P2.2 Effect of cloud processes on hurricane tracks: results from operational forecasts

Robert G. Fovell[1]* and Donald J. Boucher[2]
[1]University of California, Los Angeles
[2]Aerospace Corporation

1. Introduction

Ensemble spread is a useful proxy for forecast uncertainty. The National Hurricane Center (NHC) makes use of a multi-model ensemble with more than a dozen models of various complexity that typically yields a substantial variation in hurricane tracks. In the case of Hurricane Rita (2005), Fovell and Su (2007) demonstrated that a comparable track spread could be obtained from a single model, the WRF-ARW, via manipulation of cloud microphysics and cumulus schemes. Their simulations employed "operational resolutions" of 30 and 12 km initialized using NCEP GFS gridded forecast fields. Substantial and realistic track variations developed even in relatively short (roughly 2 day) model integrations.

Recently, Fovell, Corbosiero and Kuo (2009) used a specially modified version of the real-data WRF-ARW to examine how and why cloud microphysical assumptions influence hurricane motion. This higher resolution, idealized experiment employed a large land-free domain with uniform and fixed sea-surface temperature initialized with a calm, horizontally homogeneous environment following Jordan (1958). They showed that microphysics directly and indirectly modulates the tangential wind strength in the outer portion of the model tropical cyclones, well beyond the core. These winds are known to influence vortex self-propagation owing to the "beta drift" (Fiorino and Elsberry 1989). In the absence of strong environmental currents, the beta drift alone modulated by microphysics could produce substantial differences in cyclone position in less than 2 days.

In the present study, we employed WRF-ARW in real time during the 2008 Atlantic hurricane season to gauge the efficacy of a simple, relatively low resolution cloud physics-based ensemble for producing reasonable track variations. Position forecast skill was assessed and compared to the NHC official forecast and dynamical members of the multi-model consensus. Intensity forecasts were not examined.

*Corresponding author address: Prof. Robert Fovell, UCLA Atmospheric and Oceanic Sciences, Los Angeles, CA 90095-1565. E-mail: rfovell@ucla.edu.

2. Experimental design

The ensemble forecast system employed WRF version 3.0, with a single, fixed 9720 by 5040 km domain having 36 km resolution centered over the Caribbean. Simulations were initialized with 1° GFS forecast grids and interpolated to 33 model levels beneath a 100 mb model top. All simulations were cold starts, without bogusing, correction or data assimilation of any kind. No nesting, vortex-following meshes, or special handling of surface fluxes or drag appropriate to tropical cyclones was implemented.

The ensemble consisted of twelve members, including 6 microphysics (MP) schemes – Kessler (K), Lin (L), WSM3 (W3), WSM5 (W5), WSM6 (W6) and Thompson (T) – joined with one of two cumulus parameterizations (CP), Kain-Fritsch (KF) or Betts-Miller-Janjic (BMJ). The Grell CP was excluded owing to poor performance in a real-time trial conducted during 2007 using an earlier version of WRF. Model physics held fixed among the runs included the YSU PBL, Dudhia shortwave and RRTM longwave, and thermal diffusion land surface schemes.

This study focuses on five Atlantic storms that made landfall in the United States: Dolly, Fay, Gustav, Hanna, and Ike, spanning 65 separate initialization periods. Initially, ensemble runs were made twice daily, at 00 and 12 UTC. The frequency was increased to four per day during September which resulted in roughly half of the runs involving long-lived Hurricane Ike (see Table 1). Each simulation was integrated for 96 hours. A case was included only if the storm was more than 24 h from a U.S. landfall.

Model forecast locations for minimum sea-level pressure (SLP) were recorded hourly but the present analysis is restricted to the 0, 24, 48, 72 and 96 h lead times. Forecasts were compared to the best track positions and operational model forecasts obtained from the NHC. This analysis focuses on the official NHC forecast (OFCL) and the most skillful models for the five cases: the GFS Ensemble Mean (AEMN), the Hurricane WRF (HWRF) and the GFDL model. The "full analysis" includes every forecast time in which a closed cyclonic circulation could be detected, while the "pre-landfall" analysis

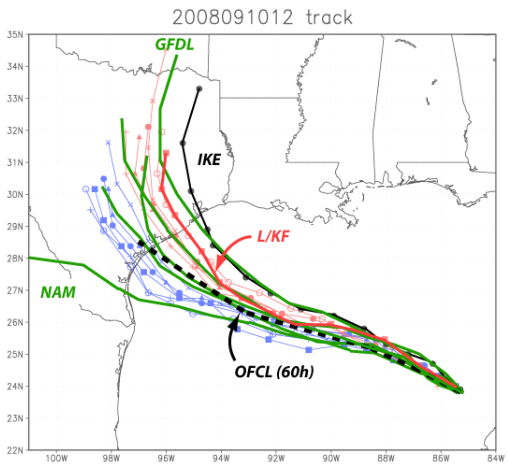

Fig. 1: 96 h forecast tracks for the 12 UTC 9/10/08 ensemble (Hurricane Ike). Red and blue indicates KF and BMJ members, respectively. Thick red line indicates L/KF track. Green lines mark GFDL, AEMN, HWRF, NAM and Navy NOGAPS (NGPS) tracks; dashed black track is OFCL forecast out to 60 h. Actual track labeled "IKE".

excludes forecasts made after the *actual* storm made landfall. The two analyses were quite similar.

Table 1: Simulation count by Atlantic storm: Full analysis/pre-landfall analysis.

storm/lead	0	24	48	72	96
Ike	26/26	26/26	26/22	22/18	18/14
Hanna	11/11	11/11	11/10	9/8	6/6
Gustav	14/14	14/14	13/11	9/8	7/6
Dolly	5/5	6/6	6/4	4/2	2/0
Fay	6/6	8/8	8/6	8/4	8/2
TOTAL	62/62	65/65	64/63	52/40	41/28
% Ike	42/42	40/40	41/42	42/45	44/50

3. Results

The initial hypotheses for this experiment were:

1. Individual ensemble members would lack skill relative to OFCL and the standard models
2. Ensemble spread would be comparable to the NHC consensus, and larger than would result from initial condition variations alone
3. A set of 4 or 5 equally skillful members would emerge to form an optimal sub-ensemble
4. This subset's ensemble mean would have comparable skill to some of the operational models

Significant skill for individual members was not anticipated owing to crude resolution and unsophisticated initialization. The initial GFS vortex was often the wrong size and intensity and was sometimes even in the wrong location. However, we believed that different biases and tendencies inherent in various cloud-related model physics could favorably or unfavorably interact with errors in both the initial tropical cyclone and its large-scale environment to produce track spreads as large as or larger than would result from initial condition perturbations alone. Further, some physics combinations would not produce acceptable forecasts, and would be culled. However, we anticipated that several members would emerge to to demonstrate an acceptable level of skill, and that the mean of this sub-ensemble would be better still. Thus, we were expecting the subset ensemble mean to be competitive with more sophisticated and expensive models, a not unusual expectation in ensemble forecasting.

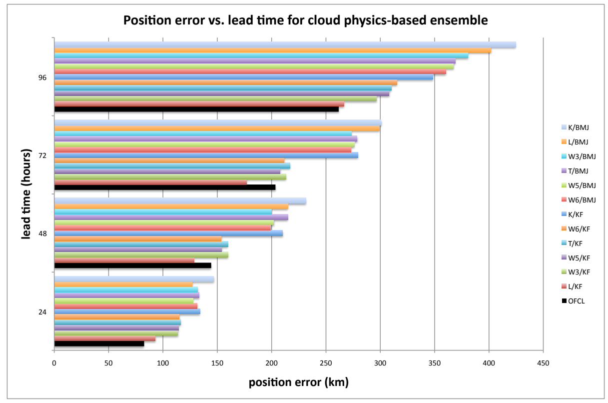

Fig. 2: Position error vs. forecast lead time for WRF cloud physics ensemble (full analysis), compared to OFCL (black bars). Members are ordered by 96 h forecast error.

Figure 1 presents one of the ensemble runs, starting 12 UTC September 10, during Hurricane Ike. In addition to the 12 MP/CP members, the official forecast and tracks from other operational models are also included. The thick black path labeled "IKE" marks the actual storm track. It is clear that while there is very substantial track variation among the MP/CP members, not unlike that among the operational runs. None of the forecasts predicted the correct landfall location, but the L/KF run's 72 h position had least error among all contenders.

Average position errors in the full analysis as a function of forecast lead time for the 12 ensemble members are shown in Fig. 2. The average OFCL error is included for comparison. It is clear that some members are systematically less skillful than others. Subsensemble means stratified by CP (Fig. 3) show that the BMJ scheme produced poorer forecasts, particularly for the longer lead times. The most poorly performing members at 96 h were K/BMJ, L/BMJ, W3/BMJ and T/BMJ.

The surprise of this ensemble is that one member (L/KF) was not only far and away the most skillful of the group, but also had position error statistics *as*

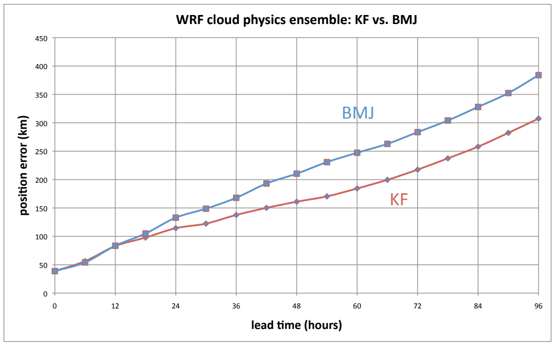

Fig. 3: Full analysis position errors for KF (red) and BMJ (blue) based runs.

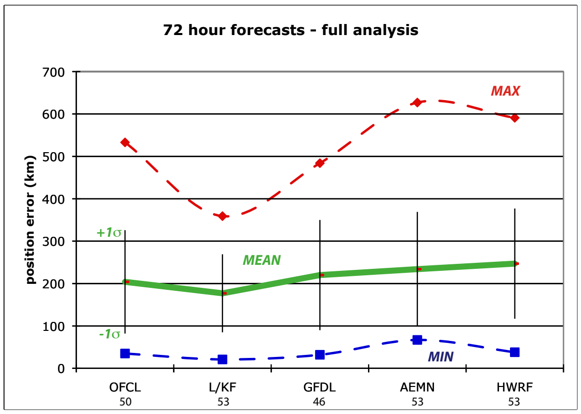

Fig. 4: 72 h forecasts from the full analysis.

good as or better than the official forecast and the best operational models at some lead times. Table 2 presents the percentage that each member had the lowest position error as a function of lead time for the pre-landfall analysis. (Numbers do not add to 100% owing to ties.) Between the 24 and 72 hour lead times, L/KF accounted for half of all "winning" forecasts. The percentage dropped to 29% by hour 96, even though L/KF retained by far the smallest average position error at that lead time (Table 3).

Table 2: Percentage of first place finishes (including ties) in the pre-landfall analysis.

member	24 h	48 h	72 h	96 h
K/KF	3	2	5	4
K/BMJ	6	6	8	14
L/KF	54	51	50	29
L/BMJ	8	2	8	15
W3/KF	3	11	0	14
W3/BMJ	8	6	0	11
W5/KF	6	11	5	4
W5/BMJ	9	4	0	4
W6/KF	5	15	5	4
W6/BMJ	5	4	0	4
T/KF	8	8	5	0
T/BMJ	11	2	15	4

Figure 4 compares the L/KF ensemble member with the OFCL, GFDL, AEMN and HWRF forecasts at 72 h for the full analysis. The number beneath each label indicates the number of cases. The means are NOT statistically distinct and the slightly lower average position error for L/KF reflected relatively fewer very bad forecasts. The statistics for the pre-landfall subset (not shown) were nearly identical, albeit with fewer cases, indicating the poor forecasts did not result from storms moving over land.

Table 3: Average position errors in the pre-landfall analysis.

member	24 h	48 h	72 h	96 h
K/KF	134	222	287	361
K/BMJ	147	241	318	429
L/KF	93	**129**	**177**	266
L/BMJ	127	214	303	395
W3/KF	114	171	245	307
W3/BMJ	132	209	284	329
W5/KF	114	164	238	322
W5/BMJ	128	210	286	350
W6/KF	115	165	241	328
W6/BMJ	131	207	285	347
T/KF	116	172	247	321
T/BMJ	133	219	289	345
OFCL	83	144	204	262
AEMN	113	189	234	**253**
HWRF	97	169	247	296
GFDL	**80**	138	220	369

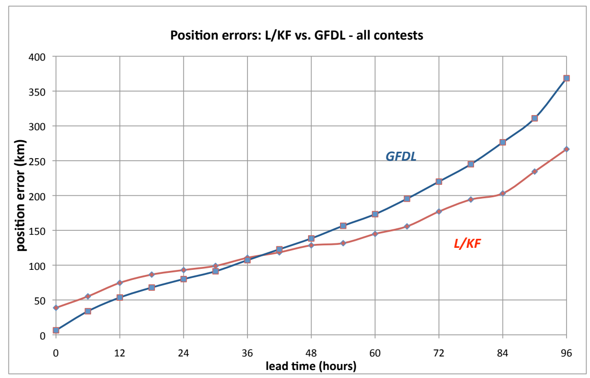

Fig. 5: Full analysis position errors for L/KF (red) and GFDL (blue) runs, against lead time for 65 contests combined.

Specifically contrasting L/KF's average position error trend with GFDL's (Fig. 5) suggests that it might have been more skillful had its initialization been better. It is likely that the L/KF run suffered from its raw, GFS initial condition. However, its error growth with time was smaller than GFDL's so that, after 36 h,, its average position error was superior. Note the number of samples available for averaging decreases with lead time (Table 1). Again, this appears to have transpired owing to a relatively smaller number of poor forecasts (not shown). Naturally, an assertion such as this begs for further proof, which is left to future work.

It should be noted that L/KF model fields were interpolated to a higher resolution mesh prior to storm centroid identification, reducing uncertainties owing to the present ensemble's relatively lower resolution compared to the GFDL model.

4. Conclusions

The interesting result is that, with this particular physics combination, the limited resolution and simply initialized WRF-ARW's forecasts were, on average, as good or better than those produced by more sophisticated models. As it is reasonable to anticipate that better forecasts would result from more accurate initializations, this result, in combination with our other work (e.g., Fovell and Su 2007; Fovell, Corbosiero and Kuo 2009) suggests that operational models could benefit from further tuning with respect to the model physics, particularly those involving cloud processes.

For 2009, another operational ensemble will be conducted for Atlantic storms, weather permitting. This ensemble will use the current (3.1) version of ARW and will emphasize better initialization, with physics choices guided by the present study's results.

5. References

Fiorino, M. J., and R. L. Elsberry, 1989: Some aspects of vortex structure related to tropical cyclone motion. *J. Atmos. Sci.,* **46**, 975-990.

Fovell, R. G., and H. Su, 2007: Impact of microphysics on hurricane track. *Geoph. Res. Lett.,* **134**, L24810, doi:10.1029/2007GL031723

Fovell, R. G., K. L. Corbosiero, and H.C. Kuo, 2009: Cloud microphysics impact on hurricane track as revealed in idealized experiments. *J. Atmos. Sci.,* **66**, 1764-1778.

Jordan, C. L., 1958: Mean soundings for the West Indies area. *J. Meteor.,* **15**, 91-97.

Acknowledgments. This work was supported by NSF grant ATM-0554765. Computing facilities were provided by the Aerospace Corporation. NHC provided best track and forecast data. Travis Wilson very ably assisted with the track error computations. Ty Fovell also assisted with some analyses.