Hurricane Forecasting: The State of the Art

H. E. Willoughby¹; E. N. Rappaport²; and F. D. Marks³

Abstract: In this article, we summarize current forecasting practice, the performance of the forecasting enterprise, and the impacts of tropical cyclones from a meteorological perspective. In the past, a forecast was considered successful if it predicted the hurricane’s position and intensity 12–72 h into the future. By the 1990s, forecast users came to expect more specific details such as spatial distributions of rainfall, winds, flooding, and high seas. In the early 21st century, forecasters extended their time horizons to 120 h. Meteorologists have maintained, homogeneous statistics on forecast accuracy for more than 50 years. These verification statistics are reliable metrics of meteorological performance. In terms of outcomes, forecasting in the late 20th century prevented 66–90% of the hurricane-related deaths in the United States that would have resulted from techniques used in the 1950s, but it is difficult to demonstrate an effect on property damage. The economic and human consequences of the response to forecasts and warnings are also poorly known. A final key concern is how to frame forecasts to address users’ needs and to elicit optimum responses.

DOI: 10.1061/(ASCE)1527-6988(2007)8:3(45)

CE Database subject headings: Hurricanes; Cyclones; Tropical regions; Weather data; Weather forecasting.

Introduction

Hurricanes are circular cyclonic storms that draw their energy from the warm tropical sea. They are distinct from middle-latitude cyclones which depend on the tropics-to-pole horizontal temperature gradient for their energy. Hurricanes are 500–1,000 km in horizontal extent, much smaller than middle latitude cyclones. The hurricane vortex generally fills the depth of the troposphere. Its effects may reach from the upper 200 m of the sea into the lower stratosphere. Hurricanes are warm core in the sense that the air near the center is warmer than the surrounding atmosphere. Because warm air is less dense than cold, this property causes their low hydrostatic central pressures, which can be as much as 10% below that in the normal tropical atmosphere. Their sustained winds can reach 165 kt ("kt" denotes a nautical mile per hour or 1.15 statute miles per hour; 165 kt=190 mi/hr =85 m s⁻¹).

To qualify as a hurricane, a tropical vortex must be over the Atlantic or northeastern Pacific east of the International Date Line and have winds of at least 64 kt (33 m s⁻¹). Weaker cyclones with winds from 34 to 63 kt (from 17 to 32 m s⁻¹) are called tropical storms. Still weaker systems are called tropical depressions if the wind undulates in a wavelike pattern that does not close back on itself. In the northwestern Pacific, meteorologists call cyclones of hurricane intensity typhoons. In the Southern Hemisphere and throughout the Indian Ocean, they are simply called cyclones. Tropical cyclone (TC) is the generic term for warm-core tropical systems with closed surface circulations.

The eye, which encloses the geometric center of the TC vortex, is characteristically tens of kilometers in radius. It generally forms as the TC crosses the threshold of hurricane intensity. The eye is often clear from a kilometer or two above the surface to the tropopause, particularly in the most intense hurricanes. Within the eye, there is invariably a stagnation point where the winds are calm, but the clear eye is not filled with calm winds, literary metaphors notwithstanding. Around the eye is a ring of convective clouds, called the eye wall or wall cloud, where water vapor drawn from the sea under the strong-wind part of the vortex condenses, transforming heat absorbed during evaporation into temperature increases that fuel the storm. The inner edge of the eye wall is also the site of the hurricane’s strongest winds. In this context, it is important to recall that hurricanes are both more intense and more compact than their middle-latitude relatives. Damaging winds are often confined within 100 km of their centers.

Unlike other relatively small atmospheric wind systems, hurricanes last a long time—from days to weeks. Normally (though by no means always), at least 12 h are required for a hurricane to change intensity or motion appreciably. Because of their circular geometry, spiral cloud bands, and evident rotation, hurricanes present dramatic pictures in animated satellite imagery. Tropical cyclones’ appearance and their natural time scale (which matches the morning and evening news cycle of broadcast journalism) cause them to be made-for-TV natural hazards.

During a typical Atlantic Hurricane Season (June–November), 10 or 11 tropical storms form, of which about 6 become hurricanes. Any individual year may vary greatly from the average. For example, single seasons have produced as few as 2 (1982) and as many as 15 (2005) hurricanes. Most Atlantic hurricanes begin as African, or easterly, waves in the unstable airflow between the deserts and grasslands of North Africa. Each year, this

¹Distinguished Research Professor, Dept. of Earth Sciences, Florida International Univ., PC 344, Miami, FL 33199. E-mail: hugh.willoughby@fiu.edu

²Deputy Director, National Hurricane Center, NOAA/NWS, 11691 SW 17th St., Miami, FL 33165. E-mail: edward.n.rappaport@noaa.gov

³Director, Hurricane Research Division, AOML/NOAA, 4301 Rickenbacker Cswy., Miami, FL 33149. E-mail: frank.marks@noaa.gov

Note. Discussion open until January 1, 2008. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on August 1, 2006; approved on November 29, 2006. This paper is part of the Natural Hazards Review, Vol. 8, No. 3, August 1, 2007. ©ASCE, ISSN 1527-6988/2007/3-45–49/$25.00.
Most victims drowned in wind-driven flooding. From 1970 through 2000, however, most hurricane-related deaths in the United States were caused by drowning in floods caused by hurricane rainfall (Rappaport 2000). This change came about because timely evacuations largely prevented drowning in the surge zone. The risk of freshwater drowning, however, has remained constant or increased slightly.

Between 1970 and 2004, the annual average number of deaths from all causes attributed to hurricanes was in the low 20s. If the forecasting and warning enterprise functioned as it did in 1950, but with 21st century population levels, the United States would lose >200 souls annually (Willoughby 2002). Overall, the risk to a U.S. resident of dying in a hurricane seemed to decrease by two orders of magnitude during the 20th century. Using an accepted value of statistical life of $4 \times 10^6$ (Blomquist 2004), the economic impact of these prevented deaths approaches $1 \times 10^9$.

Hurricane forecasting at the current state of the art, then, reduced mortality by about 90% during this interval.

Before Katrina, the American Red Cross estimated that 25,000–100,000 people could drown in a worst-case hurricane strike on New Orleans. Extension of the statistics quoted in the previous paragraph to include the mortality that actually occurred in 2005 (~1,500) increases the average annual toll to about 60. This means that modern forecasting can now claim to prevent only about two-thirds of hurricane deaths. It is ironic that the 2005 statistics degrade this metric even though the accurate forecast of the landfall 60 hours in advance saved thousands, if not tens of thousands, of lives. Despite the reemergence of storm surge as a cause of hurricane deaths, drowning in freshwater floods remains a challenge that requires both more accurate precipitation and flooding forecasts. It is clear that meteorological forecasts are one link in a chain that includes education and effective communication.

Since 1900, there were 20 years in which TCs killed no U.S. residents—15 years before 1950 and 5 since. Only three hurricane seasons, 1900, 1928, and 2005, claimed more than 1,000 lives, and 16 claimed more than 100 lives. Before Katrina, the most recent of these was in 1972, when Hurricane Agnes made landfall on the Florida Panhandle and affected the Eastern Seaboard from Virginia to Long Island. Ten of the seasons that killed more than 100 U.S. residents occurred before 1950 and 6 since 1950 (Jarrell et al. 2001). Thus, although increasing population makes it less likely that the nation will escape hurricane mortality completely in any given year, better forecasting has reduced the probability of substantial mortality. Still, in light of the disastrous 2005 season, the annual probability of a large loss of life in a hurricane (i.e., >1,000 deaths) may be as great as 1–2%.

Hurricane property damage has increased in step with increasing personal wealth, inflation of the currency, and exploding coastal development. Correction for these economic factors to calculate normalized damages yields estimates of the costs of historical hurricanes if they had occurred with present-day economic and demographic conditions. Once the economic factors are accounted for, property losses appear to be essentially constant at $6 \times 10^9$ annually ($5 \times 10^9$ before 2005) with no discernable trend resulting from better forecasts or more effective damage mitigation measures. Major hurricanes with winds of 96 kt (>50 m s$^{-1}$) compose 20% of U.S. TC landfalls, but cause 80% of property damage (Pielke and Landsea 1998). The average is less representative than we might hope. In some years—most recently 1982—zero damage has occurred. Even before 2005, it was widely recognized that a reoccurrence of the Miami Hurricane of 1926, or equivalently landfall by a Category 4 or 5 hur-
hurricane passing over the center of a 21st-century metropolis, could demolish as much as $100 \times 10^6$ in property.

A frequently cited value for the cost of preparing for a hurricane strike is $1 \times 10^6 / \text{mi}$ of coastline warned, but this number has a dubious provenance. It began as an estimate of a few tens of thousands of dollars per mile for evacuation costs and has been updated haphazardly for inflation and population increase. Despite the questionable basis of this number, the actual cost of preparations for landfall is clearly more than $0.1 \times 10^6 / \text{mi}$ (except for the most sparsely inhabited shores) and less than $10 \times 10^6 / \text{mi}$ (except for metropolitan centers). Typically, NHC warns 300–400 mi of coastline to provide a margin of safety around the affected area. A reasonable estimate for the economic cost, then, is a multiple of $100 \times 10^6$ per landfall. Preparation costs can total $>1 \times 10^8$ in an active year with many landfalls. In this context, a vital research objective is accurate accounting both for locale-specific costs due to evacuation, damage mitigation, and lost business, and for benefits in terms of lives saved and property losses prevented.

Of the 300 or so miles of coastline warned, 100 mi represents the average diameter of hurricane-force winds. The remaining 200 mi, roughly 100 on either side of the landfall point, are a conservative margin of error that takes into account uncertainty in the track forecast. The best-case reduction in warning area from more accurate track forecasts would be a few tens of miles. Anything more would probably increase the losses lost so that there would be marginal gain, or even a net loss, if NHC were to reduce the warning area too drastically. A reduction of 25 mi on each side of the landfall—representing a great increase in forecast accuracy and confidence—would save $50 \times 10^6 / \text{event}$. This value, however, is an exaggeration because the cost per mile is probably much more near the center of the warning area than near the edge. Although $50 \times 10^6$ seems like a great deal of money, it represents only $<1\%$ of the $6 \times 10^9$ annual average hurricane damage, or the economic value of $\sim 12$ deaths (although moral grounds make framing it in those terms repugnant).

The cost of the forecasting enterprise itself, including such items as NHC operations, reconnaissance aircraft, pro-rata shares of satellites and local forecast offices, and research, appears to be less than $0.1 \times 10^9$ (Willoughby 2000). An approximate balance sheet for forecasting and impacts shows a well-defined savings of $500$–$1,000 \times 10^6$ in economic impacts of hurricane-related deaths, offset by $0.1 \times 10^9$ spent on observations, forecasting, and research, and several hundred million spent on evacuation and other mitigation measures. It is difficult to document a positive impact of forecasts on property damage, although one might reasonably argue that evacuation should work as well for expensive mobile property (ships, boats, aircraft, and high-profile motor vehicles) as it does for people.

**Hurricane Forecasts**

Because TCs are compact, long-lived weather systems, forecasts of their positions and intensities—measured in terms of maximum wind—are the first steps toward characterizing the threat. Marks et al. (1998) present an authoritative, though now somewhat dated, plan for TC forecasting and related research. The highest priority has historically been predicting the cyclones’ future paths, which is called *track forecasting*. The track worries everyone in the threatened area, whereas intensity is of overwhelming concern only to those directly in the cyclone’s path. Track forecast error, the great-circle distance between the forecast and the actual verification TC position, is an objective and stable measure of track forecast quality. Forecasters in the North Atlantic Basin have maintained track forecast error statistics (Fig. 1) for the 24-h time horizon since 1954, for 48 h since 1961, and for 72 h since 1964. In response to customer-driven requirements, NHC began issuing forecasts for 96 and 120 h in 2003. Throughout the time for which verification statistics exist, forecast errors have decreased by 1–2% per year for all time horizons, with the most rapid improvement at longer durations. This steady improvement has resulted from more available observations, better assimilation of the data into models at numerical forecasting centers, incorporation of more representative physics into the models themselves, and a forecasting enterprise that aggressively translates scientific advances into new techniques. Despite steadily improving accuracy, warning areas increased during the late 20th century because emergency managers wanted to ensure that nobody was struck without warning, as well as to gain more lead time for evacuating ever-increasing coastal populations. Since 2000, the size of warning areas has decreased somewhat in response to more accurate forecasts.

Output from computer models, called *guidance*, is the primary tool for track forecasting (see, e.g., DeMaria and Gross 2003). Statistical extrapolations based on numerical predictions of global wind and pressure patterns, such as the NHC90 model (McAdie 1991), once resulted in the most accurate guidance. These statistical–dynamical schemes superseded earlier, purely statistical models based on observed weather patterns at the initial time. In the modern era, forecasters rely on purely dynamical models that integrate the Navier–Stokes equations for atmospheric motions to represent both the storm itself and its surroundings. Numerical models are structured on a computational grid with temperature, moisture, and wind tabulated in more-or-less rectangular cells. The models calculate the spatial derivatives that appear in the equations by finite differences or spectrally and extrapolate the computed time derivatives forward to predict the future weather elements on the grid. Examples of dynamical models are the Geophysical Fluid Dynamics Laboratory (GFDL) model, the United Kingdom Meteorological Office model, and the Navy Operational Global Atmospheric Prediction System. As computers become faster with larger memories, these models are able to use finer spatial resolution and more elaborate representa-
tion of physical processes to attain increasing accuracy.

All numerical predictions start with an “initial condition,” which depicts the current state of the atmosphere at the moment the calculation starts. In meteorology, the term synoptic denotes simultaneous observations taken globally, or at least over a large area. Synoptic observations by rawinsondes (balloon-borne instrument packages tracked from the ground) are a vital source of the wind, temperature, and humidity data required to define the initial condition. Since the end of World War II, observers have launched rawinsonde observation worldwide at 00 and 12 UTC (coordinated universal time). High-threat situations, such as impending TC landfall, may dictate observations at 6-h or even 3-h intervals. Interpolating rawinsonde observations to the model’s mesh and starting the calculation at a synoptic time is the simplest way to prepare a model initial condition. A key limitation to this approach is the lack of rawinsonde observations over the sea, apart from scattered islands and very few ships. Special targeted aircraft observations over the sea during the 24- to 48-h interval before landfall can reduce forecast errors by 25–30% (Aberson 2003).

Satellites can cover the entire globe, but most of their observations fall outside a window of a few hours around synoptic times. Spaceborne sensors can retrieve thermodynamic data and winds over the sea. Although the presence of local meteorological elements such as heavy rainfall or dense clouds can compromise retrievals from some sensors, data from this source have been crucial to better forecasts. Because simple interpolation in space and time is a less-than-optimum way to use observations, nonsynoptic data require elaborate four-dimensional variational data assimilation schemes. More and better observations over the sea, more powerful data assimilation tools, and more sophisticated models with finer resolution were the forces that drove the reduction in forecast errors shown in Fig. 1.

Intensity is the other aspect of the first-order characterization of TCs. Here, techniques lag about a generation behind the state of the art for track forecasting. The best intensity guidance is still the Statistical Hurricane Intensity Prediction System (SHIPS), although the same GFDL model used for track shows promise. In many situations, simple extrapolation of past trends augmented by SHIPS provides an adequate forecast. But the most threatening situations occur when a hurricane undergoes rapid intensification, as did Hurricane Opal of 1995 (Fig. 2). This process may increase maximum winds by 39 kt (>20 m s⁻¹) in 12–24 h. Rapid intensification is dangerous because the phenomenon produces nearly all the “major” (Categories 3–5) hurricanes that cause the greatest devastation and also because rapid intensification can happen too quickly for the normal forecast cycle, as in Hurricane Charley of 2004. Currently the best rapid intensification forecasting model is a statistical scheme based upon SHIPS. The tropical cyclone forecasting community recognizes the development of improved rapid intensification forecasting as a vital research objective, and it is at the top of TPC’s research agenda.

A numerical model called SLOSH (sea, lake and overland surges from hurricanes) is the basis of predicting the maximum extent of storm surge flooding. It uses a probabilistic representation of the hurricane vortex to drive storm surge over accurately depicted local topography and bathymetry. By running SLOSH for a wide range of possible tracks and intensities, forecasters produce a maximum envelope of water (MEOW) for each locale. Emergency managers use the MEOWs to delimit evacuation zones in terms of readily recognized geographical features. The strategy responsible for the reduction in U.S. hurricane mortality is timely evacuation of the surge zone, as computed by SLOSH, combined with increasingly accurate track forecasts. As the “clearance times” required for complete evacuation increase in step with growing coastal populations, the question arises whether this strategy will continue to work so well.

Quantitative forecasting of surface winds, sea and swell, rain-fall, and other local impacts has lagged behind track and storm surge. From the user’s perspective, weather is like politics—all local. Users want to know what will happen where they are. For example, when gale-force winds reach the shoreline, evacuation of barrier islands, shuttering of windows, and other mitigation efforts that require work in exposed locations must end.

Improved numerical models, observations, and data assimilation schemes will be the keys to realizing “neighborhood-level” forecasts. The meteorological research community is currently beta-testing the new weather research and forecasting (WRF) model, which will be the essential tool for attaining this goal. WRF is an open-source-code program that will be run in different forms for different purposes in a variety of settings on a gamut of computer architectures. This philosophy means that academic or government researchers can develop code modules on desktop computers and transfer them directly to the operational forecast models with limited recoding. The TC forecasting community already has rigorous protocols for model evaluation and a tradition of (generally) good-natured competition to improve forecasts.

In the interim as WRF becomes operational, areas of active research are the coupling of existing atmospheric models to ocean models to simulate storm-induced changes in the oceanic energy source, better representation of air–sea exchange processes, finer grid resolution, and improved parameterization of physical effects in general. Types and formats of forecast products are also evolving. The NWS now devotes considerable effort to probability-based forecasts and other means of indicating uncertainty. Optimum employment of probabilistic forecasts presents significant challenges in communication and interpretation for all users from professional managers to households and individuals.

Summary

Tropical cyclone forecasting is a successful enterprise with demonstrably favorable benefit-to-cost returns. Track forecasting accuracy is improving steadily, but intensity forecasting and prediction of local wind, rainfall, and sea state remain problematic.
Because more accurate tropical cyclone forecasts pose significant operational, technological, and scientific challenges, we should expect progress to be incremental, although cumulative.

In the aftermath of Katrina, forecasters are asking a number of key social science questions: What is the probability of high levels of hurricane-related mortality? What costs do warnings and evacuations impose on society? Can we demonstrate that forecasts reduce property damage and prevent loss of life? What is the optimum trade-off between warning lead time and size of the warning area for a particular forecast accuracy and degree of coastal development? What weather elements should we be forecasting? How should we communicate forecasts so that they are most valuable to users? What mix or balance of improvements in meteorology or societal and economic preparedness, including long-term policies such as building codes and land usage, will yield the greatest benefits? What long-term plans should the federal government, the private sector, and the public develop to realize these benefits, and what should be the individual and collaborative roles of these stakeholders?

Acknowledgments

The first writer’s contribution to this paper was supported by NSF Grant No. ATM-0454501. The writers are grateful to Naomi Surgi for her comments on an earlier draft.

References


