3A.3 EVALUATING TROPICAL CYCLOGENESIS FORECASTS FROM FOUR GLOBAL NUMERICAL MODELS

Daniel J. Halperin[1], Henry E. Fuelberg[1], Robert. E. Hart[1], Philip Sura[1],
Joshua Cossuth[1], Ryan Truchelut[1], Richard J. Pasch[2]
[1]Florida State University, Tallahassee, FL
[2]NOAA/NWS/NCEP/National Hurricane Center, Miami, FL

1. INTRODUCTION

Forecasts of tropical cyclones (TCs) rely greatly on output from numerical models. Each member in the suite of models used by forecasters at the National Hurricane Center (NHC) has its particular strengths and weaknesses. Some research has investigated the skill of the various models with respect to track and intensity, with the assumption that a TC already exists (e.g., Goerss 2000, Goerss et al. 2004, Sampson et al. 2008). However, little research has focused on the skill of model-derived forecasts of TC genesis (Cheung and Elsberry 2002). In some cases, a global numerical model will accurately predict tropical cyclogenesis. However, observations indicate that these models sometimes generate a TC that does not develop (Beven 1999). Conversely, a TC sometimes develops that has not been forecast. At other times, a model forecasts TC genesis for a storm that ultimately develops, but the timing of genesis is incorrect. The goal of this research is to quantify the accuracy of model-derived TC genesis forecasts. We compare results from four global models: Environment Canada Global Environmental Multiscale Model (CMC; Côté et al. 1998a, 1998b), the National Centers for Environment Prediction (NCEP) Global Forecast System (GFS; Kanamitsu 1989), the Navy Operational Global Atmospheric Prediction System (NOGAPS; Rosmond 1992), and the United Kingdom Meteorological Office global model (UKMET; Cullen 1993). We are currently working to add the European Centre for Medium-Range Weather Forecasts model (ECMWF) to our analysis.

Climatologically, we know the necessary conditions for TC genesis (Gray 1968, Tory and Frank 2010). But, a complete physical explanation for the specific processes that cause TC genesis still is largely unknown. This is evident by the number of TC genesis theories that have been proposed, and the disagreement that exists among them. Several field experiments (e.g., PRE-Depression Investigation of Cloud-systems in the Tropics (PREDICT; Montgomery et al. 2012), Genesis and Rapid Intensification Processes (GRIP), Intensity Forecasting Experiment (IFEX; Rogers et al. 2006)) have been conducted recently to evaluate these theories.

Corresponding author address: Daniel J. Halperin, Florida State University, Dept. of Earth, Ocean, and Atmospheric Science, Tallahassee, FL, 32306-4520; e-mail: djh09@fsu.edu

The operational models are not bound by the TC genesis theories or the conditions stated to be necessary for TC development. There is no "checklist" that a model must satisfy to forecast TC genesis. Additionally, due to grid spacing and computational limitations, the models are not able to resolve all of the atmospheric processes governed by the full Navier-Stokes equations. For example, convection, planetary boundary layer processes, and microphysical processes are all parameterized. Therefore, one cannot expect the models to fully resolve all of the processes that are necessary for TC genesis. Nonetheless, the models commonly indicate TCs developing in the forecast fields.

So, how well do the models predict TC genesis, and are their predictions based on the scientifically accepted conditions for genesis? While there have been a few previous studies regarding model-derived TC genesis forecasts (e.g., Briegel and Frank 1997; Chan and Kwok 1999; Cheung and Elsberry 2002; Pasch et al. 2006, 2008; Snyder et al. 2010), most resources have been dedicated to improving model-derived TC track and intensity forecasts.

The goal of this research is to quantify the accuracy of global models' forecasts of TC genesis over a period of several hurricane seasons. The results will reveal which model performs best during a given season (performance varies from year to year) and whether a given model improved its TC genesis forecasts over time. It is important not only to analyze successful forecasts but also false alarm cases and cases when a TC formed in reality but was missed by the global models. The results are sub-divided geographically and temporally to determine where and when the models perform well and poorly.

2. METHODOLOGY

A successful evaluation of TC genesis forecasts must be systematic and objective. One of the most important components of this research was to define a model tropical cyclone. Given the number of processes that are parameterized in the models, it is unreasonable to expect that a model-indicated tropical depression (TD) will exhibit the same characteristics as a TD in reality. Since one of our primary goals was to provide operational forecasters with information about predicting tropical cyclogenesis, we conducted a series of experiments with different sets of criteria and determined their hit, miss, and false alarm rates during selected hurricane seasons. The set of criteria given below produced the greatest skill and was used to define a tropical cyclone in the models:

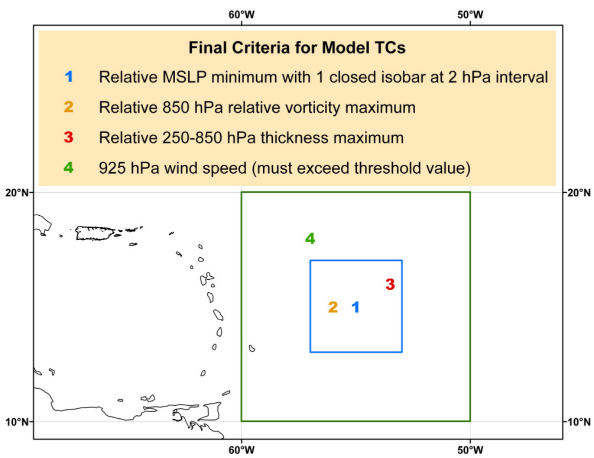

Fig. 1. Criteria used to define a model TC.

1. A relative minimum in MSLP with at least one closed isobar at a 2 hPa interval must exist,
2. A relative maximum in 850 hPa relative vorticity must be located within ± 2° latitude and longitude of the MSLP minimum. This maximum must exceed the lowest tercile value of all model-indicated TD analyses that correspond to Best-Track TDs,
3. A relative maximum in 250-850 hPa thickness must occur within ± 2° latitude and longitude of the MSLP minimum. This maximum must exceed the lowest tercile value of all model-indicated TD analyses that correspond to Best-Track TDs,
4. The wind speed at 925 hPa must exceed the model specific wind threshold at any point within ± 5° of the MSLP minimum, and
5. Criteria 1-4 must be met for at least 24 h.

Criteria 1-3 are similar to the TC definition of Chueng and Elsberry (2002), where criterion 3 is the proxy for a warm core. Criterion 4 defines an objective wind threshold, and criterion 5 addresses the temporal threshold suggested by Walsh et al. (2007). The requirement of at least one closed isobar at a 2 hPa interval removed many weak relative MSLP minima that were equatorward of 10°N, in or near the Inter Tropical Convergence Zone (ITCZ). It also removed relative MSLP minima that were not cyclones, but merely broad areas of relatively low pressure between two anticyclones. The requirement that the 250-850 hPa thickness maximum exceed the lowest tercile threshold removed many cyclones in the mid latitudes (i.e., 30-45° N) that had small thickness values and likely were not purely tropical.

Once a model-indicated TC was identified, it was tracked and classified as one of the following events:

Hit: A model-indicated TC exhibits genesis within ± 24 h of the Best-Track indicated genesis time and is located within ± 5° latitude and longitude of the Best-Track indicated genesis location.

False alarm (FA): TC genesis is forecast in the model, but does not occur within ± 24 h of a Best-Track time and is not located within ± 5° latitude and longitude of any Best-Track location. These cases are considered to be spurious vortices.

Incorrect timing (IT): TC genesis is forecast in the model, and it occurs at the same time as an existing Best-Track position, and is within ± 5° latitude and longitude of the Best-Track location; however, it is not the genesis position of a TC. Instead, it is the Best-Track position of a TC that already has formed. These cases are considered to be forecasts of actual TCs, but at an incorrect genesis time (see Fig. 2).

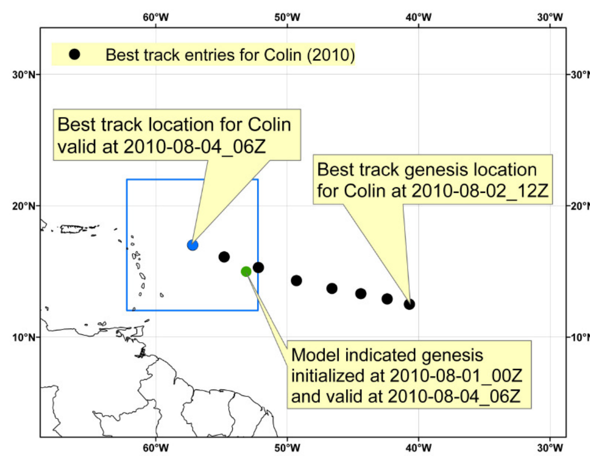

Fig. 2. Example of an incorrect timing (IT) event for Colin (2010). The model-indicated genesis time matches a Best-Track entry time, and the location of model-indicated genesis is within the 5° buffer of the Best-Track location. But, the verifying Best-Track position is a post-genesis position. Thus, it is a forecast of genesis at the wrong time.

One also must consider the scenario when a TC in the Best-Track data is not predicted by the models (miss events). Thus, in addition to calculating conditional probabilities given model-indicated genesis (hit, false alarm, IT), we also calculate a set of unconditional probabilities that consider miss events.

3. RESULTS

The period of study (2004-2011) includes some of the most active North Atlantic hurricane seasons on record. Our dataset contained 135 tropical cyclogenesis events. Cyclogenesis locations are defined as the first time that the National Hurricane Center (NHC) designated the cyclone as purely tropical. Storms that were subtropical throughout their entire life cycle were not considered.

One set of statistics is conditioned on the forecast model predicting genesis. Each model-indicated TC is classified as a hit, false alarm, or an incorrect timing event. Three conditional probabilities are defined: $P(G_y \mid F_y)$, the probability of actual genesis given that the model indicates genesis (hit); $P(G_n \mid F_y)$, the probability of no actual genesis given that the model indicates genesis (false alarm); and, $P(G_t \mid F_y)$, the probability of actual genesis, but occurring at a different time given that the model indicates genesis (incorrect timing). The other set of statistics factor in the miss events. We investigate $P(G_e)$, the probability of a hit or IT in unconditional terms. Finally, we calculate a critical success index (CSI) which is modified to include the incorrect timing events (not shown).

3.1 *CMC Results*

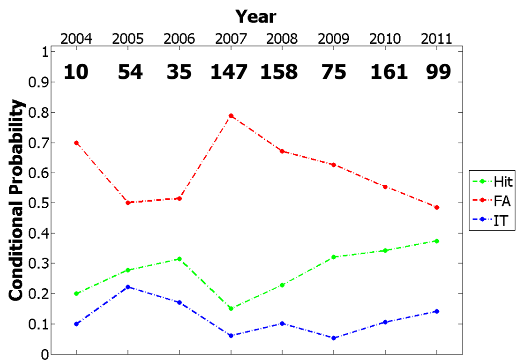

Fig. 3. Conditional probabilities of a hit (green), false alarm (red), or incorrect timing event (blue) given model-indicated TC genesis for the CMC from 2004 to 2011. Numbers under each year indicate the total number of model-indicated TC genesis events.

The CMC typically has been the most aggressive model in predicting TC genesis, especially since 2007. As a result, FAs have been an issue over the past several seasons, although steady improvements have been made. The conditional probability of a FA has decreased from near 80% in 2007 to just below 50% in 2011 (Fig. 3). Also of note is that the CMC was noticeably less aggressive in predicting TC genesis during 2011 than during 2007-2010 (not shown).

Hit and IT events depend on where TCs form, but false alarms are completely independent of Best-Track TC genesis locations. The CMC does not seem to have a preferred region of genesis—all three types of genesis events occur across the entire basin (not shown). Hits, FAs, and ITs are located across the main development region (MDR; 10-20° N, 80-20° W) and the southwestern Caribbean Sea. However, false alarms far outnumber hits over the Gulf of Mexico and the Atlantic Ocean poleward of ~25° N.

The locations of false alarms generally agree with the climatologically preferred areas of development. Most FAs that develop off the coast of Africa occur during August and September. However, during June, July, October, and November, most FAs occur closer to North America. Overall, the CMC predicts TC genesis almost anywhere in the Atlantic basin.

3.2 *GFS Results*

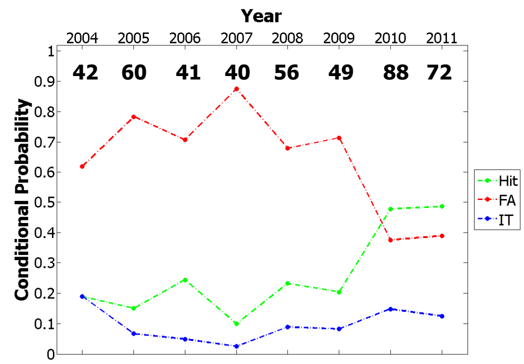

Fig. 4. As in Fig. 3, but for the GFS.

The GFS has markedly improved its TC genesis forecast performance since 2010 (Fig. 4). Much of this improvement may be attributable to a major upgrade to the model, including resolution, boundary layer scheme, and deep and shallow convection schemes. The results for 2010 and 2011 are nearly identical, which argues that the model upgrade had an impact and that the improvement in 2010 was not a "fluke."

The GFS exhibits preferred regions of TC genesis (not shown), unlike the uniformity observed with the CMC. There are clusters of false alarms off the west coast of Africa and over the southwestern Caribbean Sea. There are a large number of false alarms over the eastern Atlantic Ocean south of 10°N, which is particularly intriguing since only four TCs in the dataset actually developed south of that latitude. Also noteworthy is the lack of false alarms over the Gulf of

Mexico. Unfortunately, that area also exhibits a lack of hits and ITs, indicating that the GFS misses most of the storms that develop over the Gulf of Mexico. Conversely, the GFS predicts the genesis of African Easterly Waves (AEWs) relatively well, as indicated by a cluster of hits and ITs in the MDR.

The GFS exhibits some seasonal preferences. Most of the false alarms off the coast of Africa occur during August and September. Meanwhile, most of the false alarms over the western Caribbean Sea and near the Bahamas occur either earlier (June/July) or later (October/November). This agrees with climatology and suggests that the GFS follows the seasonal cycle.

3.3 *NOGAPS Results*

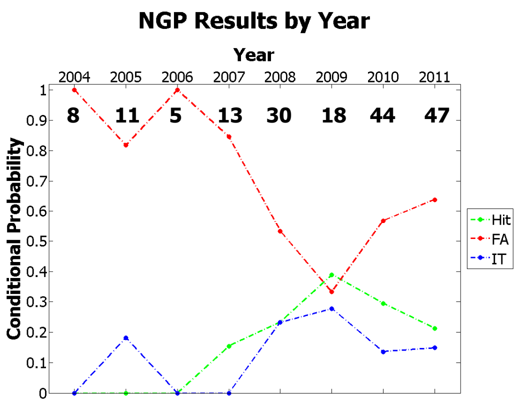

Fig. 5. As in Fig. 3, but for the NOGAPS.

The NOGAPS is the least aggressive model in terms of predicting TC genesis. Its performance seems to have peaked in 2009 when FA probabilities were near 30% (Fig. 5).

Most of NOGAPS' hit events occur west of 65° W, over the Caribbean Sea, near the Greater Antilles and near the southeast U.S. coast (not shown). There also is a small cluster of hits off the coast of Africa. The plot of false alarms reveals problems over the western Caribbean Sea, the southern Gulf of Mexico, and in the MDR. However, unlike the GFS, NOGAPS does not produce numerous false alarms south of 10° N.

NOGAPS exhibits seasonal preferences for TC genesis. During June and July there is only one false alarm in the entire dataset. Instead, the majority of the false alarms occur during August and September, coincident with the climatological peak in TC activity. These events occur throughout the basin. However, during October and November all but two false alarms occur west of 75° W, consistent with the climatologically favored regions of development during those months.

3.4 *UKMET Results*

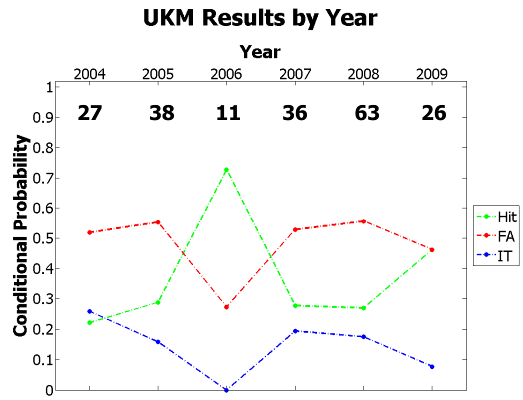

Fig 6. As in Fig. 3, but for the UKMET and only 2004-09.

With the exception of 2006, the UKMET was rather consistent in terms of TC genesis forecasting. We are seeking to procure the data for 2010 and 2011 and should have results from those seasons in the near future.

Hits, FAs, and ITs generally occur in the same regions; the majority of events occur over the MDR, and a few events are scattered over the Caribbean Sea and near the southeast U.S. The UKMET seems to follow the seasonal cycle. Very few events occur during June and July, while the majority of genesis cases occur during August and September over the MDR. Some events occur during October and November, mainly over the Caribbean Sea and Gulf of Mexico (not shown).

3.5 *Model Comparison*

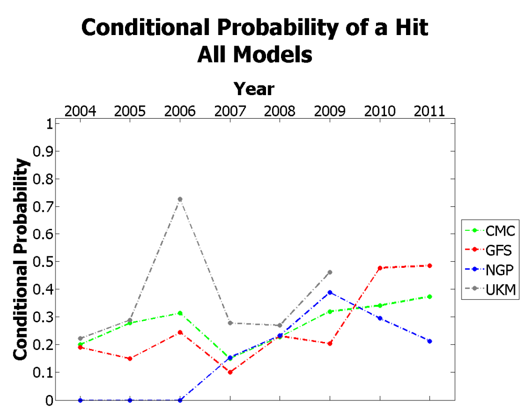

Fig. 7. Conditional probability of a hit for all models.

We plotted the conditional probability of a hit for each model and compared their relative performances

(Fig. 7). Although the UKMET outperforms its counterparts during the years when its data are available, results are not yet available for the years when the GFS noticeably improved. The CMC has shown steady improvement since 2007. The NOGAPS had its best performance in 2009.

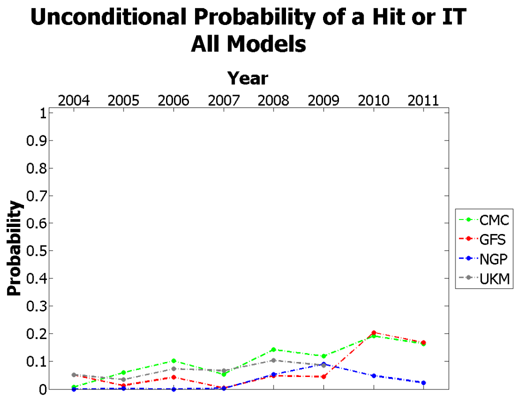

Fig. 8. Unconditional probability of a hit or IT for all models.

When miss events are considered, all of the models exhibit a noticeable decrease in performance (Fig. 8). The CMC performs relatively well, likely due to its being the most aggressive model in the study, and thereby having the fewest number of miss events. The improvement of the GFS that was noted in the conditional probabilities beginning with the 2010 season also is evident when analyzing the unconditional probabilities. The performance of the UKMET is relatively steady over time according to this metric. Again, the NOGAPS had its best performance in 2009.

3.6 *Consensus Results*

We also investigated whether better genesis forecasts occur when the same genesis event is predicted by multiple models. We documented all cases when two or more models at the same initialization time predicted genesis within 5° latitude and longitude of each other. These were considered the same genesis event. We then determined whether the genesis forecast was a hit, false alarm, or incorrect timing event. Based on the forecast location and time of genesis, the genesis event could be a hit in one model, but a FA or IT in another model. Several experiments were conducted with different combinations of models (not shown).

As an example, there were 117 instances during the period of study when the CMC and GFS both predicted the same genesis event. 25.6% of time, the forecast for both models verified as a FA (i.e., nothing ever developed). But, 74.4% of the time, a TC ultimately did develop (although the timing of genesis may have been incorrect).

Overall, if two or more models predict TC genesis, there is a relatively low probability that all of these models are predicting a false alarm. Thus, the results from the various consensus experiments are better than from any one model. A forecaster should have increased confidence in a TC genesis forecast when multiple models predict it—even if one of the models historically is aggressive.

4. SUMMARY AND CONCLUSIONS

Forecasts of TC genesis from four global models (CMC, GFS, NOGAPS, and UKMET) have been verified over eight seasons in the North Atlantic basin. Each model-indicated TC genesis event was identified, tracked, and classified as a hit, false alarm, or incorrect timing event. We also considered cases when a TC was not forecast in a model, but occurred in reality.

The results revealed several commonalities among the models. The probability of a hit, given model-indicated genesis, was greatest during 2011 for the CMC and GFS (data not available for the UKMET in 2011). However, we cannot definitively state whether this result is due to improvements in the models, environmental conditions (e.g., actual storms developing in a region where the models are more likely to predict genesis, such as the MDR), or simply coincidence. Analyzing future seasons will help answer that question. The models all appeared to generally follow the observed climatological cycle of the hurricane season. That is, relatively few genesis events occurred during June, July, October, and November. Those that did occur typically were located over the Caribbean Sea, Gulf of Mexico and near the southeast U.S. The majority of TC genesis events occurred during August and September over the MDR. Very few genesis events were forecast over the Gulf of Mexico, implying that the models typically missed TC genesis in that area. There also was a lack of genesis predictions over the eastern Caribbean Sea, consistent with the climatological scarcity of genesis in that region.

The models exhibited clear differences among themselves. The CMC is the most aggressive in forecasting TC genesis. As a result, it had the greatest number of hits, but also the greatest number of false alarms. The NOGAPS was least aggressive, under predicting TC genesis during all seasons.

Experiments with consensus forecasts revealed that false alarm rates were lower when multiple models predicted TC genesis. The consensus approach yielded better results than any one model alone. For example, when the CMC and GFS both predict the same genesis event, there is ~75% chance that genesis will occur (albeit possibly at the wrong time). This argues that forecasters should have greater confidence in TC genesis forecasts made by multiple models.

The ever-present caveat to this research is that the top performing model during one season may not be the top performer the next season. Therefore, it is impossible to predict which model will be the most reliable during an upcoming season. Nonetheless, our approach does indicate the models' past performance and whether they have been improving or degrading.

Future research will investigate why the models perform as they do. We will attempt to determine which changes or upgrades in the model yielded changes in model performance. We also will seek to determine whether particular synoptic situations are associated with relatively good or poor model performance. The ultimate goal of this research is to develop a genesis potential index to aid operational forecasters in predicting TC genesis.

5. REFERENCES

Beven, J.L., 1999: The Boguscane—A serious problem with the NCEP Medium Range Forecast model in the tropics. Preprints, *23rd Conference on Hurricanes and Tropical Meteorology,* Dallas, TX, Amer. Meteor. Soc., 845-848.

Briegel, L. M., and W. M. Frank, 1997: Large-scale influences on tropical cyclogenesis in the western North Pacific. *Mon. Wea. Rev.*, **125**, 1397–1413.

Chan, J., C. L. Roger, and H. F. Kwok, 1999: Tropical cyclone genesis in a global numerical weather prediction model. *Mon. Wea. Rev.*, **127**, 611–624.

Cheung, K. W., and R. L. Elsberry, 2002: Tropical cyclone formations over the western North Pacific in the Navy Operational Global Atmospheric Prediction System Forecasts. *Wea. Forecasting*, **17**, 800–820.

Côté, J., J. G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998a: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part II: Results. *Mon. Wea. Rev.*, **126,** 1397-1418.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998b: The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.*, **126,** 1373-1395.

Cullen, M. J. P., 1993: The Unified Forecast/Climate Model. *Meteor. Mag.*, **122,** 81-94.

Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187–1193.

Goerss, J. S., Charles R. Sampson, James M. Gross, 2004: A history of western North Pacific tropical cyclone track forecast skill. *Wea. Forecasting*, **19**, 633–638.

Gray, W. M., 1968: Global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.*, **96**, 669–700.

Kanamitsu, M., 1989: Description of the NMC global data assimilation and forecast system. *Wea. Forecasting*, **4,** 335–342.

Montgomery, M. T., and Coauthors, 2012: The pre-depression investigation of cloud systems in the tropics (PREDICT) experiment: Scientific basis, new analysis tools and some first results. *Bull. Amer. Meteor. Soc.*, **93**, 153–172.

Pasch, R. J., E. S. Blake, J. G. Jiing, M. M. Mainelli, D. P. Roberts, 2008: Performance of the GFS in predicting tropical cyclone genesis during 2007. *28th Conference on Hurricanes and Tropical Meteorology,* Amer. Meteor. Soc., Orlando, FL, 28 April-2 May 2008, 11A.7.

Pasch, R. J., P. A. Harr, L. A. Avila, J. G. Jiing, G. Elliot, 2006: An evaluation and comparison of predictions of tropical cyclogenesis by three global forecast models. *27th Conference on Hurricanes and Tropical Meteorology,* Amer. Meteor. Soc., Monterey, CA, 24-28 April 2006, 14B.5.

Rogers, R., and Coauthors, 2006: The Intensity Forecasting Experiment: A NOAA multiyear field program for improving tropical cyclone intensity forecasts. *Bull. Amer. Meteor. Soc.*, **87**, 1523–1537.

Sampson, C. R., J. L. Franklin, J. A. Knaff, M. DeMaria, 2008: Experiments with a Simple Tropical Cyclone Intensity Consensus. *Wea. Forecasting*, **23**, 304–312.

Snyder, A. D., Z. Pu, Y. Zhu, 2010: Tracking and verification of East Atlantic tropical cyclone genesis in the NCEP Global Ensemble: Case Studies during the NASA African Monsoon Multidisciplinary Analyses. *Wea. Forecasting*, **25**, 1397–1411.

Tory, K. J., and W. M. Frank, 2010: Tropical Cyclone Formation. *Global Perspectives on Tropical Cyclones: From Science to Mitigation.* J.C.L. Chan and J.D. Kepert, Eds., World Scientific, 55-91.

Walsh, K. J. E., M. Fiorino, C. W. Landsea, and K. L. McInnes, 2007: Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *J. Climate*, **20**, 2307–2314. doi: 10.1175/JCLI4074.1