but no consensus has been reached (Fernandez 1997).

6. Remarks

The above examples demonstrate that the concept of catastrophe is fundamental to atmospheric dynamics. While some of the specifics of these examples may later turn out to be incomplete or even incorrect, the general concept of catastrophe will still remain useful. When we study (quasi-) equilibria such as Rossby waves, cloud clusters, and gravity waves, we need to ask not only how they are maintained, but also if and how they can become unstable, and how they transition to other (quasi-) equilibria.

Why are there so many catastrophes in the atmosphere? The answer has two components. First, as seen from the examples in the preceding section, multiple (quasi-) equilibria exist because one or both of the forcings that maintain a quasi-equilibrium are highly nonlinear functions of a gross measure of the physical system. The atmosphere is full of factors contributing to nonlinearity. Among them are the earth's rotation, resonance, the structure of the external heating, air-sea interaction, etc. Second, there is an abundance of mechanisms that can act as triggers in moving a system from one (quasi-) equilibrium to another. Among these mechanisms are the annual cycle, the diurnal cycle (in spontaneous catastrophes) and various perturbations (in triggered catastrophes.) (The annual cycle and diurnal cycle can be viewed as triggers in a larger sense.)

The reasons for nonlinearity are different for different atmospheric catastrophes. Finding them is the core challenge of studying atmospheric catastrophes. The examples given in this review may not be sufficient to serve as a complete guide to study other atmospheric catastrophes, but they do point in the general direction.

Finally, how can we turn the knowledge gained from studying atmospheric catastrophes into practical use, such as improving the performance of forecast models? One possibility is rather than assessing the performance of global models by comparing the seasonal mean state of the model with observations—the most common practice—one should compare the catastrophes simulated by the models with those observed--such as stratospheric sudden warming and monsoon onset. Most models have difficulties in simulating these events with fidelity. Through studying the models' failures, one can strive for improvements.

References

Arnold, V. I. 1984: *Catastrophe Theory*. (Translated by R. K. Thomas), Springer-Verlag, 79 pp.

Barnes, J. R., and J. L. Hollingsworth, 1987: Dynamic modeling of a planetary wave mechanism for a Martian Polar warming. *ICARUS*, **71**, 313-334.

Brown, R. A., 1970: A secondary flow model for the planetary boundary layer. *J. Atmos. Sci.*, **27**, 742-757.

Budydo, M. I., 1969: The effect of solar radiation variation on the climate of the earth. *Tellus*, **21**, 613-629.

Chao, W. C., 1985: Stratospheric sudden warmings as catastrophes. *J. Atmos. Sci.*, **42**, 1631-1646.

Chao, W. C., 2000: Multiple quasi-equilibria of the ITCZ and the origin of monsoon onset. *J. Atmos. Sci.*, **57**, 641-651.

Chao, W. C., and B. Chen, 2001: Multiple quasi-equilibria of the ITCZ and the origin of monsoon onset. Part II. Rotational ITCZ attractors. *J. Atmos. Sci.*, **58**, 2820-2831.

Chao, W. C., and B. Chen, 2001: The origin of monsoons. *J. Atmos. Sci.*, **58**, 3497-3507.

Chao, W. C., B. Chen, and W.-K. Tao, 2003: Tropical cyclogenesis as a catastrophe. *AMS Conf. Atmos. Ocean Fluid Dynamics*, San Antonio, Preprint Vol., Paper 11.

Chao, W. C., and B. Chen, 2004: Single and double ITCZ in an aqua-planet model with constant sea surface temperature and solar angle. *Climate Dyn*., **22**, 447-459.

Chao, W. C., and L. Deng, 1998: Tropical intraseasonal oscillation, super cloud clusters and cumulus convection schemes. Part II. 3D Aqua-planet simulations. *J. Atmos. Sci*., **55**, 690-709.

Charney, J. G., 1971: Tropical cyclogenesis and the formation of the ITCZ. *Mathematical Problems of Geophysical Fluid Dynamics.* W. H. Reid, Ed., Lectures in Applied Mathematics, Vol. 13, *Amer. Math. Soc*., 355-368.

Charney, J. G. and J. G. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci*., **36**, 1205-1216.

Charney, J. G. and A. Eliassen, 1964: On the growth of the hurricane depression. *J. Atmos. Sci*., **21**, 68-75.

Church, C., 1993: Ed. *The Tornado: Its Structure, Dynamics Prediction and Hazards*. Amer. Geophys. Union, pp 637.

Drazin, P. G., 1992: *Nonlinear System*. Cambridge U. Press. 317 pp.

Fernandez, W., 1997: Martian dust storms: A review. *Earth, Moon and Planets*, **77** 19-46.

Hartmann, D. L, 1994: *Global Physical Climatology*, Academic Press. 409 pp.

Held, I. M., 1983: Stationary and quasi-stationary eddies in the extratropical troposphere: theory. Chap. 6 of *Large-scale Dynamical Processes*. Edited by B., J. Hoskins and R. P. Pearce, Academic Press, 397pp.

Holton, J. R. 2004: *An Introduction to Dynamic Meteorology*, 4th ed. Academic Press, 531 pp.

Hoskins, B. J., M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Q. J. R. M. S.*, **111**, 877-946.

Iooss, G. and D. D. Joseph, 1980: *Elementary Stability and Bifurcation Theory*. Springer-Verlag, 286 pp.

Kirtman. B. P., and E. K. Schneider, 2000: A spontaneously generated tropical atmospheric general circulation**.** *J. Atmos. Sci.,* **57**, 2080–2093.

Lorenz, E. N., 1967: *The Nature and Theory of the General Circulation of the Atmosphere*. WMO, Geneva.

Lorenz, E. 1993: *The Essence of Chaos*. U. Wash. Press.

Mak, M., 2001: Non-hydrostatic barotropic instability: Applicability to nonsupercell tornadogenesis. *J. Atmos. Sci.,* **58**, 1965–1977.

Ooyama, K., 1864: A dynamical model for the study of tropical cyclone development. *Geofis. Int*., **4**, 187-198.

Palmen, E. and C. W. Newton, 1969: *Atmospheric Circulation Systems*. Academic Press, 603 pp.

Philander, S.G.P., D. Gu, D. G. Halpern, G. Lambert, N. C. Lau, T. Li, R. C. Pacanowski, 1996: Why the ITCZ is mostly north of the equator. *J. Climate*, **9**, 2958-2972.

Poston, T, and I. N. Stewart, 1978: *Catastrophe Theory and Its Applications*. London, Pitman, 491 pp.

Randall, D. A. and M. J. Suarez, 1984: On the dynamics of stratocumulus formation and dissipation. *J. Atmos. Sci*., **41**, 3052-3057.

Rotunno, R., and K. A. Emanuel, 1987: An air-sea interaction theory for tropical cyclones. Part II: Evolutionary study using a nonhydrostatic axisymmetric numerical model. *J. Atmos. Sci*., **44**, 542-561.

Sellers, W. D., 1969: A global climate model based on the energy balance of the earth-atmosphere system. *J. Appl. Meteor*., **8**, 392-400.

Sparrow, C., 1982: The *Lorenz Equations: Bifurcations, Chaos, and Strange Attractors* Applied Math. Sci., **41**, Springer-Verlag, pp 269.

Sturrock, P. A., 1966: Explosive and non-explosive onsets of instability. *Phys. Rev., Lett.* **16**, 270-273.

Sumi, A., 1992: Pattern formation of convective activity over the aqua-planet with globally uniform sea surface temperature. *J. Meteor. Soc. Japan,* **70**, 855-876.

Thom, R., 1972: *Structural stability and morphogenesis*. (Translated by D. H. Fowler 1975) Benjamin, Reading, Mass.

Xie, S. P. and K. Saito, 2001: Formation and variability of a northerly ITCZ in a hybrid coupled AGCM: Continental forcing and oceanic-atmospheric feedback. *J Climate,* **14**, 1262-1276.

Yu, L. A, and M. M. Rienecker, 1998: Evidence of an extratropical atmospheric influence during the onset of the 1997-98 El Nino. *Geoph. Res. Let*., **25** (18), 3537-3540.