APPLICATION OF SATELLITE CLOUD-MOTION VECTORS IN HURRICANE TRACK PREDICTION

by

Alan L. Adams

B.A.E., Georgia Institute of Technology

(1971)

B.E.S., University of Texas

(1972)

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science
at the
Massachusetts Institute of Technology

February 1976

Signature of Author....................................
Department of Meteorology, Jan. 21, 1976

Certified by...
Thesis Supervisor

Accepted by..
Chairman, Departmental Committee

APPLICATION OF SATELLITE CLOUD-MOTION VECTORS IN HURRICANE TRACK PREDICTION

by

Alan L. Adams

Submitted to the Department of Meteorology on January 1976 in partial fulfillment of the requirements for the degree of Master of Science.

ABSTRACT

The representation of the mean tropospheric flow by satellite - derived cloud-motion vectors is studied for use in a barotropic hurricane prediction model. The systematic use of these vectors is considered over areas not covered by rawinsonde data to aid the initial analysis of the flow pattern. Linear regression analysis is used to develop equations for the pressure-averaged tropospheric flow from data at only 1, 2, or 3 levels. The equations are derived from a large sample of rawinsonde observations, used as simulated cloud-motion vectors, from the tropical and subtropical latitudes of the Northern Hemisphere. The performance of the regression equations on independent data is considered, as is the loss of skill when satellite winds are used in the equations instead of rawinsonde winds. The satellite data is applied, in a pilot study, to two operational SANBAR hurricane forecasts, with inconclusive results.

Thesis Supervisor: Frederick Sanders

Title: Professor of Meteorology

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

SANBAR is a barotropic hurricane prediction model that utilizes vorticity conservation in the mean tropospheric flow to predict tracks of tropical cyclones. The model was developed by Sanders and Burpee (1968) and has been discussed by Sanders (1970) and by Sanders et al (1975). SANBAR makes use of observed winds which are averaged with respect to pressure through the depth of the troposphere, defined as the layer between the 1000-mb and 100-mb surfaces. The averaged wind is represented by the weighted average of the data at the ten mandatory levels, as observed by rawinsonde.

A major factor limiting forecast accuracy in the operational use of the model at the National Hurricane Center (NHC) was the lack of data over the large oceanic areas included in the SANBAR forecast grid area. Sanders et al (1975) discussed specific cases. To guide the analysis of the wind field over the vast oceanic areas far from any rawinsonde observations, the model relies on "bogus" wind observations at 44 selected geographical locations. These bogus points are shown in Figure 1. The winds are obtained at present from consideration of many factors, including 12-hour prognostic wind and height fields, surface observations from ships, aircraft reports, and SMS satellite-derived winds. This report explores the increased and systematic use of such satellite winds over the oceans to improve the initial analysis, by determining how well the pressure-averaged flow is represented by information at one, two, or three levels.

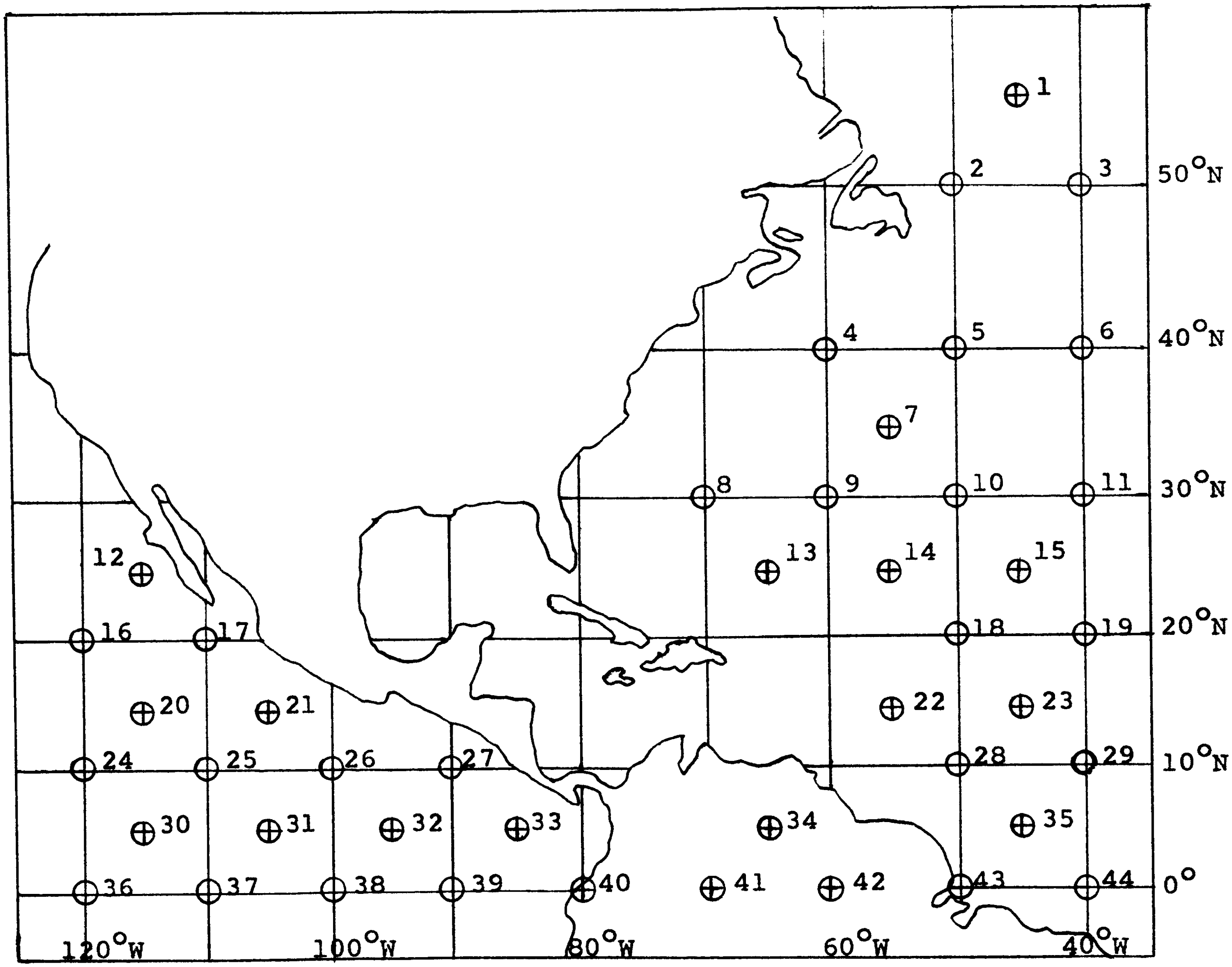


Figure 1--SANBAR Bogus Points

II. SATELLITE-DERIVED WINDS

The use of satellite photographs to track cloud motion as a means of determining wind velocity at cloud level was discussed by Hubert and Whitney (1971). They compared cloud movements from the geostationary ATS satellite imagery with rawinsonde observations of wind from nearby stations. Reasonable agreement was found, if the motions of low level and high level clouds were compared to winds in the layers from 3000 to 5000 feet and around 30,000 feet, respectively. On rare occasions when mid-level clouds can be identified, their wind vectors are assigned to the 500 mb level by the National Meteorological Center. (NMC). The median vector differences between these estimated and rawinsonde observations at low and high levels are approximately 6 knots and 12 knots, respectively (Hubert, 1975). Further discussion of satellite derived winds can be found in Appendix A.

Given good satellite coverage, it is thus possible to obtain estimates of wind flow at low levels and high levels over wide areas with possibly some idea of the mid-level flow. This information should be extremely valuable over oceanic regions for prediction of tracks of tropical storms. In the context of the SANBAR model, the question then arises how adequately one, two, or three levels of wind data can represent the mean tropospheric flow. Linear regression analysis will be used to estimate the mean flow from rawinsonde observations (used as simulated cloud-motion vectors) at one to three levels.

The idea of using satellite cloud-motion vectors to improve the bogus data is not new. Pike (1975), at NHC, developed a set of regression equations utilizing data from the low level ATOLL (Atlantic Tropical Oceanic Lower Layer) analysis and the 200 mb analysis. The ATOLL analysis is essentially the level at the top of the planetary boundary layer and utilizes ship reports, low-level satellite winds, and available 2000-foot rawinsonde winds. The 200-mb analysis is supplemented with aircraft observations and upper level satellite winds. Pike's equations for June through November are included in the section on results from linear regression analysis. They are applied to a small sample of data in the Western Atlantic, Carribean Sea, and Gulf of Mexico.

The purpose of this report is to first develop a statistically stable set of linear regression equations from a substantially larger sample of data over time and geographical location than used by Pike. The results of the linear regression analysis will then be applied to the initial data field in operational SANBAR cases, in hopes of improving the forecasts. Only satellite winds will be used at low levels while the high-level data will consist of satellite winds and aircraft reports.

III. DATA SAMPLE FOR REGRESSION ANALYSIS

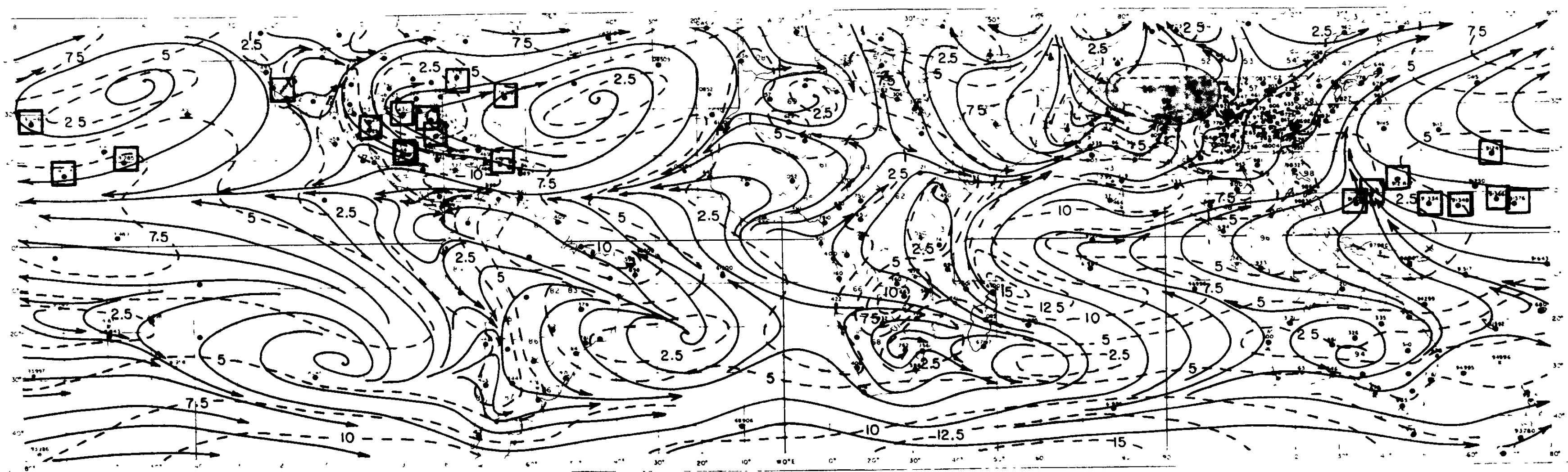
The data for the linear regression analysis consist of a sample of rawinsonde observations from 20 stations located between 0° and 35°N and 60°W westward to 130°E, as listed in Table 1. The data sample is for the months of June through October from 1971 through 1974 for each of the 20 stations. The five-month period corresponds to the period of maximum tropical storm activity in the data area. The five-year time span was chosen to create a sample of sufficient size to obtain statistically sound results, even after considerable stratification. The 20 stations were chosen to provide coverage of different wind regimes and areas of tropical storm activity. The stations are shown in Figures 2 and 3 relative to the long-term mean June-August streamline pattern for the 850-mb and 200-mb levels respectively. The streamline analyses are based on data from Newell *et al* (1972) and others.

TABLE 1

Stations For Linear Regression Analysis

Int. Index No.	Station Name	Lat.	Long.
1. 78016	Bermuda	32.2N	64.6W
2. 78526	San Juan, P.R.	18.3N	66.1W
3. 72304	Cape Hatteras, N.C.	35.3N	75.6W
4. 72202	Miami, Fla.	25.5N	80.1W
5. 72211	Tampa, Fla.	27.6N	82.3W
6. 72240	Lake Charles, La.	30.2N	93.1W
7. 72250	Brownsville, Tx.	25.6N	97.3W
8. 72295	Vandenberg, Cal.	33.9N	118.4W
9. 76644	Merida, Mex.	20.6N	89.4W
10. 91285	Hilo, Hi.	19.4N	155.3W
11. 91275	Johnston Is.	17.0N	168.3W
12. 91066	Midway	28.1N	177.2W
13. 91217	Taguac, Guam	13.3N	144.5E
14. 91245	Wake Is.	19.1N	166.3E
15. 91334	Truk	7.4N	151.8E
16. 91348	Ponape	7.0N	158.2E
17. 91366	Kwajalein	8.7N	167.6E
18. 91376	Majuro	7.1N	177.4E
19. 91413	Yap	9.3N	138.1E
20. 91408	Koror	7.2N	134.3E

LONG-TERM MEAN 850-mb FLOW. DATA FROM NEWELL(1972), LUFTHANSA(1967), SADLER(1970), and SCHWARTZKOPF(1970). ISOTACHS IN M SEC^{-1}



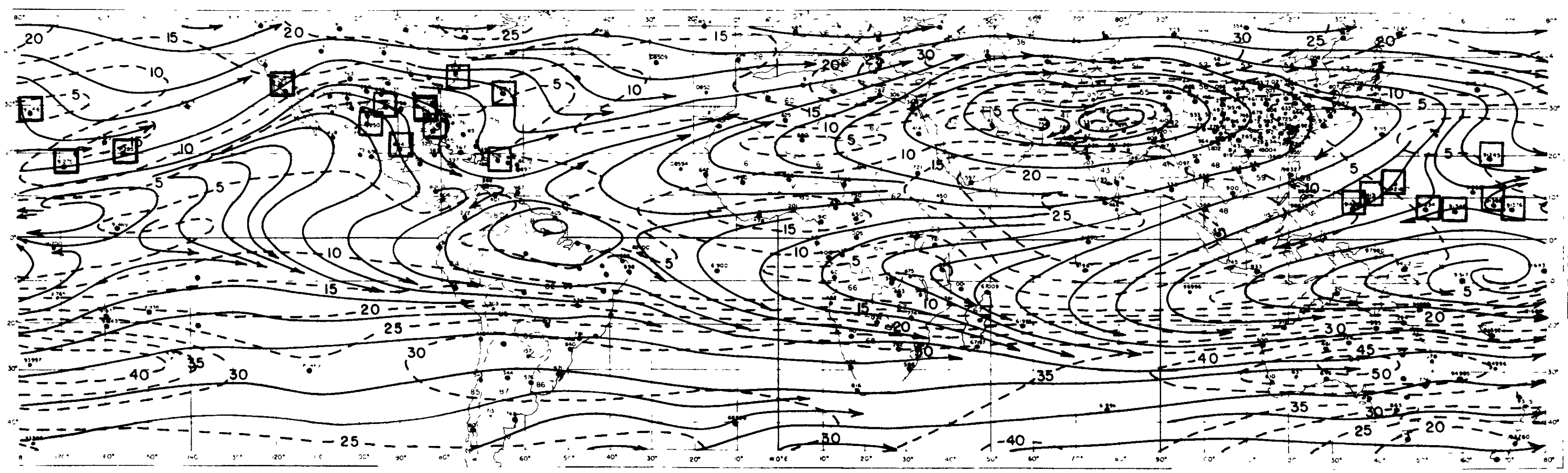
JUNE—AUGUST

SANDERS (1975)

Figure 2--Locations of Regression Analysis Data Stations Relative To 850-mb Average Flow

□--Indicates Stations

LONG-TERM MEAN 200-mb FLOW. DATA FROM NEWELL (1972), LUFTHANSA (1967), SADLER (1970) and SCHWARTZKOPF (1970). ISOTACHS IN M SEC^{-1}



JUNE — AUGUST

SANDERS (1975)

Figure 3--Locations of Regression Analysis Data Stations Relative to 200-mb Average Flow

--Indicates Stations

IV. LINEAR REGRESSION ANALYSIS

The regression equations are formed from the rawinsonde observations at the ten mandatory pressure levels. These data are used to calculate a mean wind based on the assumption that the wind vector varies linearly with pressure between levels. (Appendix B) Standard linear regression techniques are used to obtain separate regression equations for the zonal and meridional components of the mean wind. The predictors are the zonal and meridional components of the winds, respectively, at the specified number of data levels. (The term, "prediction" as used in the linear regression analysis, means the specification of the mean wind by one, two, or three levels of wind data used as "predictors.")

The forms of the prediction equations are

$$\begin{aligned} \hat{y} &= a_o + a_1x_1 \\ \hat{y} &= a_o + a_3x_3 \\ \hat{y} &= a_o + a_1x_1 + a_3x_3 \\ \hat{y} &= a_o + a_1x_1 + a_2x_2 + a_3x_3 \end{aligned} \tag{1}$$

where $\hat{y}$ is the predicted mean wind, x_i are the predictors, and a_i are the coefficients. The subscripts 0, 1, 2, and 3 refer to, respectively, the constant term, 850-mb, 500-mb, and 250-mb predictors. Further discussion of the linear regression analysis can be found in Appendix D.

V. RESULTS OF LINEAR REGRESSION ANALYSIS

The 20-station data sample produced a total of 27,421 soundings. The size and character of the sample suggested the possibility of data stratification both by location and by time. Six geographical and three time stratifications were considered and are shown in Table 2. (Appendix C)

The regression analysis determines the coefficients of the predictors, as well as the constant, a_o, in equations 1. These data are listed in Tables 3, 4, and 5 by u- and v-component and by stratification set. The ability of the resulting equations to predict the mean wind in the dependent data sample is indicated by the reduction of variance, mean-square error, and root-mean-square error (rms error). These quantities are also shown in Tables 3, 4, and 5. (Appendix D)

For comparison, Pike's equations for June through November are:

$$\hat{u}_{1000-100mb} = -0.512 + 0.561\ u_{ATOLL} + 0.399\ u_{200mb}$$

$$\hat{v}_{1000-100mb} = 0.574 + 0.269\ u_{ATOLL} + 0.265\ u_{200mb}$$

where all wind speeds are in knots. Pike's data sample most closely corresponds to geographical set 1 covering the Western Atlantic, Caribbean, and the Gulf of Mexico. The difference between the u-component equations is small, generally much less than two knots, but the v-component equations exhibit larger differences that can be as high as four knots. While Pike's meridional equation gives smaller magnitudes than set 1, the zonal equation generally enhances northerly winds. Set 1 is drawn from a substantially larger data sample than Pike's equations.

The regression equation results, as tabulated, are for the rawinsonde data. The use of satellite-derived winds operationally will cause some loss of skill and correspondingly larger rms errors due to differences in the data sources. The increase in rms error can be relatively large, but the equations still provide significant skill when compared to climatology even for differences or "errors" in the data as large as the wind itself. Further discussion of this problem can be found in the next section and in Appendix G.

Consideration of the rawinsonde-derived results provides useful information on the accuracy of the various predictor sets and stratifications. The three-predictor geographically-stratified equation sets show very high reductions of variance and, consequently, low rms errors, indicating close approximation of the mean tropospheric flow. The two-predictor equations also exhibit high reductions of variance, even though some skill is lost with the omission of the mid-level predictor. The rms errors are still acceptable as compared with the standard deviations of the mean wind shown in Table 6. The reduction of variance of the one-predictor equations shows wide variability with some values being quite low. The rms errors still indicate some skill as compared to climatology with the 250-mb set showing lower errors than the 850-mb set. (Appendix E)

The use of satellite-derived winds in the operational context suggests that particular importance attaches to the two-predictor equations, hence only these were considered for time stratification. This stratification investigates the possible influence of seasonal variation of the flow pattern, but the results in Table 5 indicate that little is to be gained.

Stratification of the data by location or by time did not produce any significant results. The coefficients, reductions of variance, and rms errors are very similar within each predictor set. This similarity, coupled with the observation that the combination stratifications are approximately the average of their constituent sets, indicates that stratification provides little additional information when compared to the sample taken as a whole. Operationally, the total sample equations (set 6) are the most useful when a single set of equations is desired. The size of the total sample, however, is so large that even after stratification, the individual sets are statistically sound. The use of the stratifications would give somewhat better data resolution than the general set if more accuracy were needed.

It is of interest to note that set 4, the Southwestern Pacific, has, in general, the smallest coefficients, reductions of variance, and rms errors in each predictor set for both u- and v-components. This is due to the location of the 8 stations south of 20^{o}N and the small day-to-day variability of the mean wind in the tropics. As expected, the time stratification for this set shows only minor seasonal variation.

The decrease in skill of the regression equations when applied to independent data will be small because of the large sample size and the small number of predictors. A sample calculation for the Southwestern Pacific with a dependent sample size of 840 statistically-independent observations shows a drop in the reduction of variance from 77.0% to 76.9% which is almost negligible. (Appendix F)

TABLE 2

DATA STRATIFICATIONS

Geographical

Set 1: 8 Stations — 11682 Observations

1) Bermuda
2) San Juan, P.R.
3) Cape Hatteras, N.C.
4) Miami, Fla.
5) Tampa, Fla.
6) Lake Charles, La.
7) Brownsville, Tx.
8) Merida, Mex.

Set 2: 4 Stations — 5594 Observations

1) Vandenberg, Cal.
2) Hilo, Hi.
3) Johnston Island
4) Midway

Set 3: 12 Stations — 17276 Observations

1) Set 1
2) Set 2

Set 4: 8 Stations — 10145 Observations

1) Taguac, Guam
2) Wake Island
3) Truk
4) Ponape
5) Kwajalein
6) Majuro
7) Yap
8) Koror

TABLE 2 (cont'd)

Set 5:	12 Stations	15739 Observations
1)	Set 2	
2)	Set 4	
Set 6:	20 Stations	27421 Observations
1)	Set 1	
2)	Set 2	
3)	Set 4	

Time

Set 1:	Geographical Set 1	
A)	June, July, August	7045 Observations
B)	September, October	4637 Observations
Set 4:	Geographical Set 4	
A)	June, July, August	6151 Observations
B)	September, October	3994 Observations
Set 6:	Geographical Set 6	
A)	June, July, August	16573 Observations
B)	September, October	10848 Observations

TABLE 3

REGRESSION EQUATIONS RESULTS

Geographical Stratifications

u Component

Set	Constant a_0	850 mb a_1	500 mb a_2	250 mb a_3	Reduction of Variance r^2	Mean Square Error (knots) $\overline{e^2}$	rms Error (knots) $\sqrt{\overline{e^2}}$
1	-0.1106	0.3087	0.3591	0.2611	0.9739	3.5713	1.8898
2	0.1453	0.2889	0.3762	0.2488	0.9647	4.1554	2.0385
3	-0.0524	0.3013	0.3653	0.2577	0.9714	3.7425	1.9346
4	-0.8016	0.2848	0.3230	0.2543	0.9240	3.5899	1.8947
5	-0.3728	0.2852	0.3506	0.2611	0.9588	3.9345	1.9836
6	-0.2237	0.2954	0.3561	0.2621	0.9684	3.8105	1.9521
1	0.3944	0.5299	------	0.3740	0.9174	11.3023	3.3619
2	0.1805	0.5216	------	0.3641	0.8723	15.0325	3.8772
3	0.3504	0.5314	------	0.3694	0.9043	12.5207	3.5385
4	-2.1777	0.4280	------	0.3233	0.7701	10.8596	3.2954
5	-1.1955	0.4685	------	0.3699	0.8613	13.2319	3.6376
6	-0.4306	0.5049	------	0.3777	0.8929	12.9088	3.5929
1	4.7984	0.7105	------	------	0.4301	77.9803	8.8306
2	7.6402	0.6140	------	------	0.2658	86.4278	9.2967
3	5.7109	0.6608	------	------	0.3625	83.4060	9.1327
4	-4.4933	0.2837	------	------	0.2422	35.7955	5.9829
5	-0.2741	0.3802	------	------	0.1837	77.8973	8.8259
6	2.3709	0.5457	------	------	0.3053	83.7696	9.1526
1	-2.0123	------	------	0.4332	0.6911	42.2673	6.5013
2	-3.8766	------	------	0.3843	0.6824	37.3869	6.1145
3	-2.5778	------	------	0.4073	0.6757	42.4291	6.5138
4	-5.9377	------	------	0.2216	0.2774	34.1328	5.8423
5	-4.6409	------	------	0.3416	0.5864	39.4687	6.2824
6	-3.4631	------	------	0.3913	0.6325	44.3146	6.6569

TABLE 4

REGRESSION EQUATIONS RESULTS

Geographical Stratifications

v Component

Set	Constant a_0	850 mb a_1	500 mb a_2	250 mb a_3	Reduction of Variance r^2	Mean Square Error (knots) $\overline{e^2}$	rms Error (knots) $\sqrt{\overline{e^2}}$
1	-0.3469	0.3024	0.3472	0.2380	0.9528	3.1279	1.7686
2	0.0331	0.2963	0.3495	0.2420	0.9604	3.4615	1.8605
3	-0.2057	0.2949	0.3522	0.2396	0.9562	3.2041	1.7900
4	-0.3600	0.3005	0.3061	0.2356	0.8855	2.7327	1.6531
5	-0.2141	0.2945	0.3291	0.2445	0.9337	3.0857	1.7566
6	-0.2683	0.3008	0.3373	0.2419	0.9445	3.0495	1.7463
1	-0.5127	0.4500	------	0.3273	0.8508	9.8870	3.1444
2	0.0918	0.4600	------	0.3596	0.8781	10.6554	3.2643
3	-0.2756	0.4456	------	0.3422	0.8591	10.3073	3.2105
4	-0.3575	0.4035	------	0.2643	0.6967	7.2387	2.6905
5	-0.1158	0.4291	------	0.3266	0.8047	9.0897	3.0149
6	-0.2933	0.4383	------	0.3277	0.8478	8.3628	2.8919
1	-1.5262	0.5146	------	------	0.3377	43.8900	6.6250
2	0.9196	0.6820	------	------	0.2511	65.4622	8.0909
3	-0.7875	0.5232	------	------	0.2826	52.4803	7.2443
4	-0.7632	0.3976	------	------	0.2367	18.2173	4.2682
5	-0.3055	0.4884	------	------	0.2270	35.9770	5.9981
6	-0.8001	0.4932	------	------	0.2777	39.6877	6.2998
1	0.8994	------	------	0.3507	0.5952	26.8257	5.1794
2	-0.2884	------	------	0.3899	0.7680	20.2794	4.5033
3	0.5521	------	------	0.3632	0.6564	25.1355	5.0135
4	0.2241	------	------	0.2618	0.4179	13.8927	3.7273
5	0.1543	------	------	0.3403	0.6306	17.1926	4.1464
6	0.4651	------	------	0.3439	0.6098	21.4400	4.6303

TABLE 5

REGRESSION EQUATION RESULTS

Time Stratification

u Component

Set	Constant a_0	850 mb a_1	250 mb a_3	Reduction of Variance r^2	Mean Square Error (knots) $\overline{e^2}$	rms Error (knots) $\sqrt{\overline{e^2}}$
1A	0.2679	0.5375	0.3552	0.9079	12.6022	3.5500
4A	-2.1727	0.4169	0.3247	0.7682	10.9493	3.3090
6A	-0.5044	0.4975	0.3666	0.8834	14.0601	3.7497
1B	0.5810	0.5211	0.3860	0.9260	10.1256	3.1821
4B	-2.1915	0.4457	0.3208	0.7750	10.6281	3.2601
6B	-0.3485	0.5148	0.3881	0.9018	11.8413	3.4411

v Component

Set	a_0	a_1	a_3	r^2	$\overline{e^2}$	$\sqrt{\overline{e^2}}$
1A	-0.5148	0.4372	0.3218	0.8331	11.0603	3.3257
4A	-0.3167	0.3867	0.2644	0.7005	7.1480	2.6736
6A	-0.1783	0.4117	0.3217	0.8034	10.8024	3.2867
1B	-0.4512	0.4674	0.3325	0.8668	8.8270	2.9910
4B	-0.4268	0.4253	0.2630	0.6918	7.3557	2.7120
6B	-0.3987	0.4691	0.3322	0.8514	8.1650	2.8574

A=June, July, August
B=September, October

TABLE 6

Standard Deviations Of The Mean Wind

Set	u Component(knots)	v Component(knots)
1	11.6975	8.1406
2	10.8497	9.3494
3	11.4382	8.5530
4	6.8729	4.8853
5	9.7687	6.8222
6	10.9811	7.4126
1A	10.4429	7.3171
1B	13.0745	9.1752
4A	6.7491	4.7720
4B	7.0450	5.0459
6A	10.0390	6.6519
6B	12.1500	8.3833

VI. COMPARABILITY OF SATELLITE DATA TO RAWINSONDE DATA

Operational use of the rawinsonde-derived regression equations presents a problem, since the predictors are now satellite-derived winds while the regression coefficients are tailored to rawinsonde data. The operational use of satellite winds will decrease the accuracy of the equations because of differences between the data sources.

To consider the effects of this difference, the satellite wind can be considered to be the sum of the rawinsonde wind and an effective error. The "true" mean wind, y, is not affected so that the only source of error will be the satellite data. The two-predictor equation will be used to investigate the effects of this error. The satellite winds at the low- and high-level are then:

$$X_1 = x_1 + e_1$$
$$X_3 = x_3 + e_3$$

where x_1 and x_3 are the rawinsonde winds and e_1 and e_3 are their respective errors.

The errors are assumed to be uncorrelated with the wind itself at each level and with each other. In the development of new regression equations, these assumptions will make all covariance quantities involving the errors equal to zero. Only the variances of the satellite winds will be affected by the error which will appear as a variance itself as shown in equations 2.

$$\begin{aligned} \overline{X_1'^2} &= (\overline{x_1'^2} + \overline{e_1'^2}) \\ \overline{X_3'^2} &= (\overline{x_3'^2} + \overline{e_3'^2}) \end{aligned} \qquad (2)$$

The results of the revised regression analysis would be new coefficients whose value would depend upon the magnitude of the error variances.

The reduction of variance for the two-predictor equation can be defined as

$$r^2 = \frac{a_1\overline{x_1'y'} + a_3\overline{x_3'y'}}{\overline{y'^2}}$$

where the variances and covariances are standard statistical quantities. Under the previous assumptions, the coefficients, a_i, will be the only quantities to change. The reduction of variance can thus be written as:

$$r^2 = \frac{\dfrac{(\overline{X_1'Y'})^2(\overline{X_3'^2} + \overline{e_3'^2}) + (\overline{X_3'Y'})^2(\overline{X_1'^2} + \overline{e_1'^2}) - 2(\overline{X_1'Y'})(\overline{X_1'X_3'})(\overline{X_3'Y'})}{(\overline{X_1'^2} + \overline{e_1'^2})(\overline{X_3'^2} + \overline{e_3'^2}) - (\overline{X_1'X_3'})^2}}{\overline{Y'^2}} \quad (3)$$

where the quantities $(\overline{x_1'^2} + \overline{e_1'^2})$ and $(\overline{x_3'^2} + \overline{e_3'^2})$ are the variances of simulated satellite-derived winds for the upper and lower levels, respectively.

The values of $\overline{e_1'^2}$ and $\overline{e_3'^2}$ are not known and can only be estimated. No matter what their value, they can be considered to be some percentage of $\overline{x_i'^2}$, and, therefore, some measure of the effect of this error can be gained by assuming $\overline{e_i'^2}$ over a range of such percentages. Table 7 shows the effects of this error on the reduction of variance and the rms error for the total sample if the same percentage of error is assumed at both levels. This assumption is for simplicity and should not be considered as a correlation between the errors at the two levels.

The effect of the satellite "error" is considerable since even a 10% difference can increase the rms error by approximately 30%. It is interesting to note, however, that an error of 100%, which increases the rms error by 125%, is still better than climatology. Although the $\overline{e_i'^2}$ are unknown, I believe them to be between 10 and 25% of the variance of the wind itself. The actual rms error when satellite winds are used in the two-predictor regression equations is, therefore, about 30 to 60% higher for the rawinsonde data. The equations, however, still exhibit reasonable skill in predicting the mean wind. Similar calculations for the one-predictor equations show increases in the rms errors of about 5 to 20% for $\overline{e_i'^2}$ equal to 25% of the wind variance. These increases indicate further loss of predictability as compared to climatology, but the equations still exhibit some skill. Operational testing of the one-predictor equations is needed to adequately evaluate their usefulness.

Improved methods of measurement should decrease the error and possibly aid in determining its true value. The details of the revised regression analysis can be found in Appendix G.

TABLE 7

Effects Of Satellite Data Error

Total sample (set 6)

Two-predictor equation

u Component

$\overline{x_1'^2} = 123.60 \text{ (knots)}^2$

$\overline{x_3'^2} = 498.16 \text{ (knots)}^2$

$\overline{y'^2} = 120.58 \text{ (knots)}^2$

$\sqrt{\overline{y'^2}} = 10.98$ (knots) (climatological standard deviation)

$\overline{e_1'^2}$ (% $\overline{x_1'^2}$)	$\overline{e_3'^2}$ (% $\overline{x_3'^2}$)	r^2	$\sqrt{\overline{e^2}}$ (knots)
0*	0*	0.893	3.59
10	10	0.815	4.72
25	25	0.721	5.80
50	50	0.605	6.90
100	100	0.457	8.09

*Original rawinsonde data values

TABLE 7
(cont'd)

Total sample (set 6)
Two-predictor equation
v Component

$\overline{x_1'^2} = 62.73 \text{ (knots)}^2$

$\overline{x_3'^2} = 283.25 \text{ (knots)}^2$

$\overline{y'^2} = 54.95 \text{ (knots)}^2$

$\sqrt{\overline{y'^2}} = 7.41$ (knots) (climatological standard deviation)

$\overline{e_1'^2}(\%\overline{x_1'^2})$	$\overline{e_3'^2}(\%\overline{x_3'^2})$	r^2	$\sqrt{\overline{e^2}}$(knots)
0*	0*	0.848	2.89
10	10	0.757	3.65
25	25	0.671	4.25
50	50	0.564	4.89
100	100	0.428	5.61

*Original rawinsonde data values

VII. SELECTION OF OPERATIONAL CASES FOR STUDY

The 1974 hurricane season was chosen for study since it was the most recent. NHC had retained the rawinsonde data base necessary to re-run some operational SANBAR forecasts with revised bogus wind data. Seven named storms occurred during 1974, providing 53 SANBAR forecasts.

The position errors between the forecast and the actual storm track were determined for these predictions. Originally 12 cases were chosen for study, 6 good and 6 bad. The criteria for a good or bad forecast is discussed by Sanders et al (1975). The rationale for choosing bad cases is obvious, since these should, hopefully, show improvement. Good cases are chosen as a check to determine if they are adversely affected by the new data.

Satellite cloud-motion vectors and commercial and reconnaissance aircraft reports for the 12 cases were obtained from the NMC data files as provided by the National Center for Atmospheric Research. Of the original 12 cases, however, 5 were discarded due to the complete absence of satellite data and replaced. The data for each case was then plotted and analyzed to obtain low- and high-level wind flow patterns.

The analysis of the data uncovered some operational problems with the satellite data that proved to be quite formidable. The most obvious and most significant problem was the poor coverage in almost every case. Some large areas of the grid were completely devoid of data while other areas lacked coverage at one of the levels. The aircraft reports and continuity helped in some cases, but large areas were still left with very insufficient coverage. Even in areas of good coverage, the satellite and aircraft data in the same region were sometimes contradictory.

The data coverage problem can be attributed to two causes, one that is inherent in satellite data and one that is unique to 1974. The basis of satellite cloud-motion vectors is, of course, tracking identifiable cloud elements. If no clouds are present over an area, then no vectors can be obtained. Tropical cumulus are very prevalent in the areas of tropical storm activity and are easily tracked even throughout the subtropical anticyclone. Cirrus clouds are less prevalent and offer fewer persistent identifiable elements. Overcast or broken layers of cloud at any level mask all lower clouds. Even when clouds are discernable at more than one level over the same area, only one level may provide suitable targets. Current editing procedure at the National Environmental Satellite Service throws out low-level clouds in the presence of high-level and discards high-level clouds whose motion do not agree with the synoptic situation (Hubert, 1975). Hubert and Whitney (1971) discuss other problems of this type.

The second problem is that the SMS-1 satellite in use during 1974 was moved to longitude 45°W to aid the Global Atmospheric Research Project Atlantic Tropical Experiment (GATE). The satellite was not available for data collection at all times and was unable to adequately cover the SANBAR area. Better coverage should be expected in the future with the increased utilization of more satellites.

These problems were so severe that of the 12 test cases, only 4 were judged useable and even they lacked sufficient coverage to revise all 44 bogus points. In two of the instances, two storms were simultaneously present in the SANBAR area so that 6 storm cases were sent to NHC for recalculation. Because of technical problems at NHC, neither of the "double storm" instances could be used. The two remaining cases were re-run. Further discussion of the selection process can be found in Appendix H.

VIII. RESULTS OF TEST CASES

The two forecasts were for tropical storm Elaine, from initial data on September 10 and 11, 1974, at 1200 GMT. As previously noted, even these two cases still suffered from inadequate data coverage. On September 10, new bogus data could be determined for only 14 points, while the September 11 case provided revised data at only 9 points. The former case was considered a good forecast and the latter was considered bad, at least at 48 hours. The original SANBAR forecast results as well as the revised forecast results are listed in Table 8, and shown in Figures 4 and 5.

The September 10 case exhibits some improvement at 24 and 48 hours, but poorer results at 72 hours. The September 11 case shows virtually no change at any time. Elaine was a weak tropical storm that moved generally ENE until September 12 at 00GMT when it abruptly moved almost northward for approximately 24 hours and then returned to the ENE direction. Neither the original nor the revised forecast was able to predict the northward turn causing the large errors.

The results of the Elaine cases are inconclusive regarding the value of the satellite winds in the initial analysis. The lack of data coverage is very evidently a prime factor. More work with better documented cases is necessary to reliably evaluate the satellite data.

TABLE 8

Forecast Results For Tropical Storm Elaine

Forecast (Date/Time)	Position Error (NM)		
	24 Hr	48 Hr	72 Hr
September 10, 1974/1200GMT			
Original Bogus	105	67	237
Revised Bogus	100	36	299
September 11, 1974/1200GMT			
Original Bogus	105	265	—
Revised Bogus	106	264	—

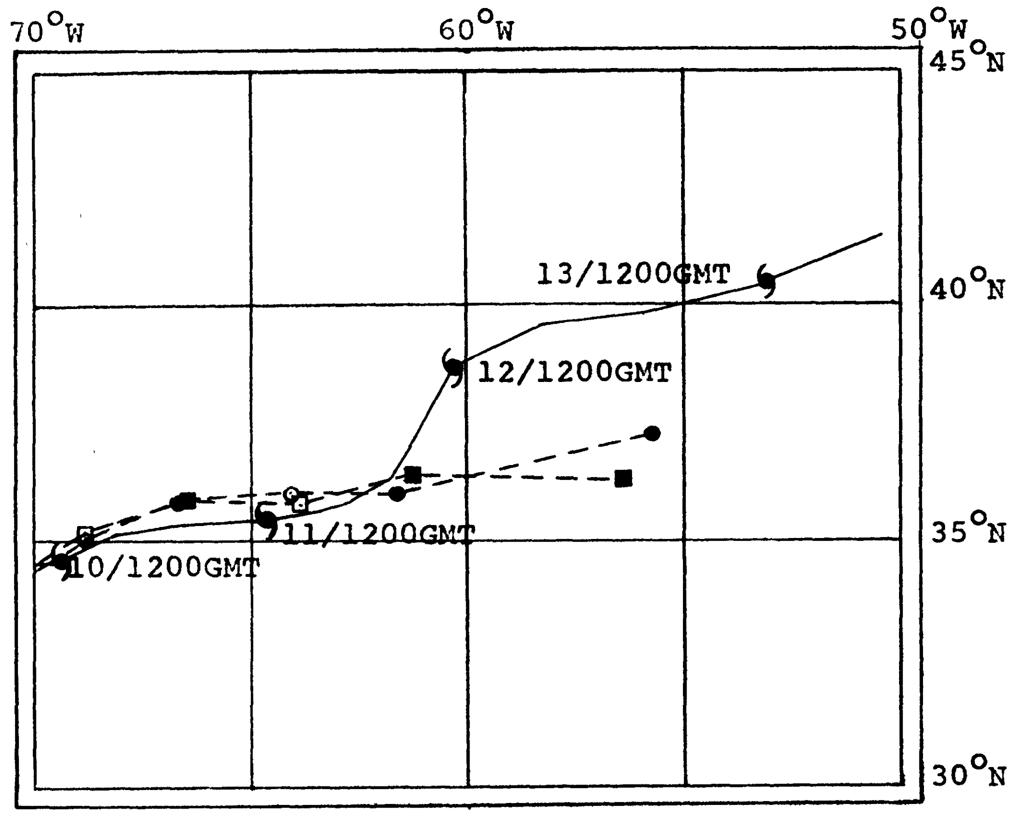


Figure 4--Elaine Forecasts From September 10, 1974, 1200GMT

Best Track of Actual Storm

Forecast With Original Bogus

Forecast With Revised Bogus

For forecasts, closed symbols indicate 24, 48, and 72 Hr positions.

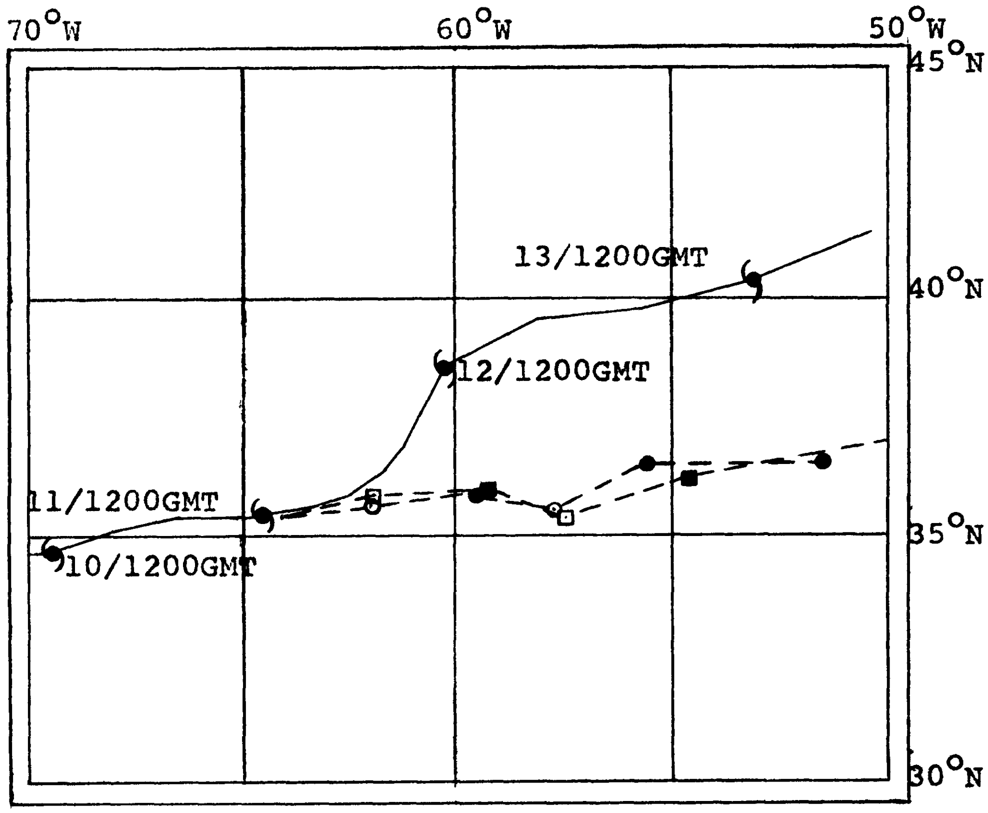


Figure 5--Elaine Forecasts From September 11, 1974,1200GMT

Best Track of Actual Storm

Forecast With Original Bogus

Forecast With Revised Bogus

For forecasts, closed symbols indicate 24, 48, and 72 Hr positions.

ACKNOWLEDGEMENTS

Many people and organizations contributed much to the research and writing of this thesis, and I would like to acknowledge:

The USAF Air Force Institute of Technology for providing the opportunity for me to pursue this study at MIT,

Professor Frederick Sanders, my advisor, who gave invaluable insight and assistance throughout the research and preparation of this study,

My wife, Cindy, for typing the many drafts and final manuscript and providing much needed moral support,

Professor E. N. Loranz who greatly aided the development of the regression analysis procedures,

Isabelle Kole for drafting some of the figures,

Mark Zimmer and the National Hurricane Center for outstanding personal and technical assistance in re-running the test cases,

L.F. Hubert and the National Environmental Satellite Service for providing technical assistance and information,

Paul Mulder and the National Center for Atmospheric Research for assistance in obtaining data for the test cases,

and the

USAF Environmental Technical Applications Center for providing other data for this study.

This project was funded by the USAF Cambridge Research Laboratory.

REFERENCES

1. Burpee, R.W., 1972: The origin and structure of Easterly Waves in the lower troposphere of North Africa. J. Atmos. Sci., 29, No. 1, pp77-90.
2. Gaertner, J.P., 1973: Investigation of forecast errors of the SANBAR hurricane track model. SM thesis, Dept. of Meteorology, Mass. Inst. of Tech., unpublished.
3. Hubert, L.F., 1975: Private communication.
4. Hubert, L.F. and L.F. Whitney, Jr., 1971: Wind estimation from geostationary-satellite pictures. Mon. Wea. Rev., 99, pp665-672.
5. Lorenz, E.N., 1975: Class notes, Dept. of Meteorology, Mass. Inst. of Tech., unpublished.
6. Newell, R.E., J.W. Kidson, D.G. Vincent, G.J. Boer. The General Circulation of the Tropical Atmosphere and Interactions with Extratropical Latitudes. Vol. 1 & 2, Mass. Inst. of Tech. Press, 1973 & 1974.
7. Pike, A.C., 1975: Private communication.
8. Sanders, F., 1959: The application of pressure-averaged flow to the problem of hurricane displacement. AFCRC-TR-59-254, AD 212263.
9. Sanders, F., 1970: Dynamic forecasting of tropical storm tracks. Trans. New York Acad. Sci., Series 11, 32, pp495-508.
10. Sanders, F. and R.W. Burpee, 1968: Experiments in barotropic hurricane track forecasting. J. Appl. Meteor., 7, pp313-323.
11. Sanders, F., A.C. Pike, and J.P. Gaertner, 1975: A barotropic model for operational prediction of tracks of tropical storms. J. Appl. Meteor., 14, No. 3, pp265-280.

APPENDIX A

Determination Of Satellite Cloud-Motion Vectors

Satellite cloud-motion vectors are derived from analysis of successive photographs of cloud patterns. Individual cloud elements, or target clouds, are identified and tracked to determine their motion and estimate the wind field in which they are embedded. More than one level of cloud can often be detected and identified to obtain wind estimates at that cloud level. The height resolution can be determined from different cloud motions over the same area, infrared measurements of cloud top temperatures, and subjective observations of cloud type, brightness and texture. The estimated heights of the clouds are subject to some uncertainties and are generally classified simply as low, middle, or high cloud levels.

Hubert and Whitney (1971) determined the heights of the lower and upper cloud layers from comparison of the motion of the target clouds to hodographs of nearby rawinsondes. The LBF or "level of best fit" was determined from the assumption that the minimum velocity difference between the balloon wind the cloud-motion vector occurred at cloud level. They found that the low-level clouds correspond best to the 3000 to 5000-ft. layer and the high-level clouds correspond to the 30,000 ft. level.

Currently, the level of the cloud is obtained by measuring the temperature of the cloud top with infrared sensors and then comparing this temperature to the vertical temperature profile obtained from the National Meteorological Center (NMC) forecast model. Tropical cumulus are very prevalent in the forecast area and are easily identified as low clouds. The low clouds are generally assigned to the 900 mb level over the oceans although the 850 mb level is also often used. The upper cloud levels correspond well to 200 mb between 0^{o} and $30^{o}N$ and to 300 mb north of $30^{o}N$. Middle level clouds are sometimes identified and are generally representative of the 500 mb level. (Hubert, 1975)

Satellite cloud-motion vectors are used operationally in some forecast models. Under certain assumptions, these winds are used at more than one level. For this reason, NMC provides the low level winds at both the 850 and 700 mb levels, the mid-level at 500 mb, and the high level at 300, 250, and 200 mb levels.

APPENDIX B

Analysis Of Data

The rawinsonde winds are inputs to a computer analysis that develops the mean wind and statistical quantities necessary to form the equation sets. The winds are first resolved into u and v components and then each component is treated separately to develop u and v regression equation sets. The 10 levels of rawinsonde data determine the mean wind components. The mean wind components and the components of the 850-, 500-, and 250-mb winds are then used to compute variances and covariances of the quatities needed to solve for the coefficients of the regression equations.

The mean wind is formed in the computer analysis subject to certain assumptions and constraints. The flow in the troposphere is pressure averaged over the 10 mandatory levels. Lower and upper level mean winds are formed from the lower four and upper six levels respectively. The mean wind is then determined by linear averaging of the two. The sounding is discarded if certain conditions are met concerning missing data: 1) if more than two lower or three upper level winds are missing, 2) if both 1000 and 850 mb winds are missing, or 3) if any four consecutive winds are missing. If the sounding is not discarded, missing winds are linearly interpolated before any computations are performed.

APPENDIX C
Data Stratification

The 20 station sample produced a total of 27,421 soundings for computation after screening by the computer analysis. Six geographical and three time stratifications were considered as shown in Table 1.

The geographical stratifications were based primarily on natural groupings of the 20 stations by their locations. Set 1 covers the western Atlantic, the Gulf of Mexico, and the Caribbean. Set 2 covers the east central Pacific from the California coast to Midway Island. Set 4 covers the southwestern Pacific islands. Sets 3, 5, and 6 are combinations of sets 1, 2 and 4.

Originally, it was hoped that the data could be stratified by latitude and by hemisphere to determine if there was any justification for such groupings. The availability of data, however, did not allow such stratification. Sets 1 and 2 are in the western hemisphere and north of 17^{o}N while Set 4 is in the eastern hemisphere but south of 19^{o}N with 6 of the 8 stations between 0^{o} and 10^{o}N. Stratification by geographical location also stratifies by latitude and hemisphere at the same time causing uncertainty as to what factors might actually contribute to any difference in the equation sets. Sets 3 and 5 were used to check if any effects of location could be detected. Set 6, which included all 20 stations as one data sample, combined all of the possible geographical effects to produce a set of equations for general use in the latitude zone from which the stations were chosen. The selection process used for the analysis would seem to restrict the use of the equations to the oceanic regions of the data area as is intended for tropical storm prediction. The effects of topography would have to be considered over land areas and would require the use of inland stations to include these effects in the development of new regression equations.

Stratification according to time was considered to investigate any seasonal variation between summer and early fall. The five month period was divided into June, July, August, and September, October to form two sets of regression equations. The geographical stratifications were maintained, and three sets were considered for time stratification. Set 1 seemed to be the most likely set to exhibit time dependence due to the location of its stations near mid latitudes. Set 4 is located deep in the tropics and primarily south of 10°N. This set should exhibit little, if any, seasonal change. Set 6 was used to combine all the geographic factors and consider time dependence on the entire sample. Set 2 was not considered due to its small size and location north of the preferred storm areas.

APPENDIX D

Details of Linear Regression Analysis

Let y be the predictand (the component of the mean wind in either the u or v direction)

Let x_i be the predictors (1, 2, or 3 predictors as required)

Define $x_o = 1$ for ease of notation

Let a_i be the coefficients of the predictors x_i with a_o being the constant term in the equation

Define e as the residual error after prediction

Define $\hat{y}$ as the predicted value of y so that $y = \hat{y} + e$

$\overline{(\ )}$ denotes an average over the sample: $\overline{x} = \frac{\Sigma x}{N}$;

N = sample size

Therefore: $$y = \sum_{i=0}^{k} a_i x_i + e$$

so that $\hat{y} = \sum_{i=0}^{k} a_i x_i$ where k can = 1, 2, or 3.

The a_i are chosen to minimize $\overline{e^2}$, the mean square error, so that $\frac{\partial \overline{e^2}}{\partial a_i} = 0$

Now: $$\frac{\partial \overline{e^2}}{\partial a_i} = 2 \overline{\frac{e \partial e}{\partial a_i}} = +\overline{ex_i} = 0$$

Application of the above forms a set of k + 1 equations:

$$a_o + \overline{x_1} a_1 + \ldots \overline{x_k} a_k = \overline{y}$$

$$\overline{x_1} a_o + \overline{x_1^2} a_1 + \ldots \overline{x_1 x_k} a_k = \overline{x_1 y}$$

$$\vdots$$

$$\overline{x_k} a_o + \overline{x_k x_1} a_1 + \ldots \overline{x_k^2} a_k = \overline{x_k y}$$

The elimination of a_o from the equations reduces the set to k equations:

$$\overline{x_1'^2}\, a_1 + \ldots\ \overline{x_1' x_k'}\ a_k = \overline{x_1' y_1'}$$

$$\overline{x_k' x_1'}\ a_1 + \ldots\ \overline{x_k'^2}\ a_k = \overline{x_k' y'}$$

Where the prime ( )' denotes the departure from average, $(x_i - \overline{x_i}) = x_i'$, $\overline{x_i'^2}$ is the variance of x_i, and $\overline{x_i' x_k'}$ is the covariance of x_i and x_k.

$$\overline{x_i'^2} = \overline{(x_i - \overline{x}_i)^2} = (\overline{x_i^2} - \overline{x}_i^2)$$

$$\overline{x_i' x_k'} = \overline{(x_i - \overline{x}_i)(x_k - \overline{x}_k)} = (\overline{x_i x_k} - \overline{x}_i \overline{x}_k)$$

The variances and covariances are evaluated from the computer analysis of the rawinsonde data and are used to determine the a_i's.

a_o is determined by:

$$a_o = \overline{y} - \sum_{i=1}^{k} a_i \overline{x}_i$$

The equations for the 1, 2, and 3 predictor sets are shown in Table D1

The reduction of variance, r^2, is defined by:

$$r^2 = 1 - \frac{\overline{e^2}}{\overline{y'^2}} = \frac{\sum_{i=1}^{k} a_i \overline{x_i' y'}}{\overline{y'^2}}$$

From this expression, $\overline{e^2}$ can be determined by:

$$\overline{e^2} = (1 - r^2)\ \overline{y'^2}$$

The reduction of variance equations used in this report are shown in Table D2.

The standard deviation is defined as the square root of the variance:

$$S = \sqrt{\overline{x_i'^2}}$$

The root mean square error is the square root of $\overline{e^2}$.

$$\text{rms error} = \sqrt{\overline{e^2}}$$

(After Lorenz, 1975)

TABLE D1

3 Predictor (850 mb, 500 mb, 250 mb)

$$\overline{x_1'^2}a_1 + \overline{x_1'x_2'}\ a_2 + \overline{x_1'x_3'}\ a_3 = \overline{x_1'y'}$$

$$\overline{x_1'x_2'}\ a_1 + \overline{x_2'}a_2 + \overline{x_2'x_3'}\ a_3 = \overline{x_2'y'}$$

$$\overline{x_1'x_3'}\ a_1 + \overline{x_2'x_3'}\ a_2 + \overline{x_3'}\ a_3 = \overline{x_3'y'}$$

$$a_o = \overline{y} - a_1\overline{x}_1 - a_2\overline{x}_2 - a_3\overline{x}_3$$

$$\hat{y} = a_o + a_1x_1 + a_2x_2 + a_3x_3$$

2 Predictor (850 mb, 250 mb)

$$\overline{x_1'^2}a_1 + \overline{x_1'x_3'}\ a_3 = \overline{x_1'y'}$$

$$\overline{x_1'x_3'}\ a_1 + \overline{x_3'^2}\ a_3 = \overline{x_3'y'}$$

$$a_o = \overline{y} - a_1\overline{x}_1 - a_3\overline{x}_3$$

$$\hat{y} = a_o + a_1x_1 + a_3x_3$$

1 Predictor

a) 850 mb

$$a_1 = \frac{\overline{x_1'y'}}{\overline{x_1'^2}}$$

$$a_o = \overline{y} - a_1\overline{x}_1$$

$$\hat{y} = a_o + a_1x_1$$

b) 250 mb

$$a_3 = \frac{\overline{x_3'y'}}{\overline{x_3'^2}}$$

$$a_o = \overline{y} - a_3\overline{x}_3$$

$$\hat{y} = a_o + a_3x_3$$

x_1 = 850 mb wind component
x_2 = 500 mb wind component
x_3 = 250 mb wind component

TABLE D2

3 Predictor (850 mb, 500 mb, 250 mb)

$$r^2 = \frac{a_1\overline{x_1'y'} + a_2\overline{x_2'y'} + a_3\overline{x_3'y'}}{\overline{y'^2}}$$

2 Predictor (850 mb, 250 mb)

$$r^2 = \frac{a_1\overline{x_1'y'} + a_3\overline{x_3'y'}}{\overline{y'^2}}$$

1 Predictor

a) 850 mb

$$r_1^2 = \frac{a_1\overline{x_1'y'}}{\overline{y'^2}}$$

b) 250 mb

$$r_3^2 = \frac{a_3\overline{x_3'y'}}{\overline{y'^2}}$$

APPENDIX E

Evaluation of The Prediction Sets

The three-predictor equation closely approximates the mean tropospheric flow as discussed in Section 5. Unfortunately, the operational usefulness of this equation is limited since the mid-level satellite wind is rarely available. The most operationally useful equation is the two-predictor, since two levels of data are routinely available. Given good data coverage at these levels, the two-predictor equation can provide reasonable values of the mean flow.

The one-predictor equation can be operationally useful in cases of good coverage at one level but little or no coverage at the other. The accuracy of the single-predictor equation is less than for the two-predictor, but still somewhat better than climatology. The rms errors from the one-predictor equation can be compared to the standard deviations of the mean wind in Table 6. The 250-mb equation sets show lower rms errors than the 850-mb sets and, therefore, provide more prediction skill. The increase in rms error due to the use of satellite data in the rawinsonde-derived equations is discussed in Appendix G.

APPENDIX F

Effects of Sampling On The Regression Analysis

The regression equations have shown reasonable skill in predicting the wind in the dependent data sample from which they were derived. The question then arises as to the ability of the equations to predict the mean wind in a new independent data sample. The equations must be evaluated to determine their reliability.

As previously defined, the reduction of variance in the dependent sample is

$$r^2 = 1 - \frac{\overline{e^2}}{\overline{y'^2}} \qquad 1)$$

The reduction of variance will decrease in a new data sample due to the process of sampling. The amount of this decrease can be characterized by a new quantity called the expected reduction of variance, ρ. The expected reduction of variance is an estimate of how well the equations will perform on a new data sample. ρ is dependent on the number of predictors in the equations, the original reduction of variance, and the number of independent observations in the dependent sample. The expected reduction of variance is then:

$$\rho = r^2 - \frac{2MN'}{(N' + 1)(N' - M - 1)} (1 - r^2) \qquad 2)$$

where r^2 is the original reduction of variance, M is the number of predictors, and N' is the adjusted sample size.

The quantity N' can be equal to N, the original sample size, but in many cases it is less than N. This is due to the fact that N' is the number of independent observations in the sample while N is simply the total number of observations. If the observations are chosen completely at random such that all are independent of the others, $N' = N$. In this study, however, the observations are chosen as consecutive rawinsonde soundings on 153 consecutive days for 5 consecutive years. This formulation suggests that there is a dependency of one observation on another implying that N' is less than N. The amount that N' is less than N is not an exact figure, but can be estimated.

The initial assumption for determining N' is that the winds in one year will not be dependent on any other year so that each year can be considered to be independent. The five month time span per year is 153 days. Since two soundings are generally made per day, the initial one year sample is 306 observations per station. However, the two (or more) daily soundings must be assumed to be dependent causing a reduction of the sample size by one-half leaving 153 potential observations.

Burpee (1972) determined that African waves in the lower troposphere have periods of 3-5 days. Based on this and other considerations of tropical flow patterns, it seems reasonable to assume that an independent observation should be obtained at least in every 5-7 days. Assuming the time scale to be 7 days gives 21 independent observations per year per station. Therefore, the value of N' will be 105 observations per station for the 5 year sample. The values of N' and are tabulated in Table F1 for each of the geographical stratifications.

The expected loss in skill of prediction is negligible due to the small number of predictors and the large number of independent observations.

TABLE F1

Expected Reduction of Variance

A)

Set	No. of Stns.	N	N'
1	8	11682	840
2	4	5594	420
3	12	17276	1260
4	8	10145	840
5	12	15739	1260
6	20	27421	2100

B)

Set	M	r^2		ρ	
		u	v	u	v
1	3	0.9739	0.9528	0.9737	0.9525
2	3	0.9647	0.9601	0.9642	0.9598
3	3	0.9714	0.9562	0.9713	0.9560
4	3	0.9240	0.8855	0.9235	0.8845
5	3	0.9588	0.9337	0.9586	0.9334
6	3	0.9684	0.9445	0.9683	0.9443
1	2	0.9144	0.8508	0.9140	0.8501
2	2	0.8723	0.8781	0.8711	0.8769
3	2	0.9043	0.8591	0.9040	0.8587
4	2	0.7701	0.6967	0.7690	0.6953
5	2	0.8613	0.8047	0.8609	0.8041
6	2	0.8929	0.8478	0.8927	0.8475
1	1/850	0.4301	0.3377	0.4287	0.3361
2	1	0.2658	0.2511	0.2623	0.2475
3	1	0.3625	0.2826	0.3615	0.2815
4	1	0.2422	0.2367	0.2404	0.2349
5	1	0.1837	0.2270	0.1824	0.2258
6	1	0.3053	0.2777	0.3046	0.2770

TABLE F1 B) (cont'd)

Set	M	r^2		ρ	
		u	v	u	v
1	1/250	0.6911	0.5952	0.6904	0.5942
2	1	0.6824	0.7680	0.6809	0.7669
3	1	0.6757	0.6564	0.6752	0.6559
4	1	0.2774	0.4179	0.2157	0.4165
5	1	0.5864	0.6306	0.5857	0.6300
6	1	0.6325	0.6098	0.6321	0.6094

APPENDIX G

Analysis Of Satellite Data Error

Inherent in the determination of satellite cloud-motion vectors is the possibility of "error" when they are compared to the actual flow at the level as defined by rawinsonde data. This error is more accurately a difference between the data sources and is possibly due to a difference in the scale of observed motion. The satellite winds are determined from cloud motions over broad areas and generally represent the large scale flow. They do, however, suffer from errors in measurement and from height uncertainty as detailed by Hubert and Whitney (1971). The rawinsonde data, as a whole, represent the large scale flow, but individual stations can often be indluenced by small scale fluctuations.

This difference or error is not simply a question of accuracy, but is a question of the applicability of the rawinsonde-derived equation to satellite data. As discussed in Section 6, the satellite wind can be considered as the sum of the rawinsonde wind and some effective error as:

$$X_i = (x_i + e_i)$$

The variances and covariances for satellite data are then:

$$\begin{aligned} \overline{X_i'^2} &= \overline{x_i'^2} + 2\overline{x_i'e_i'} + \overline{e_i'^2} \\ \overline{X_i'X_j'} &= \overline{x_i'x_j'} + \overline{x_i'e_j'} + \overline{x_j'e_i'} + \overline{e_i'e_j'} \qquad (1) \\ \overline{X_i'y'} &= \overline{x_i'y'} + \overline{e_i'y'} \end{aligned}$$

where y is the actual mean wind. Since the errors have been assumed to be uncorrelated with the wind and with each other, the covariances involving the errors will be equal to zero and equations 1 reduce to:

$$\overline{X_i'^2} = \overline{x_i'^2} + \overline{e_i'^2}$$

$$\overline{X_i'X_j'} = \overline{x_i'x_j'} \tag{2}$$

$$\overline{X_i'Y'} = \overline{x_i'y'}$$

Revised regression equation analysis for the satellite data produces new coefficients a_i for the two-predictor equations as shown:

$$a_i = \frac{(\overline{x_i'y'})(\overline{x_j'^2} + \overline{e_j'^2}) - (\overline{x_i'x_j'})(\overline{x_3'y'})}{(\overline{x_i'^2} + \overline{e_i'^2})(\overline{x_j'^2} + \overline{e_j'^2}) - (\overline{x_i'x_j'})^2}$$

where $(\overline{x_i'^2} + \overline{e_i'^2})$ and $(\overline{x_j'^2} + \overline{e_j'^2})$ are simulated satellite winds at the two levels. The new reduction of variance is defined by equation 3 in Section 6. The effects of the satellite error on the two-predictor equation are shown in Section 6, Table 7.

Analysis of the one-predictor equations produces similar results. Table G1 shows the effect of the satellite error for set 1 (Western Atlantic, Caribbean, Gulf of Mexico) and set 4 (Southwestern Pacific) and for their individual stations. The increase in the rms error varies from about 1% to about 20% for assumed realistic values of the error and can be compared to the standard deviation of the mean wind. Some variability can be seen in the rms errors within each stratification set. The stations are arranged in the table by decreasing latitude with the most northerly first within each set. Less variability is seen in low latitudes as is to be expected. The single-predictor equations show little improvement over climatology in most cases, but the 250-mb sets do exhibit modest skill at the higher latitudes. Operational testing of the one-predictor equations is necessary to actually determine their prediction skill.

TABLE G1

Effects of Satellite Error on One-Predictor Equations

u Component

$$e_i'^2 = 0.25x_i'^2$$

	850 mb				250 mb		
	Rawinsonde		Satellite		Rawinsonde		Satellite
Int. Index. No.	r^2(%)	$\sqrt{\overline{e^2}}$	$\sqrt{\overline{e^2}}$	$\sqrt{\overline{y'^2}}$	r^2(%)	$\sqrt{\overline{e^2}}$	$\sqrt{\overline{e^2}}$
72304	40	9.8	10.4	12.7	70	7.0	8.4
78016	42	8.7	9.2	11.3	63	6.9	8.0
72240	28	9.7	10.1	11.4	77	5.5	7.1
72211	34	9.1	9.6	11.2	73	5.8	7.2
72250	22	9.4	9.7	10.7	82	4.6	6.3
72202	34	8.4	8.8	10.4	72	5.5	6.7
76644	22	6.1	6.3	6.9	65	4.1	4.8
78526	34	6.2	6.5	7.6	55	5.1	5.7
Set 1	43	8.8	9.5	11.7	69	6.5	7.8
91245	35	7.3	7.7	9.1	58	5.9	6.6
91217	49	5.5	6.0	7.7	23	6.8	7.0
91413	39	4.7	5.0	6.1	18	5.5	5.6
91366	20	5.5	5.7	6.2	26	5.3	5.5
91334	30	5.1	5.3	6.1	21	5.4	5.6
91408	46	4.8	5.2	6.6	17	6.0	6.1
91376	18	5.2	5.3	5.8	24	5.0	5.2
91348	15	5.1	5.2	5.5	28	4.7	4.9
Set 4	24	6.0	6.2	6.9	28	5.8	6.1

APPENDIX H
Selection And Analysis of Study Cases

The 53 SANBAR forecasts were evaluated for position errors at 24, 48, and 72 hours with the "best track" storm locations as supplied by NHC. The best track is the official track of the storm as determined from all observations. The position error was simply calculated as the vector difference between the SANBAR forecast position and the best track position at the same time. Good and bad SANBAR forecasts were initially identified for study. The criteria for judging good from bad is that a good forecast should have a position error of less than 75, 150, and 300 NM at 24, 48, and 72 hours respectively, as discussed by Sanders et al (1975). Every "good" forecast does not meet every one of these position error values, but these criteria are generally useful for evaluation.

The analysis procedure first involved plotting the data on low-level and high-level charts. Streamline analysis was used to determine the flow patterns where possible at each level. From these patterns, two levels of data could be obtained for a bogus point and the mean wind calculated from the regression equations. The data for each bogus point had to be interpolated and was, therefore, subject to errors whose magnitudes depended largely on the quality of the data coverage. The error could be almost zero in areas of good coverage and quite large in poor coverage areas. The value of the interpolation error is difficult to evaluate, but it must be considered, in some manner.