Statistical Prediction of Integrated Kinetic Energy in North Atlantic Tropical Cyclones

Michael E. Kozar[1] and Vasubandhu Misra
Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida

[1] *Corresponding author address: Michael Kozar, Center for Ocean-Atmospheric Prediction Studies, Florida State University, 2000 Levy Avenue, Building A, Suite 292, Tallahassee, FL 32306-2741; email: mkozar@coaps.fsu.edu*

Abstract

Integrated Kinetic Energy (IKE) is a useful quantity that measures the size and strength of a tropical cyclone wind field. As a result, it is inherently related to the destructive potential of these powerful storms. In most current operational settings, there are limited resources designed to assess the IKE of a tropical cyclone, as storm track and maximum intensity are typically prioritized. Therefore, to complement existing forecasting tools, a statistical scheme is created to project fluctuations of IKE in North Atlantic tropical cyclones for several forecast intervals out to seventy-two hours. The resulting scheme, named Statistical Prediction of Integrated Kinetic Energy (SPIKE), utilizes multivariate normal regression models trained on environmental and storm-related predictors from all North Atlantic tropical cyclones occurring from 1990 through 2011. During this training interval, SPIKE outperforms persistence and is capable of explaining more than 80% of observed variance in total IKE values at a forecast interval of 12 hours, trailing down to just below 60% explained variance at an interval of 72 hours. The skill of the SPIKE model is evaluated further using bootstrapping exercises in order to gauge the predictive abilities of the statistical scheme. In addition, the performance of the SPIKE model is also evaluated for the 2012 Atlantic Hurricane Season, which notably falls outside of the training interval. Ultimately, the validation exercises return shared variance scores similar to those found in the training exercises, serving as a proof of concept that the SPIKE model can be used to project IKE values when given accurate predictor data.

1. Introduction

Over the past few decades, predicting track and intensity fluctuations of North Atlantic tropical cyclones (TCs) has been a major emphasis for operational forecasters and atmospheric researchers alike. Both track and intensity metrics are definitively connected to the risks of a tropical cyclone, as maximum wind speed (VMAX) is related to the maximum potential destructiveness of a storm's wind field (e.g. Emanuel 2005; Bell et al. 2000), and storm position is used to identify the regions at risk of experiencing the storm.

Despite the inherent usefulness of track and intensity metrics, they do not yield information on the size or overall strength of TCs. Over the past few Atlantic hurricane seasons, several notable landfalling TCs have produced more damage than otherwise would have been expected by a storm of its intensity. Hurricanes Ike (2008) and Irene (2011) both produced in excess of $15 billion in damage across the United States as noted by the National Hurricane Center's (NHC) tropical cyclone reports (Berg 2009; Avila and Cangialosi 2011) despite being rated as category two and category one storms respectively at time of landfall on the Saffir Simpson Hurricane Wind Scale ("SSHWS"; Saffir 1975; Simpson 1974). More recently, Sandy (2012) only had a VMAX of 70 knots when it made landfall as an extraordinarily large post-tropical storm, yet it caused in excess of $50 billion in total damage (Blake et al. 2013). As a result of these recent damaging storms and those from the record-setting 2005 Atlantic Hurricane Season (i.e. Hurricane Katrina), some researchers have questioned whether or not intensity metrics, such as maximum sustained winds, are best suited to communicate the risks of TCs (Kantha 2006).

Hurricanes Ike, Irene, and Sandy were all very large storms despite their weaker intensities. Therefore, it is distinctly possible that VMAX is not the sole variable tied to damage potential, especially in large tropical cyclones. Irish et al. (2008) found that storm size is significantly correlated to storm surge in Atlantic Hurricanes and recommended that although intensity

scales adequately categorize wind damage, storm size must be considered to adequately categorize damage from flooding. Of course, storm surge damage is also significantly influenced by non-meteorological variables such as coastal bathymetry, population density, and property value. Nonetheless, with all else being equal, storm size is thought to have a significant influence over storm surge risks in landfalling TCs. Overall, conclusions such as these underscore the importance of assessing TC structure in real time and also help to validate the usefulness of metrics related to TC size such as operational wind radii.

Consequently, new scales and metrics have emerged as a complement to the SSHWS that are more closely tied to the overall kinetic energy of tropical storms (Powell and Reinhold 2007; Maclay et al. 2008). Integrated kinetic energy (IKE) is one such metric that is calculated by accumulating one-half of the 10-meter wind field squared (U^2) times the density (ρ) per unit volume over a one-meter depth volume domain of a tropical cyclone (Powell and Reinhold 2007), or more specifically,

$$(1) \quad IKE = \int_v \frac{1}{2}\rho U^2 dV \, .$$

In most cases, IKE is integrated using Equation 1 over a portion of the storm volume (V) that contains wind speeds greater than a certain threshold, such as tropical storm force or hurricane force (Powell and Reinhold 2007; Misra et al. 2013).

Importantly, unlike maximum sustained wind metrics, IKE responds to changes in the overall size, strength, and intensity of a TC. With respect to the physical processes that govern damage potential, IKE is designed to scale with the stress of the wind on the ocean as well as the wind load that acts upon a structure (Powell and Reinhold 2007). Therefore, IKE and other similar kinetic energy metrics are hypothesized to correspond well to the destructive potential of TCs, particularly with regards to storm surge damage.

For example, Hurricanes Ike, Irene, and Sandy at their respective times of landfall in the United States had extremely high IKE values, despite their low intensities (Figure 1). Just prior

to landfall, each of these three storms had IKE values that would rank in the top 7.5% of IKE values for all hurricane fixes between 1990 and 2011. Sandy, in particular, reached a lifetime maximum of over 400 TeraJoules (TJ) of IKE just prior to landfall, giving it the second highest maximum IKE value in any Atlantic TC since 1990. Therefore, it is distinctly possible that forecasting IKE, as a complement to existing intensity metrics, could help to better assess the risks of landfalling North Atlantic TCs, particularly for larger and less intense storms such as these three recent examples.

Despite the potential usefulness of real-time IKE updates, there are limited resources currently available to forecasters that are specifically designed to assess the kinetic energy of a tropical cyclone. Therefore, we present a statistical model, named Statistical Prediction of Integrated Kinetic Energy (SPIKE), which can potentially be used in real time for forecasting IKE fluctuations in North Atlantic TCs. Much like the operational Statistical Hurricane Intensity Prediction Scheme (SHIPS) used for forecasting VMAX (DeMaria and Kaplan 1994;1999; DeMaria et al. 2005), SPIKE utilizes a multivariate linear regression model trained on a blend of environmental and internal storm-driven predictors, in addition to observed persistence metrics, to predict changes of IKE out to three days in the future.

2. Data

a. Historical Integrated Kinetic Energy Record

In order to calibrate a statistical regression model for IKE, it is first necessary to obtain a historical record of IKE that can be used as the calibration model's dependent variable. Gridded analyses of TC wind fields would be ideal for calculating IKE, but unfortunately datasets like H*Wind analyses (Powell et al. 1998) are discontinuous, as they are not available for every six-hourly storm fix. Alternatively, Powell and Reinhold (2007) developed an approximate set of stepwise equations to calculate estimates of IKE from operational 34-knot, 50-knot, and 64-knot

wind radii, the radius of maximum wind (RMW), and VMAX. This methodology to approximate IKE from operational wind radii has been used previously to catalogue historical IKE levels (Misra et al. 2013) and will be used again here to form a large historical record of IKE values to train and test our statistical model. Specifically, the operational wind radii from the extended best track dataset (Demuth et al. 2006) are utilized to create a six-hourly record of IKE values for all North Atlantic TCs over a twenty-two year training interval between 1990 and 2011. The resulting IKE values in the historical dataset specifically measure the integrated kinetic energy only over the portion of the wind field where the wind speeds are of tropical storm force or greater ($U > 18$ m s^{-1}). This specific IKE quantity is selected because it is hypothesized to be closely related to storm surge damage potential by Powell and Reinhold (2007).

In total, our historical IKE record comprises of 5498 six-hourly IKE values from 291 individual TCs included from 1990 to 2011 (one IKE value for each six-hourly TC fix). For storm fixes to be included in this IKE database, the storm must have maximum sustained winds of tropical storm force or greater, and the storm must be classified as a tropical or subtropical cyclone in the official best track database (i.e. extratropical and post-tropical lows are not considered). The vast majority of the fixes that meet these criteria, and are included in the historical IKE database, are located over the open ocean. However, there are a small minority of points, for which the center of the TC is located over land. These so called “land points” occur mostly over islands or very near the coastlines of mainland North America, as storms typically do not retain tropical characteristics and tropical storm intensity long after landfall. We elected not to remove these “land points” in order to include landfalling storms in our training sample, considering that TCs have the most significant societal impacts when they make landfall.

The relative frequency distribution of this large sample of six-hourly historical IKE data points is shown in Figure 1. It is immediately evident that Atlantic TCs most often have IKE values that do not exceed 25 TJ. The mean IKE value in the dataset is 34.9 TJ, and the standard deviation

is 43.0 TJ. In rare instances, TCs can briefly obtain IKE values greater than 300 TJ, typically near the end of their time as a tropical cyclone.

It should be noted that the historical wind radii dataset used to calculate historical IKE values is subject to year-to-year and storm-to-storm inconsistencies. Landsea and Franklin (2013) estimated that uncertainties for operational 30-knot, 50-knot, and 64-knot radii were quite large during the 2010 season. In fact, they estimate that on average the error bars for each radii estimate could exceed 30% of the mean size for each radii. Overall, they find that the accuracy of the historical wind radii is limited to the data available in the operational or post-storm analyses, wherein operational radii measured by aircraft reconnaissance have less uncertainty than do those measured by just satellite observations. Furthermore, Landsea and Franklin (2013) indicate that landfalling storms and more intense storms typically have less uncertainty, as these types of storms are generally observed more frequently and with better equipment, especially landfalling storms which can be observed from the ground. Therefore, the historical IKE dataset is likely subject to the same inconsistencies of data quality and quantity that is evident within the wind radii data in the extended best track data.

b. Potential Model Predictors

In addition to the historical record of IKE, a pool of potential predictors must also be created to establish the relationships that will govern the statistical model. These predictors are carefully selected based on some of the understood relationships between a storm's environment and the factors that govern the size, strength, intensity, and ultimately kinetic energy of a TC (e.g. Gray 1968; McBride 1995; Hill and Lackmann 2009; Maclay 2008; Musgrave et al. 2012).

These environmental, storm-specific, and persistence predictors are gathered from a combination of the NHC best track data set (Jarvinen, 1984) and the SHIPS developmental data set (DeMaria and Kaplan 1999). The variables derived from the NHC best track data are storm-

157 specific predictors, such as date, position, duration, intensity, and translational storm motion.
158 The environmental variables, which encompass thermodynamic, dynamic, and moisture related
159 fields known to affect TC behavior, are taken from the aforementioned SHIPS developmental
160 dataset (DeMaria and Kaplan 1999). In all, thirty-one predictors were considered for this
161 regression exercise (Table 1).

162 It should be noted that the SHIPS developmental dataset, from which the environmental
163 data is obtained, is not a forecast, and instead it utilizes analyses and reanalyses from the
164 National Centers for Environmental Prediction (NCEP) to provide estimates for the observed
165 environmental conditions experienced by each storm from genesis to dissipation. Therefore, this
166 model is trained using a so called “perfect prog” approach (e.g. Neumann 1987; DeMaria,
167 2010), wherein observations and analyses are used to drive a statistical model. Obviously, an
168 operational real-time version of SPIKE will require the predictors to be forecasted by numerical
169 models, since the analyses and reanalyses are not available for future time steps in a real-time
170 operational setting. Therefore, the SPIKE models discussed for the remainder of Sections 3-5,
171 all of which utilize observations and SHIPS developmental data, are not meant to be forecasts
172 or even hindcasts. Instead these regression models will provide an estimate for the maximum
173 potential skill of SPIKE in the idealistic scenario that the forecasted predictors exactly match the
174 future observations.

175

176 3. Regression Methodology

177 The distribution of historical six-hourly IKE values is decidedly non-Gaussian as shown in
178 Figure 1. In fact, the six-hourly total IKE values are approximately log-normally distributed,
179 which is similar to the distribution of storm size as measured by the radii of vanishing winds
180 (Dean et al. 2009). Therefore, it would be inappropriate to use linear regression to model total
181 IKE values. Instead, SPIKE will seek to predict changes of IKE from initial time out to twelve

evenly-spaced time intervals from six hours to seventy-two hours (0-6hrs, 0-12hrs, … , 0-72hrs). Ultimately, these IKE changes are more normally distributed (e.g. Figure 2), and thus, it becomes more appropriate to use multivariate linear regression. However, in order to create a model for each of these twelve forecast intervals, SPIKE regression models must be calibrated and validated separately for each forecast interval.

Based on the success of the statistical SHIPS model used to predict intensity change relative to other statistical and dynamical models (DeMaria et al. 2005), we utilize similar regression methodology to construct our SPIKE model for IKE changes. As done by DeMaria and Kaplan (1994), both the dependent and independent variables are normalized prior to training the regression model. Ultimately, by normalizing the predictors, it becomes possible to make a comparison between regression coefficients for various predictors and forecast hours (DeMaria and Kaplan 1994).

To avoid over fitting in our SPIKE regression model, predictor screening must be utilized. Once again, we follow the lead of DeMaria and Kaplan's SHIPS model (1994; 1999) by using backwards screening to objectively select the most skillful predictors. Backward predictor screening is done here by training the model upon all of the predictors for each forecast interval, and then repeatedly removing the single predictor with the least significant regression coefficient one at a time, until all of the remaining regression coefficients are significant at the p=0.01 level. Ultimately, this backward screening methodology retains a smaller subset of predictors that are used in the SPIKE model for each forecast interval. To make the SPIKE prediction model as uniform as possible across the forecast intervals, the same predictors are chosen for all intervals. These predictors are selected if their coefficients are significant at the 99% level for at least half of the forecast intervals. As a result, predictors may be used on intervals when their coefficients are not significant, but as pointed out by DeMaria and Kaplan (1994), when a predictor is not significant, its coefficient becomes small and its influence on the regression model is diminished.

It should also be noted that some of the SHIPS predictors included in Table 1 have missing or null data values scattered throughout our training interval of 1990-2011. Most notably, some of the subsurface ocean predictors (i.e. RD26) are unavailable for entire years at the beginning of the record. Therefore, to account for the limited number of missing predictor values, the regression coefficients are calibrated and the predictors are screened using a sample of all storm fixes that contain no missing data in the training interval. In other words, if a certain storm fix contains even one missing predictor, that data point will not be used in the calibration of the model. However, once the predictors are chosen and the coefficients are calculated, we estimate IKE for every storm fix to evaluate model skill. In the case that a predictor is missing, as occurs occasionally in an operational setting, we filled the missing data point with a value equivalent to the sample mean (i.e. exactly zero since the predictors are all normalized).

4. Results

This section presents the results of the SPIKE regression model. In addition to estimating fluctuations of IKE, SPIKE will also be adapted to estimate total IKE by incorporating persistence values of kinetic energy. An emphasis will be made in interpreting the physical relationships that drive the regression model in the first section, because a statistical relationship is meaningless without an understanding of the underlying physical processes. Later sections focus on evaluating the calibration and predictive skill of SPIKE with explained variance and mean error statistics as well as a standard bootstrapping exercise. Finally, the last section analyzes SPIKE's ability to project IKE values during the 2012 Atlantic Hurricane season in order to evaluate the model's skill on a season that is outside of the training interval.

a. Physical Interpretation of Selected Predictors

The predictors retained through the backward screening exercise are shown in the first column of Table 2. These predictors encompass a wide array of variables ranging from thermodynamical fields such as the depth of the 26°C isotherm to positional variables such as latitude, dynamical values like upper-level divergence, and persistence variables such as past values of IKE. For the sake of simplicity, the variables are referenced as they are abbreviated in Table 1 for the rest of this discussion.

The coefficients for these variables at selected forecast intervals are also shown in Table 2. Encouragingly, the sign of most of the predictors' coefficients do not vary with forecast hour. For example, the coefficient for PIKE is negative in all intervals suggesting that storms with higher IKE are more likely to have decreasing IKE over time. As seen in the distribution of historical IKE values, TCs most often have low values of IKE; therefore it is not terribly surprising that higher IKE storms typically weaken. The physical reasoning behind this relationship is tied to the timing of maximum IKE during a TC lifecycle. As found by Musgrave et al. (2012), TCs often exhibit storm growth (increasing IKE) through most of their lifecycle. North Atlantic TCs, in particular, maintain or increase their size even after reaching maximum intensity (Knaff et al. 2014). As a result, TCs often have their highest levels of IKE late in their lifecycle, either prior to landfall or during extratropical transition (ET) when TCs often undergo wind field expansion as the RMW moves outward and the outer wind field accelerates (Evans and Hart 2008). Obviously following landfall or the completion of ET, TCs typically weaken drastically over the hostile environments of land or the cold northern Atlantic Ocean. Therefore, the negative coefficient of PIKE can be attributed to the negative IKE change of these strong storms, and the fact that weaker storms near genesis typically will gain IKE as they become more mature, provided the environment is not too unfavorable. Similar to PIKE, PDAY has negative coefficients for all forecast intervals. The negative coefficient is tied to the fact that TCs are more likely to have periods of increasing IKE close to the peak of the season (small PDAY), when conditions are typically most favorable for TC development.

The only predictor to have a coefficient that changes sign with forecast hour is dIKE12, wherein the coefficient is positive in the shorter forecast intervals and slightly negative in the longer intervals. The positive dIKE12 coefficients in the first several forecast intervals make sense, as storms typically continue to have increasing IKE in most phases of their lifecycle (Musgrave et al. 2012), wherein a growing storm will continue to grow provided the environment remains somewhat favorable. Likewise, if a storm is in an environment unfavorable for IKE growth, the kinetic energy will likely continue to drop, at least in the short term. The slightly negative sign for the longer forecast intervals is more difficult to reason physically, but it should be noted that the coefficient is not significant to begin with, suggesting that past 12hr IKE change is only helpful for determining upcoming IKE changes in the immediate future.

In terms of the environmental predictors, some of the underlying physical relationships are immediately apparent. For example, the coefficients for VORT and D200 are positive, suggesting a direct relationship between storm growth and each of these fields. In this case, both low-level vorticity and upper-level divergence are well known conditions that are generally favorable for large scale organized convection and the formation and development of TCs (e.g. McBride 1995). Similarly, the negative coefficients for MSLP and PENV are expected, as a more intense storm and/or a storm with a larger area of low pressure will typically have higher wind speeds and increased IKE with all else being equal. Likewise, the positive coefficients for RD26 are unsurprising, as TCs induce turbulent mixing and upwelling in the upper levels of the ocean which cools SSTs through the entrainment of cooler subsurface waters (e.g. Price 1981). This SST cooling mechanism plays a significant role of slowing down TC growth and intensification, especially for slower moving storms over shallow oceanic mixed layers (Schade and Emanuel 1999). Thus, an environment with a deeper thermocline, and a higher RD26, is more resistant to the negative SST feedback mechanism, making storm growth more favorable. Finally, the positive coefficients for DTL also make sense, as TCs tend to weaken as they approach and eventually cross landmasses.

In some cases, however, the physical processes that govern the regression coefficients are less apparent. For example, the positive coefficients for SHRD and LAT seem somewhat counterintuitive to conventional TC development theories, wherein TCs favor low shear environments as well as warmer oceans which are typically found in the lower latitudes. However, as discovered by Maclay et al. (2008), TC growth can also be tied to external forcing from trough interactions and baroclinic environments over the higher latitudes. The positive coefficients for REFC, for example, reflect the positive influence of trough interactions on storm growth (Maclay et al. 2008; DeMaria et al. 1993). Likewise, the positive coefficients of LAT and SHRD are likely related to ET. Historically, ET occurs in just under half of all Atlantic TCs (Hart and Evans 2001), typically over the more sheared higher latitudes of the basin. Thus the wind field expansion from ET and the subsequent increase in IKE over the higher latitudes is likely influencing the signs of these coefficients. Finally, the negative coefficients for RHLO are particularly counterintuitive because in most cases increased low level humidity is favorable for TC development and also for increased storm size (e.g. Hill and Lackmann 2009). However, this apparent contradiction can be explained going back to the relationship between increasing IKE and extratropical transition, whereas storms undergoing ET often have an intrusion of dry air into the storm circulation (Jones et al. 2003), thus decreasing RHLO. Therefore, it is possible that lower RHLO could be associated with expanding wind fields in ET or other similar events, but additional study of the physical relationship between IKE and RHLO is clearly warranted.

b. Model Skill and Validation Tests, 1990-2011

The shared variance between the SPIKE model and IKE variability is shown for selected forecast intervals during the training period in Table 2. As is the case with SHIPS (DeMaria and Kaplan 1994), the explained variance increases with increasing forecast hour. At first, this appears counterintuitive as forecast skill typically decreases with lead time. However, the

309 average magnitude of IKE change from 1990 to 2011 is much smaller in the shorter forecast
310 intervals than in the larger forecast intervals (9 TJ for 12hr; 32 TJ for 72hr). Considering the
311 errors and biases within the historical archive of operational wind radii, the calculations for
312 observed IKE likely contain biases on the same order as, or even greater than, the IKE changes
313 themselves for the shorter forecast intervals. Therefore, the model will perform poorly at
314 explaining these smaller short-term changes that are dominated by observational biases.
315 Furthermore, the predictors used to train SPIKE in this "perfect prog" approach are observed
316 and not forecasted. Therefore, this exercise is not hurt by forecast biases and errors, which
317 would ostensibly increase with forecast hour. Instead this exercise is a proof of concept for
318 forecasting IKE changes given idealistic perfectly forecasted predictors.
319 The shared variance scores of SPIKE for predicting integrated kinetic energy changes are
320 all significant considering the large sample sizes from using thousands of storm fixes between
321 1990 and 2011. The shared variance statistics for SPIKE are particularly impressive at the
322 longer forecast intervals ($r^2 = 0.54$), where they approach and in some cases exceed the shared
323 variance levels for SHIPS and TC intensity (DeMaria and Kaplan 1994; 1999). Admittedly, the
324 model performs quite poorly in the shorter ranges, especially considering observed predictors
325 are used instead of forecasted predictors.
326 Although SPIKE is designed to predict the normally distributed quantity of integrated kinetic
327 energy change, it can still be adapted to predict total IKE values (Figure 3). This is done by
328 adding the estimate of IKE change from SPIKE to the known persistence IKE value. As a result
329 of incorporating persistence into the IKE forecast, the shared variance levels between SPIKE's
330 total integrated kinetic energy estimates are significantly higher than its estimates for just IKE
331 fluctuations. At a forecast interval of 12hrs, SPIKE can estimate total IKE with a staggering
332 explained variance of 84%. This shared variance drops off to 70% by 30 hours and a still
333 impressive 60% by a forecast interval of 72 hours (Figure 4a).

334 The use of persistence to obtain total IKE projections allows the SPIKE products to take
335 advantage of the inertial nature of IKE quantities. Whereas a point metric like maximum
336 sustained wind can and does change rapidly somewhat regularly (e.g. Kaplan and DeMaria
337 2003), kinetic energy integrated across the entire wind field does not change as rapidly.
338 Although drastic intensity changes do impact IKE values, rapid intensification (RI) events
339 typically result in a drastic increase of near surface winds over a small confined area of
340 convection near the center of the storm. Therefore, the impact on IKE during RI is typically
341 small, provided the overall size of the storm's wind field remains somewhat constant. Therefore,
342 a persistence forecast of total IKE is typically very skillful, especially in a short forecast interval.
343 However, at longer forecast intervals, persistence does not fare nearly as well (r^2 = 25% at
344 72hrs; Figure 4a). In fact, the SPIKE regression model is more skillful at estimating total IKE
345 values at a 72-hour forecast interval than is persistence at forecasting the same quantity on a
346 much shorter 30-hour interval. The lack of skill by persistence over time indicates that
347 environmental and storm-specific data must be utilized for longer-term forecasts of IKE.
348 Absolute mean errors were also calculated for the total SPIKE model by computing the
349 average magnitude of the differences between the SPIKE forecast and the observed total IKE
350 values for each of the storm fixes in the 1990-2011 training interval. These mean errors for the
351 total SPIKE model and for a persistence forecast are plotted in Figure 4b. It should be noted
352 that the mean errors for the persistence forecast are equal to the mean magnitude of IKE
353 change for each forecast interval since a persistence forecast by definition will predict a change
354 of zero TJs. Therefore, the mean magnitude of errors for the SPIKE model's total IKE
355 projections are of the same magnitude as the IKE changes themselves within the shortest
356 forecast intervals (e.g. ~10 TJ for the 12 hour forecast interval; ~15 TJ for the 24 hour forecast
357 interval). This ultimately means that a short-term total IKE forecast is not significantly better than
358 persistence. While unfortunate, this does not come as a surprise, as SPIKE has low shared
359 variance scores in these shorter forecast windows for predicting the more subtle IKE changes

(Table 2). However, a persistence forecast has significantly more error than does the SPIKE model in forecast intervals greater than 24 hours. In fact, at a forecast window of 48 hours, total IKE projections from SPIKE have a mean error magnitude of 18.6 TJ compared to a much higher 24.2 TJ from a persistence forecast. Therefore, the mean error statistics support the conclusions drawn from the shared variance statistics, wherein persistence is a tough forecast to beat within 24 hours and the SPIKE model exhibits significant skill over persistence beyond that point.

Analysis of mean squared error (MSE) values in the SPIKE model, with respect to those from a persistence forecast, further emphasize that SPIKE offers a significant skill improvement over a persistence forecast during the training interval (Figure 5). In a twelve-hour forecast window, SPIKE exhibits a 14% reduction of MSE values when compared to a persistence forecast. Two-sample bootstrapping tests using the squared errors of the SPIKE model and a persistence forecast suggest that SPIKE's improvement over persistence is not quite significant at the one-sided p=0.10 level for this short forecast window. However, at 30 hours, SPIKE exhibits a 36% improvement over a persistence forecast in terms of MSE, which is significant at the p=0.01 level according to two-sample bootstrapping tests. The percent decrease of MSE in the SPIKE model relative to persistence within the training interval continues to improve across all later forecast intervals, culminating in a 60% reduction of MSE in the 72-hour SPIKE model compared to a 72-hour persistence forecast.

In addition to simply calculating statistics over the calibration interval, some validation exercises are also performed using standard bootstrapping techniques. These bootstrapping exercises are done by training the model over a sample that is created by randomly selecting data points from the overall population of IKE and predictor data (repetition allowed). The regression coefficients from the model trained over this sample are then used over the original population to examine how much skill is lost. In the case of the SPIKE model for kinetic energy change, there is an average decrease in shared variance of 3.7% across all twelve of the

386 forecast intervals. The decrease of skill for the total kinetic energy estimates is less significant,
387 averaging less than 0.5% across all of the forecast intervals. Ultimately, these simple tests
388 indicate that SPIKE should be able to retain predictive skill when using a different sample of
389 data. However, once again, it should be noted that because developmental SHIPS data is used
390 for the predictors, there likely will be a decrease in skill when using forecasted predictors in an
391 operational setting.
392

393 ## *c. Model Performance in the 2012 Atlantic Hurricane Season*

394 The 2012 Atlantic Hurricane Season consisted of 19 tropical cyclones, ten of which were
395 hurricanes and two that eventually reached major hurricane status. A total of 395 storm fixes
396 were taken from the extended best track dataset to estimate IKE for each storm. Similar to
397 storms in the training interval from 1990 to 2011, the TCs of the 2012 Atlantic season had IKE
398 values less than 25 TJ for most of their lifetimes. However, four storms (Hurricanes Leslie,
399 Nadine, Rafael, and Sandy) all obtained in excess of 100 TJs of IKE, mostly in the latter stages
400 of their lifecycles over the middle and upper latitudes of the basin.
401 The same SPIKE models that were calibrated on the 1990-2011 data in the previous
402 sections (coefficients listed in Table 2) are utilized to project observed fluctuations of IKE during
403 the 2012 Atlantic Hurricane Season, as means of determining SPIKE's potential predictive skill
404 outside of the training interval. Predictor data for the 2012 season is once again taken from the
405 best track dataset (Jarvinen, 1984) and the SHIPS developmental data set (DeMaria and
406 Kaplan 1999). Therefore, this analysis, which does not use dynamically forecasted predictors, is
407 not meant to assess the operational skill of SPIKE. Instead, this exercise serves as an
408 evaluation of SPIKE's skill outside of the training interval when given idealistically accurate
409 predictors.

Overall, the SPIKE model exhibits comparable skill during the 2012 season when compared to the 1990-2011 training interval. The explained variance and mean absolute forecast errors are shown in Figure 6. SPIKE explains an even higher percentage of the observed total IKE variance in 2012 when compared to the training interval. At a forecast interval of 36 hours, the 2012 SPIKE model explains an astonishing 83% of the variance of total IKE values, as compared to the 67% explained by SPIKE during the longer training interval. The main reason for this apparent enhanced skill during the 2012 season, when compared to the longer training interval, likely stems from the fact that persistence corresponds extremely well with future IKE values during the 2012 season. In fact, out to 36 hours, a simple persistence forecast explains more than 80% of the variance during the 2012 season, suggesting that the kinetic energy levels of TCs during this one season were even more inertial than should be expected on average.

In terms of mean absolute errors, SPIKE does not appear to perform as well during the 2012 season when compared to the 1990-2011 training interval (Figure 6b). An increase in forecast error should be expected since the 2012 data was not in the training interval that was used to calibrate the regression coefficients in SPIKE. In this case, the error at a forecast window of 72 hours was 29.2 TJs during the entire 2012 season, which is a staggering 36% higher than the mean 72-hour error during the 22-year training interval. The highest SPIKE errors during the 2012 season unsurprisingly occurred during Hurricane Sandy. As previously mentioned, Sandy had near-record levels of IKE near the end of its lifecycle. The SPIKE model, at all forecast intervals, correctly projected that Sandy would have IKE levels above 150 TJs before reaching the Mid-Atlantic coastline. This forecasted level of IKE from SPIKE would have placed Sandy within the top 2.5% of all TC fixes during the training interval. Therefore, in many regards, the SPIKE forecast still would have been more than adequate to categorize the high damage potential of Sandy, which explains why the shared variance score of SPIKE during the 2012 season is not similarly worse than the training interval shared variance levels.

Nonetheless, SPIKE was unable to project that Hurricane Sandy would reach near-record IKE values exceeding 400 TJs. As a result, even a very high forecast of 300 TJ for Sandy would contain an enormous error magnitude of more than 100 TJs, contributing to the much higher than expected mean error levels shown in Figure 6b.

If the final dozen storm fixes of Hurricane Sandy were neglected, the SPIKE model would have significantly less mean error during the 2012 season (Figure 6b). In fact, the mean error during 2012, excluding these Sandy data points, would be strikingly similar to the mean errors during the full 22-year training interval at all forecast intervals. Mean squared error statistics (Figure 7) tell a similar story, wherein the SPIKE model improves upon a persistence forecast during the 2012 season, excluding Sandy, for all forecast hours. Like the MSE reduction analysis done for the training interval in Section 4b, SPIKE performs increasingly better than persistence with increasing lead time. The percent reduction of MSE in SPIKE relative to persistence becomes significant at the one-sided p=0.05 level at a forecast interval of 48-hours, where SPIKE has 35% less MSE than a persistence forecast. At a forecast interval of 72-hours, SPIKE has a reduction of MSE relative to persistence of nearly 45% during the 2012 season, excluding Sandy.

Overall, this exercise suggests that the SPIKE model performs well over a validation dataset such as the 2012 season. Skill relative to persistence increases with increasing lead time, as expected. Within a twenty-four hour forecast window, the high correspondence between SPIKE projections and historical IKE in 2012 can be attributed to the skill of a short-term IKE persistence forecast. However, outside of this one-day forecast interval, the SPIKE model performs significantly better than a persistence forecast in this validation exercise.

5. Conclusions

Although kinetic energy forecasts are uncommon in operations today, IKE is a metric that could be potentially very useful to forecasters. The relationship between kinetic energy metrics and the storm surge damage potential of a tropical cyclone makes forecasting this physical quantity very desirable, especially since IKE better represents damage potential in larger storms such as Sandy than does VMAX. On the other hand, it should be noted that one weakness of kinetic energy metrics is their underestimation of wind damage in smaller but intense storms such as Charley in 2004 (Maclay et al. 2008). Therefore, it is suggested that IKE forecasts should be used in conjunction with existing VMAX forecasts to maximize a forecaster's ability to assess the wide array of risks for damage in landfalling storms.

The simple statistical regression model designed here adequately estimates changes of IKE in Atlantic TCs out to 72 hours. More specifically, the model created here explains as much as 50% of the variance in historical IKE changes out to three days. More impressively, it is found that persistence can be added to SPIKE's forecast for IKE change to successfully project total values of IKE. In this regard, SPIKE can explain more than 80% of the observed variance in historical total IKE values at a 12-hour forecast interval, trailing down to near 60% at 72 hours. The increase in skill for projecting total IKE is attributed to the inertial nature of the IKE metric. The fact that persistence is a viable kinetic energy forecast in the short term could be used to the advantage of forecasters in assessing TC risks, especially considering the lack of recent improvements in forecasting the less persistent and notoriously challenging intensity metrics (Rappaport et al. 2009).

Validation tests suggest that SPIKE's high levels of skill should not degrade sharply when used in a forecast environment, provided that the predictors are forecasted accurately. For example, bootstrapping exercises revealed that SPIKE remains skillful at projecting total IKE values in historical storms, even when calibrated by resampled data. Furthermore, validation of SPIKE during the 2012 season reveals that the model can perform reasonably well when given predictor data outside of its training interval, as will be required in an operational setting.

However, it should be once again noted that none of these exercises used forecasted predictors. Thus, it is fair to expect SPIKE to have somewhat reduced skill levels in a truly operational setting because the predictors will be imperfectly forecasted for future time steps in nearly all scenarios.

Previous work done by DeMaria (2010) offers some insight with regards to the amount of skill that could be lost by moving from a "perfect prog" approach to an operational approach with imperfectly forecasted predictors. DeMaria (2010) found that the skill of a statistical-dynamical model, such as the Logistic Growth Equation Model (LGEM), did not drastically improve when given perfect large-scale environmental predictors. Since most of the predictors in our SPIKE model are indeed large-scale environmental predictors, this bodes well for the potential skill of the SPIKE model in an operational setting. However, DeMaria (2010) noted that smaller scale predictors that are more dependent upon track are more likely to affect model skill when moving from a "perfect prog" approach to an operational approach with forecasted predictors. This is relevant to SPIKE, as MSLP is used in our model. The intensity of a TC obviously is quite dependent on storm track, especially when a storm is nearing a coastal environment. Considering that MSLP is an important predictor in the SPIKE model, it is worthwhile to examine how an imperfect estimation of MSLP will affect the skill of SPIKE. Therefore, we replaced the observed MSLP predictor in our "perfect prog" approach with an MSLP persistence forecast, leaving all other "perfect" predictors the same. In most cases, a persistence forecast of MSLP will be a poor indication of future MSLP values, as TC intensity tends to change somewhat rapidly, particularly when storms near landfall. Nonetheless, the skill of the SPIKE model decreases by no more than 20% across all forecast intervals, in terms of both explained variance and MSE, when the poor persistence MSLP predictor is used in place of the "perfect" observed MSLP predictor used throughout the previous sections. In fact, a SPIKE model trained with an MSLP persistence predictor is still significantly better at projecting integrated kinetic energy than an IKE persistence forecast at the one-sided p=0.10 level in the training interval for

all forecast intervals exceeding 24 hours. This further suggests that the SPIKE model will likely offer an improvement over a persistence forecast of IKE, even when using imperfect operational predictors.

In summary, this exercise serves as a proof of concept that when given accurately forecasted predictors, it is possible to project IKE in an operational setting. This work can be built upon by using IKE forecasts to statistically estimate storm size and wind field distribution. Future work will also be focused on adapting this model into a dynamical-statistical model that can be used in real time, while also continuing to build upon our understanding of the underlying physical processes that control kinetic energy variability.

Acknowledgements

Thanks to Drs. Mark Powell, Robert Hart, Phillip Sura, Allan Clarke, Ming Ye, T.N. Krishnamurti, and James O'Brien for their helpful comments and feedback. In addition, we greatly appreciate the helpful comments provided to us during the review process by Dr. Mark DeMaria and an anonymous reviewer. This work was supported by grants from NOAA (NA12OAR4310078, NA10OAR4310215, NA11OAR4310110), and USDA (027865).

References

Avila L.A. and J. Cangialosi, 2011: Tropical Cyclone Report Hurricane Irene (AL092011) 21-28 August 2011. *National Hurricane Center*. [available online: http://www.nhc.noaa.gov/data/tcr/AL092011_Irene.pdf]

Bell, G. D., and Coauthors, 2000: Climate assessment for 1999. *Bull. Amer. Meteor. Soc.,*81, 1328–1378.

Berg, R., 2009: Tropical Cyclone Report Hurricane Ike (AL092008) 1-14 September 2008. *National Hurricane Center*. [available online: http://www.nhc.noaa.gov/pdf/TCR-AL092008_Ike_3May10.pdf].

Blake, E.S., T.B. Kimberlain, R.J. Berg, J.P. Cangialosi, and J.L. Bevin II, 2013: Tropical Cyclone Report Hurricane Sandy (AL182012) 22-29 October 2012. *National Hurricane Center*. [available online: http://www.nhc.noaa.gov/data/tcr/AL182012_Sandy.pdf]

Dean. L., K.A. Emanuel, and D.R. Chavas, 2009: On the size distribution of Atlantic tropical cyclones. *Geophys. Res. Lett.,* 36, L14803.

DeMaria, M., and J. Kaplan, 1994: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. Wea. Forecasting, 9, 209 220.

——, and ——, 1999: An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern north Pacific basins. Wea. Forecasting, 14, 326-337.

——, M. Mainelli, L.K. Shay, J.A. Knaff and J. Kaplan, 2005: Further Improvements in the Statistical Hurricane Intensity Prediction Scheme (SHIPS). Wea. Forecasting, 20, 531-543.

——, J-J. Baik, and J. Kaplan, 1993: Upper-level eddy angular momentum fluxes and tropical cyclone intensity change. *J. Atmos. Sci.*, **50,** 1133–1147.

——, 2010: Tropical Cyclone Intensity Change Predictability Estimates Using a Statistical-Dynamical Model, *29th Conf. on Hurr. And Trop. Meteor., May 10-14, Tuscon, AZ, Amer. Meteor. Soc., 9C.5.*

Demuth, J.L., M. DeMaria., and J. A. Knaff, 2006: Improvement of advanced microwave sounder unit tropical cyclone intensity and size estimation algorithms. *J. Appl. Meteorol. Climatol.,* **45**, 1573-1581.

Emanuel, K., 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686–688.

Evans, C., and R.E. Hart, 2008: Analysis of the Wind Field Evolution Associated with the Extratropical Transition of Bonnie (1998). *Mon. Wea. Rev.*, **136**, 2047-2065.

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

Hart R.E., and J.L. Evans, 2001: A Climatology of the Extratropical Transition of Atlantic Tropical Cyclones. *J. Climate*, **14**, 546-564.

583 Hill, K.A., and G.M. Lackmann, 2009: Influence of Environmental Humidity on Tropical Cyclone
584 Size. *Mon. Wea. Rev.*, **137**, 3294–3315.
585
586 Irish, J.L., D T. Resio, and J.J. Ratcliff, 2008: The influence of Storm Size on Hurricane Surge.
587 *J. Phys. Oceanogr.,* **38**, 2003-2013.
588
589 Jarvinen, B. R., C. J. Neumann, and M. A. S. Davis, 1984: A tropical cyclone data tape for the
590 North Atlantic Basin, 1886-1983: Contents, limitations, and uses. NOAA Technical
591 Memorandum NWS NHC 22, Coral Gables Florida, 21 pp.
592
593 Jones, S.C., and Coauthors, 2003: The Extratropical Transition of Tropical Cyclones: Forecast
594 Challenges, Current Understanding, and Future Directions. *Wea. Forecasting*, **18**, 1052–
595 1092.
596
597 Kantha, L., 2006: Time to replace the Saffir-Simpson Hurricane Scale? Eos, *Trans. Amer.*
598 *Geophys. Union,***87**,3–6.
599
600 Kaplan, J., and M. DeMaria, 2003: Large-Scale Characteristics of Rapidly Intensifying Tropical
601 Cyclones in the North Atlantic Basin. *Wea. Forecasting*, **18**, 1093–1108.
602
603 Knaff, J.A., S.P. Longmore, D.A. Molenar, 2014: An Objective Satellite-Based Tropical Cyclone
604 Size Climatology, *J. Climate,* **27**, 455-476.
605
606 Landsea, C.W., and J.L. Franklin, 2013: Atlantic Hurricane Database Uncertainty and
607 Presentation of a New Database Format. *Mon. Wea. Rev.*, **141**, 3576-3592.
608

609 Maclay K.S., M. DeMaria, and T.H. Vonder Haar, 2008: Tropical Cyclone Inner-Core Kinetic
610 Energy Evolution. *Monthly Weather Review* **136**:12, 4882-4898.
611
612 McBride, J. L., 1995: Tropical cyclone formation. *Global Perspectives on Tropical Cyclones,*
613 WMO/TD No. 693, Rep. TCP-38, World Meteorological Organization, 63–105.
614
615 Misra V., S. DiNapoli, and M. Powell, 2013: The Track Integrated Kinetic Energy of Atlantic
616 Tropical Cyclones. *Monthly Weather Review* **141**:7, 2383-2389.
617
618 Musgrave K.D., R. K.Taft, J. L.Vigh, B. D.McNoldy, *and* W. H.Schubert, 2012: Time evolution of
619 the intensity and size of tropical cyclones, J. *Adv. Model. Earth Syst.,* **4**, M08001.
620
621 Neumann, C.J., 1987: Prediction of tropical cyclone motion: Some practical aspects. Extended
622 Abstracts, *17th conference on Hurr. And Trop. Meteor.*, April 7-10, 1987, Miami, FL., Amer.
623 Meteor. Soc., 266-269.
624
625 Price, J.F., 1981: Upper Ocean Response to a Hurricane. *J. Phys. Oceanogr.*, **11**, 153–175.
626
627 Powell, M.D., and T.A. Reinhold, 2007: Tropical Cyclone Destructive Potential by Integrated
628 Kinetic Energy. *Bull. Amer. Meteor. Soc.*, **88**, 513–526.
629
630 ——, S.H. Houston, L. R. Amat, and N. Morisseau-Leroy, 1998: The HRD real-time hurricane
631 wind analysis system. J. Wind Eng. Ind. Aerodyn.,77–78,53–64.
632
633 Rappaport, E.N., and Coauthors, 2009: Advances and Challenges at the National Hurricane
634 Center. *Wea. Forecasting,* **24**, 395-419.

Saffir, H., 1975: Low cost construction resistant to earthquakes and hurricanes. *ST/ESA/23, United Nations*, 216 pp.

Schade L.R. and K.A. Emanuel, 1999: The Ocean's Effect on the Intensity of Tropical Cyclones: Results from a Simple Coupled Atmosphere–Ocean Model. *J. Atmos. Sci.*, **56**, 642–651.

Simpson, R. H., 1974: The hurricane disaster potential scale. *Weatherwise*, **27**, 169–186.

Figure Captions

Figure 1: Relative frequency distribution of six-hourly VMAX (panel A) and IKE (panel B) measurements in Atlantic TCs between 1990 and 2011. This sample includes 5498 fixes from 291 storms. Vertical lines are shown to indicate VMAX and IKE values for selected hurricanes just prior to a US landfall. The times of these IKE measurements are as follows: Ike 9/13/08 00Z; Irene 8/28/11 06Z; Sandy 10/29/12 18Z. The three storm points would fall in the top 45% of all TCs points in terms of VMAX and the top 7.5% of TCs in terms of IKE from 1990 through 2011.

Figure 2: Relative frequency distribution of 0-24 hour IKE changes (panel A), and normalized 0-24 hour IKE changes (panel B). The normalized changes of IKE at various forecast times is used as the target predictand for the linear SPIKE regression models

Figure 3: Scatter plot of observed total IKE values vs. estimated total IKE values from a 24-hour SPIKE regression model (blue dots). The dark black line represents the best fit line between the 4239 observed and estimated data points. The dashed line represents a perfect forecast (y=x).

Figure 4: Evaluation of SPIKE model's skill during the 1990-2011 training interval. Plot of shared variance over forecast hour (panel A) and mean absolute error over forecast hour (panel B) for total IKE. The blue line represents these metrics between the observations and the SPIKE model at each forecast hour. The red line represents these metrics between a persistence forecast and the observed IKE value a certain number of hours later.

Figure 5: Percent reduction of mean squared error (MSE) for the SPIKE model during the 1990-2011 training interval with respect to persistence. A reduction of MSE is plotted as a positive

percentage, indicating improved model skill. The model's improvement over persistence is significant at the one-sided p=0.05 level for all forecast hours greater than or equal to 24 hours and at the p=0.01 level for all forecast hours greater than or equal to 30 hours.

Figure 6: Evaluation of SPIKE model's skill during the 2012 Atlantic Hurricane season. Panel A depicts a plot of shared variance over forecast hour for total IKE measurements. More specifically, the blue line is the shared variance between the SPIKE model and observed total IKE measurements during the training interval of 1990 and 2011 reproduced from Figure 4a, the green line is the shared variance between the validation SPIKE model and observed total IKE values for all storms during the 2012 season, and the magenta line is the shared variance between observations and persistence during the 2012 season. Panel B depicts mean absolute error over forecast hour for total IKE. The blue line represents the mean error between the observations and the SPIKE model during the training period as reproduced from figure 4b. The green line represents the mean error of the validation SPIKE model for all storms in the 2012 season. Finally, the orange line represents the mean error of the validation SPIKE model for the 2012 season, excluding the final 3 days of Hurricane Sandy's lifecycle.

Figure 7: Percent reduction of mean squared error (MSE) for the SPIKE model during the 2012 season, excluding the final seventy-two hours of Hurricane Sandy's lifecycle, with respect to a persistence forecast. A reduction of MSE is plotted as a positive percentage, indicating improved skill. The model has significantly lower MSE than persistence at the p=0.05 level for all forecast hours greater than or equal to 48 hours.

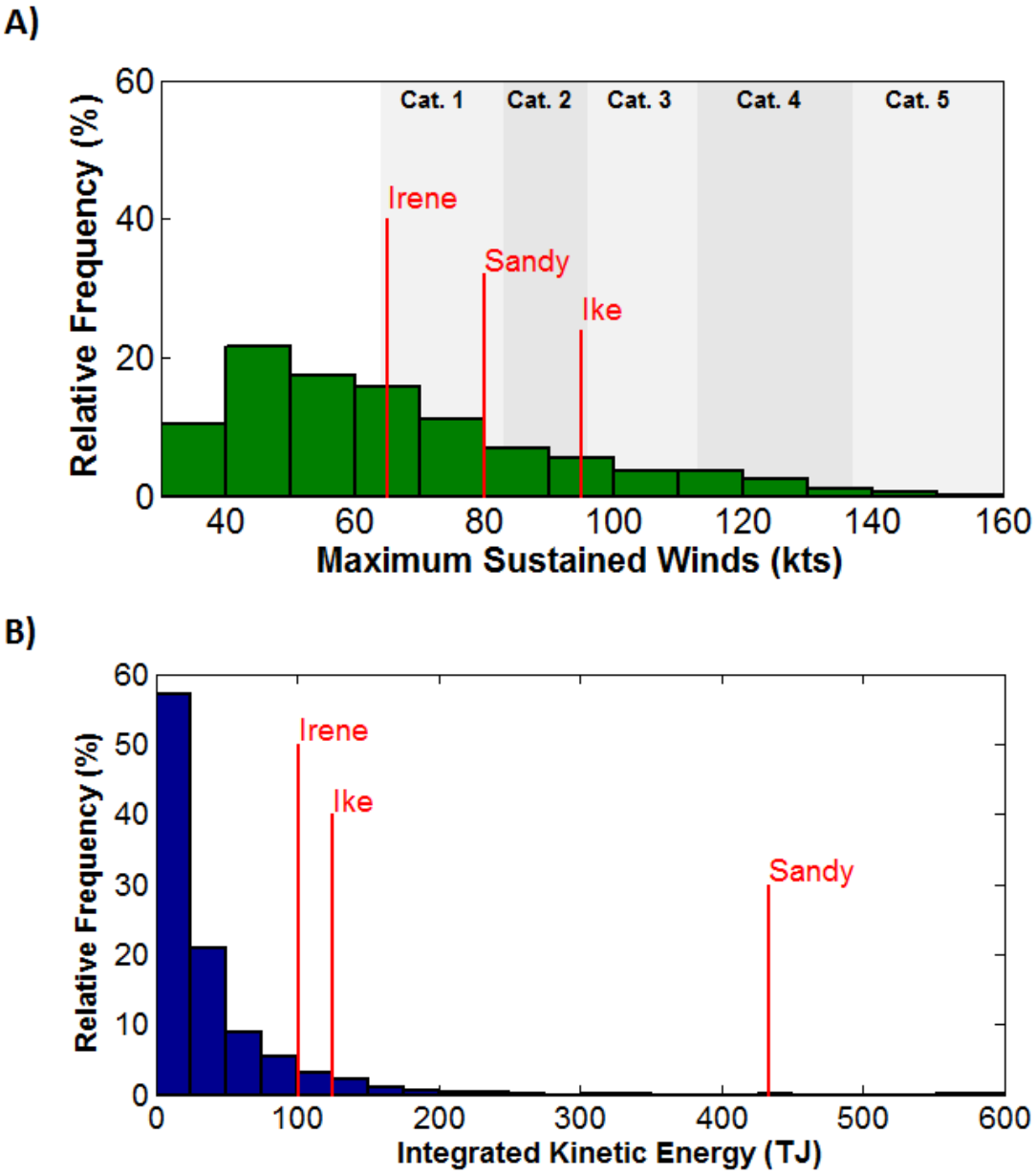
Tables and Figures

Variable	Definition	Units
PIKE	persistence of IKE	TJ
dIKE12	previous 12hr change of IKE	TJ
VMAX	maximum sustained wind speed	kts
dV12	previous 12hr change of VMAX	kts
VMPI	Difference between maximum potential intensity and VMAX	kts
LAT	latitude of storm's center	°N
LON	longitude of storm's center	-°W
DTL	distance to nearest landmass	km
UMOT	zonal translational storm motion	$m \cdot s^{-1}$
VMOT	meridional translational storm motion	$m \cdot s^{-1}$
MSLP	minimum sea level pressure	hPa
PENV	average surface pressure *(averaged from r=200-800km)*	hPa
VORT	850 hPa vorticity *(r=0-1000km)*	10^{-7} s^{-1}
D200	200 hPa divergence *(r=0-1000km)*	10^{-7} s^{-1}
SHRD	850-200 hPa shear magnitude *(r=200-800km)*	kts
SHTD	850-200 hPa shear direction *(r=200-800km)*	°
SHDC	SHRD but relative to 850 hPa center with vortex removed *(r=0-500km)*	kts
RHLO	850-700 hPa relative humidity *(r=200-800km)*	%
RHMD	700-500 hPa relative humidity *(r=200-800km)*	%
T150	150 hPa temperatures *(r=200-800km)*	°C
RD20	ocean depth of 20°C isotherm	m
RD26	ocean depth of 26°C isotherm	m
OHC	ocean heat content	$kJ \bullet cm^{-1}$
SST	sea surface temperatures	°C
EPSS	average difference between lifted surface parcel θ_e and environment θ_{es} *(r=200-800km)*	°C
TWAC	850 hpa symmetric tangential wind *(r=0-600km)*	$m \cdot s^{-1}$
TADV	low-level temperature advection by thermal wind *(r=0-500km)*	10^{-6} s^{-1}
REFC	relative eddy momentum flux *(r=100-600km)*	$m \cdot s^{-1} \bullet day^{-1}$
PEFC	planetary eddy momentum flux *(r=100-600km)*	$m \cdot s^{-1} \bullet day^{-1}$
SDAY	time after tropical storm genesis	days
PDAY	time from peak of season (Sept. 10)	days

Table 1: Variables considered for use in the SPIKE models. Many of the variables (e.g. RHLO, SHRD) originated from SHIPS and are averaged over specific areas as noted. Others were specifically created from observations for the SPIKE model. Not all of these variables are used in the final models, as many of the regression coefficients fail significance tests in the backward screening methodology.

Variable	Forecast Hour				
	12hr	**24hr**	**36hr**	**48hr**	**72hr**
PIKE	-0.53	-0.70	-0.72	-0.70	-0.70
dIKE12	0.11	*0.02*	*0.02*	*0.00*	*-0.01*
RHLO	-0.13	-0.19	-0.22	-0.23	-0.25
SHRD	0.06	0.11	0.12	0.13	0.13
DTL	0.14	0.15	0.15	0.14	0.15
D200	0.12	0.13	0.12	0.12	0.11
MSLP	-0.26	-0.37	-0.39	-0.40	-0.41
VORT	0.16	0.19	0.20	0.20	0.20
LAT	0.24	0.30	0.33	0.34	0.36
RD26	0.11	0.11	0.10	0.10	0.09
EPSS	0.06	0.08	0.09	0.09	0.11
PENV	-0.08	-0.10	-0.13	-0.14	-0.18
REFC	0.07	0.08	0.07	0.05	*0.04*
PDAY	-0.08	-0.10	-0.11	-0.10	-0.08
Sample Size	4814	4239	3756	3341	2657
Shared Variance	13%	25%	39%	43%	54%

738
739 **Table 2:** Regression coefficients for each predictor in the SPIKE model. The coefficients listed
740 in a non-italicized font are significant at 99% for that forecast hour. Sample size for the training
741 interval of 1990 to 2011 is shown below. Finally, shared variance between the SPIKE regression
742 model, and observed kinetic energy changes are listed in the bottom row.
743



744
745
746 **Figure 1:** Relative frequency distribution of six-hourly VMAX (panel A) and IKE (panel B)
747 measurements in Atlantic TCs between 1990 and 2011. This sample includes 5498 fixes from
748 291 storms. Vertical lines are shown to indicate VMAX and IKE values for selected hurricanes
749 just prior to a US landfall. The times of these IKE measurements are as follows: Ike 9/13/08
750 00Z; Irene 8/28/11 06Z; Sandy 10/29/12 18Z. The three storm points would fall in the top 45% of
751 all TCs points in terms of VMAX and the top 7.5% of TCs in terms of IKE from 1990 through
752 2011.
753

754

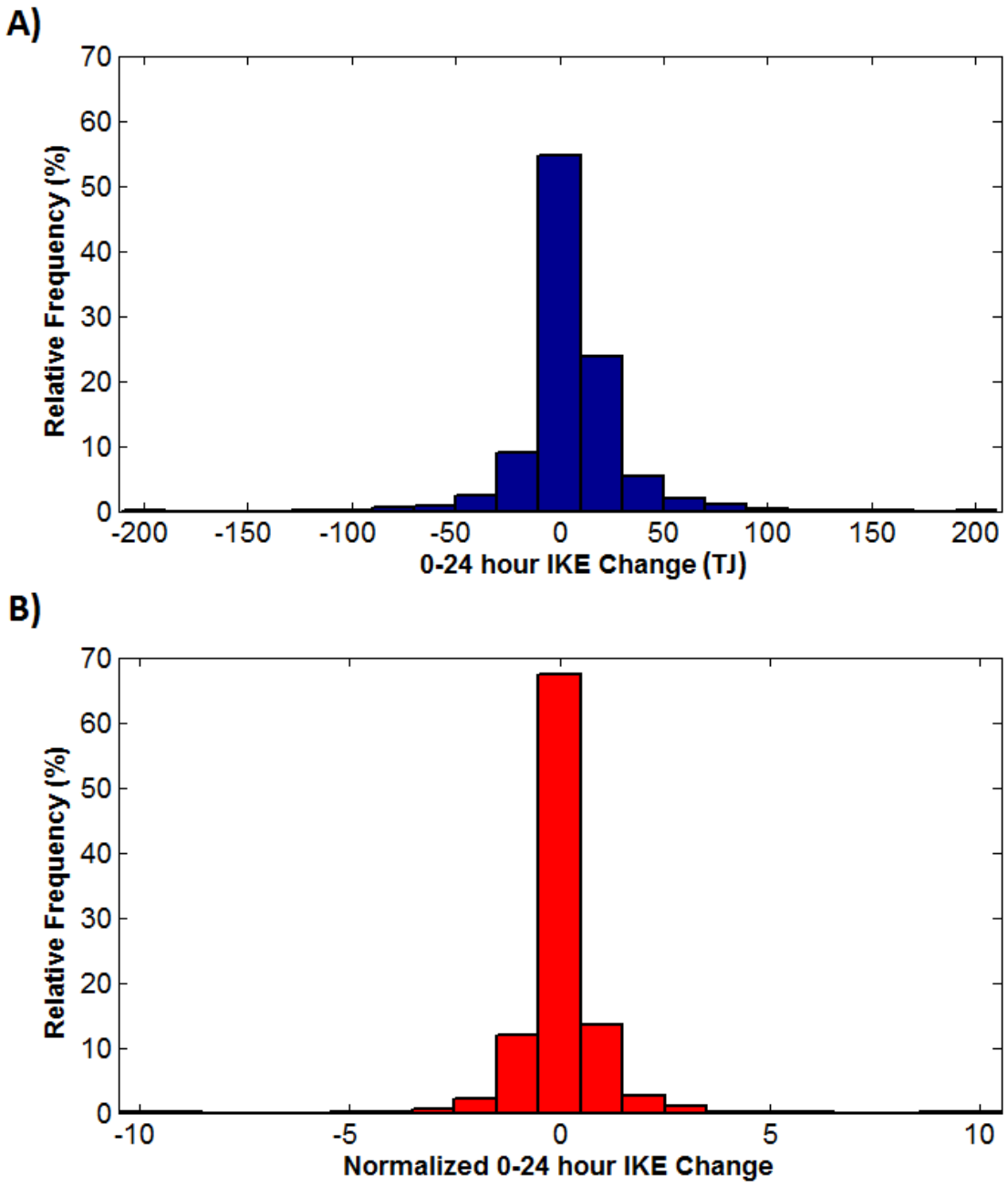


Figure 2: Relative frequency distribution of 0-24 hour IKE changes (panel A), and normalized 0-24 hour IKE changes (panel B). The normalized changes of IKE at various forecast times is used as the target predictand for the linear SPIKE regression models

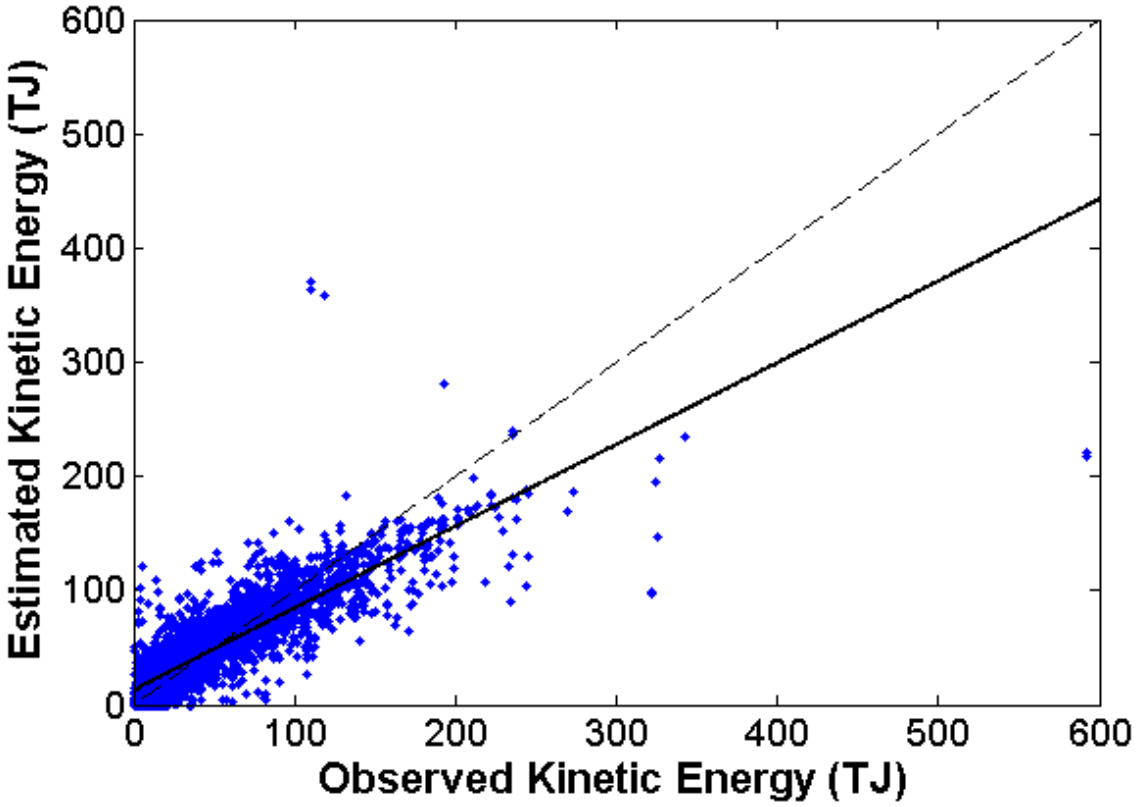


Figure 3: Scatter plot of observed total IKE values vs. estimated total IKE values from a 24-hour SPIKE regression model (blue dots). The dark black line represents the best fit line between the 4239 observed and estimated data points. The dashed line represents a perfect forecast (y=x).

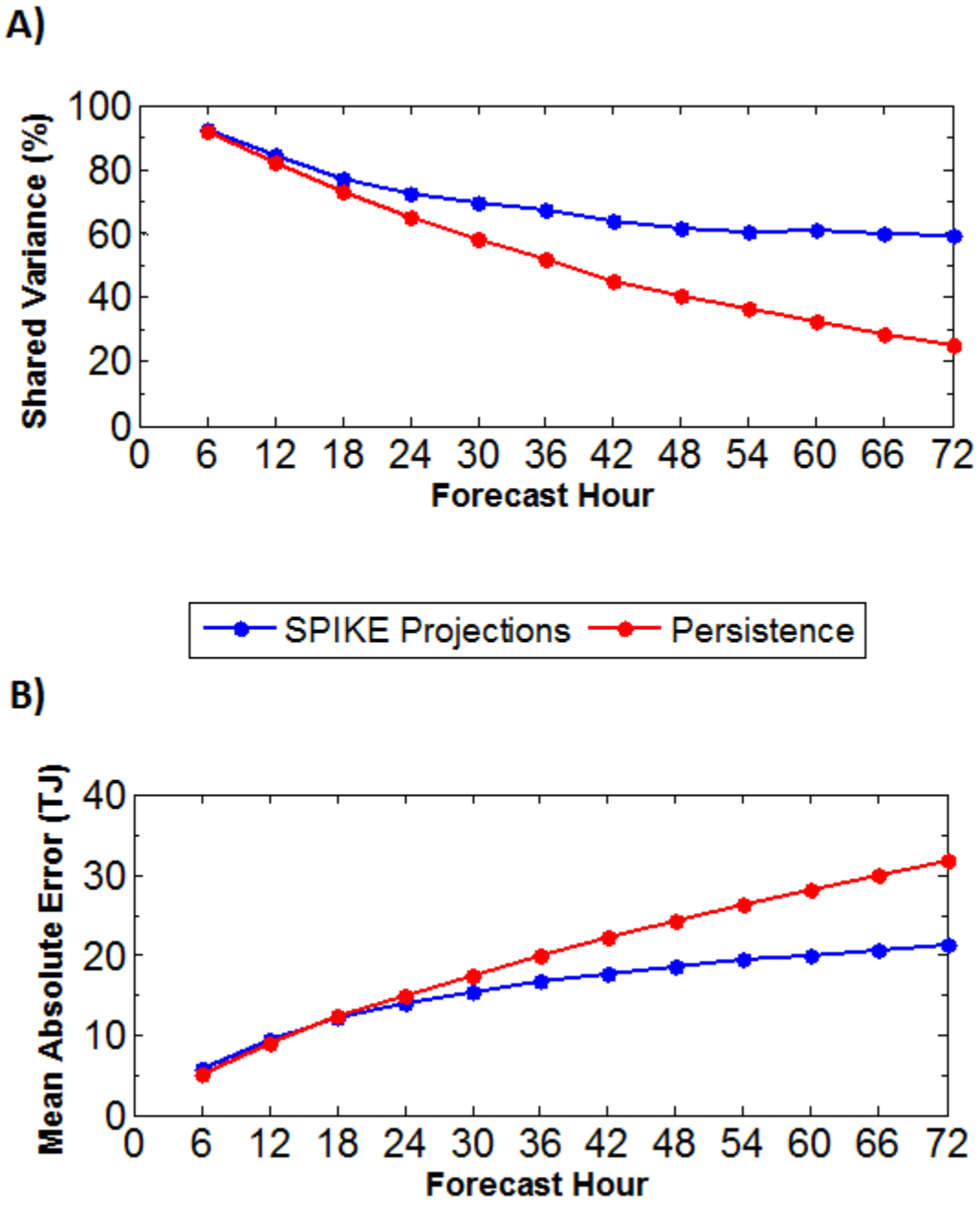


Figure 4: Evaluation of SPIKE model's skill during the 1990-2011 training interval. Plot of shared variance over forecast hour (panel A) and mean absolute error over forecast hour (panel B) for total IKE. The blue line represents these metrics between the observations and the SPIKE model at each forecast hour. The red line represents these metrics between a persistence forecast and the observed IKE value a certain number of hours later.

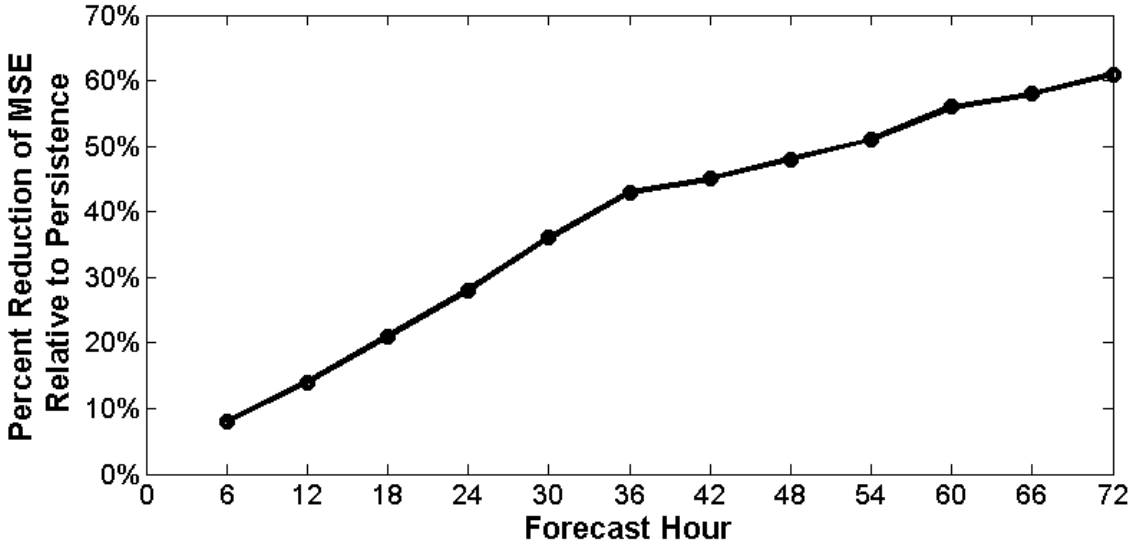


Figure 5: Percent reduction of mean squared error (MSE) for the SPIKE model during the 1990-2011 training interval with respect to persistence. A reduction of MSE is plotted as a positive percentage, indicating improved model skill. The model's improvement over persistence is significant at the one-sided p=0.05 level for all forecast hours greater than or equal to 24 hours and at the p=0.01 level for all forecast hours greater than or equal to 30 hours.

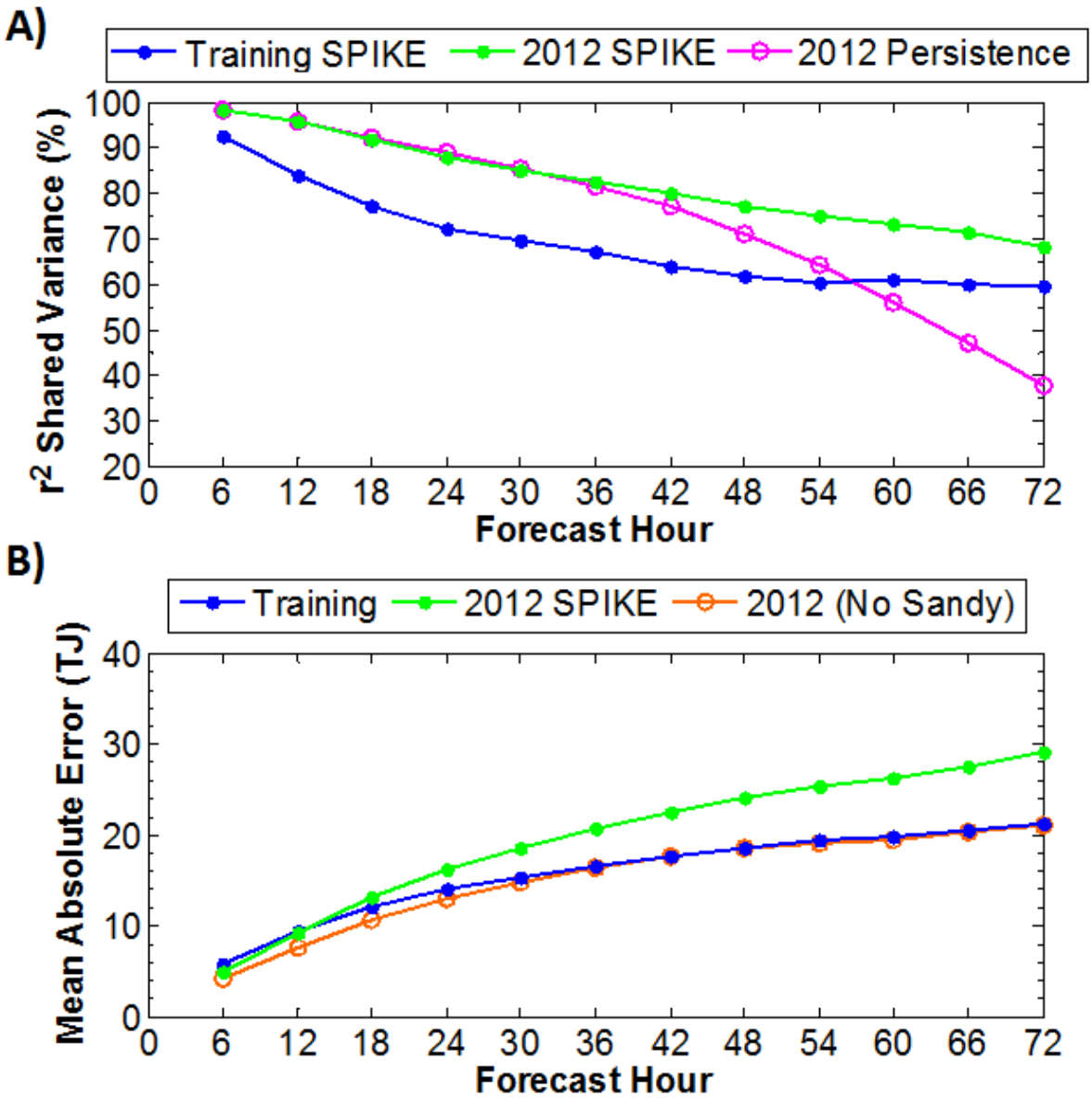


Figure 6: Evaluation of SPIKE model's skill during the 2012 Atlantic Hurricane season. Panel A depicts a plot of shared variance over forecast hour for total IKE measurements. More specifically, the blue line is the shared variance between the SPIKE model and observed total IKE measurements during the training interval of 1990 and 2011 reproduced from Figure 4a, the green line is the shared variance between the validation SPIKE model and observed total IKE values for all storms during the 2012 season, and the magenta line is the shared variance between observations and persistence during the 2012 season. Panel B depicts mean absolute error over forecast hour for total IKE. The blue line represents the mean error between the observations and the SPIKE model during the training period as reproduced from figure 4b. The green line represents the mean error of the validation SPIKE model for all storms in the 2012 season. Finally, the orange line represents the mean error of the validation SPIKE model for the 2012 season, excluding the final 3 days of Hurricane Sandy's lifecycle.

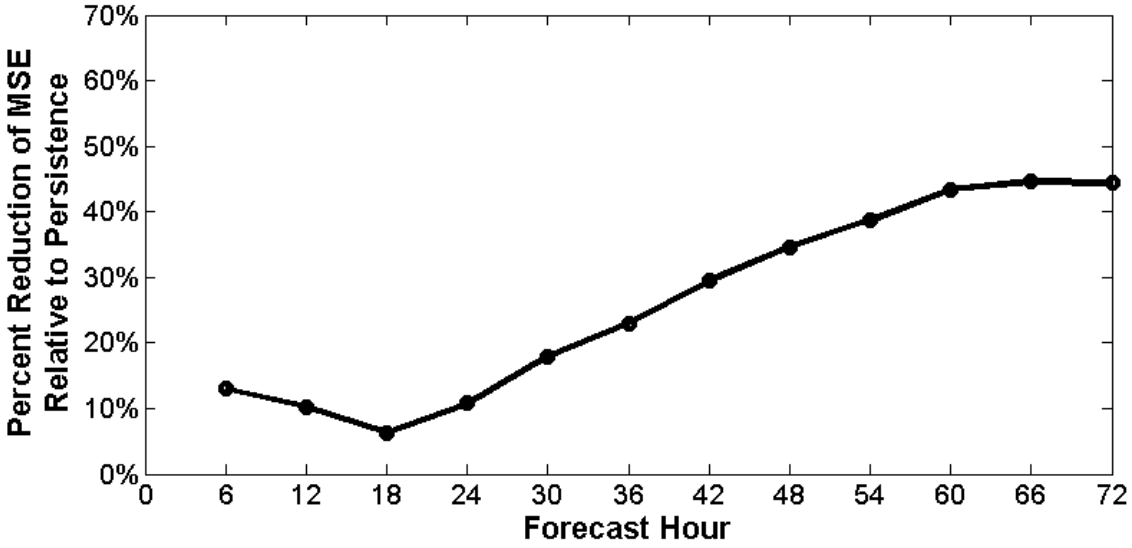


Figure 7: Percent reduction of mean squared error (MSE) for the SPIKE model during the 2012 season, excluding the final seventy-two hours of Hurricane Sandy's lifecycle, with respect to a persistence forecast. A reduction of MSE is plotted as a positive percentage, indicating improved skill. The model has significantly lower MSE than persistence at the p=0.05 level for all forecast hours greater than or equal to 48 hours.