A NEURAL NETWORK APPROACH TO PREDICT HURRICANE INTENSITY IN THE NORTH ATLANTIC BASIN

by

Anthony Veneros Castro

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

INDUSTRIAL ENGINEERING

UNIVERSITY OF PUERTO RICO

MAYAGUEZ CAMPUS

2004

Approved by:

____________________ ____________
William Hernández, Ph. D. Date
Member, Graduate Committee

____________________ ____________
Jorge E. Gonzáles, Ph. D. Date
Member, Graduate Committee

____________________ ____________
Ramon E. Vasquez Espinosa, Ph. D. Date
Member, Graduate Committee

____________________ ____________
Nazario Ramírez Beltrán, Ph. D. Date
President, Graduate Committee

____________________ ____________
Mario Ierkic-Vidmar, Ph.D. Date
Representative of Graduate Studies

____________________ ____________
Agustin Rullán Toro, Ph. D. Date
Chairperson of the Department

ABSTRACT

Upper air information and artificial neural networks (ANN) are used to predict hurricane intensity in the North Atlantic basin. Competitive neural network is used to identify analog storms to the current hurricane. Once the analog hurricanes are identified the historical NCEP reanalysis data are used along of each storm tracks to develop a set of climatology, persistence and synoptic variables. Persistence, climatological and synoptic observations of the analog hurricanes and the current storm are combined to create a training set which is used to generate nonlinear transformations and an optimization algorithm is used to identify the variables that are best correlated with storm intensity. The best variables obtained from the optimization algorithm are used to train a neural network which used Levenberg-Marquardt algorithm as a learning rule. Preliminary results show that the proposed prediction scheme is a potential tool to increase the accuracy in predicting hurricane intensity.

RESUMEN

Redes neuronales artificiales e información atmosférica son utilizadas para predecir la intensidad de los huracanes en la parte norte del Océano Atlántico. Un proceso para identificar huracanes históricos que sean análogos al huracán actual es implementado usando una red neuronal competitiva. Una vez identificado los huracanes análogos, información histórica proveniente de NCEP es usada para crear una serie de variables sinópticas, climatologicas y persistentes a lo largo de la trayectoria de cada uno de los huracanes análogos. Estas variables son combinadas con las variables del huracán actual para crear un set de entrenamiento. Un algoritmo de optimización es implementado para identificar aquellas variables que tengan la mayor correlación con la intensidad. Estas luego son usadas para implementar una red neuronal que usa el algoritmo de Levenberg-Marquardt como regla de aprendizaje. Los resultados preliminares muestran que la metodología propuesta es una herramienta potencial en los esfuerzos por aumentar la precisión en la predicción de la intensidad de los huracanes.

DEDICATORY

This thesis is dedicated to my parents Nelly and Aparicio, to my sisters Candy and Ketty, to my grandparents and to my uncle Segundo .A special dedication to my beloved wife Heidi.

ACKNOWLEDGEMENTS

I would like to sincerely express my gratitude to my advisor, Dr. Nazario Ramirez. His direction, interest, and support were essential in the development of this thesis. I like to express my thanks to Dr José González to let me work in his project which provides the financial support to obtain the master degree. Moreover, I am indebted to the University of Puerto Rico and the Industrial Engineering Department to let me achieve the goal of being a graduate student.

I give special thanks to my family. My dear father Aparicio who shows me with his actions how a good person must behave and to motivate me to continue my graduate studies. My adored mother Nelly for being the best mother ever and teaching me how to enjoy the life. To my sisters Ketty and Candy who have always taken care of me. My grandparents Ladislao and Arancidia who are like parents to me and raise me when I was a child. Also, My deceased grandmother Filomena for whom I was his favorite granddaughter and was like a mother to me. My uncle Segundo for teaching me that you must never stop learning and bringing always on me the affection for my Peru. Thanks to my new family Romero Encarnación to understand me and to give me always its support. My wife Heidi for her true love and for making me a better person everyday.

Finally, Thanks to God, our lord. For being all the time with me and for giving me the health, love and the strength necessary to achieve my goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIXES

CHAPTER I INTRODUCTION

1.1. Justification

A tropical cyclone is the generic term for a non-frontal synoptic scale low-pressure system over tropical or sub-tropical waters with organized convection and definite cyclonic surface wind circulation. It is known to form over all tropical oceans with the exception of the south Atlantic and the south Pacific east of about 140° W. Tropical cyclone in matured condition is known as hurricane in the Atlantic Ocean.

The hurricane intensity is a measure of the destructive effects over a particular place on humans and (or) structures. It is measured by the maximum 1-min sustained surface wind in the walls of the hurricane (DeMaria et al., 1994). Usually, in hurricane landfall areas the resultant damages are often extensive, especially in developed coastal areas. The principal damaging forces associated with tropical cyclones are the storm surge, floods caused by torrential rains and heavy destructive winds. Actually, the trends of human losses from hurricanes in the United States and the Caribbean have gradually decreased as a result of improved prediction techniques and warning strategies. However, property losses continue to rise because of the progressive development in vulnerable areas.

Among the reasons to conduct this investigation, the most important is to provide an operational intensity model for Puerto Rico, ready to be used when a hurricane is detected over the Atlantic Ocean. It is well known that Puerto Rico, due to its geographic location, is under continuous threats of hurricane landfalls. Hurricane George (1998) was the last one that made landfall in Puerto Rico, with sustained surface winds of 100 knots and was classified under category III. It caused considerable damage to infrastructures throughout the island. Table 1.1 shows the humans deaths and economical impact caused by Hurricane George. Therefore, to mitigate hurricane effects, it is necessary to predict the intensity of a hurricane with a high level of accuracy. The importance of an accurate hurricane intensity forecast is recognized by Gray (1997), who states that the costal population of the Unites States has grown exponentially, especially in Florida, exposing millions more people to the threat of hurricane landfalling. The required time to evacuate some highly vulnerable coastal areas is in the order of 36-48 hours, while the current time frame for meaningful hurricane intensity change guidance is in the order of 12-36 hours. These facts indicate that effective warnings for certain vulnerable and populated areas are difficult to make with the existent intensity forecast algorithms.

Another reason for this study is the inclusion of new tools especially; satellite's observations in the hurricane intensity prediction field that have given researchers a better understanding of hurricane behavior and improved intensity prediction models by including some parameters that can be used to explain better the hurricane intensity. Atmospheric Microwave Sounding Unit (AMSU) is a sensor located in satellites NOAA 15-17. This sensor has the ability to observe the inside of a hurricane and take direct measurements. AMSU data is used in this study and it is expected that the hurricane intensity model will improve its predictions.

Table 1.1 Estimates of deaths and economical damage associated to Hurricane George (Courtesy of the American Insurance Services)

LOCATION	Deaths	($ Billions)
Antigua	2	
St. Kitts and Nevis	4	0.402
U.S. Virgin Islands	0	0.050
Puerto Rico	0	1.750
Dominican Republic	380	>1.0
Haiti	209	
Bahamas	1	
Cuba	6	
United States (Mainland)		
Florida	0	0.340
Mississippi	0	0.665
Alabama	1	0.125
Louisiana	0	0.025
United States Total	**1**	**2.955**
Storm Total	602	

Gutowski et al. (1994) argued that there are three factors that modify a hurricane's intensity: sea surface temperature, atmospheric relative humidity, and the temperature difference between the surface and the lower stratosphere. From global climate models, it is known that doubling of the concentration of atmospheric carbon dioxide will likely increase sea surface temperature around the globe. DeMaria and Kaplan (1993) developed an empirical relationship between sea surface temperature and the maximum intensity of tropical cyclones in the North Atlantic basin using a 31-year sample (1962-1992).

They concluded that maximum possible intensity (MPI) of a given hurricane can be expressed as follows:

$$V = A + B * e^{C(T - T_o)}$$

Where:

V= maximum wind (m s^{-1})

T= sea surface temperature (^oC)

T_o= Reference temperature, generally 30 ^oC

A, B, C= constants

Artificial neural networks (ANN) are used in this research because ANN is a nonlinear modeling tool that can properly represent the nonlinear dynamic system inherent in the development of a hurricane intensity process. In this case, sufficient information is used and gathered from the National Center for Environmental Prediction (NCEP), Advanced Microwave Sounding Unit (AMSU), and the National Hurricane Center and Tropical Prediction Center (NHC/TPC) among others.

1.2. Objectives

The main objective of this investigation is to develop and implement an algorithm based on climatology, persistence and synoptic observations as well as neural networks algorithms to predict the hurricane intensity change at certain intervals of time (6,12,18 and 24 hours). It is expected to accomplish the following specific objectives:

- Develop an interactive database that includes hurricane historical records, upper-air information, and satellite data.
- Assess the usefulness of AMSU information to predict the hurricane intensity changes in the Atlantic basin.
- Develop a random variable selection scheme to identify the variables that explain best the relationships among the inputs and output of a nonlinear dynamic system.
- Test the capability of the neural networks as a tool to model highly nonlinear processes such as the hurricane intensity process.
- Compare the prediction capabilities of ANN with regression techniques.

1.3. Scope

A statistical model is developed to estimate the hurricane intensity in the North Atlantic Ocean. The intervals of prediction are 6, 12, 18 and 24 hours. This model uses 27 years of historical information from 1975 to 2002 stored in a database. Three types of information are used in this study: climatology, persistence and synoptic information where each one of these types are processed to obtain a unique format. The designed prediction model includes 20 meteorological variables, which are lagged and mathematically transformed to be correlated with the hurricane intensity.

1.4. Report Organization

The present study is organized in five chapters: Chapter I shows a brief justification, the main objectives and the scope of this investigation.

Chapter II introduces the literature review that supports this work. This chapter is addressed in two directions. First, the theory about hurricane intensity and intensity prediction models that are actually in operation are briefly described. The advances achieved in this field are also explored and analyzed so that they may be incorporated into this study. The inclusion of satellite observations in the hurricane intensity field is also discussed in this chapter. Second, artificial neural network is introduced and some applications of this technique in the hurricane intensity and other meteorological events are described.

Chapter III presents the entire methodology of this investigation. The proposed model is presented step by step and special attention is given to describe how it should be implemented. This chapter describes how hurricane historical data and upper air information are combined with artificial neural network algorithms to predict the hurricane intensity.

Chapter IV summarizes the results obtained when a set of hurricanes were selected for assessing the prediction algorithm. A representative random sample was selected to implement the prediction algorithm. Experiments performed for different kind of hurricanes are also presented. A complete description of the statistics obtained by the proposed intensity model is finally recapitulated.

Conclusions of this research and recommendations for future investigations are presented in Chapter V.

Finally, appendixes are also included to describe the developed program and the procedure used to select the size of the hurricane sample to test the proposed intensity model. Also, a brief description of the AMSU and its improvements over hurricane intensity are given. Finally, the Saffir-Simpson hurricane scale is explained.

CHAPTER II LITERATURE REVIEW

This chapter is aimed to describe the theories and arguments used in this study. A literature review of the most important contributions for the hurricane intensity field is described. Applications of artificial neural networks for modeling and predicting meteorological events such as hurricane intensity are also described.

The hurricane intensity prediction models are divided into three categories: 1) statistical models whose predictors are based on climatology and persistence; 2) statistical-dynamical models which, in addition of using climatology and persistence, they are also using outputs from numerical models; and 3) dynamic models whose predictions are derived from physical principles and numerical models are typically used to perform the intensive computational work.

Hope et al. (1970) developed the first operational model called HURRAN (HURRicane ANalog). By identifying previous storms that had characteristics in common with a current storm, HURRAN attempted to predict the most likely track of the current storm. Neumann (1972) introduced a model to predict hurricane tracks based on climatology and persistence called CLIPER (CLImatology and PERsistence). This model was a breakthrough in the hurricane field because it was the first operational model that used these kinds of variables. The persistence variables assumed that the integrated effects of all forces which have steered the hurricane during some past period will continue to predominate during some future period. In general, persistence is taken as the smoothed motion of the tropical cyclone in the past 12- or 24-hour period. The persistence forecast is then the linear extrapolation of this motion for the next 12, 24, 36 and 72 hours. The problem with this type of forecast is that a higher order of persistence forecast requires a better knowledge of actual and past conditions. On the other hand, a climatological forecast makes use of the temporal and spatial repetitiveness of tropical cyclone tracks produced by synoptic patterns, the simultaneous observation of pressure, temperature, wind and other meteorological parameters. Using a similar set of variables used by CLIPER, Jarvinen and Neuman (1979) developed an intensity prediction model called Statistical Hurricane Intensity Forecast (SHIFOR) which is used to predict the future intensity of the storm at 12-hour periods up to 72 hours. The predictor variables included: Julian day, initial storm intensity, intensity change during the past 12 hours, initial storm latitude and longitude, and zonal and meridional components of the storm motion vector. Ten predictor terms are included in each equation; these are usually second and third order products of the seven primary predictors listed above. The most important terms are the current intensity, the 12 hour intensity change, the Julian day and the latitude. The SHIFOR equations were developed using data from all historic storms during the period 1900-1972 that were at least 30 nautical miles from land. Thus, the SHIFOR intensity forecasts are not valid for storms less than 30 nautical miles from the coast.

It was in the 1980's when several authors recognized the importance of synoptic data in the prediction of hurricane intensity. The concept of synoptic data is used to represent simultaneous observation of atmospheric variables at different spatial geographic extensions.

Pike (1985) used geopotential height and thicknesses as predictors in an attempt to include synoptic predictors in his statistical prediction model. Also, Merrill (1987) used a wider range of synoptic predictors in an intensity prediction model but failed to provide a significant improvement over climatology and persistence. However, in the study by Merrill, DeMaria states, “The prediction model was developed for tropical cyclones over land as well as over the ocean. It’s probably that the statistical properties of storms that decay over land are quite different from the properties of storms over the ocean.”

DeMaria and Kaplan (1994) presented their model called Statistical Hurricane Intensity Prediction Scheme (SHIPS), a statistical-synoptic model, which was an improvement over SHIFOR because the authors were able to show that the average intensity error is 10-15% less than the error from a model that used only climatology and persistence (SHIFOR). In 1999, an update of SHIPS was presented (DeMaria and Kaplan, 1999). This version was considered a “statistical dynamical” model because data obtained for the first version from global model analysis was removed and synoptic predictors from a numerical models were added.

In this decade, the use of satellite data has brought a new beginning to hurricane research. DeMaria (2002) developed new improvements to the SHIPS model. Data from GOES(Geostationary Operational Environmental Satellites) infrared imagery (10.7 µm), identified more specific brightness temperatures which were previously azimuthally averaged on a 4 km, storm-centered radial grid, and Ocean heat content (OHC) data which at some depth of the ocean is important for tropical cyclone intensity changes.

The Geophysical Fluid Dynamic Laboratory (GFDL) developed a model known as GFDL model which belongs to the third category of hurricane intensity models and it was developed specifically for hurricane tracking and hurricane intensity prediction. It includes 18 sigma levels and uses a horizontal finite-difference method with three nested grids. The two inner grids move to follow the storm, and the resolution of the inner domain is 1/6 degree. The GFDL model includes convective, radiative and boundary layer parameterizations and has a specialized method for initializing the storm circulation. The initial and boundary conditions are obtained from the Aviation run of the Medium Range Forecast (MRF) model. The representation of the storm circulation in the global analysis is replaced with the sum of an environmental flow and a vortex generating by nudging the fields in a separate run of the model to an idealized vortex. This idealized vortex is based upon a few parameters of the observed storm, including the maximum wind, radius of maximum wind and outer wind radii. The environmental flow is the global analysis modified by a filtering technique which removes the hurricane circulation. The forecasts from the interpolated GFDL forecasts are known as the Geophysical Fluid Dynamic Intensity (GFDI) model. A more detailed description of the GFDL model is given by Kurihara et al. (1995).

Kidder et. al. (2000) has described the potential of the Advance Microwave Sounding Unit (AMSU) and how it can be used to predict hurricane intensity. A relationship between temperature anomalies and both the surface wind speed and central pressure of tropical cyclones was found. Several hurricanes (Bonnie, Georges, and Mitch, and Super Typhoon Zeb) were examined, and the maximum temperature anomaly was calculated. In general, the temperature anomalies closely follow both the wind speeds and the pressures. Gaps in the data are caused by the storm being located between orbital swaths or by missing AMSU data. Correlating intensity versus maximum temperature anomaly yields a correlation coefficient of 0.84 and a standard error of 19 kt. Correlating central pressure versus maximum temperature yields a correlation coefficient of 0.86 and a standard error of 12 hPa.

An artificial neural network (ANN) is a mathematical algorithm that pretends to mimic the biological brain by using mathematical models. A typical ANN consists of multiple layers of neurons interconnected with other neurons in the same or different layers so that each neuron acts as an independent processing element. Inputs and interconnection weights are processed by a nested function (typically a weighted summation) to yield a sum that is transformed to a nonlinear representation which is called a transfer function.

Two causes were responsible for the rising of neural networks. The first was the use of statistical mechanics to explain the operation of a certain class of recurrent network, which could be used as an associative memory (Hopfield, 1982). The second key development of the 1980s was the backpropagation algorithm for training multilayer perceptron networks. The most influential of the publications about backpropagation was by Rumelhart (et al. 1986) and it was the answer to the criticisms Minsky and Papert had made two decades ago.

Since then, in the last ten years, thousands of papers have been written, and many applications of neural networks have been found, especially in the field of engineering. They are used in applications such as production (Ramirez-Beltran, 1999a), chemical process control (Ramirez-Beltran et al.1999b, Ramirez-Beltran et al. 2000, 2002a), pattern recognition (Fukushima, 1988), downscaling techniques (Snell et al. 2000), and many others.

Artificial neural networks have been used in many studies related to atmospheric sciences and climate dynamics. Baik et al. (1998) claim that in their intensity model of the North Pacific that using only climatology and persistence predictors, the percent of variance explained by the neural networks model was consistently larger than that explained by the regression model at all time intervals, with an average difference of 12 %. They also pointed out the potential of their work when more sources of information are considered.

Tangang et. al (1998) applied neural network methodology to forecast the sea surface anomaly on three regions: el Niño 4, el Niño 3.5, and el Niño 3. Those regions represent the western-central, the central, and the Easter central parts of the equatorial Pacific Ocean, respectively. The inputs of the neural networks were the extended empirical orthogonal functions of the sea level pressure field that cover the Tropical Indian and Pacific Ocean and evolved over the course of one year. They have claimed that by applying spectral analysis to neural network results, it is possible to identify the important inputs and the nonlinear responses.

Ramirez-Beltrán et. al (2002b) developed a neural networks process to estimate atmospheric variables in three-dimensional space. The authors argue that because the relationship between local climatological events and large-scale phenomenon is nonlinear, neural networks can be used to explain this relationship because of the advantage of learning from data that exhibits a highly nonlinear relationship since the inherent transfer functions are nonlinear.

CHAPTER III METHODOLOGY

3.1. Introduction

Chapter III introduces the methodology of a hurricane intensity prediction algorithm. This methodology is developed under the framework of upper air information and neural networks techniques. It involves five main steps: (1) The methodology is briefly described considering data limitations and other constraints, (2) A database is designed to store information for each North Atlantic hurricane that occurred during the period of 1975 to 2002, (3) A competitive neural network procedure is used to classify hurricanes according to similar meteorological behavior, (4) A random selection scheme is developed to obtain the variables that best explain the variability of hurricane intensity, (5) A feedforward neural network is used to model and predict the hurricane intensity changes using persistence, climatology and synoptic data. Figure 3.1 shows the aforementioned procedure.

3.2. General Description of Methodology

A statistical model to predict hurricane intensity is presented in this work. The developed algorithm creates a different model at every point in time and predicts the hurricane intensity at the following time intervals: 6, 12, 18, and 24 hours. Climatology, persistence and synoptic variables are obtained using historical and actual information. Statistical and artificial intelligence techniques are used to model these meteorological variables. This model is valid for the North Atlantic Ocean so that the successful application of this technique to another basin is related to the knowledge of the variables that have influence over the hurricane intensity in this particular basin.

This model is also limited to those hurricanes that have developed on the North Atlantic basin from 1975 to 2002. This limitation occurred because the atmospheric pressure level was unknown for older hurricanes.

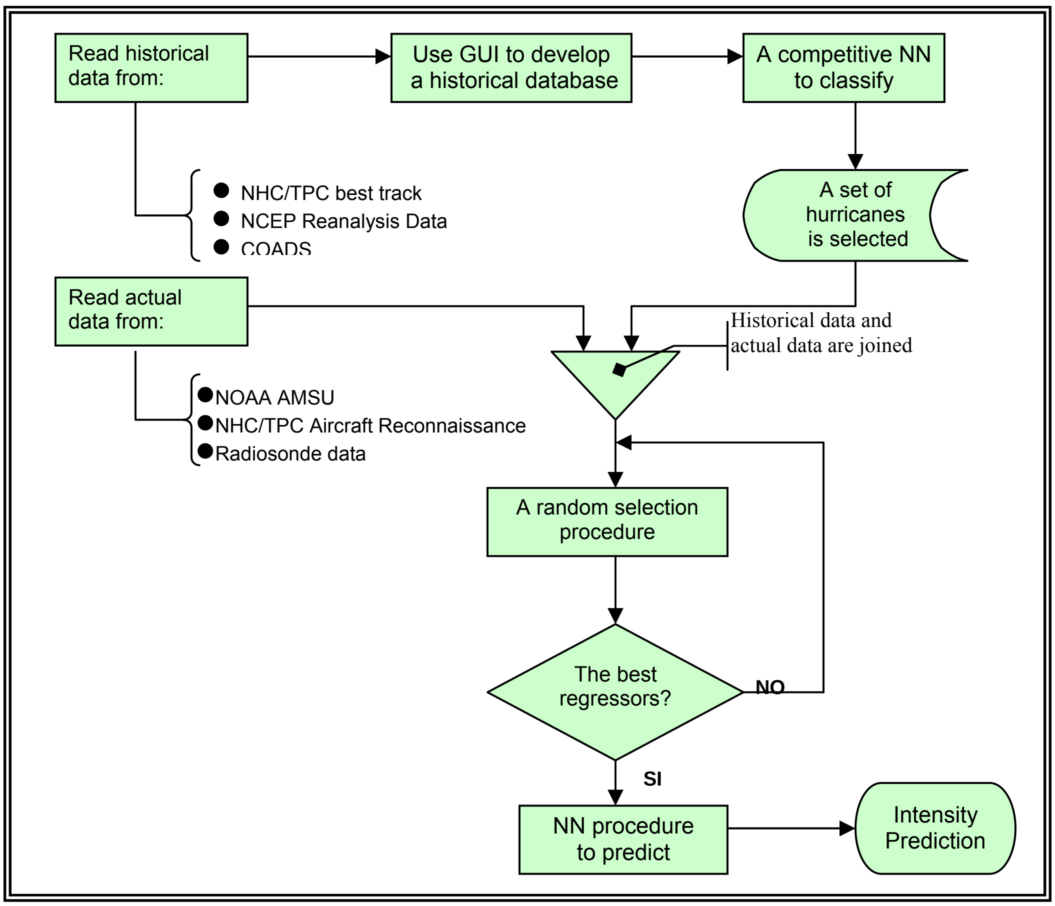


Figure 3.1 Methodology used in this study

3.3. Design of Hurricane Database

A database was developed to store information from past North Atlantic hurricanes using information from several sources, including NCEP, NHC best track and radiosonde observations. This section presents: (1) the data description; and (2) the design of the database.

3.3.1. Data

The National Hurricane Center (NHC) provides a reliable historical data set that is usually known as the best track. The best track is a comprehensive hurricane track analysis after considering all available observations and expert interpretation. Typically the observations are obtained from ships, radars, satellites, airplane reconnaissance, buoys data and other sources. The best track contains observations obtained every 6 hours and includes the following variables: hurricane location, central pressure, hurricane intensity, and storm dates. The hurricane intensity is defined as the average 1-minute maximum sustained winds at sea level. The wind speed is measured in m/sec or in knots. In this study the knot is adopted as a measure for hurricane intensity. The best tracks were obtained from the NHC for events occurred during 1995 to 2002. Before this date, the best track was collected from Unisys (http://weather.unisys.com/ hurricane/index.html), which is an online database dedicated to maintaining hurricane information, including tracks of the storms and the appropriate tracking information.

The sea surface temperature (SST) data used in this work was obtained from the Comprehensive Ocean-Atmosphere Data Set (COADS), which is the most extensive collection of surface marine data available in the world for the past two centuries and can be downloaded from the internet (http://www.cdc.noaa.gov/cdc/data.coads.1deg.html). Monthly mean values of SST are available on a $1°\text{x}1^{o}$ of resolution on the horizontal. These values are linearly interpolated in space and time and were used for estimating the SST at specific hurricane location and a particular time.

The National Center for Environmental Prediction and The National Center for Atmospheric Research (NCEP/NCAR) Reanalysis data was used to obtain the upper air observations at different pressure levels with a $2.5^{o}\text{x}2.5^{o}$ horizontal resolution. This data was obtained at every six hours along the storm track. The NCEP/NCAR reanalysis project is a state-of-the-art reanalysis/forecast system to perform data assimilation using observations from 1948 to the present. A large subset of this data is available from Climate Diagnostic Center (CDC) in its original format as well as its daily averages. The obtained variables from this source are summarized in table 3.1. More information about this data is found in its web page (http://www.cdc.noaa.gov/cdc/data.ncep.reanalysis.html).

Table 3.1 Data obtained from NCEP/NCAR Reanalysis

Variables at 17 pressure levels:	Units	Least Sig. Digit
Air temperature	K	0.1
Geopotential height	m	1.
Relative humidity	%	1.
U-wind speed	m/s	0.1
V-wind speed	m/s	0.1

Radiosonde observations are obtained from radisonde stations located in the Caribbean and in the North Atlantic coasts. Typically, radiosonde observations are obtained at every 12 hours and include the following variables: geopotential height, air temperature, dew point temperature, wind direction and speed. The data is organized and is available since 1997 at the RAOB web site (http://raob.fsl.noaa.gov/). Observations recorded before 1996 can be obtained from CD-ROM.

Satellite data provides almost a real time data and will be incorporated in the intensity prediction model. The AMSU sensor is located in the NOAA 15, 16 and 17 satellites. The AMSU has the property of being almost transparent to the clouds and consequently precise air temperature at different pressure levels can be obtained from this sensor. AMSU data comes in 48x48 km resolution on the horizontal and it is available at about every 6 hours in the Caribbean. The AMSU observations are accessible from a public ftp maintained by Cooperative Institute for Research in the Atmosphere (CIRA). Each variable is stored in one AMSU file, which is composed of three blocks: the area block, the navigation block, and the data block. The area block stores information about the file (location, date, size, etc.). The navigation block can be omitted because it stores technical information about the sensor. The data block, on the other hand, is the most important because it stores the satellite data. Table 3.2 shows AMSU variables which can be accessed by CIRA's AMSU website (http://amsu.cira.colostate.edu/).

Table 3.2 Data obtained from the Advanced Microwave Sensor (AMSU)

Extension	Variable
C01...C20	antenna temps in channels 1-20
RR	AMSU-A rain rate
RRB	AMSU-B rain rate
TPW	total precipitable water
CLW	cloud liquid water
ICE	sea ice
IC2	sea ice (with edges)
SNO	AMSU-A snow cover
SNB	AMSU-B snow cover
LAT	Latitude
LON	Longitude
THK	1000-500 hPa thickness
L07	limb-adjusted channel 7
SFC	surface type (AMSU-A)
SFB	surface type (AMSU-B)
IWP	ice water path
E23	23 GHz emissivity
E31	31 GHz emissivity
E50	50 GHz emissivity
TSF	surface temperature

A large effort has been devoted to create a friendly framework where all of this data can be merged into a single platform and to be accessed at almost any time.

3.3.3. Hurricane Database

A historical database was built to store the climatology, persistence and synoptic observations of the Atlantic hurricanes since 1975. This database is an organized structure divided in fields, where each field contains a specific type of information.

The main advantage of having a database is that the information about any hurricane that developed in the Atlantic can be accessed at any time and used according to the needs. The structure of the database is shown in figure 3.2.

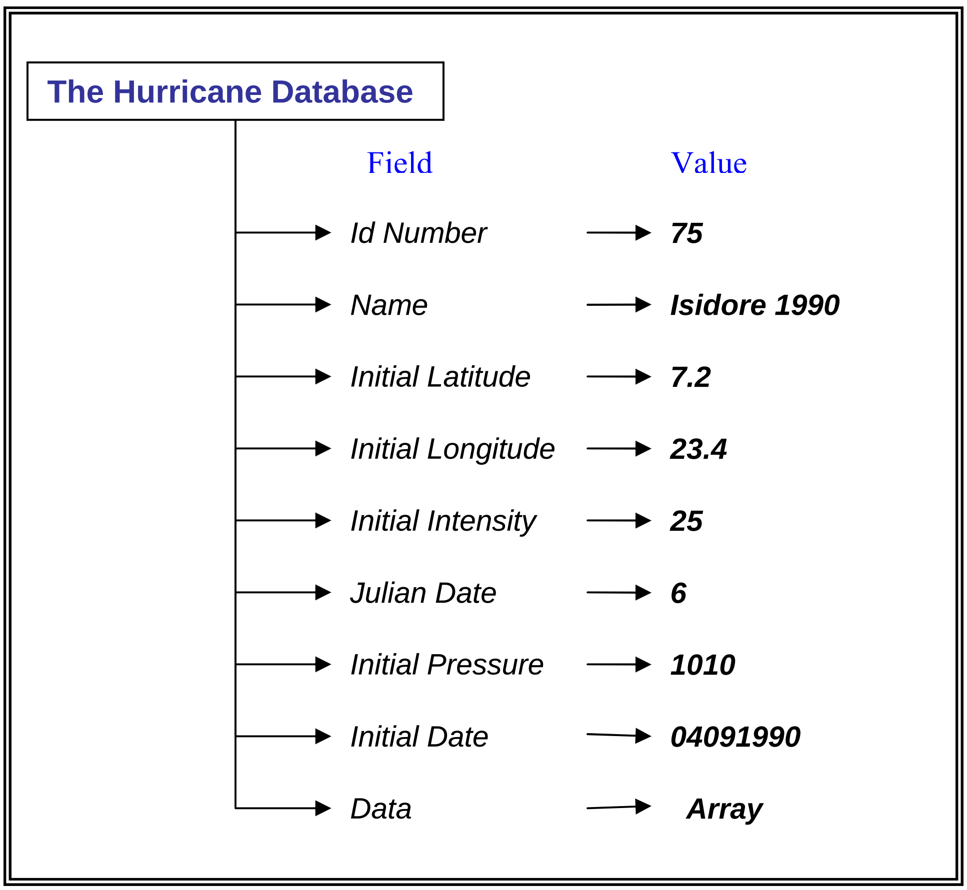


Figure 3.2 Hurricane database structure

The *Id Number* is the first field and is used to sort the hurricane data in ascendant order. The *Name* field is used to store the name of the hurricane. The *Initial Latitude* and *Initial Longitude* fields are necessary to identify the initial location of the storm. The *Initial Intensity, Julian Date, Initial Pressure* and *Initial Date* fields are utilized to save information about the initial state of the hurricane.

Table 3.3 Data Subfields

Subfield	***Source***	***Description***
1	Best track	Month of the storm
2	Best track	Day of the storm
3	Best track	Time
4	Best track	Storm Location Latitude (Slat)
5	Best track	Storm Location Longitude (SLon)
6	Best track	Storm Pressure (SPre)
7	Best track	Storm Intensity (SIn)
8	Calculated	Storm Intensity Change (SInC)
9	Calculated	Eastward comp. of storm motion (ESM)
10	Calculated	Northward comp. of storm motion (NSM)
11	Calculated	Magnitude of the Storm Motion (SM)
12	Interpolated from COADS	Sea Surface Temperature (SST)
13	Calculated	Maximum Possible Intensity (MPI)
14	Interpolated from NCEP	Eastward comp. of wind speed at 850 mb (ES850)
15	Interpolated from NCEP	Northward comp. of wind speed at 850 mb (NS850)
16	Interpolated from NCEP	Eastward comp. of wind speed at 200 mb (ES200)
17	Interpolated from NCEP	Northward comp. of wind speed at 200 mb (NS200)
18	Calculated	Vertical Wind Shear (VWS)
19	Calculated	Average angular momentum at 850 mb (M850)
20	Calculated	Average angular momentum at 200 mb (M250)
21	Calculated	K index (Kin)
22	Calculated	Total Totals (TT)

The core part of the database is found in the *Data* field, which is also divided into 21 subfields that are shown in Table 3.3. This table also shows the subfield code, the source of information and the description of the field.

The Julian Date (T) is calculated using subfields 1, 2, and 3 of *Data* field and is expressed as follows:

$$T = \text{Absolute value (Julian Date} - 253) \qquad (3.3.1)$$

The Julian Date was defined by equation (3.3.1) because the historical information shows that the largest hurricane frequency has occurred on September 10 and corresponds to the Julian date 253. The estimates of the storm intensity and the atmospheric pressure for a given hurricane are available at every 6-hour interval (subfield 6 and 7). The hurricane's displacement on zonal and meridional directions (subfield 4 and 5) are computed using the best track information obtained from the NHC.

The Storm Intensity Change is calculated by computing the difference between two consecutive values of hurricane intensity from subfield 7 and it is stored in subfield 8. The intensity change can be expressed as follows:

$$\Delta I_t = I_t - I_{t-1} \qquad (3.3.2)$$

Where:

ΔI_t= Hurricane intensity change at the time t.

I_t= Hurricane intensity at the time t.

The values used to calculate the eastward and northward movements of a hurricane were obtained from the subfields 4 and 5. They are stored in subfields 9 and 10. The meridional and zonal displacements are expressed as follows:

$\Delta Lo_t = Lo_t - Lo_{t-1}$ (3.3.3)

$\Delta La_t = La_t - La_{t-1}$ (3.3.4)

Where:

ΔLo_t= Eastward displacement at the time t.

Lo_t= Hurricane zonal location at the time t

ΔLa_t= Northward displacement at the time t.

La_t= Hurricane meridional location at the time t.

The magnitude of the storm motion is defined by the following equation:

$\Delta v = (\Delta Lo^2_t + \Delta La^2_t)^{1/2}$ (3.3.5)

The sea surface temperature (SST) is stored in subfield 12 and is estimated using time and space interpolation depending on the date and the location of the tropical cyclone. The Maximum Possible Intensity (MPI) is determined from an empirical relationship developed by DeMaria and Kaplan (1994) and is saved in subfield 13.This relationship is valid for hurricanes that have developed in the Atlantic Basin since 1950.The MPI is defined as follows:

$$MPI = A + B^{[C(SST - SST_0)]}, \quad (3.3.6)$$

Where:

A=66.5 kt,

B=108.5 kt,

C=0.1813 $^oC^{-1}$,

SST_0=30 ^oC.

The NCEP/NCAR reanalysis project is utilized to access wind speed components at 850 and 200 mb. These values are interpolated in space and time to the position and date of each tropical cyclone observation. The eastward (u_{850}) and northward (v_{850}) components of storm speed at 850 mb are stored in subfields 14 and 15 and the eastward (u_{200}) and northward (v_{200}) components of storm speed at 200 mb are saved in subfields 16 and 17, respectively.

The vertical wind shear predictor (S_t), which is the magnitude of the difference between the 850 and 200 mb wind vectors (Knaff et al., 2004), is included in subfield 18. These variables are computed along the storm track and selected at the closest grid where the storm is located. The vertical wind shear can be expressed as follows:

$$S_t = \sqrt{(u_{200} - u_{850})^2 + (v_{200} - v_{850})^2} \qquad (3.3.7)$$

Where:

S_t = Vertical wind shear

u_{850} = Eastward component of storm speed at 850 mb.

u_{200} = Eastward component of storm speed at 200 mb.

v_{850} = Northward component of storm speed at 850 mb.

v_{200} = Northward component of storm speed at 200 mb.

The change of the vertical wind shear (ΔS_t) is considered to be another synoptic variable.

An average momentum is estimated to wind speed at 200 and 850mb. This momentum is calculated to take into account the interactions between the tropical cyclone and synoptic systems. The momentum is calculated for the position and date of each tropical cyclone observation obtained from the NHC best track. It can be expressed as follows:

The average momentum at 850 mb can be estimated as follows:

$$m_{8_j} = \frac{1}{4} r_j \sum_{i=1}^{4} x_i^{(j)} \tag{3.3.8}$$

$$M_{8_t} = \frac{1}{3} \sum_{j=1}^{3} m_{8_j} \tag{3.3.9}$$

The average momentum at 200 mb is defined as follows:

$$m_{2_j} = \frac{1}{4} r_j \sum_{i=1}^{4} x_i^{(j)} \tag{3.3.10}$$

$$M_{2_t} = \frac{1}{3} \sum_{j=1}^{3} m_{2_j} \tag{3.3.11}$$

where:

r_j = distance between the center of the storm and the position where the wind speed (x_i) is obtained (j=1, 2, 3).

$r_1 = 400\,\text{km}$

$r_2 = 600\,\text{km}$

$r_3 = 800\,\text{km}$

$x_i^{(j)}$ = the i^{th} wind speed (i=1, 2, 3, 4) at the j^{th} position. (j=1, 2, 3.)

m_{8_j} = 850 mb. partial momentum at j^{th} location.

m_{2_j} = 200 mb. partial momentum at j^{th} location.

M_{8_t} = 850 mb. momentum at the time t.

M_{2_t} = 200 mb. momentum at the time t.

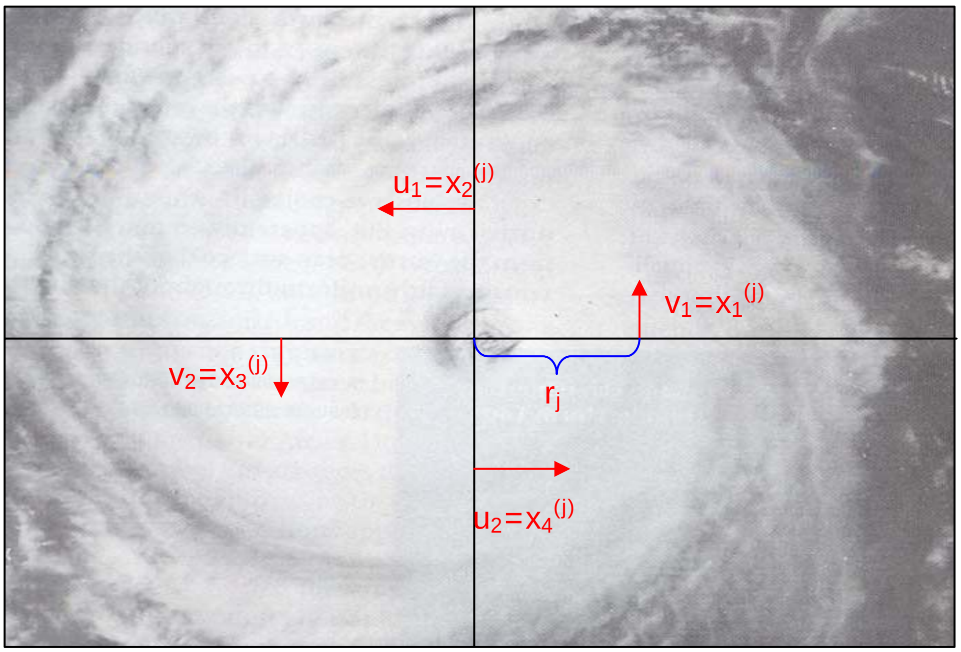


Fig 3.3 Graphical representation of the Momentum variable

The K index is a measure of thunderstorm potential based on vertical temperature lapse rate, moisture content of the lower atmosphere, and the vertical extent of the moist layer. This variable is stored in subfield 20 of the *Data* field. The K index is defined as follows:

$$K = (T_{850} - T_{500}) + \left[T_{d850} - \left(T_{700} - T_{d700}\right)\right] \tag{3.3.7}$$

Where:

T_{850} = Air temperature at 850 mb.

T_{500} = Air temperature at 500 mb.

T_{d850} = Dew point temperature at 850 mb.

T_{700} = Air temperature at 700 mb.

T_{d700} = Dew point temperature at 700 mb.

The higher the K index the greater the likelihood of thunderstorm development. Table 3.4 shows the K index and its probability description.

Table 3.4 K index values

K index	Thunderstorm Probability
< 20	None
20 to 25	Isolated
26 to 30	Widely Scattered
31 to 35	Scattered
> 35	Numerous

The Total Totals (TT) is the last subfield included in the *Data* field. This value is used to identify potential areas of thunderstorm development. General threshold values for Total Totals range from 44 for isolated convection to greater than 55 for numerous thunderstorms. Total Totals can be expressed as follows:

$$VT = T_{850} - T_{500} \tag{3.3.8}$$

$$CT = T_{d850} - T_{500} \tag{3.3.9}$$

$$TT = VT + CT \tag{3.3.10}$$

where:

T_{850} = Air temperature at 850 mb.

T_{500} = Air temperature at 500 mb.

T_{d850} = Dew point temperature at 850 mb.

3.4. Identifying analog hurricanes

A competitive neural network is used to identify analog hurricanes; where an analog is defined as a storm that best resemble the meteorological behavior of the current storm. A competitive neural network is an algorithm that learns associations from observations by identifying similarities among their properties. Once learned, associations allow networks to classify input vectors into clusters or families. This is considered as an unsupervised classification technique because no target variable or response variable is needed, i.e. the algorithm learns by identifying similarities among the provided input variables.

This section is organized as follows. The first part describes the competitive neural network and the second part the neural network will be applied to identify the analog hurricanes.

3.4.1 Description of a Competitive Neural Network

A competitive neural network is composed generally of two layers (Hagan et al., 1996). The first layer computes the direction and other properties of the input patterns and the second layer determines which of the prototype vector is closest to the input vectors.

The first layer is based on a single instar, which is a type of neural network that is capable of performing pattern recognition and is able to recognize only one pattern. To recognize more than one pattern, a set of instars is used. The input/output expression for the instar net is:

$$a = \text{hard}\lim(Wp + b) = \text{hard}\lim({}_1w^T p + b) \qquad (3.4.1)$$

where W represents a matrix of vectors which wants to be recognized, b is set equal to the number of elements in input vector(p), and hard lim is a transfer function that assign the number one if its net input reaches a given threshold, otherwise its outputs will be zero. This rule allows a neuron to perform a classification of the input patterns.

The instar will be activated whenever the inner product between the weight vector and the input is greater than or equal to –b:

$$_1w^T p \geq -b \qquad (3.4.2)$$

It has been shown that for two vectors of constant length, the inner product will achieve the largest value when they point in the same direction.

If the following relation is set:

$$b = -\|{}_1w^T\|\|p\| \qquad (3.4.3)$$

then the instar will only be active when p focus in exactly the same direction as $_1w$ ($\theta = 0$, where θ is the angle between the vectors $_1w^T$ and p). Thus, the neuron will recognize only the pattern $_1w$. To recognize more than one pattern, a variation of the procedure mentioned above has been implemented as follows:

Given the following input vectors:

$$\{p_1, p_2, \ldots, p_Q\}$$

Where:

$$p_1 = [p_{11} \quad p_{12} \quad \cdots \quad p_{1R}] \; ; \; p_2 = [p_{21} \quad p_{22} \quad \cdots \quad p_{2R}] \; ; p_Q = [p_{Q1} \quad p_{Q2} \quad \cdots \quad p_{QR}]$$

The weight matrix, W^1, and the bias vector, b^1, for Layer 1 will be:

$$W_{RxS}{}^1 = \begin{bmatrix} {}_1W^T \\ {}_2W^T \\ {}_3W^T \\ \vdots \\ {}_SW^T \end{bmatrix} \quad , \; b_{Rx1}{}^1 = \begin{bmatrix} R \\ R \\ R \\ \vdots \\ R \end{bmatrix}$$

Where each row of W^1 represents a prototype vector that is needed to be recognized and each element of b^1 is set equal to the number of elements in each input vector (R). The upper subscript in W and b represents the first layer. The number of neurons, S, is equal to the number of prototype vectors which will be identified as Q. The upper subscript T represents the transpose operation. Each row of W^1 can be expressed as follows:

$${}_1W = \begin{bmatrix} {}_1W_{11} \\ {}_1W_{12} \\ {}_1W_{13} \\ \vdots \\ {}_1W_{1Q} \end{bmatrix} , \; {}_2W = \begin{bmatrix} {}_2W_{21} \\ {}_2W_{22} \\ {}_2W_{23} \\ \vdots \\ {}_2W_{2Q} \end{bmatrix} , \; \ldots \; {}_SW = \begin{bmatrix} {}_SW_{S1} \\ {}_SW_{S2} \\ {}_SW_{S3} \\ \vdots \\ {}_SW_{SQ} \end{bmatrix}$$

Thus, the output of the first layer is:

$$a^1 = W^1 p + b^1 = \begin{bmatrix} {}_1W^T p_1 + R \\ {}_2W^T p_2 + R \\ {}_3W^T p_3 + R \\ \vdots \\ {}_SW^T p_Q + R \end{bmatrix} \qquad (3.4.4)$$

It should be noted that the output of the first layer, a^1, is equal to the inner products of the prototype vectors with the input in addition of the constant R. These inner products indicate how close each of the prototype patterns is to the input vector.

The second layer is called competitive layer and it is initialized using the outputs of the first layer. In this layer, the neurons compete with each other to determine a winner. The winning neuron indicates which category of input was presented to the network (each prototype vector represents a category).

The first layer output, a^1, is used to initialize the second layer.

$$a^2(0) = a^1 \tag{3.4.5}$$

Then the second-layer output is updated according to the following recurrence relation:

$$a^2(t+1) = \text{poslin}(W^2 a^2(t)) \tag{3.4.6}$$

where the transfer function poslin is defined as follows:

$$a = \text{poslin}(n) = \begin{cases} 0, \text{if } n < 0 \\ n, \text{otherwise} \end{cases}$$

The second-layer weights W^2 are set so that the diagonal elements are 1, and the off-diagonal elements have a small value as follows:

$$w_{ij}^2 = \begin{cases} 1, & \text{if} \quad i = j \\ -\varepsilon, & \text{otherwise} \end{cases} \qquad \text{where } 0 < \varepsilon < \frac{1}{S-1} \tag{3.4.7}$$

This matrix produces an effect called lateral inhibition, in which the output of each neuron has an inhibitory effect on all of the neurons.

At this point the network has reached a steady state. The index of the second-layer neuron with a stable positive output is the index of the prototype vector that best matched the input. This process is called the "winner-take-all competition" since only one neuron will have a nonzero output.

The Kohonen learning rule is used to train the weights because it allows the weights of a neuron to learn from an input vector. It can be described as follows:

$$ {}_1w(q) = {}_1w(q-1) + \alpha(p(q) - {}_1w(q-1)) \qquad (3.4.8) $$

where the weights at the iteration q are updated using the weights at the iteration $q-1$. Thus, the row of the weight matrix that is closest to the input vector moves toward the input vector. It moves along a line between the old row of the weight matrix and the input vector, as shown in the following graphic:

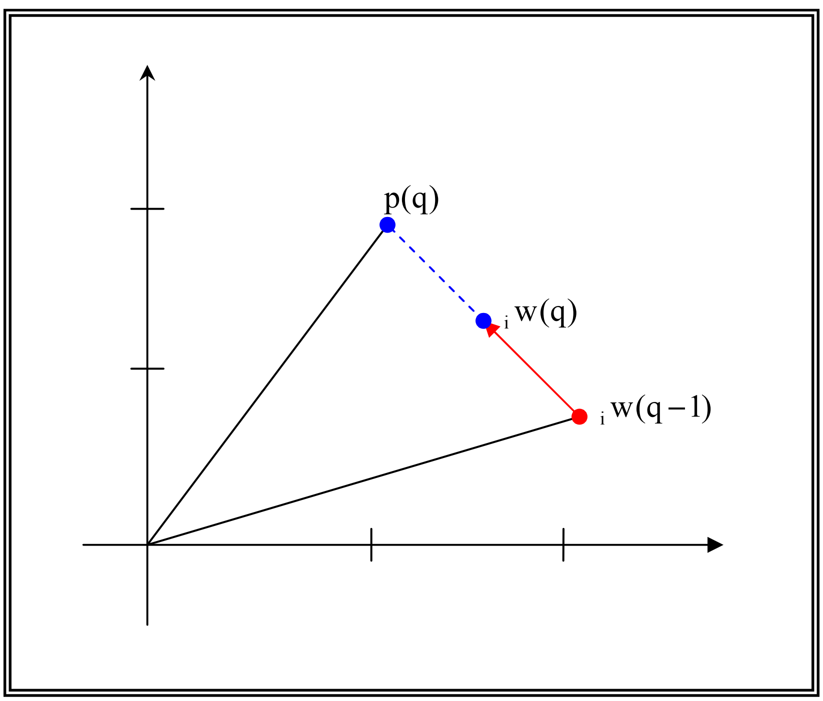


Figure 3.4 Graphical representation of the Kohonen Rule

3.4.2 Application of the Competitive Neural Network

The prediction scheme uses historical data to identify the analog hurricanes. The analogs are based on climatology and persistence variables. At the early stage of the storm, the analog set is derived based on the first six observations, and as soon as the storm life increases the number of observations for selecting analogs will also increase by one unit at a time up to fifteen observations. When more than fifteen observations are available the size of the moving window is maintained fixed to 15 observations. The period from 1975 to 2002 was selected to identify the analogs because most of the synoptic variables were completed

A self organized ANN with the Kohonen learning rule was designed to identify the storm analogs to the current hurricane. Ten neurons were used to characterize each persistence variable of each hurricane. The characteristics of the persistence variables of the current storm are the key to identify the potential analog set. The majority voting rule was used to identify a single code for each hurricane. All hurricanes that have the same code to the current storm were selected to be processed to a second ANN. The second ANN has the Kohonen learning rule and two neurons and again the assigned code to the current storm was used to select the set of analog storms. This process is repeated three times in order to increase the sample size and to derive a robust estimation, i.e., three sets of analog hurricanes were identified at every point in time.

The implemented procedure includes four major steps and they will be described as follows:

1. Once a hurricane is detected NHC will collect a set of parameters (D_A) at every 6 hours since the hurricane detection time until the current time (*t*), defined by the actual time of the hurricane in process. Figure 3.4 shows the time sequence of data collection. The parameters to be used for the identification process of analog hurricanes are: the Julian date, hurricane location (latitude and longitude), hurricane intensity, and hurricane direction. The continuous line in figure 3.4 shows the known intensity magnitudes of the current hurricane up to the current time and the dotted line shows the possible development of hurricane intensity in the near future. The parameter sets (D_{Pi}) are associated to the historical hurricanes stored in the database that are extracted for the same storm life interval of the current hurricane.

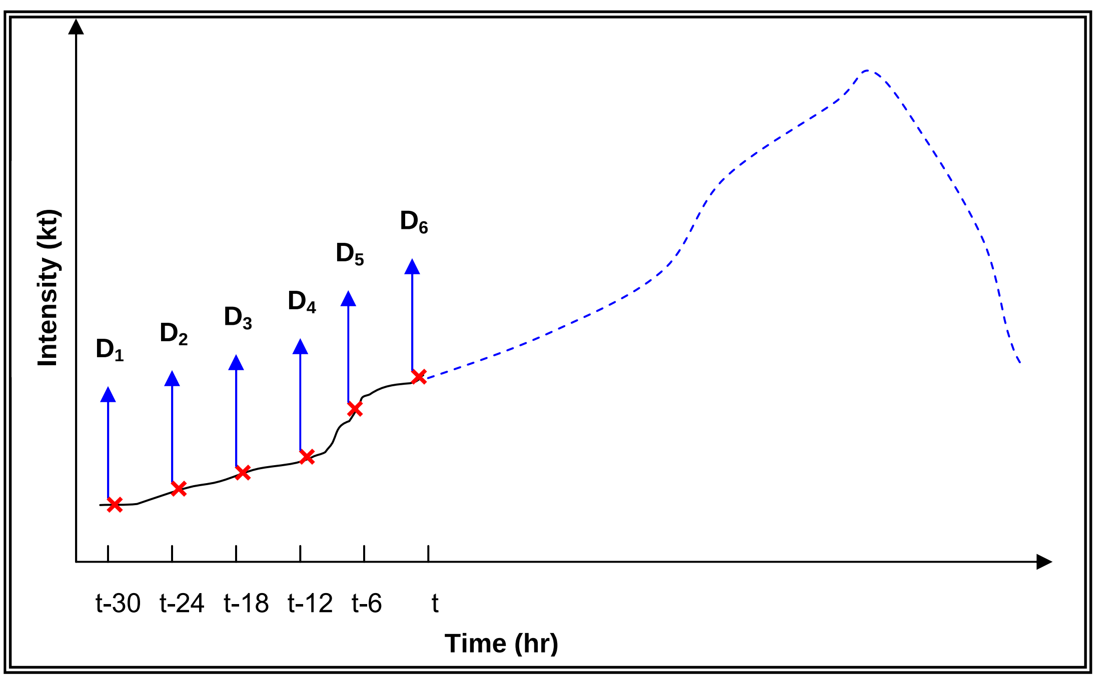


Figure 3.5 Data obtained from the current hurricane

The set parameter, D_A, is a matrix whose columns are the D_i vectors (i=1, 2…, 6) and a single column can be expressed as follows:

$$D_i = \begin{bmatrix} Ju_i \\ La_i \\ Lo_i \\ I_i \\ \alpha_i \end{bmatrix} \qquad D_A = \left[D_1\ D_2 \cdots D_6\right]$$

$$D_A = \begin{bmatrix} Ju_1 & Ju_2 & \dots & Ju_6 \\ La_1 & La_2 & \dots & La_6 \\ Lo_1 & Lo_2 & \dots & Lo_6 \\ I_1 & I_2 & \dots & I_6 \\ \alpha_1 & \alpha_2 & \dots & \alpha_6 \end{bmatrix}$$

Where:

D_1 = Vector with hurricane's information at time *t-30* hrs, and D_2 is the vector at time t-24 hrs, and so on as shown in figure 3.4

Ju_1=Julian Date at the time *t-30* hrs

La_1= Hurricane location latitude at the time *t-30* hrs

Lo_1= Hurricane location longitude at the time *t-30* hrs

I_1= Hurricane intensity at the time *t-30* hrs

α_1= Hurricane direction at the time *t-30* hrs

D_A= Matrix of information for the current hurricane.

D_P is the set that contains information for each of the past hurricane (D_{Pi}) stored in the database and can be expressed as follows:

$$D_{Pi,1} = \begin{bmatrix} Ju_{Pi,1} \\ La_{Pi,1} \\ Lo_{Pi,1} \\ I_{Pi,1} \\ \alpha_{Pi,1} \end{bmatrix}$$

$$D_{Pi} = \begin{bmatrix} D_{Pi,1} & D_{Pi,2} & \dots & D_{i,6} \end{bmatrix} \text{ for } i = 1,2,\dots,n$$

$$D_{Pi} = \begin{bmatrix} Ju_{Pi,1} & Ju_{Pi,2} & \dots & Ju_{Pi,6} \\ La_{Pi,1} & La_{Pi,2} & \dots & La_{Pi,6} \\ Lo_{Pi,1} & Lo_{Pi,2} & \dots & Lo_{Pi,6} \\ I_{Pi,1} & I_{Pi,2} & \dots & I_{Pi,6} \\ \alpha_{Pi,1} & \alpha_{Pi,2} & \dots & \alpha_{Pi,6} \end{bmatrix}$$

$$D_P = \begin{bmatrix} D_{P1} & D_{P2} & \dots & D_{Pi} & \dots & D_{Pn} \end{bmatrix}$$

where:

i = hurricane index in the database

n = number of the hurricanes in the database

$D_{P_{i,j}}$ = vector that contains historical information for hurricane i (i=1,2,…, n) in the time j (j=1,2,…,6)

D_{P_i} = matrix that collect information for the hurricane i

D_T is a matrix formed by the union of the past data set (D_P) and current data set (D_A). This set is used in step 2 and can be expressed as follows:

$$D_T = \begin{bmatrix} D_P & D_A \end{bmatrix}$$

$$D_T = \begin{bmatrix} Ju_{P1,1} & \dots & Ju_{P1,6} & \dots & Ju_{Pi,1} & \dots & Ju_{Pi,6} & \dots & Ju_{Pn,1} & \dots & Ju_{Pn,6} & Ju_1 & Ju_2 & \dots & Ju_6 \\ La_{P1,1} & \dots & La_{P1,6} & \dots & La_{Pi,1} & \dots & La_{Pi,6} & \dots & La_{Pn,1} & \dots & La_{Pn,6} & La_1 & La_2 & \dots & La_6 \\ Lo_{P1,1} & \dots & Lo_{P1,6} & \dots & Lo_{Pi,1} & \dots & Lo_{Pi,6} & \dots & Lo_{Pn,1} & \dots & Lo_{Pn,6} & Lo_1 & Lo_2 & \dots & Lo_6 \\ I_{P1,1} & \dots & I_{P1,6} & \dots & I_{Pi,1} & \dots & I_{Pi,6} & \dots & I_{Pn,1} & \dots & I_{Pn,6} & I_1 & I_2 & \dots & I_6 \\ \alpha_{P1,1} & \dots & \alpha_{P1,6} & \dots & \alpha_{Pi,1} & \dots & \alpha_{Pi,6} & \dots & \alpha_{Pn,1} & \dots & \alpha_{Pn,6} & \alpha_1 & \alpha_2 & \dots & \alpha_6 \end{bmatrix}$$

2. A competitive neural network (CNN) was implemented to classify the input set (D_T) into a preliminary set (D_S). To accomplish this task a number (S) of prototype vectors (W), which are selected in a random way, were defined so that the CNN can learn to detect similarities among the provided data set D_T. This step can be represented mathematically as follows:

$$D_S = \mathrm{Cnn}\,(W_{S\,x\,5} D_{T_{5\,x\,Q}}) \tag{3.4.9}$$

where:

W= prototype vectors

s= class type (s=1, …, S)

q= qth observation (q=1, …, Q)

D_T=is the reunion of historical observations (n) and actual observations.

D_S is a row vector with the same number of columns as D_T and its values fluctuate between 1 and S, which means that each single observation of every one of the hurricanes that composed D_T is classified as follows:

$$D_S = \lfloor P_{1,1}\ P_{1,2}\ P_{1,3}\ P_{1,4}\ P_{1,5}\ P_{1,6}\ ;\dots\ ;P_{i,1}\ P_{i,2}\ P_{i,3}\ P_{i,4}\ P_{i,5}\ P_{i,6};\dots\ ;P_{Q,1}\ P_{Q,2}\ P_{Q,3}\ P_{Q,4}\ P_{Q,5}\ P_{Q,6}\rfloor$$

where $P_{i,j}$ is the observation j (j=1,…,6) of the hurricane i (i=1,…,Q), and takes values between 1 and S. For instance, if S is equal to10, D_S may have the following distribution:

$$D_S = [4,4,3,4,4,7;\ 3,5,3,3,3,8;\ldots;4,9,3,3,3,3\ ;\ 6,5,6,6,6,9;\ \ldots\ ;3,2,3,5,10,3\ ]$$

The D_S vector means that the set composed of historical data and actual data (D_T) has been classified according to each one of the observations. In this way, the first observation of the first hurricane that composed D_T is classified as class 4; the second observation of the first hurricane is classified as the class 4, and so on.

3. A majority voting procedure was implemented to get a unique outcome from the generated information (D_S) by the competitive neural network. After the six observations that correspond to each one of the hurricanes of the dataset (D_T) were classified in S classes, a voting procedure is used to count the decision of each observation. If a majority decision is found, then the decision procedure will determine that the hurricane under analysis belongs to that majority class. The winner class would be the one that has the majority of the votes, and if there is tie any one could be the winner. This rule can be defined as follows:

$$V = \max(D_S) = \max(P_{i,j}) \quad \forall\, i\ (i = 1,\ldots, Q)\ ; \forall j\ (j = 1,\ldots,6) \qquad (3.4.10)$$

Using this rule, the vector V may be expressed as follows:

$$V = [4\ ;\ 3\ ;\ \ldots\ ;3\ ;6\ ;\ \ldots;3]$$

The first element of vector V indicates that the first hurricane in D_T belongs to class 4, because of the majority voting rule; the second element of vector V means that the second hurricane in D_T belongs to class 3, this process is repeated over and over until the last value is found, and it represents the current hurricane, which in this case, belongs to class 3. Therefore, the hurricanes that have the same class to the current hurricane are selected to be the first set of analogous hurricanes (D_N). It follows:

$$D_N = [D_{P2}, \ldots, D_{P35}, \ldots, D_A]$$

$$D_N = \begin{bmatrix} Ju_{P2,1} & \cdots & Ju_{P2,6} & \cdots & Ju_{P35,1} & \cdots & Ju_{P35,6} & \cdots & Ju_1 & \cdots & Ju_6 \\ La_{P2,1} & \cdots & La_{P2,6} & \cdots & La_{P35,1} & \cdots & La_{P35,6} & \cdots & La_1 & \cdots & La_6 \\ Lo_{P2,1} & \cdots & Lo_{P2,6} & \cdots & Lo_{P35,1} & \cdots & Lo_{P35,6} & \cdots & Lo_1 & \cdots & Lo_6 \\ I_{P2,1} & \cdots & I_{P2,6} & \cdots & I_{P35,1} & \cdots & I_{P35,6} & \cdots & I_1 & \cdots & I_6 \\ \alpha_{P2,1} & \cdots & \alpha_{P2,6} & \cdots & \alpha_{P35,1} & \cdots & \alpha_{P35,6} & \cdots & \alpha_1 & \cdots & \alpha_6 \end{bmatrix}$$

The D_N matrix shows that the hurricane (D_{P2}) in the database identified by ID-number 2 is analogous to the current hurricane and also to the hurricane identified by ID-number 35 and D_A represents the actual hurricane. It is important to notice that only two hurricanes are used to explain the idea of how the procedure works, but this does not happen in practice because of the great amount of past hurricanes that can be analogs. .

4. To prevent that a large number of hurricanes can be selected by the first competitive process, a second competitive neural network is developed to ensure that only the hurricanes that have the maximum degree of associations with the current hurricane are kept in the final set of analogous hurricanes (D_F).

$$D_F = Cnn(W_{e X 6} D_{N_{6 X p}}) \quad (3.4.11)$$

where:

p= value that varies between 1 and the number of hurricane in set D_N

D_N= first set of analogous hurricanes

W= matrix of prototype vectors

e = 2 (number of classes for the reclassified process)

The prototypes vectors are defined in this way to assure that if a hurricane is classified as the same class as the current hurricane in the first instance and, in the reclassified process, this hurricane is again classified as the class of the current hurricane, then it is included in the final set of analogous hurricanes (D_F).

Table 3.5 Variables included in the final set of analog hurricanes

$S_{1,1,k}$	***Description***
1	Storm Pressure(SPre)
2	Storm Intensity Change(SInC)
3	Eastward comp. of storm motion(ESM)
4	Northward comp. of storm motion(NSM)
5	Module of the Storm Motion(SM)
6	Sea Surface Temperature(SST)
7	Maximum Possible Intensity(MPI)
8	Eastward component of wind speed at 850 mb(ES850)
9	Northward component of wind speed at 850 mb(NS850)
10	Eastward component of wind speed at 200 mb(ES200)
11	Northward component of wind speed at 200 mb(NS200)
12	Vertical Wind Shear (VWS)
13	Average angular momentum at 850 mb (M850)
14	Average angular momentum at 200 mb (M250)
15	K index(Kin)
16	Total Totals(TT)
17	Maximum possible intensity (MPI) minus Initial Intensity
18	VWS change

However, if the hurricane is not reclassified as the same class as the current hurricane, then it is removed from the final set of analogous hurricanes (D_F).The final set of analogous hurricanes (D_F) is composed of the five variables of each one of the past hurricanes and the current hurricane. In addition to these variables, another set of variables (D_X), shown in table 3.5 is added for each one of the past hurricanes and is calculated for the current hurricane. The set D_F is augmented with the inclusion of D_X, which is composed of synoptic and persistence variables that were not considered in the classification process. Thus, the final set of analogous hurricanes (D_F) can be expressed as a computer iterative statement as i.e., not as an algebraical statement:

$$D_F = (D_F)^T$$

$$D_F = \begin{bmatrix} D_F & D_X \end{bmatrix}$$

$$D_F = \begin{bmatrix}
Ju_{1,1} & La_{1,1} & Lo_{1,1} & I_{1,1} & \alpha_{1,1} & S_{1,1,1} & \cdots & S_{1,1,18} \\
\vdots & \vdots & \vdots & & & \vdots & & \vdots \\
Ju_{1,6} & La_{1,6} & Lo_{1,6} & I_{1,6} & \alpha_{1,1} & S_{1,6,1} & \cdots & S_{1,6,18} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\
Ju_{n,1} & La_{n,1} & Lo_{n,1} & I_{n,1} & \alpha_{n,1} & S_{n,1,1} & \cdots & S_{n,1,18} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\
Ju_{n,1} & La_{n,6} & Lo_{n,6} & I_{n,1} & \alpha_{n,1} & S_{n,6,1} & \cdots & S_{n,6,18} \\
Ju_{1} & La_{1} & Lo_{1} & I_{1} & \alpha_{1} & S_{1,1} & \cdots & S_{1,18} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\
Ju_{6} & La_{6} & Lo_{6} & I_{6} & \alpha_{6} & S_{6,1} & \cdots & S_{6,18}
\end{bmatrix}$$

where $S_{i,j,k}$ is the k^{th} synoptic or persistence variable (Table 3.4.1) for the i^{th} hurricane and for the j^{th} observation and $S_{p,q}$ is the q^{th} synoptic or persistence variable for the current hurricane and for the p^{th} observation.

3.5. Random Variable Selection Process

A variable selection procedure is implemented to choose among the variables generated in the previous process, those that best explained the hurricane intensity behavior. The variable selection technique has been widely used throughout the years to find an appropriate number of regressors that can help to reduce the efforts of data collection and model maintenance.

This section is organized as follows: first, the estimation of the regression coefficient using the method of least squares will be briefly described; second, the random variable selection will be applied to select the variables that explain best the hurricane intensity

3.5.1 Estimation of regression coefficients

In general, given a single variable (*y*) dependent on *k* independent variables, for example, x_1, x_2, …, x_k, the relationship between these variables is characterized by a mathematical model called regression model which can be expressed as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \qquad (3.5.1)$$

The parameters β_j, j=0, 1, 2,…, k are called the regression coefficients.

The method of least squares chooses the β's in equation3.5.2 so that the sum of the squares of the errors, $\sum \varepsilon^2$, is minimized. The least squares estimators can be derived as follows:

$$\mathrm{S}(\beta_0, \beta_1, \ldots, \beta_\mathrm{k}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \varepsilon_\mathrm{i}^2 \qquad (3.5.2)$$

$$= \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2$$

The function S in a matrix form can be given as follows:

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) \qquad (3.5.3)$$

$$= y'y - 2\beta' X' y + \beta' X' X \beta$$

The least-squares estimators must satisfy

$$\left.\frac{\partial S}{\partial \beta}\right|_{\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

Then, the least-squares estimators of β are:

$$\hat{\beta} = (X'X)^{-1} X' y \qquad (3.5.4)$$

3.5.2. Application of the random variable selection procedure

The regression technique described above is used to correlate the intensity of a current storm with climatological and synoptic variables of analogous saved on the set (D_F) obtained in the section 3.4 and the process includes the following steps:

1. The analogous set (D_F) was divided in two subsets: one is called the response variable (Y), composed of the hurricane intensity known up to time *t,* and the other is called the predictors variables represented by the matrix (X) and its elements are the variables of the analogous. The matrix representation can be expressed as follows:

$$Y = \begin{bmatrix} I_{1,1} \\ \vdots \\ I_{1,6} \\ \vdots \\ I_{n,1} \\ \vdots \\ I_{n,6} \\ I_1 \\ \vdots \\ I_6 \end{bmatrix} \qquad X = \begin{bmatrix} La_{1,1} & Lo_{1,1} & S_{1,1,1} & \cdots & S_{1,1,18} \\ \vdots & \vdots & \vdots & & \vdots \\ La_{1,6} & Lo_{1,6} & S_{1,6,1} & \cdots & S_{1,6,18} \\ \vdots & \vdots & \vdots & & \vdots \\ La_{n,1} & Lo_{n,1} & S_{n,1,1} & \cdots & S_{n,1,18} \\ \vdots & \vdots & \vdots & & \vdots \\ La_{n,6} & Lo_{n,6} & S_{n,6,1} & \cdots & S_{n,6,18} \\ La_{1} & Lo_{1} & S_{1,1} & \cdots & S_{1,18} \\ \vdots & \vdots & \vdots & & \vdots \\ La_{6} & Lo_{6} & S_{6,1} & \cdots & S_{6,18} \end{bmatrix}$$

Where $I_{1,1}$ is the first intensity observation for the first analog hurricane that composed of the set Y, $I_{n,1}$ is the first intensity observation for the last analog hurricane that composed of the set Y, n is the number of analogs hurricanes, and I_1 is the first intensity observation for the current hurricane.

2. A lead time (t_g) was defined so that the correlations between the dependent variable (Y) and the independent variable (X) could be lagged by t_g periods of time. The value t_g varied between 1 and 4, where t_g=1 indicates that the lag period is six hours, and t_g= 4 indicates that the lag period is 24 hours. Considering the following information:

$$Y = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \end{bmatrix} \text{ and } X = \begin{bmatrix} La_1 & Lo_1 & S_{1,1} & \cdots & S_{1,18} \\ La_2 & Lo_2 & S_{2,1} & \cdots & S_{2,18} \\ La_3 & Lo_3 & S_{3,1} & \cdots & S_{3,18} \\ La_4 & Lo_4 & S_{4,1} & \cdots & S_{4,18} \\ La_5 & Lo_5 & S_{5,1} & \cdots & S_{5,18} \\ La_6 & Lo_6 & S_{6,1} & \cdots & S_{6,18} \end{bmatrix}$$

and the lead time (t_g) is set at 4. This lead lime indicates that the dependant variable(Y) at the time *t* is explained using information(X) at the time *t*-4. Thus, it follows:

$$Y = \begin{bmatrix} I_5 \\ I_6 \end{bmatrix} \text{ and } X = \begin{bmatrix} La_1 & Lo_1 & S_{1,1} & \cdots & S_{1,18} \\ La_2 & Lo_2 & S_{2,1} & \cdots & S_{2,18} \end{bmatrix}$$

The first four observations in the dependent variables are eliminated because of the effect of the lag. The last four observations in the independent variables are saved and removed from the matrix to be used at the prediction stage. Then, the application of a given lead time (t_g) can be described mathematically as follows:

$$Y = \begin{bmatrix} I_{t_g+1} \\ I_{t_g+2} \\ \vdots \\ I_{t-1} \\ I_t \end{bmatrix} \text{ and } X = \begin{bmatrix} La_1 & Lo_1 & S_{1,1} & \cdots & S_{1,18} \\ La_2 & Lo_2 & S_{2,1} & \cdots & S_{2,18} \\ \vdots & \vdots & \vdots & & \vdots \\ La_{(t-1)-t_g} & Lo_{(t-1)-t_g} & S_{(t-1)-t_g,1} & \cdots & S_{(t-1)-t_g,18} \\ La_{t-t_g} & Lo_{t-t_g} & S_{t-t_g,1} & \cdots & S_{t-t_g,18} \end{bmatrix}$$

$$\forall\, t - t_g \geq 1$$

Table 3.6 Lag value using in the model

Lag Value(t_g)	Time (hours)
2	12
3	18
4	24

3. After step two was completed, a series of mathematical transformations were used in order to explore a possible nonlinear relationship between the regressors (X) and independent variable(Y). In order to accomplish this task, a

number of models were defined so that each model had a particular mathematical transformation as follows:

$$X^{i}_{a,(n+1)*b} = \begin{bmatrix} X_{a,b} & F^{1}(X)_{a,b} & \cdots & F^{n}(X)_{a,b} \end{bmatrix}$$

where *i* is the number of the model, a is the number of observations in the model, b is the number of variables in the model and the exponent *n* takes values between 1 and 4 as summarized in table 3.7

Table 3.7 Mathematical transformations used in the prediction process

n	**Transformation**
1	Logarithm
2	Quadratic
3	Inverse
4	Power of three

Table 3.8 Mathematical transformations for the regressors

Model(M_r)	***Response***	***Regressors***	***Number of variables***
1	Y	X	20
2	Y	X log(X)	40
3	Y	X X^2	40
4	Y	X X^{-1}	40
5	Y	X X^3	40
6	Y	X log(X) X^2	60
7	Y	X log(X) X^{-1}	60
8	Y	X X^2 X^{-1}	60
9	Y	X log(X) X^2 X^3	80
10	Y	X X^2 X^{-1} X^3	80
11	Y	X log(X) X^{-1} X^3	80
12	Y	X log(X) X^2 X^{-1} X^3	100

Table 3.8 shows the original set of regressors (X), which have changed into a M_r set of regressors (r=1,…, 12). It should be noted that the number of variables increase as soon as a new mathematical transformation is added to the original set of regressors (X).

The amount of data is manageable when the number of variables is small (i.e., less than twenty). However, it becomes burdensome when the number of variables is large (i.e., greater than thirty).

4. To overcome this shortcoming, a random variable selection scheme was developed. This process has the ability to select the regressors that best fit the dependent variable, in this case the hurricane intensity. The procedure can be described as follows: First, the regressors of a selected model (M_r) are divided into *n* subsets of m variables, using the following rule:

   Given the regressor set (X) of the model M_r with *a* observations (rows) and *b* variables (columns), then the number of new variables (*m*) for each *n* subset is calculated as follows:

   - If the number of variables (*b*) is less than the twenty percent of the number of observations (*a*), then the number of new variables (*m*) per subset is set equal to *b*; otherwise, the number of new variables(*m*) is rounded to nearest integer of the twenty percent of the number of observations(*a*).

   The number of subsets in the selection process is calculated as follows:

   - If the modulus of the division between the number of variables(b) and the new variables(m) is equal to zero, then the number of subsets is equal to this division; otherwise, the number of subset is equal to this division plus one, as shows in the following code:

     If mod (b/m) = 0
     Then n=b/m

Else

n=b/m

n=n+1

End

where mod = function to calculate the modulus after division

Then, the set X is divided in *n* subset if the conditional is true as follows:

$$X_{a,b} = \left[X^{1}_{a,m} \quad X^{2}_{a,m} \quad \cdots X^{i}_{a,m} \quad \cdots \quad X^{n-1}_{a,m} \quad X^{n}_{a,m}\right]$$

But if the conditional is not true:

$$X_{a,b} = \left[X^{1}_{a,m} \quad X^{2}_{a,m} \quad \cdots \quad X^{i}_{a,m} \quad \cdots X^{n-1}_{a,m} \quad X^{n}_{a,m} \quad X^{n+1}_{a,\mathrm{mod}\left(\frac{b}{m}\right)}\right]$$

Second, each one of these *n* or *n+1* regressor subsets and its corresponding response (Y) is adjusted using a Matlab program called *Stepwisefit*, which is specially designed to fit regression models using stepwise regression. Stepwise regression is the combination of two procedures called forward and backward regression and is used to find a satisfactory number of regressors that best fit to a given response variable when the number of regressors is large but smaller that the number of observations.

To fit a regressor set ($X^{i}_{a,m}$), the stepwisefit function is executed as many times as the set requires. However, two conditions must be satisfied: first, the number of runs for the stepwisefit function has to be less or equal than a constant that gives the maximum number of runs. Second, the final variables chosen by the selection procedure must be less or equal than a number of maximum variables defined by the forecaster. To this work the maximum number of variables allowed by the system was seven variables.

The stepwisefit program will stop when both conditions must be fulfilled. Then, only the best regressors for each one of the *n* or *n+1* sets will be selected and collected to create a new set of regressors called the best subset (X^r_{BS}) and corresponding to the model *r*. The following figure shows the procedure used:

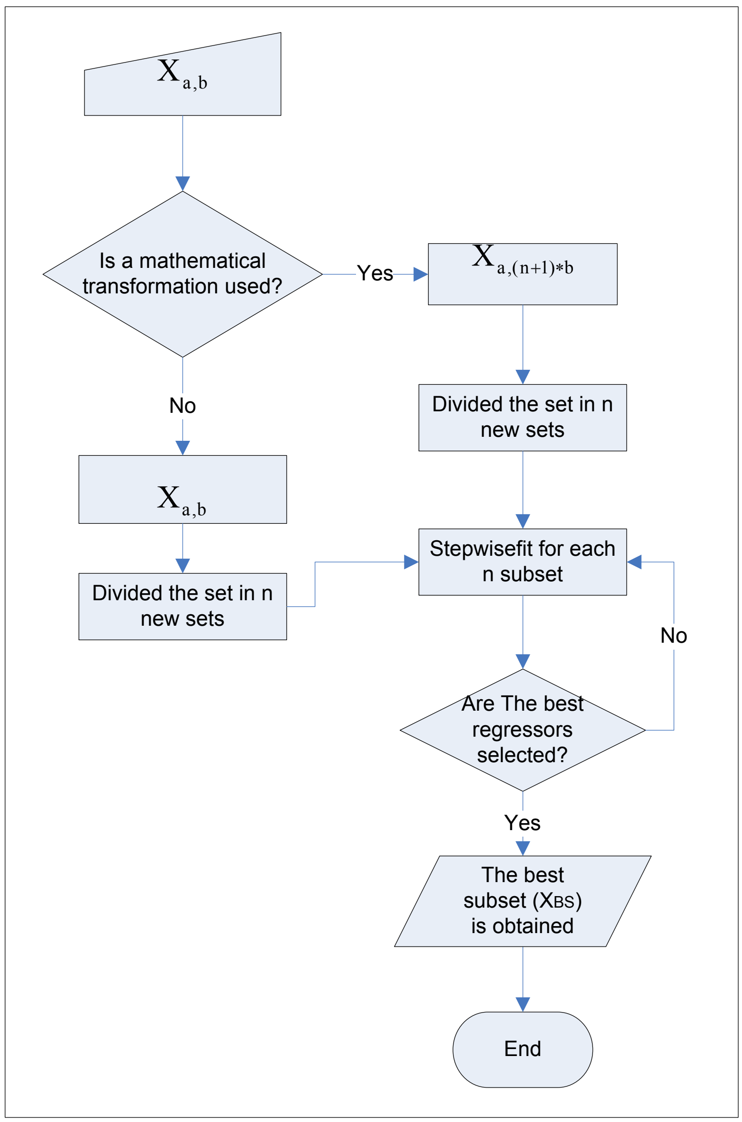


Figure 3.6 Stepwise selection procedure

The R_a^2 is used to measure the adequacy for each *n* subset. This is a statistic used to compare the adjustment of different regression models and to avoid the difficulty of interpreting the classic coefficient of determination (R^2). The R_a^2 does not necessarily increase as additional regressors are introduced into the model and it is better when its value is close to one (Montgomery et al.,1992) .Moreover, a multicollinearity test is performed for the best subset ($\mathrm{X}_{\mathrm{BS}}^{\mathrm{r}}$) to discard the near-linear dependence among the best subset variables. The maximum and minimum eigenvalues associated to matrix ($\mathrm{X}_{\mathrm{BS}}^{\mathrm{r}\,'}\mathrm{X}_{\mathrm{BS}}^{\mathrm{r}}$) are used to calculate the following index:

$$\mathrm{k} = \frac{\lambda_{\max}}{\lambda_{\min}}$$

Generally, if the index is less than 100, there is no serious problem with multicollinearity. If the index is between 100 and 1000, it implies moderate to strong multicollinearity. If *k* exceeds 1000, however, severe multicollinearity is indicated. If a multicollinearity problem is not present in the best subset$\left(\mathrm{X}_{\mathrm{BS}}^{\mathrm{r}}\right)$, then the process is finished and the best subset is ready to be used in the intensity prediction; otherwise, the best subset is discarded and the number of available models (M_r) for calculating a best subset is reduced by one unit. In summary, there are three sets of analogous hurricanes, and each set generates twelve different best subsets, and consequently, the total number of best subsets is 36 at each time interval. The best subset will be saved to be used in the prediction process.

3.6. Intensity Prediction Using Feedforward Neural Network

A feedforward neural network model (Figure 3.6) is characterized by receiving input information to accomplish a modeling identification task without processing feedback information. The training patterns are presented to the network model several times until eventually the algorithm determines the optimal weights and biases that minimize the deviation between the network outputs and the established targets. The feedforward neural network model uses a variation of the standard backpropagation algorithm as the learning rule, which is based on the steepest descent algorithm. The errors are used to modify the searching direction and the gradient is computed at each layer starting from the last layer and finishing with the first layer. This is the reason for the backpropagation name.

This section is organized as follows: in the first part a variation of the standard backpropagation algorithm called the Levenberg-Marquardt is briefly described and in the second part this algorithm is applied to predict the hurricane intensity using the best subset obtained in the previous section.

3.6.1 Description of the Levenberg-Marquardt Backpropagation Algorithm

The following lines briefly describe the use of this algorithm for training multilayer networks. This algorithm, which is described in details in Hagan and Menhaj (1994), is preferred over the standard backpropagation algorithm because it is the fastest algorithm tested for multilayer neural networks (Hagan et al., 1996).The standard backpropagation depends to find a optimum solution in the selection of the learning rate which is easy to select when the network has a single layer but it gets complex when the network has more than two hidden layers. The Levenberg-Marquardt algorithm to overcome this problem generates a learning rate which changes between the steepest descent and Newton algorithm depending on whether the value of the function is near to the optimum solution or not. A key drawback of Levenberg-Marquardt over standard backpropagation is the storage requirement. The algorithm must store the approximate Hessian matrix which is an $r_x r$ matrix where r is the number of parameters (weights and bias) in the network.

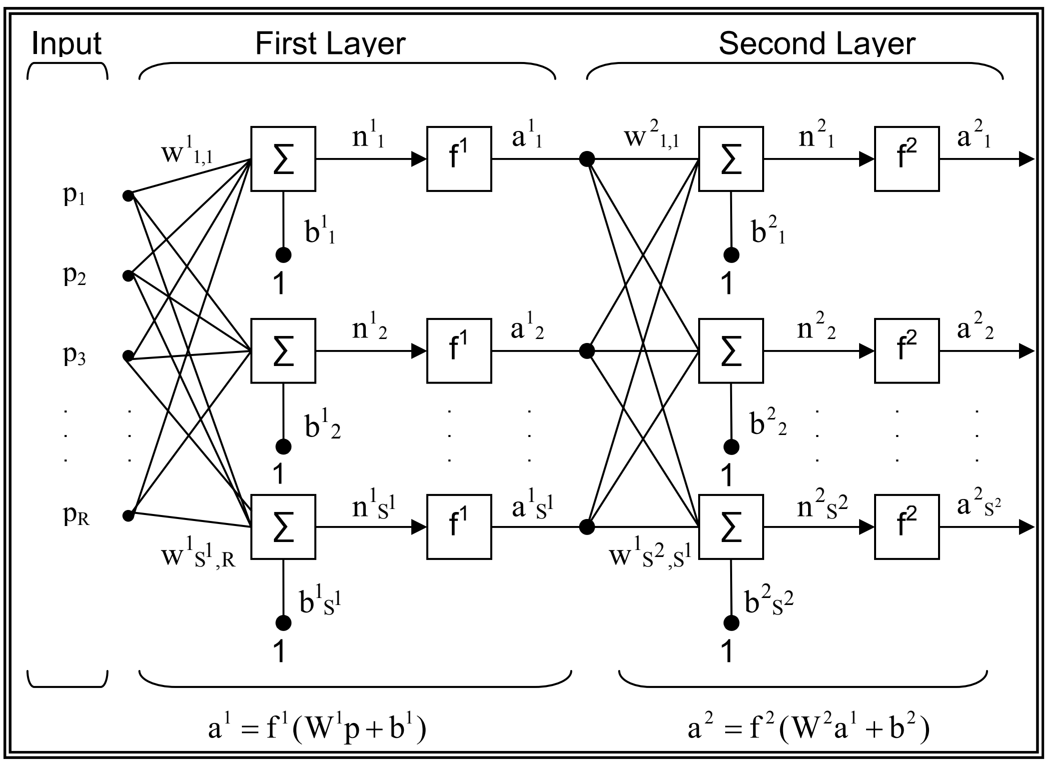


Figure 3.7 Feedforward Neural Network used in this work

The iterations of the Levenberg-Marquardt backpropagation algorithm can be summarized as follows:

Given $\{p_1, t_1\}, \{p_2, t_2\}, \cdots, \{p_Q, t_Q\}$ where p_q is an input to the network and t_q is the corresponding target output.

1. Presents all inputs to the network and compute the corresponding networks outputs and the errors.

$$a^0 = p_1 \tag{3.6.1}$$

$$a^{m+1} = f^{m+1}(W^{m+1}a^m + b^{m+1}) \text{ for } m = 0,2,\cdots,M-1 \tag{3.6.2}$$

$$a = a^M \tag{3.6.3}$$

Where a^0 is the neural network output for the input layer and p_1 is the neural network input, a^{m+1} is the output for the layer $m+1$, f^{m+1} is the transfer function for the layer $m+1$, W^{m+1} is the weights matrix for the same layer and b^{m+1} is the bias vector for the layer $m+1$, a^M is the output for the last layer.

2. Compute the sum of squared errors over all inputs, F(x), using the following equations:

$$F(x) = \sum_{q=1}^{Q} (t_q - a_q)^T (t_q - a_q) \tag{3.6.4}$$

3. Compute the Jacobian matrix using the equation 3.6.5. Calculate the sensitivities with the recurrence relations (equation 3.6.7), after initializing with equation 3.6.6. Augment the individual matrices into the Marquardt sensitivities matrix using 3.6.8. Compute the elements of the Jacobian matrix with equation 3.6.9 and 3.6.10.

$$J(x) = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial w_{1,1}^{1}} & \frac{\partial e_{1,1}}{\partial w_{1,2}^{1}} & \cdots & \frac{\partial e_{1,1}}{\partial w_{S^1,R}^{1}} & \frac{\partial e_{1,1}}{\partial b_{1}^{1}} & \cdots \\ \frac{\partial e_{2,1}}{\partial w_{1,1}^{1}} & \frac{\partial e_{2,1}}{\partial w_{1,2}^{1}} & \cdots & \frac{\partial e_{2,1}}{\partial w_{S^1,R}^{1}} & \frac{\partial e_{2,1}}{\partial b_{1}^{1}} & \cdots \\ \vdots & \vdots & & \vdots & \vdots & \\ \frac{\partial e_{S^M,1}}{\partial w_{1,1}^{1}} & \frac{\partial e_{S^M,1}}{\partial w_{1,2}^{1}} & \cdots & \frac{\partial e_{S^M,1}}{\partial w_{S^1,R}^{1}} & \frac{\partial e_{S^M,1}}{\partial b_{1}^{1}} & \cdots \\ \frac{\partial e_{1,2}}{\partial w_{1,1}^{1}} & \frac{\partial e_{1,2}}{\partial w_{1,2}^{1}} & \cdots & \frac{\partial e_{1,2}}{\partial w_{S^1,R}^{1}} & \frac{\partial e_{1,2}}{\partial b_{1}^{1}} & \cdots \\ \vdots & \vdots & & \vdots & \vdots & \end{bmatrix} \qquad (3.6.5)$$

$$S_{q}^{\sim M} = -\dot{F}^{M}(n_{q}^{M}) \qquad (3.6.6)$$

$$S_{q}^{\sim M} = -\dot{F}^{M}(n_{q}^{M})(W^{m+1})^{T} S^{\sim m+1} \qquad (3.6.7)$$

$$S^{\sim m} = \left[S_{1}^{\sim m} \middle| S_{2}^{\sim m} \middle| \cdots \middle| S_{Q}^{\sim m} \right] \qquad (3.6.8)$$

$$[J]_{h,l} = \frac{\partial v_h}{\partial x_l} = \frac{\partial e_{k,q}}{\partial w_{i,j}^{m}} = \frac{\partial e_{k,q}}{\partial n_{i,q}^{m}} \text{ x } \frac{\partial n_{i,q}^{m}}{\partial w_{i,j}^{m}} = s_{i,h}^{\sim m} \text{ x } \frac{\partial n_{i,q}^{m}}{\partial w_{i,j}^{m}} = s_{i,h}^{\sim m} \text{ x } a_{j,q}^{m-1} \qquad (3.6.9)$$

$$[J]_{h,l} = \frac{\partial v_h}{\partial x_l} = \frac{\partial e_{k,q}}{\partial b_{i}^{m}} = \frac{\partial e_{k,q}}{\partial n_{i,q}^{m}} \text{ x } \frac{\partial n_{i,q}^{m}}{\partial b_{i}^{m}} = s_{i,h}^{\sim m} \text{ x } \frac{\partial n_{i,q}^{m}}{\partial b_{i}^{m}} = s_{i,h}^{\sim m} \qquad (3.6.10)$$

where J(x) is the Jacobian matrix, $S^{\sim M}$ is the sensitivity which describes a recurrence relationship between this and the sensitivity at layer $m+1, S^{\sim m}$ is the Marquardt sensitivities matrix and $[J]_{h,l}$ is the element of the Jacobian matrix that corresponds to the h^{th} row and l^{th} column.

4. Use equation 3.6.11 to obtain Δx_k

$$\Delta x_k = -\left[J^T(x_k)J(x_k) + \mu_k I\right]^{-1} J^T(x_k)v(x_k) \qquad (3.6.11)$$

5. Recalculate the sum of squared errors using $x_k + \Delta x_k$. If this new sum of squares is smaller than that computed in step 2, then divide μ (small value, e.g., μ=0.01) by ϑ (larger than one, e.g., ϑ=10), let $x_{k+1} = x_k + \Delta x_k$ and go back to step 1. If the sum of squares is not reduced, then multiply μ by ϑ and go back to step 4.

6. The process continues until the difference between the network response and the target function reaches some acceptable level. This algorithm converged when the sum of square errors has been reduced to some error goal.

3.6.2 Application of the Levenberg-Marquardt Backpropagation Algorithm

The best subsets (X^r_{BS}) obtained in the selection variable procedure are used as input data for this algorithm which is used to estimate the hurricane intensity in a determined interval of time (t_g).

The procedure used to train the neural network can be divided in five major tasks:

1. **Assembling the Training Data**

The current inputs (X^r_{BS}) and their target vectors (hurricane intensity) are arranged in the following way:

$$X^r_{BS} = \begin{bmatrix} x^r_{1,1} & x^r_{1,2} & \cdots & x^r_{1,n-1} & x^r_{1,n} \\ x^r_{2,1} & x^r_{2,2} & \cdots & x^r_{1,n-1} & x^r_{1,n} \\ \vdots & \vdots & & \vdots & \vdots \\ x^r_{(t-1)-t_g,1} & x^r_{(t-1)-t_g,2} & \cdots & x^r_{(t-1)-t_g,n-1} & x^r_{(t-1)-t_g,n} \\ x^r_{t-t_g,1} & x^r_{t-t_g,2} & \cdots & x^r_{(t-1)-t_g,n-1} & x^r_{t-t_g,n} \end{bmatrix} \text{ and } Y = \begin{bmatrix} I_{t_g+1} \\ I_{t_g+2} \\ \vdots \\ I_{t-1} \\ I_t \end{bmatrix}$$

where $X^{r}_{t-t_g,n}$ is the *n* best regressors for the model *r* at the *t-t_g* time and I_t is the intensity at the time *t,* X^{r}_{BS} is the best subset of regressors for the model *r.*

2. Create the Network Object

A Matlab routine is used to create the network object. The function Newff creates a feedforward neural network (NN) and also initializes the weights and biases of the network; therefore the network is ready for training. The function can be expressed as follows:

NET =newff (pr,[s_1 s_2 … s_{n1}],{tf1 tf2 … tfn},btf)

where:

pr =is a nx2 matrix of min and max values for $(X^{r}_{BS})^T$

s_i = size of i^{th} layer, for nl layers

tfn_i = transfer function of i^{th} layer

btf = backpropagation network training function, in this case Levenberg-Marquardt algorithm.

3. Train the Network

Once the network weights and biases had been initialized, the network is ready to be trained. The training process requires a training set (network inputs and target outputs). During training, the weights and biases of the network are adjusted to minimize the sum of square errors (F(x)).The training process is implemented using a Matlab function called Train that has the following parameters:

[net ,tr] = train (NET, p, t)

where:

NET=original network

p = network inputs

t = network targets

net= new network

tr = training record

In this study, the training process is divided into 2 stages: One, to find out the optimum number of neurons in the hidden layer and the transfer function for the hidden and output layer. At this stage, a number of neurons was previously defined to the hidden layer and two arrays of transfer function for the hidden and output layers (first, Log-Sigmoid and Purelin, and second, Tan-Sigmoid and Purelin). Then, an iterative process is developed to test each one of the neurons with each one of the transfer function arrays. Use the sum of square error as a performance index and, select those that have the minimum sum of square errors.

Once the optimum number of neurons and the transfer functions have been found, a random search is performed to obtain the best initial point using the sum of square errors as a performance index. The weights and bias that correspond to the point that had the minimum sum of square errors are loading by the neural network that up to this point is ready to simulate new inputs.

5. Evaluate the Network Response to New Inputs

Up to this point, the NN has been trained using the inputs (X^r_{BS}) and compared with the known hurricane intensity (Y) to minimize the performance function. The next step is to present current hurricane parameters to the trained NN for obtaining a hurricane intensity prediction at the desired lead time ($t+t_g$) as follows:

$$X^r_{New} = \begin{bmatrix} x^r{}_{t,1} & x^r{}_{t,2} & \cdots & x^r{}_{t,n-1} & x^r{}_{t,n} \end{bmatrix}$$

$$I_{t+t_g} = \text{sim}(\text{net}, (X^r_{New})^T)$$

where X^r_{New} represents the hurricane parameters at the present time (t), I_{t+t_g} is the predicted intensity in the lead time(t_g) and *sim* is a Matlab function used to evaluate the neural network (net) when new input values are provided.

Finally, it is important to notice that each best subset (X^r_{BS}) is used to train an NN three times and a hurricane intensity prediction was obtained. Since the NN is a nonlinear optimization algorithm and highly dependent on the initial point, the NN will provide different results after every training process. Thus, an individual best subset was used to perform three predictions and its median was selected as the prediction for the best subset. The best prediction from the 36 subsets is selected based on attempting to avoid the multicollinearity problem and minimize the NN mean square error.

CHAPTER IV EXPERIMENTAL RESULTS

This chapter presents the results obtained by the application of the proposed intensity prediction model to a set of hurricanes from the North Atlantic Ocean. . This set was selected to contain hurricanes from different Saffir-Simpson's categories so that the model can be evaluated by using different types of intensification patterns: fast and slow intensification as well as re-intensification patterns.

The sample size used in this work was 16 hurricanes because they are a representative sample of the hurricane population used in this work (150 hurricanes) and was determined using the concepts of minimizing the sampling errors as shown in Appendix B. Table 4.1 shows the results from the sampling assessing procedure.

The selection criterion of the hurricanes included in the testing sample was based in two aspects. First, hurricanes that have happened in the last few years were evaluated to take advantages of the climatology of the hurricanes stored in the developed hurricane database. Second, hurricanes from the season 96-97 were used to compare the results from the proposed intensity model with those obtained by existing intensity models.

Several experiments were conducted to evaluate the ability of the model to deal with different kind of hurricanes. Strong and typical hurricane cases are first presented, followed by a hurricane with high rate of intensification. Next, the rapid intensity reduction rate is presented and a hurricane with re-intensification behavior is presented at the end. The statistics for the hurricane sample are finally summarized.

Table 4.1 Hurricane sample used to test the proposed prediction model

No	Hurricane Name	Year	Maximum Intensity(kt)	Saffir Simpson Scale
1	Isabel	2003	145	Category 5
2	Isidore	2002	110	Category 3
3	Lili	2002	125	Category 4
4	Kyle	2002	75	Category 1
5	Erin	2001	105	Category 3
6	Felix	2001	100	Category 3
7	Michelle	2001	120	Category 4
8	Ericka	1997	110	Category 3
9	Danny	1997	70	Category 1
10	Isidore	1996	100	Category 3
11	Frank	1996	105	Category 3
12	Lili	1996	100	Category 3
13	Hortense	1996	120	Category 4
14	Marco	1996	65	Category 1
15	Edouard	1996	125	Category 4
16	Bertha	1996	100	Category 3

4.1 Experiment with Strong Hurricane Intensity

An important aspect for any intensity prediction model is the capability of modeling hurricanes that reaches the strongest category (category 4 or 5). The proposed model was tested using a strong hurricane, hurricane Lili (2002) ,which crossed western Cuba as a category two and reached category four on October 3, 2002 while it was over the Gulf of Mexico and approaching to wind speed of 125 knots. It made landfall on the Louisiana coast as a category one. Figure 4.1 shows the official intensity (dotted line) for hurricane Lili given by the National Hurricane Center (NHC) and the forecast intensity at 12 hours (continuous line) obtained by the proposed model. The average absolute prediction error at 12 hours interval was 9.7 knots and it was computed along of the storm.

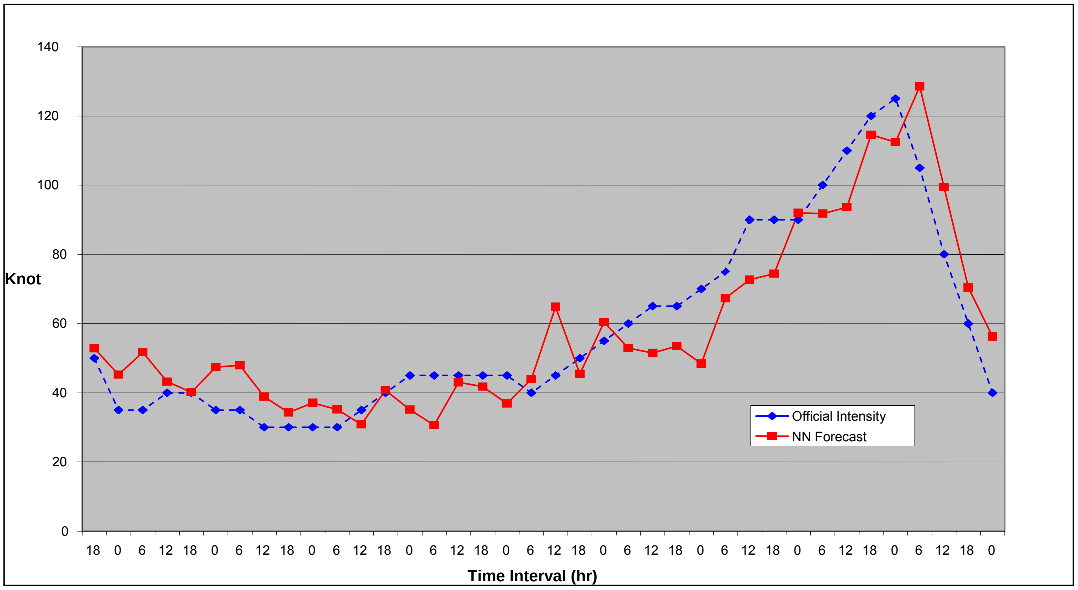


Figure 4.1 Intensity predictions for hurricane Lili at 12 hours (October, 2002)

Figure 4.2 shows the model fitting performance when the prediction interval was 24 hours for hurricane Lili. The average absolute prediction error was 12.5 knots. It should be noted that the larger the prediction interval the large the prediction error. This simple experiment shows that the suggested prediction model is capable of representing the intensity of strong hurricanes.

4.2 Experiment with Typical Hurricane Intensity

The majority of the hurricanes that have occurred on the North Atlantic Ocean can be classified as typical hurricanes, because of the hurricane intensity level. Hurricane Felix (2001) falls into this category. This hurricane remained over the open waters of the eastern Atlantic Ocean, but briefly threatened the Azores Islands. The average absolute prediction error at 12 hours interval was 5.62 knots, which was computed along the entire storm track and using the intensity measured by the NHC as the observed values.

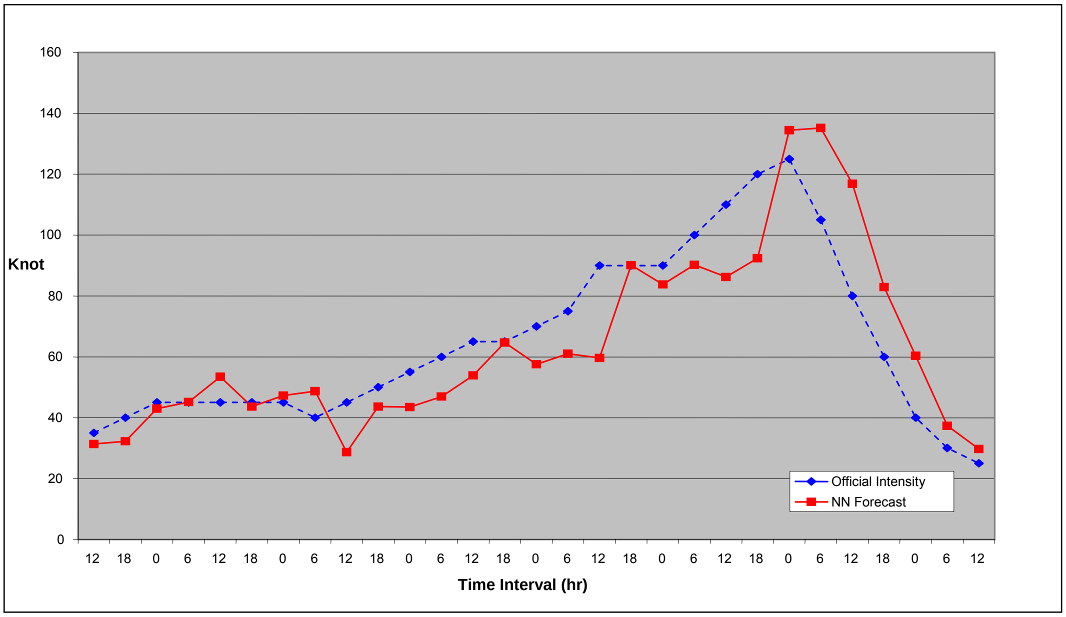


Figure 4.2 Intensity predictions for hurricane Lili at 24 hours (October, 2002)

Figure 4.3 shows the outputs for hurricane Felix (2001) at 12 hours. It was a hurricane that reached its peak at 100 knot and was categorized as hurricane of category three. The average absolute intensity error at 12 hours interval prediction for this hurricane using the proposed intensity model was 5. 62 knots along of the storm.

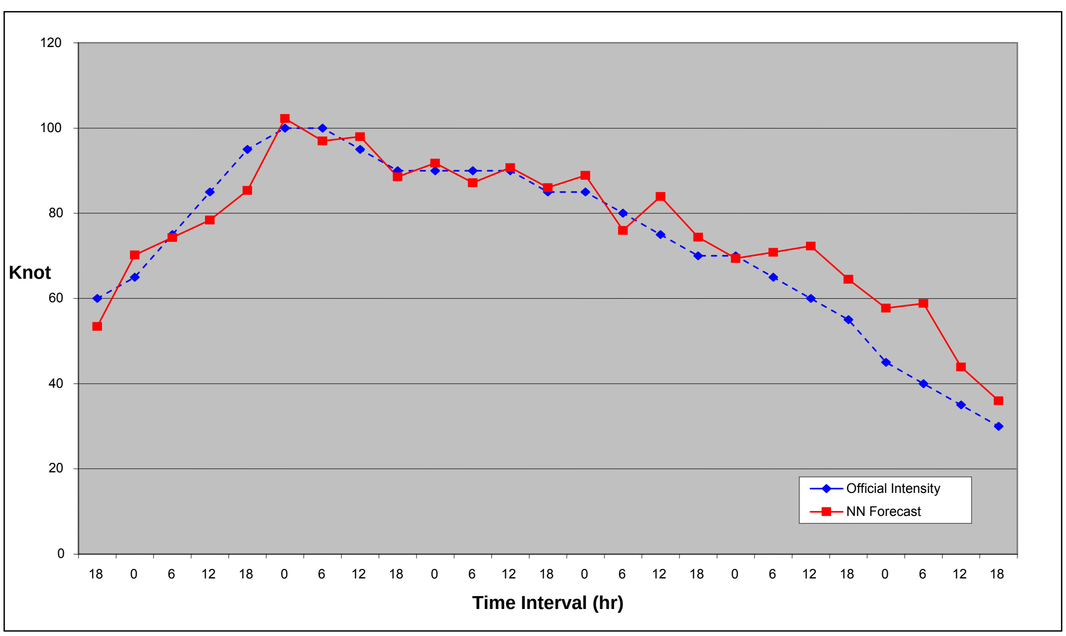


Figure 4.3 Intensity predictions for hurricane Felix at 12 hours (September, 2001)

The proposed model was also evaluated with a typical hurricane at 24 hours. Figure 4.4 shows the results. The average absolute intensity error at 24 hours interval prediction for this hurricane using the proposed intensity model was 7.88 knots along of the storm.

It can be seen that the proposed algorithm has the potential to predict better the intensity of a typical hurricane (category 2 or 3 in Saffir-Simpson scale) than the intensity of a strong hurricane. However, the results obtained (12.5 knot) when a strong hurricane (Lili, 2002) was tested at 24 hours is in its average less than the results obtained by SHIPS which was 16 knots. (http://www.nhc.noaa.gov/2002lili.shtml)

4.3 Fast Intensity Change and Re-Intensification Experiments

The characteristics that are most difficult to deal with in the prediction of the hurricane intensity are the fast intensification, rapid reduction and re-intensification. The proposed model was tested with hurricanes that exhibit at least one of these conditions. The hurricane Hortense (1996) crossed the southwestern region of Puerto Rico and the eastern top of the Dominican Republic under category one, and the associated floods killed at least 21 people. Hortense which reached category four status with a peak intensity of 120 knots was used to implement the model when a rapid intensification was presented. Figure 4.5 shows the results for hurricane Hortense at 12 hours; the dotted line illustrates the rapid intensification forecast. The average absolute intensity error for this hurricane was 8.15 knots at 12 hours interval prediction along of the storm.

It can be seen that the outputs from proposed model followed the official outputs up to the point where Hortense increases rapidly its intensity. Despite of the errors in this period seemed to be larger (10 -15 knot) for 12 hours, the proposed model tried to follow the observed intensity but it failed to make a robust intensity estimation. After Hortense has reached its intensity peak, the proposed model made a suitable prediction with small errors(less than 7 knot).

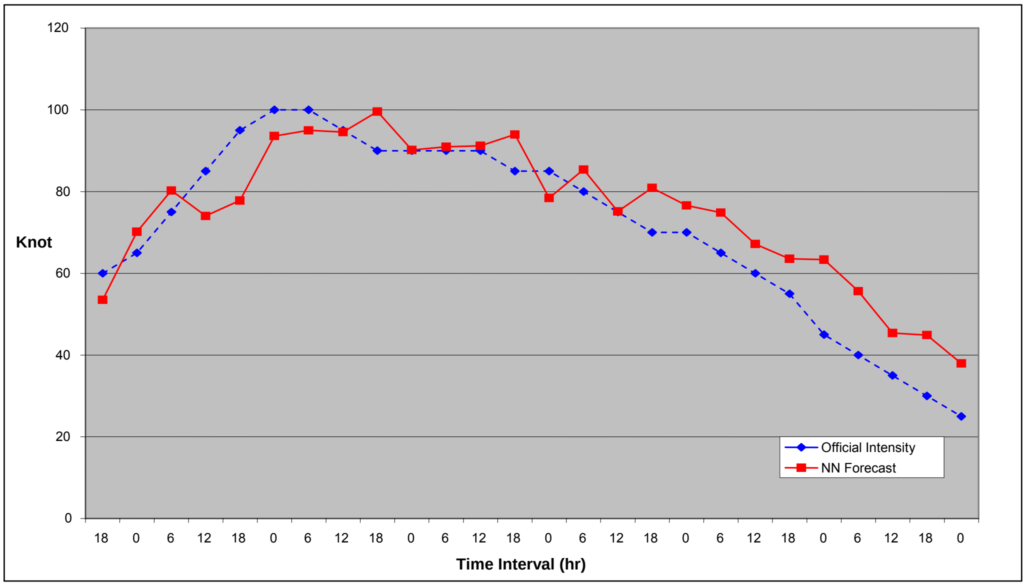


Figure 4.4 Intensity predictions for hurricane Felix at 24 hours (September, 2001)

The rapid reduction is also another attribute that is difficult to predict for any intensity hurricane model. The forecasting for hurricane Frank (1996) which was originated in Cape Verde was a hurricane that moved across the Atlantic during the peak of the hurricane season. This hurricane at 12 hours prediction interval was used to test the prediction methodology under fast intensity reduction.

The hurricane Frank was evaluated because after it reached its intensity peak; it underwent to a fast intensity changing from 100 knots to 35 knots in 18 hours. Figure 4.6 shows the results, dotted line illustrates the period of fast intensity reduction. The average absolute intensity error of this hurricane was 7.8 knots when it was evaluated along of its trajectory using 12 hours as an interval prediction.

The proposed model up to the moment of the fast intensity reduction has been forecasted with average errors of 4 knot, but the effects of this phenomenon made that the final average error went up to almost 8 knot. Again, the skill of the model was affected by especial condition that was not presented in all hurricanes but when it happened, it made the proposed model lose some of its capabilities.

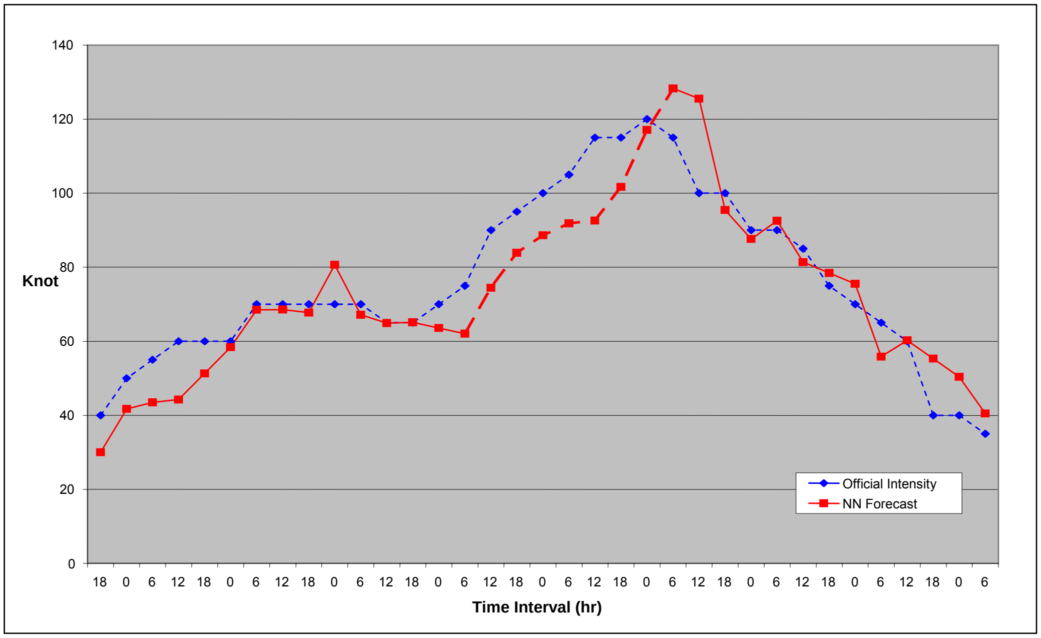


Figure 4.5 Forecasting for Hortense when rapid intensification is presented (September, 1996)

A characteristic that is also hard to predict is the re-intensification. This behavior challenges the model to response as soon as it is detected. Re-intensification means that the hurricane has gained (lost) enough strength to rise (decrease) its intensity. Hurricane Isidore (2002) was a slow-moving tropical cyclone that hit the northern Yucatan Peninsula of Mexico as a category 3 and it was used to test the proposed model when the re-intensification was presented. The re-intensification was probably the most hazardous of the three conditions before mentioned because it challenges the model to look out the new hurricane intensity behavior that could be increasing or decreasing.

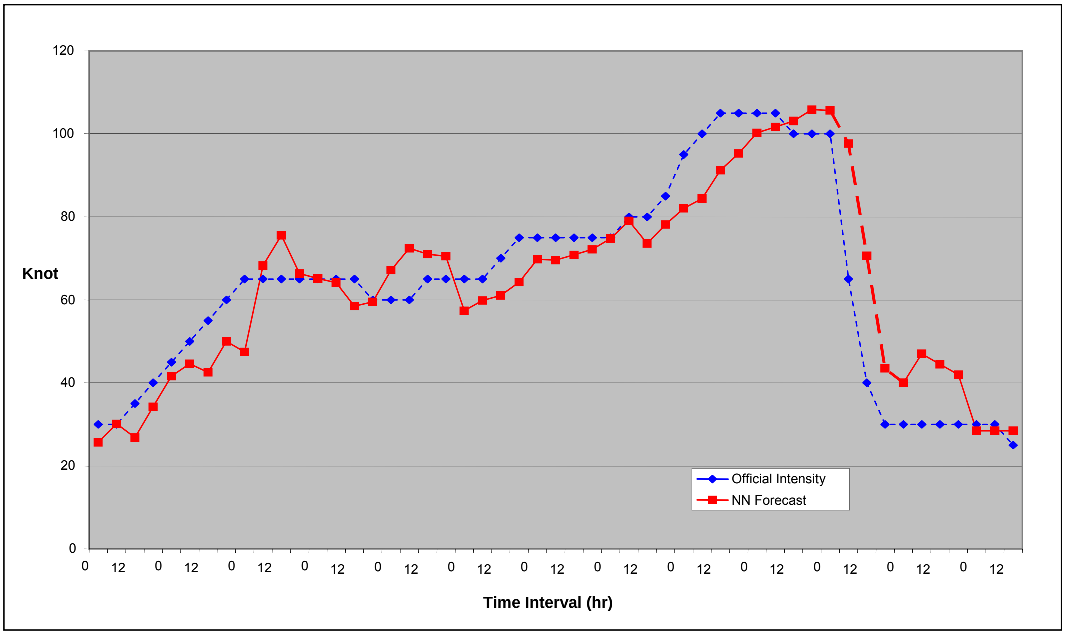


Figure 4.6 Forecasting for Frank when rapid reduction is presented (September, 1996)

The dotted line in figure 4.7 show the re-intensification process obtained from NHC and continuous line the forecast obtained by using the suggested prediction methodology.

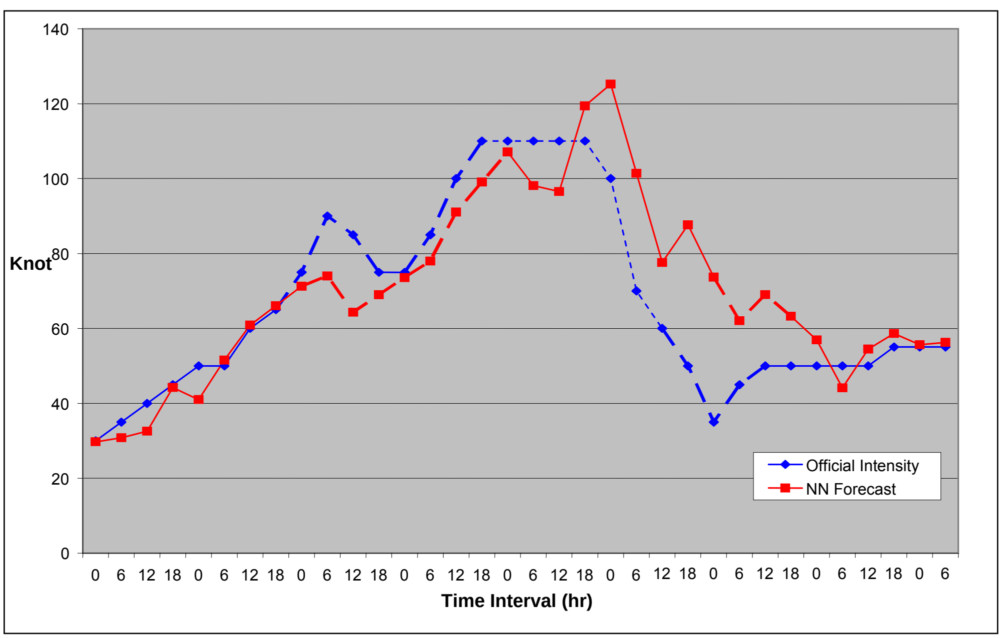


Figure 4.7 Re-intensification of Hurricane Isidore (September, 2002)

The average absolute intensity error for hurricane Isidore was 10.51 knots at 12 hours interval prediction. These results show that the model underestimated the hurricane intensity, i.e, the model does not have the same capability to predict the rapid changes in intensity. Future work may be concentrated the efforts to predict fast intensity changes.

4.4 Summary of Results

The average intensity prediction errors at 12 hours for each one of the hurricanes that composed of the selected sample to test the proposed intensity model are shown in table 4.2.The number of cases used to forecast is also shown in this table.

Table 4.2 Average absolute prediction error at 12 hours for the hurricane sample

No	Hurricane Name	No Cases	Average error (kt)
1	Isabel	44	7.35
2	Isidore	34	10.59
3	Lili	39	9.7
4	Kyle	28	5.86
5	Erin	30	7.5
6	Felix	26	5.62
7	Michelle	19	9.32
8	Ericka	53	5.49
9	Danny	35	4.67
10	Isidore	21	6.03
11	Frank	51	7.8
12	Lili	45	5.46
13	Hortense	35	8.15
14	Marco	30	8.02
15	Edouard	54	6.42
16	Bertha	42	7.27
Number of Cases and Total Avg. Error		**588**	**7.2**

Similarly, the average intensity prediction errors for 24 hours of prediction interval are shown in table 4.3.

Table 4.3 Average absolute prediction error at 24 hours for the hurricane sample

No	Hurricane Name	No Cases	Average Error(kt)
1	Isabel	44	12.46
2	Isidore	32	17.3
3	Lili	31	12.5
4	Kyle	23	11.8
5	Erin	26	10.32
6	Felix	24	7.88
7	Michelle	16	10.44
8	Ericka	50	9.27
9	Danny	34	9.05
10	Isidore	17	12.94
11	Frank	49	12.4
12	Lili	44	9.48
13	Hortense	33	11.37
14	Marco	26	11.64
15	Edouard	50	9.4
16	Bertha	39	10.81
Number of Cases and Total Avg. Error		**539**	**11.19**

The contribution of the variables (predictors) used in the intensity forecast of the hurricane sample (table 4.1) at 12 hours is shown in figure 4.7. The predictors are described in table 4.4.

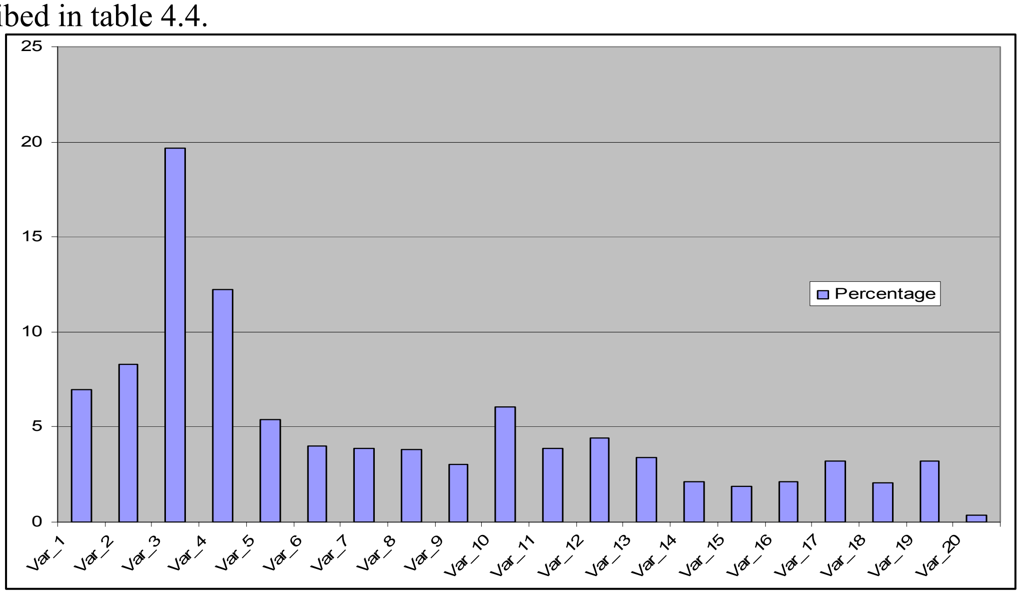


Figure 4.7 Variable contribution for the hurricane sample at 12 hours

Table 4.4 Variable contribution for the hurricane sample at 12 hours

Variable	Name	Percentage
Var_1	Storm Location Latitude	6.956
Var_2	Storm Location Longitude	8.296
Var_3	Storm Pressure	19.703
Var_4	Storm Intensity Change	12.248
Var_5	Storm Northward Displacement	5.381
Var_6	Storm Eastward displacement	4.000
Var_7	Module of the Storm Motion	3.862
Var_8	Sea Surface Temperature	3.841
Var_9	Maximum Possible Intensity	3.011
Var_10	U850	6.027
Var_11	V800	3.894
Var_12	U200	4.395
Var_13	V200	3.416
Var_14	Vertical Wind Shear(VWS)	2.127
Var_15	Average angular momentum at 850 mb	1.857
Var_16	Average angular momentum at 200 mb	2.09
Var_17	Total Total	3.237
Var_18	Kindex	2.054
Var_19	MPI-Initial Intensity	3.218
Var_20	VWS Change	0.385

The contribution of the variables (regressors) used in the intensity forecast of the hurricane sample (table 4.1) at 24 hours is shown in figure 4.8 and is described in table 4.5.

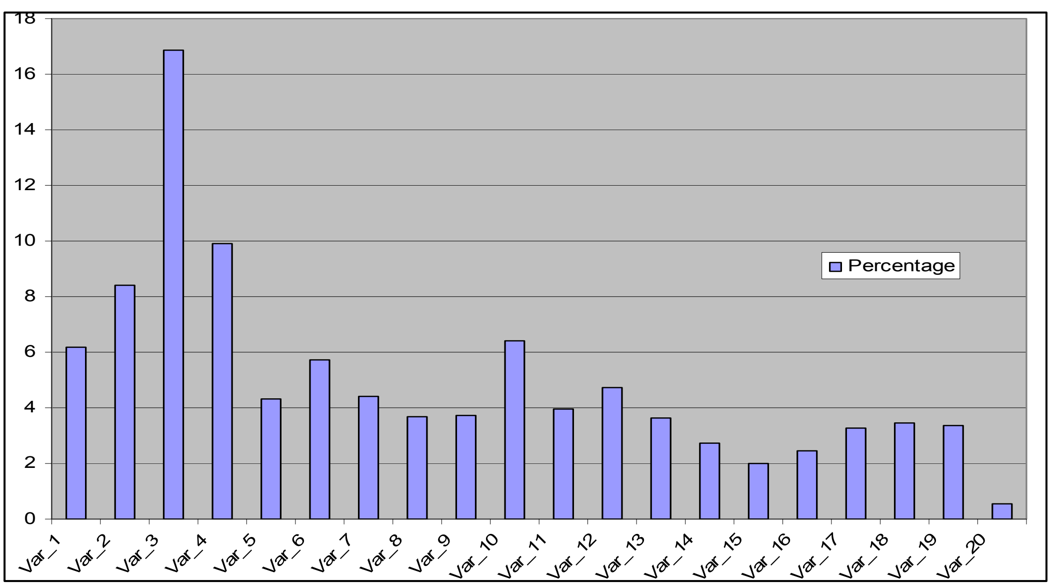


Figure 4.8 Variable contribution for the hurricane sample at 24 hours

Table 4.5 Variable contribution for the hurricane sample at 24 hours

Variable	Name	Percentage
Var_1	Storm Location Latitude	6.202
Var_2	Storm Location Longitude	8.411
Var_3	Storm Pressure	16.868
Var_4	Storm Intensity Change	9.911
Var_5	Storm Northward Displacement	4.302
Var_6	Storm Eastward displacement	5.713
Var_7	Module of the Storm Motion	4.407
Var_8	Sea Surface Temperature	3.675
Var_9	Maximum Possible Intensity	3.705
Var_10	U850	6.405
Var_11	V800	3.972
Var_12	U200	4.728
Var_13	V200	3.640
Var_14	Vertical Wind Shear(VWS)	2.743
Var_15	Average angular momentum at 850 mb	1.99
Var_16	Average angular momentum at 200 mb	2.46
Var_17	Total Total	3.277
Var_18	Kindex	3.476
Var_19	MPI-Initial Intensity	3.376
Var_20	VWS Change	0.549

The most important variables in this work were: the storm pressure, the intensity change, the storm location (longitude and latitude) and the eastward component of wind speed at 850 mb. The contributions of the variables used in this study are shown in table 4.4 and 4.5.

The storm pressure variable measured in the hurricane's eye has been the most important predictor in this work. This result is not surprised since the relation between hurricane intensity and hurricane pressure is directly proportional. The storm latitude and longitude variables have also proved their advantages to explain the hurricane intensity. These results can be explained using the idea that the smaller the central pressure, the greater winds that surrounded the hurricane.

The intensity change is the second important variable that explains the hurricane intensity for the proposed model. This result comes in agreement with the results found by DeMaria (1997) who pointed out that the intensity change provides a pattern of future hurricane's behavior which means that hurricane that has intensified in the past 6 hours is likely to continue intensifying.

It has been pointed out that the sea surface temperature (SST) is an important predictor in the hurricane intensity models (Merrill, 1987; etc.) but alone it only gives an upper bound on the hurricane intensity. However, SST in this study is a variable that exhibits a marginal contribution in the intensity model. The causes of this performance could be the explanation given by Merrill or the lack of real time SST because the SST used in this study was obtained from historical records and was interpolated in space and time for the current hurricane. The next predictor called maximum possible intensity which is a function of SST exhibits also marginal contribution. This results is probably caused by the lack of real time observations of sea surface temperature.

The contribution of different mathematical transformations described in section3.5 and used in the prediction intensity model is described in table 4.6.

Table 4.6 Model contribution for the intensity prediction

Model	Model Contribution at 12 hours (%)	Model Contribution at 24 hours (%)
Linear	62.378	55.522
Logarithm	11.699	13.406
Quadratic	6.396	9.945
Inverse	14.283	15.271
Cubic	5.243	5.855

Results have shown that the performance of the proposed model for different types of hurricanes. The performance of the model cannot be compared with existing models because of limitation of published results. However, a comparison with small sample size was conducted. Table 4.8 was developed using hurricanes from the season 96-97 (Table 4.7) and the information published by DeMaria (http://www.nhc.noaa.gov/aboutmodels.shtml). Table 4.8 shows the comparison between the performance of the proposed model (NN Model) and the models used by the National Hurricane Center during the season 96-97, in this case the Statistical Hurricane Intensity Forecast (SHIFOR), the Statistical Hurricane Intensity Prediction Scheme (SHIPS) and the Geophysical Fluid Dynamic Intensity (GFDI).

Table 4.7 Hurricanes used to compare the proposed model and NHC's models

No	Hurricane Name	Year	Maximum Intensity	Category
1	Ericka	1997	110	Category 3
2	Danny	1997	70	Category 1
3	Isidore	1996	100	Category 3
4	Frank	1996	105	Category 3
5	Lili	1996	100	Category 3
6	Hortense	1996	120	Category 4
7	Marco	1996	65	Category 1
8	Edouard	1996	125	Category 4
9	Bertha	1996	100	Category 3

Table 4.8 Comparison between the proposed model and NHC's models

		12 Hours		24 Hours	
Model	Season	No Cases	Average error (kt)	No Cases	Average error(kt)
NN Model	96-97	366	6.59	343	10.71
SHIFOR	96-97	305	8.2	270	11.4
SHIPS	96-97	305	8.1	270	11
GFDI	96-97	305	9.3	270	11.6

Table 4.9 shows the relative improvement by the proposed model over the official NHC's models. It noticed that the proposed model achieved a considerable improvement at 12 hours interval over the other models used by NHC but the improvement is not the same at 24 hours interval.

Table 4.9 Proposed models' improvement over NHC's models

Model	At 12 hours (%)	At 24 hours (%)
SHIFOR	19.63	6.05
SHIPS	18.64	2.64
GFDI	29.14	7.67

Figure 4.9 shows the fitness of the NN model (small dotted line) over the multiple linear regression model (continuous line) when both models were tested using the same variables to predict the hurricane intensity. Also, the official forecast is showed (dotted line). This figure shows that the NN model has a better behavior in the prediction stage than the Regression model which is also the same results obtained by Baik et al. (1998).

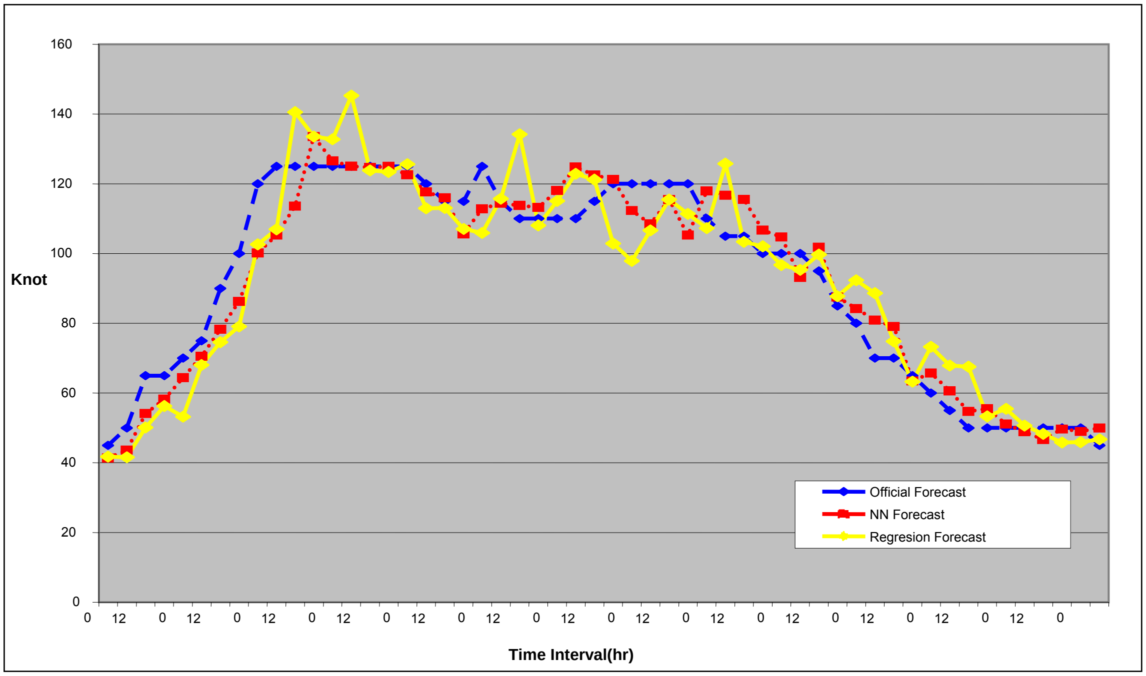


Figure 4.9 NN model enhancements over regression model for hurricane Edouard at 12 hours (September, 1996)

Table 4.10 displays the average absolute errors obtained by the proposed NN model and those obtained by the multiple linear regression model. Furthermore, the percentages of this improvement are shown.

Table 4.10 Proposed model improvements over multiple linear regression model

		At 12 Hours		At 24 Hours	
Model	**Season**	**No Cases**	**Average error (kt)**	**No Cases**	**Average error (kt)**
NN Model	96-97	366	6.59	343	10.71
Regression Model	96-97	366	8.82	343	12.53
NN improvement over Regression (%)			**25.3**		**14.53**

A 95 % confidence interval was developed to take into account the natural variation of the neural network's outputs at the prediction stage. The intensity prediction for each point in time was calculated 100 times varying the initial point in the neural network training. The cumulative distribution of the prediction sample was fitted to the closest distribution function using as a measure of fitness the Kolmogorov – Smirnov test. Once, the distribution function that best adjust the sample under study was detected, an interval of confidence was calculated using the parameters that correspond for this adjusted distribution function. Hurricane Isaac (Sept, 2000) was used to implement this concept. Isaac was a hurricane that was formed over Cape Verde and developed a long, parabolic path over the eastern half of the Atlantic. Its maximum sustained winds reached an estimated 120 knot. The Kolmogorov-Smirnov test was used to know the distribution function that best fit with the cumulative distribution of the prediction sample. Figure 4.10 shows the confidence interval implemented for the intensity prediction of hurricane Isaac at 12 hours. The complete procedure to calculate the confidence interval for a given hurricane intensity prediction is given in Appendix E.

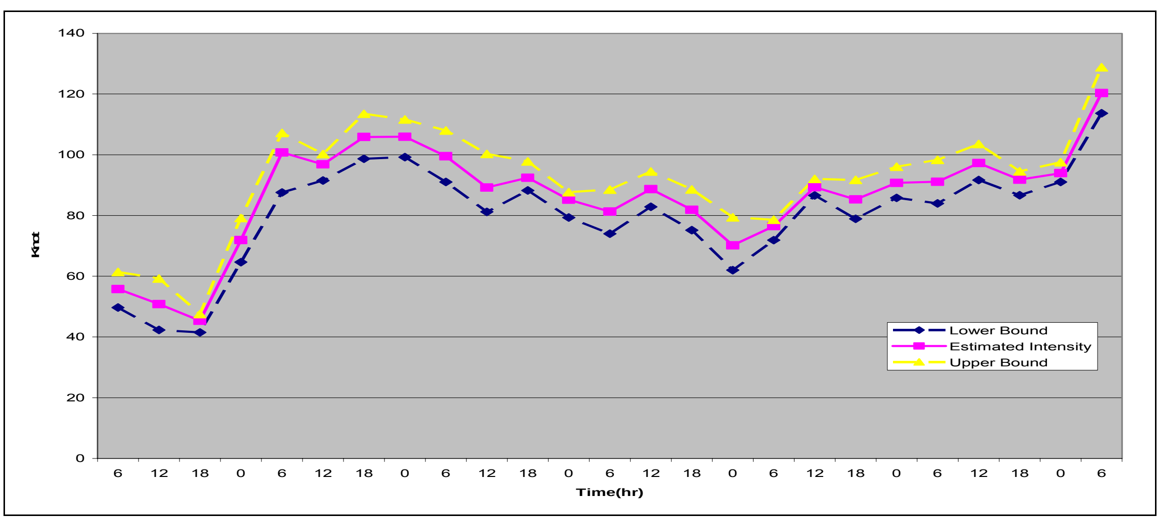


Figure 4.10 Confidence Interval for hurricane Isaac (Sept, 2000)

CHAPTER V CONCLUSIONS AND RECOMMENDATIONS

The key contribution of this study is the development of a hurricane intensity prediction model for the North Atlantic basin. A combination of statistical and artificial neural networks techniques is used to predict the intensity at 6, 12, 18 and 24 hours. This type of approach has not been addressed before in the literature to solve this problem.

Preliminary results were derived on the basis of studying a representative sample and these results show that the proposed intensity model has a potential skill to reduce the hurricane prediction errors for a lead time of 12 and 24 hours. The ability of the model to predict the intensity reduces when the hurricane is classified as a strong hurricane (figure 4.1 and figure 4.2). This lack of prediction capabilities is probably caused by the absence of a variable that can anticipate the fast changes on intensity in a period of time greater than 12 hours.

Atlantic Hurricanes for the season 96-97 were used to compare the proposed intensity model with other intensity models. The results (table 4.7) showed that the reduction of the intensity errors at 12 hours was in the order of 1.6 knot which represents 18 % of improvement over the NHC' model that had the minimum intensity error (table 4.8). In 24 hours prediction interval, the improvement was in the order of 0.29 knot which is an improvement of 6%.

In this study, a number of the best subsets formed by the union of the variables (predictors) were generated to relate with the hurricane intensity. Some of the best subsets were linear and others were non-linear. The contribution of the linear and non-linear best subsets in the intensity hurricane prediction (table 4.6) is evaluated. At 12 hours prediction interval, it can be noted that the intensity model had basically a linear behavior.

The non-linear transformations (logarithm, quadratic and inverse) were used with less frequency. The results changed when the prediction interval was 24 hours; the logarithm transformation has increased its participation in 13 %, the quadratic transformation has also increased its participation in 35 % and the linear participation has decreased by 10%. This maybe means that the hurricane intensity at intervals of equal and less of 12 hours could have a linear behavior and probably just linear predictors were needed to predict the hurricane intensity. However, when the prediction interval is greater or equal than 24 hours, the hurricane intensity behavior must include non-linear predictors, since the linear predictors only explained 50 % of the hurricane intensity variability.

The improvement of the neural network technique to model the hurricane intensity over multiple linear regression, which is the technique used in the previous intensity studies, was assessed in this work. The results (table 4.10) showed that the errors at 12 hours were reduced in 25 % and the errors at 24 hours were reduced in almost 15 % which clearly suggested the potential of the neural network over the multiple regression technique to model this kind of meteorological phenomenon.

The importance of synoptic information has been proved in this work. The eastward component of wind speed at 850 mb is among the most relevant variables for the proposed intensity model. This result confirmed the tendency pursued by the newest hurricane intensity models (SHIPS, GFDI, etc.) which have increased their attention in synoptic information in the last few years.

An important aspect in this work is the selection of the analog hurricanes for a given tropical storm. This study used the competitive neural network as the technique to perform this selection. In the future, a different unsupervised and supervised technique may be used to improve the classification algorithm. Supervised techniques based on a preliminary classification may be an alternative to improve the analog identification procedure. One of the most relevant supervised techniques that can be used in the future is the Learning Vector Quantization (LVQ)

The addition of new predictors is definitely a work that must be done in the future. In this study, the capability of AMSU to detect the hurricane and to relate its brightness temperature in some channels with hurricane intensity has been proved. The next task to do is the inclusion of this data to the intensity model. Moreover, AMSU data can be used to correlate with NCEP Reanalysis Data and to obtain temperature at different pressure levels that can be used to calculate the Kindex and Total Total. The resolution at AMSU data is 48 km in the horizontal and the NCEP is 276.43 km.

The entire computational code for the proposed intensity model was developed in Matlab to take advantage of its toolboxes which contains predefined routines that make the programming more pleasant. A disadvantage of dealing with Matlab was the computational time that takes to run a program which used so many routines. To overcome this problem, two servers with double processors Xeon and one Gigabyte of memory RAM were used to run the program.

REFERENCES

Baik, J. and H. Hwang, 1998: Tropical Cyclone Intensity Prediction Using Regression Method and Neural Network, *Journal of the Meteorological Society of Japan*, 5, 711-717.

DeMaria, M and J. Kaplan, 1993: Sea surface temperature and the maximum intensity of Atlantic tropical cyclones. *Journal of Climate* , 7, 1324-1334.

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

DeMaria, M and J. Kaplan, 1999: An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin and eastern north Pacific basins. *Weather and Forecasting*, 14, 326-337.

DeMaria, M., M. Mainelli and L. Shay, 2003: Improvements in real-time statistical tropical cyclone intensity forecasts using satellite data. *83rd American meteorological society annual meeting*.

Gray, W., Scheafer, J., and C. Landsea, 1997: Climate trends associated with multidecadal variability of Atlantic hurricane activity. *Hurricanes, Climate and Socioeconomic Impacts*, 15-53.

Gutowski, W., McMahon, G., Schluchter, S., and P. Kirshen, 1994: Effects of Global Warming on Hurricane - Induced Flooding, *Journal of Water Resources Planning and Management*, 120, 176-185.

Hagan, M., and M. Menhaj, 1994: Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks,* 5, 6.

Hagan, M., H. Demuth and M. Beale, 1996: Neural network design. PWS Publishing Co., Boston.

Hope, J., and C. Neumann, 1970: An operational technique for relating the movement of existing tropical cyclones to past tracks. *Mon. Wea. Rev.*, 98, 925-933.

Hornik, K., M. Stinchcombe and H. White, 1989: Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-366.

Jarvinen, B. R. and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity (SHIFOR). *NOAA Tech. Memo. NWS NHC*, 10, 22.

Jarvinen, B. R. and M. Davis, 1984: A tropical cyclone data tape for the North Atlantic basin, 1886-1983: contents, limitations and uses. *NOAA Tech. Memo. NWS NHC-22*.

Knaff, J. M. DeMaria and C. Sampson, 2004: An introduction to the statistical typhoon intensity prediction scheme (STIPS). *26th Conference on hurricanes and tropical meteorology*.

Kidder, S. and M. Goldberg, 2000: Satellite analysis of tropical cyclones using the Advanced Microwave Sounding Unit (AMSU). *Bulletin of the American Meteorological Society*, 81, 1241-1259.

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

Lansea, C., 1992: A climatology of intense (or major) Atlantic hurricanes. *Mon.Wea. Rev.*, 121, 1703-1713.

Merrill, R. T., 1987: An experiment in statistical prediction of tropical cyclones intensity change. *NOAA Tech. Memo. NWS NHC-34.*

Mo, T., 1999: AMSU-A antenna pattern corrections. IEEE Trans. Geosci. Remote Sens., 37,103-112.

Montgomery, D.C., E.A. Peck and G. C. Vining, 2001: Introduction to linear regression analysis, 3erd Ed., John Wiley & Sons, New York.

Neumann, C. J., 1972: An alternate to the HURRAN tropical cyclone forecast system. *NOAA Tech. Memo. NWS SR-62*.

Neumann, C. J., 1985: Trends in forecasting the tracks of Atlantic tropical cyclones: An update through 1984. *Preprints, Ninth Conf. on Probability and Statistics in Atmospheric Science, Amer. Meteor*. Soc. 109-115.

Pike, A. C., 1985: Geopotential heights and thicknesses as predictors of Atlantic tropical cyclone motion and intensity. *Mon. Wea. Rev*., 113, 931-939.

Ramirez-Beltran, N.D. and H. Rodriguez, 2000: A Neural Network Approach to Estimate the Chemical Activity Coefficients. *Proceedings of Group Technology and Cellular Manufacturing*, World Symposium 349-354.

Ramirez-Beltran, N., Winter, A., Chong, F., Veneros, A., 2003: Neural networks estimate atmospheric variables in the Caribbean basin. *83rd American meteorological society annual meeting-Third Conference on Artificial Intelligence.*

Sheets, R., 1990: The National Hurricane Center-Past, present and future. *Wea. Forecasting*, 5, 185-232.

Tangang, F., Tang, B., and W. Hsieh, 1998: Forecasting ENSO Events: A Neural Network-Extended EOF Approach., *Journal of Climate*, 11, 29-41.

Wark, D. Q., 1993: Adjustments of TIROS Operational Vertical Sounder data to a vertical view. NOAA Tech. Rep. NESDIS 64, Washington, DC, 36 pp.

APPENDIX A

COMPUTER PROGRAM DESCRIPTION

A computer program has been developed to predict hurricane intensity at different intervals of time. The Matlab environment is used in order to take advantage of its toolboxes, especially designed to deal with statistical and artificial intelligence problems. The computer program is divided into two major modules:

A.1 The Hurricane Database

A graphical user interface (GUI) was implemented to generate a hurricane database to provide enough data for the hurricane intensity model so that a robust estimation can be made. This feature implies that the data is always updated. Another reason to develop the GUI was to let the user had a friendly communication with the machine and avoided input data mistakes. The database is created using several sources of information as shown in Figure 4.1. The complete description of the sources of information for the hurricane database is given in Section 3.3.

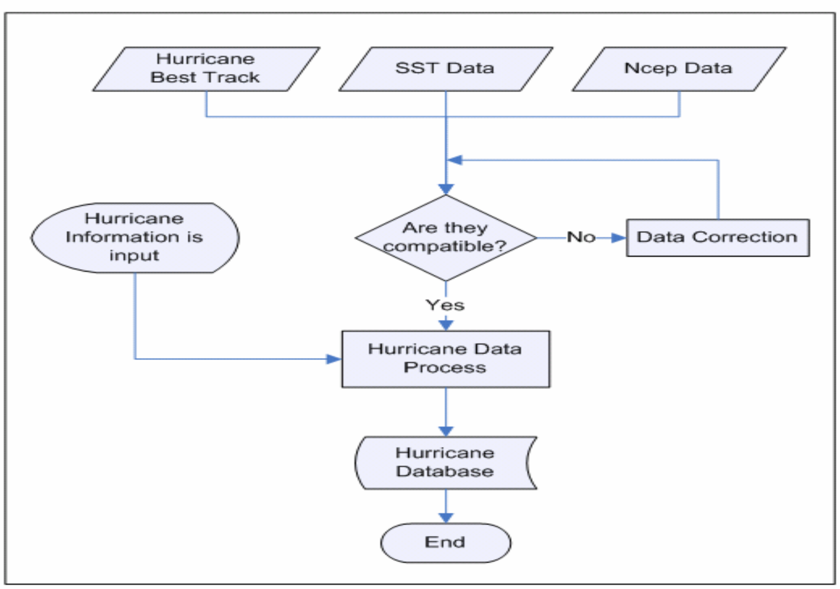


Figure A.1 Information used to created the hurricane database

Recall that the hurricane best track stored the hurricane trajectory and others parameters such as pressure and time. The Sea Surface Temperature (SST) data is used to obtain the SST along the location of the storm. The NCEP data is used to generate the synoptic variables utilized in this study.

The hurricane database interface is composed of a list of events that has a defined purpose. Figure 4.2 shows the hierarchical tree of these events:

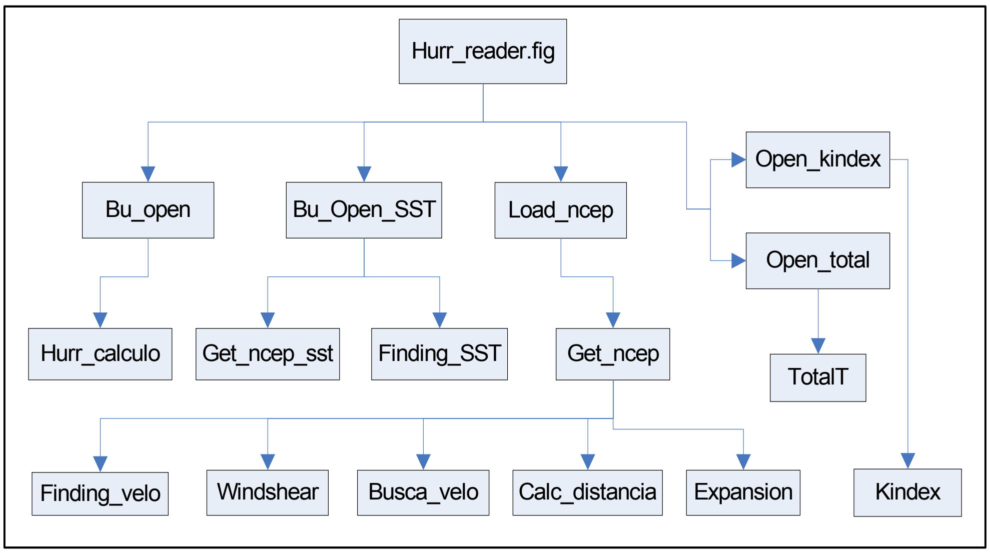


Figure A.2 Relationships among hurricane database's procedures

The main difference between a Matlab script and Matlab graphical user interface(GUI) is that the latter is ruled by events and the first is composed of a number of functions, so that each one of these must be defined as an independent file. However, in a Matlab GUI, all of these events are written in a common file. This feature makes it more attractive for the programmer to work with GUI than the original Matlab script.

The first step to execute the hurricane database GUI is to create a new record clicking the button *New* (event new*)* to invoke the procedure *create_new*. This procedure creates a new record to input a new hurricane. Then, the user must proceed to fill the needed fields manually as shown in figure A.3.After that, the user is ready to call the different events defined in the figure A.2 (*Bu_open, Bu_open_SST, Load_ncep, Open_kindex and Open_total*).

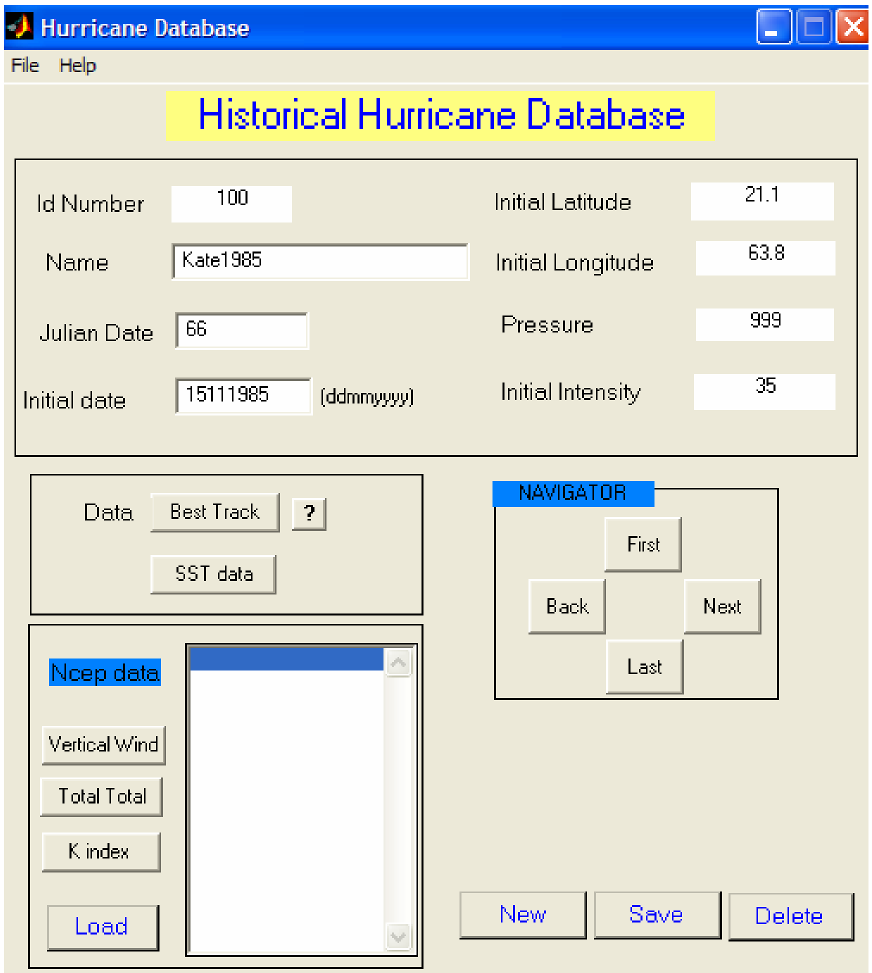


Figure A.3 Hurricane Database Graphical User Interface (GUI)

The event *Bu_open* is invoked when the *Best Track* button is clicked and used to call the hurricane best track previously defined in an ASCII file; this event brings on the procedure *Hurr_calculo* necessary to calculate the differences in hurricane intensity, latitude and longitude. The *SST data button* is used to call the event *Bu_open_SST* which initializes two procedures: *Get_ncep_sst* used to read and convert the Sea Surface Temperature (SST) binary file into ASCII file and *Finding_SST* used to capture the nearest SST for the position and date of each tropical cyclone observation.

The NCEP data are added to the database when the user called *Load_ncep* event clicking the *Vertical Wind* button. This event called the *Get_ncep* procedure which is used read NCEP binary data and convert it into a readable ASCII format. In addition to this action, *Get_ncep* called other procedures such as: *Finding_velo* is needed to get the nearest wind speed at 200 and 800 mb for position and date of each hurricane observation. *Busca_velo* is used to decompose the wind speed at 200 and 800 mb in their eastward and northward component. *Windshear* calculates the Vertical Wind Shear (VWS) using the wind speed components. *Expansion* procedure calculates the momentum at different distances from the hurricane's eye at 200 and 800 mb. *Calc_distancia* procedure is employed to obtain the distances used in the previous procedure.

The *Open_total* event is activated by clicking the *Total Total* button by the user. This event called the *TotalT* procedure which is used to calculate the Total Total index as is defined in section 3.3.2. The event *Open_index* is initialized by clicking the *K index* button and called the *Kindex* procedure which performs a series of calculations to obtain the K index as is defined in section 3.3.2.

Up this point, the fields defined in table 3.3 have been filled or calculated so that the next step is to save these and continue with another hurricane. To do this action, the user must click the *Save* button to assure the process has finished successfully. Is important to notice that the GUI's source code is not published because of the amount of space required but it is available at *a_veneros@yahoo.com*.

A.2 The Hurricane Prediction Framework

The hurricane database was developed using Matlab GUI to assure that no mistakes have done during input data process and let a friendly communication between user and computer. Despite of these advantages, the time consumed to execute a Matlab GUI is usually larger than a Matlab script and when a program consists of a series of repetitive steps is better developed using the later mode. The hurricane prediction model falls into this category because it is made up by a number of predefined routines which are executed as many times as the forecaster needs.

The hurricane intensity model is composed of four main sections as shown in Figure A.3:

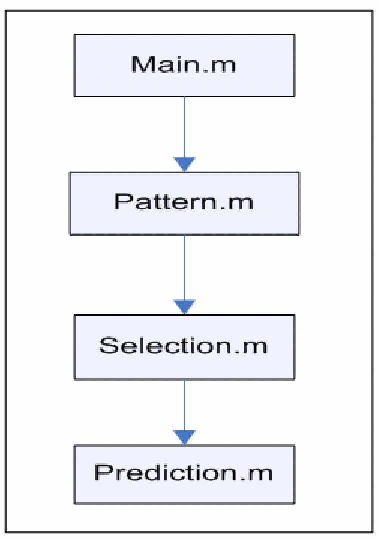


Figure A.3 Hurricane Intensity Model Functions

2.1. Main.m

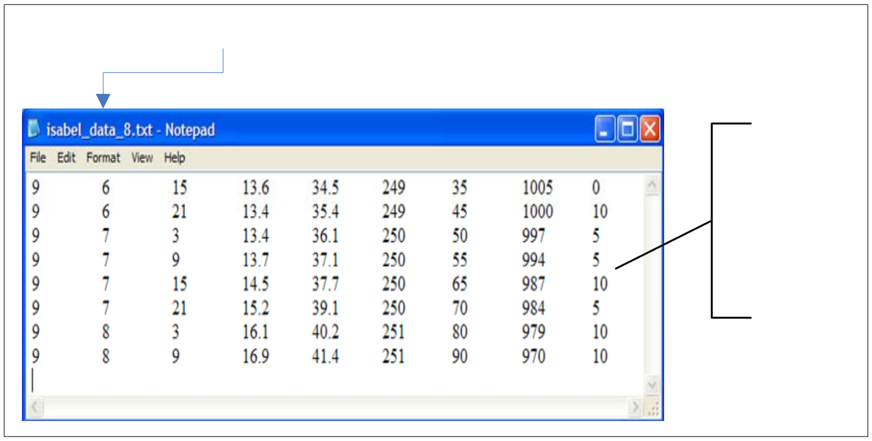
The *Main* program gives the initial settings to run the hurricane intensity model. The tasks of this program are summarized as follows:

1. The hurricane database mentioned in the previous section is loaded. Moreover, initial and final observations are defined to establish the numbers of files used in step 2. Finally the number of replicates, number of classes and number of epochs for the classification process are set.

2. The data of the current hurricane that are used to obtain the analogous hurricanes are loaded in this step. Using the format shown in figure A.4, a set of ASCII files is loaded. These files have a generic name as follows:

   HurricaneName_data_NroObservation.txt

   Where *HurricaneName* is the name of a given hurricane, *data* is a word used to indicate that this file contains hurricane data, *NroObservation* indicates how many observations are presented in this file and *.txt* gives the extension of this file.



9	6	15	13.6	34.5	249	35	1005	0
9	6	21	13.4	35.4	249	45	1000	10
9	7	3	13.4	36.1	250	50	997	5
9	7	9	13.7	37.1	250	55	994	5
9	7	15	14.5	37.7	250	65	987	10
9	7	21	15.2	39.1	250	70	984	5
9	8	3	16.1	40.2	251	80	979	10
9	8	9	16.9	41.4	251	90	970	10

Figure A.4 Data input to initialize the hurricane model

3. After the data have been uploaded, the second routine called Pattern is initialized using the following code:

   For *i=initial_observation* to *final_observation*

   Data=load (HurricaneName_data_*i*.txt);

   For *j=1*:*nro_replicates*

   [*Y, X*]=Pattern (*Data, Hurricane_Database, Nro_Class, Nro_Epochs*)

Hurricane name ir

End

End

where the routine *Pattern* has been executed as many times as the range defined by *initial_observation* and *final_observation* variables. The second For Loop is repeated up the value of *nro_replicates* variable.

2.2. Pattern.m

This function is used to find a set of analogous hurricanes of a current hurricane. The sentence that is employed to call this routine is again repeated here:

[Y, X]=Pattern (*Data, Hurricane_Database, Nro_Class, Nro_Epochs*)

where *Data* store the current hurricane data known until the actual time (*t*), *Hurricane_Database* stores the data of the hurricanes data since 1975, *Nro_Class* is the number of classes to perform the classification using Competitive Neural Network, *Nro_Epochs* is the number of epochs to be used in the Competitive Neural Network

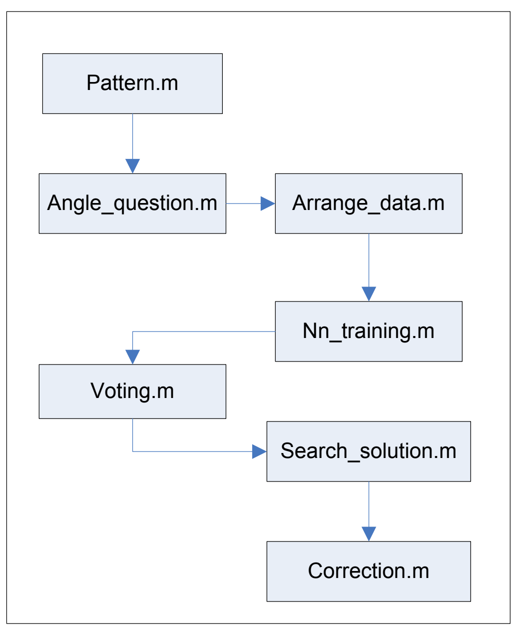


Figure A.5 Pattern Routine and Its Subroutines

This routine is composed of a number of subroutines (Figure A.5) where each one of these has a defined purpose. A brief description of these subroutines is given as follows:

2.2.1. Angle_question.m

Function used to calculate the hurricane's direction given two hurricane's positions. It is based in the calculus of the arc tangent of two distances (d_1, d_2). For two positions of the hurricane's trajectory at time *t* and *t-6* hr, the hurricane's direction from P_{t-6} to P_t is given by α as shown in figure A.6:

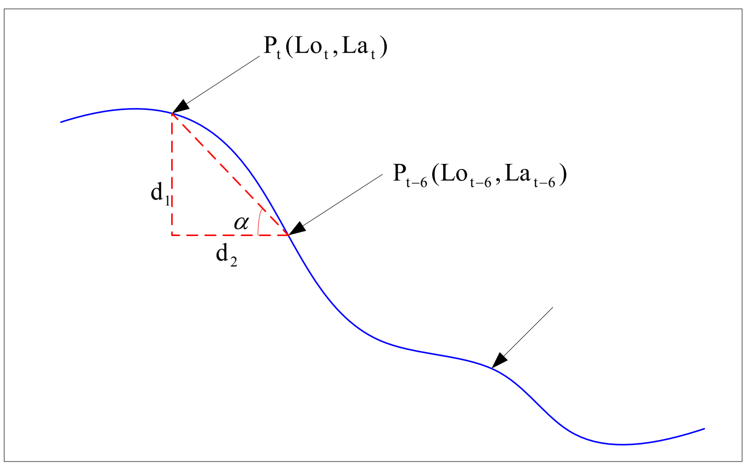


Figure A.6. Hurricane's direction procedure

2.2.3. Arrange_data.m

Function used to sort the historical data and mix with actual data. This function generates the input for the Competitive Neural Network. It also uses as a preliminary filter to reject hurricanes that don not have the same format as the current hurricane data.

2.2.4. Nn_training .m

This is the core part of Pattern routine because it is here where the Neural Network (NN) to classify is created and trained using the input generated by the previous function The Matlab Neural Network Toolbox is used to create the NN with the following code:

Pr = minmax (P);
Net = newc (Pr,*classes*);
Net.trainParam.epochs = *epochs*;
Net = train (Net,P);
Solution = sim (Net,P);

where P is matrix of inputs composed of historic data and actual data, Pr is a matrix that contains a maximum and minimum value of each column of P, Net is the NN created with the Matlab function *newc* using two parameters Pr and the number of classes defined at the beginning of *Pattern* routine. The *train* Matlab function is used to train the Net and the solution is obtained using the Matlab function called *sim*.

2.2.5 Voting.m

This function is used to obtain a unique solution using the majority voting concept explained in section3.4.The output of this function is a row vector where each element of the vector represents a hurricane and its value is the selected class.

2.2.6 Search_solution.m

Function that is called by the previous function to identify the hurricanes that have the same class as the current hurricane, in other words, the analogous hurricanes for the current hurricane.

2.2.7 Correction.m

This function is used to complete the current hurricane data with synoptic variables. These variables are calculated from the past data using a interpolation for each variable in time and space. This procedure is explained at details in section 3.4

2.3. Selection.m

Up to this point, a set of analogous hurricanes have been selected. The *Selection* routine is used to find out a correlation between the variables that composed that set and the hurricane intensity. The multiple regression technique and the stepwise selection model are used to develop this routine and the output of this function is the input of the last routine called *Prediction*. The figure A.7 shows the several sub functions used to this routine.

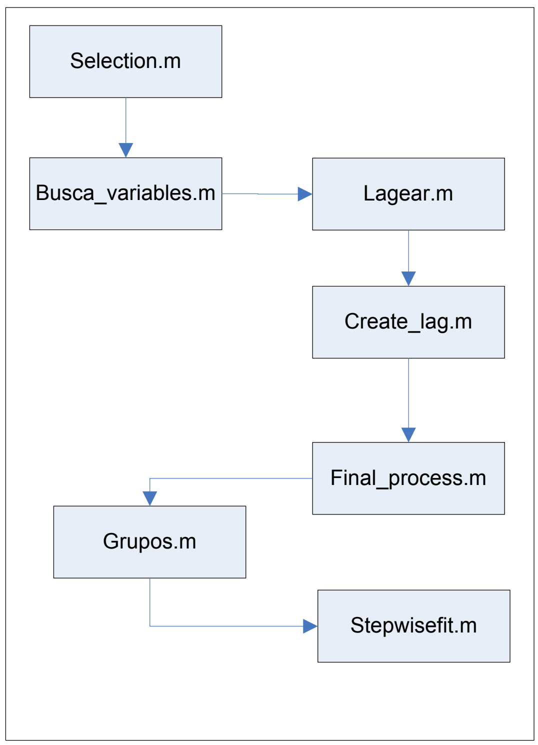


Figure A.7 Sub functions used in *Selection* routine

2.3.1. Busca_variables.m

This function is used to label the variables. This is a necessary task because the selection procedure is randomized which means that the variables are changing their position in the set at each iteration. However, the user must know the correct name and position of each variable at the end of the process.

2.3.2. Create_lag.m

This function is implemented to create the lag between the regressors and the response variable. The lead time is chosen by the user and it depends of the desired prediction time. More information of how the lag works is found in section 3.5.

2.3.3. Final_process.m

This function is applied to set a number of required parameters to run the selection procedure. It is an important function because it stores the function *Grupos* Stepwisefit.

2.3.4. Grupos.m

Function that is called by the preceding function and is used to divide the original set of variables into *n* subsets of *m* variables where *n* and *m* are calculated by the program following the rules described in section 3.5.

2.3.5. Stepwisefit.m

Having the original set of variables divided in *n* subsets of *m* variables, the *Stepwisefit* function uses stepwise regression to model the response variable Y as a function of the predictor variables represented by the columns of the matrix X. The result is a vector *b* of estimated coefficient values for all columns of X. The *b* value for a column not included in the final model is the coefficient that you would obtain by adding that column to the model. This function has the following syntax:

$$[b, se, pval, inmodel, stats] = stepwisefit(X, Y, 'penter', p\_in, 'premove', p\_out, 'iter', 1000)$$

where:

- *se* is a vector of standard errors for *b*.
- *pval* is a vector of p-values for testing whether *b* is 0.

- *inmodel* is a logical vector, whose length equals the number of columns in X, specifying which predictors are in the final model. A 1 in position j indicates that the jth predictor is in the final model; a 0 indicates that the corresponding predictor in not in the final model.
- *stats* is a structure containing additional statistics.
- *penter* is the maximum p-value for a predictor to be added. The default is 0.05.
- *premove* is the minimum p-value for a predictor to be removed. The default is 0.10.
- *iter* is the maximum number of steps to take (default is no maximum)

Up to this point, the stepwisefit has selected the best subset to explain the hurricane intensity. The next step is predicted the hurricane intensity using Artificial Neural Network (ANN).

2.4. Prediction.m

The best subset of regressors obtained in the previous function is used as an input for this procedure. An ANN is created and trained using Matlab functions and the intensity prediction values are saved in ASCII file. However, a number of necessary steps must be accomplished in order to obtain a reliable solution. Figure A.8 shows the required routines for this procedure.

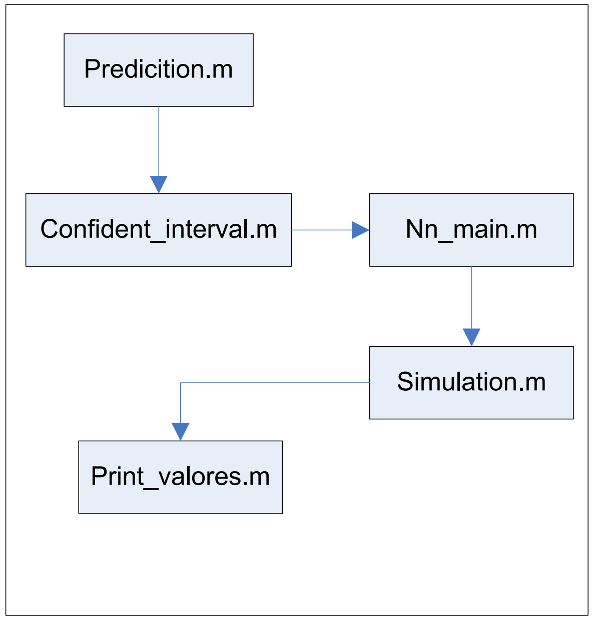


Figure A.8 Sub functions used in *Prediction* routine

2.4.1. Confident_interval.m

Function employed to create an interval of confidence to restrict the forecast value. The idea is to calculate a minimum and maximum limits using intensity information known up to the current time so that if the predictable intensity falls into this interval, it is valid and is used for the intensity prediction model. But it does not, it is discarded

2.4.2. Nn_main.m

Using this function, a feed-forward backpropagation network is created to predict the hurricane intensity. The inputs of this function are basically the best regressors obtained during the selection procedure. The target is the intensity lagged in its respective lead time. A Matlab function called *newff* is used to create the network as follows:

$$\mathrm{net} = \mathrm{newff}\left(\mathrm{pr}, [\mathrm{s}_1 \ \mathrm{s}_2 \cdots \mathrm{s}_{\mathrm{n}1}], \{\mathrm{tf}_1 \ \mathrm{tf}_2 \cdots \mathrm{tf}_{\mathrm{n}1}\}, \mathrm{btf}, \mathrm{pf}\right)$$

where:

- *net* is the feed-forward backpropagation network created recently.
- *pr* is the $r_x 2$ matrix of minimum and maximum values for *r* input elements.
- s_i is the size of the i^{th} layer, for $n1$ layers.
- tf_i is the transfer function of i^{th} layer.
- btf is the backpropagation network training function.
- *pf* is the performance function.

Moreover, a training function is used to train and update the weighting and bias for each neuron. A Matlab function called *train* is used as follows:

$$[net, tr, y] = train(NET, p, t)$$

where:

- NET is the network created using *newff.*
- *p* is a matrix that contains the network inputs.
- *t* is a vector that contains the network targets.
- *net* is the new neural network after the training process
- *tr* is a vector that stores the performance ot the training process
- *y* is a vector that contains the network outputs.

Up to this point, the NN is ready to calculate its intensity prediction, the next step is to simulate the network using current information to obtain results at the time defined by the lead time

2.4.3. Simulation.m

This function is applied to simulate the trained NN with actual information. The syntax is shown as follows:

$$y = \mathrm{sim}(\mathrm{net}, \mathrm{r})$$

Where *y* is the simulated output applying the model *net* and *r* is the input vector that contains information don not know for the model at the training time. For the intensity prediction model, *y* is the forecast intensity and *r* is the current information about the hurricane under study.

2.4.5. Print_valores.m

This function is used to print the results in an ASCII file. It also keeps information about the best regressors and the neural network and regression performance indexes.

APPENDIX B

DETERMINATION OF HURRICANE SAMPLE SIZE

Determining sample size is a very important issue because samples that are too large may waste time, resources and money, while samples that are too small may lead to inaccurate results.

When sample data is collected and the sample mean $\bar{x}$ is calculated, that sample mean is typically different from the population mean μ. This difference (δ) between the sample and population means can be thought of as an error. A formula that will determine the appropriate sample size (n) for a given type II error (β), type I error (α) and δ is shown as follows (Montgomery et al., 1995):

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{\delta^2}$$

Using a sample intensity mean of 50 kt, a standard deviation of 25 kt and δ is defined as 20 kt which is the difference allows between the sample mean and the population mean. The type I error probability is 0.05 and the type II error probability is 0.1.Then, the sample size is:

$$n = \frac{(1.96 + 1.28)^2 * 25^2}{20^2} = 16.4 = 16 \text{ cases}$$

This procedure determines the sample size required when estimating the mean of a normal distribution. Assuming that the standard deviation of the normal distribution equals 25 kt, 16 cases are required to have a 90.0% chance of rejecting the hypothesis that μ=50 kt when the true μ=65 kt.

APPENDIX C

ADVANCED MICROWAVE SOUNDING UNIT (AMSU)

The Advanced Microwave Sounding Unit (AMSU) was launched on May 13, 1998 and became the first microwave sounder with complete temperature sounding capability from the lower troposphere to the upper stratosphere. The AMSU is well suited for the observation of tropical cyclones because its measurements are not significantly affected by the ice clouds that cover tropical storms. It is a cross-track, line-scanned instrument designed to measure scene radiances in 15 discrete frequency channels which permit the calculation of the vertical temperature profile from about 3 milibars (45 km) pressure height to the Earth's surface (figure C.1).

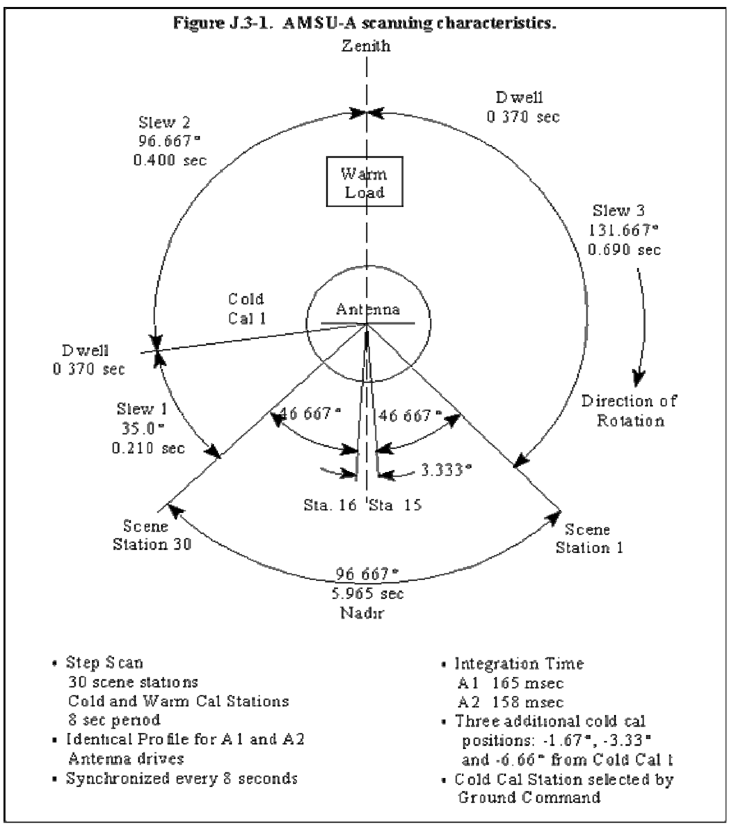


Figure C.1 AMSU-A scanning characteristics

At each channel frequency, the antenna beamwidth is a constant 3.3 degrees (at the half power point). Thirty contiguous scene resolution cells are sampled in a stepped-scan fashion every eight seconds, each scan covering 50 degrees on each side of the sub satellite path. These scan patterns and geometric resolution translate to a 50 km diameter cell at nadir and a 2,343 km swath width from the 833 km nominal orbital altitude.

The main tropical cyclone parameters of interest to the forecaster are storm location and movement, thermal anomalies, wind speeds, and rain rate. While other satellite instruments can be used to estimate these parameters, the AMSU is the first satellite instrument that has the potential to measure all of them. Since clouds are nearly (but not completely) transparent to microwave radiation, the AMSU can measure the above parameters even through the central dense overcast that prevents visible and infrared satellite instruments from making these measurements.

The AMSU has significantly improved spatial resolution, radiometric accuracy, and the number of channels over the previous Microwave Sounding Unit (MSU; figure C.2) that has been used for tropical cyclone analysis. The filled gray ellipses illustrate the 110-km resolution of the MSU. The black outlined ellipses illustrate the 48-km resolution of the AMSU-A instrument. The black dots mark the centers of the scan spots of the 16-km resolution AMSU-B instrument.

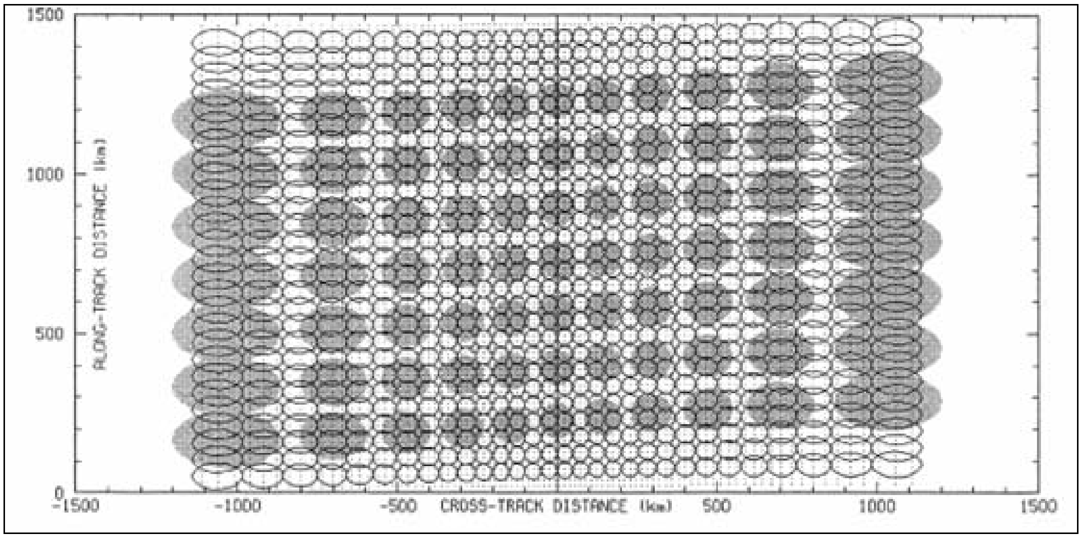


Figure C.2 Resolution improvements of AMSU over MSU

The AMSU complements the much more frequent and higher-resolution observations of the geostationary satellites to give a more complete description of tropical storms.

An application of the upper-level warm temperature anomalies is in assessing the intensity of tropical cyclones (maximum 1-min average wind speed at 10 m). Using data from previous microwave instruments, several investigators have examined the relationship between temperature anomalies (figure C.3) and the surface wind speed and central pressure of tropical cyclones (e.g., Kidder et al. 1978, 1980; Velden and Smith 1983; Velden 1989; Velden et al. 1991). The much higher spatial resolution of the AMSU allows one to more accurately estimate the storm intensity. The maximum temperature anomaly near the center of the storm was related with surface wind speeds and central pressures obtained from operational track data.

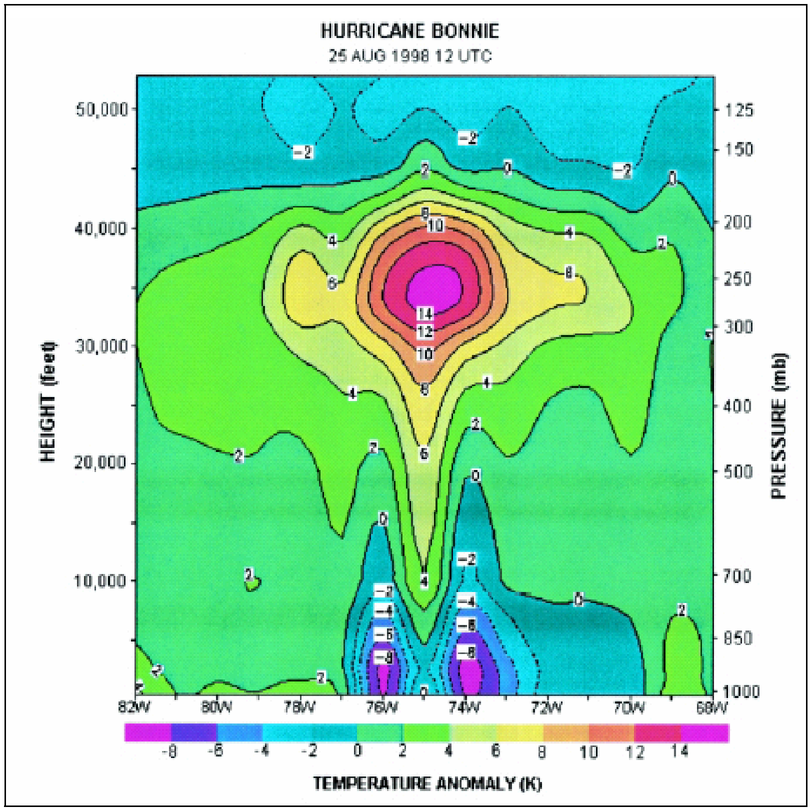


Figure C.3 Relationship between temperature anomaly and pressure

Figure C.4 8 shows for the hurricane Bonnie (1998) that the temperature anomalies closely follow both the wind speeds and the pressures. Gaps in the data are caused by the storm being located between orbital swaths or by missing AMSU data.

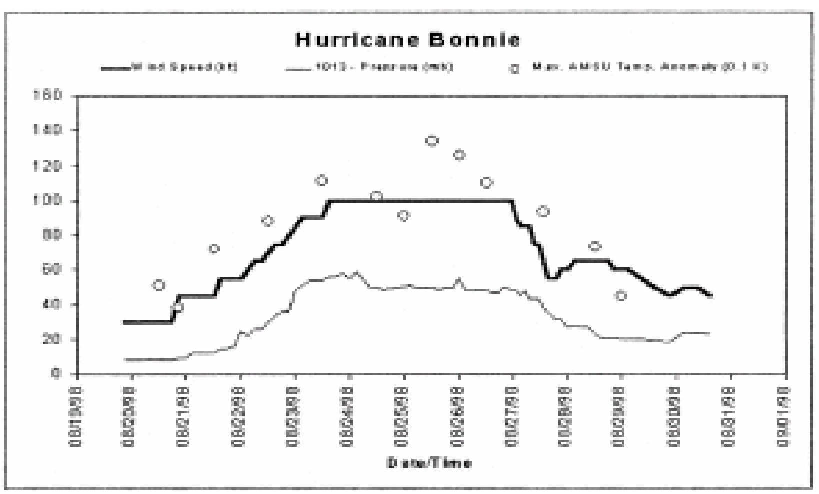


Figure C.4 Plot of wind speed, central pressure, and maximum temperature anomaly (retrieved from AMSU data)

To prove the effect of the hurricanes over the AMSU-A data, the following experiment was developed: Amsu-A data was obtained for the Julian date 253(Sept 10, 2003) in the Atlantic Ocean. The same day, the hurricane Isabel was located in the coordinates of 21.1 degrees of latitude and 50.4 degrees of longitude with sustained winds of 115 knot. After two corrections (Mo, 1999; Wark, 1993), the brightness temperature of AMSU-A (channel 7) is plotted to find out the hurricane's position (figure C.5).

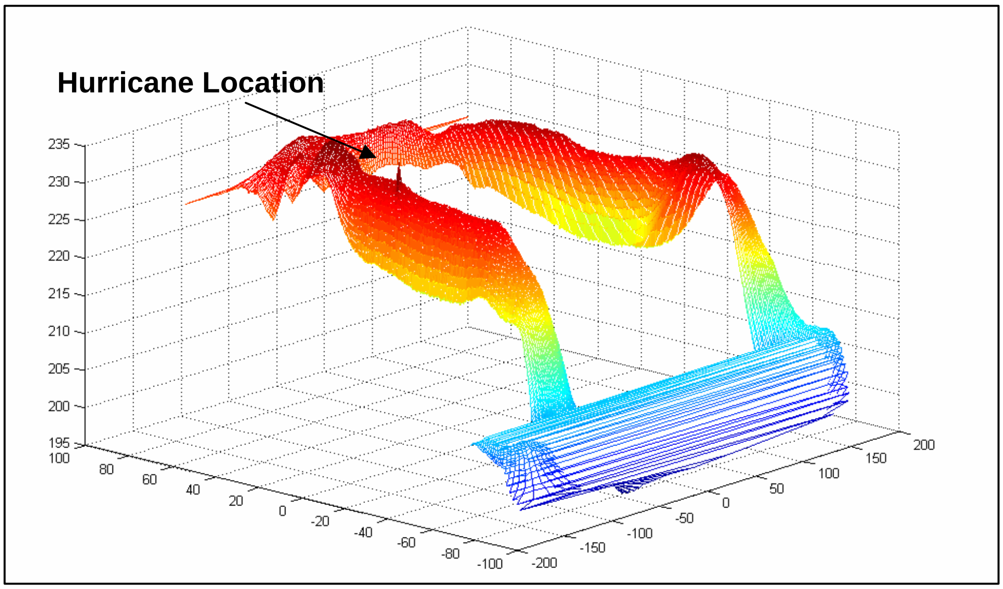


Figure C.5 Amsu data (channel 7) for hurricane Isabel (03/0910)

This figure showed that the hurricane had clearly an effect in the brightness temperature's behavior. To see with more details this effect, an area was extracted (figure C.6) from the original view. The result (figure C.7) showed clearly the effects of the hurricane over the brightness temperature at 250 mb and proved the results obtained by Kidder et al. (2000).

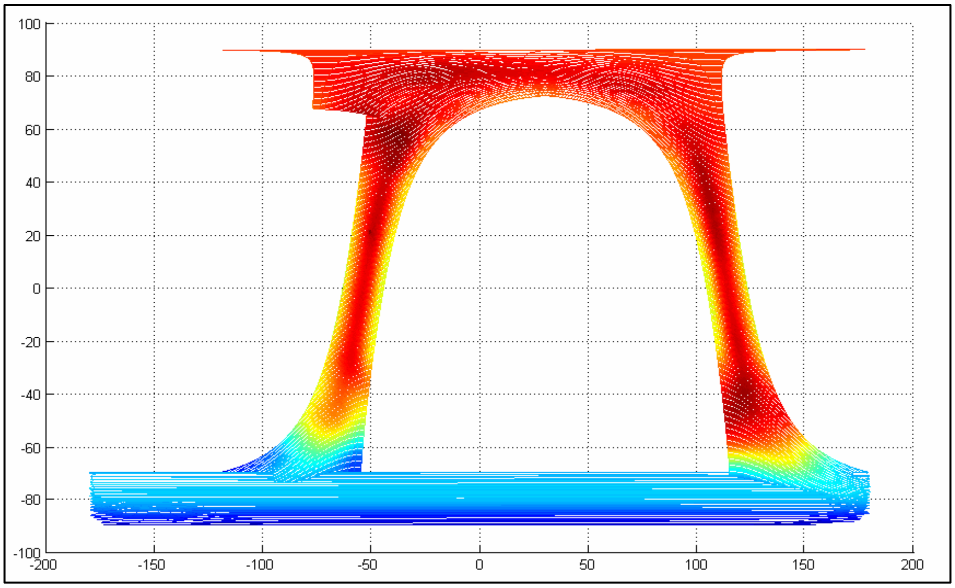


Figure C.5 Aerial view of Amsu data (channel 7) for hurricane Isabel (03/0910)

The next task in this attempt to use the AMSU data in the hurricane intensity prediction was to prove the relation obtained by Kidder et al. (2000) who demonstrated a correlation among the hurricane intensity, hurricane pressure and temperature at 250 mb obtained from AMSU.

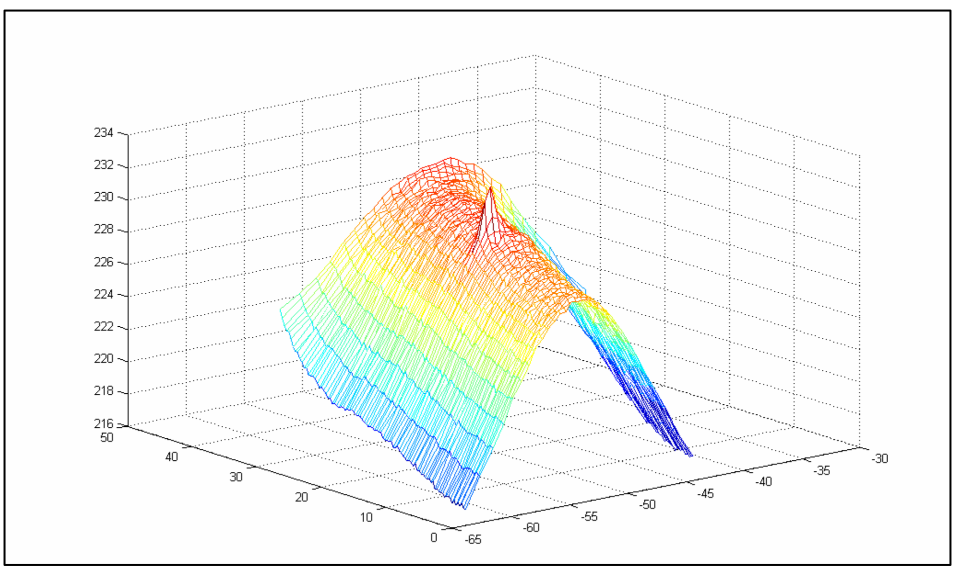


Figure C.6 Extracted area from AMSU (channel) for hurricane Isabel (03/09/10)

Up to this moment the methodology to obtain the temperature at different pressure levels from AMSU is under development, instead temperature from NCEP at

250 m was used to prove this relationship. The experiment can be summarized as follows: first, intensity and pressure observations for 6 days (4 observations per day since 8 to 13 Sept, 2003) were taken for the hurricane Isabel. For each one of these observations, a NCEP file that contained the temperature at 250 mb was obtained. Second, a program was developed to get the five nearest NCEP temperature that surround the hurricane position and the following array was created:

$$\text{data} = \begin{bmatrix} I_1 & P_1 & T_{1,1} & T_{1,2} & T_{1,3} & T_{1,4} & T_{1,5} \\ I_2 & P_2 & T_{2,1} & T_{2,2} & T_{2,3} & T_{2,4} & T_{2,5} \\ \vdots & \vdots & & & \vdots & \vdots & \vdots \\ I_{24} & P_{24} & T_{24,1} & T_{24,2} & T_{24,3} & T_{24,4} & T_{24,5} \end{bmatrix}$$

where I_i is the i^{th} intensity observation ($i = 1,\ldots,24$), P_i is the i^{th} pressure observation, and $T_{i,j}$ is i^{th} temperature observation for the j^{th} position ($j = 1,\ldots,5$).

Third, a linear regression was developed to relate the intensity (I_i) with pressure (P_i) and the 250 milibars temperatures ($T_{i,j}$). The results (figure C.7 and C.8) obtained using Statgraphics 5.1 showed that both variables were significant at 95 % and the variability of the intensity is explained at 97.4 %.

Multiple Regression Analysis

Dependent variable: Intensity

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	326.215	176.919	1.84386	0.0794
Pressure	-0.955882	0.0371247	-25.7479	0.0000
T4	2.98957	0.694011	4.30768	0.0003

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	5909.07	2	2954.53	440.25	0.0000
Residual	140.931	21	6.71102		
Total (Corr.)	6050.0	23			

R-squared = 97.6706 percent
R-squared (adjusted for d.f.) = 97.4487 percent
Standard Error of Est. = 2.59056

Figure C.7 Output from the multiple regression procedure

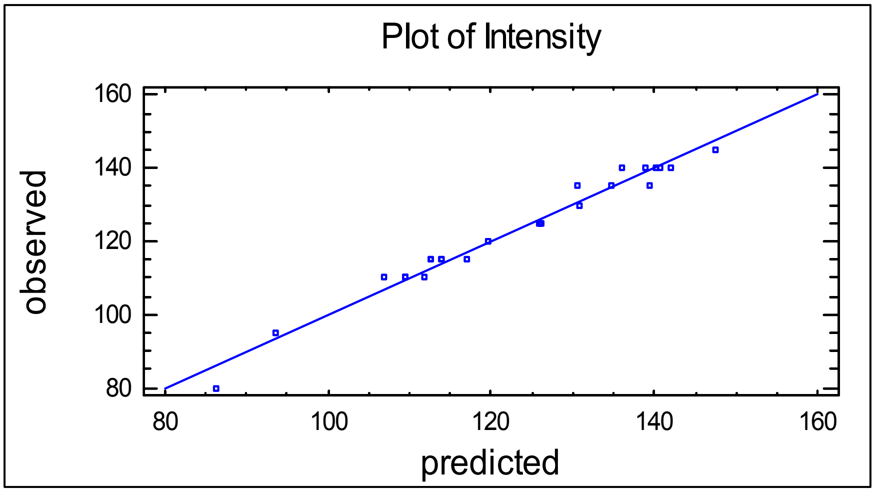


Figure C.8 Observed and predicted intensity obtained from the regression

Figure C.9 is used to judge the relative magnitude of the residuals with respect to the explanatory power of the hurricane pressure. The same analysis is performed for the temperature and shown in figure C.10

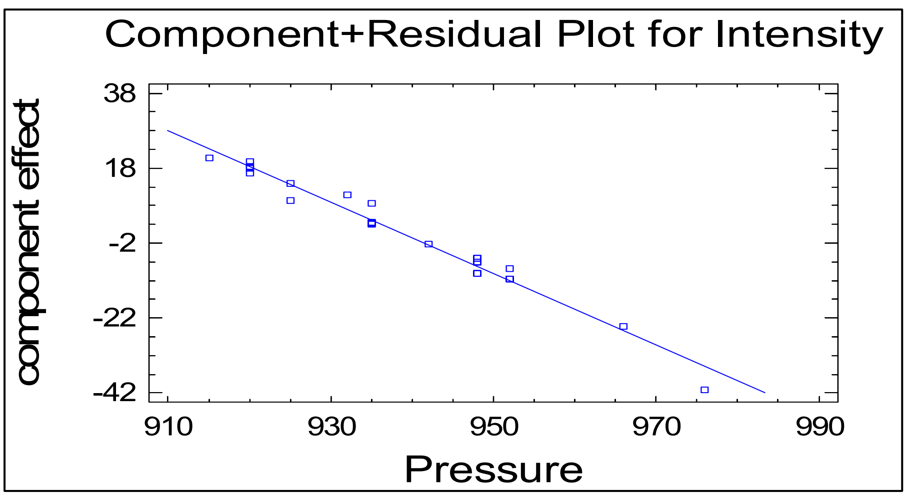


Figure C.9 Component effect plot for the hurricane pressure

This analysis has proved again the importance of the hurricane pressure in the intensity prediction. It has also given a potential variable that can be used in the future to estimate the hurricane intensity.

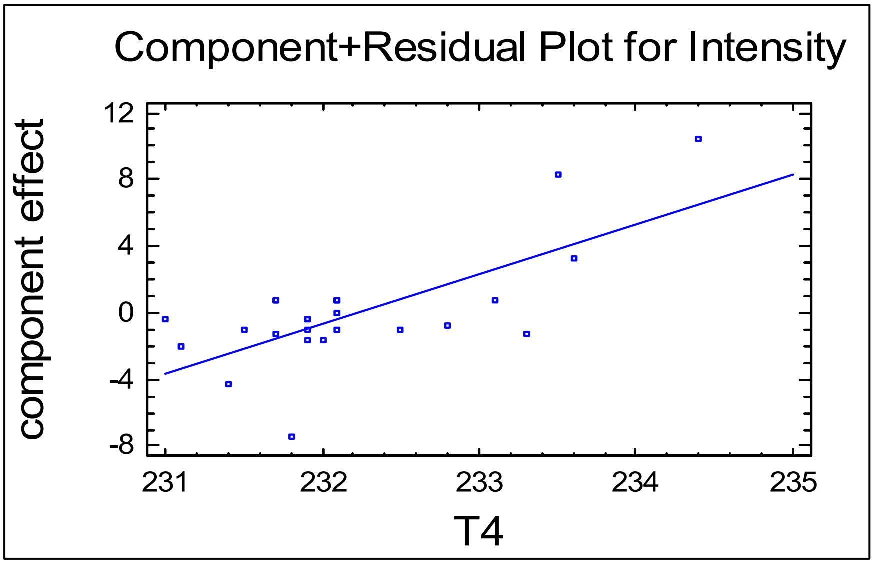


Figure C.10 Component effect for the hurricane temperature at 250 mb.

APPENDIX D

THE SAFFIR -SIMPSON HURRICANE SCALE

The Saffir-Simpson Hurricane Scale is a 1-5 rating based on the hurricane's present intensity. This is used to give an estimate of the potential property damage and flooding expected along the coast from a hurricane landfall. Wind speed is the determining factor in the scale, as storm surge values are highly dependent on the slope of the continental shelf in the landfall region.

D.1 Hurricane Category One

Hurricane winds between 64 and 82 knot. Storm surges generally 4-5 ft above normal .No real damage to building structures. Damage primarily to unanchored mobile homes and trees. Some damage to poorly constructed signs. Also, some coastal road flooding and minor pier damage. Hurricanes Allison of 1995 and Danny of 1997 were Category One hurricanes at peak intensity.

D.2 Hurricane Category Two

Hurricane winds between 83 and 95 knot. Storm surges generally 6-8 ft above normal. Some roofing material, door, and window damage of buildings. Considerable damage to trees with some trees blown down. Considerable damage to mobile homes, poorly constructed signs, and piers. Coastal and low-lying escape routes flood 2-4 hours before arrival of the hurricane center. Small craft in unprotected anchorages break moorings. Hurricane Bonnie of 1998 was a Category Two hurricane when it hit the North Carolina coast, while Hurricane Georges of 1998 was a Category Two Hurricane when it hit the Florida Keys and the Mississippi Gulf Coast.

D.3 Hurricane Category Three

Hurricane winds between 96 and 113 knot. Storm surges generally 9-12 ft above normal. Some structural damage to small residences and utility buildings with a minor amount of curtain wall failures. Damage to trees with foliage blown off trees and large trees blown down. Mobile homes and poorly constructed signs are destroyed. Low-lying escape routes are cut by rising water 3-5 hours before arrival of the center of the hurricane. Flooding near the coast destroys smaller structures with larger structures damaged by battering from floating debris. Terrain continuously lower than 5 ft above means sea level may be flooded inland 8 miles (13 km) or more. Evacuation of low-lying residences with several blocks of the shoreline may be required. Hurricanes Roxanne of 1995 and Fran of 1996 were Category Three hurricanes at landfall on the Yucatan Peninsula of Mexico and in North Carolina, respectively.

D.4 Hurricane Category Four

Hurricane winds between 114 and 135 knot. Storm surges generally 13-18 ft above normal. More extensive curtain wall failures with some complete roof structure failures on small residences. Shrubs, trees, and all signs are blown down. Complete destruction of mobile homes. Extensive damage to doors and windows. Low-lying escape routes may be cut by rising water 3-5 hours before arrival of the center of the hurricane. Major damage to lower floors of structures near the shore. Terrain lower than 10 ft above sea level may be flooded requiring massive evacuation of residential areas as far inland as 6 miles (10 km). Hurricane Luis of 1995 was a Category Four hurricane while moving over the Leeward Islands. Hurricanes Felix and Opal of 1995 also reached Category Four status at peak intensity.

D.5 Hurricane Category Five

Hurricane winds greater than 135 knot. Storm surges generally greater than 18 ft above normal. Complete roof failure on many residences and industrial buildings. Some complete building failures with small utility buildings blown over or away. All shrubs, trees, and signs blown down. Complete destruction of mobile homes. Low-lying escape routes are cut by rising water 3-5 hours before arrival of the center of the hurricane. Major damage to lower floors of all structures located less than 15 ft above sea level and within 500 yards of the shoreline. Massive evacuation of residential areas on low ground within 5-10 miles (8-16 km) of the shoreline may be required. Hurricane Mitch of 1998 was a Category Five hurricane at peak intensity over the western Caribbean. Hurricane Gilbert of 1988 was a Category Five hurricane at peak intensity and is one of the strongest Atlantic tropical cyclones of record.

APPENDIX E
CONFIDENCE INTERVAL FOR THE HURRICANE INTENSITY PREDICTION

A confidence interval for the prediction of the hurricane intensity was developed to take into account the inherent variation of the neural network during its training stage due especially to the selection of the initial point.

An interval of confidence of an unknown parameter θ is an interval of the form $1 \leq \theta \leq u$ where the end-points *l* and *u* depend on the numerical value of the statistic $\hat{\theta}$ for a particular sample and on the sampling distribution of $F(\hat{\Theta})$. Since different samples will produce different values of $\hat{\theta}$ and, consequently, different values of the end-points *l* and *u*, these end-points are values of random variables, say, L and U, respectively. From the sampling distribution of $F(\hat{\Theta})$, then, to determine values of L and U such that the following probability statement is true:

$$P(L \leq \theta \leq U) = 1 - \alpha$$

where $0 \leq \alpha \leq 1$.Thus, the probability of selecting a sample that will produce an interval containing the true value of θ is $1 - \alpha$.

The procedure used to calculate the interval of confidence is explained as follows:

1. The intensity prediction for each instance of time was repeated 100 times varying the initial point in the neural network' stage.
2. The cumulative distribution of the prediction sample was fitted to the closest distribution function using as a measure of fitness the Kolmogorov – Smirnov test.

The Kolmogorov - Smirnov test calculates the maximum distance between the cumulative distribution of the data and the cumulative distribution function of the fitted distribution. This calculation is a nonparametric method that tests the overall goodness of fit between the distribution of the data and a given distribution. The Kolmogorov – Smirnov can be expressed as follows:

$$D = \max\left|F(x) - \hat{F}(x)\right|$$

This test involves two distribution functions $F(x)$ and $\hat{F}(x)$ where $F(x)$ is the cumulative probability density function of the population and $\hat{F}(x)$ is the cumulative probability density function of the sample. If the two have been drawn from the same population distribution, the cumulative distributions of both samples may be expected to be close to each other; if the two cumulative distributions differ, this suggests that the samples come from different populations.

3. Once, the distribution function that best adjusted the sample under study was detected, an interval of confidence was calculated using the parameters that correspond for this adjusted distribution function.

The prediction of the intensity for Hurricane Isaac (2000) at 12 hours was used to calculate a confidence interval using the procedure describes above. Figure E.1 shows the predictions that have been calculated 100 times and the distribution function that best fit the sample was used to obtain the parameter estimates of this fitted distribution, and a 95 % confidence interval was calculated. Table E.1 shows the values obtained for the hurricane Isaac.

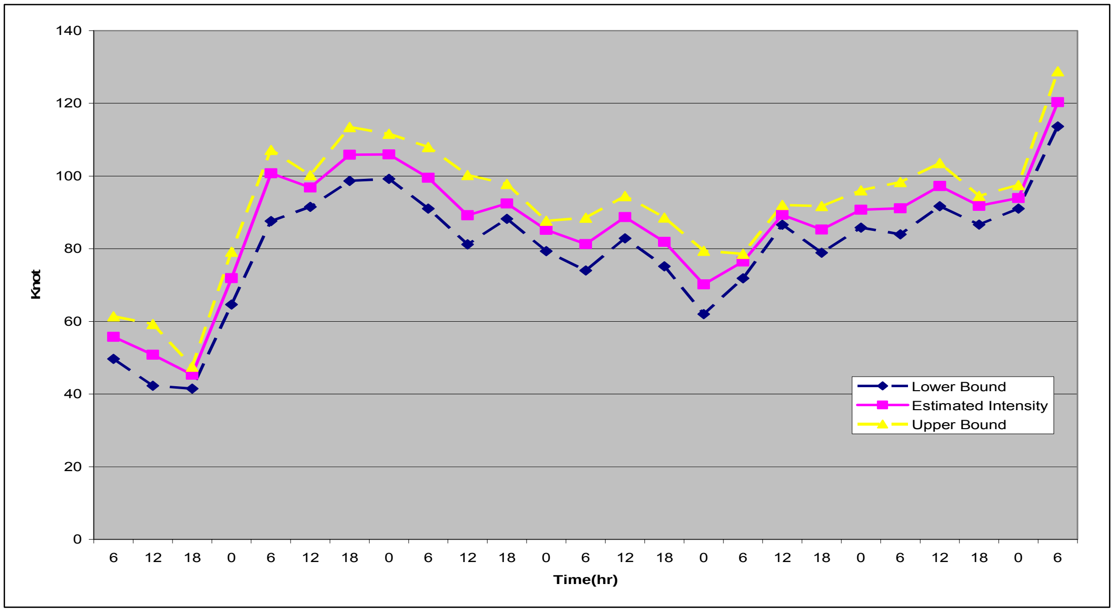


Figure E.1 Confidence interval at 95% for the hurricane Isaac (Sept, 2000)

A histogram was developed to show the frequency of the identified probability density functions. Figure E.2 shows that the majority of the predictions follow a non normal distribution.

Table E.1 Hurricane Isaac (2000) interval of confidence (upper and lower bound)

Month	Day	Time	Latitude	Longitude	Observed Intensity	Lower Bound	Predicted Intensity	Upper Bound
9	23	6	13.9	32.3	55	49.69	55.76	61.38
9	23	12	14.3	33.2	70	42.305	50.834	59.264
9	23	18	14.6	34.2	85	41.48	45.332	47.55
9	24	0	14.9	35	105	64.64	71.904	79.17
9	24	6	15.1	35.8	100	87.59	100.78	107.19
9	24	12	15.5	36.8	100	91.51	96.845	100.24
9	24	18	15.8	37.8	100	98.66	105.87	113.52
9	25	0	16.3	38.6	95	99.23	105.95	111.61
9	25	6	16.7	39.5	95	91.05	99.536	108.02
9	25	12	17.2	40.4	90	81.2	89.221	100.3
9	25	18	17.6	41.2	90	88.24	92.453	97.82
9	26	0	17.9	42	90	79.34	85.198	87.69
9	26	6	18.3	42.9	85	73.99	81.297	88.52
9	26	12	18.6	43.9	75	82.88	88.693	94.5
9	26	18	19.1	45	75	75.159	81.878	88.597
9	27	0	19.6	46	80	61.97	70.15	79.41
9	27	6	20.4	47	85	71.89	76.4615	78.61
9	27	12	21	48.1	90	86.63	89.3455	92.05
9	27	18	21.9	49.5	95	78.89	85.301	91.71
9	28	0	22.8	50.6	100	85.82	90.7315	96.06
9	28	6	23.8	52	105	83.966	91.1515	98.337
9	28	12	25	52.9	110	91.68	97.2525	103.58
9	28	18	26.6	54.2	120	86.65	91.805	94.52
9	29	0	28	55.1	115	91.03	93.9645	97.53
9	29	6	29.7	55.9	110	113.65	120.35	128.831

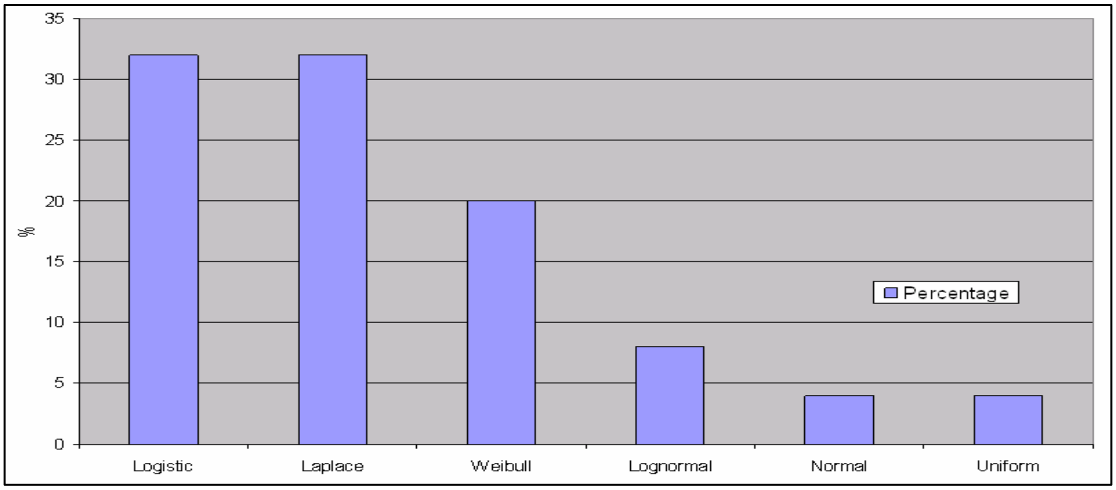


Figure E.2 Histogram for the distribution function used in the development of the hurricane Isaac confidence interval

For the 25 intensity predictions, the Laplace and Logistic distribution were used 8 times to fit the distribution of the intensity prediction sample, the Weibull distribution was used 5 times, the Lognormal distribution, and the Normal and Uniform distribution were used one time each.

Figure E.3 shows four examples of the most representative distribution functions used in the development of the confidence interval for the hurricane Isaac.

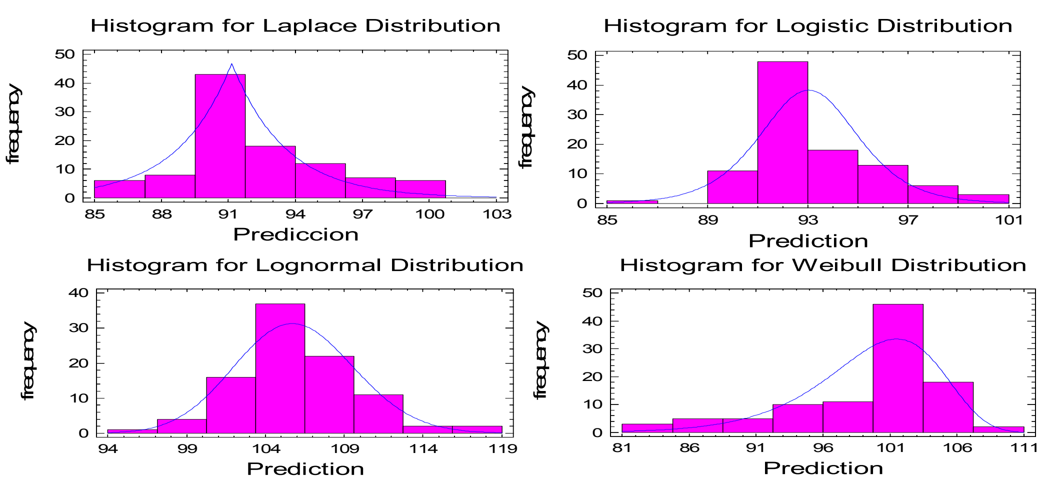


Figure E.2 Histogram for the most representative distribution function used in the development of the hurricane Isaac confidence interval