1992-06

# A real time sharpening of NOGAPS predictions of mid-latitude Central Pacific cyclones.

Harrington, Edward J.

Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/23760

Unclassified

curity classification of this page

## REPORT DOCUMENTATION PAGE

1a Report Security Classification Unclassified

1b Restrictive Markings

2a Security Classification Authority

2b Declassification Downgrading Schedule

3 Distribution Availability of Report
Approved for public release; distribution is unlimited.

4 Performing Organization Report Number(s)

5 Monitoring Organization Report Number(s)

6a Name of Performing Organization
Naval Postgraduate School

6b Office Symbol (*if applicable*) OR

7a Name of Monitoring Organization
Naval Postgraduate School

6c Address (*city, state, and ZIP code*)
Monterey, CA 93943-5000

7b Address (*city, state, and ZIP code*)
Monterey, CA 93943-5000

8a Name of Funding Sponsoring Organization

8b Office Symbol (*if applicable*)

9 Procurement Instrument Identification Number

8c Address (*city, state, and ZIP code*)

10 Source of Funding Numbers

| Program Element No | Project No | Task No | Work Unit Accession No |
|---|---|---|---|
| | | | |

11 Title (*include security classification*) A REAL TIME SHARPENING OF NOGAPS PREDICTIONS OF MID-LATITUDE CENTRAL PACIFIC CYCLONES

12 Personal Author(s) Edward J. Harrington

13a Type of Report
Master's Thesis

13b Time Covered
From To

14 Date of Report (*year, month, day*)
June 1992

15 Page Count
68

16 Supplementary Notation The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17 Cosati Codes

| Field | Group | Subgroup |
|---|---|---|
| | | |
| | | |

18 Subject Terms (*continue on reverse if necessary and identify by block number*)
NOGAPS modification, probability ellipses, regression models

19 Abstract (*continue on reverse if necessary and identify by block number*)

A Modifying Model is developed which sharpens the 24 hour position forecast issued by the Navy Operational Global Atmospheric Prediction model (NOGAPS) 24 hours into a selected, mid-latitude, Central Pacific cyclone. The technique involves measuring cyclone characteristics within the first 24 hours and using these values in regression equations to provide improved forecasts for the next 24 hour position forecast. Generally, the modified position forecasts are to the left and ahead of the NOGAPS position forecasts along the anticipated track of the cyclone. Probability ellipses about the Modifying Model estimates cover about 50 to 60 percent of the area of the corresponding NOGAPS probability ellipses. Only cyclones in the deepening phase (central pressure decreasing) or forecast to be in the deepening phase are utilized in the data base. The Modifying Model is sufficiently simple that shipboard personnel can make the computations in real time.

20 Distribution Availability of Abstract
☒ unclassified unlimited ☐ same as report ☐ DTIC users

21 Abstract Security Classification
Unclassified

22a Name of Responsible Individual
R.R. Read

22b Telephone (*include Area code*)
(408) 646-2382

22c Office Symbol
OR Re

DD FORM 1473,84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete

security classification of this page

Unclassified

Approved for public release; distribution is unlimited.

# A Real Time Sharpening of NOGAPS
# Predictions of Mid-Latitude Central
# Pacific Cyclones

by

Edward J. Harrington
Lieutenant, United States Navy
B.A., College of the Holy Cross, 1984

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 1992

# ABSTRACT

A Modifying Model is developed which sharpens the 24 hour position forecast issued by the Navy Operational Global Atmospheric Prediction model (NOGAPS) 24 hours into a selected, mid-latitude, Central Pacific cyclone. The technique involves measuring cyclone characteristics within the first 24 hours and using these values in regression equations to provide improved forecasts for the next 24 hour position forecast. Generally, the modified position forecasts are to the left and ahead of the NOGAPS position forecasts along the anticipated track of the cyclone. Probability ellipses about the Modifying Model estimates cover about 50 to 60 percent of the area of the corresponding NOGAPS probability ellipses. Only cyclones in the deepening phase (central pressure decreasing) or forecast to be in the deepening phase are utilized in the data base. The Modifying Model is sufficiently simple that shipboard personnel can make the computations in real time.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# I. INTRODUCTION

## A. BACKGROUND

Ships at sea are extremely vulnerable to weather conditions. Limited and somewhat crude instruments (barometer, anemometer) are available to help the mariner measure the weather nearby and anticipate imminent conditions, but there is little of a practical nature that he can do to anticipate weather conditions well in advance of his present position and take early avoiding action. Reports and messages from local and national weather services provide irreplaceable information on cyclone activity and other weather related phenomena which allow the ship at sea to take early avoiding action.

The Navy Operational Global Atmospheric Prediction Model (NOGAPS) is a state of the art, fluid dynamic atmospheric general circulation model that simulates the characteristics of the atmosphere [Ref. 1]. The model is executed twice daily on a mainframe supercomputer at Fleet Numerical Oceanographic Center to analyze and predict the atmosphere from the surface to the middle-stratosphere. A separate Vortex Tracking Program is used to interpret the NOGAPS analyzed and forecast representation of sea-level pressure mid-latitude cyclones. Every 12 hours a new set of predictions is issued, covering successive 12 hour periods up to 120 hours from the time of the current cyclone forecast.

Some general, predictive tendencies in the NOGAPS model have been discovered through detailed analysis by various researchers. Given certain conditions, predictions may follow specific patterns. This, in turn, leads to a rough pattern of prediction errors associated with specific cyclone qualities. For example, Harr et al. 1992, [Ref 2 : p. 14.] note that maximum underforecasting errors (forecast central pressure higher than the actual central pressure) occur over the Central Pacific (CPAC) region of climatological maximum cyclone deepening. That is, the predictions of central pressure tend to be higher than actual when the cyclone's central pressure is decreasing toward its lowest point.

## B. PROBLEM DESCRIPTION

Knowlege of these errors is qualitative and general. The ability to estimate the errors in a given forecast is totally dependent on the skill and experience of the forecaster. Although research is being conducted to remove or further reduce biases in the NOGAPS model, there are no tools which provide forecast error estimates, given the specific parameters of a unique cyclone. However, a simple, statistical model has been developed from a data base of previous cyclones to adjust the future position predictions. Using a data base of cyclones from 1989 and 1990 it utilizes parameters obtainable in the first 24 hours, to predict the next 24 hour cyclone position. The prediction is in the form of a deviation from the NOGAPS position forecast and can be computed by the ship receiving the NOGAPS forecast, or incorporated into the forecast issued by the forecaster.

## C. KEY TERMS AND CONCEPTS

The position error between the NOGAPS forecast and the actual cyclone position is described by three sides of a right triangle: Forecast Error, Track Error and Distance Error Along Track (see Figure 1). Forecast Error (FPE) is defined as the straight line distance from the actual position to the forecast position, measured in nautical miles (nm). Track Error (TKE) is defined as the distance from the forecast position to the nearest position along the actual track of the cyclone, measured in nautical miles. Distance Error Along Track (DEAT), measured in nautical miles, is defined as the distance between the point on the actual track closest to the forecast position (pt A) and the actual position (pt C).

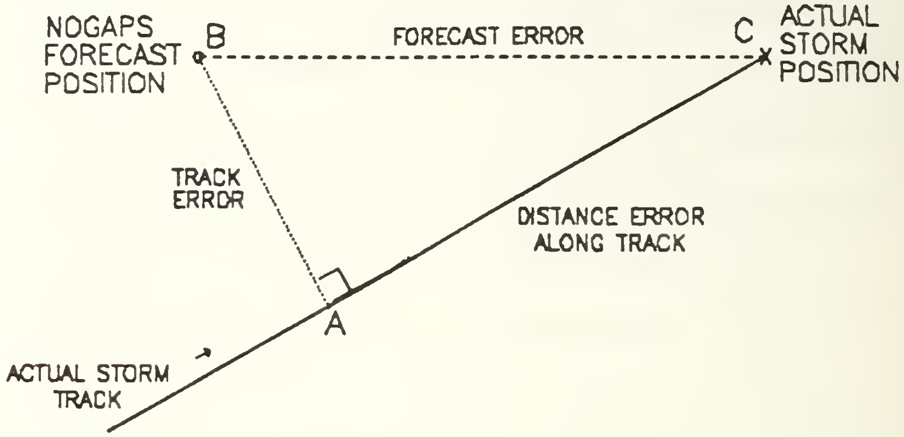


Figure 1. Components of Position Error

As can be seen from Figure 1, Track Error is measured from the predicted position (B) to the nearest position along the actual track (A). Track errors to the left (right) are considered negative (positive) and Distance Errors Along Track ahead of (behind) the NOGAPS forecast are considered positive (negative).

## D. ERROR ESTIMATE COMPOSITION

A point estimate prediction for the deviation of the actual position from the NOGAPS predicted position will be developed in terms of an anticipated TKE and DEAT. A Distance Error Along Track (DEAT) prediction yields one point along the intended cyclone track. From this point a predicted Track Error (TKE) right or left will result in a position forecast. The Modifying Model position will be developed by applying the predicted DEAT and predicted TKE to the NOGAPS forecast. The track error estimate is applied in the same direction as its sign indicates, while the DEAT estimate is applied in the opposite direction. A negative DEAT estimate indicates that the NOGAPS prediction will lag behind the actual position. Therefore, the NOGAPS position must be advanced in the positive DEAT direction to reduce the expected error. For example, a Track Error estimate of +10nm moves the NOGAPS estimate 10nm to the right of track, but a DEAT estimate of -47nm results in the NOGAPS position being advanced 47nm along track. The point estimate is computed by adding both estimates as in vector addition, starting at the NOGAPS predicted position. Figure 2 displays how the Modifying Model point estimate is applied to a cyclone with these error estimates.

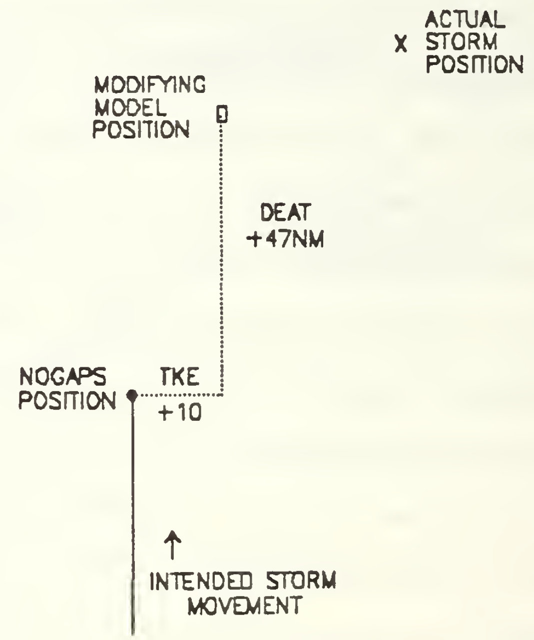


Figure 2. Applying the Modifying Model Estimate

Forecast Error (FPE) is not used because it is not a vector, and would generate ambiguity. At best it can be used to create an area forecast and not a point estimate. For example, an FPE prediction of 60 nautical miles (nm) can be anywhere on a circle of radius 60nm from the forecast position. If an FPE prediction is combined with either a TKE or DEAT estimate an ambiguity developes between four possible points (see Figure 3.). A decision would have to be made concerning which predictions to use.

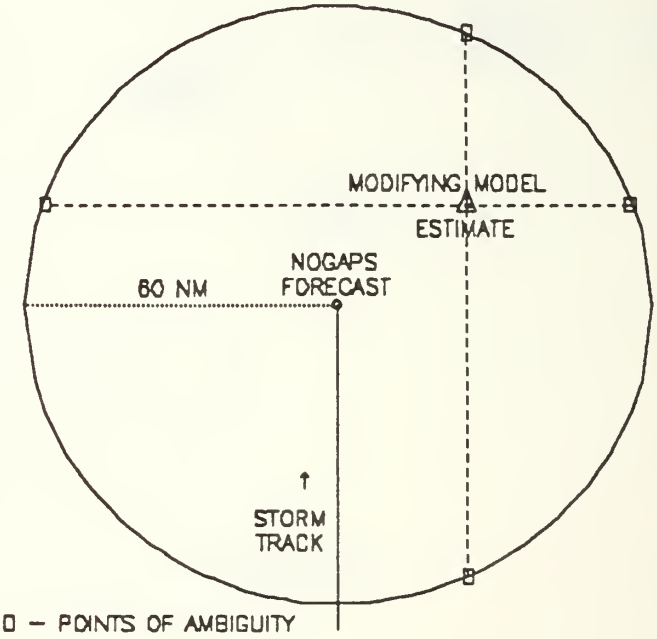


Figure 3. Ambiguity Illustration using FPE Estimate

The modelling goal is to provide a modified forecast based on an earlier actual position error measured from the NOGAPS forecast. To provide meaningful results, the focus of the model is narrow. Its application is limited to Central Pacific, mid-latitude cyclones during the third 24 hour position forecast of a deepening or predicted to be deepening cyclone. Meaningful variance reduction is possible using this stratification. A similar approach could be used during cyclone formation if analysis of that aspect of the prediction is desired.

The effect of twelve parameters on each of the three position error components is studied. Each parameter comes from either a cyclone or forecast characteristic, measurable within the first 24 hours of forecasting. Regression model predictions of Track Error and Distance Error Along Track are developed to provide a point estimate for a modified position forecast. Additionally, probability ellipses, based on the Bivariate Normal distribution, and following the general principles identified in Ref. 3, are created about each point estimate to allow for model variability and provide a reasonable area of probability. Regression model coefficients are estimated using the Jacknife technique and a set of 20 new cyclones from 1991 is used to validate the model.

The results are encouraging and indicate that the simple models can provide useful enhancements of the NOGAPS predictions. If the results are utilized in a sensible manner, both the forecaster and the ship at sea will reap benefits.

## E. DATA BASE SELECTION

A large data base of mid-latitude, Central Pacific cyclones from 1989-1990 have been studied. Over 200 summaries have been reviewed and 57 have been selected to form the data base for the analysis.

Clearly, 57 cyclones is not a large data base. While there are well over 200 candidate cyclones, the majority are unuseable for one of the following reasons: (1)

missing data, (2) deepening phase too short. The first requires little discussion except to mention that missing data occur as a result of either an omision of a 24 hour forecast or an omission of the cyclone evaluation at a given 12 hour increment. The second requires more explanation.

The Modifiying Model is concerned only with predicting the Track Error and Distance Error Along Track during the deepening phase of the cyclone. The deepening phase is determined by cyclone central pressure. While the central pressure is decreasing, the cyclone is deepening. This phase ends when the lowest central pressure is observed. Therefore, in order for a cyclone to be accepted into the data base, it must satisfy the following critieria: (1) the cyclone must deepen through the first 24 hours (when the first two 24 hour forecasts are made), and (2) the third 24 hour forecast must predict continued deepening. If a cyclone takes longer than 48 hours to reach its lowest central pressure and there are no missing data, it is placed in the data base. However, there is difficulty in those cases for which the third 24 hour central pressure prediction calls for continuing cyclone deepening but the cyclone does not actually deepen any further. These cyclones may not have a deepening phase of at least 48 hours but, the model must be applied since the pending 24 hour forecast (although incorrect) calls for deepening below the pressure reached 24 hours into the cyclone. When the model is applied, the user has no idea what the low pressure will be or when it will occur. The reader is referred to Appendix A which contains descriptions of four cyclones; two which were accepted into the data base and two rejected cyclones.

Previous research (Harr et al, 1992) has shown that position predictions follow distinctly different patterns during the deepening and filling (increasing central pressure) phases. Thus, mixing cyclone predictions and observations from cyclones which are deepening and filling is likely to degrade modelling efforts.

## F. PARAMETER DESCRIPTION

The following list contains 15 parameters extracted from each cyclone summary. Numbers 1-12 can be obtained within the first 24 hours of cyclone formation and represent the independent variables considered in the modelling of the third 24 hour forecast position error. Parameters 13-15 represent the components of the third position error after evaluation. The model establishes a position prediction based on predictions of parameters 13 and 14.

1. Initial Cyclone Latitude
2. Initial Cyclone Longitude
3. Initial Central Pressure (ICP)
4. Central Pressure After 24 Hours
5. Cyclone Central Pressure Drop in First 24 Hours
6. First 24 Hour Central Pressure Forecast Error
7. First 24 Hour Forecast - Track Error (TKE1)
8. First 24 Hour Forecast - Distance Error Along Track (DEAT1)
9. First 24 Hour Forecast - Forecast Error (FPE1)
10. Latitude Change in First 24 Hours
11. Longitude Change in First 24 Hours
12. Ratio of Longitude Change to Latitude Change (SLOPE)
13. Third 24 Hour Forecast - Track Error (TKE)
14. Third 24 Hour Forecast - Distance Error Along Track (DEAT)
15. Third 24 Hour Forecast - Forecast Error (FPE)

All parameters are easily measurable by any reasonably knowlegeable person. An experienced navy quartermaster would have no difficulty.

The first four parameters require no explanation and are read directly from the cyclone summary. Parameter five is simply the difference between three and four. Parameter six is determined by comparing the first 24 hour central pressure prediction with the actual central pressure 24 hours from the time of the forecast. Positive values are given if the predicted central pressure is higher than the actual. Parameters seven, eight and nine are all computed from the difference in actual cyclone position after 24 hours and the predicted position as per the definitions previously given. Parameter(s) 10 (11) is the difference between the initial latitude (longitude) and the latitude (longitude) after 24 hours. Parameter 12 is the ratio of parameter 11 to 10 and gives an indication as to the general direction of the cyclone. Parameters 13, 14 and 15 are computed in the same manner as 7, 8 and 9 using the actual cyclone position at 48 hours and the 24 hour prediction made 24 hours into the cyclone.

# II. POSITION ESTIMATE DEVELOPMENT

## A. OUTLIER REMOVAL

Three outliers which affect the DEAT Model and two which affect the Track Error Model have been removed from the data base. This reduces the number of cyclones to 52. These data points were removed only after careful analysis of residual plots and consideration of the random nature of weather. A detailed, step by step summary of the procedures used to isolate and remove outliers is discussed in Appendix B.

## B. MODEL COMPONENTS

Variable selection is performed separately for each of the two components. The resulting model assumes independence of the errors. The same variables were used in a multivariate regression version. The results and performance comparisons appear in Appendix C. The multivariate regression version did not outperform the present modelling.

The Track Error estimate is the product of a multiple regression model using two variables: First 24 hour Track Error **(TKE1)** and Initial Central Pressure **(ICP)**. Analysis of the residuals vs fitted values and residuals vs parameter values reveals normality and constant variance and satisfies the necessary regression assumptions (See Appendix B).

The tendency is for the Track Error estimate to be negative (left of track) unless the Initial Central Pressure (ICP) is relatively high and the value of TKE1 is slightly negative or positive in any magnitude. Negative values of TKE1 and lower ICP's produce negative Track Error estimates. Negative Track Error estimates (left of track) are more commonplace and support the general conclusions of Harr et al, 1992 [ Ref. 2, p 16.].

The DEAT estimate is the result of a simple regression model utilizing parameter 12 (Ratio of the Longitude Change to the Latitude Change, **SLOPE**). It has been extremely difficult to find any parameters which have any influence on DEAT. There does not appear to be a strong relationship to any of the parameters except for the Slope. The necessary assumptions in the regression model are satisfied: normality and constant variance of the residuals (see Appendix B).

## C. JACKKNIFE PROCEDURE

The final regression model coefficients are determined using the Jacknife procedure. The Jackknife technique is adopted since it allows for validation during the model building phase and has bias reducing properties [Ref. 4]. The data base is partitioned into k sets (13 in this application) each with n elements (four). The regression model is developed 13 times successively using 48 cases each time. A different set of four data points is excluded each time. The regression model coefficients and constant term values are recorded for each of the 13 cases. Residuals are computed using the four excluded data points and the fitted values from the regression model developed using the 48 cases that do not contain those data points. Thirteen psuedo values are obtained. A psuedo value is the difference between 13 times the regression model coefficients from the entire data base (52 data points) and 12 times the regression model coefficients from the data set without the deleted four data points. The final regression model coefficients are the average of the psuedo values. The variables and formulation of the Jackknife procedure are described in the following:

k = number of data sets within complete data set (13),

$y_{\forall}$ = regression coefficient using all 52 data points,

$y_j$ = regression model coefficient with jth data set excluded,

$y_{j*}$ = jth estimate of regression coefficent (psuedo value),

$y_{j*} = k \times y_{\forall} - (k - 1) \times y_j$ ,

$y_*$ = final regression coefficient estimate,

$y_* = (\sum_{i=1}^{k} v_{j*})/k$ ,

$S^2 = [\sum_{j=1}^{k} (y_{j*} - y_*)^2]/(k - 1)$,

$S_*$ = standard error of regression coefficients, and

$S_* = S/\sqrt{k}$ .

## D. COMPLETED MODELS

The final, jacknifed regression coefficients for the TKE and DEAT models are as follows:

$$\hat{TKE} = -2213.73 + .303(TKE1) + 2.183(ICP)$$

and

$$\hat{DEAT} = -31.16 - 2.76|SLOPE|.$$

## E. MODEL LIMITATION

It is apparent from the final regression model equation for the DEAT that the model does not allow for predictions of positive values, yet, a small fraction of all cyclones do produce positive values of DEAT. The closest value to zero which can be reached is -31.16. This occurs when the value of slope is 0. While this constraint on the model is restrictive, the statistical properties of alternative models are not as good. Another option is use of the untransformed slope value, which produced the following regression model:

$$DE\hat{A}T3 = -53.26 + 2.32(SLOPE).$$

This regression model has the ability to predict positive values of DEAT as slope values become larger than +23. This option was not adopted because there were not enough slope values greater than +23 in the data base. In fact, the only slope value larger than +23 is 28.33 (Cyclone 32) which corresponds to a DEAT of -98nm. Since the overwhelming tendency is for the DEAT to be negative (over 85 percent), and of the nine cyclones with positive DEAT's the largest slope value is 7.25, it is not considered prudent to expect larger positive slope values to produce positive DEAT values. Should future data reveal a correlation between large, positive slope values and positive values of DEAT, a model modification should be considered.

Table 1. JACKKNIFE RESULTS FOR THE TKE MODEL

| | CONST | | TKE1 | | ICP | |
|---|---|---|---|---|---|---|
| | $y_\forall = -2467.36$ | | $y_\forall = 0.27$ | | $y_\forall = 2.43$ | |
| j | $y_{(j)}$ | $y_{(j)\times}$ | $y_{(j)}$ | $y_{(j)\times}$ | $y_{(j)}$ | $y_{(j)\times}$ |
| 1 | -3593.42 | 6400.34 | 0.354 | 0.122 | 3.55 | -6.37 |
| 2 | -2564.22 | -804.06 | 0.30 | 0.5 | 2.53 | 0.77 |
| 3 | -2218.78 | -3222.14 | 0.315 | 0.395 | 2.19 | 3.15 |
| 4 | -2287.3 | -2742.5 | 0.32 | 0.36 | 2.25 | 2.73 |
| 5 | -2328.6 | -2453.4 | 0.32 | 0.36 | 2.295 | 2.415 |
| 6 | -1908.8 | -5392 | 0.34 | 0.22 | 1.88 | 5.32 |
| 7 | -2246.99 | -3024.67 | 0.32 | 0.36 | 2.2 | 3.08 |
| 8 | -2279.95 | -2793.95 | 0.399 | -0.193 | 2.25 | 2.73 |
| 9 | -1924.58 | -5281.54 | 0.34 | 0.22 | 1.89 | 5.25 |
| 10 | -2272.99 | -2842.67 | 0.297 | 0.521 | 2.24 | 2.8 |
| 11 | -1933.3 | -5220.5 | 0.36 | 0.08 | 1.9 | 5.18 |
| 12 | -2305.4 | -2615.8 | 0.27 | 0.71 | 2.27 | 2.59 |
| 13 | -2852.57 | 1214.39 | 0.33 | 0.29 | 2.82 | -1.26 |
| | $B_0 = -2213.73$ | | $B_1 = .303$ | | $B_2 = 2.183$ | |
| | $S_\times = 874.9$ | | $S_\times = .063$ | | $S_\times = .87$ | |

Table 2. JACKKNIFE RESULTS FOR DEAT MODEL

|  | CONST |  | SLOPE |  |
|---|---|---|---|---|
|  | $y_{\forall} = -32.24$ |  | $y_{\forall} = -2.58$ |  |
| j | $y_{(j)}$ | $y_{(j)\times}$ | $y_{(j)}$ | $y_{(j)\times}$ |
| 1 | -27.91 | 2.4 | -2.67 | -3.3 |
| 2 | -36.55 | -66.72 | -2.54 | -2.26 |
| 3 | -26.95 | 10.08 | -3.61 | -10.82 |
| 4 | -32.51 | -34.4 | -2.57 | -2.5 |
| 5 | -32.56 | -34.8 | -2.58 | -2.58 |
| 6 | -35.94 | -61.84 | -2.53 | -2.18 |
| 7 | -32.89 | -37.44 | -2.55 | -2.34 |
| 8 | -31.0 | -22.32 | -2.6 | -2.74 |
| 9 | -34.44 | -49.84 | -2.63 | -2.98 |
| 10 | -28.54 | -2.64 | -1.835 | 3.38 |
| 11 | -33.1 | -39.12 | -2.59 | -2.66 |
| 12 | -32.20 | -31.92 | -2.57 | -2.5 |
| 13 | -32.78 | -36.56 | -2.56 | -2.42 |
|  | $B_0$ = -31.16 |  | $B_1$ = -2.76 |  |
|  | $S_{\times}$ = 6.42 |  | $S_{\times}$ = 0.815 |  |

# III. PROBABILITY ELLIPSE GENERATION

## A. FORMULATION AS A BIVARIATE NORMAL

There is a high degree of variability in the point estimates, both in the NOGAPS forecasts and the Modifying Model forecasts. The forecaster and ship at sea can use an enhanced picture of how the cyclone is likely to deviate from the forecast position. A probability ellipse can be constructed about the Modifying Model point estimate to serve this purpose. The ellipse is constructed using a bivariate normal distribution where the X and Y variables are, respectively, the Track Error estimate and the Distance Error Along Track estimate. Probability ellipses can also be generated for the NOGAPS forecasts.

The ellipses for the Modifying Model are generated using the following residuals from the track error model ($\hat{e}_x$) and the distance error along track residuals ($\hat{e}_y$):

$e_x(i)$ = Actual $TKE_i$ - Modifying Model $TKE_i$,

$\hat{e}_x(1), \hat{e}_x(2), \ldots, \hat{e}_x(52)$ - estimated TKE residuals,

$[\sum_{i=1}^{52} \hat{e}_x]/52 \approx 0$,

$e_x \sim N(0, MSE_x)$,

and

$e_y(i)$ = Actual $DEAT_i$ - Modifying Model $DEAT_i$,

$\hat{e}_y(1), \hat{e}_y(2), \ldots, \hat{e}_y(52)$ - estimated DEAT residuals,

$[\sum_{i=1}^{52} \hat{e}_y]/52 \approx 0$,

$e_y \sim N(0, MSE_y)$.

Thus, the Bivariate Normal is assumed for each pair of residuals ( $\hat{e}_x(i), \hat{e}_y(i)$ ),

$$\begin{pmatrix} \hat{e}_x \\ \hat{e}_y \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, C\right]$$

and

$$C = \begin{bmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{bmatrix}$$

where C is the Covariance Matrix and

$$S_x^2 = \frac{1}{n-1}\,[\sum_{j=1}^{n} \hat{e}_x(j)^2] = 3207$$

$$S_y^2 = \frac{1}{n-1}\,[\sum_{j=1}^{n} \hat{e}_y(j)^2] = 3253$$

$$S_{xy} = \frac{1}{n-1}\,[\sum_{j=1}^{n} \hat{e}_x(j)\hat{e}_y(j)] = 997.5.$$

The density function for the bivariate normal is

$$f(x,y) = ce^{-(\frac{1}{2})Q}$$

where c = constant and

$$Q = [\hat{e}_x \quad \hat{e}_y]C^{-1}\begin{bmatrix} \hat{e}_x \\ \hat{e}_y \end{bmatrix}.$$

Under the assumption of normally distributed errors, Q is a Chi-Squared random variable with two degrees of freedom ( $\chi^2_{(2)}$ ). The size of the ellipse for a desired probability level, $1 - \alpha$, is equal to Q = $k^2$ when $\alpha = P[\chi^2_{(2)} > k^2]$. Using the basic equation of an ellipse and the assumption of independence, the points on the ellipse can be generated ($x_1$ =TKE estimate, $y_1$ = DEAT estimate ), and

$$\frac{(x - x_1)^2}{S_x^2} + \frac{y - y_1{}^2}{S_y^2} = k^2$$

which implies that

$$y = y_1 \pm S_y \sqrt{k^2 - \frac{(x - x_1)^2}{S_x^2}}\ .$$

The variable y represents an envelope of points corresponding to the series of points on either side of the TKE estimate. The choice of $\alpha$ determines $k^2$, which in turn determines the size of the ellipse. The major and minor axis lengths can be determined by selecting the value of y when x = $x_1$, and the value of x when y = 0. For example, using a significance level of $\alpha$ = .25, $k^2$ = 2.78 and the major (x) axis (axis along TKE) was 95.1 and the minor axis (y) (axis along DEAT) was 94.4 nautical miles. These distances can be easily marked about the point estimate, referenced to the intended track, and curves marking the ellipse can be drawn by shipboard personnel.

## B. COMPARISON OF NOGAPS AND MODIFYING MODEL ELLIPSES

Probabilty ellipses can be generated for the NOGAPS forecast in the same manner. For the NOGAPS estimate, $x_1 = y_1 = 0$. The variances of the NOGAPS Track Error residuals and the Distance Error Along Track residuals were much greater than in the Modifying Model . The values are shown below:

$$S_x^2 = \frac{1}{n-1}[\sum_{j=1}^{n}\hat{e}_x(j)^2] = 4420,$$

$$S_y^2 = \frac{1}{n-1}[\sum_{j=1}^{n}\hat{e}_y(j)^2] = 7787,$$

and

$$S_{xy} = \frac{1}{n-1}[\sum_{j=1}^{n}\hat{e}_x(j)\hat{e}_y(j)] = 2596.5.$$

Therefore, a probability ellipse corresponding to the same $\alpha$ value as for the Modifying Model Ellipse is much larger. For the case where $\alpha = .25$, the major (y) axis ( axis along DEAT) was 145.7 and the minor (x) axis (axis along TKE) was 109.8 nautical miles. Figure 4 shows a comparison of the NOGAPS ellipse and the Modifying Model ellipse using $\alpha = .25$. and shows how the Modifying Model ellipse is much more like a circle than the NOGAPS probability ellipse. The area comparisons for the 75 percent ellipses are as follows: NOGAPS = 50,259 $nm^2$ and Modifying Model = 28,203 $nm^2$.

Since the Modifying Model ellipses are smaller than the NOGAPS ellipses, a nice advantage can be obtained through their utilization. However, it should be demonstrated that the smaller ellipses perform to the desired level. In other words, the accuracy of the Modifying Model ellipses should be verified. This task is undertaken in the next chapter with fresh data, but was, also, started utilizing a weaker form of cross validation from the results of the Jackknife procedure.

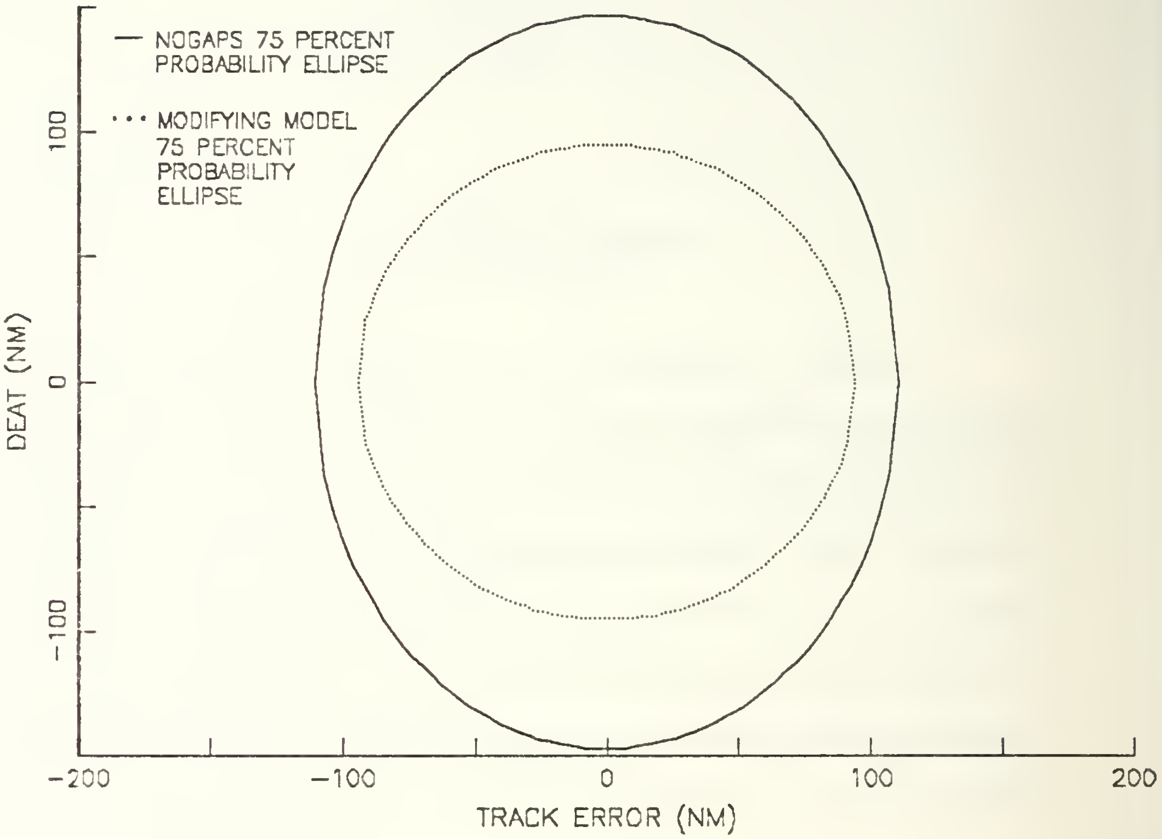


Figure 4. Comparison of NOGAPS and Modifying Model Ellipses

## C. PROBABILITY ELLIPSE VALIDATION

Using the partition of the data set accomplished during the Jackknife procedure in Chapter II, predictions of the deleted four data points from the remaining 48 cases are made. The residuals are computed and each pair (one from each model) was plotted in the Cartesian plane. A 75 percent probability ellipse was plotted about the pairs of residuals using the formulation previously developed. The 75 percent probability ellipse should, therefore, contain about 39 of the 52 residual pairs. As can be seen from Figure 5, only 12 residual pairs are clearly outside the ellipse and two are on the boundary. Thus, approximately 75 percent of the data points are contained within the ellipse and this portion of the validation may be considered successful.

A 95 percent probability ellipse is, also, plotted. Seven points are clearly not contained within the ellipse with two more on the boundary. Therefore, this ellipse seems to perform at less than a 95 percent rate of efficiency (approximately 83 percent). Similar performance is noted with the validation data.

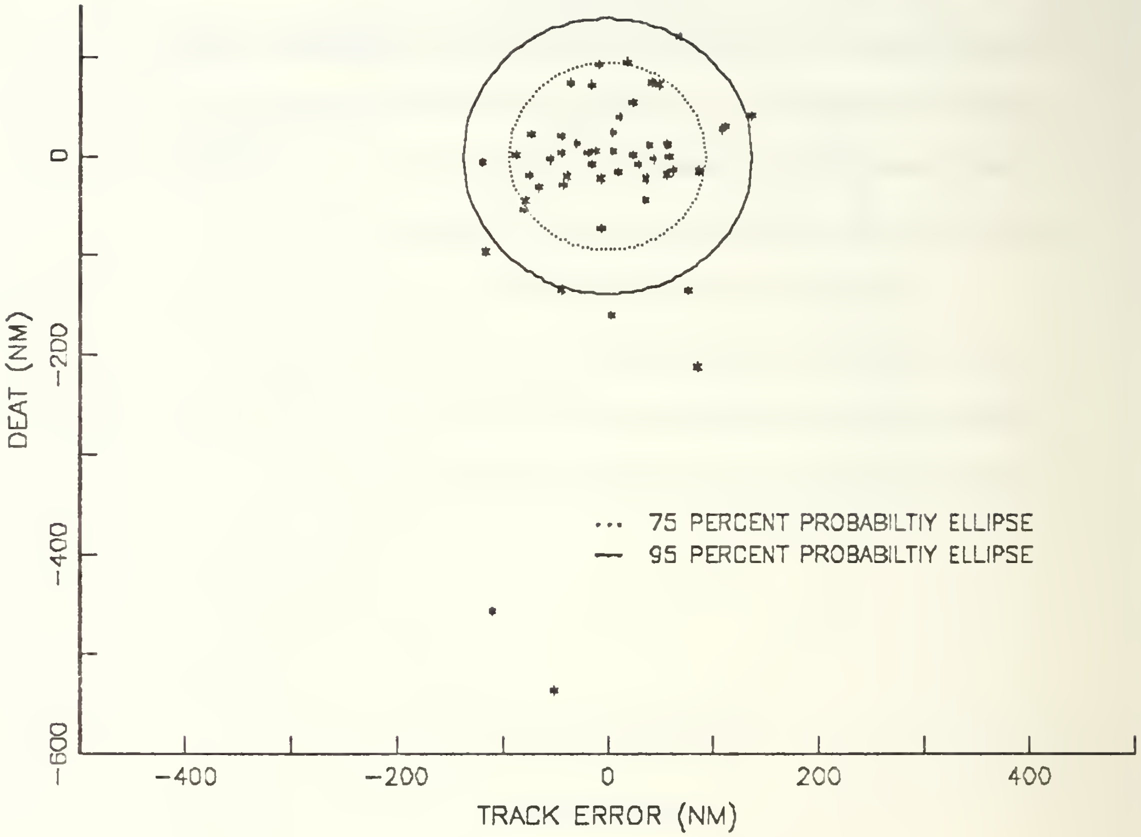


Figure 5. Cross Validation of Probability Ellipses

# IV. MODEL VALIDATION

## A. DATA SET

A data set of 20 cyclones from 1991, which met the same criteria as the cyclones in the data base, is used for separate validation. They seem to provide an adequate representation of the mix which occur over an extended time period. In passing it is noted that five of the 20 contain positive DEAT's, compared with nine of the 57 in the original data base and eight out of 52 in the data base after removal of outliers. Each variable required in both models is derived from the 20 cyclone summaries. Utilizing each model, a position estimate is developed for each 48 hour position, based on the 24 hour cyclone position prediction made 24 hours into the cyclone.

## B. RESULTS

For each cyclone there is a NOGAPS estimate, the Modifying Model estimate and the actual cyclone position. The NOGAPS TKE and DEAT , the Modifying Model TKE and DEAT predictions and the resulting FPEs for each model are computed and compared. They are displayed in Table 3.

Since the Forecast Error component of the position error represents the straight line distance from the predicted position to the actual, it was chosen as the primary measure of effectiveness of the Modifying Model, though several other MOE's will be discussed. The difference between the sum of the forecast errors over all 20 cyclones for the NOGAPS estimates and the Model estimates was -353nm. Thus, the Modifying Model improved each forecast by an average of 18nm, or 20 percent over all 20 cases.

The hypothesis that the Forecast Errors in both cases are equal could be rejected at the $\alpha = .02$ significance level using a Paired T Test and the individual FPEs (columns five and six of Table 3) for each method of position estimation [Ref 5]. Therefore, with

Table 3. COMPARISON OF NOGAPS AND MODIFYING MODEL ERRORS

| 1991 Cyclones | NOGAPS | | MODIFYING MODEL | | NOGAPS | MODIFYING MODEL |
|---|---|---|---|---|---|---|
| | TKE | DEAT | TKE | DEAT | FPE | FPE |
| 1 | 68 | -65 | 5 | -31 | 95 | 72 |
| 2 | 30 | -59 | 10 | -47 | 66 | 23 |
| 3 | -78 | -45 | -25 | -58 | 90 | 55 |
| 4 | -9 | -167 | -41 | -40 | 167 | 131 |
| 5 | 25 | 20 | -8 | -45 | 32 | 73 |
| 6 | -40 | -60 | -26 | -37 | 72 | 27 |
| 7 | -35 | -64 | -10 | -57 | 73 | 26 |
| 8 | -75 | 38 | -68 | -38 | 84 | 76 |
| 9 | -6 | 17 | -10 | -37 | 18 | 54 |
| 10 | -12 | -18 | -5 | -34 | 22 | 17 |
| 11 | -12 | -30 | 0 | -42 | 32 | 17 |
| 12 | -102 | -10 | -35 | -43 | 102 | 75 |
| 13 | -73 | -210 | -22 | -39 | 222 | 178 |
| 14 | -22 | -78 | -20 | -45 | 81 | 33 |
| 15 | -65 | -252 | -24 | -45 | 260 | 211 |
| 16 | -42 | -18 | -43 | -33 | 46 | 15 |
| 17 | 67 | 0 | -22 | -101 | 67 | 135 |
| 18 | -5 | -45 | -1 | -38 | 45 | 8 |
| 19 | 35 | 76 | -23 | -35 | 84 | 125 |
| 20 | -50 | -52 | -69 | -34 | 72 | 26 |

a high degree of confidence one can conclude that the Forecast Errors are smaller using the Modifying Model. The Paired T test results are given below:

$$T = \frac{\overline{D} - 0}{S/\sqrt{n}},$$

where

$D_i =$ NOGAPS $FPE_i$ - MODIFYING MODEL $FPE_i$ ,

$\overline{D} = (\sum_{i=1}^{n} D_i)/n =$ 17.65,

$$S^2 = [\sum_{i=1}^{n}(D_i - \overline{D})^2]/(n-2) \quad = \quad 1279.5,$$

$$S/\sqrt{n} = 7.999,$$

$$T = 2.206, \ v = 19,$$

and

$$\alpha < .02 \ .$$

The validation results can be analyzed in several other ways. For example, the Forecast error is smaller using the Modifying Model in 16 out of the 20 cases for an effectiveness rating of 80 percent.

Additionally, the NOGAPS prediction can be viewed as the center of an (x,y) coordinate system, with four quadrants corresponding to each pair of possible positive and negative values of the TKE and DEAT (Figure 6). Each Modifying Model estimate adjusts the NOGAPS prediction and establishes a position in one of these four quadrants. In 14 out of 19 cases in the validation set, the Modifying Model places the cyclone position in the correct quadrant. In the remaining case, the Track Error estimate is zero but the DEAT estimate is in the correct direction.

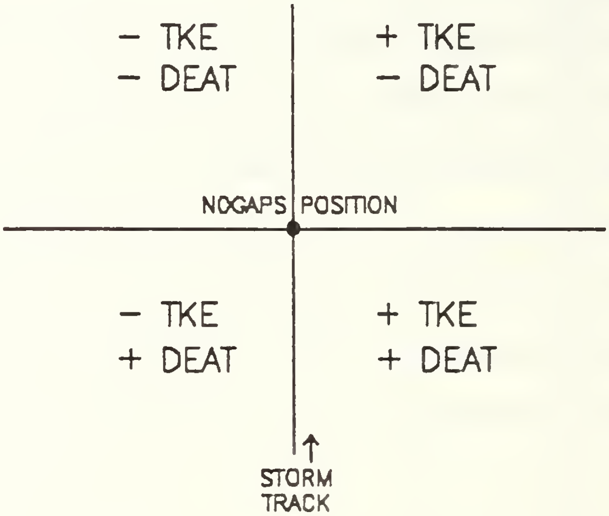


Figure 6. Illustration of Quadrants in Position Estimation

## C. TRACK ERROR ANALYSIS

The Modifying Model did well in correctly anticipating the Track Error direction. In 15 out of 19 cases, the model forecast a track error on the correct side of track. In the remaining case, the Modifying Model forecast no track error when the actual is a small error left of track. Anticipated Track Error direction can be particularly useful if decisions must be made concerning manuevering around a cyclone. In each of the four cases in which the track error direction is incorrectly forecast, the actual track error is to the right of track (positive). The model only forecast track errors to the right of track twice and was correct both times. Thus, if a cyclone in the validation set had a track error left of track (negative) it was correctly forecast 100 percent of the time and, if the Modifying Model forecast a Track Error right of track, it was correct 100 percent of the time. Obviously, these occurrences do not represent absolutes but do give some insight into the workings of the Track Error model. A Track Error prediction right of track is less common, but more likely to be correct when it is made.

The Modifying Model Track Error estimate is closer to the actual track error in 14 of the 20 cases, with four of the six due to incorrect predictions of the track error direction. Six predictions are within 10nm of the actual track error and in the correct direction. Another five are within 20nm.

The Track Error between the Modifying Model prediction and the actual position can be computed by subtracting the Track Error estimate of the Modifying Model (column three from Table 3) from the Track Error of the NOGAPS prediction (column one from Table 3). The sum of these differences can be compared to the sum of the errors from column one to determine the amount of error reduction which is achieved. The result is 23 percent.

## D. DISTANCE ERROR ALONG TRACK ANALYSIS

The DEAT model predicts the correct direction of the Distance Error Along Track 15 out of 20 times. Each time there is a positive DEAT, the Modifying Model produces an incorrect prediction. The Modifying Model DEAT produces an estimate closer to the actual DEAT in 14 of the 20 cases. Five of the six poorly forecast cases are those in which the actual DEAT is positive.

Of the fifteen cases in which the direction of the DEAT is correctly forecast, the Modifying Model overforecast the DEAT in six cases and underforecast the DEAT in nine. Of the nine cases which are underforecast, three involve cases in which the actual DEAT is larger than 160nm (in the negative direction). Therefore, among the cases in which the actual DEAT is moderate, there is no discernable bias in the DEAT forecasts of the Modifying Model.

The DEAT component of the error between the Modifying Model estimate and the actual position can be computed by subtracting the DEAT estimate of the Modifying Model (column four of Table 3) from the DEAT of the NOGAPS forecast (column two of Table 3). The sum of these differences can be compared to the sum of the actual DEAT's (column one) to determine the amount of DEAT reduction which is achieved. The result is a 13% reduction.

## E. PROBABILITY ELLIPSE PERFORMANCE

The performance of the probability ellipses discussed in Chapter III is analyzed on all 20 cases in the validation set. Three different probability ellipses are studied: 50, 75 and 95 percent. Eleven out of 20 cyclone positions are contained within the 50 percent ellipse, for a performance rate of 55 percent. Fifteen are within the 75 percent ellipse (75 percent performance rate) and 18 are contained within the 95 percent ellipse (90 percent performance rate). These results are very favorable. See Figure 7.

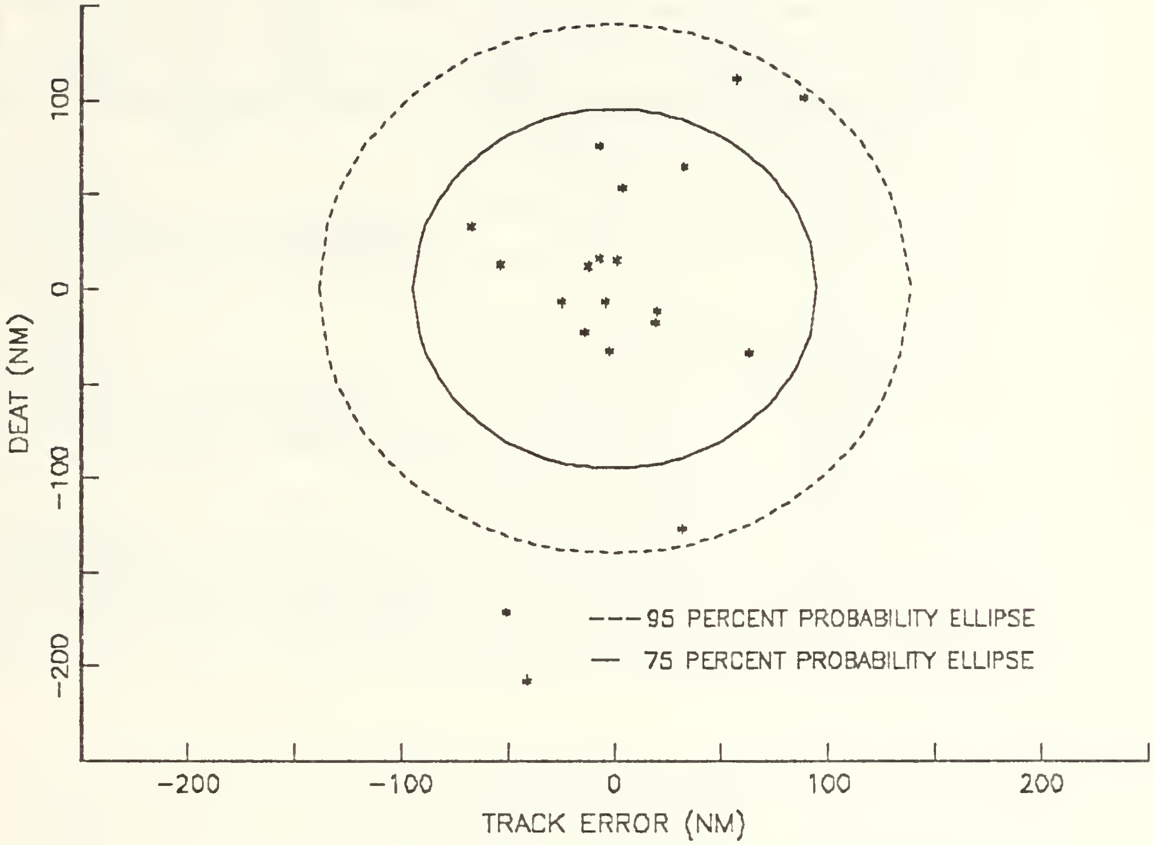


Figure 7. Probability Ellipse Performance

Probability ellipses about the NOGAPS estimates contain more of the actual cyclone positions than the Modifying Model estimates. However, this is attributable to the much larger probability ellipses about the NOGAPS model estimates. Ellipse size comparisons are listed below in Table 4.

Table 4. NOGAPS AND MODIFYING MODEL ELLIPSE COMPARISON

| | NOGAPS | | Modifying Model | |
|---|---|---|---|---|
| $\alpha$ level | Major Axis | Minor Axis | Major Axis | Minor Axis |
| 0.5 | 103 | 77.6 | 66.8 | 67.2 |
| 0.25 | 145.7 | 109.8 | 94.4 | 95.1 |
| 0.05 | 214.1 | 161.3 | 138.7 | 139.7 |

# V. CONCLUSION

The modelling introduced in this thesis has a positive impact on the accuracy of NOGAPS cyclone position predictions. Using several different Measures of Effectiveness, summarized in Table 5, the modelling reduces the magnitude of the component errors by approximately 20 percent and anticipates the error tendencies in their various forms with an accuracy rate of approximately 75 percent.

Table 5. MOE SUMMARIES FROM THE VALIDATION DATA

| MOE | RATIO | Percentage |
|---|---|---|
| FPE Reduction | N/A | 20% |
| Position Estimate in Correct Quadrant | 14/19 | 74% |
| Modifying Model FPE < NOGAPS FPE | 16/20 | 80% |
| TKE Reduction | N/A | 23% |
| TKE Prediction in Correct Direction | 15/19 | 79% |
| Modifying Model TKE < NOGAPS TKE | 14/20 | 70% |
| DEAT Reduction | N/A | 13% |
| DEAT Prediction in Correct Direction | 15/20 | 75% |
| Modifying Model DEAT < NOGAPS DEAT | 15/20 | 75% |

A simple illustration provides the best evidence of how the modelling can have a significant impact on the decision making process of the at sea commander when a manuevering decision must be made in response to the presence of a cyclone. Figure 8 depicts a hypothetical situation in which the present position of a storm, its 24 hour predicted position and therefore, its predicted course, and ship's position are plotted. Additionally, a circle of radius 150nm representing the hypothetical radius of 30 kt winds about the predicted 24 hour position is, also, plotted. The cyclone is predicted to move at a speed of 18nm per hour, perpendicular to the ship's intended course. The ship is 510nm from the 24 hour predicted postion and desires to procede at 15kts. If the ship does not alter course and/or speed it will enter the circle of 30 kt winds at the exact time the storm is forecast to reach the 24 hour position. Clearly, this is not prudent and a manuevering decision must be made.

Prudent seamanship precludes speeding up and manuevering to cross ahead of the intended cyclone path. Therefore, the decision involves how best to let the cyclone pass ahead and then manuever to regain the desired track. Using only the plot in Figure 8, the shipboard commander may choose to maintain course and reduce speed enough to remain safely outside the 30kt envelope or he may choose to adjust course to the right and steer for the bottom edge of the 30kt envelope.

In Figure 9, the Modifying Model estimate is applied to the NOGAPS position and a 75 percent probability ellipse is constructed about the Modifying Model position. The estimated range of 30 kt winds (150nm) is plotted from the boundary of the probability ellipse. The picture has now changed dramatically. What seemed to be a safe manuever, given the first illustration, may be hazardous. If remaining outside the circle of 30 kt winds is very important to the shipboard commander, a better course of action would be to loiter in the present area or turn futher right to more safely pass astern of the cyclone.

The plots utilizing the Modifying Model are easy to construct following a simple checklist and, easy to read. They provide the decision maker with more information upon which to base his manuevering decisions and do not seek to replace the NOGAPS position predictions. They work in conjunction with NOGAPS to provide the best possible information.

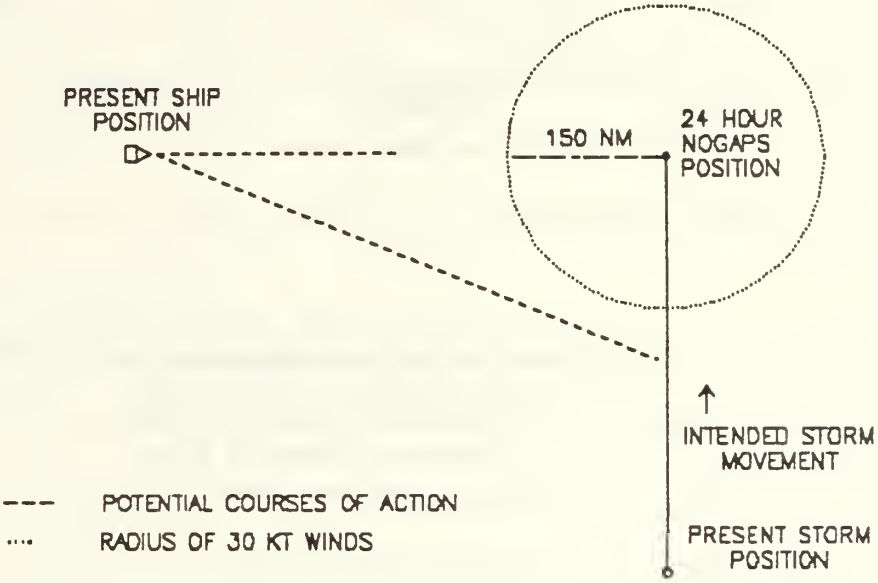


Figure 8. Decision Making Plot using only NOGAPS

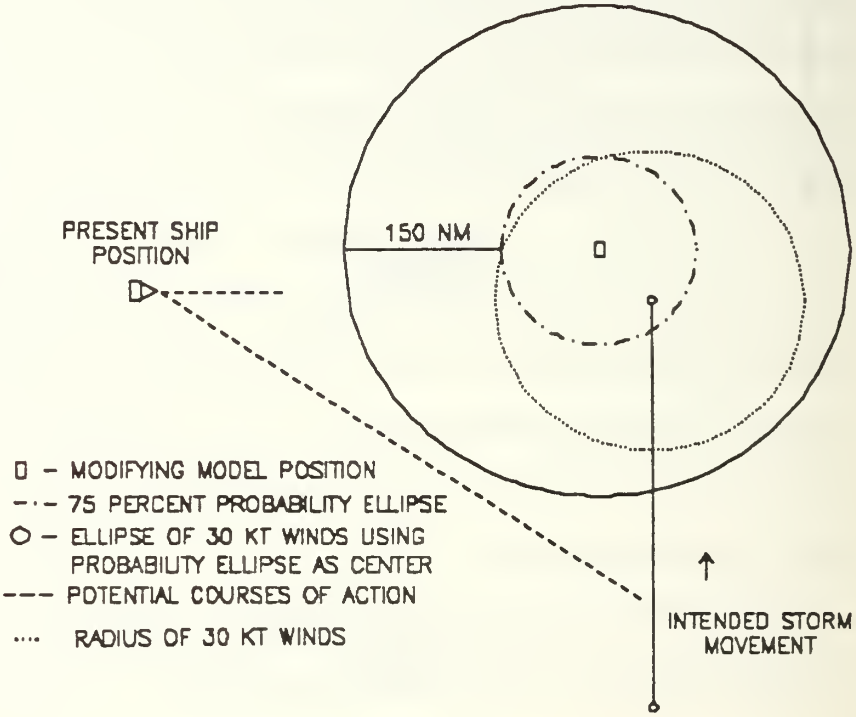


Figure 9. Decision Making Plot with Modifying Model Enhancement

# APPENDIX A. SAMPLE CYCLONES

The following are descriptions of four cyclones which were considered for inclusion in the data base. Tables 6 and 7 depict cyclones which were accepted into the data base and Tables 8 and 9 represent cyclones which were rejected. The predictions are for 24 hours from the time of the associated observation. The accuracy of a prediction can be evaluated by comparing it to the observation recorded 24 hours later. For example, using Table 6, at 0000Z on 28 October 1989 the central pressure was observed to be 1007mb. The prediction of central pressure for 24 hours from that time (0000Z, 29 October) is 1001mb. The actual central pressure at that time is 1002mb which corresponds to the observation at 89102900. Thus, the error of the central pressure prediction is – 1mb.

Table 6 displays a cyclone which does not have a deepening phase of 48 hours. It is accepted into the data base because the third 24 hour forecast issued at 89102900 calls for continued deepening below the observed central pressure at that time. The cyclone in Table 7 is accepted because the deepening phase is 48 hours. The cyclone in Table 8 is rejected because the third 24 hour prediction at 90031712 calls for the central pressure to increase. Additionally, the deepening phase of the cyclone is only 12 hours. The cyclone in Table 9 has a deepening phase of only 24 hours and the third 24 hour forecast at 89100912 predicts filling (increasing central pressure). It is therefore, rejected.

Table 6. ACCEPTED CYCLONE

| | YY/MM/DD/HH | CENTRAL PRESSURE |
|---|---|---|
| OBSERVATION | 89102800 | 1007 |
| PREDICTION | 89102800 | 1001 |
| OBSERVATION | 89102812 | 1004 |
| PREDICTION | 89102812 | 998 |
| OBSERVATION | 89102900 | 1002 |
| PREDICTION | 89102900 | 996 |
| OBSERVATION | 89102912 | 985 |
| PREDICTION | 89102912 | 993 |
| OBSERVATION | 89103000 | 986 |
| OBSERVATION | 89103012 | 989 |

Table 7. ACCEPTED CYCLONE

| | YY/MM/DD/HH | CENTRAL PRESSURE |
|---|---|---|
| OBSERVATION | 89112212 | 1007 |
| PREDICTION | 89112212 | 999 |
| OBSERVATION | 89112300 | 1001 |
| PREDICTION | 89112300 | 1000 |
| OBSERVATION | 89112312 | 998 |
| PREDICTION | 89112312 | 993 |
| OBSERVATION | 89112400 | 996 |
| OBSERVATION | 89112412 | 992 |
| OBSERVATION | 89112500 | 995 |

Table 8. REJECTED CYCLONE

| | YY/MM/DD/HH | CENTRAL PRESSURE |
|---|---|---|
| OBSERVATION | 90031612 | 992 |
| PREDICTION | 90031612 | 978 |
| OBSERVATION | 90031700 | 970 |
| PREDICTION | 90031700 | 976 |
| OBSERVATION | 90031712 | 970 |
| PREDICTION | 90031712 | 981 |
| OBSERVATION | 90031800 | 971 |
| PREDICTION | 90031800 | 982 |
| OBSERVATION | 90031812 | 977 |
| PREDICTION | 90031812 | 990 |
| OBSERVATION | 90031900 | 987 |
| OBSERVATION | 90031912 | 993 |

Table 9. REJECTED CYCLONE

| | YY/MM/DD/HH | CENTRAL PRESSURE |
|---|---|---|
| OBSERVATION | 89100812 | 984 |
| PREDICTION | 89100812 | 966 |
| OBSERVATION | 89100900 | 968 |
| PREDICTION | 89100900 | 971 |
| OBSERVATION | 89100912 | 962 |
| PREDICTION | 89100912 | 978 |
| OBSERVATION | 89101000 | 965 |
| PREDICTION | 89101000 | 989 |
| OBSERVATION | 89101012 | 974 |
| PREDICTION | 89101100 | 993 |
| OBSERVATION | 90031900 | 987 |
| OBSERVATION | 90031912 | 993 |

# APPENDIX B. OUTLIER REMOVAL

## A. DISTANCE ERROR ALONG TRACK MODEL

Starting with the Distance Error Along Track Model and referencing the plots of studentized residuals (Figure 10), all points are roughly within two standard deviations of the mean with the exception of three. From the complete data summary listed in Appendix D, these points correspond to cyclones 9, 14 and 41. Cyclones 9 and 14 are two of the three poorest forecasts in the entire data base and contain the second and third largest DEAT's. Unlike the largest DEAT in the data set (cyclone 44), which has an extremely high value of Slope and, therefore, a better prediction resulting in a small residual value, the Slope values for cyclones 9 and 14 are moderate and small, respectively.

The DEAT for cyclone 41 is over 50 percent larger than the next largest positive DEAT in the data set, which itself is 66 percent larger than the third largest positive DEAT in the data set. It was removed to leave a data set which contains all residuals within two standard deviations of the mean.

Additionally, Figure 10 shows two plots of the residuals versus the Normal Distribution for all 57 observations. Both the Chi-Squared and Kolmogorov-Smirnov statistics support rejection of the assumption of normal residuals if $\alpha = .10$ . When the three cyclones are removed from the data base, the residual plots are greatly enhanced (see Figure 11) and the hypothesis of normal residuals can not be rejected at any reasonable significance level.

## B. TRACK ERROR MODEL

With respect to the Track Error model, two cyclones, 45 and 51, are considered to be outliers. The TKE of Storm 45 is the largest TKE in the data base by 47nm and 23 percent larger than the second largest. It corresponds to a small negative TKE1 (-25) and ICP (1006mb) right at the average. Cyclone 51 has one of the highest positive TKE's, but not the largest. It is unique because no other cyclone in the data base with a positive TKE has as large a negative TKE1. In fact, only 9 out of the 28 cases with positive TKE's had negative values of TKE1. Cyclone 1 has the next most radical difference between a negative TKE1 and positive TKE. Clearly, this difference does not have the magnitude of that of cyclone 51.

Additionally, residual plots (Figure 12) show both cyclones 45 and 51 to have residuals larger than 2.5 standard deviations from the mean, while all others plot within two standard deviations. The residual plots for the TKE model versus the normal distribution support the normality assumption.

Finally, Figure 13, shows the residual plots of the TKE model with storms 45 and 51 removed (52 data points) and Figure 14 displays the residual plots for the DEAT model after all outliers have been removed.

## C. SUMMARY OF MODEL SPECIFIC DATA

The following two tables summarize the changes in the major regression parameters over the course of the process of outlier removal. Note that as the data base goes from 57 to 54 points, outliers affecting the DEAT model are being removed, while in moving from 54 points to 52 data points outliers affecting the TKE model are removed. Therefore, the model parameter changes in Table 10 are more dramatic between 57 and 54 data points, and the more dramatic changes occur between 54 and 52 data points in Table 11.

Table 10. CHANGES IN DEAT MODEL AS OUTLIERS ARE REMOVED

| | DEAT 57 pts | DEAT 54 pts | DEAT 52 pts |
|---|---|---|---|
| $B_0$ | -36.42 | -31.88 | -32.24 |
| $B_1$ | -2.58 | -2.57 | -2.58 |
| $R^2$ | 0.2392 | 0.3511 | 0.3569 |
| S | 73.12 | 56.99 | 57.52 |
| F (df) | 17.291 (2,54) | 28.136 (2,51) | 27.75 (2,49) |
| Sig Level | 0.000113 | 0.000002 | 0.000003 |

Table 11. CHANGES IN TKE MODEL AS OUTLIERS ARE REMOVED

| | TKE 57 pts | TKE 54 pts | TKE 52 pts |
|---|---|---|---|
| $B_0$ | -2085.6 | -1895.6 | -2344.2 |
| $B_1$ | 0.318 | 0.304 | 0.325 |
| $B_2$ | 2.05 | 1.86 | 2.31 |
| $R^2$ | 0.1838 | 0.1739 | 0.2614 |
| S | 66.99 | 65.14 | 56.19 |
| F (df) | 6.08 (3,53) | 5.37 (3,50) | 8.67 (3,48) |
| Sig Level | 0.0042 | 0.0077 | 0.0006 |

## D. STATISTICAL TEST FOR OUTLIERS

A statistical method is used to test each data point's significance as an outlier to bolster the assertions made in the preceeding paragraphs concerning outlier removal. The test involves deleting one data point to determine the difference between the model coefficient estimates without this one data point (using 51 data points) and model coefficient estimates using all 52 data points [Ref 6 : pp. 113-117]. This process is repeated for each point in the data set. A T-Statistic is generated for determining the significance level of this difference. Fortunately, a general formula can be used to compute the T-statistic for each data point without having to run n different regressions. The variables and formulation are listed in the following:

n = number of distinct data points (57),

$p'$ = # independent variables + 1 (intercept term),

$x_i = p' \times 1$ matrix of independent variables for the ith case,

$B_{-i}$ = regression coefficients generated from n-1 cases excluding the ith,

$y_i$ = ith response variable,

$\tilde{y}_i$ = estimate of the ith response using regression estimates from n-1 cases (excluding the ith case),

$\tilde{y}_i = x_i \hat{B}_{-i}$,

$\hat{\sigma}^2_{-i}$ = variance estimate with the ith observation removed,

and

$\hat{Var}(\tilde{y}_i) = \hat{\sigma}^2_{-i} x_i (X^T_{-i} X_{-i})^{-1} x_i^T$.

If there are no outliers then, $E(y_i - \tilde{y}_i) = 0$ for all i. A Student's T-test of the hypothesis $E(y_i - \tilde{y}_i) = 0$ is as follows:

$$t_i = \frac{y_i - \tilde{y}_i}{\hat{\sigma}_{-i}\sqrt{1 + x_i (X^T_{-i} X_{-i})^{-1} x_i}}, \quad v = n - p' - 1.$$

The T-statistic formula can be converted to the following general formula

$$t_i = r_i \left( \frac{n - p' - 1}{n - p' - r_i^2} \right)^{1/2}$$

where $r_i$ = ith Studentized residual.

This T-statistic is computed for each of the 57 points in the data set for each model. Each of the data points mentioned previously as outliers (three in the DEAT model and two in the TKE model) generates T-statistics which are significant at the $\alpha = .02$ level. No other points in either model obtain this level of significance.

## E. TEST FOR INFLUENTIAL POINTS

Due to the nature of the regression plot of DEAT versus Slope (see Figure 15), there is some concern over the influence on the model, of the two large, negative values of DEAT, corresponding to two large values of Slope (Cyclones 44 and 50). Cook's Distance formula is used to determine if either point has an unduely large influence on the model. The formula quantifies, in a meaningful way, the magnitude of the difference between the model estimates of the regression coefficients with all data points and the coefficient estimates excluding the ith. Cook's Distance, $D_i$, roughly represents the squared distance between the two estimates scaled by the variance of the data [Ref. 6: pp. 106-109.]. The larger the value of $D_i$ the larger the influence of the data point. The general rule is that if the value of $D_i$ is larger than 1, the ith case may be judged to be influential.

The general formula for computing $D_i$ follows:

$$D_i = \frac{1}{p'} r_i^2 \left( \frac{v_{ii}}{1 - v_{ii}} \right).$$

where

$v_{ii} = x_i(X_{-i}^T X_{-i})x_i^T$ = diagonal elements of the Hat matrix,

and

$\frac{v_{ii}}{1 - v_{ii}}$ = the distance from the ith, deleted data point, to the means of the remaining, n-1 data points.

If the distance from the deleted data point to the n-1 means is large, a large value of $D_i$ may result. A large value of $D_i$ may also occur when $r_i$ is large. While the values of $D_i$ for the two data points in question are high relative to the other data points, they are not close to 1.00 (.79 and .58) and therefore, are not determined to be influential cases.

The TKE Model was not tested for influential points since the residual plots followed a much more symmetrical and traditional pattern.

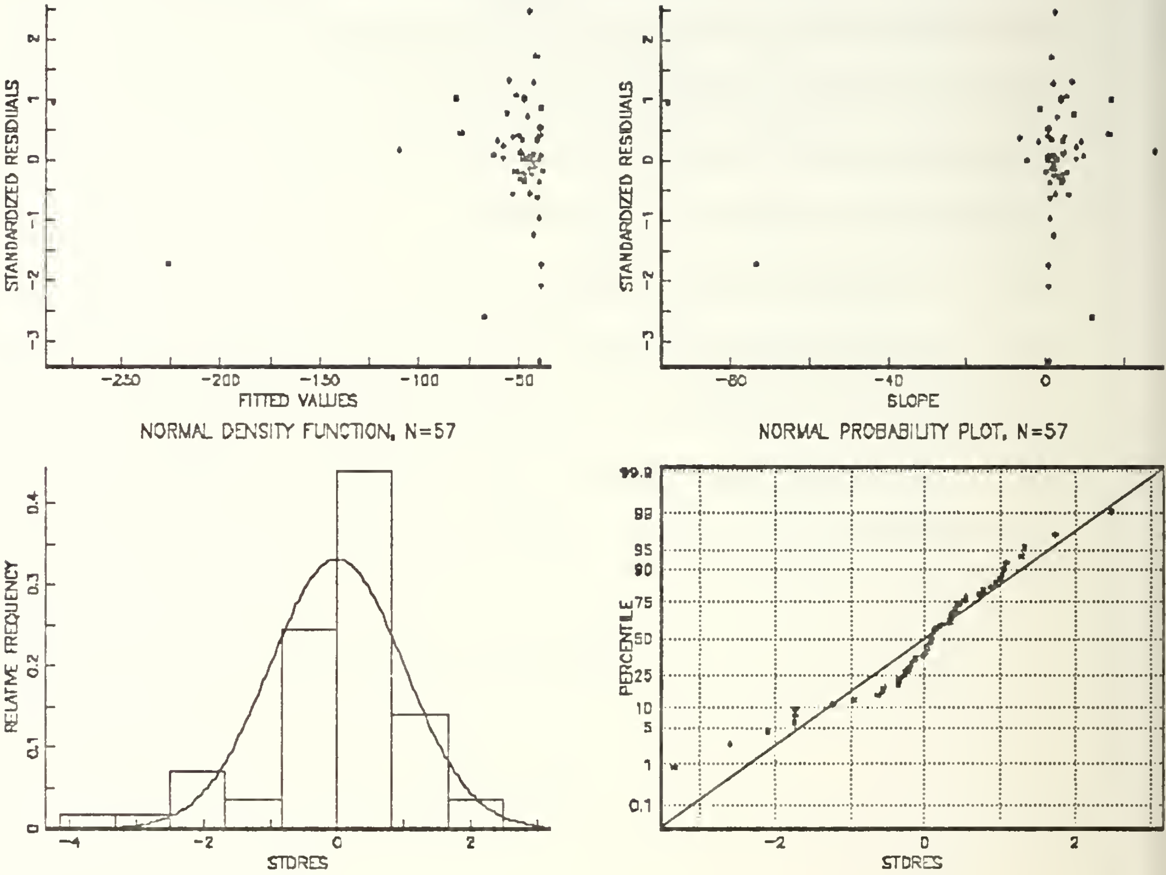


Figure 10. DEAT Model - 57 Data Points

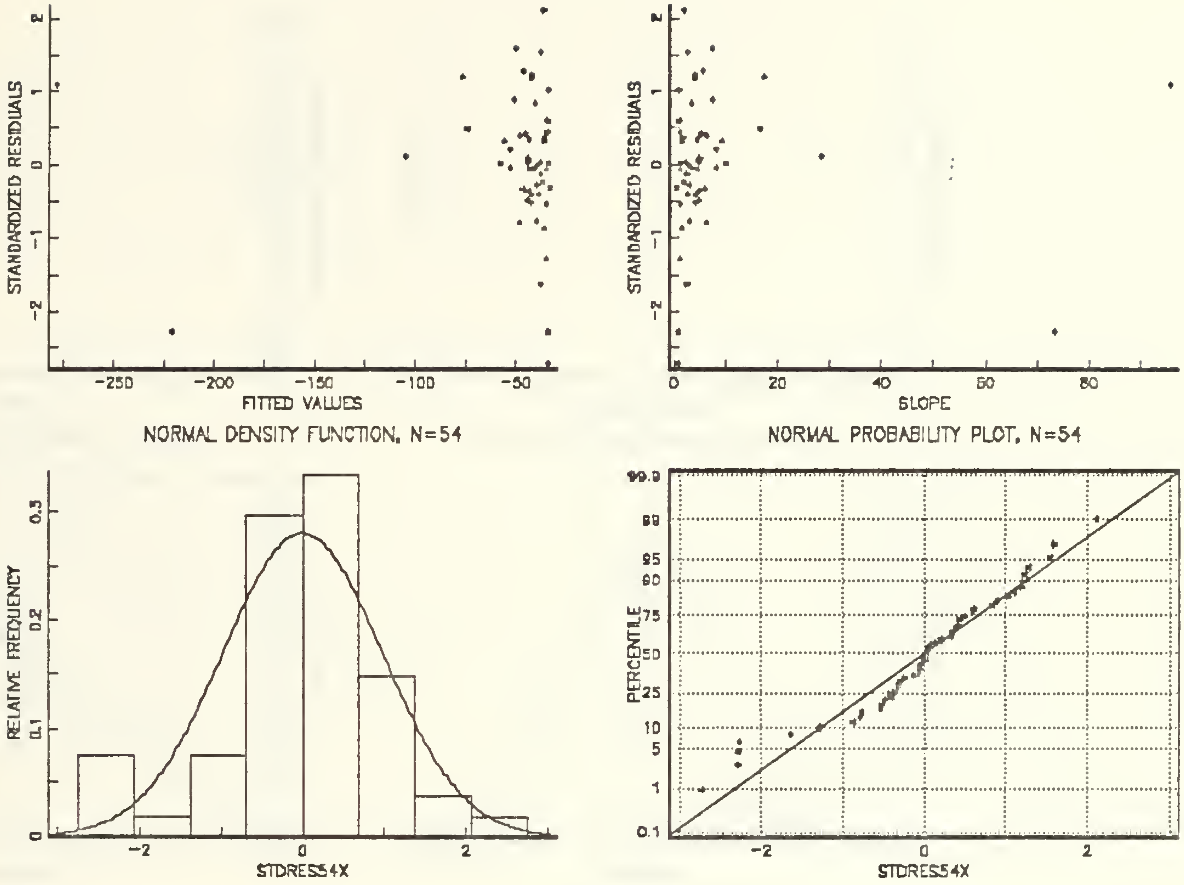
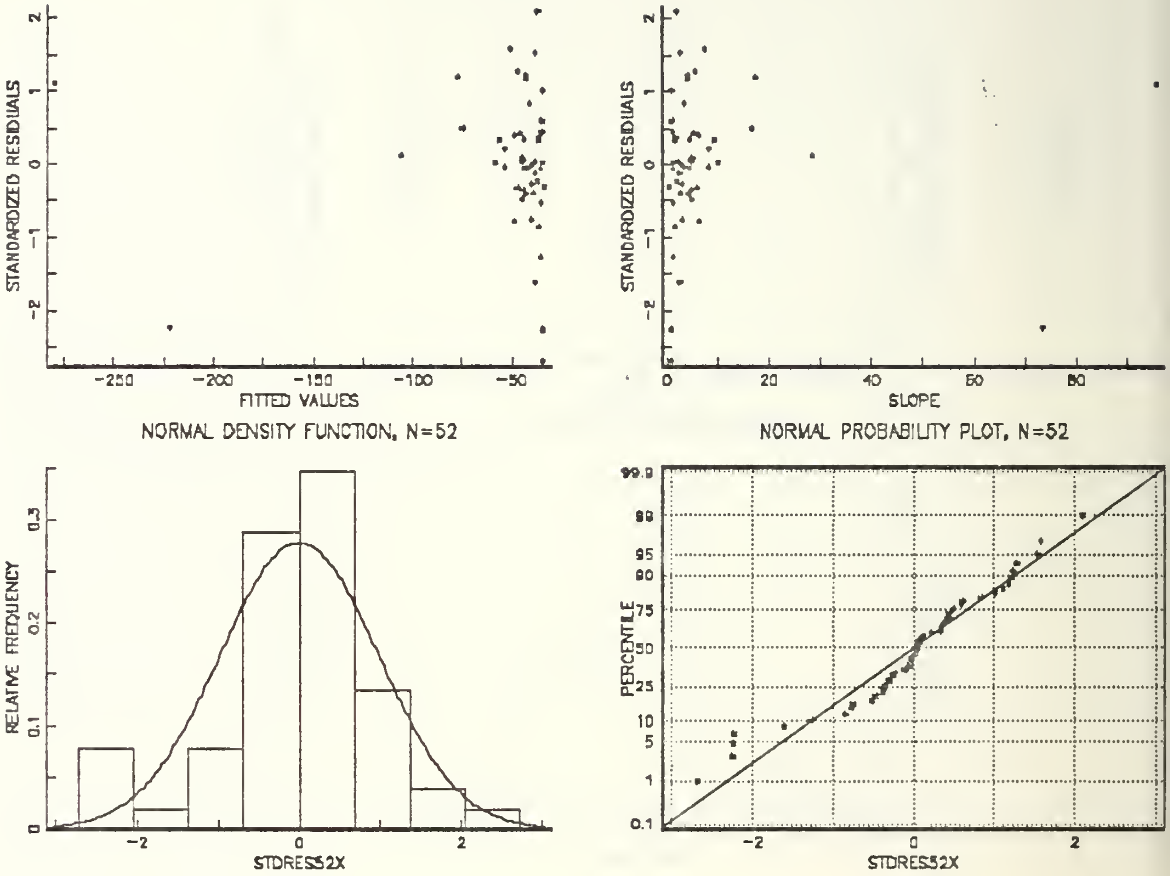


Figure 11. DEAT Model - 54 Data Points



NORMAL DENSITY FUNCTION, N=52

NORMAL PROBABILITY PLOT, N=52

Figure 12. DEAT Model - 52 Data Points

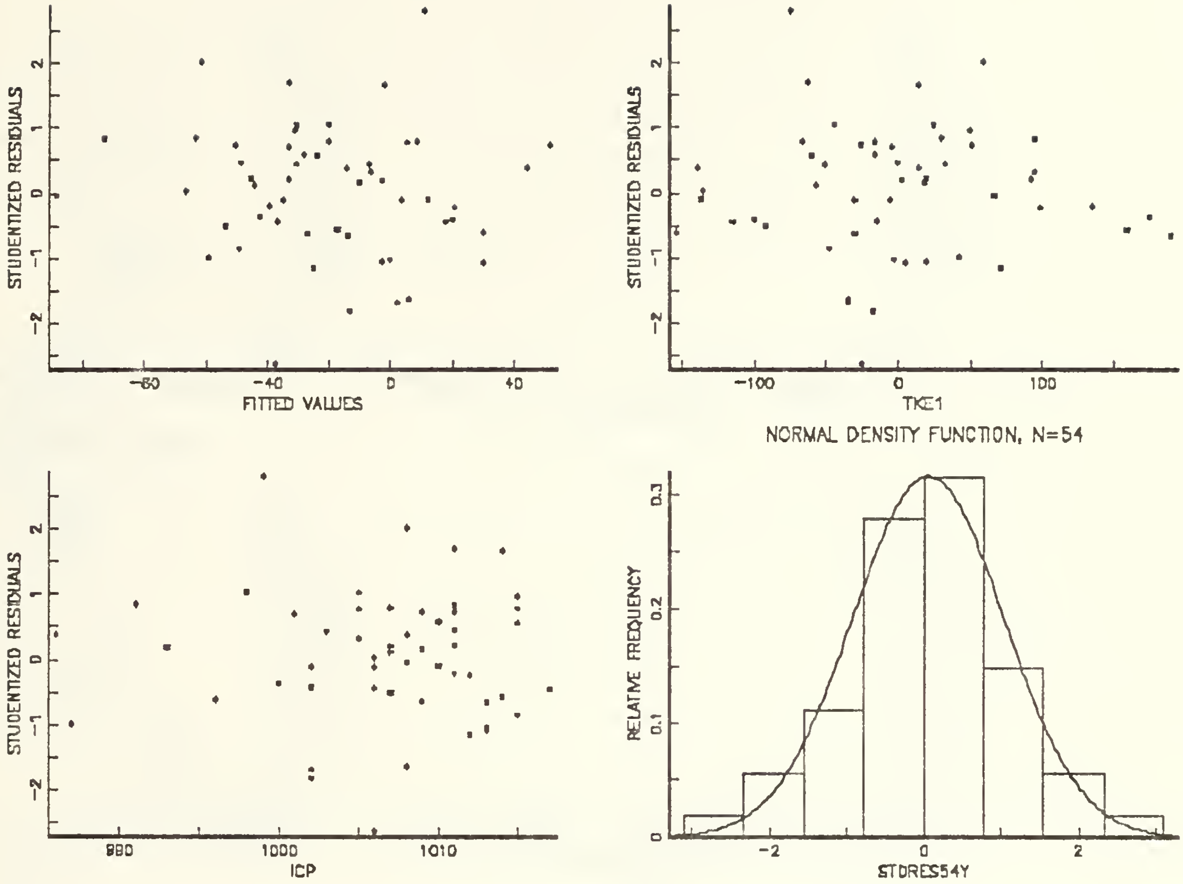


Figure 13. TKE Model - 54 Data Points

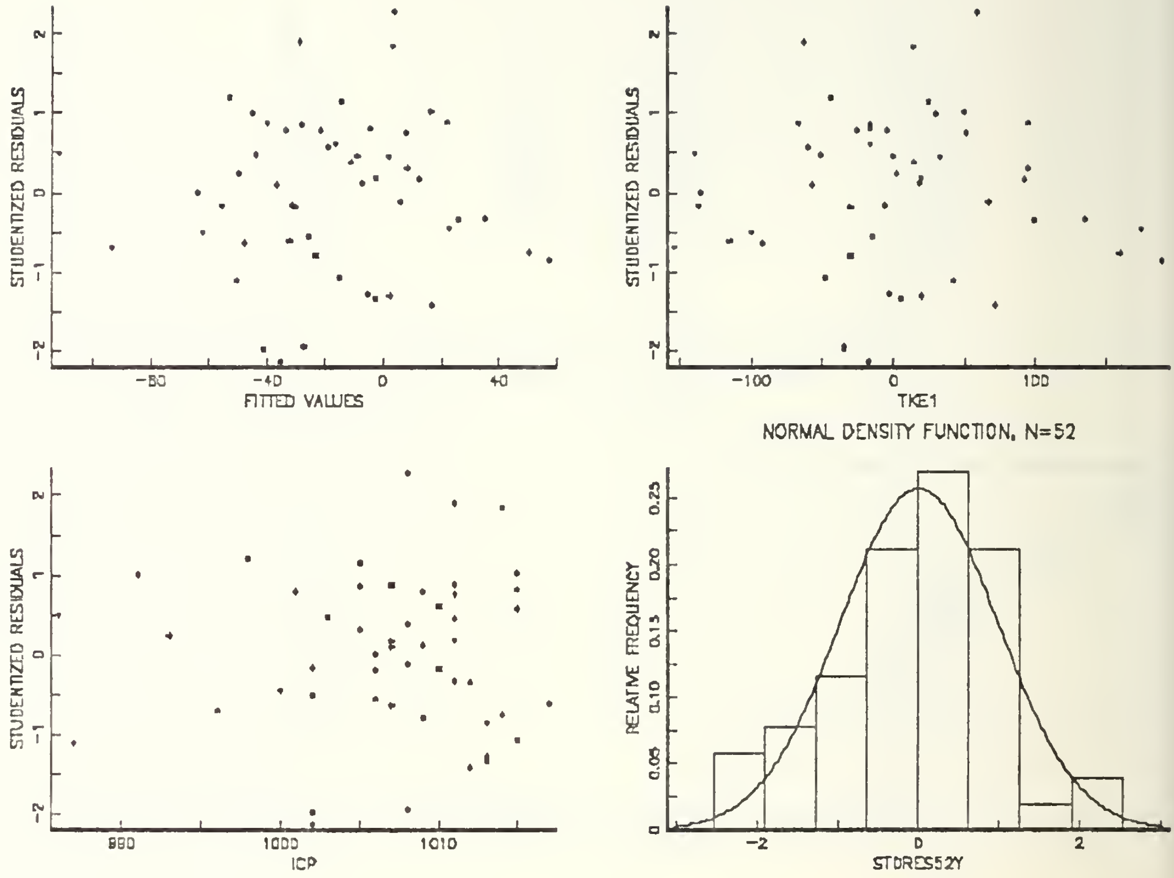


Figure 14. TKE Model - 52 Data Points

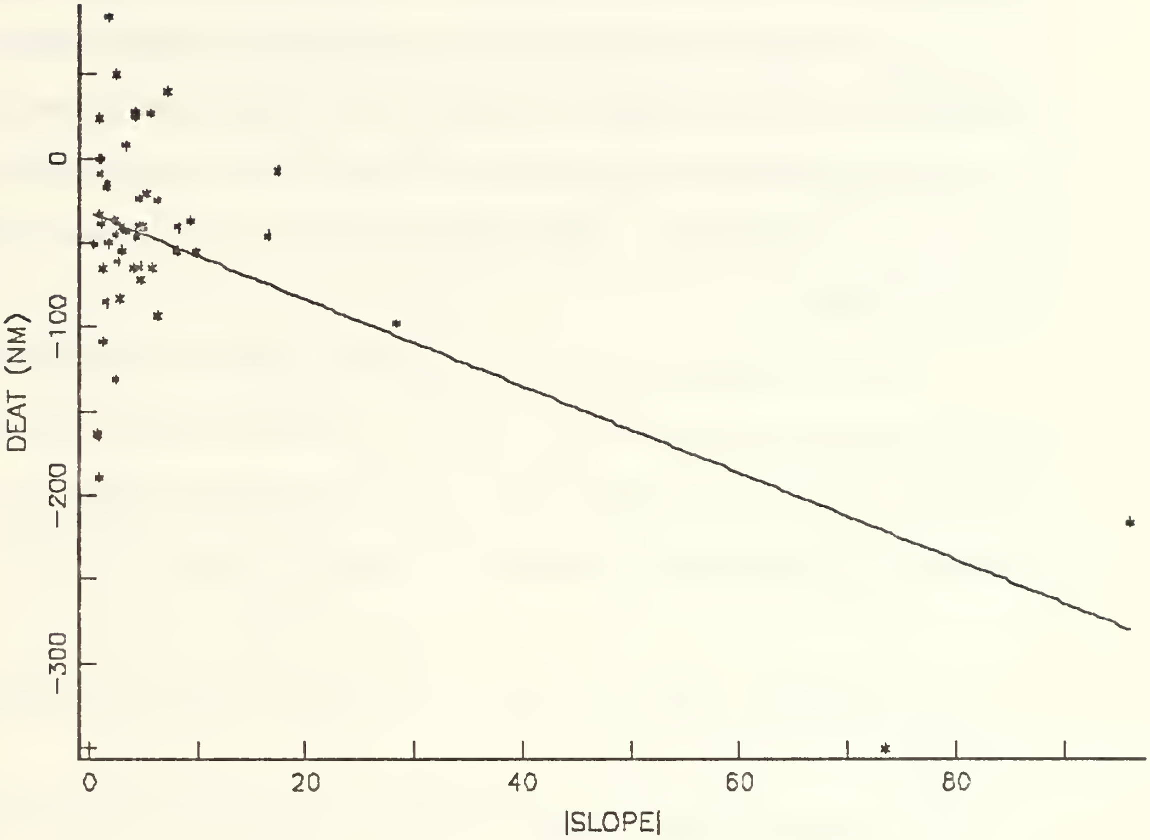


Figure 15. DEAT Regression Model Scatter Plot

# APPENDIX C. MULTIVARIATE ANALYSIS

As mentioned in Chapter II, a multivariate analysis has been conducted in order to examine the effect of the correlation between the DEAT and TKE [Ref. 7 ] and for comparison with the model which was adopted. The adopted model was tested more thoroughly and assumes independence between the TKE and DEAT. The correlation between the NOGAPS TKE and DEAT was 0.19. The correlation between the TKE and DEAT, after applying the Modifying Model estimate, is 0.09. Due to these low correlation values, particularly in the Modifying Model case, the independence assumption is considered tenable.

Each of the three independent variables, TKE1, ICP and SLOPE, used in the regression models previously developed, is used in a composite analysis to estimate TKE and DEAT. The same 52 data points and Jackknife procedure is employed. The final regression model coefficients for the multivariate regression model are as follows:

**DEAT Composite:** $\hat{DEAT} = -1624.7 + 0.31(TKE1) + 1.6(ICP) - 0.2(SLOPE)$

and

**TKE Composite:** $\hat{TKE} = -1091.6 + 0.04(TKE1) + 1.05(ICP) - 2.1(SLOPE)$.

Cross validation is conducted using the 52 pairs of residuals from the jackknife technique. The residuals are plotted and have the same general pattern as the model assuming independence between TKE and DEAT (see Figure 16). Figure 16, also, shows a comparison between a 75 percent multivariate probability ellipse (A) and the independence model 75 perecent probability ellipse (B). The multivariate ellipse is tilted slightly to account for the correlation between the TKE and DEAT. Thus, unlike the previously developed ellipse, which is oriented along the intended cyclone track, the multivariate ellipse is tilted off the intended cyclone track and, is more difficult to plot. Also, the multivariate probability ellipse appears less like a circle.

As indicated in Figure 16, the multivariate model does not appear to perform as well as the adopted model. The 75 percent probability ellipse contains only 69 percent of the points (36 out of 52). Figure 17 is a composite of figure 16A and 16B with the residual pairs from both models plotted. An association between model residuals seems to exist when the residuals are large in both models. There appears to be less association when the residuals are small. Figure 17, also, shows that the multivariate ellipse covers slightly less area than the independent models' ellipse. Proportionately increasing the multivariate ellipse to the same size would place only one more residual pair in the ellipse and not raise the percentage to 75 percent.

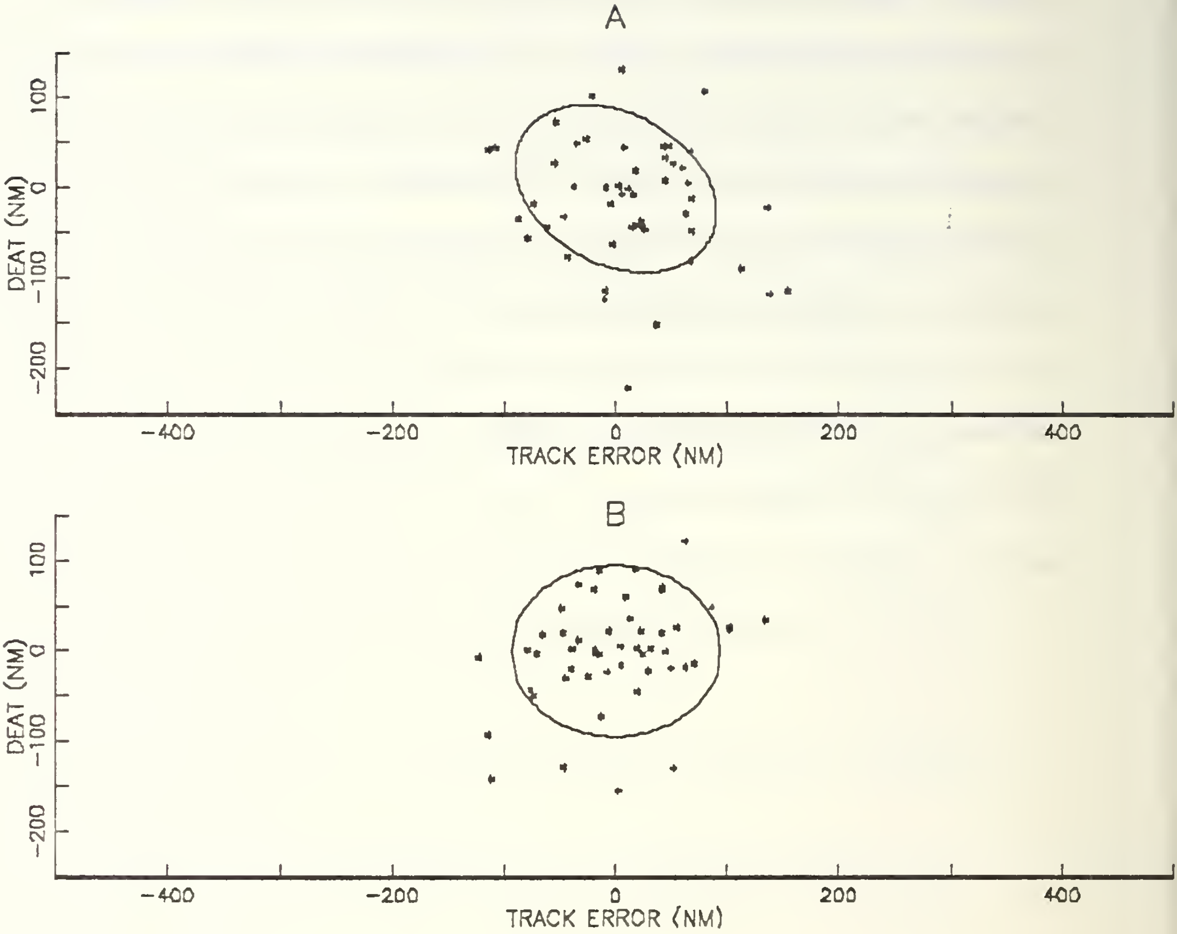


Figure 16. Multivariate Probability Ellipse Comparison

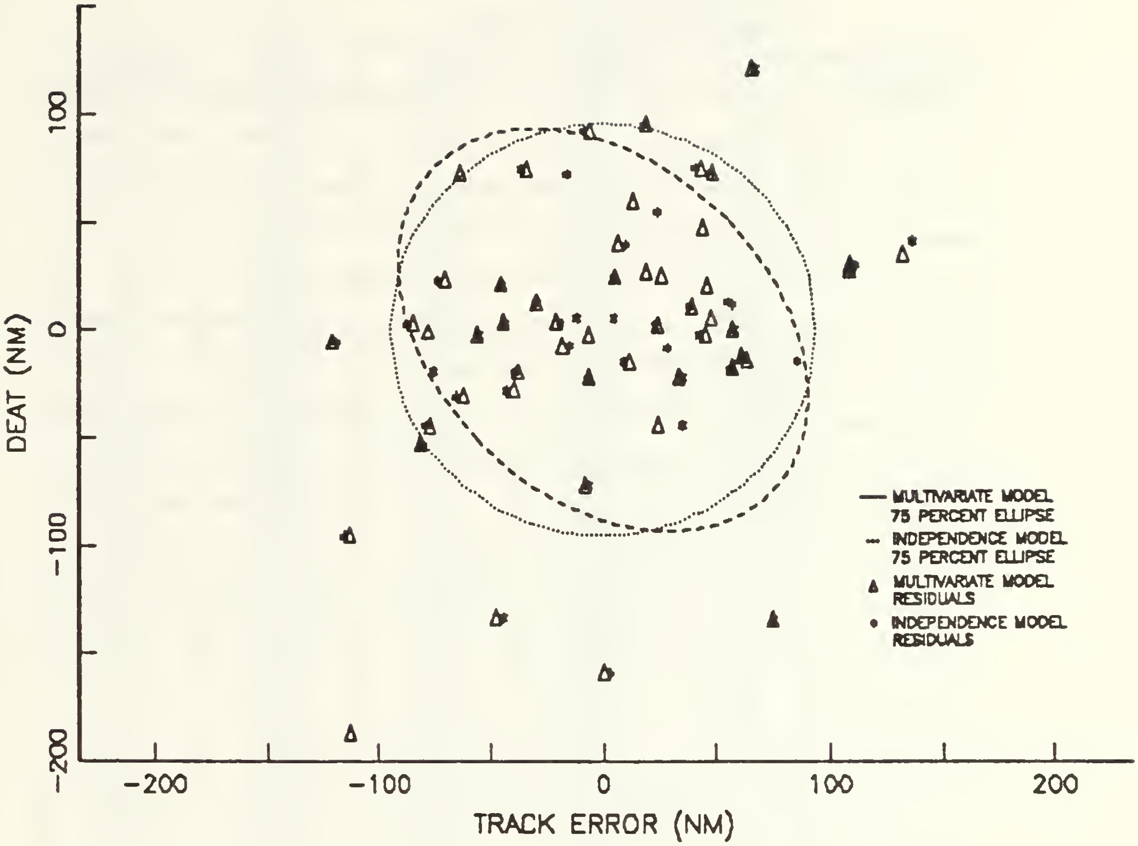


Figure 17. Figures 16A and 16B superimposed

# APPENDIX D. DATA BASE

The following table gives a complete list of all five variables for each of the 57 cyclones in the original data base.

Table 12. COMPLETE DATA SET

| Cyclone | TKE | DEAT | TKE1 | ICP | SLOPE |
|---|---|---|---|---|---|
| 1 | 77 | -46 | -52 | 1011 | 16.5 |
| 2 | -70 | -85 | 20 | 1013 | 1.51 |
| 3 | 13 | -61 | -60 | 1015 | 2.575 |
| 4 | 7 | -35 | 100 | 1012 | 2.38 |
| 5 | 11 | -93 | 15 | 1008 | 6.35 |
| 6 | 106 | -21 | 15 | 1014 | 5.36 |
| 7 | -67 | -23.5 | -30 | 1009 | 4.56 |
| 8 | -64 | -14.5 | -135 | 1006 | 1.67 |
| 9 | 141 | -254 | 90 | 1018 | 12 |
| 10 | -77 | -83 | 5 | 1013 | 2.875 |
| 11 | -17 | -25 | -50 | 1003 | -6.27 |
| 12 | 8 | 24 | 20 | 1011 | -0.93 |
| 13 | 26 | 40 | 95 | 1005 | 7.25 |
| 14 | -105 | -279 | 10 | 1009 | 1.11 |
| 15 | 20 | -40 | -16 | 1005 | 2.92 |
| 16 | 10 | -65 | 190 | 1013 | 1.19 |
| 17 | 50 | 84 | 25 | 1005 | 1.75 |
| 18 | 22.5 | -7.5 | -25 | 1009 | 17.375 |
| 19 | 8 | -33 | 160 | 1014 | 0.98 |
| 20 | -62 | -56 | 72 | 1012 | 9.92 |
| 21 | -84 | -40 | -139 | 986 | 4.575 |
| 22 | -65 | 50 | -137 | 1010 | 2.46 |
| 23 | 0 | -65 | 67 | 1008 | 4 |
| 24 | -112 | -37 | 42 | 987 | 9.33 |
| 25 | -151 | -131 | -35 | 1002 | 2.4 |

| | | | | | |
|---|---|---|---|---|---|
| 26 | 17 | -43 | 0 | 1011 | 3.3 |
| 27 | -2.5 | -71.5 | 175 | 1000 | 4.75 |
| 28 | -154 | -45 | -17 | 1002 | 2.4 |
| 29 | -90 | -41 | -100 | 1002 | 8.22 |
| 30 | -56 | 27 | -15 | 1006 | 5.61 |
| 31 | 11 | -164 | 31 | 991 | 0.81 |
| 32 | 18 | -98 | -16 | 1010 | 28.33 |
| 33 | -31 | -189 | -57 | 1007 | 0.875 |
| 34 | 14 | -51 | -44 | 998 | 0.44 |
| 35 | 17 | -38 | 135 | 1011 | 1.05 |
| 36 | -83 | -64 | -92 | 1007 | 4.68 |
| 37 | -132 | -163 | -154 | 996 | 0.71 |
| 38 | 51 | 28 | 52 | 1011 | 4.18 |
| 39 | 72 | -65 | 96 | 1011 | 5.79 |
| 40 | -40 | -108 | -30 | 1006 | 1.245 |
| 41 | -63 | 134 | 47 | 1013 | 2.98 |
| 42 | -66 | -42 | -115 | 1017 | 4.88 |
| 43 | 11 | -17 | -4 | 1001 | -1.63 |
| 44 | -135 | -350 | -35 | 1008 | -73.5 |
| 45 | -201 | 0 | -25 | 1006 | 7.5 |
| 46 | -40 | -46 | -5 | 1002 | -4.32 |
| 47 | 27 | -9 | 33 | 1011 | 1 |
| 48 | 74 | -50 | 50 | 1015 | 1.79 |
| 49 | 22 | -55 | 93 | 1007 | 2.95 |
| 50 | -36 | -216 | 3 | 993 | -96 |
| 51 | 120 | -72 | -75 | 999 | 4.16 |
| 52 | 41 | 7.5 | -16 | 1015 | 3.34 |
| 53 | -75 | 25 | -48 | 1015 | 4.18 |
| 54 | 131 | 0 | 60 | 1008 | 0.84 |
| 55 | 0 | 0 | 19 | 1009 | 1.15 |
| 56 | 9 | -40 | -67 | 1007 | 5 |
| 57 | -76 | -55 | -3 | 1013 | 8.11 |

# LIST OF REFERENCES

1. Hogan, T.F.,and Rosmond, T.E., "The Description of the Navy Operational Global Atmospheric Prediction System's Spectral Forecast Model," *Monthly Weather Review* , V.119, pp.1786-1815, 1991.

2. Harr, P.A., Elsberry, R.L., Hogan, T.F., and Clune, W.M., "Forecasts of North Pacific Maritime Cyclones with the Navy Operational Global Atmospheric Prediction System," paper to appear in *Weather and Forecasting* , v.6, September 1992.

3. Naval Environmental Prediction Research Facility Contractor Report CR 78-01, *Tropical Cyclone Strike Probability Forecasting* , by J.D. Jarrell, pp. 2-1 - 2-17, December 1978.

4. Mosteller, F., Tukey, J.W., *Data Analysis and Regression* , 1st ed., Addison-Wesley Publishing Co., pp.134-148, 1977.

5. Mendenhall, W., Schaeffer, R., Wackerly, D., *Mathematical Statistics with Applications* , Duxbury Press, 2nd ed., 1981, pp.493-500.

6. Weisberg, S., *Applied Linear Regression* , John Wiley & Sons, Inc., 1980.

7. Anderson, T.W., *An Introduction to Multivariate Statistical Analysis,* pp.178-226., John Wiley & Sons, Inc., 1958.

# INITIAL DISTRIBUTION LIST

| | | No. Copies |
|---|---|---|
| 1. | Defense Technical Information Center<br>Cameron Station<br>Alexandria, VA 22304-6145 | 2 |
| 2. | Library, Code 52<br>Naval Postgraduate School<br>Monterey, CA 93943-5002 | 2 |
| 3. | Professor Robert R. Read<br>Code OR/Re<br>Naval Postgraduate School<br>Monterey, CA 93943 | 1 |
| 4. | Professor Glenn F. Lindsay<br>Code OR/Li<br>Naval Postgraduate School<br>Monterey, CA 93943 | 1 |
| 5. | William Clune<br>Fleet Numerical Oceanographic Center<br>Monterey, CA | 2 |
| 6. | Professor Pat Harr<br>Code OR/Ha<br>Naval Postgraduate School<br>Monterey, CA 93943 | 1 |