

Article

## Multivariate Analysis of MODerate Resolution Imaging Spectroradiometer (MODIS) Aerosol Retrievals and the Statistical Hurricane Intensity Prediction Scheme (SHIPS) Parameters for Atlantic Hurricanes

Mohammed M. Kamal <sup>1,\*</sup>, Ruixin Yang <sup>2</sup> and John J. Qu <sup>2,3</sup>

<sup>1</sup> School of Physics, Astronomy, and Computational Sciences (SPACS), College of Science, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA

<sup>2</sup> Department of Geography and Geoinformation Science, College of Science, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA; E-Mails: ryang@gmu.edu (R.Y.); jqu@gmu.edu (J.Q.)

<sup>3</sup> Environmental Science and Technology Center (ESTC), College of Science, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA

\* Author to whom correspondence should be addressed; E-Mail: mkamal@gmu.edu; Tel.: +1-703-582-8825; Fax: +1-703-971-0314.

Received: 1 August 2012; in revised form: 7 September 2012 / Accepted: 14 September 2012 / Published: 24 September 2012

---

**Abstract:** MODerate Resolution Imaging Spectroradiometer (MODIS) aerosol retrievals over the North Atlantic spanning seven hurricane seasons are combined with the Statistical Hurricane Intensity Prediction Scheme (SHIPS) parameters. The difference between the current and future intensity changes were selected as response variables. For 24 major hurricanes (category 3, 4 and 5) between 2003 and 2009, eight lead time response variables were determined to be between 6 and 48 h. By combining MODIS and SHIPS data, 56 variables were compiled and selected as predictors for this study. Variable reduction from 56 to 31 was performed in two steps; the first step was via correlation coefficients (cc) followed by Principal Component Analysis (PCA) extraction techniques. The PCA reduced 31 variables to 20. Five categories were established based on the PCA group variables exhibiting similar physical phenomena. Average aerosol retrievals from MODIS Level 2 data in the vicinity of UTC 1,200 and 1,800 h were mapped to the SHIPS parameters to perform Multiple Linear Regression (MLR) between each response variable against six sets of predictors of 31, 30, 28, 27, 23 and 20 variables. The deviation among the predictors Root Mean Square Error (RMSE) varied between 0.01 through 0.05 and,

therefore, implied that reducing the number of variables did not change the core physical information. Even when the parameters are reduced from 56 to 20, the correlation values exhibit a stronger relationship between the response and predictors. Therefore, the same phenomena can be explained by the reduction of variables.

**Keywords:** Atlantic hurricane; dust aerosol; humidity; optical depth; MODIS; SHIPS; intensity change

---

## 1. Introduction

Hurricane forces cause enormous natural disasters. Fortunately, the destruction capacity can be predicted ahead. Satellite measurements for hurricanes and the vast amount of data gathered by hurricane hunters; enable us to measure the force and to track hurricanes. National Oceanic and Atmospheric Administration (NOAA) hurricane hunter [1] airplanes fly directly into the hurricane eye to collect important data about the hurricane. Based on the wind speed, pressure and humidity received by the airplane, forecasters can explain whether the hurricane is weakening or intensifying. Factors such as vertical wind shear [2–5], atmospheric moisture [6,7], air temperature [6], sea surface temperature [8,9] and dust aerosols [6,7,10] may also impact the intensity [11,12] of the hurricane after it has formed. To sustain a strong hurricane, water temperature above 80°F and warm water depths of 150 feet are needed while strong vertical shear in the atmospheric horizontal winds around the hurricane dampen its force [1]. Extremely dry conditions in the mid-atmosphere may act as an agent of taming hurricane force as well [6,10]. Also, Houze *et al.* [13] reported the dynamics of the internal structure of the vortex are responsible for hurricane intensity changes, and they suggested improvements on physical understanding in forecasting hurricane intensity modeling of the internal structure of the vortex.

Satellite observations, hurricane hunters' data collection and numerical weather predictions have advanced the forecasting of hurricane tracks over the last few decades. However, there have been limited improvements in forecasting hurricane intensity [1,14]. Among the models used in the National Hurricane Center (NHC) for hurricane intensity forecasting, the Statistical Hurricane Intensity Prediction Scheme (SHIPS) model is known as the most trusted in regard to intensity forecast models [15] based on the 2011 National Hurricane Center Forecast Verification Report. SHIPS database provides values of parameters related to Tropical Cyclones (TC), but there is a lack of information on dust aerosols which also affect hurricane intensity.

Rosenfield *et al.* [7,16] described the relationship between the intensity change and the sum of anthropogenic aerosols which was calculated as the Aerosol Optical Thickness (AOT) for black carbon (BC), organic carbon (OC), dust (DU) and sulfate (SU). The sum of DU, BC and OC is called "Pollution" while and Total AOT (TAOT) as the sum of DU, BC, OC and SU [16]. Although studies demonstrated the roles of both the SHIPS parameters, aerosol related parameters on the TC intensity changes, the combining roles is not commonly investigated. In his study, the response variable (intensity change) was examined against the "pollution" and "TAOT" [16] and CCNO [7]. Zhang *et al.* proposed a new physical mechanism by conducting simulations with CCN added at the periphery of a

TC to demonstrate large amounts of CCN can influence the eyewall development [17]. Zipser *et al.* discussed an improved understanding of the linkage between AEWs, the SAL, and tropical cyclogenesis by pointing out (a) the difference between AEWs that develop into TCs and those that do not (b) the fate of the AEW by the roles of SAL and (c) vertical distribution, microphysical and optical properties characteristics in composition of the African dust [18]. Gao *et al.* [19] studied the influence of air pressure, temperature, relative humidity, and wind velocity on predicting air pollution from MODIS AOT data without employing SHIPS parameters. Braun, S.A. [20] concluded that the Saharan Air Layer (SAL) is just one of many possible influences and can be both positive and negative and emphasized that aerosol is not the major negative influence on hurricanes. Khain *et al.* [21] came up with an additional mechanism which is related to the TC circulation and described that aerosols significantly affect the spatial distribution of cloudiness and hydrometeor contents. It is imperative to know the intensity change of the hurricane force in advance based on temperature, moisture, vertical shear as well as aerosol retrievals. In this paper, therefore, we focused on the important relationship based on analyzing hurricane intensity change records and the combination of MODIS aerosol retrievals and SHIPS parameters over the North Atlantic spanning several hurricane seasons.

The combination of SHIPS [22–24] and MODIS [25,26] variables created a large set of variables. It is difficult to clearly explain the physical processes with such a large number of variables. Therefore, a reduction of variables in two steps (1) Correlation Coefficients (cc) and (2) Principal Component Analysis (PCA) is introduced. Step 1 is a selection process for screening. The idea of step 2 is to describe the same meaningful physical phenomena by a smaller set of derived variables which will be linear combinations of the original variables. Reducing the number of variables may lead to some loss of original information of the dataset. However, PCA makes this loss minimal and will present a precise meaning without losing original information.

## 2. Data Source

In December 1999, a new generation multi-spectral satellite (Terra, EOS AM-1) was launched carrying the first MODIS sensor. The second MODIS sensor was launched on the Aqua (EOS PM-1) platform on May 2002. Both MODIS sensors onboard Terra and Aqua platforms have been used to monitor the environment continuously in a wide range of spectral frequencies from the blue to the thermal infra-red range. MODIS is an exceptional source for monitoring the Earth's water cycle and environment as both Terra and Aqua satellites have a sun-synchronous orbit at 705 km height. Aqua in ascending mode crosses the equator daily at 1:30 p.m. while Terra, in descending mode, crosses the equator at 10:30 a.m. daily [27].

The MODIS aerosol product measured over the ocean [28,29] is retrieved based on an algorithm for the remote sensing of tropospheric aerosol, and it is different from the aerosol over land [30]. MODIS observed reflectances were matched to a lookup table of pre-computed reflectances for a wide range of normally observed aerosol conditions for both algorithms [30].

The reflectance is calculated from the geometry pertaining to the state of the ocean [31]. Better ocean surface characterization enables [31] the use of reflectances at seven wavelengths (0.47, 0.56, 0.65, 0.86, 1.24, 1.64, and 2.13  $\mu\text{m}$ ) in the retrieval algorithm. The retrieved aerosol products are then represented by the best fit between observed reflectance and the lookup table [31].

Aerosol measurements from MODIS over the oceans, such as aerosol optical thickness and aerosol size distribution can be retrieved from the daily Level 2 data at the spatial resolution of a 10 km × 10 km pixel array at nadir from MODIS Atmospheric Product website [25]. These Level 2 aerosol data products, MOD04\_L2 and MYD04\_L2 [31] are collected from the Terra and Aqua platforms respectively [32].

SHIPS data was collected based on DeMaria *et al.* [22–24] and data files can be found via the Internet [33]. SHIPS model combines climatology, atmospheric environmental parameters, and sea surface temperature as its predictors to forecast intensity changes using a multiple regression scheme [16].

The National Hurricane Center (NHC) of National Weather Service (NWS) issues public advisories for Atlantic tropical cyclones every six hours. Based on the NHC website [34], Table 1 was compiled to describe the anatomy of the twenty four selected hurricanes between the years 2003 and 2009. The time frame for this selection was chosen for example only. Spatial and temporal data for all hurricanes were collected focusing on the hurricane center while it is moving towards the west and north-west above the ocean. Hurricanes near landfall were not in the scope of this study.

**Table 1.** Twenty four hurricanes selected between 2003 and 2009 based on the category 3, 4 and 5 and their lifespan.

Year	CAT	Hurricane	Life Span	Wind Speed (mph)	Pressure (mb)	Start [LAT LON]	End [LAT LON]
2003	4	Fabian	27 Aug–8 Sep	125	939	14.60–31.50	49.80–39.20
2003	5	Isabel	6–19 Sep	140	920	14.00–34.00	42.00–80.70
2003	3	Kate	25 Sep–7 Oct	110	952	11.70–38.30	49.30–45.80
2004	3	Alex	31 Jul–6 Aug	105	957	30.60–78.60	47.50–34.60
2004	4	Charley	9–15 Aug	124	941	11.70–61.10	43.00–69.00
2004	4	Frances	25 Aug–9 Sep	125	935	11.20–36.00	41.40–79.40
2004	5	Ivan	2–24 Sep	145	910	9.70–29.10	31.00–94.90
2004	3	Jeanne	13–28 Sep	110	985	16.00–60.40	37.00–80.30
2004	4	Karl	16–24 Sep	120	938	11.40–32.80	47.30–40.40
2005	4	Dennis	5–13 Jul	130	930	12.50–63.10	38.60–86.80
2005	4	Emily	11–21 Jul	135	929	10.80–42.90	25.00–101.20
2005	5	Katrina	23–31 Aug	150	902	23.20–75.50	41.10–81.60
2005	3	Maria	1–10 Sep	100	960	19.00–46.10	43.60–38.60
2005	5	Rita	18–26 Sep	150	897	22.00–69.70	40.80–86.80
2005	5	Wilma	15–25 Oct	150	882	17.60–78.80	41.70–62.80
2006	3	Gordon	11–20 Sep	105	955	20.20–54.50	39.20–16.60
2006	3	Helene	12–24 Sep	105	954	12.50–23.00	40.90–37.50
2007	5	Dean	13–23 Aug	145	918	12.00–31.60	20.50–100.00
2007	5	Felix	31 Aug–5 Sep	145	929	11.80–58.60	14.00–87.00
2008	3	Bertha	3–20 Jul	105	948	12.60–22.70	51.30–35.70
2008	4	Gustav	25 AUG–04 Sep	130	941	15.50–70.10	35.60–93.20
2008	4	Ike	1–14 Sep	125	935	17.60–39.50	36.40–92.50
2009	4	Bill	15–24 Aug	115	945	11.50–34.00	48.60–50.20
2009	3	Fred	7–12 Sep	105	958	12.50–24.50	17.70–33.70

Table 2 has the selected aerosol retrievals (over the ocean only) for the MOD04\_L2 and MYD04\_L2 Scientific Data Set (SDS), List of Collection 051 [28].

**Table 2.** MODerate Resolution Imaging Spectroradiometer (MODIS) Aerosol retrievals.

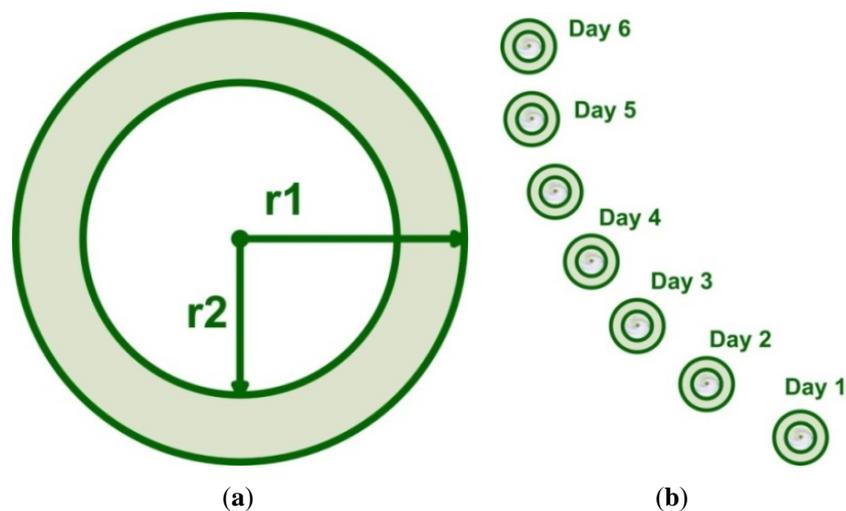
Description	Name
Effective Optical Depth Best Ocean	AOT
Mass Concentration for Best and Average Solutions	MCO
Effective Radius of Both Solutions at 0.55 $\mu\text{m}$	ERO
Column Number of Cloud Condensation Nuclei (CCN) of Both Solutions at 0.55 $\mu\text{m}$	CCNO
Asymmetry Factor for Best Solution <sup>1</sup>	AFBO
Backscattering Ratio of Best Solution <sup>1</sup>	BRBO
Mean Reflectances <sup>1</sup>	MRO

<sup>1</sup> at 7 bands 0.47, 0.55, 0.66, 0.86, 1.24, 1.63, and 2.13  $\mu\text{m}$ .

### 3. Methodology

Pixels close to the hurricane center are usually covered by clouds, making it impossible to retrieve AOT with MODIS measurements. Thus, for this study, a unique technique was developed to select spatial coordinates to investigate aerosol retrievals as shown in Table 2 around each hurricane. Two concentric circles with radii  $r_1$  and  $r_2$ , as shown in Figure 1(a), were drawn with a common center. These circles were drawn to be approximately at the hurricane eye. The spatial regions for this analysis were chosen between the two concentric circles called an annulus. The concentric circle annulus thickness can be adjusted by varying the radii  $r_1$  and  $r_2$ . In this study  $r_1$  and  $r_2$  were selected as 8 and 5 degrees respectively to produce a ring with 3 degrees annulus size. The selected region is far away from the center of the hurricane, but still around the hurricane edge, and can generate enough valid remote sensing measurements for analysis.

**Figure 1.** (a) MODIS data was collected within the annulus of the two concentric circles with radii  $r_1$  and  $r_2$ . (b) Data collection following the motion of a hurricane (for example).



The phenomena being investigated are three-dimensional in case of the variables such as Relative Humidity and Temperature where data is available between sea level and top of the atmosphere (between 100 to 1,000 mb). Vertical Shear and Wind are also three-dimensional phenomena. Although the MODIS sensor on both the Aqua and Terra satellites provides a measure of the vertically integrated dust concentration [20], the vertical distribution of the dust frequency was not considered in this study. In this study we selected MODIS Atmospheric retrievals as two-dimension at seven wavelengths (0.47, 0.56, 0.65, 0.86, 1.24, 1.64, and 2.13  $\mu\text{m}$ ), therefore, circles were used instead of spheres. Therefore, for this investigation of aerosol, retrievals around a hurricane that involves “concentric circles” with the hurricane eye is appropriate.

We investigated whether the 3 to 4 degrees of annulus size would be an appropriate spatial coordinate selection process because aerosol parameters were retrieved around each hurricane by following the direction of motion of a hurricane. Since linear motion of a hurricane is very slow, for example a hurricane’s forward speed averages around 15–20 mph [35], selecting a large annulus size would mostly overlap the spatial region while retrieval happens every day at 1,200 and 1,800 h. Again, selecting a narrow annulus size such as 1 to 2 degrees would introduce significant error while averaging the values within the annulus. Therefore, 3 degrees is the best selection for this study.

MODIS aerosol data at 0.55  $\mu\text{m}$  was averaged in the vicinity of 1,200 and 1,800 h and associated with the corresponding SHIPS data at 1,200 and 1,800 for each day. This technique was employed on a spatial area for studying all 24 hurricanes between the day they formed and the day they dissipated. The center of the concentric circle corresponds to the approximate location of the hurricane core. The angle within this region was spaced out into 36 segments of  $10^\circ$  each. Data for each  $10^\circ$  segment was retrieved and averaged resulting 36 data points at a particular time and date. The readings from these 36 segments were then further averaged to present a final average to demonstrate the values between  $0^\circ$  and  $360^\circ$ . For this analysis, this concentric circles center was programmed to move with the hurricane center for 1,200 and 1,800 h.

Aerosol retrieval variables (Table 2) were retrieved around each hurricane by following the direction of motion of a hurricane as illustrated in Figure 1(b). The response variables “Future Difference (FD) or Intensity Change” has been calculated based on the following formula:  $FD_{\text{future}} = VMAX_{\text{future}} - VMAX_{\text{current}}$ , for example,  $FD_{06} = VMAX_{06} - VMAX_{\text{current}}$ , where VMAX is the maximum 1-min wind speed. Similarly,  $FD_{12}$ ,  $FD_{18}$ ,  $FD_{24}$ ,  $FD_{30}$ ,  $FD_{36}$ ,  $FD_{42}$  and  $FD_{48}$  are calculated. We started our analysis by combining the 49 SHIPS parameters from DeMaria *et al.* [22–24] with the 7 aerosol retrievals as shown in Table 2. Correlation analysis was performed for the intensity change lead time at 06, 12, 18, 24, 30, 36, 42 and 48 h (which are basically the eight response variables between  $FD_{06}$  and  $FD_{48}$ ). For each  $FD_{\text{future}}$  set, correlation analysis will be performed with each of the 56 variables to determine the correlation coefficient (cc) between each variable and the  $FD_{\text{future}}$ . Variables having small correlation ( $|cc| < 0.165$ ) were filtered out. These correlation based filtering create the first set of predictor, Predictor\_1 which comprised 31 variables.

As the second step of data reduction, PCA is carried out on selected variable groups. The reduction of variables for each group by PCA is described in Table 3. Prior to carrying out PCA on the five categories, variables were normalized to avoid skewness caused by units of the variables.

**Table 3.** Reduction of Variables by Principal Component Analysis (PCA).

Category	Variables before PCA	Variables after PCA
Aerosol	AOT MCO CCNO	Aero-PC1 Aero-PC2
Wind	V20C U200 U20C TWAC TWXC	Wind-PC1 Wind-PC2 Wind-PC3
Relative Humidity	RHLO RHMD R000	RH-PC1 RH-PC2
Shear	SDDC SHDC SHGC SHRD SHRG SHRS SHTD SHTS	Shear-PC1 Shear-PC2 Shear-PC3 Shear-PC4
Temperature	SST T250 T200 RD20 ENEG ENSS	Temp-PC1 Temp_PC2 Temp-PC3

PCA is known as a variable reduction procedure and is useful when variables are significantly correlated. In each group, the variables describe the same physical mechanisms. The numbers of some group variables were shrunk to a reduced number of principal components. Although the details may be different among variables, their overall trends are the same based on their values. Therefore, using PCA to identify a reduced number of variables in the same group is a natural step. In this case, AOT, MCO, CCNO variables reduced to Aero-PC1 and Aero-PC2 for the Aerosol group and presented in the combination as Predictor\_2. Similarly, for the Wind group, V20C U200 U20C TWAC TWXC are reduced to Wind-PC1 Wind-PC2 Wind-PC3 and presented as a combination of Predictor\_3.

There will be some loss of information when a variable reduction was performed, therefore, when this technique was applied we made sure to select group of variables which exhibit similar physical phenomena to minimize loss of information. We have analyzed MODIS aerosol retrievals and SHIPS parameters for 24 hurricanes spanning 7 hurricane seasons. By combining MODIS and SHIPS data, 56 variables were compiled and selected as predictors. Variable reduction from 56 to 31 was performed via correlation coefficients. Among these 31 variables, some are highly correlated or “redundant” with one another. For example, Sea Surface Temperature, Air Temperature and Ocean depth of the (20 and 26 °C) isotherm for the Temperature group are usually very strongly correlated. Therefore, one or two of these variables or the combination of the variables (potentially for a newly defined, more representative variable) could be used as a substitution for all the others. For our study we selected the variable which is most likely to be the direct cause of categorical response and relevant to the hurricane intensity studies and of course highly correlated.

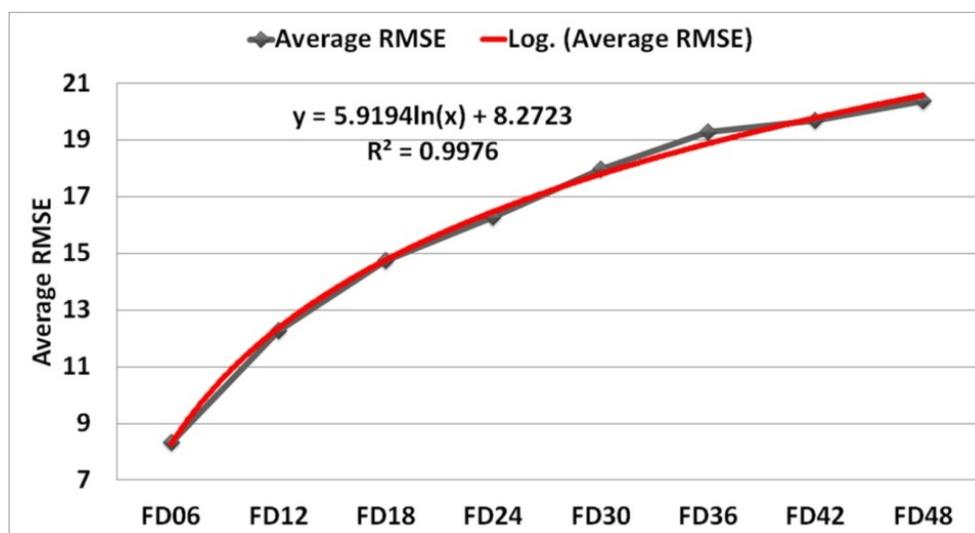
Identification and comparison of the impact of our approach on uncorrelated and correlated variables described in Table 4 by considering, for example, Aerosol and Temperature components. Among the original set of Aerosol variables (AOT AFBO BRBO MRO MCO ERO CCNO) only (AOT MCO CCNO) were highly correlated. We have excluded (AFBO BRBO MRO ERO) variables because they were not correlated as highly as (AOT MCO CCNO). Similarly, the original temperature set of variables was (E000 EPOS EPSS T000 RD26 T150 SST T250 T200 RD20 ENEG ENSS) and only (SST T250 T200 RD20 ENEG ENSS) were highly correlated. Uncorrelated variables (E000 EPOS EPSS T000 RD26 T150) were excluded from this analysis. For comparison we performed PCA on both correlated and uncorrelated variables followed by Multiple Linear Regression (MLR) by the Predictor\_2 and Predictor\_6 at FD48 and presented in the Table 4.

**Table 4.** Multiple Linear Regression (MLR) results for uncorrelated aerosol and temperature variables.

Response	Predictors	R <sup>2</sup>	Adjusted R <sup>2</sup>	F value	RMSE	Residual Error
FD48	Uncorr_Predictor_2	64.80%	60.90%	16.85	20.43	417.30
	Corr_Predictor_2	65.00%	61.10%	17.00	20.37	414.90
	Uncorr_Predictor_6	63.40%	60.80%	24.69	20.45	418.00
	Corr_Predictor_6	63.70%	61.20%	25.02	20.36	415.00

When comparing the results we see for Predictor\_2 and Predictor\_6, R<sup>2</sup>, Adjusted R<sup>2</sup> and F Values had decreased for uncorrelated case while RMSE had creased increased. This illustrated that uncorrelated variables had lost more information than the correlated variables.

**Figure 2.** Average root mean square error (RMSE) for each future difference (FD).



For the Relative Humidity group, RH-PC1 RH-PC2 principal components were extracted from the variables RHLO RHMD R000 and presented as Predictor\_4. Predictor\_5 and Predictor\_6 were presented similarly for the Shear and Temperature groups. For each predictor the combination of variables along with the principal components are shown in Table 5. The PEFC REFC Z850 PENC MSLP PSLV was excluded from PCA because they described diverse physical processes. The PCA technique can only produce outcomes with very limited benefits from such a data set.

The average RMSE for the six Predictors for 06, 12, 18, 24, 30, 36, 42 and 48 h was found to be 8.33, 12.28, 14.76, 16.27, 17.96, 19.28, 19.69 and 20.38 respectively as illustrated in Figure 2. The RED line is the logarithmic fit for the eight data points showing significant R<sup>2</sup> value. The variation among the Predictors RMSE varied between 0.01 through 0.05. This small variation suggests that reducing the number of variables did not change the core physical information. Therefore, the same phenomena can be explained by the reduction of a variable.

**Table 5.** Reduction of variables.

Predictor Set Name	Variable Set	Action Performed
Original Set	MSLP INCV SST DTL PHCN U200 U20C V20C E000 EPOS ENEG EPSS ENSS RHLO RHMD RHHI PSLV Z850 D200 REFC PEFC T000 R000 Z000 TWAC TWXC PENC SHDC SDDC SHGC DIVC T150 T200 T250 SHRD SHTD SHRS SHTS SHRG PENV VMPI VVAV VMFX VVAC IR00 IRM3 RD20 RD26 RHCN AOT AFBO BRBO MRO MCO ERO CCNO	From the 56 variables, less correlated predictors ( $ cc  \leq 0.165$ ) were filtered out to present Predictor_1 (31 variables)
Predictor_1	AOT MCO CCNO PENC V20C MSLP PEFC PSLV U200 U20C Z850 REFC RHLO RHMD R000 SDDC SHDC SHGC SHRD SHRG SHRS SHTD SHTS SST T250 TWAC TWXC T200 RD20 ENEG ENSS	PCA on AOT MCO CCNO to reduce them into Aero-PC1 Aero-PC2 make Predictor_2 (30 variables)
Predictor_2	Aero-PC1 Aero-PC2 PENC V20C MSLP PEFC PSLV U200 U20C Z850 REFC RHLO RHMD R000 SDDC SHDC SHGC SHRD SHRG SHRS SHTD SHTS SST T250 TWAC TWXC T200 RD20 ENEG ENSS	PCA on V20C U200 U20C TWAC TWXC to reduce them into Wind-PC1 Wind-PC2 Wind-PC3 make Predictor_3 (28 variables)
Predictor_3	Aero-PC1 Aero-PC2 Wind-PC1 Wind-PC2 Wind-PC3 PENC MSLP PEFC PSLV Z850 REFC RHLO RHMD R000 SDDC SHDC SHGC SHRD SHRG SHRS SHTD SHTS SST T250 T200 RD20 ENEG ENSS	PCA on RHLO RHMD R000 to reduce them into RH-PC1 RH-PC2 Predictor_4 (27 variables)
Predictor_4	Aero-PC1 Aero-PC2 Wind-PC1 Wind-PC2 Wind-PC3 PENC MSLP PEFC PSLV Z850 REFC RH-PC1 RH-PC2 SDDC SHDC SHGC SHRD SHRG SHRS SHTD SHTS SST T250 T200 RD20 ENEG ENSS	PCA on SDDC SHDC SHGC SHRD SHRG SHRS SHTD SHTS to reduce them into Shear-PC1 Shear-PC2 Shear-PC3 Shear-PC4, Predictor_5 (23 variables)
Predictor_5	Aero-PC1 Aero-PC2 Wind-PC1 Wind-PC2 Wind-PC3 PENC MSLP PEFC PSLV Z850 REFC RH-PC1 RH-PC2 Shear-PC1 Shear-PC2 Shear-PC3 Shear-PC4 SST T250 T200 RD20 ENEG ENSS	SST T250 T200 RD20 ENEG ENSS reduce to Temp-PC1 Temp_PC2 Temp-PC3 make Predictor_6 (20 variables).
Predictor_6	Aero-PC1 Aero-PC2 Wind-PC1 Wind-PC2 Wind-PC3 PENC MSLP PEFC PSLV Z850 REFC RH-PC1 RH-PC2 Shear-PC1 Shear-PC2 Shear-PC3 Shear-PC4 Temp-PC1 Temp_PC2 Temp-PC3	

Each parameter used in this Table 4 is provided in the Appendix based on [33].

#### 4. Results and Discussion

As shown in Table 6 PCA for AOT, MCO and CCNO the cumulative results explain the variability for the first two components as 80.4% and 98.4%. For V20C U200 U20C TWAC TWXC, the top three principal components demonstrate a variability of 52.8%, 82.8% and 98.4%. When PCA was performed on RHLO RHMD R000, we have the cumulative variability for the first two components as 66.67% and 95.20%. PCA for SDDC SHDC SHGC SHRD SHRG SHRS SHTD SHTS gives variability for the four components as 50.7%, 72.0%, 81.4% and 88.8%. For SST T250 T200 RD20 ENEG ENSS, PCA results give variability for the first three components as 58.5%, 76.1% and 91.2%.

Dimensionality reduction infers loss of information; therefore, the goal is to preserve as much information as possible by minimizing difference between the higher (original) and lower dimensional

variables representation. One of the commonly used methods to determine lower dimensional variables is the principal component analysis (PCA), in which the principal components, linear combination of the originals, ranked based on the contribution to the total variance, are chosen as new variables. The first few new variables are responsible for interpreting most of the physical phenomena described by the original variables that have been reduced.

**Table 6.** Principal components for Aerosol, Wind, Relative Humidity, Vertical Shear and Temperature.

Category	Component	Cumulative % Variability
Aerosol	Aero-PC1	80.40%
	Aero-PC2	98.40%
Wind	Wind-PC1	52.80%
	Wind-PC2	82.80%
	Wind-PC3	98.40%
Relative Humidity	RH-PC1	66.67%
	RH-PC2	95.20%
Vertical Shear	Shear-PC1	50.70%
	Shear-PC2	72.00%
	Shear-PC3	81.40%
	Shear-PC4	88.80%
Temperature	Temp-PC1	58.50%
	Temp-PC2	76.10%
	Temp-PC3	91.20%

Two Aerosol principal components explain 98.4% (loss = 1.6%) variability when variable reduction happened from three to two. For Wind, five variables were reduced to three principal components which resulted in a cumulative variability of 98.4% (loss = 1.6%). When PCA was performed on three Relative Humidity variables, it gave us cumulative variability for the two principal components as 95.20% (loss = 4.8%). PCA for eight Shear variables gives variability for the four reduced components as 88.8% (loss = 11.2%). Six Temperature variables were reduced to three components with 91.2% (loss = 8.8%) cumulative variability.

Reducing the variables does not always lead to a better result, but it is expected that the result should be comparable to that with original variables. Reduction of variables removes irrelevant features and dampens noise; it also leads to more comprehensible model because the model involves fewer variables [36].

For the aerosol category the first two components have the proportionality of 0.80 and 0.18 respectively. Most of the weight is on the Aero-PC1 component which is about four times larger than Aero-PC2. For the Wind category, the first component is less than two times the second component and over three times larger than the third component. The proportion for Humidity shows that the first component is twice as large as the second component. Shear has a proportion of about 51% for the first component. The first component of the temperature has about 59% weight. When comparing the first component of the Aerosol, Wind, Relative Humidity, Shear and Temperature we found that aerosol

had the highest proportion followed by Humidity, Temperature, Wind and Shear. Therefore, aerosol might have some influence based on the first component comparison.

In Table 6, the cumulative variability percentage for each extracted component is presented, where the cumulative % threshold was set at 88%.

MLR technique was applied for the model forecast lead time of 06, 12, 18, 24, 30, 36, 42 and 48 h. For each  $FD_{future}$ , six predictor sets (Predictor\_1 through Predictor\_6) variables were analyzed. Table 7 shows the common measures of MLR, and from this table, we can see for FD06,  $R^2$  varied between 15.0% and 18.8% which is about 25.3% variation. For FD12,  $R^2$  varied between 24.4% and 27.8% which is about 22.7% variation. The smallest variation for FD48 is 9.3% between the highest and lowest values.

**Table 7.** Response and Predictor Variables.

Response	Predictors	$R^2$	Adjusted $R^2$	F value	RMSE	Residual Error
<b>FD06</b>	Predictor_1	18.80%	9.60%	2.05	8.32	69.23
	Predictor_2	18.80%	9.90%	2.12	8.31	68.98
	Predictor_3	17.70%	9.40%	2.13	8.33	69.38
	Predictor_4	17.60%	9.60%	2.20	8.32	69.26
	Predictor_5	16.40%	9.60%	2.41	8.32	69.22
	Predictor_6	15.00%	9.10%	2.52	8.35	69.64
<b>FD12</b>	Predictor_1	27.80%	19.60%	3.40	12.27	150.50
	Predictor_2	27.80%	19.90%	3.52	12.25	150.10
	Predictor_3	26.90%	19.50%	3.63	12.28	150.90
	Predictor_4	26.80%	19.70%	3.77	12.27	153.40
	Predictor_5	25.40%	19.30%	4.17	12.30	151.20
	Predictor_6	24.40%	19.10%	4.60	12.31	151.60
<b>FD18</b>	Predictor_1	37.90%	30.80%	5.39	14.76	217.70
	Predictor_2	37.90%	31.10%	5.59	14.73	217.00
	Predictor_3	37.10%	30.80%	5.84	14.77	218.00
	Predictor_4	37.10%	31.00%	6.08	14.74	217.20
	Predictor_5	36.10%	30.80%	6.91	14.76	217.80
	Predictor_6	34.90%	30.30%	7.63	14.81	219.40
<b>FD24</b>	Predictor_1	45.40%	39.30%	7.36	16.27	264.60
	Predictor_2	45.40%	39.40%	7.62	16.24	263.80
	Predictor_3	44.70%	39.10%	7.99	16.29	265.30
	Predictor_4	44.60%	39.20%	8.29	16.27	264.80
	Predictor_5	43.80%	39.30%	9.57	16.27	264.60
	Predictor_6	43.10%	39.10%	10.78	16.29	265.40
<b>FD30</b>	Predictor_1	50.90%	45.40%	9.18	17.92	321.10
	Predictor_2	50.90%	45.50%	9.50	17.90	320.30
	Predictor_3	50.20%	45.10%	9.96	17.97	322.70
	Predictor_4	49.70%	44.90%	10.19	18.01	324.20
	Predictor_5	49.40%	45.20%	11.96	17.94	322.00
	Predictor_6	48.30%	44.70%	13.31	18.04	325.40

Table 7. Cont.

Response	Predictors	R <sup>2</sup>	Adjusted R <sup>2</sup>	F value	RMSE	Residual Error
FD36	Predictor_1	54.90%	49.80%	10.76	19.25	370.40
	Predictor_2	54.60%	49.90%	11.14	19.22	369.50
	Predictor_3	54.10%	49.40%	11.64	19.32	373.40
	Predictor_4	53.70%	49.20%	11.95	19.36	374.80
	Predictor_5	53.40%	49.60%	14.07	19.27	371.70
	Predictor_6	53.00%	49.70%	16.04	19.28	371.60
FD42	Predictor_1	61.80%	57.50%	14.31	19.64	385.80
	Predictor_2	61.70%	57.50%	14.75	19.64	385.80
	Predictor_3	61.20%	57.20%	15.59	19.70	388.10
	Predictor_4	60.70%	56.90%	15.92	19.78	391.10
	Predictor_5	60.60%	57.30%	18.82	19.68	387.30
	Predictor_6	60.20%	57.40%	21.51	19.67	387.10
FD48	Predictor_1	65.10%	61.10%	16.46	20.38	415.30
	Predictor_2	65.00%	61.10%	17.00	20.37	414.90
	Predictor_3	64.80%	61.20%	18.20	20.35	414.00
	Predictor_4	64.30%	60.80%	18.55	20.45	418.20
	Predictor_5	64.00%	61.10%	21.84	20.38	415.30
	Predictor_6	63.70%	61.20%	25.02	20.36	415.00

Let us select FD48 as the response and explanatory variables as Original set of 56, Predictor\_1 as 31 and Predictor\_6 as 20. For the Original variables, 55 degrees of freedom (DF) provide us with RMSE = 19.47, R<sup>2</sup> = 71.1%, R<sup>2</sup> (adj) = 64.5%, F = 10.72 and P = 0.000. For Predictor\_1, DF = 30, RMSE = 20.38 R<sup>2</sup> = 65.1%, R<sup>2</sup> (adj) = 61.1%, F = 16.46 and P = 0.000. For Predictor\_6, DF = 19 RMSE = 20.36 R<sup>2</sup> = 63.7% R<sup>2</sup> (adj) = 61.2% F = 25.02 P = 0.000. One interesting finding is that the adjusted R<sup>2</sup> with 20 variables is larger (or equal to) the corresponding value with 31 variables. At least in this special case, reducing the number of variables does not reduce the effectiveness of the MLR model but increases the efficiency.

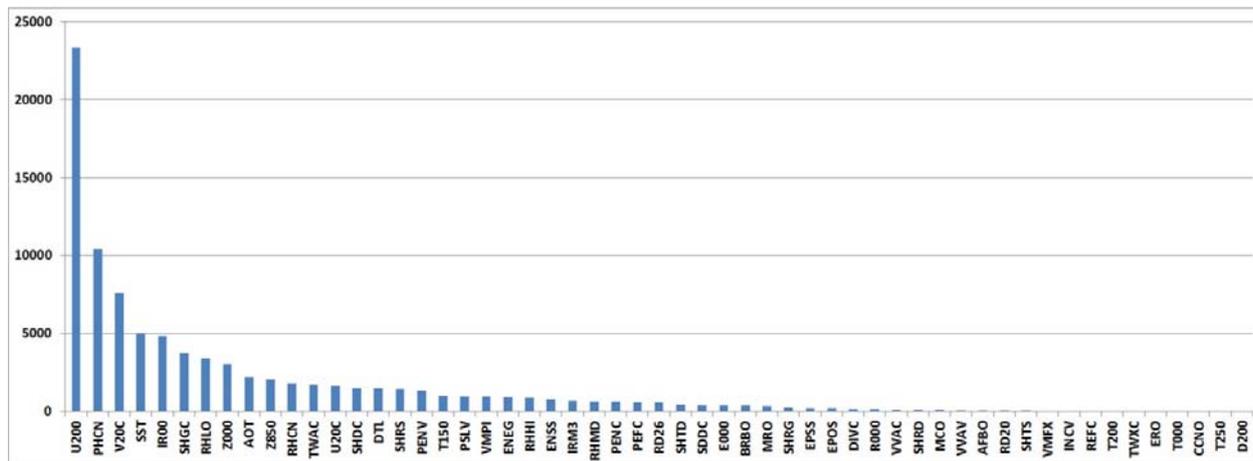
Figures 3–5 illustrates the contribution factor based on the MLR performed between FD48 and the 55 original variables (MSLP was taken out from the analysis because its contributing factor was high), Predictor\_1 of 31 variables and Predictor\_6 of 20 variables respectively. We found aerosol, wind, humidity, shear and temperature all contributing factors in the regression equation. Based on Figure 5, the Predictor\_6 plot, the ranking for the contribution was found as (1) Wind, (2) Aerosols, (3) Shear, (4) Relative Humidity, and (5) Temperature components. Further breakdown, as in Figure 4, showed that U200 and PHCN has the highest contribution then V20C followed by SST.

Figure 3 tells us about the effect of the SHIPS and MODIS variables used on the FD48. R<sup>2</sup> = 71.1% indicating that about 71% of the variation in FD48 can be accounted for by the 56 predictors. Based on the results of the Sequential Sum of Squares we can see components such as zonal winds, estimated ocean heat content and sea surface components are the greatest contributors to the MLR. This is also true for the intensity changes at 06, 12, 18, 24, 30, 36 and 42 h.

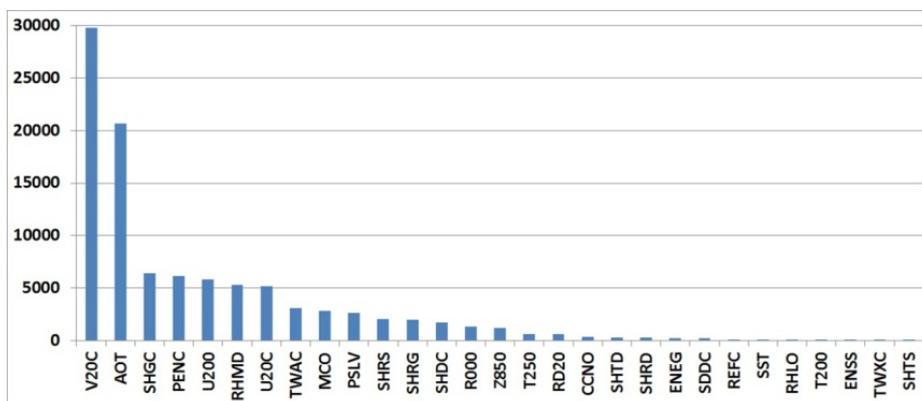
In Figure 4, R<sup>2</sup> = 65.1% indicating that about 65% of the variation in FD48 can be accounted for by the 31 explanatory variables. The contribution factor in this case is mainly governed by the same

variables as shown in Figure 4 except we see Aerosol Optical Thickness (AOT) and relative humidity are playing significant roles as well. We also see Shear and Eddy play important roles in the case of Predictor\_1.

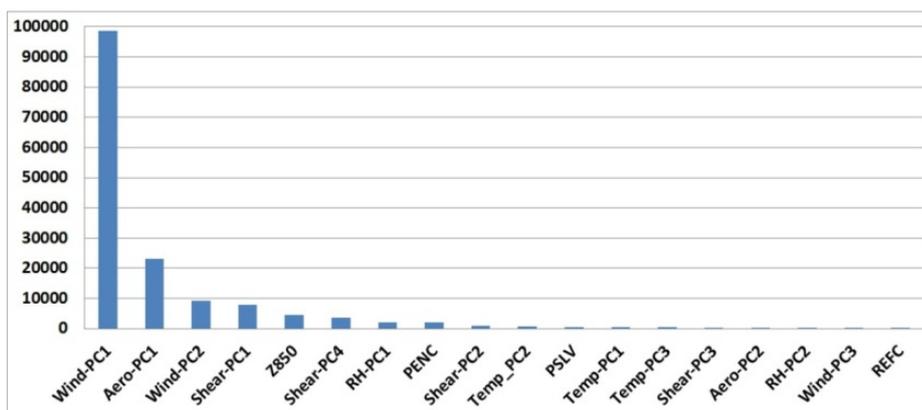
**Figure 3.** Contribution Factors on the MLR between the FD48 and Original set of 55 variables. MSLP was removed from the analysis.



**Figure 4.** Contribution Factors on the MLR between the FD48 and Predictor\_1 which has 31 variables.



**Figure 5.** Contribution Factors on the MLR between the FD48 and Predictor\_6 which has 20 variables.



The effect of the SHIPS and MODIS variables used on the FD48 as illustrated in Figure 5,  $R^2 = 63.7\%$  indicating that about 64% of the variation in FD48 can be accounted for by the 20 explanatory variables. The contribution factor in this case is governed by tangential and zonal wind in addition to AOT and RH.

Figure 6 shows the  $R^2$  and adjusted  $R^2$  values along with the RMSE and the Residual Errors for the MLR performed between the eight response variables and six predictor sets.

At 48 h forecast intervals as in Figure 6,  $R^2$ , adjusted  $R^2$  and RMSE are the largest and at 06 h, the smallest was recorded. The range of values of  $R^2$ , adjusted  $R^2$  and RMSE between 06 and 48 h for Predictor\_6 were found to be (15.0% and 63.7%), (9.1% and 61.2%), (8.35 and 25.02) and (69.64 and 415.0) respectively. The RMSE and Residual errors found negligible for all six predictors. However, significant  $R^2$  values were found to be larger when considering the 42 and 48 h lead time for longer forecast intervals. This may be due to the results of discretization of the intensity of values as per DeMaria *et al.* [22,23] and the regressions for the shorter forecast intervals may have been exposed to some noise [22,23].

In Figure 7 “Residuals vs. fits” are presented to show the residuals vs. the fitted values at FD06 and FD48. Residuals varied between  $\pm 10$  for FD06 whereas for FD48 its  $\pm 50$  and FD06 has lesser outlier than FD48.

In addition, for this study MODIS Aerosol Retrievals were averaged, therefore, it is important to articulate the statistical uncertainty for the three variables used in the Aerosol PCA. For example, the quoted uncertainties for Fabian 2003 found for AOT ( $0.23 \pm 0.02$ ), MCO ( $15.54 \pm 1.89$ ) and CCNO ( $3.98 \pm 0.781$ )  $\times 10^8$  when 95% confidence interval was considered.

**Figure 6.**  $R^2$  values and RMSE at eight lead time positions 06, 12, 18, 24, 30, 36, 42 and 48 h.

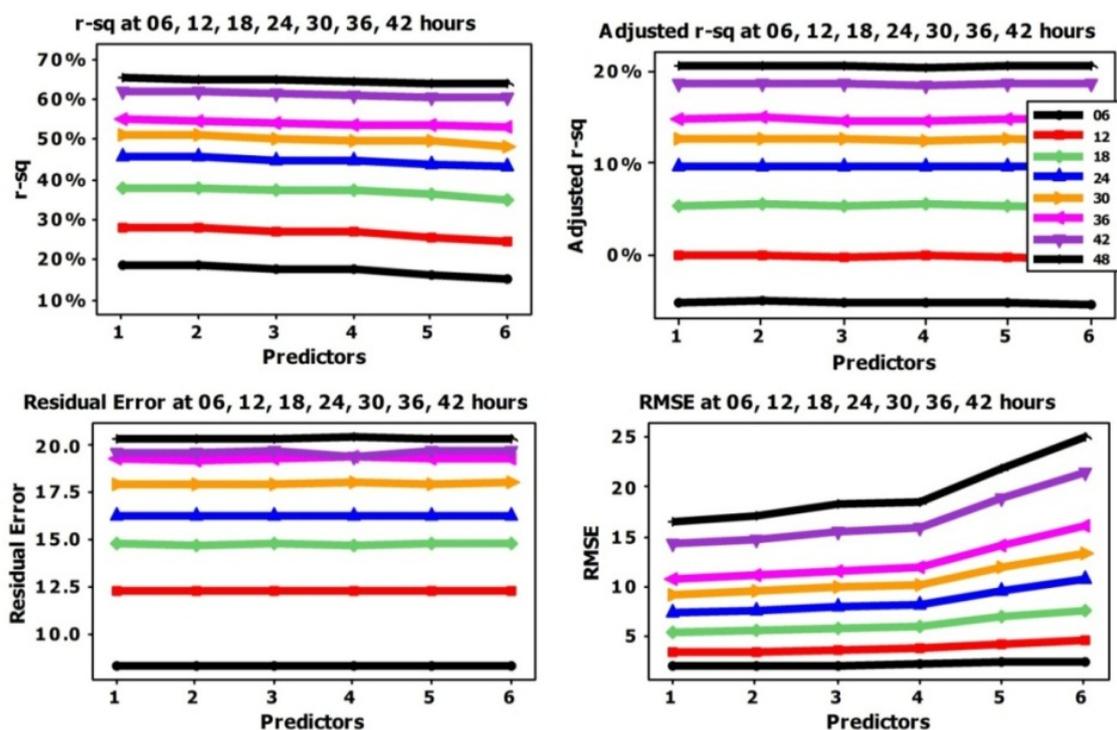
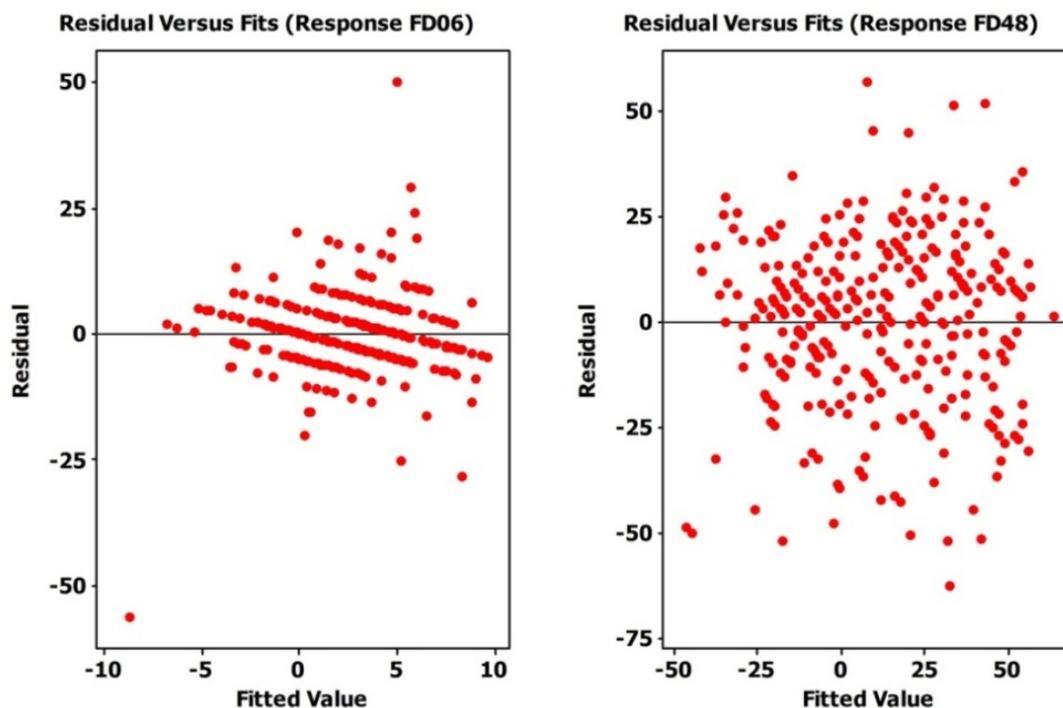


Figure 7. Residual plots for FD06 and FD48.



## 5. Conclusions

By combining MODIS and SHIPS data, 56 variables were compiled and selected as predictors for this study. Variable reduction from 56 to 31 was performed via correlation coefficients (cc) followed by Principal Component Analysis (PCA) extraction techniques to further reduce these 31 variables to 20. Among the 31 variables, PCA candidates were selected for the variables describing the same physical mechanism and the PCA procedure reduces the numbers from 3–8 to 1–4 for each group of variables. Five categories: wind, aerosols, shear, relative humidity, and temperature components were established by reducing 56 variables to 20. Aerosol, wind, humidity, shear and temperature are all contributing factors in the regression equation with the ranking for the contribution found to be (1) Wind, (2) Aerosols, (3) Shear, (4) Relative Humidity, and (5) Temperature components. Indicating that aerosols predictor surpass the other predictors especially shear. However, from a dynamics point of view, it is impossible for aerosol to be more important than shear and temperature. The aerosol rank preceded the shear, which could be because our sample size was too small (306 data points) when compared to the original SHIPS dataset (over 6,000 data points) and inadvertently the value ranges of shear and temperature are not large. As a result, the limited variance in those parameters makes it difficult to demonstrate the importance of those parameters. This is practically similar to a study with other parameter values being controlled. When the coefficient of variations (cv) was calculated we found cv for AOT 40.29%, Wind 37.61%, Shear 35.50%, SST 3.65% and Relative Humidity (RH) 6.8%. SST and RH cv values are so low that we can consider the experiment to be controlled at a specific value. In the same sense, it is not surprising to that AOT was the second dominated factor in this study because AOT are of the largest variability. When MLR is performed on all 56 variables (without any variable reduction) as illustrated in Figure 3, interestingly, we see that aerosol is ranked in the last place. The original parameter describing aerosol effects are not a good choice. The linear

combination of the original variables gives a much better description because of the much higher variance in the derived variable. As a result, although the AOT role is not among the first few parameters in the MLR model with all variables, the combined aerosol parameter plays a dominant role in the limited model.

There are plenty of benefits for overcoming the curse of dimensionality. Original variables may demonstrate better results but the reduced variables gave similar results with much lower dimensionality and improved efficiency. For computational purposes, improved efficiency is much more important than highly precise results.

One interesting finding is that the adjusted  $R^2$  with Predictor\_6, 20 variables is larger than (or equal to) the corresponding value with Predictor\_1 of 31 variables. At least in this special case, reducing the number of variables does not reduce the effectiveness of the MLR model but increases the efficiency.

The variation among the Predictors RMSE varied between 0.01 through 0.05. This implies that reducing the number of variables did not change the core physical information because variation is from the mean for all sets of predictors and very small. Therefore, the same phenomena can be explained by the reduction of the variable.  $R^2$  values were found to be larger when considering the 42 and 48 h lead time.  $R^2$ , adjusted  $R^2$ , RMSE and residual error among Predictor 1 through 6 was negligible. The RMSE and residual errors difference among the six predictor groups were found to be negligible.

## Acknowledgments

We acknowledge the MODIS mission scientists and associated NASA personnel for the production of the data used in this research effort. The HDF Group, <http://www.hdfgroup.org>, for posting code for numerous commercial and non-commercial platforms of powerful data visualization and analysis on HDF files. The authors are grateful to Mark DeMaria for the SHIPS data. The authors would also like to thank the reviewers for their valuable and constructive comments for improving the manuscript.

## References

1. NOAA. *Hurricane Hunters*. Available online: <http://flightscience.noaa.gov/> (accessed on 28 June 2012).
2. Elsberry, R.L.; Jeffries, R.A. Vertical wind shear influence on tropical cyclone formation and intensification during TCM-92 and TCM-93. *Mon. Weather Rev.* **1996**, *124*, 1374–1387
3. Black, M.L.; Gamache, J.F.; Marks, F.D.; Samsury, C.E.; Willoughby, H.E. Eastern Pacific Hurricanes Jimena of 1991 and Olivia of 1994: The effects of vertical shear on structure and intensity. *Mon. Weather Rev.* **2002**, *130*, 2291–2312.
4. DeMaria, M.M. The effect of vertical shear on tropical cyclone intensity change. *J. Atmos. Sci.* **1996**, *53*, 2076–2087.
5. Rogers, R.; Chen, S.S.; Tenerelli, J.E.; Willoughby, H.E. A numerical study of the impact of vertical shear on the distribution of rainfall in Hurricane Bonnie (1998). *Mon. Weather Rev.* **2003**, *131*, 1577–1599.
6. Kamal, M.M.; Qu, J; Hao, X. A study of dust aerosols impact on hurricanes with multi-sensors measurement from space. *Open Remote Sens. J.* **2012**, *5*, 73–82.

7. Rosenfeld, D.; Woodley, W.L.; Khain, A.; Cotton, W.R.; Carrió, G.; Ginis, I.; Golden, J.H. Aerosol effects on microstructure and intensity of tropical cyclones. *Bull. Amer. Meteorol. Soc.* **2012**, *93*, 987–1001.
8. DeMaria, M.; Kaplan, J. Sea surface temperature and the maximum intensity of the Atlantic tropical cyclones. *J. Climate* **1994**, *7*, 1324–1334.
9. Evans, J.L. Sensitivity of tropical cyclone intensity to sea surface temperature. *J. Climate* **1993**, *6*, 1133–1140.
10. Dunion, J.P.; Velden, C.S. The impact of the Sahara air layer on Atlantic tropical cyclone activity. *Bull. Amer. Meteorol. Soc.* **2004**, *85*, 353–365.
11. Holland, G.J. The maximum potential intensity of tropical cyclones. *J. Atmos. Sci.* **1997**, *54*, 2519–2541.
12. Emanuel, K.A. The maximum intensity of hurricanes. *J. Atmos. Sci.* **1988**, *45*, 1143–1155.
13. Houze, R.A., Jr.; Cetrone, J.; Brodzik, S.R.; Chen, S.S.; Zhao, W.; Lee, W.; Moore, J.A. Stossmeister, G.J.; Bell, M.M.; Rogers, R.F. The hurricane rainband and intensity change experiment: Observations and modeling of Hurricanes Katrina, Ophelia, and Rita. *Bull. Amer. Meteorol. Soc.* **2006**, *87*, 1503–1521.
14. Krishnamurti, T.N.; Han, W.N.; Jha, B.; Bedi, H.S. Numerical prediction of Hurricane Opal. *Mon. Weather Rev.* **1998**, *126*, 1347–1363.
15. Cangialosi, J.P.; Franklin, J.L. *2011 National Hurricane Center Forecast Verification Report*; National Hurricane Center, NOAA: Miami, FL, USA, 2011. Available online: [http://www.nhc.noaa.gov/verification/pdfs/Verification\\_2011.pdf](http://www.nhc.noaa.gov/verification/pdfs/Verification_2011.pdf) (accessed on 30 July 2012).
16. Rosenfeld, D.; Clavner, M.; Nirel, R. Pollution and dust aerosols modulating tropical cyclones intensities. *Atmos. Res.* **2011**, *102*, 66–76.
17. Zhang, H.; McFarquhar, G.M.; Cotton, W.R.; Deng, Y. Direct and indirect impacts of Saharan dust acting as cloud condensation nuclei on tropical cyclone eye wall development. *Geophys. Res. Lett.* **2009**, doi:10.1029/2009GL037276.
18. Zipser, E.J.; Twohy, C.H.; Tsay, S.C.; Thornhill, K.L.; Tanelli, S.; Ross, R.; Krishnamurti, T.N.; Ji, Q.; Jenkins, G.; Ismail, S.; *et al.* The saharan air layer and the fate of African easterly waves—NASA’s AMMA field study of tropical cyclogenesis. *Bull. Amer. Meteorol. Soc.* **2009**, *90*, 1137–1156.
19. Gao, J.; Zha, Y. Meteorological influence on predicting air pollution from modis-derived aerosol optical thickness: A case study in Nanjing, China. *Remote Sens.* **2010**, *2*, 2136–2147.
20. Braun, S.A. Reevaluating the role of the Saharan Air Layer in Atlantic tropical cyclogenesis and evolution. *Mon. Weather Rev.* **2010**, *138*, 2007–2037.
21. Khain, A.; Lynn, B.; Dudhia, J. Aerosol effects on intensity of land falling hurricanes as seen from simulations with the wrf model with spectral bin microphysics. *J. Atmos. Sci.* **2010**, *67*, 365–384.
22. DeMaria, M.; Kaplan, J. A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. *Weather Forecast.* **1994**, *9*, 209–220.
23. DeMaria, M.; Kaplan, J. An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins mark. *Weather Forecast.* **1999**, *14*, 326–337.
24. DeMaria, M.; Mainelli, L.K.; Shay, J.A.; Knaff, K.J. Further improvements in the statistical hurricane intensity prediction scheme (SHIPS). *Weather Forecast.* **2005**, *20*, 531–543.

25. NASA. *MODIS Atmosphere*. Available online: [http://modis-atmos.gsfc.nasa.gov/MOD04\\_L2/index.html](http://modis-atmos.gsfc.nasa.gov/MOD04_L2/index.html) (accessed on 26 June 2012).
26. Madhavan, S.; Qu, J.; Xiong, J. Comparison Study between MODIS Terra and Aqua for AOT Retrieval over Ocean. In *Proceedings of 2008 IEEE International Geoscience and Remote Sensing Symposium*, Boston, MA, USA, 6–11 June 2008; Volume 3, pp. 515–518.
27. NASA. *MODIS Website*. Available online: <http://modis.gsfc.nasa.gov/> (accessed on 30 July 2012).
28. Tanré, D.; Kaufman, Y.J.; Herman, F.; Mattoo, S. Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. *J. Geophys. Res.* **1997**, *102*, 971–988.
29. Kaufman, Y.J.; Tanré, D.; Remer, L.A.; Vermote, E.F.; Chu, A.; Holben, B.N. Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *J. Geophys. Res.* **1997**, *102*, 17051–17067.
30. King, M.D.; Menzel, W.P.; Kaufman, Y.J.; Tanré, D.; Gao, B.C.; Platnick, S.; Ackerman, S.A.; Remer, L.A.; Pincus, R.; Hubanks, P.A. Cloud and aerosol properties, precipitable water, and profiles of temperature and humidity from MODIS. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 442–458.
31. Remer, L.A.; Kaufman, Y.J.; Mattoo, S.; Martins, J.V.; Ichoku, C.; Levy, R.C.; Kleidman, R.G.; Tanré, D.; Chu, D.A.; Li, R.-R; *et al.* The MODIS aerosol algorithm, products, and validation. *J. Atmos. Sci.* **2005**, *62*, 947–973.
32. NASA Goddard Space Flight Center. *LAADS Web*. Available online: <http://ladsweb.nascom.nasa.gov> (accessed on 26 June 2012).
33. DeMaria, M. *SHIPS Data Repository*. Available online: <ftp://rammftp.cira.colostate.edu/demaria/SHIPS/> (accessed on 26 June 2012).
34. NOAA. National Weather Service. *National Hurricane Center*. Available online: <http://www.nhc.noaa.gov/> (accessed on 26 June 2012).
35. National Hurricane Center. *Hurricane Basics*. Available online: <http://hurricanes.noaa.gov/pdf/hurricanebook.pdf> (accessed on 2 September 2012).
36. Tan, P-N.; Steinbach, M.; Kumar, V. Data. In *Introduction to Data Mining*; Chapter 2; Addison-Wesley: Boston, MA, USA, 2006; pp. 50–52.

## Appendix

Selected SHIPS parameters based on the website at [33].

Name	Description
SST	Climatological SST (deg C $\times$ 10) vs. time
RHLO	850–700 mb relative humidity (%) vs. time (200–800 km)
RHMD	700–500 mb relative humidity (%) vs. time (200–800 km)
RHHI	500–300 mb relative humidity (%) vs. time (200–800 km)
SHRS	850–500 mb shear magnitude (kt $\times$ 10) vs. time
VMAX	The current maximum wind intensity in kt
MSLP	Mean sea level pressure (hPa)
INCV	Intensity change (kt) –18 to –12, –12 to –6, ... 114 to 120 hr.
SST	SST (deg C $\times$ 10) vs. time
DTL	Distance to nearest major land mass (km) vs. time

Cont.

Name	Description
PHCN	Estimated ocean heat content ( $\text{kJ}/\text{cm}^2$ ) from climo OHC and current SST anomaly. Designed to fill in for RHCN when that is missing.
U200	200 mb zonal wind ( $\text{kt} \times 10$ ) vs. time ( $r = 200\text{--}800$ km)
U20C	Same as U200 but for $r = 0\text{--}500$ km)
V20C	Same as U20C, but for the v component of the wind
E000	1,000 mb theta_e ( $r = 200\text{--}800$ km) vs. time ( $\text{deg K} \times 10$ )
EPOS	The average theta_e difference between a parcel lifted from the surface and its environment (200–800 km average) vs. time ( $\text{deg C} \times 10$ ). Only positive differences are included in the average
ENEG	Same as EPOS, but only negative differences are included. The minus sign is not included.
EPSS	Same as EPOS, but the parcel theta_e is compared with the saturated theta_e of the environment
ENSS	Same as ENEG, but the parcel theta_e is compared with the saturated theta_e of the environment
PSLV	Pressure of the center of mass (mb) of the layer where storm motion best matches environmental flow ( $t = 0$ only)
Z850	850 mb vorticity ( $\text{sec}^{-1} \times 10^7$ ) vs. time ( $r = 0\text{--}1,000$ km)
D200	Same as above for 200 mb divergence
REFC	Relative eddy momentum flux convergence ( $\text{m}/\text{sec}/\text{day}$ , 100–600 km avg) vs. time
PEFC	Planetary eddy momentum flux convergence ( $\text{m}/\text{sec}/\text{day}$ , 100–600 km avg) vs. time
T000	1,000 mb temperature ( $\text{deg C} \times 10$ ) (200–800 km average)
R000	1,000 mb relative humidity (200–800 km average)
Z000	1,000 mb height deviation (m) from the US standard atmosphere
TWAC	0–600 km average symmetric tangential wind at 850 mb from NCEP analysis ( $\text{m}/\text{sec} \times 10$ )
TWXC	Maximum 850 mb symmetric tangential wind at 850 mb from NCEP analysis ( $\text{m}/\text{sec} \times 10$ )
PENC	Azimuthally averaged surface pressure at outer edge of vortex ( $(\text{mb} - 1,000) \times 10$ )
SHDC	Same as SHRD but with vortex removed and averaged from 0–500 km relative to 850 mb vortex center
SDDC	Heading (deg) of above shear vector
SHGC	Same as SHRG but with vortex removed and averaged from 0–500 km relative to 850 mb vortex center
DIVC	Same as D200, but centered at 850 mb vortex location
T150	200 to 800 km area average 150 mb temperature ( $\text{deg C} \times 10$ ) vs. time
T200	Same as above for 200 mb temperature ( $\text{deg C} \times 10$ )
T250	Same as above for 250 mb temperature ( $\text{deg C} \times 10$ )
SHRD	850–200 mb shear magnitude ( $\text{kt} \times 10$ ) vs. time (200–800 km)
SHTD	Heading (deg) of above shear vector
SHTS	Heading of above shear vector
SHRG	Generalized 850–200 mb shear magnitude ( $\text{kt} \times 10$ ) vs. time (takes into account all levels)
PENV	200 to 800 km average surface pressure ( $(\text{mb} - 1,000) \times 10$ )
VMPI	Maximum potential intensity from Kerry Emanuel equation (kt)
VVAV	Average (0 to 15 km) vertical velocity ( $\text{m}/\text{s} \times 100$ ) of a parcel lifted from the surface where entrainment, the ice phase and the condensate weight are accounted for. Note: Moisture and temperature biases between the operational and reanalysis files make this variable inconsistent in the 2001–2007 sample, compared 2,000 and before.
VMFX	Same as VVAV, but a density weighted vertical average.

Cont.

Name	Description
VVAC	Same as VVAV but with soundings from 0–500 km with GFS vortex removed
IRXX	Same as IR00 below, but generated from other predictors (not satellite data). These should only be used to fill in for IR00 as needed. Predictors from GOES data (not time dependent). The 17 values in this record are as follows: (1) Time (hr × 10) of the GOES image, relative to this case (2) Average GOES ch 4 brightness temp (deg C × 10), r = 0–200 km (3) Stan. Dev. of GOES BT (deg C × 10), r = 0–200 km (4) Same as (2) for r = 100–300 km (5) Same as (3) for r = 100–300 km (6) Percent area r = 50–200 km of GOES ch 4 BT < −10 C (7) Same as (6) for BT < −20 C (8) Same as (6) for BT < −30 C (9) Same as (6) for BT < −40 C (10) Same as (6) for BT < −50 C (11) Same as (6) for BT < −60 C (12) max BT from 0 to 30 km radius (deg C × 10) (13) avg BT from 0 to 30 km radius (deg C × 10) (14) radius of max BT (km) (15) min BT from 20 to 120 km radius (deg C × 10) (16) avg BT from 20 to 120 km radius (deg C × 10) (17) radius of min BT (km)
IR00	
IRM3	Same as IR00 but at three hours before initial time
RD20	Ocean depth of the 20 deg C isotherm (m), from satellite altimetry data
RD26	Ocean depth of the 26 deg C isotherm (m) from satellite altimetry data
RHCN	Ocean heat content (kJ/cm <sup>2</sup> ) from satellite altimetry data

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).