NCEP NOTES

Statistical, 5-Day Tropical Cyclone Intensity Forecasts Derived from Climatology and Persistence

JOHN A. KNAFF

CIRA/Colorado State University, Fort Collins, Colorado

MARK DEMARIA

NOAA/NESDIS, Fort Collins, Colorado

CHARLES R. SAMPSON

Naval Research Laboratory, Monterey, California

JAMES M. GROSS

Tropical Prediction Center, Miami, Florida

26 March 2002 and 6 September 2002

ABSTRACT

Tropical cyclone track forecasting has improved recently to the point at which extending the official forecasts of both track and intensity to 5 days is being considered at the National Hurricane Center and the Joint Typhoon Warning Center. Current verification procedures at both of these operational centers utilize a suite of control models, derived from the ''climatology'' and ''persistence'' techniques, that make forecasts out to 3 days. To evaluate and verify 5-day forecasts, the current suite of control forecasts needs to be redeveloped to extend the forecasts from 72 to 120 h. This paper describes the development of 5-day tropical cyclone intensity forecast models derived from climatology and persistence for the Atlantic, the eastern North Pacific, and the western North Pacific Oceans. Results using independent input data show that these new models possess similar error and bias characteristics when compared with their predecessors in the North Atlantic and eastern North Pacific but that the west Pacific model shows a statistically significant improvement when compared with its forerunner. Errors associated with these tropical cyclone intensity forecast models are also shown to level off beyond 3 days in all of the basins studied.

1. Introduction

Since 1964, the National Hurricane Center (NHC) has been producing operational tropical cyclone (TC) forecasts of track and intensity (maximum 1-min sustained wind speed to the nearest 5 kt) through 72 h in the Atlantic tropical cyclone basin. In 1988, NHC also began issuing tropical cyclone forecasts in the eastern North Pacific tropical cyclone basin. Before 1988, the San Francisco Weather Service Forecast Office issued eastern North Pacific hurricane forecasts (Sheets 1990). Likewise, the Joint Typhoon Warning Center (JTWC) makes track and intensity forecasts through 72 h in the Southern Hemisphere, north Indian Ocean, and the western North Pacific. Many of the current guidance models used by forecasters at NHC and JTWC provide forecasts of track and intensity out to and beyond 5 days. A common approach to assessing the merit and skill of TC forecasts is to compare them with what is referred to as a ''control'' forecast derived from a combination of climatology and persistence (CLIPER). [See Panofsky and Brier (1968) for a basic discussion about meteorological forecast verification. Climatology, as used here, refers to a forecast based on climatological averages, and persistence is a forecast that the current conditions will continue.] There currently is a set of control forecast models designed for making intensity forecasts out to 3 days. To evaluate and to verify 5-day intensity forecasts, updated versions of these CLIPER-

Corresponding author address: John A. Knaff, NOAA/Cooperative Institute for Research in the Atmosphere, Colorado State University, Foothills Campus, Fort Collins, CO 80523-1375.
E-mail: knaff@cira.colostate.edu

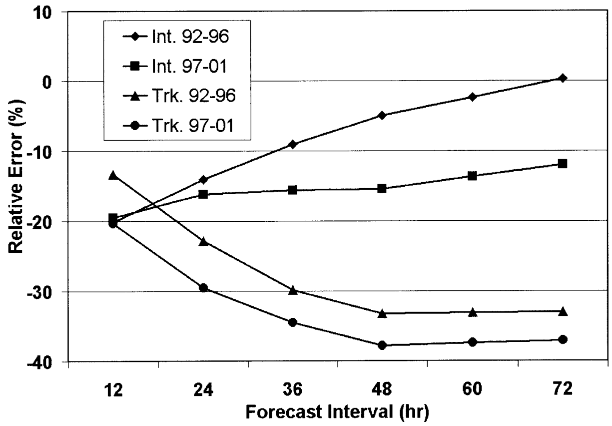

FIG. 1. Relative error of the official NHC track (Trk.) and intensity (Int.) forecasts for 1992–96 and 1997–2001. The zero line represents a CLIPER-based control forecast, so that negative relative errors represent skillful forecasts.

based intensity models that make forecasts through 120 h are needed.

Operational track forecasts have shown considerable improvement over the past several decades (McAdie and Lawrence 2000). In recent years, the operational track forecasts improve upon those from CLIPER-type models by about 30%–40% at 3 days (e.g., Gross 1999) and, thus, have considerable forecast skill. These recent improvements in tropical cyclone track forecasting skill have prompted the U.S. Navy and other agencies to request that NHC and JTWC provide TC guidance beyond 72 h (OFCM 2001). As a result of this request, both NHC and JTWC have developed plans to provide forecasts through 5 days.

Intensity forecasting skill lags that of track forecasting (DeMaria and Kaplan 1999), although some progress has been made in the past few years. Figure 1 shows the official NHC track and intensity forecast errors relative to their respective CLIPER control forecasts [3-day statistical hurricane intensity forecast (SHIFOR; Jarvinen and Neumann 1979) for intensity, CLIPER (Neumann 1972) for track], indicated by the zero line, for two periods: 1992–96 and 1997–2001. In this figure, greater skill is indicated by larger negative values of relative error. This figure shows that the intensity forecast skill is considerably less than that for track but that the skill has improved over the past decade. These results come from NHC's Atlantic basin forecasts; a similar situation exists in the eastern and western North Pacific (Gross 1999; JTWC 2002a). Given that there is some intensity forecast skill at 3 days, NHC and JTWC are also evaluating the extension of their intensity forecasts to 5 days. The creation of updated 5-day versions of CLIPER-based control models of TC intensity is an integral part of these plans, and these models will be used to evaluate the skill of the operational 5-day intensity predictions.

This purpose of this paper is to describe the development of the 5-day versions of the CLIPER-type intensity models for the Atlantic, eastern North Pacific (including the central Pacific), and the western North Pacific tropical cyclone basins. The use of climatology combined with persistence to make control tropical cyclone track and intensity forecasts has a long history (Neumann 1972; Jarvinen and Neumann 1979; Merrill 1980; Chu 1994; Aberson 1998). These models typically have been developed using data that describe the past history of tropical cyclones to create a set of multiple linear regression equations for the prediction. Existing 3-day CLIPER-based intensity forecast models include SHIFOR for the Atlantic and eastern North Pacific (Jarvinen and Neumann 1979) and the statistical typhoon intensity forecast (STIFOR) described in Chu (1994). It is important to note that while these models are used to assess intensity forecasting skill they also, at times, are used to provide real-time forecast guidance because of the difficulty involved in TC intensity forecasting. Expanding on these past efforts, this paper documents the development, statistical characteristics, and performance of 5-day tropical cyclone intensity forecast models following the CLIPER approach. The databases are described in section 2, the statistical development is presented in section 3, and the new models are evaluated in section 4.

2. Datasets

The historical track and intensity information used to develop updated versions of SHIFOR and STIFOR come from the ''best track'' datasets for each basin. NHC's best-track digital database is described by Davis et al. (1984) for the eastern North Pacific and Jarvinen et al. (1984) for the Atlantic. These databases contain date, time, location, and intensity of tropical cyclones and subtropical cyclones that reached an intensity of greater than 34 kt. The intensity archived in these historical datasets as well as operational intensity forecasts are recorded to the nearest 5 kt (2.58 m s^{-1}) at 6-h intervals. For this reason, model formulation as well as any discussion of intensity in this paper will be given in units of knots (1 kt = 0.52 m s^{-1}). Also, for the purposes of this paper, the definitions of tropical depression, tropical storm, subtropical storm, hurricane, and typhoon will follow those in Elsberry (1987). The term ''nondeveloping depression'' refers to a tropical depression that never reached tropical storm or subtropical storm status.

The current operational version of SHIFOR used at NHC was developed using best-track data for the period of 1967–87 and included all tropical cyclones and nondeveloping depressions in the Atlantic basin whose initial position is west of 5°E and south of 45°N. An examination of the number of nondeveloping depressions over 1967–99 shows that the number has declined over the past decade or so (Fig. 2). This change is likely the result of changes in operational procedures related to

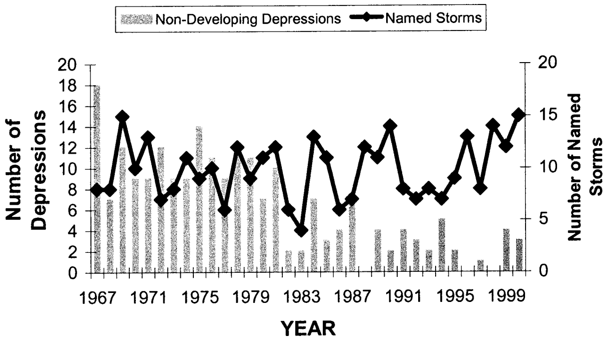

FIG. 2. Time series of the number of annual nondeveloping depressions and named storms in the Atlantic tropical cyclone basin. Note the decrease in numbers of nondeveloping depressions with time while the trend of named storms shows little overall trend.

classification of weaker systems and not of atmospheric trends in weak tropical cyclones. For this reason, the development of a 5-day version of Atlantic SHIFOR uses best-track data from 1967 to 1999 but does not include nondeveloping depressions. As a result, all of the tropical cyclones in the developmental dataset eventually reach an intensity of 35 kt or stronger, becoming tropical storms, subtropical storms, or hurricanes. In addition, to remove the potential influence of land, all cases within 50 km of land were excluded. The choice of the starting date (1967) also minimizes potential effects of positive intensity biases documented in major hurricanes (Landsea 1993).

In the eastern North Pacific, the original 3-day version of SHIFOR used best-track data for the period of 1967–87, including nondeveloping depressions. The development of a 5-day SHIFOR in this basin uses the best-track data starting in 1975 and continuing through 1999, but without the nondeveloping depressions. One justification for this change is that routine aerial reconnaissance has never been routinely available in this tropical cyclone basin and reliable techniques for satellite intensity estimates (Dvorak 1972, 1975, 1984) were not used operationally until the mid-1970s (Neumann et al. 1993). For this reason, the estimated intensities of tropical cyclones, especially for weaker and very intense systems and thus the intensity-change climatology, is likely less reliable prior to 1975. Also, a record of nondeveloping depressions in this basin does not exist prior to 1989. To expand the developmental dataset to include the central Pacific basin, initial positions were allowed to be east of 165°W and south of 35°N. To limit the possible effects of land on intensity, cases are excluded if the tropical cyclone is within 50 km of land.

The digital database that contains historical track and intensity estimates for the western North Pacific is maintained by JTWC and the Naval Research Laboratory in Monterey, California, and contains both nondeveloping depressions and all other tropical cyclones (JTWC 2002b). The nondeveloping depressions are included in the development of this model because there is a long-standing operational requirement to provide warnings on tropical depressions occurring in the western North Pacific, including the South China Sea, when intensities exceed 24 kt (Guard et al. 1992), and the record of nondeveloping depressions appears to be consistent throughout the 1967–2000 sample. Unlike in models developed for the Atlantic and eastern North Pacific, 2000 was used in this model development because this model was developed slightly later; hence, only 2001 data were available for independent evaluation. All best tracks for the period 1967–2000 are used in the development of the 5-day version of STIFOR, because aerial reconnaissance was routinely available in this basin prior to 1987. Data were utilized if initial storm positions were located between 80°E and the date line, south of 45°N, and not within 50 km of land.

3. Statistical methods

Model development closely follows the method described in Jarvinen and Neumann (1979) for the development of 3-day SHIFOR. The dependent variable (predictand) is the change in intensity from the initial conditions (DELV), which is predicted using 7 primary independent variables (predictors) and 28 secondary predictors constructed from the squares and cross products of the 7 primary predictors. All predictors are constructed in the same way for each basin, except for the variable containing the date, which is constructed from the yearday minus the yearday corresponding to the peak of the seasonal activity in that basin (Neumann 1993). The absolute value of this date variable is used in the eastern and western North Pacific tropical cyclone basins because it was better related to intensity change in those basins. Also, note that west longitude is treated as an absolute value in the eastern North Pacific and Atlantic. The primary predictors are as follows:

1) date information (JDAY):
    (a) Atlantic—the initial yearday minus yearday 253,
    (b) eastern North Pacific—the absolute value of the initial yearday minus yearday 238, and
    (c) western North Pacific—the absolute value of the initial yearday minus yearday 248;
2) latitude (LAT), 0°–90°N;
3) longitude (LON), 180°–0° in the Atlantic and eastern North Pacific (always positive), 0°–180° in the western North Pacific;
4) zonal speed of the storm (U) (kt), westerly motion is positive;
5) meridional speed of the storm (V) (kt), northerly motion is positive;
6) current intensity (VMAX) (kt); and
7) 12-h change in intensity (DVMX) (kt).

This method results in a pool of 35 potential predictors from which the best predictor combinations can be selected.

Variable selection for multiple-regression schemes can be accomplished through a number of methods. Three such methods are combined to select predictors for the models developed here. The methods are forward selection, backward selection, and stepwise selection. In forward selection, an attempt to add a predictor to the model is made (forward step). The predictor is retained if the probability associated with its relationship to the dependent variable determined from an *F* test (*p* value) is greater than a predefined level PIN. The model initially has no predictors, or individual predictors can be forced into the model prior to the first forward step. During typical forward selection, several forward steps would continue until none of the remaining potential predictors had *p* values less than PIN. Backward selection, as the reader may anticipate, involves an attempt to remove predictors from the model (backward step). A predictor is removed if its *p* value is less than a predefined level POUT. All the predictors initially would be forced into the model. In backward selection, backward steps are typically continued until all of the retained predictors have *p* values greater than POUT. The final variable selection method used in developing these models is referred to as stepwise selection. In stepwise selection, a backward step is attempted using POUT; if no variable is removed, a forward step is attempted using PIN. The combination of a backward step followed by a forward step is referred to as a stepwise step. In stepwise selection, only forced predictors enter the model initially, and stepwise steps continue until none of the predictors can be removed from the model and none of the remaining potential predictors can be added. Forcing predictors into the model refers to the process of giving forced predictors preference in the selection procedure. If a set of predictors is forced into the model, all of these predictors must be removed or kept before other potential predictors are allowed into the model (IMSL 1987). One can also combine these stepping procedures and force certain predictors into a given model to create regression models with desired properties. Such a procedure is used in this study. In the development of the regression equations for each basin, it is desired that continuity between predictors used at successive times be maintained as much as possible and that primary predictors are given preference in entering the model equations for the first forecast interval (12 h).

Models for each basin are developed using the same procedures but with different statistical significance levels (PIN and POUT) for the predictor selections. The significance levels were increased to limit the number of predictors used in the model. Because of the number of cases used to create these models, the typical significance levels used for the development of this type of regression model (e.g., PIN = 0.01, POUT = 0.02) were too easily achieved and predictors were entering the model that accounted for a very limited additional reduction in overall variance. Because the average intensity changes in the Atlantic tend to be smaller than in the eastern North Pacific or western North Pacific, the signal-to-noise ratio is lower in the Atlantic basin and significance levels (PIN, POUT) are set lower there.

Forecast equations are developed using multiple linear regression in which the predictand is DELV and independent variables are those available in the potential predictor pool. The predictors at each forecast interval are chosen using a procedure designed to improve the forecast continuity from one interval to the next and to provide a preference for the selection of primary predictors over quadratic combinations for the first forecast interval (12 h). The first step is to choose from the primary variables (1–7) for the 12-h forecast in a forward-selection process. PIN is set to 1×10^{-5} in the North Atlantic and 1×10^{-7} for the eastern and western North Pacific models for this forward selection. Once primary predictors have been chosen, they are forced into the model, and secondary predictors are then allowed to enter the model using a forward-selection procedure with a PIN and POUT equal to 1×10^{-6} and 1×10^{-10} for the Atlantic and eastern/western North Pacific models, respectively. To remove primary predictors that have lost their statistical significance, a backward-selection procedure is performed that removes all predictors that have a probability of being by chance greater than 2×10^{-5} and 2×10^{-10} in the Atlantic and eastern/western North Pacific, respectively. Last, the predictors in the model following the backward selection are forced into the model and a stepwise-selection procedure passes through the remaining potential predictors a final time using the same significance levels as the previous backward and forward steps, thus adding any remaining potential predictors made significant by the previous backward selection.

For forecast equations with lead times greater than 12 h, the predictors chosen for the forecast equation at the previous time are given preference in the same way primary predictors are given preference for 12-h forecast equations. This procedure was shown to provide more continuity among predictors and predictions with differing time lags than other variable selection procedures and resulted in between 4 and 12 predictors being chosen for each forecast equation.

The statistical formulation of these models differs from their predecessors SHIFOR and STIFOR in several important ways. First, the selection of predictors is different than the methods used by both Jarvinen and Neumann (1979) and Chu (1994), who included only predictors that explained at least 0.5% of the predictand variance. The predictor-selection procedure used in the development of these models also allows the forecast equations to evolve over the 120-h forecast period by allowing the predictors and number of predictors to change. Chu (1994) also chose to use VMAX instead of DELV as the predictand, a shortcoming that may explain some of the STIFOR forecast error characteristics (discussed later).

A common way of displaying predictors used in a set

TABLE 1. Predictors and associated normalized coefficients for the new 5-day Atlantic SHIFOR model at forecast times of 12–120 h. The number of individual predictors used for each forecast is given in parentheses.

	Forecast (h) (Predictor)									
	12 (6)	24 (8)	36 (7)	48 (7)	60 (8)	72 (8)	84 (7)	96 (5)	108 (7)	120 (8)
LAT									0.31	
LON	0.10	0.12	0.13	0.12	0.11	0.08				0.94
U	0.22	0.21								
VMAX	−0.26	−0.24								
DVMAX	0.66	0.71	0.67	0.57	0.44	0.39	0.11			
JDAY^2	−0.07	−0.10	−0.12	−0.13	−0.12	−0.12	−0.09	−0.10	−0.11	−0.11
LAT^2			0.25	0.30	0.36	0.40	0.46	0.45		
LAT × VMAX		−0.13	−0.54	−0.64	−0.71	−0.76	−0.83	−0.85	−0.60	−0.18
LON^2										−0.65
LON × VMAX										−0.70
U^2					−0.08	−0.10	0.11	0.13	0.16	0.16
$U \times V$									0.11	0.14
U × VMAX	−0.38	−0.36	−0.17	−0.18	−0.18	−0.18	−0.17	−0.15	−0.21	−0.16
VMAX^2									−0.24	
VMAX × DVMAX		−0.27	−0.39	−0.41	−0.36	−0.27	−0.26			

of regression equations is to show them in terms of their normalized regression coefficients. Normalized regression coefficients are expressed in terms of standard deviations and are created when all the predictors and the predictand have been normalized prior to calculating regression coefficients. To normalize a variable, the sample mean is subtracted from each member of the sample and the result is then divided by the sample standard deviation. A simple interpretation of a list of normalized regression coefficients is that the larger the normalized coefficient is, the greater is the individual influence of the predictor on the forecast of the predictand.

The predictors, along with their normalized regression coefficients, used in the Atlantic forecast equations are shown in Table 1. From the possible 35 predictors, only 15 are utilized in any of the forecast equations. Continuity is held among predictors, except for the 108- and 120-h time periods, in which different predictors begin entering the regression equations. Two predictors, JDAY^2 and U × VMAX, are used in every forecast equation. The combinations of cross-product terms involving LAT and LON represent spatial distributions of climatological intensity change based upon initial latitude and longitude and, to some degree, the current intensity and forward motion. An example of the spatial distributions of 48-h DELV predicted by these equations in the Atlantic basin, assuming the yearday is 250, VMAX = 50 kt, DVMAX = 0 kt, U = 10 kt, and V = 2 kt, is shown in Fig. 3a. The JDAY term accounts for median intensity changes from the seasonal peak of tropical cyclone activity.

Table 2 shows the normalized coefficients for predictors used in various forecast equations for the eastern North Pacific. From the possible 35 predictors, only 13 are utilized in any of the forecast equations. Predictors have forecast-to-forecast continuity, except for the 84- and 120-h time periods, in which different predictors again begin entering the model. The use of V × DVMAX at 84 h is a transitory feature in the statistics and was tested in verification studies and found to be important. At 120 h, the change in predictors suggests that 5-day forecasts are best estimated by damped persistence, because the decay in storm intensity is primarily a function of VMAX (with a negative coefficient). Other terms at 120 h are related to spatial distributions (i.e., LAT and LON). One of the predictors, JDAY × LAT, is used in every forecast equation. The cross-product terms involving LAT and LON for this basin create spatial distributions of climatological intensity change based upon initial latitude and longitude, date, and, to some degree, the current intensity and meridional motion. Figure 3b shows an example of the spatial distribution of DELV for a 48-h forecast made using the same assumption used in Fig. 3a. JDAY accounts for median intensity changes from the peak of the season.

Normalized coefficients and associated predictors used in various forecast equations for the western North Pacific are shown in Table 3. Fourteen predictors are utilized in any forecast equation. Three predictors (VMAX, LAT × VMAX, and VMAX^2) are used in every forecast equation. Cross-product terms that contain LAT and LON are also related to spatial distributions of climatological intensity change in this basin (Fig. 3c). The patterns are based upon initial latitude and longitude and, to some degree, the current intensity. In this basin the JDAY × VMAX term accounts for intensity changes relative to the peak of the season, suggesting that both intense storms and storms occurring farther from the peak of the season are forecast to weaken more rapidly.

4. Model evaluation

In this section, the statistical characteristics of the forecast made by these three models will be discussed.

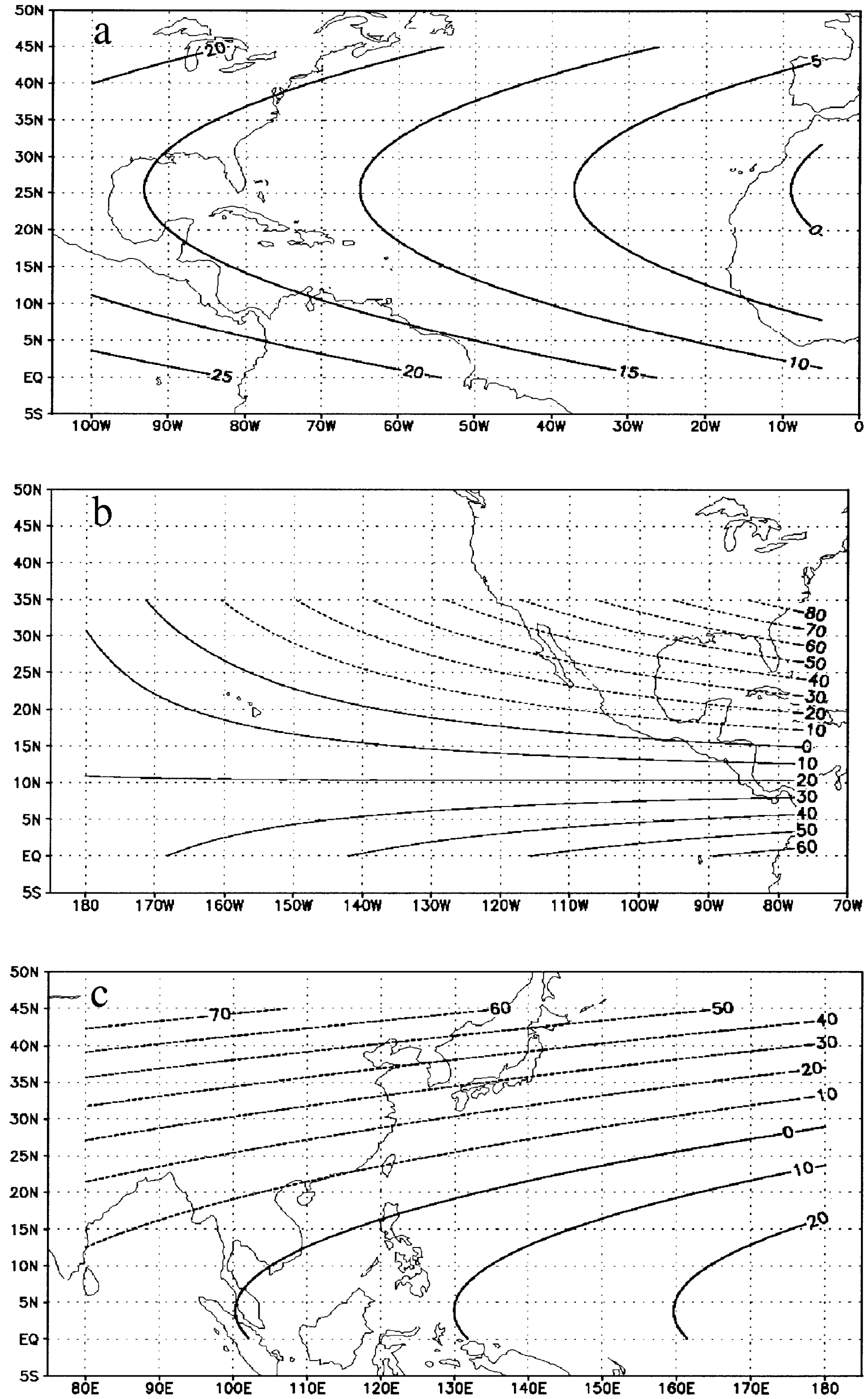

FIG. 3. Examples of the resulting spatial distribution of DELV (kt) calculated from the cross-product terms containing LAT and LON in the 48-h forecasts in the (a) North Atlantic, (b) eastern North Pacific, and (c) western North Pacific tropical cyclone basins. To calculate these distributions, the yearday = 250 (7 Sep), VMAX = 50 kt, DVMAX = 0 kt, U = 10 kt, and V = 2 kt. The contour intervals are 5 kt in the North Atlantic and 10 kt in the other basins.

The term ''forecast ability'' is used in this section to describe the statistical goodness of these forecasts, because by definition these control forecasts, derived using the CLIPER approach, possess no skill. In this same context, this definition does not imply that forecasts using the CLIPER approach have no utility, but rather that they are simply what is expected to occur based on past events, which may at times be an appropriate fore-

TABLE 2. Predictors and associated normalized coefficients for the new 5-day eastern North Pacific SHIFOR model at forecast times 12–120 h. The number of individual predictors used for each forecast is given in parentheses.

	Forecast (h) (Predictor)									
	12 (6)	24 (7)	36 (8)	48 (9)	60 (7)	72 (7)	84 (8)	96 (5)	108 (5)	120 (4)
LAT	−0.18	−0.22	−0.23	−0.69						
LON	−0.06	−1.13	−1.47	−1.67						
VMAX										
DVMAX	0.76	0.77	0.71	0.63	0.68	0.63	0.58			−0.64
JDAY × LAT	−0.10	−0.12	−0.14	−0.15	−0.15	−0.15	−0.14	−0.14	−0.13	
LAT × LON				0.56	−0.28	−0.27	−0.24	−0.25	−0.21	−0.10
LAT × *V*							−0.09	−0.10	−0.11	−0.18
LAT × DVMAX	−0.30	−0.38	−0.42	−0.41	−0.51	−0.52	−0.45			−0.10
LON^2		0.98	1.30	1.25						
LON × VMAX	−0.24				−0.23	−0.27	−0.33	−0.29	−0.31	
V^2			−0.07	−0.08	−008	−0.08				
V × DVMAX							−0.09			
$VMAX^2$		−0.33	−0.39	−0.44	−0.31	−0.30	−0.27	−0.31	−0.32	

cast. Both dependent and independent results will be examined in this section. As part of this examination, a comparison of the 3-day version of these models with the newly derived 5-day versions will be conducted using independent operational data. The errors and bias associated with these forecasts will be discussed through the entire 120-h forecast cycle.

There are two specific questions that need to be addressed in these new models. First, what are the statistical characteristics of these models? To address this question, root-mean-square errors (rmse), biases (average errors), and the *R* squared (R^2) created from the developmental dataset for each of these models are created and discussed in the context of other control models. Here R^2 is the linear coefficient of multiple correlation and measures the percentage of the total variance of the predictand that is explained by the predictors (Panofsky and Brier 1968). Another important question is how these forecasts perform independently and operationally. Using operational track and intensity information, forecasts can be made and evaluated, acknowledging that operational information (location, speed, intensity, etc.) can be different than the best-track data used for verification. For the Atlantic and eastern North Pacific models an independent evaluation can be performed for both the 2000 and 2001 hurricane seasons. Both the east Pacific and Atlantic versions of 5-day SHIFOR were run operationally in 2001, and the forecasts for 2000 were generated using operational input. For the 2000 and 2001 hurricane seasons, a direct comparison with the 3-day versions of the SHIFOR and official NHC forecasts can be accomplished using a homogeneous sample. In the western North Pacific, 5-day STIFOR was run operationally only for the later part of the 2001 typhoon season (after 20 July). Forecasts for the early part of the 2001 typhoon season were created using operational input so that the 5-day STIFOR could be compared with the 3-day STIFOR (Chu 1994) using a homogeneous sample for the entire 2001 season.

TABLE 3. Predictors and associated normalized coefficients for the new 5-day western North Pacific STIFOR model at forecast times 12–120 h. The number of individual predictors used for each forecast is given in parentheses.

	Forecast (h) (Predictor)									
	12 (9)	24 (9)	36 (10)	48 (10)	60 (10)	72 (10)	84 (12)	96 (12)	108 (9)	120 (7)
JDAY									−0.22	−0.22
LAT	0.10	0.10	0.37	0.46	0.43	0.39	0.41	0.39		
LON	1.38	1.99					2.04	1.89	1.68	1.46
VMAX	0.45	0.57	0.59	0.57	0.59	0.60	0.55	0.52	0.28	0.33
DVMAX	0.43	0.32								
JDAY × VMAX	−0.20	−0.27	−0.28	−0.28	−0.29	−0.29	−0.27	−0.25		
LAT^2			−0.28	−0.39	−0.37	−0.34	−0.39	−0.38	−0.13	−0.13
LAT × VMAX	−0.83	−1.04	−1.03	−0.96	−0.95	−0.93	−0.81	−0.72	−0.36	−0.42
LAT × DVMAX			−0.18	−0.24	−0.25	−0.23	−0.21	−0.19	−0.13	
LON^2	−1.45	−2.06	−0.5				−1.96	−1.81	−1.55	−1.31
LON × VMAX	0.96	1.22	1.18	1.03	0.93	0.82	0.54	0.37		
LON × DVMAX			0.43	0.45	0.42	0.37	0.32	0.29	0.18	
$VMAX^2$	−0.56	−0.76	−0.82	−0.77	−0.74	−0.68	−0.55	−0.47	−0.46	−0.43
$DVMAX^2$				−0.06	−0.06	−0.07	−0.07	−0.06		

TABLE 4. Statistical characteristics including R^2 (variance explained), rmse, and the number of cases used to develop the 5-day climatology and persistence models, STIFOR (North Atlantic and eastern North Pacific), and SHIFOR.

	12 h	24 h	36 h	48 h	60 h	72 h	84 h	96 h	108 h	120 h
					North Atlantic, 1967–99					
R^2	28.3	30.8	33.3	35.5	37.8	40.6	42.5	45.4	50.5	55.5
Rmse	6.6	10.6	13.5	15.5	16.9	17.4	17.8	17.6	16.4	15.0
No.	6004	5277	4620	4038	3534	3098	2725	2402	2108	1843
					Eastern North Pacific, 1975–99					
R^2	40.4	46.7	50.2	54.0	55.0	55.9	57.6	57.0	57.4	58.3
Rmse	6.6	10.3	13.2	14.7	16.3	17.1	17.2	17.8	17.8	17.4
No.	8899	8127	7375	6644	5941	5282	4667	4094	3581	3118
					Western North Pacific, 1967–2000					
R^2	40.4	45.0	47.4	50.6	52.8	54.3	57.4	58.3	58.9	59.8
Rmse	5.8	9.7	12.6	14.4	15.7	16.8	16.9	17.4	17.9	18.0
No.	25 883	24 428	22 786	21 014	19 266	17 570	15 959	14 428	12 990	11 626

a. Developmental data statistics

The statistical characteristics of the 5-day SHIFOR models and the 5-day STIFOR model are a by-product of the development of these models. The rmse and values of R^2 associated with each model along with the number of data points used to develop these models are shown in Table 4. The number of forecast cases for all regression models decreases from the initial time because as lead time increases more storms dissipate, resulting in approximately a 2/3 reduction in cases by 120 h. Rmse tends to increase with increasing forecast time but to level off and even to decrease beyond 84 h in the Atlantic and 108 h in the eastern North Pacific. The maximum rmse are found in the western North Pacific at 120 h, but the value of 18 kt is a common maximum error for all of these tropical cyclone basins. Values of R^2 increase with increasing forecast time in all of the basins, with the largest values approaching 60% of the variance in the west Pacific basin at 120 h. In tests using independent data, we expect the performance of the regression equations to degrade in a mean-case scenario following equations discussed in Knaff and Landsea (1997) that account for sample size, number of predictors, and the value of R^2. This method, however, does not account for errors in the data or the effects of extreme cases as is discussed in Mielke et al. (1996). Degradation is expected to be minimal because of the large number of data points used in development of these equations. For example, using Knaff and Landsea's equations, the regression equation that has the least data supporting it (8 predictors, 1843 forecast cases) is in the North Atlantic at 120 h, and the anticipated degradation is from $R^2 = 55.5$ and an rmse $= 15.0$ kt to $R^2 = 55.1$ and an rmse $= 15.1$ kt. A much larger degradation will likely occur as a result of errors in operational input data, particularly those inputs related to current and past intensity measurements that have the greatest uncertainty.

If these models are developed by excluding terms involving DVMAX, a climatology control model can be created. Because these models do not have knowledge of the persistence, the early forecasts (12–84 h) have lower values of R^2 than the control models developed using the CLIPER approach. Later forecasts (beyond 84 h) are nearly identical. The values of R^2 are compared for models developed using climatology and CLIPER in Figs. 4a–c. These results indicate that the CLIPER approach produces a superior control model.

b. Independent and operational comparison with their predecessors

Forecast comparisons between 5-day versions of SHIFOR (Atlantic and eastern North Pacific) and STIFOR and their 3-day predecessors were made using independent operational data during 2000 and 2001. Operational input data come from the Automated Tropical Cyclone Forecast System (Sampson and Schrader 2000). Using these data, which have errors inherent in operations, independent forecasts were made for the North Atlantic, eastern North Pacific, and western North Pacific using the newly created 5-day models.

Results of these comparisons between the 3-day version SHIFOR and the newer 5-day SHIFOR in the Atlantic and eastern North Pacific are shown in Table 5. Mean absolute errors and biases (mean errors) of the two models are similar in the Atlantic for both years. Differences were not statistically significant using a one-tailed Student's t test. Sample sizes were adjusted for 24-h serial correlation, producing an effective sample size, following the method described in Franklin and DeMaria (1992). Biases for the newer version of SHIFOR tend to be smaller and close to zero in 2000 and larger and positive in 2001, suggesting that predicted intensities of the 5-day model are on average larger. In the east Pacific, the two versions of SHIFOR are again very similar, with errors that are not significantly different. The biases again suggest that the new version of SHIFOR tends to forecast slightly greater intensities at the longer leads than its predecessor.

An observation made while verifying, but not shown for brevity, is that the 5-day SHIFOR models tend to

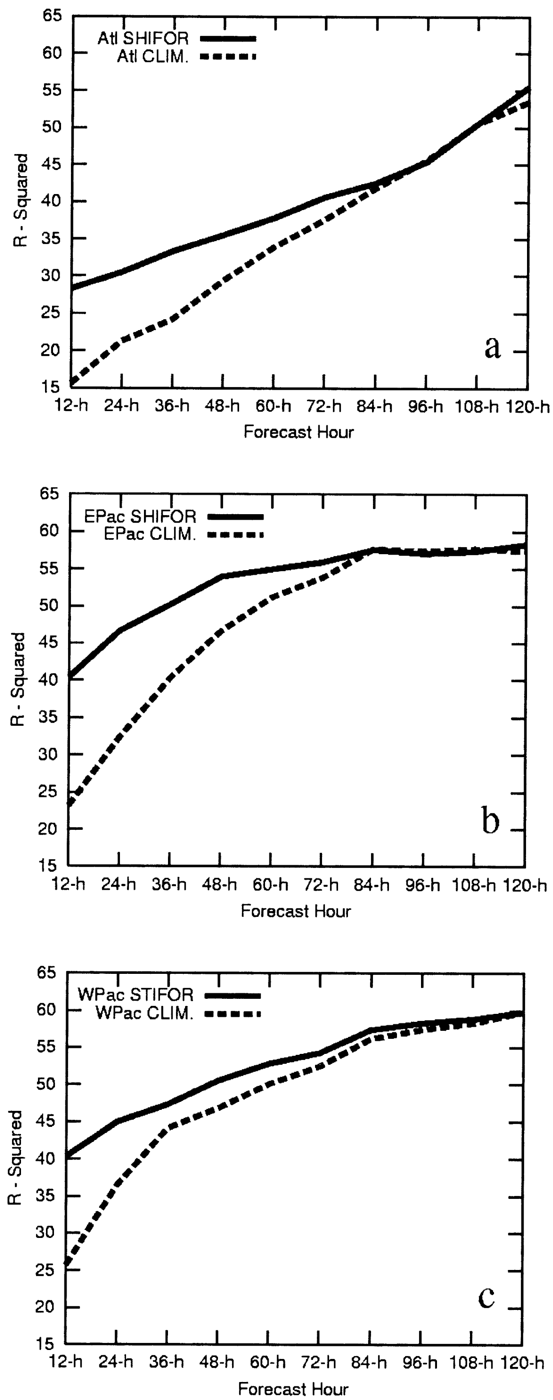

FIG. 4. Comparison between the R^2 values associated with the 5-day CLIPER models (Atlantic SHIFOR, eastern North Pacific SHIFOR, and western North Pacific STIFOR) and similar models derived using the climatological information only (CLIM.). Comparisons are shown for the (a) North Atlantic, (b) eastern North Pacific, and (c) western North Pacific.

predict larger intensity changes, straying further from the initial conditions than their 3-day predecessors. This characteristic, although it creates forecasts with a greater spread from the initial intensities, does not result in greater or significantly different intensity forecast errors in these basins. The biases during this 2-yr verification, however, are generally less negative in both basins with the newer statistical formulation.

In the western North Pacific, 2001 is used to make a comparison between the 5-day version of STIFOR and the 3-day STIFOR (Chu 1994). Table 6 lists the mean absolute errors and biases associated with STIFOR (3 day) and the newly developed 5-day STIFOR. The mean absolute errors associated with the 5-day STIFOR are significantly smaller for all forecast time periods and are much smaller beyond 36 h, suggesting that the statistical formulation of the new 5-day STIFOR is superior. Statistical significance was measured using a one-tailed Student's *t* test and the sample size was adjusted for serial correlation using a 24-h time window. An important consequence of this finding is that the better forecasts made by 5-day STIFOR are more difficult to outperform and will ''raise the bar'' for achievement of skillful tropical cyclone intensity forecasting in the western North Pacific. The biases are, as was seen in the Atlantic and east Pacific, less negative. Also similar to the Atlantic and east Pacific, the new STIFOR model produces forecasts that vary further from their initial conditions, as was the case for the SHIFOR models (not shown).

There are two factors that likely result in the new model producing superior forecasts. The first factor is the choice of DELV as the predictor instead of VMAX. It was found that DELV is an easier parameter to predict by comparing three versions of the model with the following predictands: DELV, VMAX, and the 12-hourly change in intensity (''INCV''). Of these predictands, DELV produced much better hindcasts. The second factor is the inclusion of quadratic terms as done in Jarvinen and Neumann (1979), so that more variance is explained relative to the model with only linear terms.

c. Intensity forecasts beyond 72 h

One of the main purposes for updating these simple models was to provide a verification tool for intensity forecasts through 5 days. We examined mean absolute error and bias characteristics of these models with respect to their tropical cyclone basins. Error characteristics of the 5-day SHIFOR and 5-day STIFOR models versus persistence (in which the intensity is held constant out to 5 days) for forecasts made at 12–120 h are examined using operational data for 1997–2000 to initialize the models (Fig. 5). These results are not completely independent, but general characteristics of model performance can still be evaluated. Results show that the Atlantic 5-day SHIFOR mean absolute intensity errors maximize at 84 h and then begin to decrease

TABLE 5. Mean absolute errors of intensity forecasts made by the 3-day SHIFOR models and the 5-day SHIFOR models in the North Atlantic and eastern North Pacific tropical cyclone basins. The years 2000 and 2001 are considered in this comparison. Input data are independent of the data used to develop the 5-day forecast models. The verification includes all storms and depressions.

	12 h		24 h		36 h		48 h		72 h	
Statistic	3 day	5 day	3 day	5 day	3 day	5 day	3 day	5 day	3 day	5 day
North Atlantic, 2000										
Errors	8.4	8.5	14.0	13.7	18.2	18.2	21.5	21.0	20.7	20.9
Biases	0.1	0.6	0	1.4	−1.5	0.3	−2.2	−0.3	−2.2	−0.6
No.	308		275		244		214		171	
North Atlantic, 2001										
Errors	8.1	8.3	12.2	12.2	14.8	15.0	17.1	16.7	20.4	20.1
Biases	1.2	1.8	1.6	2.7	1.3	3.4	0.9	3.5	−0.3	4.3
No.	396		372		348		326		276	
East Pacific, 2000										
Errors	7.2	6.8	11.3	11.1	14.4	14.6	16.8	17.2	18.3	18.3
Biases	−0.4	0.2	0.1	1.4	0.8	2.7	1.6	3.3	2.3	3.6
No.	298		266		239		208		153	
East Pacific, 2001										
Errors	7.2	7.1	11.4	11.4	14.6	14.1	16.8	15.9	18.5	19.1
Biases	0.2	0.5	0.4	0.9	0.2	0.9	0.4	0.3	1.0	2.6
No.	303		271		238		211		160	

through 120 h while forecasts made with persistence continue to increase slightly even after 84 h. This behavior is seen in the eastern and western North Pacific, where errors level off at 96 h with nearly constant errors thereafter. This behavior is similar to the behavior of the rmse in Table 4. The behavior of these errors may be partially explained by two factors: 1) the standard deviation of DELV begins to level off at the longer lead times and 2) the average time to peak intensity is less than 5 days, averaging 4.1, 3.4, and 4.8 days in the Atlantic (1967–99), eastern North Pacific (1975–99), and western North Pacific (1967–2000), respectively. The biases calculated from the models initialized with 1997–2000 operational data (Fig. 6) become larger and more negative as the forecast time increases for both the North Atlantic and western North Pacific, indicating that Atlantic and western North Pacific storms on average are underforecast. The Atlantic 5-day SHIFOR has a tendency to underforecast, more so than even the persistence-based forecast. On the contrary, the biases remain very close to zero through 60 h and then become slightly negative for the rest of the 5-day period in the eastern North Pacific, indicating that storms are very well forecast through 60 h, with a tendency to overforecast at longer leads. The relatively small biases in the eastern North Pacific likely result from a common weakening behavior resulting from encounters with cooler SSTs and vertical wind shear to the west. This basin is also the only one of the three for which the average DELV is negative. Furthermore, positive biases resulting from the persistence-based forecast in the eastern North Pacific also indicate that the sample tends to weaken with increasing forecast lead.

Independent errors and biases for forecasts beyond 72 h are examined and compared with persistence during 2000 and 2001 in the Atlantic and eastern North Pacific and during 2001 (beginning 21 July) in the western North Pacific. Table 7 shows the mean absolute forecast errors and biases for the Atlantic and eastern North Pacific. These results show that the 5-day SHIFOR models perform well beyond 72 h, producing forecasts that are significantly better than persistence. Also evident is that 2000 and 2001 have distinctly different characteristics. It appears that 2001 was an easier year in which to beat the 5-day SHIFOR control models in both basins, as indicated by larger errors.

Table 8 shows mean absolute forecast errors and biases for the 5-day STIFOR model and persistence at times beyond 72 h derived using independent data. All forecasts were run operationally starting 21 July 2001 for the western North Pacific. Five-day STIFOR produces forecasts that are significantly better than persis-

TABLE 6. Mean absolute errors of intensity forecasts made by the 3-day STIFOR model and the 5-day STIFOR model in the western North Pacific for 2001. This verification includes all storms and depressions.

	12 h		24 h		36 h		48 h		72 h	
Statistic	3 day	5 day	3 day	5 day	3 day	5 day	3 day	5 day	3 day	5 day
Errors	7.6	7.0	12.8	11.2	17.5	14.6	21.3	17.6	26.8	21.5
Biases	−1.8	−1.9	−3.0	−2.4	−4.6	−4.3	−7.2	−6.1	−13.0	−8.4
No.	688		645		597		545		442	

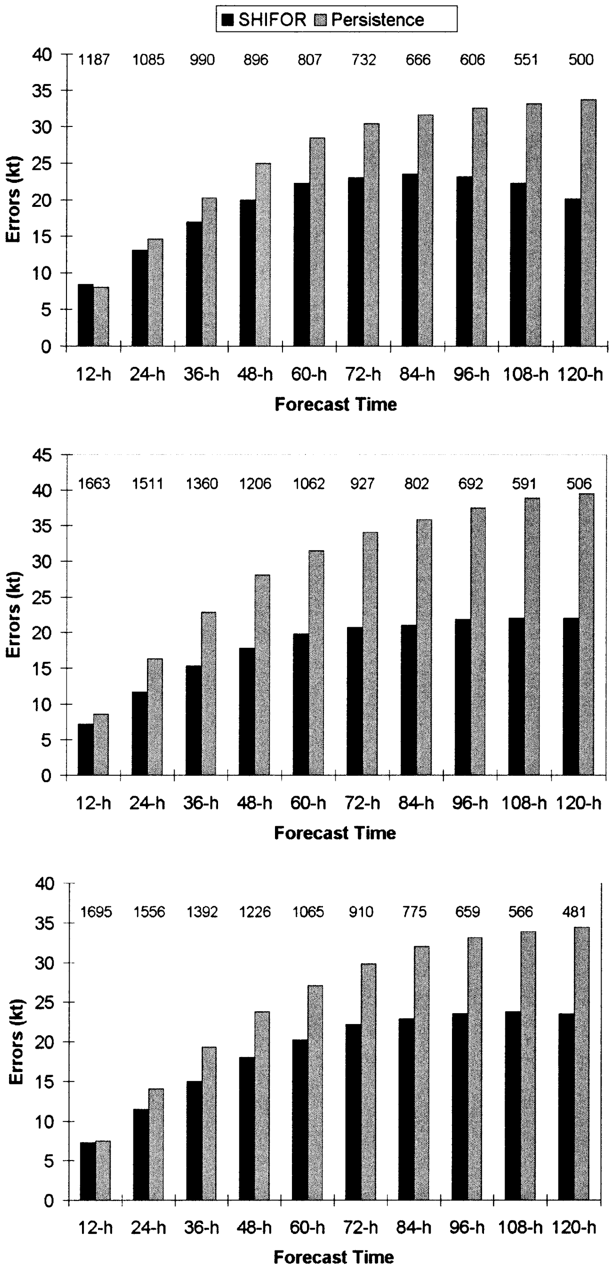

FIG. 5. Mean absolute intensity forecast errors (kt) from the 5-day SHIFOR and STIFOR models through 120 h, initialized using operational data for 1997–2000, as compared with a forecast of persistence of initial conditions. Errors from the (top) North Atlantic, (middle) eastern North Pacific, and (bottom) western North Pacific are shown. The number of cases is shown above the bars in each panel.

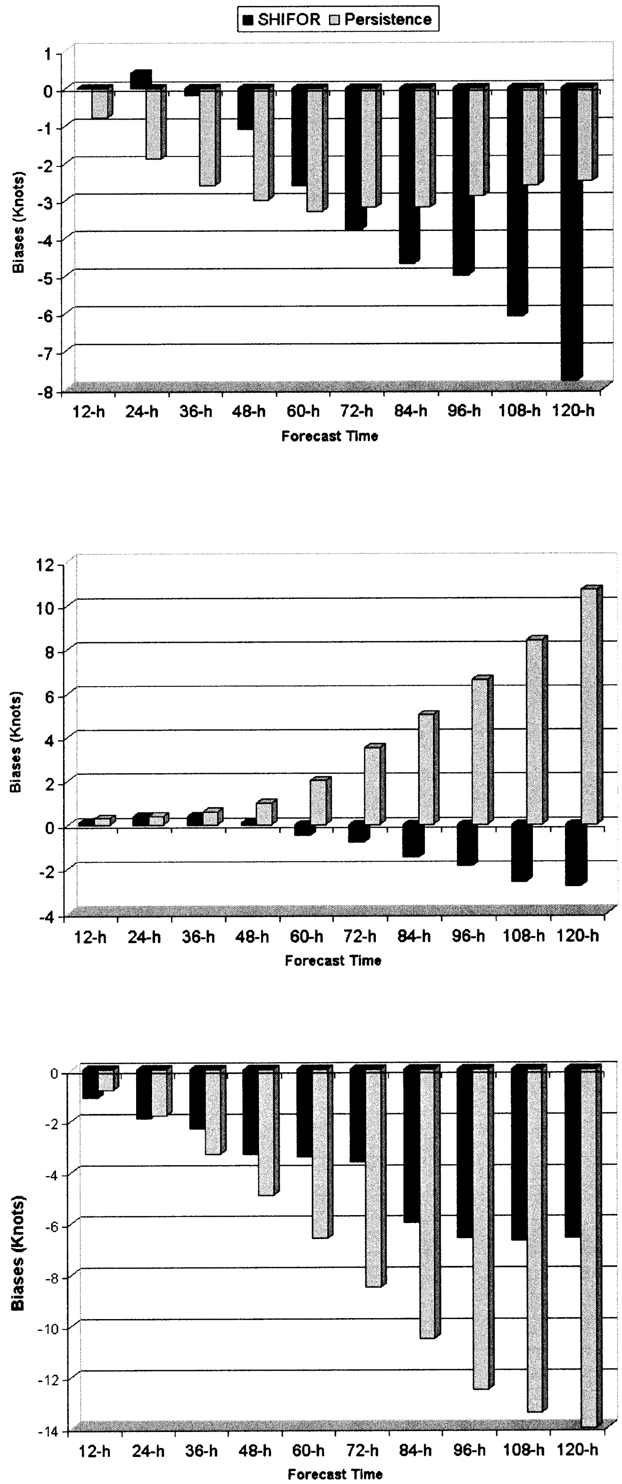

FIG. 6. Intensity forecast biases (kt) from the 5-day SHIFOR and STIFOR models through 120 h, initialized using operational data for 1997–2000, as compared with a forecast of persistence of initial conditions. Biases from the (top) North Atlantic, (middle) eastern North Pacific, and (bottom) western North Pacific are shown. The number of cases is the same as in Fig. 5.

tence alone, reducing errors by a factor of 2 even at these long leads. Comparing the errors for persistence and 5-day STIFOR with values shown in Figs. 5 and 6 also suggests that 2001 was an easy year in which to beat the 5-day STIFOR forecasts.

Performance of these simple 5-day intensity forecasts models, derived using the CLIPER approach, at time

TABLE 7. Mean absolute errors of intensity forecasts beyond 72 h made by the 5-day SHIFOR model (SHF5) and persistence (PER) in the North Atlantic and eastern North Pacific. The years 2000 and 2001 are considered in this comparison, because the input data are independent of the data used to develop the 5-day forecast models. This verification included all storms and depressions. Note that in 2001 the model was run operationally.

	84 h		96 h		108 h		120 h	
Statistic	SHF5	PER	SHF5	PER	SHF5	PER	SHF5	PER
				North Atlantic, 2000				
Errors	21.0	26.8	19.3	24.5	17.6	23.2	17.6	22.9
Biases	−0.8	−3.4	0.8	−2.7	1.1	−2.9	−0.5	−3.9
No.	172		158		144		130	
				North Atlantic, 2001				
Errors	21.7	25.7	22.1	27.6	20.9	28.2	20.9	28.2
Biases	3.9	−8.0	3.4	−9.3	3.4	−10.4	2.1	−11.4
No.	249		221		197		173	
				East Pacific, 2000				
Errors	17.5	28.9	16.2	28.8	15.6	27.8	14.2	27.4
Biases	2.9	1.3	2.0	1.4	2.5	3.3	3.9	7.3
No.	127		104		86		71	
				East Pacific, 2001				
Errors	20.4	30.7	19.0	31.7	17.8	32.6	15.9	32.8
Biases	4.2	4.4	3.0	5.7	3.2	7.5	2.6	9.2
No.	135		111		91		75	

periods beyond 72 h is somewhat unexpected. These models produce forecasts that reduce errors associated with persistence by between 30% and 50% and have characteristics that are similar to the developmental datasets. Operational mean absolute forecast errors beyond 72 h have values between 14 and 24 kt. These errors are similar to the values at 72 h and have a tendency to become nearly constant at 5 days. This error saturation likely indicates the limit of predictability for this method.

5. Concluding remarks

To evaluate 5-day intensity forecasts, control models that make forecasts through 120 h are needed. This paper describes development and testing of newly developed TC intensity models that utilize the CLIPER approach to make forecasts through 5 days for the Atlantic, eastern North Pacific, and the western North Pacific basins. These models serve as a new control that represents the no-skill threshold for TC intensity forecasting skill at NHC and JTWC as those institutions begin to evaluate 5-day tropical cyclone forecasts.

The newly developed models produce forecast error statistics statistically very similar to their 3-day predecessors in the Atlantic and east Pacific but represent a statistically significant improvement over their predecessors in the western North Pacific. In essence, the bar that represents skillful intensity forecasts in the Atlantic and eastern Pacific has changed little, but it has been dramatically raised in the western North Pacific basin. Large differences between the 3-day STIFOR model and the newer 5-day STIFOR model in the western North Pacific likely are the result of model formulation (the use of quadratic predictors and a more easily explained predictand).

The new 5-day versions of these models also tend to produce intensity change forecasts that deviate further from the initial intensity than their predecessors in all three basins. These forecasts are different in an individual sense but produce nearly identical error statistics in the Atlantic and eastern North Pacific. It is interesting to note how well all of these models perform beyond 72 h. Errors at 72 h are at least 2 times as large as errors at 12 h, but errors at 120 h are on the same order and in some cases actually decrease relative to 72-h errors. This observation is likely due to two factors: 1) the standard deviation of DELV begins to level off at the longer lead times and 2) the average time to peak intensity is less than 5 days in all basins.

TABLE 8. Mean absolute error of intensity forecasts made by the STIFOR model (ST5D) and persistence (PER) in the western North Pacific for 21 Jul–31 Dec 2001. Verification included all storms and depressions.

	84 h		96 h		108 h		120 h	
Statistic	ST5D	PER	ST5D	PER	ST5D	PER	ST5D	PER
Errors	22.6	40.4	23.6	40.4	23.6	41.8	23.7	42.2
Biases	−9.6	−17.7	−10.6	−20.6	−10.9	−13.7	−10.0	−25.5
No.	363		321		277		243	

With these models now running as part of the operational suite at both NHC and JTWC, it will be interesting to see how well the official forecasts and other models that produce 5-day intensity forecasts compare with these simple models over the next several years. It remains to be seen if operational 5-day forecasts of hurricane track and intensity prove to be useful enough to become a permanent part of the operational TC forecasting package. It is hoped that these baseline forecasts will continue to provide verification information as well as, at times, useful guidance and in general to continue the excellence exemplified by their predecessors.

Acknowledgments. The authors thank the three reviewers for their helpful and constructive comments concerning this manuscript and acknowledge the work of Arthur Pike who created the 3-day versions of SHIFOR used in the Atlantic and eastern North Pacific. The Office of Naval Research and the Insurance Friends of the National Hurricane Center, Inc., supported portions of this research.

REFERENCES

Aberson, S. D., 1998: Five-day tropical cyclone track forecasting in the North Atlantic basin. *Wea. Forecasting,* **13,** 1005–1015.

Chu, J.-H., 1994: A regression model for the western North Pacific tropical cyclone intensity forecasts. NRL Memo. Rep. 7541-94-7215, Naval Research Laboratory, 33 pp. [Available from Naval Research Laboratory, 7 Grace Hopper Ave., Monterey, CA 93943-5502.]

Davis, M. A. S., G. M. Brown, and P. Leftwich, 1984: A tropical cyclone data tape for the eastern and central North Pacific basins, 1949–1983. NOAA Tech. Memo. NWS NHC-25, 15 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]

DeMaria, M., and J. Kaplan, 1999: An updated Statistical Hurricane Intensity Prediction Scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting,* **14,** 326–337.

Dvorak, V. F., 1972: A technique for the analysis and forecasting of tropical cyclone intensities from satellite pictures. NOAA Tech. Memo. NESDIS 36, 15 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]

——, 1975: Tropical cyclone intensity analysis and forecasting from satellite imagery. *Mon. Wea. Rev.,* **103,** 420–430.

——, 1984: Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. NESDIS 11, 47 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]

Elsberry, R. L., 1987: Observation and analysis of tropical cyclones. *A Global View of Tropical Cyclones,* R. L. Elsberry, Ed., Office of Naval Research, 1–11. [Available from Naval Postgraduate School, Monterey, CA 93943.]

Franklin, J. L., and M. DeMaria, 1992: The impact of Omega dropwindsonde observations on barotropic hurricane track forecasts. *Mon. Wea. Rev.,* **120,** 381–391.

Gross, J. M., 1999: 1998 National Hurricane Center forecast verification. *Minutes of the 53d Interdepartmental Hurricane Conf.,* Biloxi, MS, Office of the Federal Coordinator for Meteorological Services and Supporting Research, B24–B63. [Available from Office of the Federal Coordinator for Meteorological Services and Supporting Research, 8455 Colesville Rd., Suite 1500, Silver Spring, MD 20910.]

Guard, C. P., L. E. Carr, F. H. Wells, R. A. Jefferies, N. D. Gural, and D. K. Edson, 1992: Joint Typhoon Warning Center and the challenges of multibasin tropical cyclone forecasting. *Wea. Forecasting,* **7,** 328–352.

IMSL, 1987: *FORTRAN Subroutines for Statistical Analysis.* International Mathematical and Statistical FORTRAN Library, 1232 pp.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity for the North Atlantic basin. NOAA Tech. Memo. NWS NHC-10, 22 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]

——, ——, and M. A. S. Davis, 1984: A tropical cyclone data tape for the North Atlantic basin, 1886–1983: Contents, limitations, and uses. NOAA Tech. Memo. NWS NHC-22, 21 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]

JTWC, cited 2002a: JTWC TC climatology tables—northwestern Pacific. [Available online at http//www.npmoc.navy.mil/jtwc/climostats/Statclimo.html.]

——, cited 2002b: Tropical cyclone best track data site. [Available online at http://www.npmoc.navy.mil/products/jtwc/best_tracks/.]

Knaff, J. A., and C. W. Landsea, 1997: An El Niño–Southern Oscillation climatology and persistence (CLIPER) forecasting scheme. *Wea. Forecasting,* **12,** 633–652.

Landsea, C. W., 1993: A climatology of intense (or major) Atlantic hurricanes. *Mon. Wea. Rev.,* **121,** 1703–1713.

McAdie, C., and M. B. Lawrence, 2000: Improvements in tropical cyclone track forecasting in the Atlantic basin, 1970–98. *Bull. Amer. Meteor. Soc.,* **81,** 989–997.

Merrill, R. T., 1980: A statistical tropical cyclone motion forecasting system for the Gulf of Mexico. NOAA Tech. Memo. NWS NHC 14, 21 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]

Mielke, P. W., Jr., K. J. Berry, C. W. Landsea, and W. M. Gray, 1996: Artificial skill and validation in meteorological forecasting. *Wea. Forecasting,* **11,** 153–169.

Neumann, C. J., 1972: An alternate to the HURAN (hurricane analog) tropical cyclone forecasting system. NOAA Tech Memo. NWS SR-62, 23 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]

——, 1993: Global overview. *Global Guide to Tropical Cyclone Forecasting,* G. J. Holland, Ed., WMO Tech. Doc. 560, Rep. TCP-31, 1-1–1-43.

——, B. R. Jarvinen, C. J. McAdie, and J. D. Elms, 1993: Tropical cyclones of the North Atlantic Ocean, 1871–1992. Historical Climatology Series, No. 6-2, 4th rev., NESDIS, Asheville, NC, 193 pp. [Available from National Climatic Data Center, 151 Patten Ave., Asheville, NC 28801.]

OFCM, 2001: 55th IHC action items. *Minutes of the 55th Interdepartmental Hurricane Conf.,* Orlando, FL, Office of the Federal Coordinator for Meteorological Services and Supporting Research, 4–8. [Available from Office of the Federal Coordinator for Meteorological Services and Supporting Research, 8455 Colesville Rd., Suite 1500, Silver Spring, MD 20910.]

Panofsky, H. A., and G. W. Brier, 1968: *Some Applications of Statistics to Meteorology.* The Pennsylvania State University, 224 pp.

Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (version 3.2). *Bull. Amer. Meteor. Soc.,* **81,** 1231–1240.

Sheets, R. C., 1990: The National Hurricane Center—past, present and future. *Wea. Forecasting,* **5,** 185–232.