PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2010), 62A, 633–646

TELLUS

Hurricane Forecasts with a Mesoscale Suite of Models

By T. N. KRISHNAMURTI[1]*, SANDEEP PATTNAIK[2], MRINAL K BISWAS[1], ED BENSMAN[3], MELANIE KRAMER[1], NAOMI SURGI[4] and T. S. V. VIJAYA KUMAR[4],
[1]*Department of Earth, Ocean and Atmospheric Sciences, Florida State University, Tallahassee, FL 32310, USA;* [2]*Indian Institute of Tropical Meteorology, Pune, India*; [3]*University of Notre Dame, Notre Dame, IN, USA;* [4]*NCEP/EMC/NOAA, 5200 Auth Road, Camp Springs, MD, USA*

(Manuscript received 7 October 2009; in final form 27 May 2010)

ABSTRACT

A suite of mesoscale models are being used in the present study to examine experimental forecast performance for tracks and intensity of hurricanes covering the years 2004, 2005 and 2006. Fifty-eight storm cases are being considered in the present study. Most of the mesoscale models are being run at a horizontal resolution at around 9 km. This includes the WRF (two versions), MM5, HWRF, GFDL and DSHP. The performances of forecasts are evaluated using absolute errors for storm track and intensity. Our consensus forecasts utilize ensemble mean and a bias corrected ensemble mean for these member models on the mesoscale and the large-scale model suites. Comparing the forecast statistics for the mesoscale suite, the large-scale suite and the combined suite we find that the mesoscale suite provided the best track forecasts for 60 and 72 h. However, the forecast from the combined suite of model were also very close to the track errors of the mesoscale at 60 and 72 h. Overall track forecast errors were least for the combined suite. The intensity forecasts of the bias corrected ensemble mean of the mesoscale suite were comparable to DSHP and GFDL at the later part of the forecast periods.

1. Introduction

This study addresses consensus forecasts for the hurricane seasons of 2004, 2005 and 2006. Given two suites of models (large scale and mesoscale) it is possible to construct consensus forecasts of the track and intensity of hurricanes. This paper makes use of the well-known consensus ensemble mean (ENSM), which is a simple average of all the member models and the bias corrected ensemble mean (BCEM) (Appendix A). Unlike the previously discussed superensemble the present study does not utilize a training phase. The new aspect of the present work is in the area of consensus forecasts from a suite of mesoscale models with a horizontal resolution of roughly 9 km. We also carry out a comparison of results on hurricane forecasts for the years 2004, 2005 and 2006 from a suite of large scale and mesoscale models and from a mix of the two suites of models. Operational hurricane intensity forecasts have not seen any major improvements in the last 17 yr, Fig. 1, which cites the official statistics on year-to-year skills of intensity forecasts. Basically this illustration suggests that not much improvement has been seen for 1–5 d forecasts of intensity in operations.

Both outer and inner core influences are considered important for the intensity problem Krishnamurti et al. (2005). Hurricane intensity forecasts seem to be influenced by many factors. The hurricane intensity issue is being addressed by the research community to explore sensitivity with respect to a number of areas such as: dry air intrusions, dust incursions (Dunion and Velden, 2004), vertical shear, heat content of the ocean (Shay et al., 2000), cloud microphysics (Pattnaik and Krishnamurti 2007a,b), inner core organization of convection, diabatic potential vorticity, scale interactions among the cloud and the hurricane scales (Krishnamurti et al., 2005), angular momentum issues, inner core dynamics, inner core hot towers and vortex Rossby waves. This list is in fact quite extensive and most studies have to limit themselves to a few focused areas. Data coverage in the inner core, in spite of available remote sensing from satellites and aircraft reconnaissance, is limited. The mesoscale models are beginning to address some of these aforementioned issues in a limited manner. The data sets outside the inner core are currently being reasonably handled by large-scale models from their variational data assimilation. This helps in defining the outer angular momentum of the storm environment reasonably well. The inflow channels in a hurricane's lower level bring this large outer angular momentum into the storm's interior.

*Corresponding author.
e-mail: tkrishnamurti@fsu.edu
DOI: 10.1111/j.1600-0870.2010.00469.x

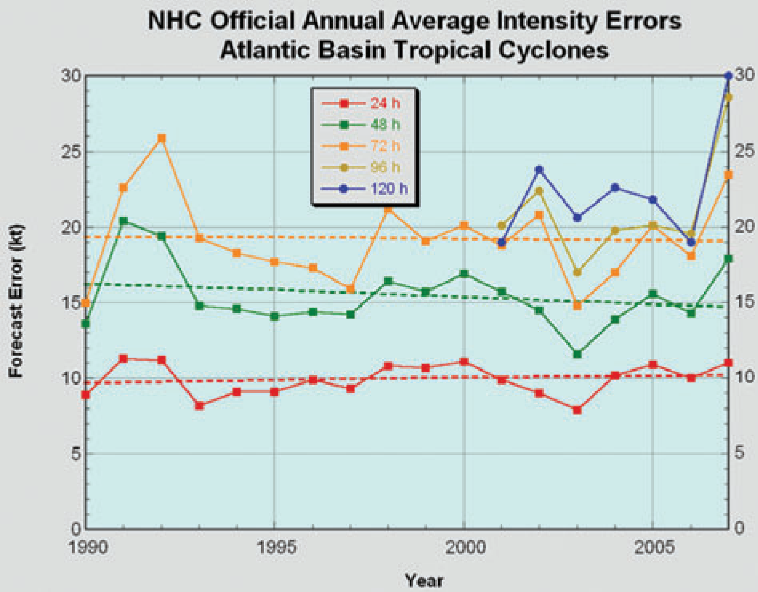

Fig. 1. National Hurricane Center Official Annual average intensity forecast errors (in kt) in the Atlantic basin from 1990 to 2007. The forecast errors of 5, 10, 15, 20, 25, 30 kt corresponds to approximately 2.5, 5, 7.5, 10, 12.5 and 15 m s^{-1}. Adapted from NHC (Franklin 2006).

Even processes such as sensitivity of clouds to microphysics appear to modify cloud growth and in the estimates of these torques which in turn impact the eventual intensity (Pattnaik and Krishnamurti, 2007a,b). The mesoscale models (Houze et al., 2007) provide the opportunity to incorporate inner core data sets from hurricane aircraft reconnaissance and ocean temperatures at a high resolution. Some of these models also carry an inner nest at a very high resolution where the microphysical impacts on cloud growth can be incorporated.

Consensus forecasts are carried out using subjective methods based on experience and using objective ensemble methods. The subjective methods of experienced forecasters at NHC have shown unusual expertise in a rather consistent manner. A recent example of intensity forecast from the multimodels is illustrated in Fig. 2. The abscissa of the curve denotes hours of forecast and the ordinate denotes skill with respect to Statistical Hurricane Intensity Forecasts SHIFOR (Knaff et al., 2003). SHIFOR is a statistical model based on climatology and persistence. It is used as a baseline for intensity prediction by the NHC. This is a summary of all hurricanes of the year 2005. ICON, a simple mean of two models (Previous cycle GFDL and DSHP) carried the highest skill among all member models. The member models are

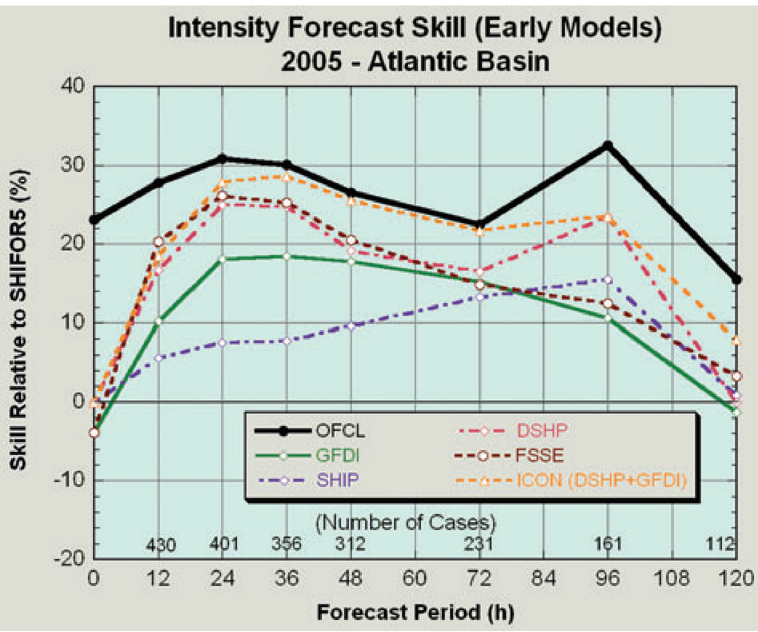

Fig. 2. 2005 Intensity skills for different models during 2005 Atlantic Hurricane season. The ordinate shows skills with respect to SHIFOR5 and the abscissa shows the forecast lead times. Adapted from NHC (Franklin 2006).

evaluated annually and change from year to year. [Decay-SHIPS (DSHP) is a version of Statistical Hurricane Intensity Prediction Scheme (SHIPS) which includes an inland decay component. Since land interactions result in weakening, the Decay-SHIPS provide more accurate intensity forecasts when cyclones make landfall.] The message clearly is that, if a consensus forecast performs better than the best model it does not necessarily imply a usefulness of the product since the official statement admits that no improvements in hurricane intensity were possible in the last 25 yr.

2. Data usage and selection of storms

The selection of hurricanes, (Table 1) for this study was entirely based on data files provided to us by the EMC/HWRF modelling group. This storm list is not all inclusive; this is an experimental research project where forecasts for nearly all important storms of 2004, 2005 and 2006 seasons are included. The intention here is to cover as many storms that were included in the EMC/HWRF data tapes. In this paper, we shall show a relative comparison for forecasts of the storms made from a suite of models listed in Table 2. The number of cases selected in our study, shown in Table 2, do carry an inadvertent bias. More of these forecast cases were for storms that showed intensification in the first 72 h time frame and a lesser number covered weakening cases. This happened as a consequence of the selection of the initial states. Our results thus largely address storms that were intensifying. This is a limitation for the choice of our initial states. We expect to remedy this from a selection of a much larger initial database in a future study. The cases we have selected are however quite identical to the forecast database used by the National Hurricane Center in their operations (personal communication with Dr James Franklin of NHC).

Table 1. Name, year, duration and domains of tropical storms and hurricanes

Year	Storm name	IC for 72 h FCST (00/12 UTC)	Domain
2006	Chris	1 Aug	55W–75W, 10N–30N
	Debby	22 Aug	20W–50W, 5N–35N
	Ernesto	27 Aug	65W–85W, 10N–30N
	Gordon	13 Sep	45W–65W, 15N–40N
	Helene	17 Sep	40W–60W, 15N–35N
	Isaac	30 Sep	50W–70W, 25N–50N
2005	Dennis	7 July	65W–95W, 10N–35N
	Emily	15 July	60W–90W, 5N–30N
	Harvey	3 Aug	50W–70W, 20N–40N
	Irene	13 Aug	60W–80W, 20N–40N
	Katrina	27 Aug	75W–95W, 20N–40N
	Maria	4 Sept	45W–65W, 20N–40N
	Nate	6 Sept	55W–70W, 20N–40N
	Ophelia	8 Sept	65W–85W, 25N–45N
	Philippe	18 Sept	50W–70W, 10N–35N
	Rita	21 Sept	80W–100W, 20N–35N
	Stan	02 Oct	80W–100W, 10N–25N
	Wilma	22 Oct	70W–90W, 15N–35N
2004	Charley	10 Aug	60W–87W, 5N–22N
	Danielle	15 Aug	25W–45W, 10N–35N
	Frances	1 Sept	65W–85W, 15N–40N
	Frances	2 Sept	65W–85W, 15N–40N
	Gaston	28 Aug	65W–85W, 25N–45N
	Hermine	30 Aug	60W–75W, 30N–50N
	Ivan	12 Sept	75W–95W, 15N–35N
	Ivan	13 Sept	75W–95W, 15N–35N
	Jeanne	24 Sept	65W–85W, 20N–35N
	Karl	18 Sept	30W–55W, 10N–30N
	Lisa	25 Sep	35W–55W, 10N–30N

Note: All storm cases are run for 0000 and 1200 UTC.

3. Mesoscale suite of models

Table 2 provides a brief description of the five mesoscale models of our suite.

3.1. MM5-A

One version of MM5 model (MM5-A) is configured having the planetary boundary layer and radiation parametrization schemes and surface layer parametrization, that is, Blackadar PBL (Blackadar 1976, 1979; Zhang and Anthes 1982) radiation (Dudhia, 1989) and multilayer soil temperature (5-layer) model, respectively. The Betts–Miller cumulus parametrization (Betts, 1986; Betts and Miller, 1986) with Goddard microphysics parametrization schemes (Tao and Simpson, 1989; Tao et al., 1993; Lin et al., 1983) are used for MM5-A. The model vertically, from the top of the model (50 hPa) to the surface, there are 23σ levels. MM5A have horizontal resolution of 9 km.

3.2. Advanced research weather research and forecasting model (WRFARW A and B)

The WRF-ARW model is being developed as a collaborative effort among the NCAR Mesoscale and Microscale Meteorology Division (MMM), and NCEP's Environmental Modeling Center (EMC). We have used the following physics options for this model: Radiation schemes Longwave: rapid radiative transfer model (rrtm) (Mlawer et al., 1997) Shortwave: Dudhia scheme (Dudhia 1989; Grell, 1993), Surface physics: Monin-Obukhov (Janjic) scheme (Monin and Obukhov, 1954), Land surface model: five layer thermal diffusion (Skamarock et al., 2005) Planetary boundary layer scheme: Mellor-Yamada-Janjic (MYJ) TKE PBL (Janjic, 1994, 2002) Convection scheme: Kain-Fritsch (new Eta) scheme (Kain and Fritsch, 1993), Explicit moisture scheme: WRF six-class graupel scheme (WSM6) (Hong et al.,

Table 2. Comparison of mesoscale models of different parameters

Parameters	MM5A	WRFA	WRFB	HWRF	GFDL
Horizontal resolution	9 km	9 km	9 km	27/9 km	30/15/7.5
Vertical sigma layers	23	31	28	43	42
Initial and boundary conditions	GFS initial conditions and forecasts (T382L64)	GFS initial conditions and forecasts (T382L64)	GFS initial conditions and forecasts (T382L64)	GFS initial conditions and forecasts (T382L64)	GFS initial conditions and forecasts
Radiation parametrization	Dudhia	RRTM and Dudhia	RRTM and Dudhia	GFDL scheme	Schwarz-kopf-Fels scheme
Cumulus parametrization	Betts and Miller,	Kain–Fritsch (new Eta) scheme	Betts–Miller–Janjic	Simplified Arakawa Schubert	Arakawa Schubert
Microphysics parametrization	Goddard	WRF 6-class	Ferrier	Ferrier	Ferrier
Planetary boundary parametrization	Blackadar	Mellor-Yamada-Janjic TKE	Yonsei University (YSU)	GFS Non-Local PBL	GFS Non-Local PBL
Land surface model	Multilayer Soil model	5 layer thermal diffusion	5 layer thermal diffusion	GFDL Slab Model	Slab Model
Duration of forecast	72 h	72 h	72 h	96 h	120 h
Nest	Single	Single	Single	Two	Triply
Bogus vortex/vortex relocation	No	No	No	Yes	Yes
Model top	10 hPa	10 hPa	10 hPa	50 hPa	50 hPa

2004; Hong and Lim, 2006). Simulation from a second version of the WRF model (i.e. WRFB) were conducted at the Center for Research Computing at the University of Notre Dame by one of our collaborators. The WRFB model consisted of similar physics options as WRFA, with the exceptions: YSU Planetary boundary layer, Betts–Miller–Janjic convective parametrization (Janjic, 1994; Betts and Miller, 1993) and Ferrier (1994) explicit moisture parametrization schemes. These two models (WRFA and WRFB) have a horizontal resolution of 9 km.

3.3. Hurricane WRF (HWRF)

The HWRF model which is included in our suite of mesoscale models was run at National Center for Environmental Protection (NCEP). The operational version is used for running 2004, 2005 and 2006 storms. The outermost grid of the HWRF model has a 27 km resolution that carried an outer domain which is 75° latitude by 75° longitude. The innermost grid has a nine km resolution and 43 staggered hybrid sigma levels. The HWRF model is a coupled two-way double nested model with a moving nest. The atmospheric component of the HWRF model was coupled with the Princeton Ocean Model (POM) (Blumberg and Mellor, 1987). The HWRF model uses a pressure hybrid coordinate (Simmons and Burridge, 1981). The Global Forecast System (GFS)/GFDL (Geophysical Fluid Dynamics Laboratory) surface, boundary layer physics, and GFDL/GFS radiation options were incorporated for some of the physics options. The Simplified Arakawa Schubert (SAS) for the cumulus convection scheme (Arakawa and Schubert, 1974; Grell 1993), Ferrier (2005) for microphysics parametrization, GFS non-local planetary boundary layer (Troen and Mart, 1986; Hong and Pan, 1996) scheme with surface layer physics of Moon et al. (2007) and GFDL radiation scheme are the physics options used in HWRF model integration. The sea surface temperatures are obtained from Princeton Ocean Model (POM) for the Atlantic basin. The model uses initial and boundary conditions from the GFS model, which are subjected to relocation and bogusing of the vortex structure using 3DVAR data assimilation using high-resolution Grid Statistical Interpolation (GSI). Additional details of HWRF model is available from Tuleya and Gopalakrishnan (2006).

For a weak storm, the initial analysis is a blend of the first guess and the composite storm. The weighting is based on the observed storm intensity and the vertical structure of the first guess storm. A balanced storm environment is created by first adjusting the wind field based on the TPC observation, then adjust other fields based on physical constraints. Instead of assuming a gradient wind balance (as in the GFS analysis), the surface pressure is expressed as a function of the non-linear gradient wind stream function H(F). Once the surface pressure is adjusted, the hydrostatic balance is used to adjust the vertical temperature fields.

3.4. GFDL model

The GFDL hurricane forecast model is always considered an important member model because of its high performance level. This is a triply nested grid point model with the innermost grid which corresponds to roughly 7.5 km. The inner grid spreads over 5° latitude by 5° longitude area. The next two outer grids carry resolutions of one sixth and one half-degree resolutions. The smallest resolvable wave ($2\Delta x$) in the innermost grid is of the order of 15 km and is still not sufficiently high to resolve the inner core details of a hurricane adequately. This model carries 42 vertical levels in the conventional sigma coordinate. The physics modules includes the simplified Arakawa Schubert cumulus parametrization (Pan and Wu, 1995), a non local planetary boundary layer scheme (Troen and Mahrt, 1986), the model includes microphysics based on the work of Ferrier (2005) this directly impacts the large scale condensation in the model.

A current version of the GFDL model is coupled to POM for its ocean modelling, which was developed by Blumberg and Mellor (1987). This coupling permits wind induced upwelling and cooling of the ocean in the trail of strong winds for moving hurricanes. The strong evaporation and moisture supply from the ocean to the atmosphere in hurricanes has been well tested in its definition of the interface. This ocean model carries 23 vertical sigma levels at a horizontal resolution of roughly one-sixth degree latitude/longitude. An important component of the GFDL model is its initial vortex specification. This is described in Kurihara et al., (1993, 1995). The removal of a hurricane in the parent analysis provided by GFS is an important start for this vortex specification. This scheme next prepares a bogus vortex with inputs from NHC parameters.

In terms of performance the use of the innermost high resolution nested grid has improved the intensity forecasts (for day 3 of forecasts) by as much as 2 m s^{-1} (Bender et al., 2007). GFDL provides 6 hourly forecasts, in real time, to the NHC operations during the hurricane season for the Atlantic, Caribbean, Gulf of Mexico and the eastern Pacific sectors.

The aforementioned mesoscale models were integrated out (by the modellers and FSU) out to 72 h. This includes the entire 58 storm cases, covering years 2004–2006 that formed over the Atlantic basin over different domains those are summarized in Appendix A. The model outputs from these models were stored at 6 h intervals. All of the mesoscale models' initial and boundary conditions were derived from GFS forecasts.

4. Comparison of results from suites of large scale and mesoscale models

Here we present a comparison of forecast errors (track and intensity) for a suite of large-scale operational models and those from the mesoscale suite of the above models. The large-scale operational models include the Global Forecasting System (GFS) of National Center for Environment Prediction (NCEP), the Navy Operational Global Prediction System (NOGAPS) and United Kingdom Met Office model (UKMet). Table 2 provides lists of models that were used for forecasts of tracks and intensity for the mesoscale suite of models and Table 3 provides some present model configurations for the large scale suite of models. These comparisons are, here, limited to 3 d since we wanted to examine the capabilities of regional nested models for shorter range intensity forecasts. Absolute track errors (km) and absolute intensity errors (m s^{-1}) are compared in Figs. 3a and b. Here we shall present the results of a bias corrected ensemble mean (Methodology is presented in Appendix A) for these two suites of models. This includes a summary of results for all 6 hourly forecasts for the hurricane season for the years 2004, 2005 and 2006, these carried with 22, 24 and 12 cases, respectively. All calculations were done for homogeneous comparisons, that is, error calculations were made when all of the member model's forecasts were simultaneously present. These results show that for 3 d forecasts, the intensity from the mesoscale suite carried somewhat lower absolute errors (except 12 h forecasts) compared to the large scale suite. The intensity errors reduce by as much as 2 m s^{-1} by 72 h forecasts. This systematic reduction of intensity error from the large sample of forecasts appears possible from the mesoscale suite. We examined this performance of the mesoscale (bias removed ensemble mean) against the performance of each member model of the large scale suite, Figs. 4a and b . Here we see that the bias removed ensemble mean for the mesoscale suite generally outperforms each of its member model forecasts (later shown in Fig. 8) for individual storms. This large-scale suite of models is currently being used

Table 3. Model configurations of the large scale suite of models

Models	Model physics	Horizontal grid spacing	Vertical levels	Cumulus parametrizations	Data assimilation
GFS	Hydrostatic spectral	∼35 km	64	Simplified Arakawa Schubert	3-D Var, GSI, GDAS analysis
NGPS	Hydrostatic spectral	∼55 km	30	Emmanuel	3-D Var, NAVDAS analysis
UKMet	Non-hydrostatic grid point	∼40 km	50	Gregory/Rountree	4-D Var

Courtesy: NHC website (http://www.nhc.noaa.gov/modelsummary.shtml).

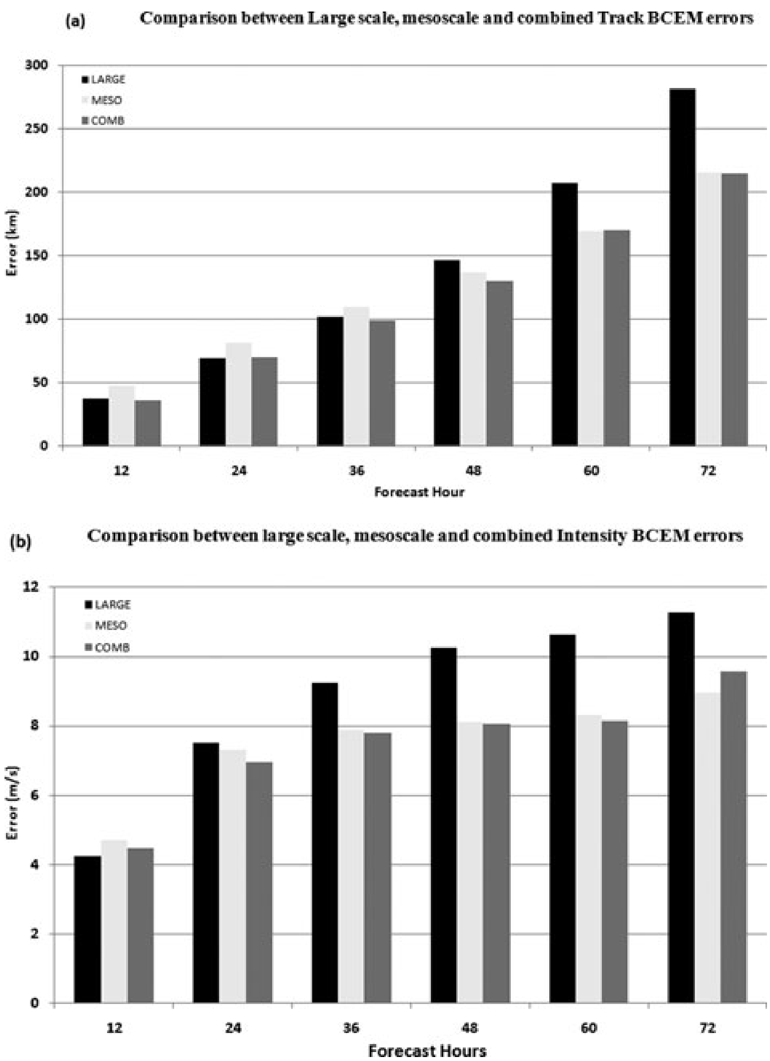

Fig. 3. Comparison of (a) track errors (km) and (b) intensity errors (m/ s^{-1}) between bias corrected large scale, mesoscale and using both large and mesoscale suite of models.

in operations at the National Hurricane Center. Although this is a small improvement for intensity forecasts, with further improvements of the member models of the mesoscale suite, we can expect further improvements. The large-scale models with their coarse resolutions, cannot be expected to perform better than the mesoscale models for days 1 and 2 of forecasts. Our results show that there is however a major spin up problem in these mesoscale models. The initial data for all models come from the 3D Var assimilation of the GFS model and when interpolated and inserted into mesoscale models they undergo a drastic spin up. We did not assimilate the mesoscale data sets for each model separately on the mesoscale, which remains a problem for the future. Thus it was not surprising that the large-scale models, each of which carry its own data assimilation, performed better in the initial 2 d compared to the mesoscale models. These large-scale models do incorporate the dropwindsonde data, which is transmitting via Global Telecommunication Systems (GTS), in their respective data assimilation.

The mesoscale suite of models provide better forecasts for storm tracks compared to the large scale models for hours 48, 60 and 72 (Fig. 3a). During the first 36 h the large-scale models carry track errors of the order of 10 km or less compared to the mesoscale suite. Overall we find that the results from these two suites of models are closely comparable for the first 48 h. At hour 72 of forecasts the absolute track error of the mesoscale suite is roughly 80 km less than that for the large-scale models. Given the large sample of forecasts (made at intervals of every 6 h) for the years 2004, 2005 and 2006, these results seem important.

We next combined all of the mesoscale and the large -scale models to arrive at a combined bias removed ensemble mean. This exercise was completed covering all of the hurricanes provided in Table 1 of the year 2004, 2005 and 2006. This combination of all nine models provided the best track forecasts for hours 12 to 72. This combined product was somewhat superior to the bias corrected ensemble means of the large and the mesoscale

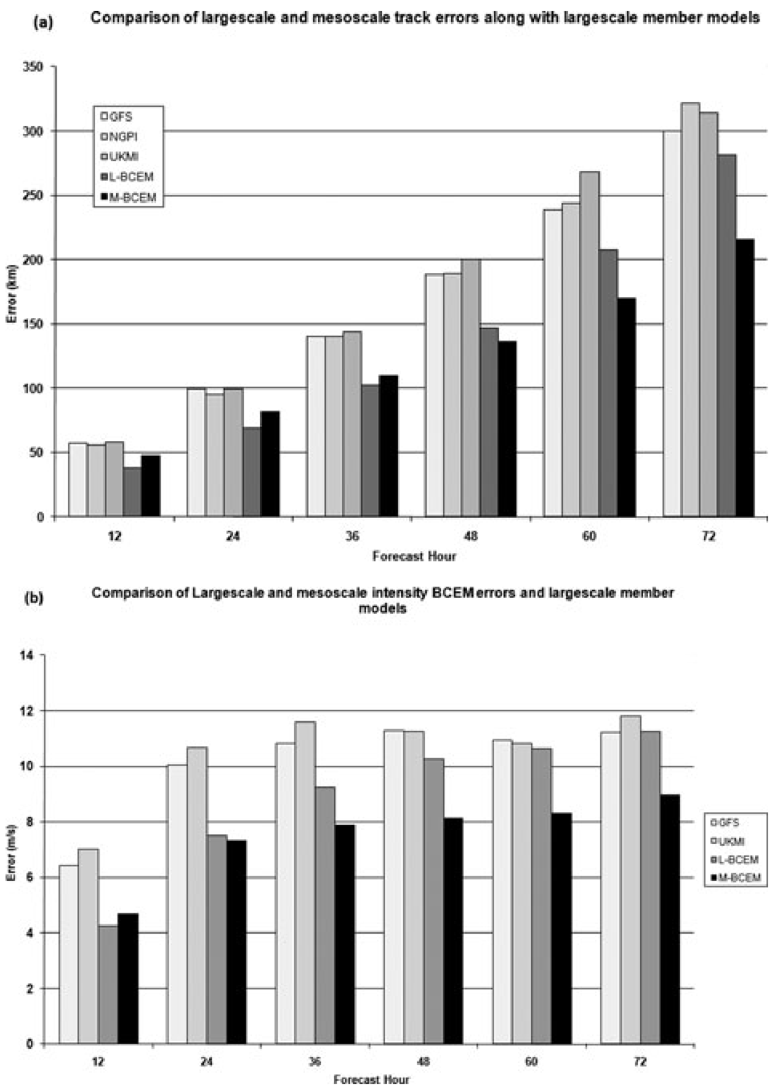

Fig. 4. Comparison of large scale and mesoscale bias corrected ensemble mean errors with GFS and UKMI individual model errors for (a) track (in km) and (b) intensity (m s^{-1}).

suites. The 60 and 72 h forecasts from this product were comparable to the mesoscale suites' bias corrected ensemble mean. For intensity forecasts we did not see a clear superiority for this combined product. The combined product had less error compared to the bias corrected ensemble mean of the large and the mesoscale suites through 60 h forecasts. At 72 h the mesoscale suite carried the least forecast errors. Since these results are based on a fairly large number of hurricane forecasts the message conveyed by the combined suite of models is important. There are examples where the bias corrected ensemble mean of the mesoscale models were smaller than those of the large scale models, when we combine all models it is not guaranteed that the bias corrected ensemble mean of these combined models would show the least errors. This depends on the spread of errors of each of the member models of this entire suite. We would recommend the use of the combined suite for track forecasts through day 3. For intensity forecasts we recommend the use of the combined suite for 1 and 2 d forecasts and the mesoscale for day 3 forecasts.

Models are constantly evolving and changing, hence the above recommendations are expected to change in the coming years. We have compared these results from the combined suite with those from the best member model. The track forecast error at day 3 can be improved by as much as 50 km compared to the best model. For intensity forecasts it should be possible to reduce the errors with respect to the best model by as much as 3 m s^{-1}. Although this is a small systematic improvement, it is expected that as member models improve we can expect a further overall reduction of such errors.

5. Forecast performance of member models

In this, we describe the performance for each of the member models of the mesoscale suite with respect to their bias

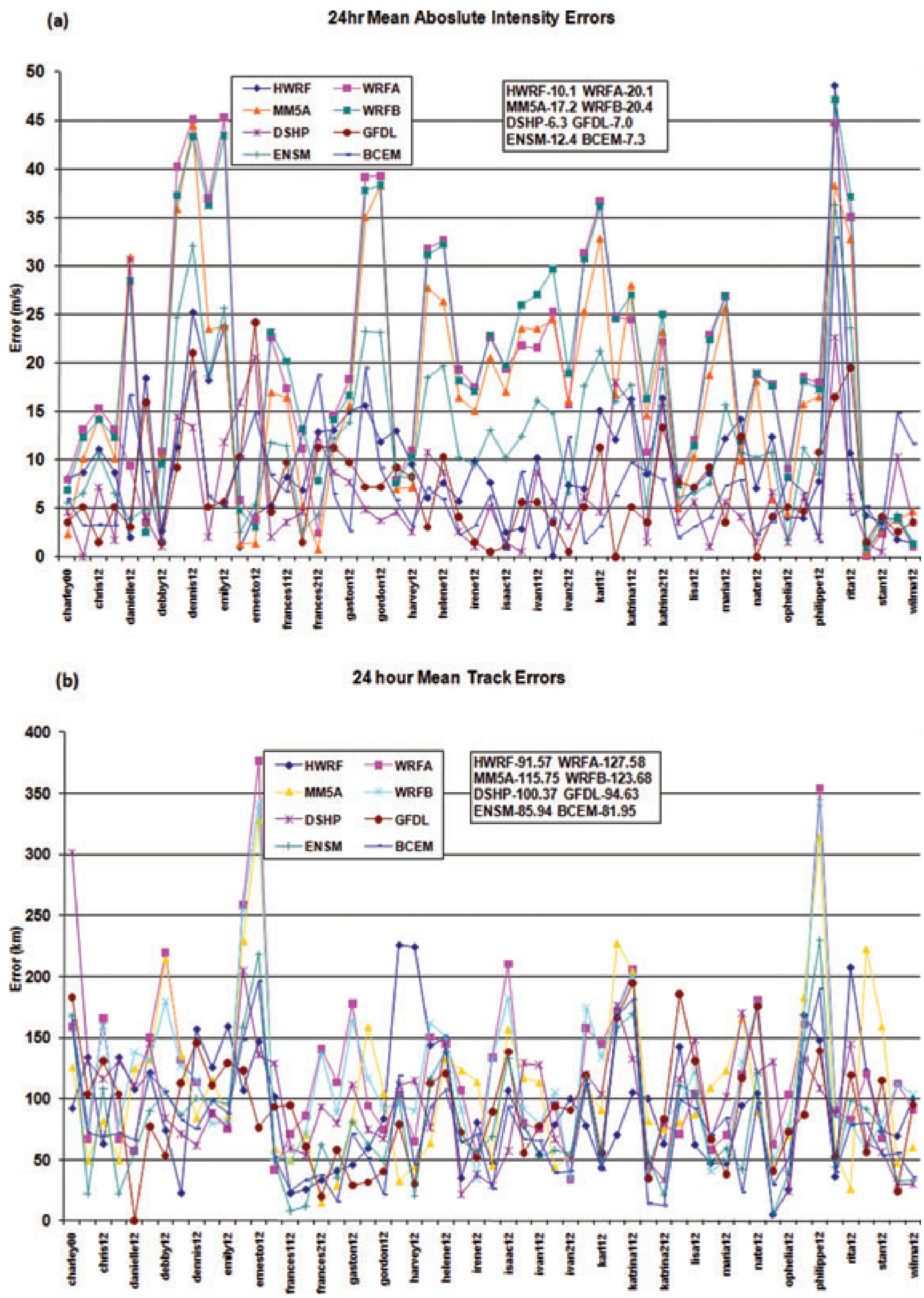

Fig. 5. Twenty-four hour mean absolute errors for member models, ensemble mean and bias corrected ensemble mean for (a) intensity (m s^{-1}) and (b) track (km). In the inset the average errors for 24 h forecast are shown for member models, ensemble mean and bias corrected ensemble mean. The storm names are shown in the *x*-axis followed by the initial time. The storms which have two initial dates as shown in table is named as <storm_name>2<XX> where XX is the initial time.

removed ensemble mean. Figs. 5–7 shows the absolute track errors (km) and the absolute errors for intensity (m s^{-1}) for forecasts of all hurricanes of the year 2004, 2005 and 2006 as shown in Table 1 for forecasts at the end of days 1, 2 and 3. Figures 5–7 includes the error distributions of each member model, the ensemble mean and the bias corrected ensemble mean of the mesoscale suite. These illustrations also include the models that are included in Table 2 and the mean error for each model.

The bias corrected ensemble mean (BCEM) clearly provides one of the better forecasts compared to the member models in a consistent manner. In evaluating the results in Figs. 5–7 we note that the BCEM of the mesoscale suite can be ranked 1, 1 and 1 (ENSM was very close to BCEM) respectively for 1, 2

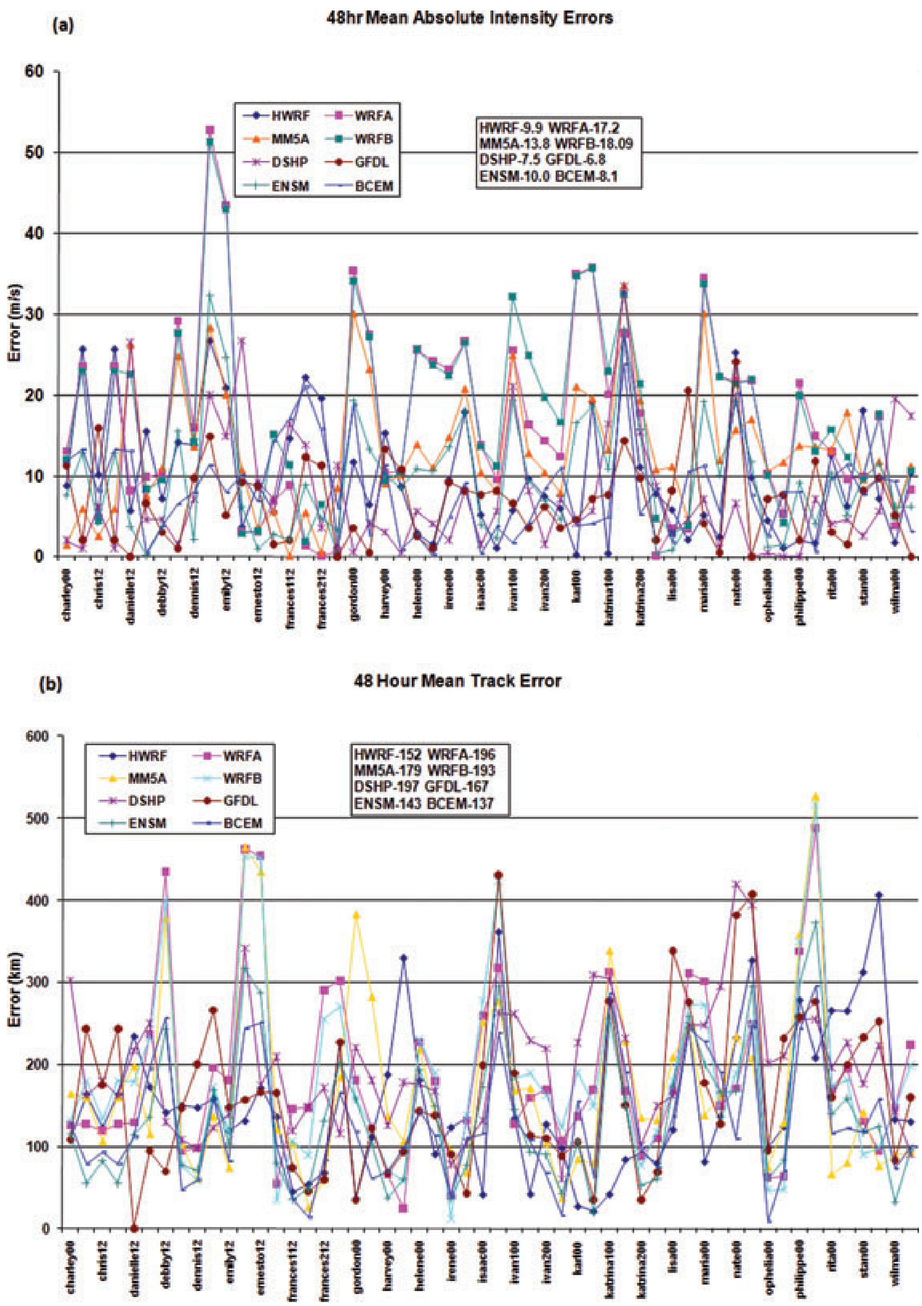

Fig. 6. Same as in Fig. 5 except for 48 h forecast.

and 3 d forecasts for the track errors with respect to the member models and the ensemble mean. The corresponding rankings for the intensity errors for the mesoscale suite are 3, 3 and 2, respectively.

In this illustration we see a clustering of model errors, that is, most models do well for some storms and do poorly for others. The clustering of errors reflects initial state dependence of model errors. Most of these models use the GFS and that clustering can partly be attributed to the common use of GFS based assimilation. The large intersection of lines in Figs. 5–7 is indicative of a lack of consistency of performance of member models from one forecast to the next. Consistent improvement in performance is necessary for improving statistical post processing methods such as the ensemble mean and the BCEM. Such inconsistencies in performance arise from resolution, data coverage, and response of data assimilation to the data coverage, model physics, and boundary condition and nesting strategies. Further studies are required to locate the source of such errors and require a major effort, Krishnamurti et al., (2004).

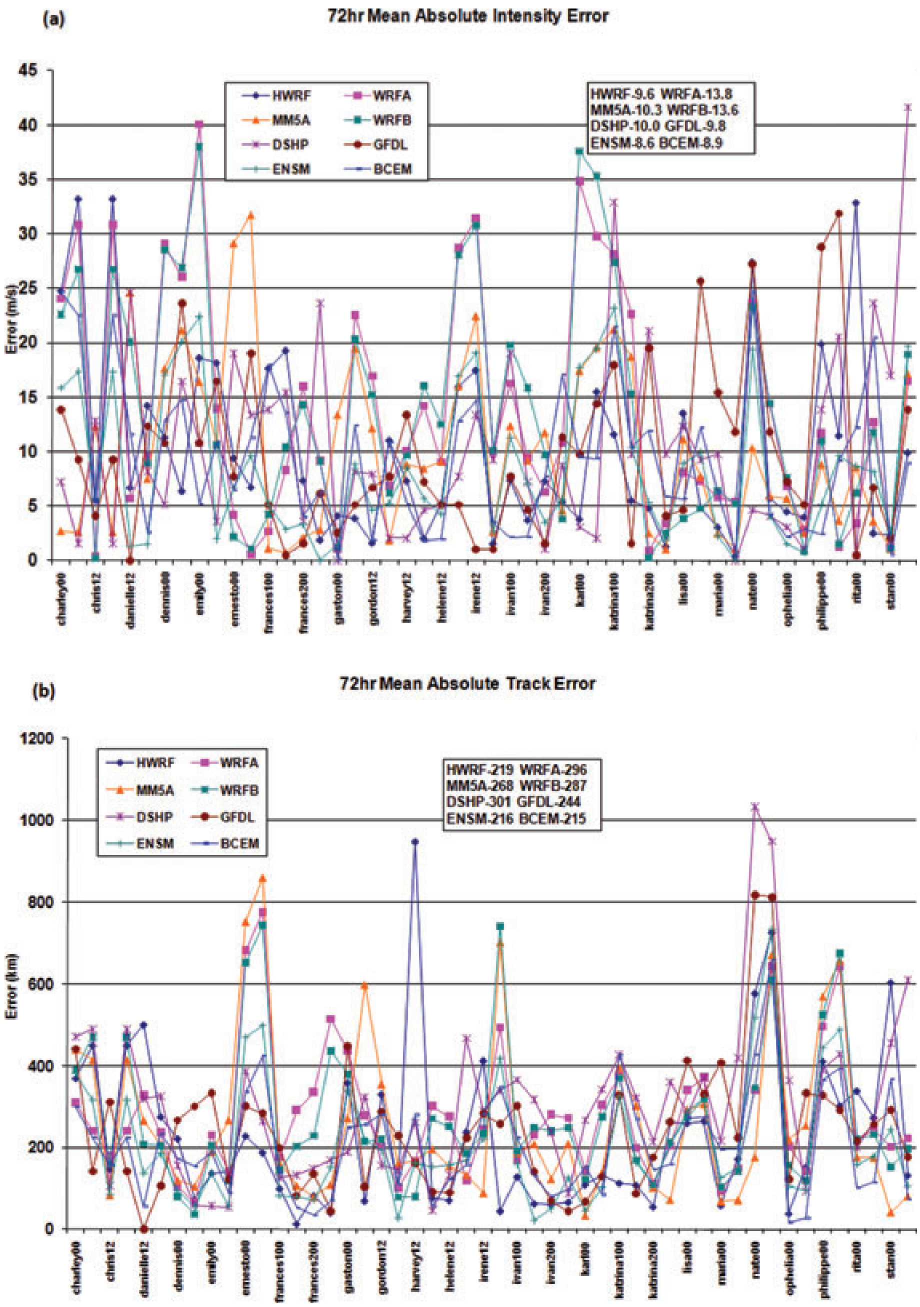

Fig. 7. Same as in Fig. 5 except for 72 h forecast.

For 24 h forecasts, large intensity errors were noted for hurricane Dennis, Gaston, Karl of 2004 and Philippe of 2005. These intensity errors were as large as 35–45 m s^{-1}. The intensity errors for DSHP's for these storms were in the range of 5–15 m s^{-1} and those for bias corrected ensemble mean were between 10 and 15 m s^{-1}. Further work is clearly needed to reduce such errors for two of our choice of models the WRFA and WRFB that contributed the largest errors. The range of errors for HWRF was generally lower than those for the other regional mesoscale models. The intensity errors at 48 h were generally less than 30 m s^{-1}. for most forecasts (except those for Dennis). The error histories for intensity forecasts at 72 h were similar to those at 24 h. The two models WRFA and WRFB exhibited consistently large intensity errors. The initial large errors of the model are largely attributed to the spin up issue that was apparent in these forecasts. One of the possible contributors was our choice of the GFS 3DVAR for the provision of the initial states for these models. The choice of physics in these two models was not entirely different from those of some of our other models. This is one of the most difficult questions to answer when one is asked why the performance of a model X is so good or is so bad. Diagnosing this problem is not a simple matter given that the behaviour of a certain physical parametrization within a model depends strongly on other features of the model within which the physics reside, Krishnamurti and Sanjay (2003). These errors arise for a host of reasons such as data sets, data assimilation, response of model physics to these data sets and impacts of model physics on other components of model physics (such as PBL and cumulus parametrizations) as well as dynamics, finite difference procedures (e.g. Arakawa C-grid versus Arakawa E-grid), air–sea fluxes and a myriad of non-linear couplings among dynamics and physics. Such modelling errors can be somewhat reduced from the use of a large number of member models and collective bias corrections.

The differences in track errors among model for 24, 48 and 72 h, Figs. 5–7 shows a few well-marked spikes. For 24 h forecasts Emily of 2004 and Ophelia (2005) carried large errors of the order of 350–375 km. By 48 h as many as six forecasts had errors in excess of 400 km. There were seven forecasts by hour 72 that had errors in excess of 600 km. These forecast errors describe the nature of the performance of the mesoscale suite of models that we have used.

We have noted that a consistent pattern of errors (from storm to storm) is reduced by the bias correction proposed here. The bias corrected ensemble mean reduces the track errors significantly as shown in Figs. 5–7 for 24, 48 and 72 h.

6. Three year summary of skill forecasts from the mesoscale model suite

Figure 8 shows the summary of track errors for all of the hurricanes of the year 2004, 2005 and 2006. The bias corrected ensemble mean shown as the last bar for these forecasts from 12 to 72 h carries the least error among all of the mesoscale member models and the ensemble mean for track forecasts. The HWRF provides some of the best track forecasts among the member models of the mesoscale suite. Overall the bias corrected ensemble mean improves the forecast performance for tracks by roughly 10–20 km. For intensity forecasts the bias corrected ensemble mean compares close to the best models (DSHP and GFDL) and the ensemble mean at 60 and 72 h. The DSHP, GFDL, ensemble mean and the bias corrected ensemble mean provide the best guidance for intensity forecasters. The differences in errors of intensity forecasts for these four models were within 1–2 m s^{-1}.

7. Concluding remarks

This paper reflects our first attempts at using a suite of mesoscale model towards improving experimental hurricane forecasts. Numerous modelling uncertainties still remain within each of the member models of the mesoscale suite. A considerable amount of further research is needed to find the best among the available physics and microphysics options within each mesoscale model for the specific application on hurricane forecasts. Numerous U.S. research groups have been addressing these issues. Taking one set of selected experimental options we were able to put together versions of these mesomodels described in Table 2. With those sets of models we completed forecast runs for most of the named hurricanes of the 2004, 2005 and the 2006 seasons (as shown in Table 1). All models share the same assimilated initial fields, provided by the HWRF group. These include the reconnaissance aircraft based dropwindsonde and flight level data sets in and around the hurricanes, these are brought into the HWRF initial state from the operational GFS data assimilation that includes these data sets.

In parallel to these experiments we also compiled hurricane forecasts from a suite of large-scale models. These were mostly the large-scale suite of operational hurricane forecast model used by the National Hurricane Center in their operations. We examined the absolute errors for track positions and maximum wind speed for all of the hurricanes for the years 2004 through 2006, in Table 1. This provided us with the opportunity to compare the consensus forecasts of the mesoscale suite of models with those of the large-scale suite and from a mix of the two suites of models. That information based on 58 storm forecast cases was quite revealing. We found that the best track forecasts, with least position errors, were obtained from the mix of the two suites of models. For intensity forecasts we found that forecasts through 48 h a mix of the two suites provided the least forecast errors. However for 60 and 72 h, the least error emerged from the mesoscale suite of models. This showed that the large-scale suite of models carried somewhat larger errors in the 48–72 h forecast time frame.

We have also compared the time histories of the track and intensity errors for each of the member model covering all of the years 2004, 2005 and 2006. We noted that some models consistently performed poorly while some others provided superior forecasts. There was also a random behaviour in these performances. This was revealed by the intersections of time history plots as a function of time. This was especially true when the errors were moderate. This type of non-systematic behaviour of intermodel skill is due to random errors arising from a number of factors that are not easy to remedy from the use of a bias corrected ensemble mean. The bias removed ensemble mean performs very well in reducing systematic errors. The statistics on tracks and intensity, we have presented here, all pertain to the Atlantic, Caribbean and Gulf of Mexico hurricanes. For other basins, Vijaya Kumar et al. (2003), we have noted some differences in the statistical weights for the member models, as should be expected. Emanuel (1999) concluded that after a storm reaches tropical-storm strength the intensity is dependent on the thermodynamics of the upper ocean layer and of the atmosphere

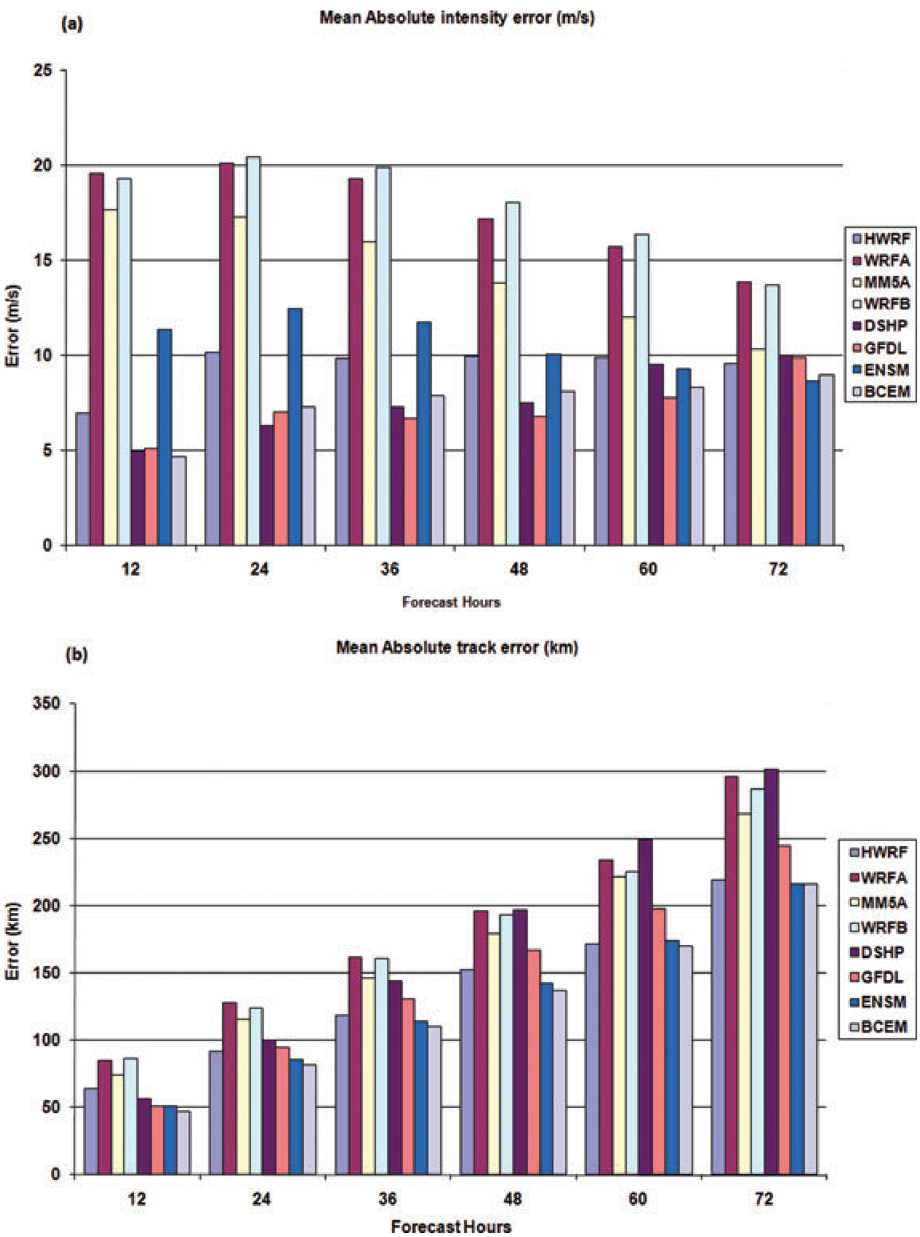

Fig. 8. Mean absolute errors for (a) intensity (in m s^{-1}) (b) track (in km) for member models, ensemble mean and bias corrected ensemble mean of the mesoscale suite of models.

along with other factors which are different for different basins and exhibits spatial and temporal variations.

This methodology can be extended to study precipitation forecasts after a hurricane makes landfall. Cartwright and Krishnamurti (2007) showed that higher skills can be obtained with a suite of mesosocale models for heavy precipitation studies including hurricanes. We shall address this issue in our future work.

8. Acknowledgments

The authors thank Drs. Richard Pasch, James Franklin, Chris Landsea, Mike Fiorino for their critical comments on this manuscript, those comments helped immensely in improving this paper. This work was supported by grants from NSF grants ATM-0553491 and ATM-0419618, SAIC Subcontract 4400105430 from NOAA, FIU Subcontract 120001515–01 from

NOAA, NASA grant NNG05GH81G. The in house model runs were done in Linux Cluster provided by HPC-FSU. Computations at Notre Dame were supported by the Northwest Indiana Computational Grid (NWICG).

9. Appendix A: Bias corrected ensemble mean methodology

The bias of a forecast F_{ij} (at a geographical location *i,j*) is given by $\overline{F_{ij}} - \overline{O_{ij}}$ where $\overline{O_{ij}}$ is observed value and the bar denotes a time mean. Then if a new forecast F_{Nij} is made then the bias corrected forecast is $F_{BCij} = F_{nij} + (\overline{F_{ij}} - \overline{O_{ij}})$. If there are n models then the bias corrected mean of the member models is given by $F_{BCEMij} = \sum_{k=1}^{n} \frac{1}{n}.F_{BCij})_k$.

References

Arakawa, A. and Schubert, W. H. 1974. Interaction of a cumulus cloud ensemble with the large-scale environment. Part I. *J. Atmos. Sci.* **31**, 674–701.

Bender, M. A., Ginis, I., Tuleya, R., Thomas, B. and Marchok, T. 2007. The operational GFDL coupled hurricane–ocean prediction system and a summary of its performance. *Mon. Wea. Rev.* **135**, 3965–3989.

Betts, A. K. 1986. A new convective adjustment scheme. Part I: observational and theoretical basis. *Quart. J. Roy. Meteor. Soc.* **112**, 677–691.

Betts, A. K. and Miller, M. J. 1986. A new convective adjustment scheme. Part II: single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.* **112**, 693–709.

Betts, A. K. and Miller, M. J. 1993. The Betts–Miller scheme. In: *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr*. 46, Amer. Meteor. Soc., 107–121.

Blackadar, A. K. 1976. Modelling the nocturnal boundary layer. In: *Preprints of Third Symposium on Atmospheric Turbulence and Air Quality, Raleigh, NC, 19–22 October 1976*. American Meteorological Society, Boston, 46–49.

Blackadar, A. K. 1979. High resolution models of the planetary boundary layer. In: *Advances in Environmental and Scientific Engineering* Volume 1 (eds J. R. Pfafflin and E. N. Zeigler). Gordon and Breach, London, 50–85.

Blumberg, A. F. and Mellor, G. L. 1987. A description of a three-dimensional coastal ocean circulation model. *Three-dimensional Coastal Ocean Models* Volume 4 (ed. N. Heaps). Amer. Geophys. Union, 1–16.

Cartwright, T. J. and Krishnamurti, T. N. 2007. Warm season mesoscale superensemble precipitation forecasts in the southeastern United States. *Wea. Fore.* **22**, 873–886.

Dudhia, J. 1989. Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.* **46**, 3077–3107.

Dunion, J. P. and Velden, C. S. 2004. The impact of the saharan air layer on atlantic tropical cyclone activity. *Bull. Am. Meteor. Soc.* **85**, 353–365.

Emanuel, K. A. 1999. Thermodynamic control of hurricane intensity. *Nature* **401**, 665–669.

Ferrier, B. S. 1994. A double-moment multiple-phase four-class bulk ice scheme. Part I: description. *J. Atmos. Sci.* **51**, 249–280.

Ferrier, B. S. 2005. An efficient mixed-phase cloud and precipitation scheme for use in operational NWP models. *EOS, Trans. Am. Geophys. Un.*, **86**, (Spring Meeting Suppl.), Abstract A42A-02.

Franklin, J. L. 2006. 2005 National Hurricane Center Forecast Verification Report, National Hurricane Center, Miami, FL (http://www.nhc.noaa.gov/verification/pdfs/Verification_2005.pdf).

Grell, G. A. 1993. Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.* **121**, 764–787.

Hong, S.-Y. and Lim, J. J.-O. 2006. The WRF single moment 6-Class Microphysics scheme (WSM6). *J. Korean Met. Soc.* **42**, 129–151.

Hong, S.-Y. and Pan, H.-L. 1996. Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.* **124**, 2322–2339.

Hong, S.-Y., Dudhia, J. and Chen, S.-H. 2004. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.* **132**, 103–120.

Houze, R. A., Chen, S. S., Smull, B. F., Lee, W.-C. and Michael, M. Bell 2007. Hurricane intensity and eyewall replacement. *Science* **315** (5816), 1235. doi:10.1126/science.1135650.

Janjic, Z. I. 1994. The step-mountain Eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.* **122**, 927–945.

Kain, J. S. and Fritsch, J. M. 1993. Convective parameterization for mesoscale models: the Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.* Volume 46. Amer. Meteor. Soc., 165–170.

Knaff, J. A., DeMaria, M., Sampson, C. R. and Gross, J. M. 2003. Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecast.* **18**, 80–92.

Krishnamurti, T. N. and Sanjay, J. 2003. A new approach to the cumulus parameterization issue. *Tellus* **55A**, 275–300.

Krishnamurti, T. N., Sanjay, J., Mitra, A. K. and Kumar, T. S. V. 2004. Determination of forecast errors arising from different components of model physics and dynamics. *Mon. Wea. Rev.* **132**, 2570–2594.

Krishnamurti, T. N., Pattnaik, S., Stefanova, L., Kumar, T. S. V. V., Mackey, B. P. and co-authors. 2005. The hurricane intensity issue. *Mon. Wea. Rev.* **133**, 1886–1912.

Kurihara, Y., Bender, M. A. and Ross, R. J. 1993. An initialization scheme of hurricane models by vortex specification. *Mon. Wea. Rev.* **121**, 2030–2045.

Kurihara, Y., Bender, M. A., Tuleya, R. E. and Ross, R. J. 1995. Improvements in the GFDL hurricane prediction system. *Mon. Wea. Rev.* **123**, 2791–2801.

Lin, Y. L., Farley, R. and Orville, H. D. 1983. Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.* **22**, 1065–1092.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J. and Clough, S. A. 1997. Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.* **102**, 16663–16682, D14.

Monin, A. S. and Obukhov, A. M. 1954. Basic laws of turbulent mixing in the surface layer of the atmosphere (in Russian)*Contrib. Geophys. Inst. Acad. Sci. USSR* **151**, 163–187.

Moon, I. J., Ginis, I., Hara, T. and Thomas, B. 2007. A Physics-Based Parameterization of Air-Sea Momentum Flux at High Wind Speeds and Its Impact on Hurricane Intensity Predictions. *Mon. Wea. Rev.* **135**, 2869–2878.

Pan, H.-L. and Wu, W.-S. 1995. Implementing a mass flux convection parameterization package for the NMC medium-range forecast model. NMC Office Note 409, 43 pp. Available at: NCEP, 5200 Auth Road, Washington, DC 20233..

Pattnaik, S. and Krishnamurti, T. N. 2007a. Impact on cloud microphysical processes on hurricane intensity, part 1: control run. *Met. Atmos. Phys.* **97**, 117–126.

Pattnaik, S. and Krishnamurti, T. N. 2007b. Impact on cloud microphysical processes on hurricane intensity, part 2: sensitivity experiments. *Met. Atmos. Phy.* **97**, 127–147.

Shay, L. K., Goni, G. J. and Black, P. G. 2000. Effects of a warm oceanic feature on hurricane opal. *Mon. Wea. Rev.* **128**, 1366–1383.

Simmons, A. J. and Burridge, D. M. 1981. An energy and angular-momentum conserving finite-difference scheme and hybrid vertical coordinates. *Mon. Wea. Rev.* 109, 758–766.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M. and co-authors. 2005. A description of the advanced research WRF version 2. NCAR Tech. Note 468+STR, 88 pp.

Tao, W. K. and Simpson, J. 1989. Modeling study of a tropical squall-type convective line. *J. Atmos. Sci.* **46**, 177–202.

Tao, W. K., Simpson, J., Sui, C. H., Ferrier, B., Lang, S. and co-authors. 1993. Heating, moisture, and water budgets of tropical and midlatitude squall lines: comparisons and sensitivity to longwave radiation. *J. Atmos. Sci.* **50**, 673–690.

Troen, I. and Mahrt, L. 1986. A simple model of the atmospheric boundary layer: sensitivity to surface evaporation. *Bound.-Layer Meteor.* **37**, 129–148.

Tuleya, R. E. and Gopalakrishnan, S. G. 2006. Hurricane Model Transitions to operations at NCEP/EMC. Year One Progress Report August 1, 2005 – June 30 2006. Available at: http://www.nhc.noaa.gov/jht/05–07reports/tuleya_yr1_annualrpt06.pdf.

Vijaya Kumar, T. S. V., Krishnamurti, T. N., Fiorino, M. and Nagata, M. 2003. Multimodel Superensemble Forecasting of Tropical Cyclones in the Pacific. *Mon. Wea. Rev.* **131**, 574–583.

Zhang, D. L. and Anthes, R. A. 1982. A high-resolution model of the planetary boundary layer—sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteorol.* **21**, 1594–1606.