

## Further Improvements to the Statistical Hurricane Intensity Prediction Scheme (SHIPS)

MARK DEMARIA

*NOAA/NESDIS, Fort Collins, Colorado*

MICHELLE MAINELLI

*Tropical Prediction Center, Miami, Florida*

LYNN K. SHAY

*RSMAS/MPO, University of Miami, Miami, Florida*

JOHN A. KNAFF

*Colorado State University/CIRA, Fort Collins, Colorado*

JOHN KAPLAN

*NOAA/Hurricane Research Division, Miami, Florida*

(Manuscript received 21 April 2004, in final form 27 October 2004)

### ABSTRACT

Modifications to the Atlantic and east Pacific versions of the operational Statistical Hurricane Intensity Prediction Scheme (SHIPS) for each year from 1997 to 2003 are described. Major changes include the addition of a method to account for the storm decay over land in 2000, the extension of the forecasts from 3 to 5 days in 2001, and the use of an operational global model for the evaluation of the atmospheric predictors instead of a simple dry-adiabatic model beginning in 2001.

A verification of the SHIPS operational intensity forecasts is presented. Results show that the 1997–2003 SHIPS forecasts had statistically significant skill (relative to climatology and persistence) out to 72 h in the Atlantic, and at 48 and 72 h in the east Pacific. The inclusion of the land effects reduced the intensity errors by up to 15% in the Atlantic, and up to 3% in the east Pacific, primarily for the shorter-range forecasts. The inclusion of land effects did not significantly degrade the forecasts at any time period. Results also showed that the 4–5-day forecasts that began in 2001 did not have skill in the Atlantic, but had some skill in the east Pacific.

An experimental version of SHIPS that included satellite observations was tested during the 2002 and 2003 seasons. New predictors included brightness temperature information from Geostationary Operational Environmental Satellite (GOES) channel 4 (10.7  $\mu\text{m}$ ) imagery, and oceanic heat content (OHC) estimates inferred from satellite altimetry observations. The OHC estimates were only available for the Atlantic basin. The GOES data significantly improved the east Pacific forecasts by up to 7% at 12–72 h. The combination of GOES and satellite altimetry improved the Atlantic forecasts by up to 3.5% through 72 h for those storms west of 50°W.

---

*Corresponding author address:* Mark DeMaria, NOAA/NESDIS/ORA, CIRA/CSU, W. Laporte Ave., Fort Collins, CO 80523.  
E-mail: Mark.DeMaria@noaa.gov

## 1. Introduction

The National Hurricane Center (NHC) has a suite of skillful global and regional prediction models available as guidance for tropical cyclone track prediction. However, operational intensity models have considerably less skill than the track models (DeMaria and Gross 2003). For intensity, the three primary models are the simple Statistical Hurricane Intensity Forecast (SHIFOR) model (Jarvinen and Neumann 1979; Knaff et al. 2003), which uses climatology and persistence to make a prediction; the Statistical Hurricane Intensity Prediction Scheme (SHIPS), which includes storm environmental conditions in addition to climatology and persistence; and the National Centers for Environmental Prediction (NCEP) version of the Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model (Kurihara et al. 1998). The SHIFOR model is primarily used as a benchmark for evaluating the skill of the official NHC intensity forecasts and those from the objective models. The GFDL model became operational in 1995, and a version that uses input from the U.S. Navy global model became available at NHC in 1999. The purpose of this paper is to document recent changes to the SHIPS model, evaluate its performance relative to the SHIFOR and GFDL models, and describe a recent effort to reduce the skill gap between the SHIPS intensity model and track models through the inclusion of new predictors from satellite observations.

The SHIPS model has been run operationally for storms in the Atlantic basin since 1991 and in the east Pacific basin since 1996. The model characteristics and its performance through 1997 were documented in DeMaria and Kaplan (1994a) and DeMaria and Kaplan (1999, hereafter DK99). Since 1997 significant modifications have been made including the addition of an adjustment factor to account for the decay over land, the extension of the forecasts from 3 to 5 days, and changes to the large-scale model from which many of the predictors are derived. Over the past several years SHIPS has become one of the primary guidance models that NHC uses for their operational intensity forecasts. For this reason there is a need to document the model changes relative to the version described by DK99 and to perform a verification of the operational SHIPS intensity forecasts.

The predictors for SHIPS include climatology and persistence, atmospheric environmental parameters (vertical shear, etc.), and sea surface temperature (SST), but contain relatively little information about the storm itself. Fitzpatrick (1997) showed that statistical intensity forecasts for western North Pacific tropical cyclones could be improved by including predictors

from infrared satellite data. In this paper, the potential for improving SHIPS using predictors from Geostationary Operational Environmental Satellite (GOES) infrared ( $10.7\ \mu\text{m}$ ) imagery will be evaluated.

The ocean influence on intensity change is included in SHIPS by evaluating the SST at the storm center. The SST is determined from the weekly  $1^\circ$  latitude–longitude analyses described by Reynolds and Smith (1993). The SST is then used to estimate the maximum potential intensity (MPI) from the empirical formula developed by DeMaria and Kaplan (1994b) for the Atlantic and by Whitney and Hobgood (1997) for the east Pacific. Many studies have shown that the upper-ocean heat content through the depth of the  $26^\circ\text{C}$  isotherm is important for tropical cyclone intensity change (e.g., Shay et al. 2000), and that the SST is not always a good measure of the total heat content. A method to estimate the oceanic heat content (OHC) from *European Remote Sensing Satellite-2 (ERS-2)* and *TOPEX/Poseidon* satellite altimetry observations was developed by Mainelli-Huber (2000), and an archive of oceanic heat content analyses for the prestorm environment of most Atlantic tropical cyclones from 1995 to the present was produced. Real-time OHC analyses were implemented at NHC beginning in 2002. The possibility of improving the SHIPS forecasts using OHC predictors will also be evaluated.

The modifications to the SHIPS model from 1997 to 2003 are described in section 2, and an evaluation of the model performance is presented in section 3. In section 4, the evaluation of a parallel version of SHIPS with predictors from GOES and satellite altimetry data is described.

## 2. The operational SHIPS model

The 1997 operational version of the SHIPS model was described in DK99. Beginning that year, SHIPS was converted to a statistical–dynamical model where some of the atmospheric predictors were determined from a numerical forecast model. A multiple linear regression was used to develop the model equations, where the dependent variable is the intensity change from 0 to 12, 0 to 24, . . . , or 0 to 72 h (separate regressions for each time interval), and the independent variables are parameters believed to be important for intensity change. All predictors must be significant at the 1% level (using a standard  $F$  statistic) for at least one forecast period. For forecast consistency, the same set of predictors is used at all forecast intervals. The inclusion of predictors that are not significant at the 1% level does not present a problem because all of the included predictors are significant at more than half of

the forecast intervals (e.g., for the 2003 model 83% of the coefficients passed the 1% significance test), and the magnitudes of the predictors tend to become small at the forecast intervals when they are not significant. The model for the east Pacific uses the same predictors as for the Atlantic version but with different coefficients. The model has undergone significant changes since 1997 as summarized in the remainder of this section.

#### a. Dependent sample

The 1997 model was developed from all storm cases from 1989 to 1996 that reached at least tropical storm strength. The sample was restricted to cases where the storm track remained over water for the entire interval for which the dependent variable was calculated. For each year since 1997 the cases from the previous year were added to the sample and the prediction coefficients were rederived. The sample for the 2003 season included all Atlantic storms from 1989 to 2002. As will be described in section 3, the rules that NHC uses to perform their annual verifications have changed over the years. The dependent sample for SHIPS was modified to remain consistent with these verification rules. For this purpose tropical depressions that never got strong enough to be named were added in 2001 and subtropical storms were added in 2003.

#### b. Track changes

The real-time forecasts in 1997 used the track from the operational Limited Area Sine Transform Barotropic (LBAR) model (Horsfall et al. 1997). This choice sometimes presented a problem because the LBAR track was not always consistent with the NHC official forecast track. Beginning in 1999 the LBAR track was replaced by the official NHC track forecast. Because the SHIPS model is usually run before the NHC track forecast is completed, the NHC track forecast from the previous forecast cycle (6 h old, but corrected to match the current storm position) is used. Beginning in 2001 the official NHC forecasts were extended from 72 to 120 h on an experimental basis. In 2003 the 5-day forecasts became operational. SHIPS was modified to provide intensity forecasts out to 120 h beginning in 2001.

#### c. Land correction

The SHIPS developmental sample excludes cases that crossed major landmasses. Beginning in 2000 the real-time forecasts over land were modified in a post-processing step using the simple empirical decay model described by Kaplan and DeMaria (1995). The land

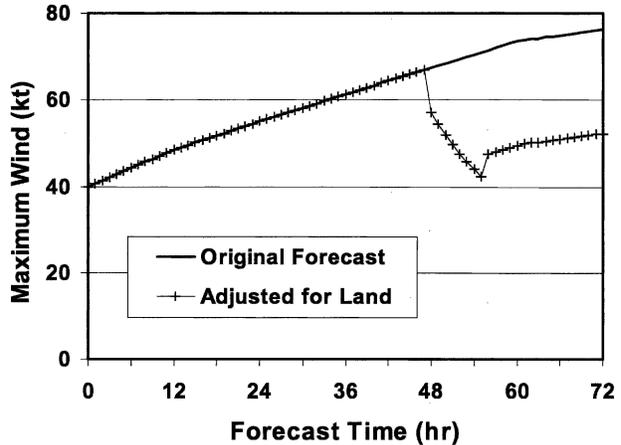


FIG. 1. The original SHIPS forecast and the forecast adjusted for movement over land for Hurricane Isidore beginning at 1200 UTC on 18 Sep 2002. Isidore crossed western Cuba between 48 and 55 h.

modification procedure begins with the SHIPS prediction without land effects. The forecast positions and intensities (every 12 h) are interpolated to 1-h intervals, and the overland portions of the track are identified. The empirical decay model is then applied to the part of the track that is over land. If the forecasted storm moves back over the water, the remaining portion of the prediction is adjusted so that the intensity changes are the same as for the unadjusted forecast. It should be pointed out that this method introduces some error for cases that move back over the water because the remainder of the prediction following the landfall is based upon statistical relationships that did not take into account the reduction in the intensity due to interaction with land. A method is being developed to correct this problem by reformulating the model with dependent variables that are the intensity change in 12-h intervals (0–12, 12–24, etc.) rather than from zero to the forecast time (0–12, 0–24, etc.).

Figure 1 shows the original and adjusted SHIPS 72-h forecasts for Hurricane Isidore beginning at 1200 UTC on 18 September 2002. The forecast track began in the western Caribbean, moved over western Cuba from 48 to 55 h, and then moved into the Gulf of Mexico. The decay model reduces the maximum winds by a factor of 0.9 at landfall to account for land – sea differences in surface roughness, and then applies an exponential decay formula for the portion of the track over land. When the track point is south of 36°N, the decay coefficients for the southeastern United States are applied and for points north of 40°N the coefficients for the New England area (Kaplan and DeMaria 2001) are used. A linear combination of the coefficients is used

TABLE 1. Predictors used in SHIPS 1997–2003. The predictors that are evaluated at the beginning of the forecast period are static (S), and predictors that are evaluated along the forecasted track of the storm are time dependent (T). An X indicates that the predictor was used in that year and a dash (—) indicates it was not used that year.

Predictor	Static (S) or time dependent (T)	Year						
		1997	1998	1999	2000	2001	2002	2003
1) Absolute value of (Julian day – peak season value)	S	X	X	X	X	X	X	—
2) Gaussian function of (Julian day – peak value)	S	—	—	—	—	—	—	X
3) Initial maximum winds	S	—	X	X	X	X	X	X
4) Max wind change during the past 12 h	S	X	X	X	X	X	X	X
5) Initial max winds times previous 12-h change	S	—	—	—	—	—	—	X
6) Zonal component of storm motion	S	—	X	X	X	X	X	X
7) Pressure level of storm steering	S	—	—	—	—	X	X	X
8) 200-hPa divergence	S	—	X	X	X	—	X	X
9) 200-hPa eddy momentum flux convergence	S	—	—	—	—	—	X	—
10) 200-hPa eddy momentum flux convergence	T	X	X	X	X	X	—	—
11) Max potential intensity – current intensity	T	X	X	X	X	X	X	X
12) 850–200-hPa vertical shear	T	X	X	X	X	X	X	X
13) 200-hPa zonal wind	T	X	X	X	X	X	X	—
14) 200-hPa temperature	T	X	X	X	X	X	X	X
15) 850–700-hPa relative humidity	T	—	—	—	—	X	X	—
16) 500–300-hPa relative humidity	T	—	—	—	—	—	—	X
17) 850-hPa relative vorticity	T	X	X	X	X	X	X	X
18) Surface – 200-hPa $\theta_e$ deviation of lifted parcel	T	—	—	—	—	—	—	X
19) Vertical shear times sine of storm latitude	T	X	X	X	X	X	X	X
20) Square of potential – current intensity	T	X	X	X	X	X	X	X
21) Initial intensity time shear	T	—	—	—	—	—	X	X

between these two latitudes. When the storm moves back over the water, the intensity is divided by 0.9, which accounts for the intensity increase at 55 h in the adjusted forecast in Fig. 1. Because the real-time track forecasts are not perfect, both the original (unadjusted) and adjusted SHIPS intensity predictions are provided to the forecasters. The forecasts adjusted for decay over land will be referred to as D-SHIPS.

#### d. Model fields

The SHIPS model utilizes a “perfect prog” approach (Kalnay 2003) where the predictors are calculated from analyses along the best-track positions in the dependent sample to develop the coefficients but applied using the NCEP global model along a forecast track in real time. For the 1997 model, operational analyses were used for the model development. Beginning in 2003 the operational analyses from 1989 to 2000 were replaced by the NCEP–National Center for Atmospheric Research (NCAR) reanalyses to obtain a more complete and consistent dataset.

When SHIPS was converted to a statistical–dynamical model in 1997, atmospheric predictors for the real-time runs were determined from a simple 10-layer dry-adiabatic prediction model where the storm circulation was filtered from the initial condition. This filtering procedure eliminated the difficulty of separat-

ing the storm from its environment during the model forecast. Because the 10-layer model had very limited accuracy, the predictors from this model were used only out to 48 h. When the SHIPS forecasts were extended to 5 days in 2001, the adiabatic model fields out to 48 h were replaced by 5-day forecasts from the NCEP global forecasting system (GFS).

#### e. Predictor changes

The SHIPS forecasts are verified at the end of every hurricane season. Based upon the model performance, modifications are made to the predictors almost every year. New predictors are tested if there is some physical reasoning behind them and they are added only if they pass the underlying 1% statistical significance test for their coefficients. Predictors are sometimes removed each year because they no longer pass the significance test. Modifications to the SHIPS predictors since 1997 are summarized below, and the physical reasoning for the new predictors is briefly described.

Some of the SHIPS predictors such as the initial storm intensity are “static,” in that they are evaluated only at  $t = 0$  h. The same value of a static predictor is used as input to the intensity change forecast for each time interval. Other predictors such as the vertical shear are time dependent, where they are averaged along the storm track. Table 1 summarizes the predic-

tors used in the operational SHIPS model from 1997 to 2003.

The 1997 predictors are described in detail in DK99. In 1998, three new static predictors were added: the initial maximum wind, the zonal component of the storm motion, and the 200-hPa divergence. The initial maximum wind provides information about the organizational stage of the storm. The zonal component of motion distinguishes between storms in easterly versus westerly basic currents. The divergence was added as a measure of synoptic-scale forcing. The divergence could have been included as a time-dependent predictor, but tests showed that there was a large difference between the perfect prog divergence and the divergence from the model forecasts. The 1998 predictors were also used in 1999 and 2000.

In 2001, one static predictor (the 200-hPa divergence) was eliminated and one was added, and one time-dependent predictor was added. During the 2000 Atlantic season there were two storms (Debby and Joyce) that experienced strong vertical shear and appeared to decouple in the vertical based upon their appearance in GOES satellite imagery. Although the environmental conditions later became more favorable, these two storms never reintensified. It also appeared that these storms began to move with the low-level flow after they decoupled in the vertical. To help account for highly sheared systems, a method was developed to find the weights on the horizontal winds from 1000 to 150 hPa surrounding the storm (200–800-km annular average) that provide the best match to the observed storm motion. The center of mass determined by these weights is referred to as the steering layer pressure (SLP). For storms that became decoupled in the vertical, the SLP occurs at a higher pressure, and it was found that this variable is negatively correlated with intensity change. Dunion and Velden (2004) used multispectral imagery from GOES to identify the Sahara air layer (SAL), which is characterized by very dry air in the 850–500-hPa layer. Their results show that the SHIPS model overforecasted intensification for storms that interacted with the SAL (including Joyce and Debby). To represent SAL effects in SHIPS, a low-level (850–700 hPa) relative humidity predictor from the NCEP global model was added. Although the results of Dunion and Velden suggest that the NCEP global analysis does not always represent the magnitude of the low-level drying associated with the SAL, this predictor was found to be statistically significant. An upper-level (500–300 hPa) humidity predictor was also tested but was not significant.

Only minor modifications were made to the predictors in 2002. The 200-hPa momentum flux was con-

verted from a time-dependent to a static predictor because the real-time estimates of the parameter were much larger than those from the perfect prog variables. This inconsistency may have been due to changes in the resolution of the GFS that were implemented that year. This change to a static predictor was implemented in mid-August of 2002 when the problem was first identified. The divergence predictor was again included in 2002 since it passed the significance test. One new predictor (the product of the initial intensity and shear) was added because the response of a strong storm to shear is different than that of a weak storm.

Several changes were made to the predictors in 2003. The 200-hPa momentum flux and zonal wind component were eliminated, one static predictor was added and one was modified, and one time time-dependent predictor was modified and another was added. The new static predictor was the product of the previous 12-h intensity change and the current intensity, which helps to reduce the contribution of the persistence term for storms that are already very intense. The form of the static Julian day predictor was modified because the original form overly penalized very early and very late season storms. The new Julian day predictor is given by  $\exp\{-[(J_d - P_d)/R_d]^2\}$ , where  $J_d$  is the Julian day,  $P_d$  is the peak day of the hurricane season from climatology (day 253 for the Atlantic, day 238 for the east Pacific), and  $R_d$  is a time scale that controls the width of the influence of this term. By experimentation it was found that  $R_d = 25$  days provided the best fit for the Atlantic and east Pacific cases. The time-dependent low-level humidity parameter was replaced by the upper-level humidity variable. This change was related to the use of the reanalysis fields in the dependent sample, which have different moisture properties than some of the older operational NCEP analyses. The poor performance of SHIPS for Tropical Storm Dolly in 2002 (and similar storms in the sample), which did not intensify despite what appeared to be a low-shear environment with fairly warm water, motivated the testing of a new time-dependent thermodynamic predictor. For this purpose the surface equivalent potential temperature ( $\theta_e$ ) is calculated and then assumed to be constant for a lifted parcel. The positive differences between the  $\theta_e$  of this lifted parcel and that of the original NCEP profile are then averaged from 1000 to 200 hPa. This variable is a crude representation of the atmospheric stability.

#### *f. Averaging areas*

The atmospheric predictors in Table 1 are averaged over horizontal areas. In the 1997 model the vertical shear and momentum fluxes were averaged from  $r = 0$  to 600 km, and the 200-hPa temperature, zonal wind,

and divergence, and the 850-hPa vorticity were averaged from  $r = 0$  to 1000 km. These variables are meant to measure the storm environment rather than the storm itself. This separation was not a problem when the 10-layer model was used for these predictors because the storm circulation was filtered from the initial and forecast fields. However, when the 10-layer model was replaced by the GFS in 2001, the separation between the storm circulation and the environment became problematic because the global model contains a representation of most tropical cyclones. This problem is further complicated by the fact that SHIPS uses the NHC official forecast track, which can differ significantly from the storm track in the GFS. The vertical shear is especially sensitive to these differences because the storm will make a substantial contribution to the shear if the center point for the calculation is not located at the storm center. To help alleviate this problem, the area used to calculate all of the atmospheric predictors (except 200-hPa divergence and momentum fluxes and 850-hPa vorticity) was modified to an annulus from a radius of 200–800 km. The size of the inner circle was based upon the typical differences between the official and GFS tracks out to 3 days. The radii for divergence and vorticity were maintained at 0–1000 km because these are even more sensitive to the storm circulation. The radii for momentum flux was kept at 0–600 km because the dependent sample showed that there is no relationship between the fluxes outside of 600 km and intensity change.

### 3. Validation of operational forecasts

A database of real-time forecasts from all guidance models (track and intensity) is maintained by NHC. A “best track” of positions and intensities is also produced from a postanalysis of all available information. Unless otherwise indicated, all verifications described below used the NHC model forecast and best-track archives. The intensity forecasts are evaluated by computing the mean absolute error in knots. The SHIPS forecast errors will also be compared to those from the SHIFOR and GFDL models, and to the official NHC intensity forecasts. The significance of the difference between average errors from different models is determined by the method described in DK99. Standard statistical procedures for comparing the means of two data samples are applied, where the sample size is reduced to account for serial correlation. The 95% level is the threshold for statistical significance.

SHIFOR uses climatology and persistence to forecast intensity, and will be used as a benchmark for evaluating the skill of other forecasts. If the average errors

from a given model are smaller than those from SHIFOR, then that model is considered to be skillful. The relative error (RE) defined by

$$RE = 100(E_{\text{shifor}} - E_{\text{model}})/E_{\text{shifor}}, \quad (1)$$

where  $E_{\text{model}}$  and  $E_{\text{SHIFOR}}$  are the mean absolute error from a given model and SHIFOR, respectively, will be used as a quantitative measure of skill. Here, RE is the percentage reduction of the model errors relative to SHIFOR, where positive values indicate forecast skill. When the NHC forecasts were extended to 5 days on a test basis in 2001, SHIFOR was updated and extended from 3 to 5 days (Knaff et al. 2003). The updated SHIFOR model will be referred to as SHIFOR5.

NHC performs a yearly validation of their official forecasts and their operational guidance models. For many years NHC restricted their verification sample to cases of at least tropical storm strength (34 kt), and extratropical and subtropical cases were excluded. Beginning in 2002, NHC expanded their verification rules to include all systems that were classified as at least a tropical depression in the best track. In addition, NHC began naming subtropical storms in 2002, using the same list of names previously restricted to purely tropical systems. For this reason, subtropical storms are included in the verification, but extratropical systems are still excluded. In the discussion below, the traditional NHC sample selection procedure will be referred to as the “old rules,” and the procedure implemented in 2002 (with depression stage and subtropical systems added) will be referred to as the “new rules.” The official NHC forecast is available at 12, 24, 36, 48, 72, 96, and 120 h, and all verification results are restricted to these times even though the SHIPS forecasts are also available at 60, 84, and 108 h.

To evaluate the long-term trends in the SHIPS skill, the Atlantic forecasts were verified using the old NHC rules for the periods 1991–96 and 1997–2003. These periods were chosen because SHIPS was converted to a statistical–dynamical model in 1997. The official NHC and the SHIFOR intensity forecasts were also included for comparison and for the calculation of RE.

Figure 2 shows that SHIPS did not have any significant skill at any forecast interval for the 1991–96 sample. In contrast, SHIPS has significant skill at all time intervals out to 72 h for the 1997–2003 sample. This result indicates that the SHIPS modifications described above have improved the performance of the model. Figure 2 also shows that the NHC official intensity forecasts for the 1991–96 sample were skillful only out to 36 h, but were skillful out to 72 h for the 1997–2003 sample. Thus, over the last several years, NHC has demonstrated improved intensity forecast skill, which is

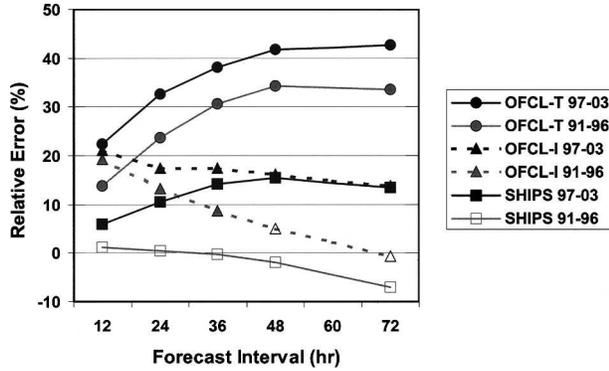


FIG. 2. The skill (relative to SHIFOR) of the SHIPS and official NHC intensity forecasts (OFCL-I) for 1991–96 and 1997–2003 Atlantic samples. The skill (relative to CLIPER) of the official NHC track forecasts (OFCL-T) is shown for reference. Solid (open) symbols indicate that the skill was (was not) statistically significant at the 95% level.

probably due to the availability of skillful model guidance.

Figure 2 also shows the skill of NHC official track forecasts [RE as determined by comparison of the average track errors to those from the climatology and persistence (CLIPER; Neumann 1972) track model]. For the 1997–2003 sample, the 12-h NHC track and intensity skill are comparable. By 36 h, the track skill is twice that of the intensity skill, and by 72 h it is three times greater. Thus, although some progress has been made in the ability to forecast tropical cyclone intensity change over the last decade, there is still a long way to go.

Figure 3 shows the RE of the SHIPS and NHC forecasts for the east Pacific with the old verification rules. It was not possible to show the RE for 1991–96 because SHIPS for the east Pacific was not developed until 1996. The SHIPS skill for the east Pacific is generally less

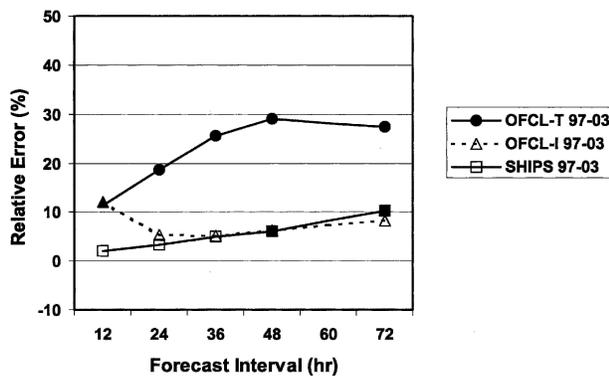


FIG. 3. Same as in Fig. 2 but for the eastern North Pacific for the 1997–2003 sample.

than that for the Atlantic, and was statistically significant only at 48 and 72 h. The skill of the NHC official track and intensity forecasts is also less than that for the Atlantic. These reductions in skill are probably due to the fact that the track and intensity of the east Pacific storms are generally better behaved than in the Atlantic and, thus, are easier to predict using climatology and persistence. Figure 3 also shows that, similar to the Atlantic, the track forecast skill is much greater than the intensity skill except at 12 h.

Figure 4 shows the yearly average SHIPS and NHC official (OFCL) intensity forecast errors from 1991 to 2003. As shown in Fig. 2, the skill of the SHIPS and NHC Atlantic intensity forecasts improved from 1991 to 2003. Figure 4 shows that there is also a downward trend in the mean absolute error, although it is not quite as obvious as the skill increase in Fig. 2. As described by McAdie and Lawrence (2000), it is sometimes difficult to detect long-term trends in tropical cyclone forecasts because the forecast difficulty varies from year to year. One way to account for forecast difficulty is to normalize the errors with respect to climatology and persistence forecasts. The relative error defined by (1) includes this normalization. However, even without the normalization, the slopes of the linear trend lines in Fig. 4 for SHIPS (OFCL) are  $-0.39$ ,  $-0.42$ , and  $-0.39$  ( $-0.33$ ,  $-0.32$ , and  $-0.47$ )  $\text{kt yr}^{-1}$  at 24, 48, and 72 h, indicating that the intensity forecasts have improved with time.

The only other intensity guidance model routinely available for the Atlantic and east Pacific is the GFDL primitive equation hurricane model, which became operational in 1995 (Kurihara et al. 1998). Figure 5 shows the relative error of the SHIPS and GFDL intensity forecasts for a homogeneous sample of cases from 1997 to 2003, using the old verification rules. The GFDL

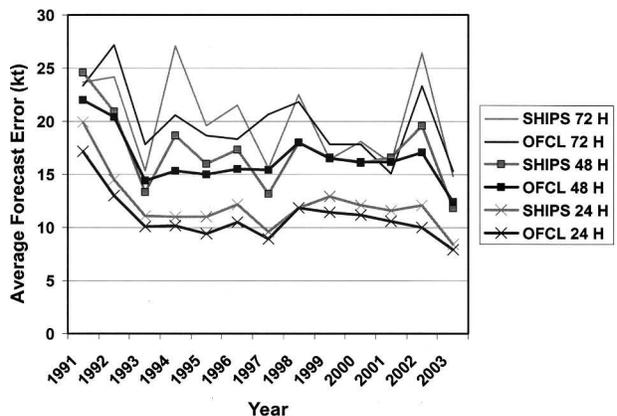


FIG. 4. The yearly average 24-, 48-, and 72-h intensity errors (kt) for the official NHC and SHIPS forecasts, 1991–2003.

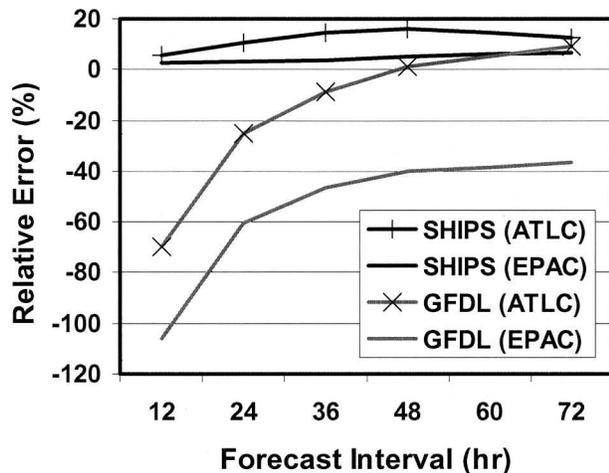


FIG. 5. The skill (relative error) for homogeneous samples of the SHIPS and NCEP/GFDL intensity forecasts for the Atlantic and eastern North Pacific for 1997–2003.

model did not have any forecast skill in either the east Pacific basin or the Atlantic basin through 48 h. The GFDL model had some skill at 72 h in the Atlantic, although it was not statistically significant at the 95% level. The lack of intensity skill of the GFDL model indicates the difficulty of initializing a primitive equation model with a strong circulation and where boundary layer and convective processes are of first-order importance. A number of GFDL model improvements are being implemented including new parameterization schemes and increased horizontal and vertical resolution. Thus, the lack of skill in Fig. 5 might not be representative of future versions of the GFDL model.

A method to account for the storm decay after landfall was added to SHIPS in 2000 (D-SHIPS). Although it might seem obvious that including the decay over land would improve the intensity forecasts, the operational forecast tracks are not perfect. Thus, there are errors in the timing and location of landfall. Also, the empirical inland decay model is a highly simplified representation of the physical processes that occur. Figure 6 shows the improvement of the D-SHIPS forecasts relative to the SHIPS forecasts for the Atlantic and east Pacific for 2000–03. For this comparison the new NHC verification rules were applied to increase the number of cases over land. Figure 6 shows that the inclusion of the effects of land improves the Atlantic forecasts by about 15% at 24–48 h. These improvements were statistically significant out to 72 h. In contrast, the east Pacific forecast improvements were 3% or less, and none was statistically significant. This result is not surprising since the east Pacific storms interact with land much less frequently than do Atlantic storms. The in-

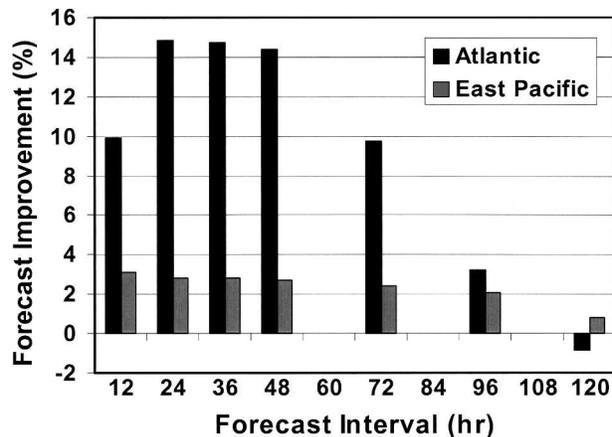


FIG. 6. The percent improvement (reduction in mean forecast error) of the decay version of SHIPS relative to the version without the decay factor. All cases from 2000 to 2003 are included using the “new” NHC verification rules.

clusion of land does not significantly improve the Atlantic forecasts at day 4 and results in small (but not statistically significant) degradation at 120 h. It is likely that the track forecast errors at these longer ranges introduce a source of error that cancels any improvement due to the inclusion of land effects.

The SHIPS forecasts were extended to 5 days beginning in 2001, using the same methodology as for 12–72 h. Figure 7 shows the mean absolute intensity errors for 2001–03 from D-SHIPS and SHIFOR5. For the Atlantic, the D-SHIPS intensity errors grow approximately linearly with time, but the SHIFOR5 errors level off after 72 h. By 5 days the Atlantic D-SHIPS errors are larger than SHIFOR5, indicating no skill. A possible

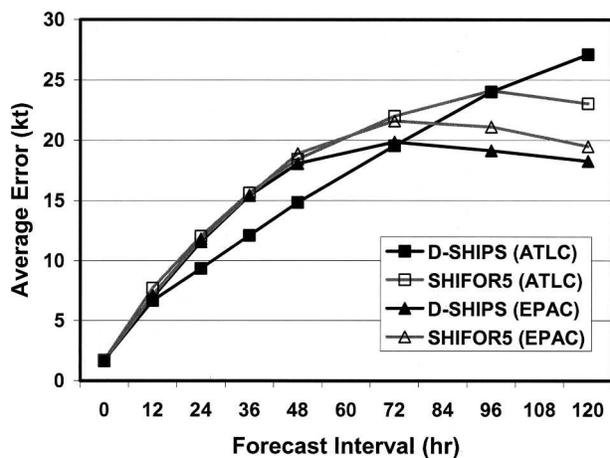


FIG. 7. The average forecast errors (kt) of the 5-day versions of the SHIFOR model and decay SHIPS model for a homogeneous sample from 2001 to 2003 for the Atlantic and eastern North Pacific.

TABLE 2. The mean absolute error (kt) and the 95th percentile of the absolute error (kt) of the Atlantic SHIFOR5, D-SHIPS, and official forecasts for the 2001–03 sample. Also shown are the relative errors (%) of the D-SHIPS and official mean and 95th percentile errors.

	Forecast interval (h)						
	12	24	36	48	72	96	120
<b>SHIFOR5</b>							
Mean error	7.7	12.0	15.6	18.4	22.0	24.1	23.0
95th percentile error	22	33	41	45	50	55	52
<b>D-SHIPS</b>							
Mean error	6.7	9.3	12.1	14.8	19.5	24.0	27.0
95th percentile error	18	25	31	36	46	58	62
Mean RE	13.6	22.3	22.5	19.5	11.1	0.4	−17.7
95th percentile RE	18.2	24.2	24.4	20.0	8.0	−5.5	−19.2
<b>Official</b>							
Mean error	5.6	9.0	11.4	14.2	18.2	19.9	21.2
95th percentile error	15	25	30	35	45	55	60
Mean RE	26.8	25.1	27.3	22.8	17.3	17.6	8.0
95th percentile RE	31.8	24.2	26.8	22.2	10.0	0.0	−15.3
Sample size	947	852	765	683	536	428	347

explanation for the continued growth of the D-SHIPS errors is that the track errors make a significant contribution to the intensity prediction. The median error of the 5-day Atlantic track forecast used by SHIPS was about 520 km, which is a significant fraction of the outer radius of the area over which most of the predictors are calculated. To further investigate the issue, the 5-day track errors were correlated with the 5-day intensity errors. This analysis confirmed that there is a positive correlation between the track and intensity errors.

In contrast to the Atlantic, the east Pacific D-SHIPS errors in Fig. 7 level off after 72 h in a manner similar to that of the SHIFOR5 errors, and are less than the SHIFOR5 errors at days 4 and 5, indicating forecast skill. The median 5-day track error for the east Pacific was 450 km, which is less than that of the Atlantic track. Thus, the impact of track errors on intensity errors is less in the east Pacific. Also, the most severe impact of the track error on the intensity forecast occurs when landfall is forecast but does not occur or vice versa. There are considerably fewer storms that move near land in the east Pacific, which may also explain why the model has skill at the longer time intervals in the east Pacific.

From a user perspective the very large intensity errors are the most problematic. To give a better idea of the tails of the distributions, the 95th percentile of the D-SHIPS, SHIFOR5, and official NHC errors were computed for the sample with 5-day forecasts (2001–03) using the new verification rules (Table 2). Also shown is the RE of the mean and 95th percentile of the D-SHIPS and official forecasts. This table shows that D-SHIPS improves upon the climatology and persistence mean error by up to 22.5% at 12–72 h, again indicating

forecast skill. The official forecasts were skillful out to 5 days. The RE of the 95th percentile error is similar to that of the mean error. Thus, D-SHIPS improves the most difficult forecasts by about the same percentage as the total forecast sample. Table 2 also shows that the 95th percentile D-SHIPS and official errors are quite large. This result again indicates that considerable improvement in intensity forecasting is still needed.

The 95th percentiles of the east Pacific forecasts were also examined for the 2001–03 sample using the new verification rules (not shown). Results were fairly similar to those from the Atlantic in that the percent improvements of the 95th percentile error were similar to the mean error improvements.

#### 4. Inclusion of satellite data

Most of the predictors in Table 1 are measures of the storm environment so there is little direct information about the storm itself. To provide additional storm information, an experimental version of SHIPS was developed that includes predictors from GOES infrared imagery. The GOES data should provide information about the structure of the deep convection near the storm center. Cloud-top brightness temperatures do not always directly relate to convective updrafts on a “pixel by pixel” basis due to sloping eyewalls, cirrus outflow, and other factors. However, validation of infrared-based precipitation methods indicates that there is a statistical relationship between the area-integrated cold cloud tops and precipitation in the storm inner core (e.g., Scofield et al. 2001). Also, the lack of cold cloud-top temperatures is well correlated with the lack of deep convection.

The GOES data were obtained from the archive maintained by the Cooperative Institute for Research in the Atmosphere (CIRA). Since 1995, channel 4 infrared ( $10.7\ \mu\text{m}$ ) imagery from GOES-East and GOES-West for nearly all Atlantic and east Pacific tropical cyclones has been collected (Zehr 2000). The channel 4 brightness temperatures ( $T_B$ ) were azimuthally averaged on a 4-km, storm-centered radial grid. The  $T_B$  standard deviations from the azimuthal average were also calculated at each radius. In all cases, the GOES data were within 1 h of the time from the corresponding SHIPS case.

Several studies have shown that the oceanic heat content (OHC) relative to the depth of the  $26^\circ\text{C}$  isotherm is important for tropical cyclone intensity change (e.g., Shay et al. 2000). Through the retrieval and analysis of high-resolution blended surface height anomaly (SHA) altimetry data, the Reynolds SST weekly analysis, and in combination with a hurricane season (June–November) averaged oceanic climatology, the OHC relative to  $26^\circ\text{C}$  water was estimated by applying the approach described by Mainelli-Huber (2000). OHC analyses in the prestorm environments of most Atlantic cases were available back to 1995 over a domain that extends from  $0^\circ$  to  $40^\circ\text{N}$  and  $100^\circ$  to  $50^\circ\text{W}$  and over the entire Atlantic basin since 2002. OHC analyses are not yet available for the eastern Pacific.

The incorporation of SHA altimetry data into OHC estimations has changed over the last decade. Prior to 1999 only TOPEX/Poseidon (T/P) altimeter data were operationally accessible. The T/P satellite provides a 10-day repeat cycle; however, data are spaced up to  $3^\circ$  apart in longitude. Earlier studies by Mainelli-Huber (2000) revealed that the poor spatial resolutions would, at times, cause a misrepresentation of oceanic features. During the summer of 2002 the T/P was replaced by *Jason-1*, which provides the same temporal and spatial resolution.

The *European Remote Sensing Satellite-2 (ERS-2)* radar altimetry data became available in 1999. The *ERS-2* provided a higher spatial resolution dataset with a 35-day repeat cycle. In late June of 2003, the *ERS-2* became inoperable. However, most of the experimental SHIPS runs occurred while both T/P (or *Jason-1*) and *ERS-2* were available. The 10-day T/P (or *Jason-1*) and 15-day *ERS-2* data were objectively analyzed following the approach of Mariano and Brown (1992), which provides estimates on an evenly spaced  $0.5^\circ$  latitude–longitude grid. Currently, the Geosat Follow-On (GFO) altimeter data (with a 17-day repeat cycle) have become fully operational and have replaced the *ERS-2*. All OHC estimates incorporated into SHIPS are now a blend of *Jason-1* and GFO. The real-time OHC analy-

ses are updated daily at NHC using the previous 10 days of altimetry data over a domain that includes the entire Atlantic basin to  $60^\circ\text{N}$ .

If the satellite predictors were combined with the other SHIPS predictors, it would be necessary to reduce the sample size because these new data were only available back to 1995, and over only a limited part of the Atlantic basin. To overcome this problem, a two-step prediction procedure is applied. In the first step the SHIPS model with the full data sample is derived as before. Then, the difference between the SHIPS predictions and the observed intensity changes are used as the dependent variables in a second multiple regression, where parameters from the GOES and altimetry data are the independent variables. This second regression will be referred to as the perturbation model.

The perturbation model includes two GOES predictors and one OHC predictor. The GOES predictors are 1) the percent of the area (pixel count) from 50 to 200 km from the storm center where  $T_B$  is colder than  $-20^\circ\text{C}$  and 2) the standard deviation of  $T_B$  (relative to the azimuthal average) averaged from 100 to 300 km. Several other GOES predictors were tested, but the two described above explained the most variance in the perturbation model. The GOES standard deviation predictor is actually the product of the  $T_B$  standard deviation and the initial intensity. This variable provided a better fit in the perturbation model and helps account for the variability of the  $T_B$  standard deviation as a function of storm intensity. The altimetry predictor is the OHC above  $50\ \text{KJ cm}^{-2}$ , averaged along the storm track. A better fit of the perturbation model was obtained using this threshold. In fact, the OHC predictor was not statistically significant without the threshold. The GOES parameters are static predictors and the altimetry parameter is a time-dependent predictor. The OHC generally exceeds  $50\ \text{KJ cm}^{-2}$  in the Caribbean, the Loop Current, in warm-core eddies in the Gulf of Mexico, and along the Gulf Stream. For the eastern Pacific model, only the two GOES predictors were included in the perturbation model.

The perturbation model was run in real time during the 2003 hurricane season for all Atlantic and east Pacific forecast cases, but was not ready in time for the 2002 season. In the 2002–03 evaluations described below, the 2003 cases are real-time forecasts and the 2002 cases are postseason runs that used input that would be available in real time (GFS forecasts and NHC operational tracks and initial intensity estimates).

Figure 8 shows the percent improvement in the D-SHIPS forecasts for the 2002–03 samples due to the satellite data. For the east Pacific, the inclusion of the GOES predictors improves the forecasts by as much as

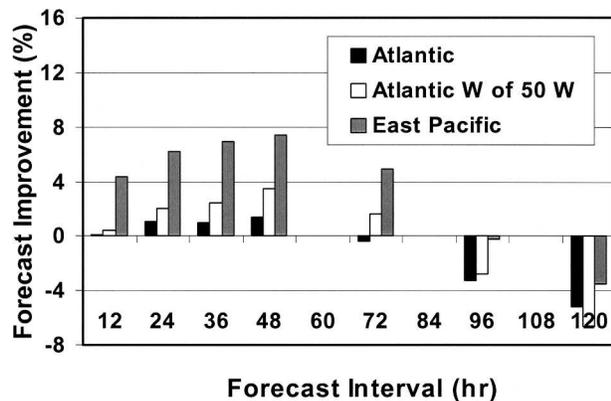


FIG. 8. The percent improvement of the SHIPS intensity forecasts with satellite input relative to the operational version without the satellite data for the 2002–03 samples in the east Pacific, the Atlantic, and the Atlantic for storms initially west of 50°W.

7.4% at 12–72 h. All of these improvements were statistically significant. By 120 h, the east Pacific forecasts with the GOES data were slightly degraded, but the differences in the average errors were not statistically significant. In contrast to the east Pacific, the inclusion of the satellite data (GOES and OHC predictors) in the Atlantic resulted in only a very slight improvement at 12–48 h and some degradation at the longer forecast intervals. None of the Atlantic improvements was statistically significant.

Because the GOES data improved the east Pacific forecasts but the combination of OHC and GOES data had a neutral effect on the Atlantic forecasts, it might be suspected that the OHC predictor was canceling the positive impact of the GOES data on the Atlantic forecasts. To investigate this possibility, all of the Atlantic forecasts were rerun with the contributions to the perturbation intensity changes from the OHC predictor set to zero. This procedure was repeated but with the contributions from the GOES predictors set to zero. Because of potential cancellation between the impacts of the OHC and GOES predictors, the sum of the improvements from the forecasts with only GOES and only OHC predictors does not add up to the improvements when both are included. However, this analysis illustrates whether one or the other of the predictors is dominating the improvements or degradations. Results showed that the improvements in the short-range Atlantic forecasts (12–24 h) and degradations at the longer periods were roughly evenly divided between the contributions from the GOES and OHC predictors. Thus, it does not appear that the OHC predictors were interfering with the GOES predictors.

The perturbation SHIPS model for the Atlantic was developed from cases that were nearly all west of 50°W

due to the limitation in the size of the OHC analysis area. However, the perturbation model was applied to all Atlantic cases in the operational forecasts from 2002 to 2003. To determine if this mismatch between the developmental and operational samples had any impact, the 2002–03 forecasts were verified for only those cases where the storm was west of 50°W. The improvements for this subsample in Fig. 8 are nearly twice as large as for the total sample at 12–48 h, although the forecasts were still degraded at 96 and 120 h. Similar to the total sample, the GOES and OHC data contributed about equally to the short-range improvements and longer-range degradations. This result suggests that an effort should be made to obtain OHC analyses over the entire Atlantic basin or restrict the application of the perturbation model to the same area as used to develop the model. Efforts are under way to obtain the climatology of OHC analyses over the entire Atlantic basin for the developmental sample.

Figure 8 shows that the satellite data degraded the Atlantic and east Pacific forecasts at 4 and 5 days. The version of SHIPS with the satellite data included became operational for the 2004 season. Results from a preliminary verification for the Atlantic and east Pacific (through the end of October 2004) are very encouraging and are even better than what was shown in Fig. 8. The improvements with the satellite data are similar to those in Fig. 8 out to 72 h, but there is no negative impact at 96 and 120 h. The developmental datasets for the 2004 perturbation model had much larger sample sizes at 4 and 5 days due to the addition of the 2002 and 2003 cases, which likely helped to eliminate the degradation of the longer-range forecasts.

## 5. Concluding remarks

Modifications to the NHC operational SHIPS intensity model for each year from 1997 to 2003 were described. Major changes include the addition of a method to account for the storm decay over land in 2000, the extension of the forecasts from 3 to 5 days in 2001, and the replacement of a simple dry-adiabatic model with the NCEP operational global model for the evaluation of the atmospheric predictors, beginning in 2001. Several new predictors were also described. Verification of the SHIPS forecasts showed the following results: 1) for the period 1997–2003, the SHIPS forecasts had statistically significant skill (relative to forecasts based upon climatology and persistence) out to 72 h in the Atlantic, and at 48 and 72 h in the east Pacific; 2) the operational NCEP dynamical intensity forecast (GFDL) model did not have any significant forecast skill for this same period; 3) The 4- and 5-day SHIPS

forecasts that began in 2001 were not skillful in the Atlantic but showed some modest skill in the east Pacific; 4) the inclusion of the effects of the decay over land beginning in 2000 reduced short-range Atlantic (East Pacific) intensity errors by up to 15% (3%) but had a neural impact after 72 h; 5) the mean and 95th percentiles of the SHIPS intensity errors were documented.

An experimental version of SHIPS that included satellite observations was tested during the 2002 and 2003 seasons. New predictors included brightness temperature information from GOES channel 4 (10.7  $\mu\text{m}$ ) imagery, and oceanic heat content (OHC) estimates inferred from satellite altimetry observations. The OHC estimates were only available for the Atlantic basin. The GOES data significantly improved the east Pacific forecasts by up to 7% at 12–72 h. The combination of GOES and satellite altimetry improved the Atlantic forecasts by up to 3.5% through 72 h for those storms west of 50°W.

To provide perspective on the current state of intensity forecasting, the skill of the NHC official intensity forecasts was compared with the track forecast skill. The Atlantic intensity and track skill are comparable at 12 h, but by 36 h (72 h) the track skill is twice (three times) the intensity forecast skill. Thus, intensity forecasting still has a long way to go.

Several modifications are planned for the SHIPS model to improve its performance. Methods are being developed to include predictors from aircraft flight-level observations. It is anticipated that the inner-core wind field structure information will provide additional short-range forecast skill for those storms that have potential to make landfall along the U.S. coastline. The GOES data are currently included in a very simple way (pixel counts and brightness temperature standard deviations averaged over large areas). The radial patterns of the GOES data near the storm center are being examined in greater detail using an empirical orthogonal function (EOF) approach. A daily SST analysis with a 50-km horizontal resolution under development at NCEP is being compared with the weekly 100-km Reynolds's SST analyses currently used by SHIPS (Berg et al. 2004). In addition, experimental daily SST analyses from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) instrument are showing potential for improving the SHIPS forecasts, especially in the east Pacific basin. Another potential use of satellite observations is to improve the analysis of the storm thermodynamic environment. Dunion and Velden (2004) have shown that total precipitable water analysis from microwave imagery has the potential to provide a better method for identifying

the SAL than current global model analyses. In the longer term, new sounders planned for the next-generation National Oceanic and Atmospheric Administration (NOAA) polar-orbiting and geostationary satellite systems should also be of utility. Finally, neural network techniques are being explored to determine if they can improve the SHIPS forecasts, as suggested by Baik and Paek (2000) for the western North Pacific.

*Acknowledgments.* This work was partially supported by a grant to CIRA from the NOAA U.S. Weather Research Program Joint Hurricane Testbed. L. K. Shay was supported by the National Science Foundation through Grant ATM-01-08218. Tom Cook helped to develop the real-time satellite-derived heat content analyses and Inger Solheim and James Kossin assisted with the processing of the GOES data. Jack Dostalek, Miles Lawrence, Richard Pasch, Fran Holt, and three anonymous reviewers provided useful comments on an earlier version of the manuscript. The views, opinions, and findings in this report are those of the authors and should not be construed as an official NOAA and/or U.S. government position, policy, or decision.

#### REFERENCES

- Baik, J.-J., and J.-S. Paek, 2000: A neural network model for predicting typhoon intensity. *J. Meteor. Soc. Japan*, **78**, 857–869.
- Berg, R., C. Sisko, and M. DeMaria, 2004: High resolution SST in the SHIPS model: Improving operational guidance of tropical cyclone intensity forecasts. Preprints, *26th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 536–537.
- DeMaria, M., and J. Kaplan, 1994a: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209–220.
- , and —, 1994b: Sea surface temperature and the maximum intensity of Atlantic tropical cyclones. *J. Climate*, **7**, 1324–1334.
- , and —, 1999: An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **14**, 326–337.
- , and J. M. Gross, 2003: Evolution of tropical cyclone forecast models. *Hurricane! Coping with Disaster*, R. Simpson, Ed., Amer. Geophys. Union, 103–126.
- Dunion, J. P., and C. S. Velden, 2004: The impact of the Saharan air layer (SAL) on Atlantic tropical cyclone activity. *Bull. Amer. Meteor. Soc.*, **85**, 353–364.
- Fitzpatrick, P. J., 1997: Understanding and forecasting tropical cyclone intensity change with the Typhoon Intensity Prediction Scheme (TIPS). *Wea. Forecasting*, **12**, 826–846.
- Horsfall, F. M., M. DeMaria, and J. M. Gross, 1997: Optimal use of large-scale boundary and initial fields for limited-area hurricane forecast models. Preprints, *22d Conf. on Hurricanes and Tropical Meteorology*, Fort Collins, CO, Amer. Meteor. Soc., 571–572.
- Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of

- tropical cyclone intensity change. NOAA Tech. Memo. NWS NHC-10, 22 pp.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.
- Kaplan, J., and M. DeMaria, 1995: A simple empirical model for predicting the decay of tropical cyclone winds after landfall. *J. Appl. Meteor.*, **34**, 2499–2512.
- , and —, 2001: On the decay of tropical cyclone winds after landfall in the New England area. *J. Appl. Meteor.*, **40**, 280–286.
- Knaff, J. A., M. DeMaria, C. R. Sampson, and J. M. Gross, 2003: Statistical 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92.
- Kurihara, Y., R. E. Tuleya, and M. A. Bender, 1998: The GFDL Hurricane Prediction System and its performance in the 1995 hurricane season. *Mon. Wea. Rev.*, **126**, 1306–1322.
- Mainelli-Huber, M., 2000: On the role of the upper ocean in tropical cyclone intensity change. M.S. thesis, Dept. of Meteorology and Physical Oceanography, University of Miami, 73 pp.
- Mariano, A. J., and O. B. Brown, 1992: Efficient objective analysis of heterogeneous and nonstationary fields via parameter matrix. *Deep-Sea Res.*, **7**, 1255–1271.
- McAdie, C. J., and M. B. Lawrence, 2000: Improvements in tropical cyclone track forecasting in the Atlantic basin, 1970–98. *Bull. Amer. Meteor. Soc.*, **81**, 989–998.
- Neumann, C. J., 1972: An alternate to the HURRAN tropical cyclone forecast system. NOAA Tech. Memo. NWS SR-62, 22 pp.
- Reynolds, R. W., and T. M. Smith, 1993: An improved real-time global sea surface temperature analysis. *J. Climate*, **6**, 114–119.
- Scotfield, R. A., M. DeMaria, and R.M del Alfaro, 2001: Space-based rainfall capabilities in hurricanes offshore and inland. Preprints, *Symp. on Precipitation Extremes: Prediction, Impacts and Responses*, Albuquerque, NM, Amer. Meteor. Soc., 297–301.
- Shay, L. K., G. J. Goni, and P. G. Black, 2000: Effects of a warm oceanic feature on Hurricane Opal. *Mon. Wea. Rev.*, **128**, 1366–1383.
- Whitney, L. D., and J. S. Hobgood, 1997: The relationship between sea surface temperatures and maximum intensities of tropical cyclones in the eastern North Pacific Ocean. *J. Climate*, **10**, 2921–2930.
- Zehr, R. M., 2000: Tropical cyclone research using large infrared data sets. Preprints, *24th Conf. on Hurricanes and Tropical Meteorology*, Fort Lauderdale, FL, Amer. Meteor. Soc., 486–487.