

NPSOR-91-029

# NAVAL POSTGRADUATE SCHOOL

Monterey, California



**PRELIMINARY RESULTS FROM THE ANALYSIS OF  
WIND COMPONENT ERROR**

Donald P. Gaver  
Patricia A. Jacobs

September 1991

Approved for public release; distribution is unlimited.

Prepared for:

Naval Oceanographic and Atmospheric Research Laboratory  
Monterey, CA 93943-5006

FedDocs  
D 208.14/2  
NPS-OR-91-029

Fed Docs  
D 208.14/2: NPS-DR-91-029

NAVAL POSTGRADUATE SCHOOL  
MONTEREY, CALIFORNIA 93945

**NAVAL POSTGRADUATE SCHOOL,  
MONTEREY, CALIFORNIA**

Rear Admiral R. W. West, Jr.  
Superintendent

Harrison Shull  
Provost

This report was prepared in conjunction with research funded by the Naval Oceanographic and Atmospheric Research Laboratory, Monterey, California.

This report was prepared by:

**REPORT DOCUMENTATION PAGE**

Form Approved  
OMB No. 0704-0188

REPORT SECURITY CLASSIFICATION <b>UNCLASSIFIED</b>		1b. RESTRICTIVE MARKINGS	
SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION /AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.	
DECLASSIFICATION/DOWNGRADING SCHEDULE			
PERFORMING ORGANIZATION REPORT NUMBER(S) NPSOR-91-029		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
NAME OF PERFORMING ORGANIZATION Naval Postgraduate School	6b. OFFICE SYMBOL (If applicable) OR	7a. NAME OF MONITORING ORGANIZATION Naval Oceanographic and Atmospheric Research Laboratory	
ADDRESS (City, State, and ZIP Code) Monterey, CA 93943		7b. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5006	
NAME OF FUNDING/SPONSORING ORGANIZATION Naval Oceanographic and Atmospheric Research Laboratory	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N6846291WR 10023 and N6846291WR 10035	
ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5006		10. SOURCE OF FUNDING NUMBERS	
		PROGRAM ELEMENT NO.	PROJECT NO.
		TASK NO.	WORK UNIT ACCESSION NO.

TITLE (Include Security Classification)  
*Preliminary Results from the Analysis of Wind Component Error*

PERSONAL AUTHOR(S)  
Donald P. Gaver and Patricia A. Jacobs

TYPE OF REPORT Technical Report	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, month day) September, 1991	15. PAGE COUNT
------------------------------------	--	---	----------------

SUPPLEMENTARY NOTATION

COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Gaussian model with log-linear scale parameter; maximum likelihood; Newton's method
FIELD	GROUP	SUB-GROUP	

ABSTRACT (Continue on reverse if necessary and identify by block number)  
Estimation of mean square prediction error of wind components is required in the optimal interpolation (OI) process in numerical prediction of atmospheric variables. Statistical models with log-linear scale parameters which include covariates are described for the prediction error. Data from February and April of 1991 are used to fit the model parameters and to study the predictive ability of the models. This preliminary investigation indicates that observational and best guess wind components can be helpful in predicting mean square prediction error for wind components.

DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
NAME OF RESPONSIBLE INDIVIDUAL P. Gaver		22b. TELEPHONE (Include Area Code) (408) 646-2605	2c. OFFICE SYMBOL OR/Gv



# PRELIMINARY RESULTS FROM THE ANALYSIS OF WIND COMPONENT ERROR

by

P. A. Jacobs and D. P. Gaver

## 0. INTRODUCTION

Numerical meteorological models are used to assist in the prediction of weather. Each run of a numerical model produces forecasts of meteorological variables which are used as preliminary predictions of the future values of these variables. These initial predictions are referred to as **first-guess** values. In this paper first-guess values will refer to the most recent 12 hour forecasts.

In certain areas of the world observations of the values of forecasted variables become available, in our case the observations become available 12 hours after the first-guess values are computed. Prior to the next run of the numerical model a multivariate optimal interpolation analysis updates a first-guess value of a variable by adding to it a weighted observed value of the variable if it is available. The weight multiplying the observed value depends on estimates of the squared error of the first-guess value and the squared error of the observation; cf. Goerss et al. [1991, a, b]. Thus it is of importance to predict such first-guess squared errors.

The general problem of modeling and predicting mean square errors is important but not widely studied; see Efron (1986) and Jorgenson (1987). In the next section statistical models for the error of the first-guess are introduced. The models assume the error of the first-guess has mean 0 but has a scale parameter that is log-linear with suitable covariates, i.e. explanatory or regression variables.



Results are reported concerning the estimation of model parameters, and model cross-validation and predictive ability for  $u, v$  wind component data from the months of February and April 1991. The data consist of measurements and 12 hour forecasts (first-guess values) from 93 stations in North America, 25N–75N. The forecasts are produced using the NOGAPS Spectral Forecast Model; cf. Hogan et al. Each station has measurement and first-guess values for every 12 hours; there are some missing observations. The first-guess values are subtracted from measurement values (if available) to obtain observations of the error of the first-guess. The results appear in Sections 3 and 4 and in Appendices B, C and D.

The results indicate that estimates of the variance of the error of first-guess wind components can be improved by using covariates which are functions of the wind components. Covariates using observed values of the wind components appear to have more predictive ability than those using first-guess values. Further exploratory work is needed to determine the degree with which these statistical results can be used to improve the forecasting ability of the numerical model.

## 1. THE MODELS

Let

$U_0(t)$  = observed  $u$ -wind component at time  $t$

$U_f(t)$  = first-guess  $u$ -wind component at time  $t$

$V_0(t)$  = observed  $v$ -wind component at time  $t$

$V_f(t)$  = first-guess  $v$ -wind component at time  $t$

$$r(t) = \left[ (U_0(t) - U_0(t-1))^2 + (V_0(t) - V_0(t-1))^2 \right]^{\frac{1}{2}}$$

$$s(t) = \left[ U_0(t)^2 + V_0(t)^2 \right]^{\frac{1}{2}}$$

$$Y(t) = U_0(t) - U_f(t) \quad \text{or} \quad Y(t) = V_0(t) - V_f(t)$$

The models considered are as follows:

## NORMAL MODELS:

### One Variable Models

1.  $\{Y(t)\}$  are independent normally distributed random variables with mean 0 and variance

$$\sigma_1^2(1;t) = \exp\{\alpha_1(1) + \beta_1(1)r(t)\}. \quad (1)$$

2.  $\{Y(t)\}$  are independent normally distributed random variables with mean 0 and variance

$$\sigma_1^2(2;t) = \exp\{\alpha_1(2) + \beta_1(2)s(t)\}. \quad (2)$$

### Two Variable Model

3.  $\{Y(t)\}$  are independent normally distributed random variables with mean 0 and variance

$$\sigma_2^2(t) = \exp\{\alpha + \beta_1 r(t) + \beta_2 s(t)\}. \quad (3)$$

## CAUCHY MODELS:

While many measurement errors of physical quantities are approximately normal, especially "in the middle" of their distribution, there can well be thicker-than-normal tails and occasional extreme outliers. These attributes can have seriously degrading effects in regression-like problems; cf. Mosteller and Tukey (1977), Huber (1981) and Hampel (1986). The Cauchy distribution is a symmetric distribution with thicker tails than those of the normal distribution. Distributions with long straggling tails have the tendency to produce outlying values. The following models use the Cauchy distribution to represent and suitably compensate for more-thick-tailed measurement error than that of the Normal distribution.



## One Variable Models

4.  $\{Y(t)\}$  are independent Cauchy random variables with scale parameter

$$\sigma_1^2(1;t) = \exp\{\alpha_1(1) + \beta_1(1)r(t)\}. \quad (4)$$

5.  $\{Y(t)\}$  are independent Cauchy random variables with scale parameter

$$\sigma_1^2(2;t) = \exp\{\alpha_1(2) + \beta_1(2)s(t)\}. \quad (5)$$

## Two Variable Model

6.  $\{Y(t)\}$  are independent Cauchy random variables with scale parameter

$$\sigma_2^2(t) = \exp\{\alpha + \beta_1 r(t) + \beta_2 s(t)\}. \quad (6)$$

The form of the Cauchy density function with scale parameter  $\sigma$  that is used is

$$f(y) = \frac{1}{\pi\sigma} \left[ 1 + \frac{y^2}{\sigma^2} \right]^{-1} \quad \text{for } -\infty < y < \infty.$$

## 2. ESTIMATION OF PARAMETERS

For both normal and Cauchy models, the model parameters are estimated by maximum likelihood. A system of equations is obtained by setting the first partial derivative with respect to each parameter of the  $\ln$  likelihood function equal to zero. The system of equations is solved numerically using Newton's method to obtain the maximum likelihood estimates. The procedure for the normal models is given in Appendix A.

### 3. THE DATA ANALYSIS—FEBRUARY DATA

#### 3.1 Observed Wind Covariate Models

In this subsection we report an assessment of the goodness of fit and cross-validation for the normal models (1)–(3) using observational wind components as covariates. There are six analyses; one for the  $u$ -wind component (respectively  $v$ -wind component) for each pressure level height. Each analysis proceeds along the same lines. In what follows by data we mean triples  $\{y(t), r(t), s(t)\}$ .

In each analysis the data are randomly divided into two sets called DA and DB without regard to the values of the data.

The maximum likelihood parameter estimates for each model (1)–(3) are obtained for each set DA and DB and for all the data. The estimated values appear in Table 1. The estimated variances  $\sigma_1^2(1,t)$ ,  $\sigma_1^2(2,t)$ ,  $\sigma_2^2(t)$ , are computed for the parameters estimated from DA and DB using (1)–(3) for each data point in DA and DB.

The models are for the variances of the observations rather than the observations themselves. One possible procedure to assess goodness-of-fit and cross-validate the models is by binning the data. To assess models (1) and (3) the data  $(y(t), r(t), s(t))$  are binned into 10 bins based on ordering the values of  $r(t)$  from smallest to largest. The data in the first bin correspond to the smallest values of  $r(t)$ ; the data in the 10th bin correspond to the largest values of  $r(t)$ . Each bin contains about  $\frac{1}{10}$ th of the data with the 10th bin containing a few more data. The averages of the estimated variances for models (1) and (3) are computed for each bin. The average  $y(t)^2$  is also computed for each bin.

To assess models (2) and (3) the same procedure is used but the binning is based on the values of  $s(t)$ .

Figures 1-24 present graphs of the  $\ln$  [average  $y(t)^2$ ] in each bin versus  $\ln$  [average estimated variance] in each bin for models (1) and (3) and models (2) and (3). Figures 1, 5, 9, 13, 17, 21 (respectively 2, 6, 10, 14, 18, 22) show the logarithm of the average of the  $y(t)^2$  values of DA (respectively DB) versus the logarithm of the average of the estimated variances for each bin using the estimated parameters from DA (respectively DB). If a model were perfect, a point should be close to the 45° line shown.

Figures 3, 7, 11, 15, 19, 23, (respectively 4, 8, 12, 16, 20, 24) present graphs of  $\ln$  average  $y(t)^2$  of DA (respectively DB) versus  $\ln$  average estimated variances using parameters estimated using data DB (respectively DA). Once again if the model were perfect, the points would be close to the 45° line.

Since the two-variate model (3) is shown with both one-variate models, it is possible to obtain some idea of the effect of the two different sets of bins on the  $\ln$  averages. In particular, the graphs corresponding to the 500 Mb height winds, Figures 9-16, show that the display of  $\ln$  averages can be quite sensitive to which variate is used to do the binning.

Keeping this binning sensitivity in mind, the figures suggest the following concerning the models using observed winds as covariates. It appears that of the two one-variate models, model (1) which uses  $r(t)$  as the covariate is the better. The two-variate model (3) appears not much better than model (1). If wind speed is used as the single covariate, it appears to overstate the variance; the addition of the second covariate  $r(t)$  in this case seems to tend to make the estimated variance smaller and bring the  $\ln$  average predicted variance in a bin closer to the  $\ln$  average  $y^2$  in the bin.

Preliminary examination of ln average  $y^2$  in bins and ln average model variances in bins for the Cauchy models suggests that the Cauchy models result in little or no improvement over the results of the normal model. The results of the Cauchy models will not be reported here.

Another way to assess goodness of fit and to cross validate is to evaluate the ln-likelihood for the different models at the parameter estimates. Larger values of the ln-likelihood suggest better model fit; cf. Cox and Hinkley [1974].

Table 2 presents the values of the ln-likelihood up to addition and multiplication of constants for the parameter estimates of Table 1; the function being evaluated is

$$\tilde{\ell} = -n\alpha - \sum_{i=1}^n \underline{x}_i \underline{\beta} - \sum_{i=1}^n y_i^2 \exp\{-\alpha + \underline{x}_i \underline{\beta}\}. \quad (7)$$

where  $\underline{x}_i \underline{\beta} = \sum_j x_{ij} \beta_j$ . The values of  $\tilde{\ell}$  are presented for data DA (respectively DB) using the parameters fit using DA (respectively DB); these are values assessing goodness of fit; since maximum likelihood is the estimation procedure, the largest value of  $\tilde{\ell}$  in each of these two rows is the one corresponding to the two-variate model. Values of  $\tilde{\ell}$  are also presented for data DA (respectively DB) using the parameters fit using DB (respectively DA); these are values assessing cross-validation. The underlined value in each row is the maximum value in that row; the corresponding model provides the best model fit. The bold italicized value in each row is the maximum value for the two one-variate models; the corresponding one-variate model provides the best model fit between the two one-variate models.

**TABLE 1. NORMAL MODELS  
PARAMETER ESTIMATES  
OBSERVED WIND COVARIATES**

Pressure Height	Wind Comp.	Data Set	One-Variate Models				Two-Variate Models			
			$r(t)$		$s(t)$		$\ln \text{MSE} = \alpha + \beta_1 r(t) + \beta_2 s(t)$			
			$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta_1$	$\beta_2$	
850	$u$	A	2.02	0.054	1.94	0.050	1.70	0.040	0.040	
		B	2.09	0.050	1.76	0.066	1.63	0.027	0.058	
		ALL	2.06	0.052	1.85	0.058	1.66	0.034	0.049	
850	$v$	A	2.19	0.040	1.59	0.080	1.51	0.015	0.076	
		B	2.05	0.051	1.68	0.071	1.56	0.028	0.062	
		ALL	2.12	0.045	1.64	0.076	1.53	0.022	0.069	
500	$u$	A	2.29	0.045	2.45	0.018	2.11	0.040	0.011	
		B	2.18	0.054	2.19	0.029	1.84	0.046	0.020	
		ALL	2.23	0.050	2.32	0.024	1.97	0.043	0.015	
500	$v$	A	2.31	0.039	2.27	0.023	1.99	0.033	0.018	
		B	2.24	0.042	2.14	0.028	1.89	0.034	0.021	
		ALL	2.28	0.041	2.21	0.025	1.94	0.034	0.019	
250	$u$	A	3.12	0.034	2.48	0.034	2.22	0.023	0.030	
		B	2.95	0.039	2.45	0.032	2.17	0.026	0.027	
		ALL	3.04	0.036	2.46	0.033	2.20	0.024	0.028	
250	$v$	A	3.01	0.031	2.36	0.033	2.13	0.021	0.029	
		B	2.97	0.032	2.28	0.035	2.12	0.021	0.028	
		ALL	2.98	0.031	2.31	0.034	2.12	0.021	0.029	

$$r(t) = [((u(t) - u(t-1))^2 + (v(t) - v(t-1))^2)]^{1/2}$$

$$s(t) = [u(t)^2 + v(t)^2]^{1/2}$$

NOTE: Data are divided into two sets randomly without regard to data values. One set is called A; the other is called B.

TABLE 2. NORMAL MODELS  
VALUES OF LN-LIKELIHOOD  
OBSERVED WIND COVARIATES

Pressure Height	Wind Comp.	Data Set	Model	Constant	One-Variate Models		Two- Variate Models
					$r(t)$	$s(t)$	
850	$u$	A	A	-7695.9	-7596.5	-7591.6	<u>-7538.1</u>
		B	B	-7746.9	-7661.8	-7560.5	<u>-7540.3</u>
		B	A	-7747.5	-7663.9	-7571.1	<u>-7553.2</u>
		A	B	-7696.4	-7598.5	-7601.9	<u>-7551.7</u>
850	$v$	A	A	-7759.1	-7703.4	-7505.5	<u>-7498.5</u>
		B	B	-7707.6	-7614.0	-7512.8	<u>-7489.4</u>
		B	A	-7708.2	-7620.6	-7515.4	<u>-7497.5</u>
		A	B	-7759.7	-7710.3	-7508.2	<u>-7507.4</u>
500	$u$	A	A	-9454.3	-9314.6	-9405.3	<u>-9299.2</u>
		B	B	-9518.7	-9296.8	-9376.7	<u>-9239.6</u>
		B	A	-9519.5	-9303.2	-9397.9	<u>-9257.4</u>
		A	B	-9455.2	-9320.4	-9425.2	<u>-9315.8</u>
500	$v$	A	A	-9317.7	-9317.1	-9243.0	<u>-9174.9</u>
		B	B	-9258.3	-9140.7	-9161.1	<u>-9091.7</u>
		B	A	-9259.1	-9142.8	-9165.1	<u>-9094.2</u>
		A	B	-9318.4	-9219.2	-9247.5	<u>-9177.5</u>
250	$u$	A	A	-11265.4	-10657.4	-10431.5	<u>-10344.8</u>
		B	B	-10782.7	-10389.8	-10249.0	<u>-10149.2</u>
		B	A	-10829.6	-10403.9	-10261.7	<u>-10162.1</u>
		A	B	-11319.3	-10673.4	-10445.1	<u>-10358.6</u>
250	$v$	A	A	-10417.8	-10259.4	-10094.9	<u>-10032.0</u>
		B	B	-10783.1	-10181.5	-10050.1	<u>-9960.1</u>
		B	A	-10814.9	-10182.5	-10051.8	<u>-9961.3</u>
		A	B	-10446.4	-10260.3	-10096.3	<u>-10033.2</u>

The models considered are  $\{Y_i\}$  are independent normal with mean 0 and variance

$$\sigma_i^2(t) = e^\alpha \quad (\text{Constant variance}) \quad (8)$$

and models (1)–(3).

The two-variate model (3) maximizes the cross-validation value of  $\tilde{\ell}$  for data DA (respectively DB) with a model using parameters fit using DB (respectively DA). This suggests that both  $r(t)$  and  $s(t)$  together have predictive ability.

For the one-variate models (1) and (2) the cross-validation value of  $\tilde{\ell}$  for DA (respectively DB) using the parameters fit using DB (respectively DA) are equally divided as to whether  $r(t)$  by itself or  $s(t)$  by itself produces the higher value of  $\tilde{\ell}$ . This suggests that neither variate by itself has obviously better predictive value than the other. The goodness of fit values of  $\tilde{\ell}$  for the one-variate models using DA (respectively DB) have a higher value of  $\tilde{\ell}$  associated with  $s(t)$  the majority of the time. This suggests that  $s(t)$  by itself provides a better description of the data than  $r(t)$  by itself.

Comparing the value of  $\tilde{\ell}$ ,  $\tilde{\ell}_c$ , for DA (respectively DB) using the constant variance model (8) fit using DA (respectively DB) with the corresponding cross-validation value of  $\tilde{\ell}$  for DA (respectively DB) using models (2), (3) fit using DB (respectively DA) indicates the following. The values of  $\tilde{\ell}$  for models (2) and (3) fit with the other half of the data are larger than the corresponding value  $\tilde{\ell}_c$  for the constant variance model fit using the data to be modeled. This indicates that both models (2) and (3) fit with the other half of the data describe the data better than the best constant variance model (8) fit with the same data it is used to summarize.

### 3.2 First Guess Wind Covariate Models

In this section we report the results of using models (1)–(3) and (8) with first guess winds as covariates; the two covariates considered are

$$r_f(t) = \left[ \left( U_f(t) - U_f(t-1) \right)^2 + \left( V_f(t) - V_f(t-1) \right)^2 \right]^{\frac{1}{2}}$$

and

$$s_f(t) = \left[ U_f(t)^2 + V_f(t)^2 \right]^{\frac{1}{2}}.$$

The analysis is the same as in the previous subsection. The data sets DA and DB are the same as those in the previous subsection in each case.

The values of the parameter estimates appear in Table 3. The corresponding values of  $\tilde{\ell}$  appear in Table 4. Once again the underlined value of  $\tilde{\ell}$  is the largest value in each row; the bold italicized value  $\tilde{\ell}$  is the largest value between the two one-variate models.

In all but two cases the values of  $\tilde{\ell}$  for the observed wind covariates are larger than those for the first-guess wind covariates. This suggests that the observed wind components have better predictive and descriptive value than the first guess wind components.

Table 4 also indicates the following results concerning models using first guess wind covariates. Between the two one-variate models (1) and (2) the one-variate model using first guess wind speed always has the greater  $\tilde{\ell}$ -value. This suggests that first guess wind speed alone has better predictive and descriptive value than  $r_f(t)$  alone. The cross-validation values of  $\tilde{\ell}$  for data DA (respectively DB) using parameters fit with DB (respectively DA) are maximized about half the time using the one-variate model with  $s_f(t)$ . The other times the maximal  $\tilde{\ell}$  is associated with the two-variate model.



**TABLE 3. NORMAL MODELS  
PARAMETER ESTIMATES  
FIRST GUESS WIND COVARIATES**

Pressure Heights	Wind Comp.	Data Set	One-variate Models				Two Variate Models		
			$r_f(t)$		$s_f(t)$		$\alpha$	$\beta_1$	$\beta_2$
			$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta_1$	$\beta_2$
850	$u$	A	2.52	-0.006	2.23	0.025	2.30	-0.023	0.031
		B	2.48	0.004	2.21	0.029	2.26	-0.013	0.032
		ALL	2.50	-0.0007	2.22	0.027	2.28	-0.017	0.032
850	$v$	A	2.54	-0.004	2.34	0.017	2.40	-0.016	0.021
		B	2.41	0.013	2.27	0.021	2.28	-0.001	0.022
		ALL	2.47	0.005	2.31	0.019	2.34	-0.008	0.021
500	$u$	A	2.61	0.023	2.35	0.023	2.25	0.015	0.021
		B	2.68	0.019	2.48	0.017	2.39	0.014	0.016
		ALL	2.65	0.021	2.42	0.020	2.31	0.014	0.018
500	$v$	A	2.71	0.006	2.40	0.018	2.39	0.0008	0.017
		B	2.76	-0.002	2.29	0.022	2.35	-0.009	0.023
		ALL	2.73	0.002	2.34	0.020	2.37	-0.004	0.020
250	$u$	A	4.02	-0.005	2.66	0.037	2.67	-0.002	0.037
		B	3.46	0.021	3.01	0.022	2.78	0.019	0.021
		ALL	3.74	0.009	2.79	0.031	2.69	0.008	0.031
250	$v$	A	3.49	0.007	3.00	0.017	2.93	0.006	0.017
		B	3.53	0.016	2.87	0.026	2.60	0.018	0.026
		ALL	3.50	0.012	2.92	0.022	2.75	0.013	0.022

TABLE 4. NORMAL MODEL  
FIRST GUESS WIND COVARIATES  
VALUE OF LIKELIHOOD

Pressure Height	Wind Comp.	Data Set	Model	Constant	One Variate Models		Two Variate Models
					$r_f(t)$	$s_f(t)$	
850	$u$	A	A	-7695.9	-7695.1	<u>-7667.6</u>	<u>-7657.9</u>
		B	B	-7746.9	-7746.6	-7708.8	<u>-7705.6</u>
		B	A	-7747.5	-7749.5	-7709.8	<u>-7708.8</u>
		A	B	-7696.4	-7697.7	-7668.6	<u>-7660.9</u>
850	$v$	A	A	-7759.1	-7758.7	-7745.4	<u>-7741.1</u>
		B	B	-7707.6	-7703.7	-7685.5	<u>-7685.5</u>
		B	A	-7708.2	-7711.3	<u>-7687.3</u>	-7691.8
		A	B	-7759.7	-7765.3	-7747.3	<u>-7746.7</u>
500	$u$	A	A	-9454.3	-9433.7	-9391.1	<u>-9383.0</u>
		B	B	-9518.7	-9505.7	-9481.9	<u>-9475.2</u>
		B	A	-9519.5	-9507.5	-9486.2	<u>-9479.1</u>
		A	B	-9455.2	-9435.5	-9395.4	<u>-9387.0</u>
500	$v$	A	A	-9317.7	-9316.0	-9281.2	<u>-9281.2</u>
		B	B	-9258.3	-9258.2	-9202.6	<u>-9199.5</u>
		B	A	-9259.1	-9261.5	<u>-9205.6</u>	-9206.2
		A	B	-9318.4	-9319.7	<u>-9284.3</u>	-9288.4
250	$u$	A	A	-11265.4	-11263.9	-10907.3	<u>-10907.0</u>
		B	B	-10782.7	-10745.7	-10684.5	<u>-10653.5</u>
		B	A	-10829.6	-10846.5	<u>-10739.9</u>	-10745.7
		A	B	-11319.3	-11371.6	<u>-10987.7</u>	-11035.7
250	$v$	A	A	-10417.8	-10414.2	-10349.3	<u>-10346.7</u>
		B	B	-10783.1	-10758.9	-10622.4	<u>-10587.8</u>
		B	A	-10814.9	-10796.9	-10658.7	<u>-10640.3</u>
		A	B	-10446.4	-10446.4	<u>-10379.2</u>	-10389.1

Comparing the value of  $\tilde{\ell}$ ,  $\tilde{\ell}_c$ , for DA (respectively DB) using the constant variance model (8) fit using DA (respectively DB) with the cross-validation value of  $\tilde{\ell}$  for DA (respectively DB) using models (2), (3) fit using DB (respectively DA) indicates the following. The values of  $\tilde{\ell}$  for models (2) and (3) fit with the other half of the data are always larger than the corresponding value  $\tilde{\ell}_c$  for the constant variance model fit using the data to be modeled. This suggests that both models (2) and (3) fit with the other half of the data describe the data somewhat better than the best constant variance model (8) fit with the data to be described.

In summary, based on values of  $\tilde{\ell}$ , when first guess winds are used as covariates it appears that the one-variate model using first guess wind speed is an attractive choice for predictive purposes. When observational winds are used as covariates, the two-variate model appears to have the best predictive value.

Assessing goodness of fit and cross-validation using values of  $\tilde{\ell}$  has the advantage of not being sensitive to binning. However,  $\tilde{\ell}$  may be sensitive to data sets DA and DB. Further work needs to be done to develop procedures to assess goodness of fit and for cross-validation. Procedures based on bootstrapping or jackknifing hold some promise.

#### 4. THE DATA ANALYSIS—APRIL AND FEBRUARY DATA

In this section we report results of an assessment of goodness of fit for the normal models (1)–(3) for April data. We also report results concerning using a model whose parameters are fit using February data (respectively April) data to model April data (respectively February) data.

#### 4.1 Observed Wind Covariate Models

In this subsection we report results for normal models (1)–(3) using observed wind components as covariates. There are six analyses; one for the  $u$ -wind component (respectively  $v$ -wind component) for each pressure height.

Table 5 shows the values of the parameter estimates for both the February data and April data. Table 6 shows the values of  $\tilde{\ell}$  for February data (respectively April data) using parameters fit using February data (respectively April data). Values of  $\tilde{\ell}$  are also presented for February data (respectively April data) using parameters fit using April data (respectively February data). Once again, larger values of  $\tilde{\ell}$  indicate better model fit. The underlined value in each row is the maximum value in that row. The bold italicized value in each row is the maximum value of  $\tilde{\ell}$  for the two one-variate models.

The values of  $\tilde{\ell}$  for February data (respectively April data) using parameters fit using April data (respectively February data) are maximized by the two-variate model in all but one case; between the two one-variate models  $\tilde{\ell}$  is the maximized half the time for the model involving  $s(t)$ .

Comparing the value of  $\tilde{\ell}$ ,  $\tilde{\ell}_c$ , for the model of constant variance (8) for February (respectively April) data fit using February (respectively April) data with that for the prediction value of  $\tilde{\ell}$  for the models (2)–(3) for February (respectively April) data fit using April (respectively February) data indicate the following. The values of  $\tilde{\ell}$  for models (2) and (3) fit with data from the other month are always larger than the corresponding values of  $\tilde{\ell}_c$  fit with the data of the same month. This suggests that models (2) and (3) fit using data from the other month have predictive value over a model of constant variance fit using the data that is to be modeled.

**TABLE 5. NORMAL MODELS  
PARAMETER ESTIMATES  
OBSERVED WIND COVARIATES**

Pressure Height	Wind Comp.	Data Set	One-Variate Models				Two-Variate Models		
			$r(t)$		$s(t)$		ln MSE = $\alpha + \beta_1 r(t) + \beta_2 s(t)$		
			$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta_1$	$\beta_2$
850	$u$	Feb.	2.06	0.052	1.85	0.058	1.66	0.034	0.049
		Apr.	1.86	0.084	1.69	0.086	1.44	0.053	0.068
850	$v$	Feb.	2.12	0.045	1.64	0.076	1.53	0.022	0.069
		Apr.	1.83	0.089	1.69	0.090	1.42	0.062	0.065
500	$u$	Feb.	2.23	0.050	2.32	0.024	1.97	0.043	0.015
		Apr.	2.12	0.058	2.20	0.030	1.80	0.049	0.023
500	$v$	Feb.	2.28	0.041	2.21	0.024	1.94	0.034	0.019
		Apr.	2.02	0.065	1.97	0.041	1.66	0.052	0.027
250	$u$	Feb.	3.04	0.036	2.46	0.033	2.20	0.024	0.028
		Apr.	2.77	0.044	2.69	0.027	2.33	0.037	0.018
250	$v$	Feb.	2.98	0.031	2.31	0.034	2.12	0.021	0.029
		Apr.	2.73	0.041	2.61	0.027	2.26	0.034	0.019

$$r(t) = [((u(t) - u(t-1))^2 + (v(t) - v(t-1))^2)]^{1/2}$$

$$s(t) = [u(t)^2 + v(t)^2]^{1/2}$$

TABLE 6. NORMAL MODELS  
VALUES OF LN-LIKELIHOOD  
OBSERVED WIND COVARIATES

Pressure Height	Wind Comp.	Data Set	Model	Constant	One-Variate Models		Two- Variate Models
					$r(t)$	$s(t)$	
850	$u$	Feb.	Feb.	-15443.1	-15259.3	<u>-15157.3</u>	<u>-15084.9</u>
		Apr.	Apr.	-15709.1	-15311.6	<u>-15169.9</u>	<u>-15043.6</u>
		Apr.	Feb.	-15713.2	-15368.2	<u>-15238.7</u>	<u>-15117.7</u>
		Feb.	Apr.	-15447.0	-15321.2	<u>-15240.3</u>	<u>-15171.9</u>
850	$v$	Feb.	Feb.	-15467.0	-15320.8	<u>-15019.6</u>	<u>-14992.1</u>
		Apr.	Apr.	-15819.1	-15320.6	<u>-15307.2</u>	<u>-15106.4</u>
		Apr.	Feb.	-15827.8	-15429.9	<u>-15386.7</u>	<u>-15251.5</u>
		Feb.	Apr.	-15475.3	-15428.6	<u>-15105.9</u>	<u>-15121.1</u>
500	$u$	Feb.	Feb.	-18973.4	-18614.5	-18792.2	<u>-18547.4</u>
		Apr.	Apr.	-18504.1	-18083.1	-18270.7	<u>-17969.4</u>
		Apr.	Feb.	-18528.9	-18093.4	-18280.6	<u>-17989.1</u>
		Feb.	Apr.	-18999.9	-18625.3	-18804.3	<u>-18576.6</u>
500	$v$	Feb.	Feb.	-18576.4	-18358.8	-18406.2	<u>-18267.9</u>
		Apr.	Apr.	-18698.2	-17869.0	-18014.4	<u>-17733.0</u>
		Apr.	Feb.	-18699.0	-17938.0	-18086.9	<u>-17796.2</u>
		Feb.	Apr.	-18577.1	-18421.3	-18480.2	<u>-18345.2</u>
250	$u$	Feb.	Feb.	-22073.2	-21054.7	<u>-20687.1</u>	<u>-20500.6</u>
		Apr.	Apr.	-22712.2	-20364.3	-20658.3	<u>-20195.9</u>
		Apr.	Feb.	-22712.4	-20439.9	-20703.3	<u>-20276.4</u>
		Feb.	Apr.	-22073.4	-21127.5	<u>-20726.0</u>	<u>-20591.9</u>
250	$v$	Feb.	Feb.	-21216.0	-20441.4	<u>-20145.7</u>	<u>-19992.6</u>
		Apr.	Apr.	-21205.7	-20006.9	-20274.8	<u>-19840.6</u>
		Apr.	Feb.	-21240.2	-20065.6	-20338.7	<u>-19924.9</u>
		Feb.	Apr.	-21252.5	-20498.8	<u>-20190.2</u>	<u>-20070.5</u>

## 4.2 First Guess Wind Covariate Models

In this section we report results for normal models (1)–(3) using first guess wind components as covariates.

Table 7 shows the values of the parameter estimates for both February data and April data. Table 8 shows the values of  $\tilde{\ell}$  for February data (respectively April data) using parameters fit using February data (respectively April data). Values of  $\tilde{\ell}$  are also presented for February data (respectively April data) using parameters fit using April data (respectively February data). The underlined value in each row is the maximum value in that row. The bold italicized value in each row is the maximum value of  $\tilde{\ell}$  for the two one-variate models.

The values of  $\tilde{\ell}$  for the observed wind covariates are larger than those for the first guess wind covariates except for two values associated with the one-variate model using  $s(t)$  to model  $u$ -wind component error at the 250 mb height for the model using parameters fit with the same data. This suggests that the observed wind covariates provide a better model of the data both in terms of goodness-of-fit and prediction.

The values of  $\tilde{\ell}$  for February data (respectively April data) using parameters fit using April data (respectively February data) are maximized about half the time by the two-variate model and the other half the time by the one-variate model using the first guess wind speed  $s(t)$ .

**TABLE 7. NORMAL MODELS  
PARAMETER ESTIMATES  
FIRST GUESS WIND COVARIATES**

Pressure Heights	Wind Comp.	Data Set	One-variate Models				Two-Variate Models		
			$r_f(t)$		$s_f(t)$		$\alpha$	$\beta_1$	$\beta_2$
			$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta_1$	$\beta_2$
850	$u$	Feb.	2.50	-0.0007	2.22	0.027	2.28	-0.017	0.032
		Apr.	2.42	0.022	2.19	0.041	2.19	-0.002	0.041
850	$v$	Feb.	2.47	0.005	2.31	0.019	2.34	-0.008	0.021
		Apr.	2.46	0.019	2.23	0.039	2.24	-0.004	0.040
500	$u$	Feb.	2.65	0.021	2.41	0.020	2.32	0.014	0.018
		Apr.	2.50	0.031	2.23	0.030	2.11	0.019	0.028
500	$v$	Feb.	2.73	0.002	2.34	0.020	2.37	-0.004	0.020
		Apr.	2.30	0.061	1.98	0.045	1.74	0.040	0.040
250	$u$	Feb.	3.74	0.009	2.79	0.031	2.69	0.008	0.031
		Apr.	4.00	-0.010	3.48	0.014	3.63	-0.021	0.016
250	$v$	Feb.	3.50	0.012	2.92	0.022	2.75	0.013	0.022
		Apr.	3.26	0.025	2.80	0.026	2.68	0.015	0.024



**TABLE 8. NORMAL MODEL  
VALUE OF LN-LIKELIHOOD  
FIRST GUESS WIND COVARIATES**

Pressure Height	Wind Comp.	Data Set	Model	Constant	One-Variate Models		Two- Variate Models
					$r_f(t)$	$s_f(t)$	
850	$u$	Feb.	Feb.	-15443.1	-15443.0	<u>-15376.0</u>	<u>-15365.0</u>
		Apr.	Apr.	-15709.1	-15695.3	-15593.3	<u>-15593.1</u>
		Apr.	Feb.	-15713.2	-15714.0	<u>-15620.6</u>	-15624.2
		Feb.	Apr.	-15447.0	-15471.3	<u>-15413.2</u>	<u>-15409.9</u>
850	$v$	Feb.	Feb.	-15467.0	-15466.0	<u>-15431.8</u>	<u>-15429.5</u>
		Apr.	Apr.	-15819.1	-15808.8	<u>-15716.0</u>	<u>-15715.7</u>
		Apr.	Feb.	-15827.8	-15824.2	<u>-15759.4</u>	<u>-15757.7</u>
		Feb.	Apr.	<u>-15475.3</u>	-15846.6	<u>-15493.2</u>	-15490.5
500	$u$	Feb.	Feb.	-18973.4	-18940.3	<u>-18875.1</u>	<u>-18860.2</u>
		Apr.	Apr.	-18504.1	-18455.5	<u>-18331.3</u>	<u>-18312.9</u>
		Apr.	Feb.	-18528.9	-18473.2	<u>-18351.3</u>	<u>-18332.6</u>
		Feb.	Apr.	-18999.9	-18956.6	<u>-18897.5</u>	<u>-18885.5</u>
500	$v$	Feb.	Feb.	-18576.4	-18575.9	<u>-18485.3</u>	<u>-18484.1</u>
		Apr.	Apr.	-18698.2	-18509.1	<u>-18289.2</u>	<u>-18208.2</u>
		Apr.	Feb.	-18699.0	-18683.4	<u>-18423.3</u>	-18441.4
		Feb.	Apr.	<u>-18577.1</u>	-18805.8	<u>-18656.5</u>	-18778.6
250	$u$	Feb.	Feb.	-22073.2	-22061.3	<u>-21624.3</u>	<u>-21613.7</u>
		Apr.	Apr.	-22712.2	-22699.3	<u>-22641.2</u>	<u>-22609.5</u>
		Apr.	Feb.	<u>-22712.4</u>	-22739.2	-22906.5	-22987.4
		Feb.	Apr.	-22073.4	-22139.0	<u>-21792.9</u>	-21863.5
250	$v$	Feb.	Feb.	-21216.0	-21190.5	<u>-20988.0</u>	<u>-20957.8</u>
		Apr.	Apr.	-21205.7	-21142.9	<u>-20919.7</u>	<u>-20899.6</u>
		Apr.	Feb.	-21240.2	-21181.4	<u>-20927.1</u>	<u>-20902.2</u>
		Feb.	Apr.	-21252.5	-21232.3	<u>-20994.4</u>	<u>-20961.4</u>

A comparison of the value of  $\tilde{\ell}$ ,  $\tilde{\ell}_c$ , for the constant variance model of February (respectively April) data fit using the same month February (respectively April) data and the prediction values of  $\tilde{\ell}$  for models (1)–(3) of February (respectively April) data fit using the other month of April (respectively February) indicate the following. A little fewer than half the time  $\tilde{\ell}_c$  is smaller than the corresponding values of  $\tilde{\ell}$  for models (1)–(3) fit with the other month's data. This suggests that the first-guess wind speed models fit using the other month's data may not describe the data as well as a constant variance model fit using the data being modeled. This may be an indication that models fit using first-guess February wind (respectively April wind) data are not good predictors of April (respectively February) wind component error.

### 4.3 Conclusions

Models (2) and (3) using observed wind components as covariates and fit using February (respectively April) data appear to have predictive value for April (respectively February) data. It is less clear if models (1)–(3) using first-guess wind components as covariates and fit using February (respectively April) data have predictive value for April (respectively February) wind component error data. It might be that models (1)–(3) fit with first-guess data from other Aprils (respectively Februarys) are better predictors of April (respectively February) wind component error. Alternatively, if first-guess winds are to be used as predictors, it might be worthwhile to develop a procedure to update the fitted model parameters using new data as it comes in.

## REFERENCES

- D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Chapman and Hall, London, 1974.
- B. Efron (1986) "Double exponential families and their use in generalized linear regression." *J. Amer. Statist. Assoc.*, **81** pp. 709-721.
- J. S. Goerss and P. A. Phoebus (1991a). *The Multivariate Optimum Interpolation Analysis of Meteorological Data at FNOG*. NOARL Report Number 31, Naval Oceanographic and Atmospheric Research Laboratory, Stennis Space Center, MS.
- J. S. Goerss and P. A. Phoebus (1991b). "The Navy's operational atmospheric analysis." To appear in *Weather and Forecasting*.
- R. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York, 1986.
- T. F. Hogan and T. E. Rosmond (1991) "The description of the Navy operational global atmospheric prediction system's spectral forecast model. *Monthly Weather Review*, **119**, No. 8, pp. 1786-1815.
- P. J. Huber, *Robust Statistics*, Wiley, New York, 1981.
- B. Jorgenson (1987) "Exponential dispersion models (with discussion)." *J. Royal Statist. Soc.*, **49**, pp. 127-162.
- F. Mosteller and J. W. Tukey, *Data Analysis and Regression*, Addison-Wesley, 1977.

## APPENDIX A

### MAXIMUM LIKELIHOOD ESTIMATION FOR THE NORMAL MODEL

Let  $Y_1, Y_2, \dots, Y_n$  be independent normal random variables with mean 0 and variances

$$\sigma_i^2 = \exp\left\{\alpha + \sum_{j=1}^P x_{ij}\beta_j\right\} = \exp\{\alpha + \underline{x}_i\underline{\beta}\} \quad i = 1, \dots, n \quad (\text{A.1})$$

where  $(x_{i1}, \dots, x_{ip})$  are fixed explanatory variables associated with  $Y_i$ .

The likelihood function for this model is

$$L(\alpha, \underline{\beta}; \underline{y}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(\alpha + \underline{x}_i\underline{\beta})\right\} \exp\left\{-\frac{1}{2}y_i^2 \exp\{-(\alpha + \underline{x}_i\underline{\beta})\}\right\}. \quad (\text{A.2})$$

Hence, the ln-likelihood function is

$$\ell(\alpha, \underline{\beta}; \underline{y}) = \frac{1}{2} \left[ -n\alpha - \sum_{i=1}^n \underline{x}_i\underline{\beta} - \sum_{i=1}^n y_i^2 \exp\{-(\alpha + \underline{x}_i\underline{\beta})\} \right] - n \frac{1}{2} \ln 2\pi. \quad (\text{A.3})$$

Computing partial derivatives of  $\ell$  with respect to  $\alpha$  and  $\beta_j$  results in

$$\frac{\partial}{\partial \alpha} \ell(\alpha, \underline{\beta}; \underline{y}) = \frac{1}{2} \left[ -n + \sum_{i=1}^n y_i^2 \exp\{-(\alpha + \underline{x}_i\underline{\beta})\} \right] \quad (\text{A.4})$$

$$\frac{\partial}{\partial \beta_j} \ell(\alpha, \underline{\beta}; \underline{y}) = \frac{1}{2} \left[ -\sum_{i=1}^n x_{ij} + \sum_{i=1}^n y_i^2 \exp\{-(\alpha + \underline{x}_i\underline{\beta})\} x_{ij} \right]. \quad (\text{A.5})$$

Setting  $\frac{\partial}{\partial \alpha} \ell = 0$  results in the equation

$$e^\alpha = \frac{1}{n} \sum_{i=1}^n y_i^2 \exp\{-\underline{x}_i\underline{\beta}\}. \quad (\text{A.6})$$

Setting  $\frac{\partial}{\partial \beta_j} \ell = 0$  and replacing  $e^\alpha$  by (A.6) yields the equation

$$0 = f_j(\underline{\beta}) = -\bar{x}_j \sum_{i=1}^n y_i^2 e^{-x_i \underline{\beta}} + \sum_{i=1}^n y_i^2 e^{-x_i \underline{\beta}} x_{ij} \quad (\text{A.7})$$

where  $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ .

Further,

$$\frac{\partial}{\partial \beta_k} f_j(\underline{\beta}) = +\bar{x}_j \sum_{i=1}^n y_i^2 e^{-x_i \underline{\beta}} x_{ik} - \sum_{i=1}^n y_i^2 e^{-x_i \underline{\beta}} x_{ij} x_{ik}. \quad (\text{A.8})$$

If  $f_k(\underline{\beta}) = 0$ , then

$$\bar{x}_k \sum_{i=1}^n y_i^2 e^{-x_i \underline{\beta}} = \sum_{i=1}^n y_i^2 e^{-x_i \underline{\beta}} x_{ik}. \quad (\text{A.9})$$

Substituting (A.9) into (A.8) yields

$$\frac{\partial}{\partial \beta_k} f_j(\underline{\beta}) \approx - \sum_{i=1}^n y_i^2 e^{-x_i \underline{\beta}} (x_{ij} x_{ik} - \bar{x}_j \bar{x}_k). \quad (\text{A.10})$$

An iteration of a Newton procedure to solve the system of equations  $0 = f_j(\underline{\beta})$ , ( $j = 1, \dots, p$ ) yields the system of linear equations

$$0 = f_j(\underline{\beta}) \approx f_j(\underline{\beta}^0) + \sum_{k=1}^P \left[ \frac{\partial}{\partial \beta_k} f_j(\underline{\beta}^0) \right] [\beta_k - \beta_k^0] \quad (\text{A.11})$$

$$= \sum_{i=1}^n y_i^2 e^{-x_i \underline{\beta}^0} (x_{ij} - \bar{x}_j) - \sum_{k=1}^P \sum_{i=1}^n y_i^2 e^{-x_i \underline{\beta}^0} [x_{ij} x_{ik} - \bar{x}_j \bar{x}_k] (\beta_k - \beta_k^0) \quad (\text{A.12})$$

where  $\underline{\beta}^0$  is the current value for  $\underline{\beta}$ . This system of linear equations is solved for  $\{\beta_k\}$ . The Newton procedure is iterated until it converges. The resulting

$\{\hat{\beta}_k\}$  are the maximum likelihood estimates of  $\{\beta_k\}$  and the estimate of  $\alpha$  is obtained from (A.6).

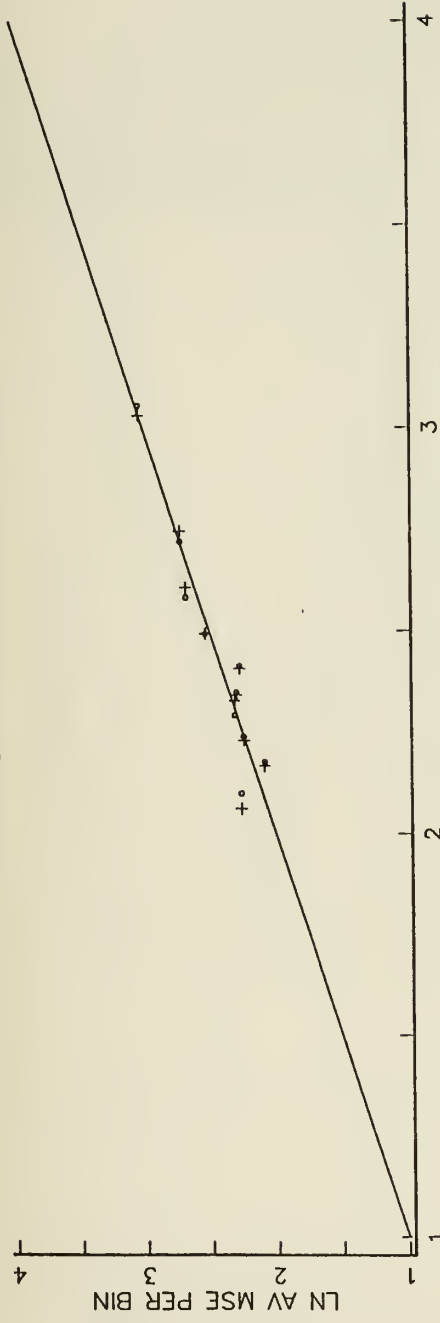
To summarize, the procedure to estimate  $\alpha$  and  $\{\beta_j\}$  is as follows

1. Start with  $\beta_j^0 = 0, j = 1, \dots, p$ .
2. Solve the linear system of equations (A.12) for  $\{\beta_k\}$ .
3. If  $\max_j \frac{|\beta_j - \beta_j^0|}{\beta_j^0} < 0.001$  stop and set  $\hat{\beta}_j = \beta_j$ ; otherwise set  $\beta_j^0 = \beta_j$  and return to step 2.
4. Compute

$$\hat{\alpha} = \ln \frac{1}{n} \sum_{i=1}^n y_i^2 \exp\{-x_i \hat{\beta}\}$$

850 MB U WIND; MODEL A ON DATA A; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

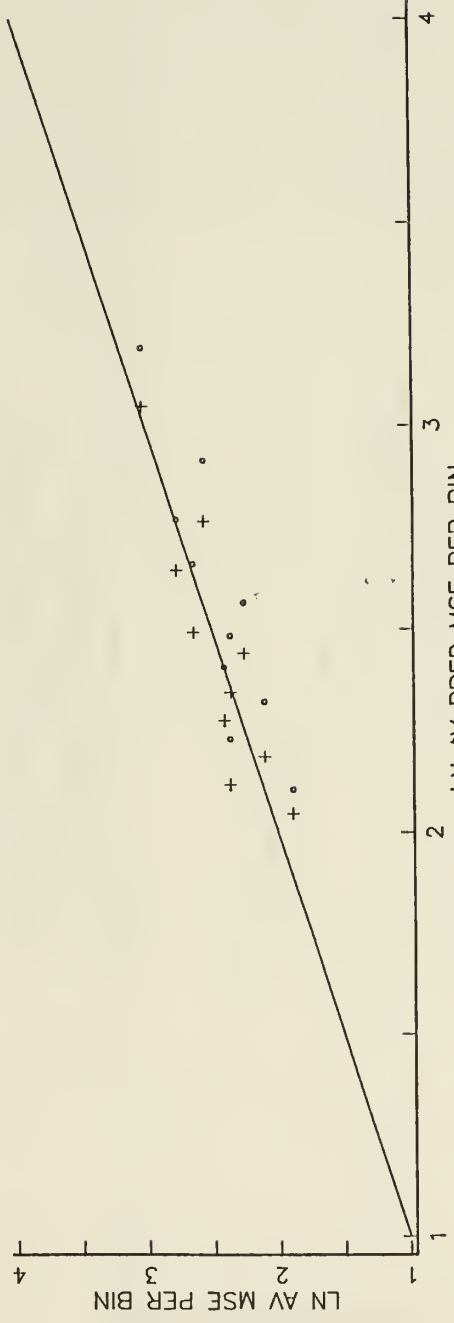
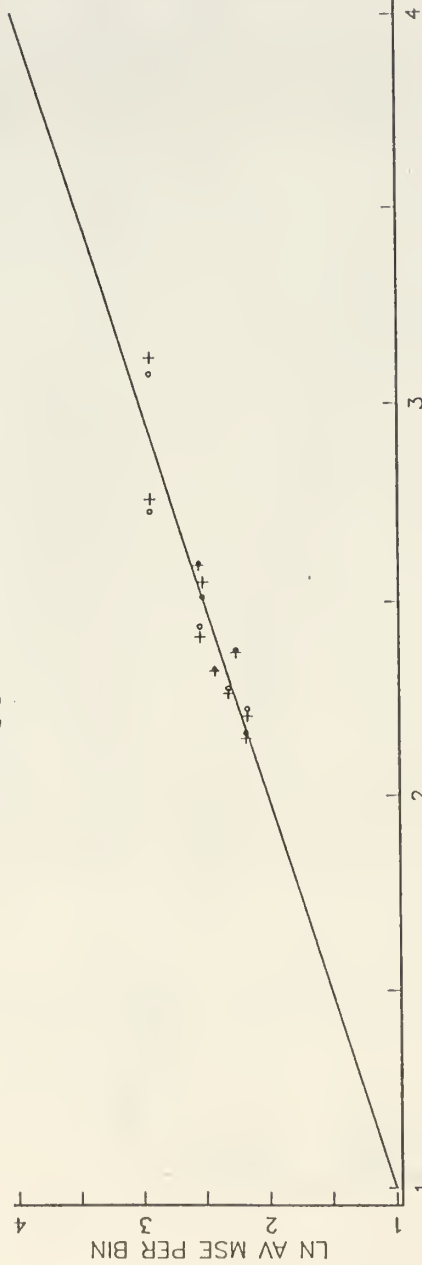


Figure 1

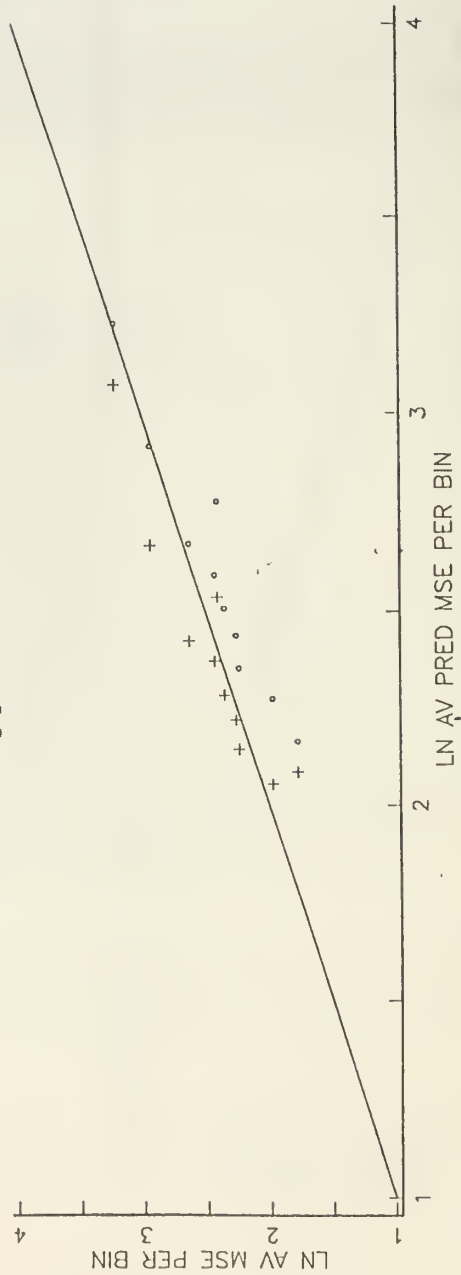
850 MB U WIND; MODEL B ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 2



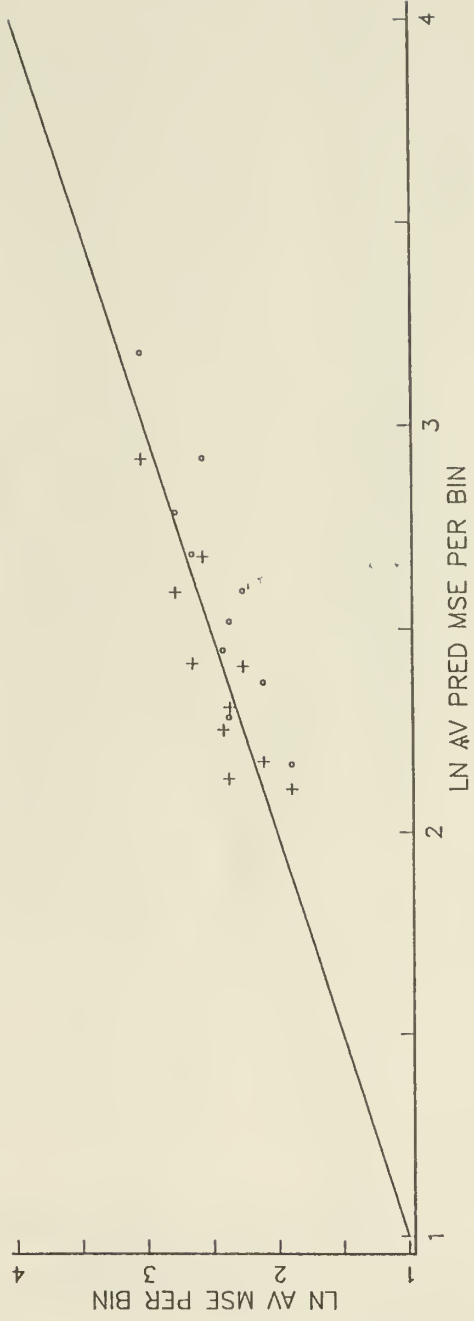
850 MB U WIND; MODEL B: ON DATA A; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 3

850 MB U WIND; MODEL A ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



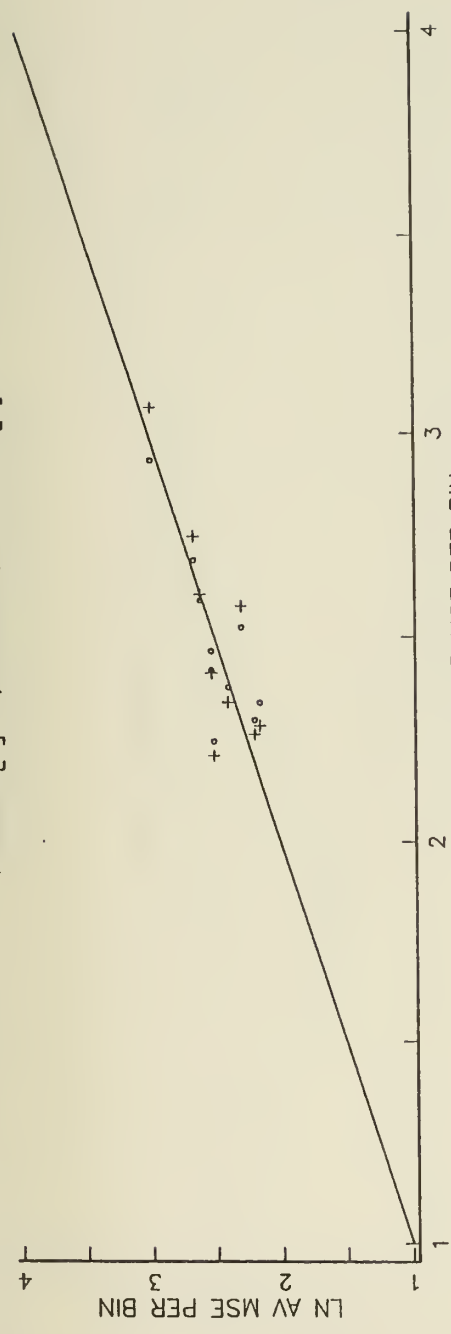
LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



850 MB V WIND; MODEL A ON DATA A; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN  
1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

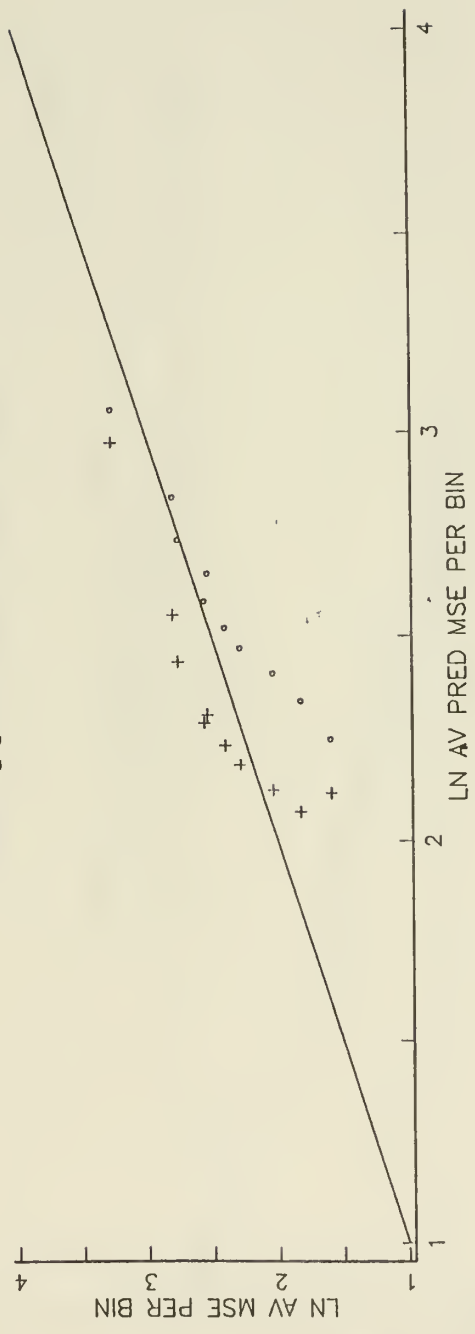


Figure 5

850 MB V WIND; MODEL B ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

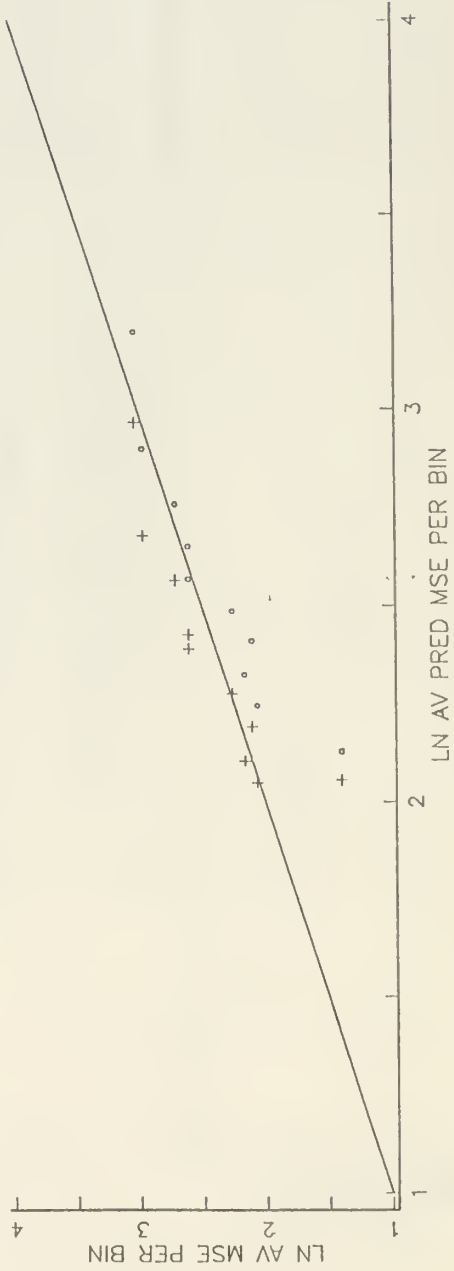
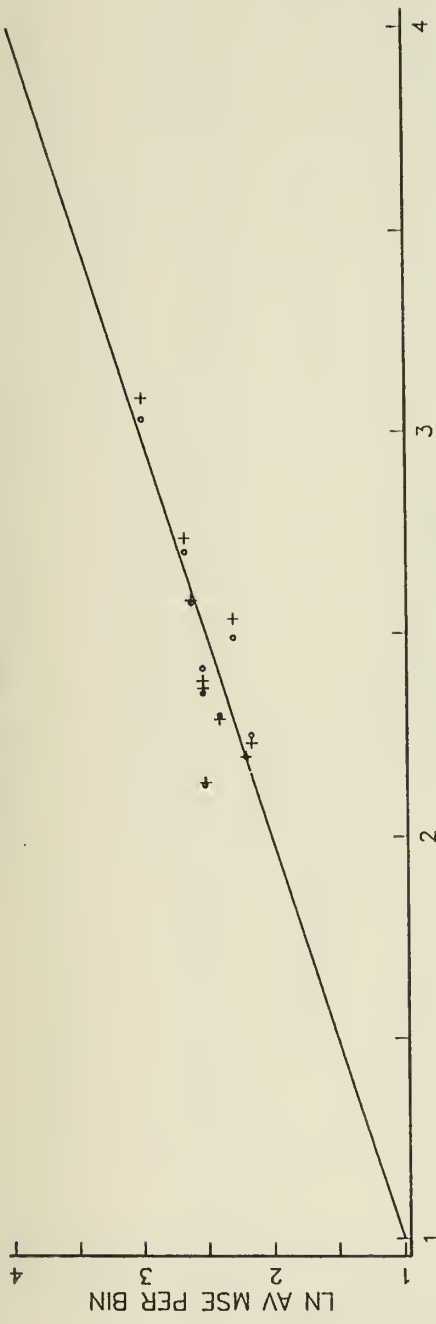


Figure 6

850 MB V WIND; MODEL B ON DATA A; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

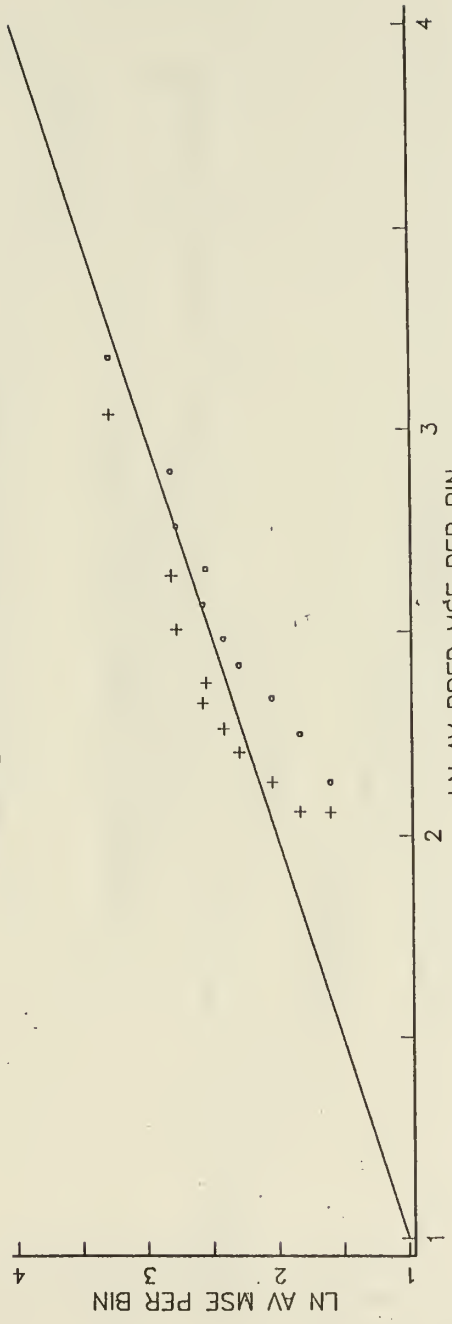


Figure 7

850 MB V WIND; MODEL A ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

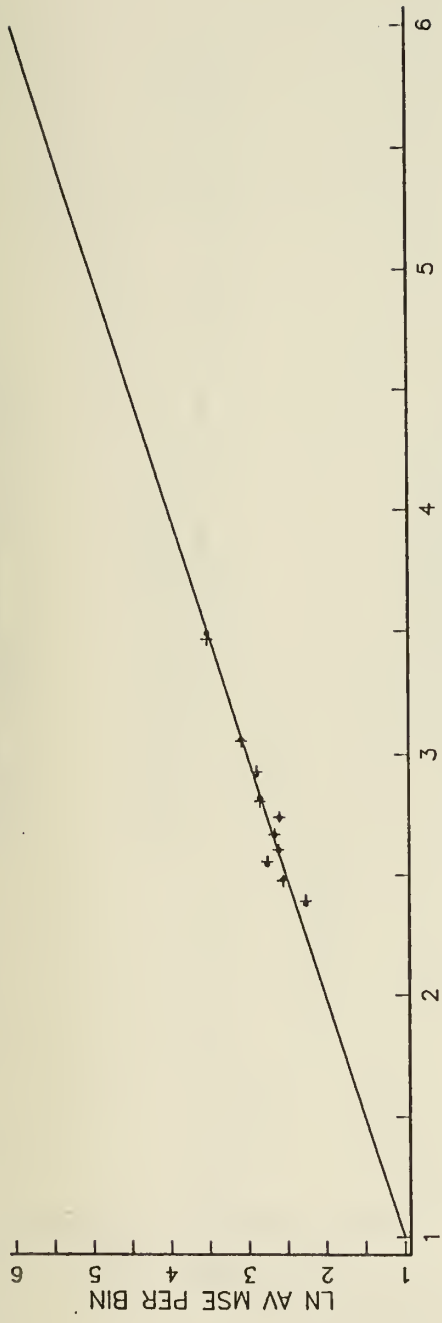


LN AV PRED MSE PER BIN

Figure 8

500 MB U WIND; MODEL A ON DATA A; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



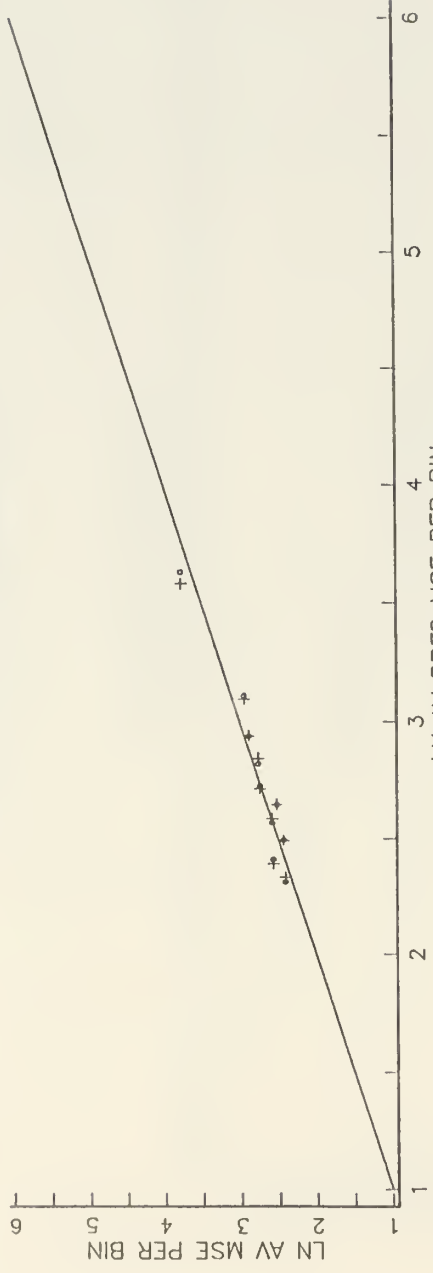
1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



Figure 9

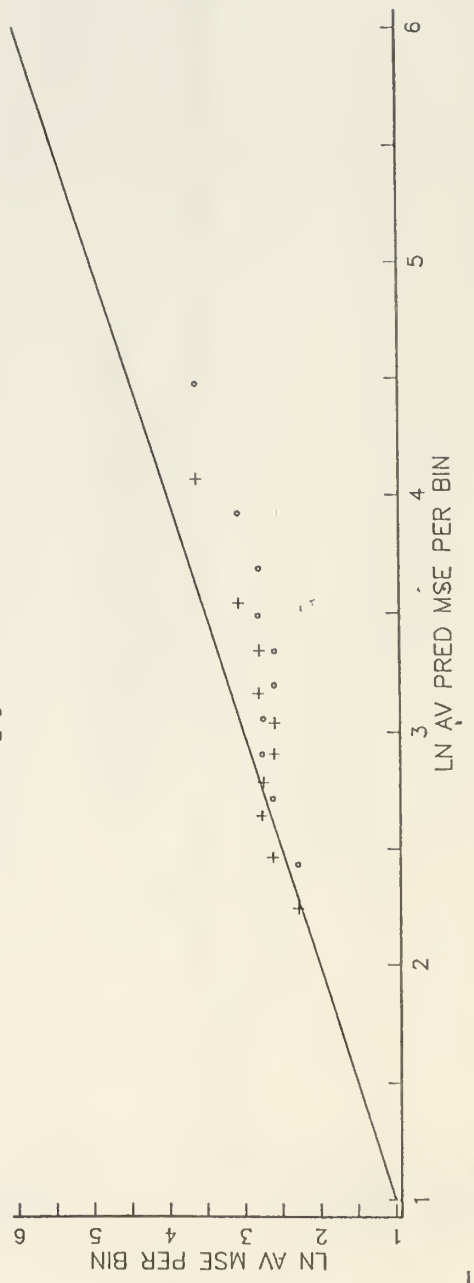
500 MB U WIND; MODEL B ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



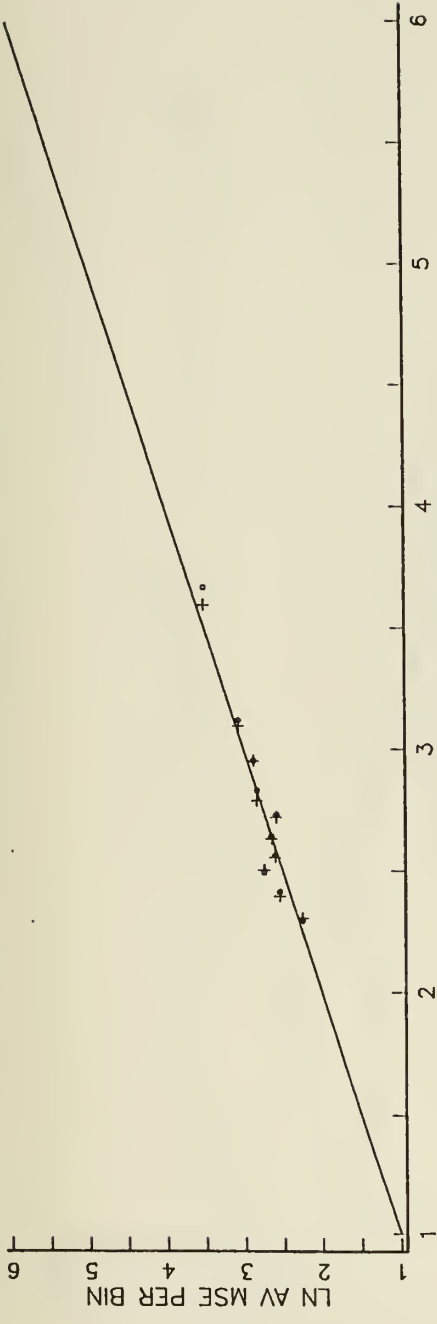
LN AV PRED MSE PER BIN

Figure 10



500 MB U WIND; MODEL B ON DATA A; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 11

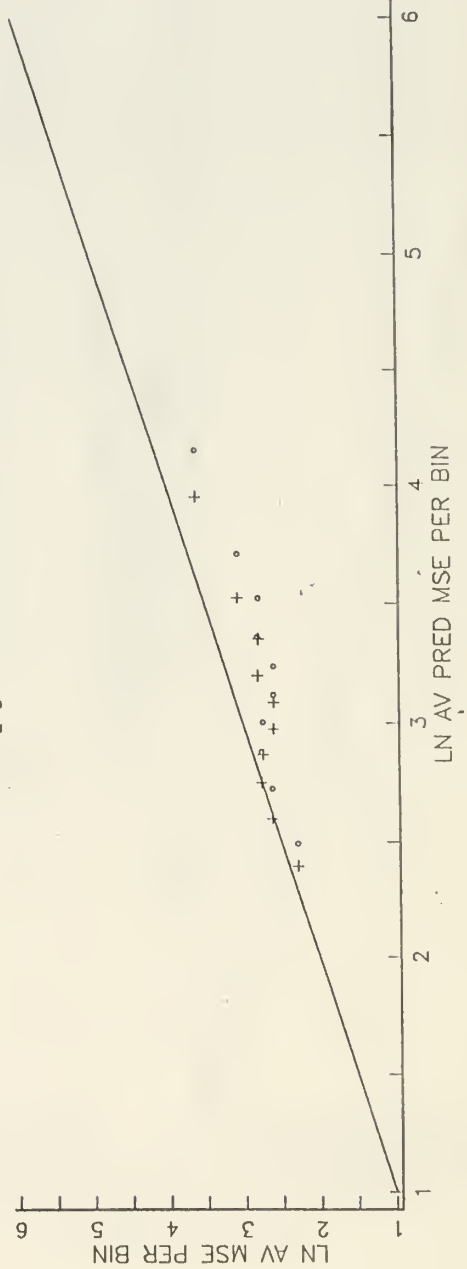
500 MB U WIND; MODEL A ON DATA B; FEB OBS WIND

1VAR=R[T]=°; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=°; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 12

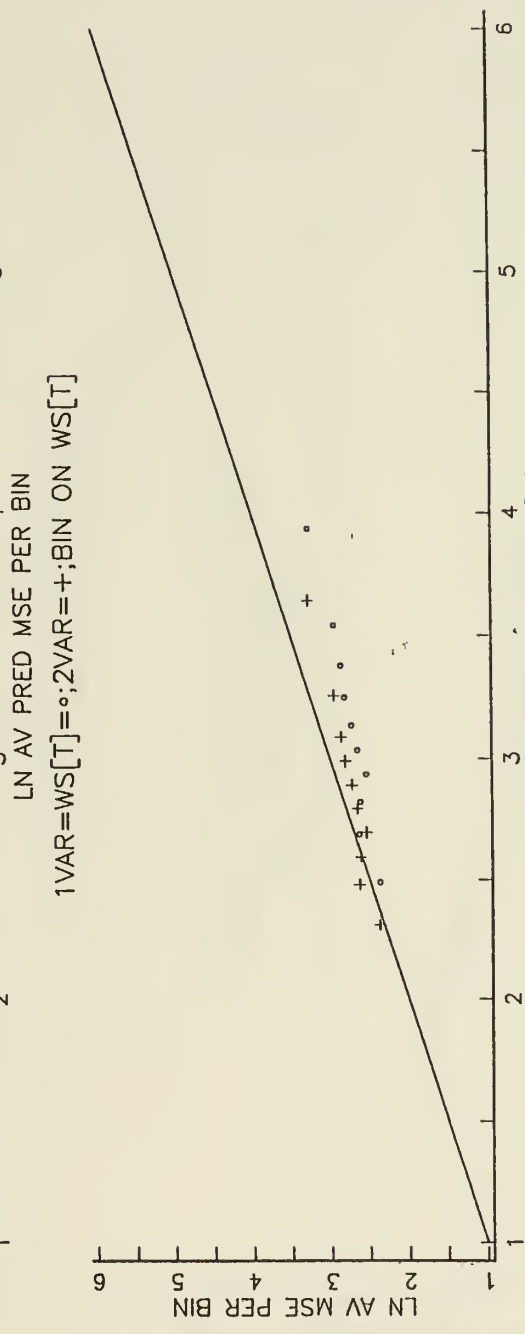
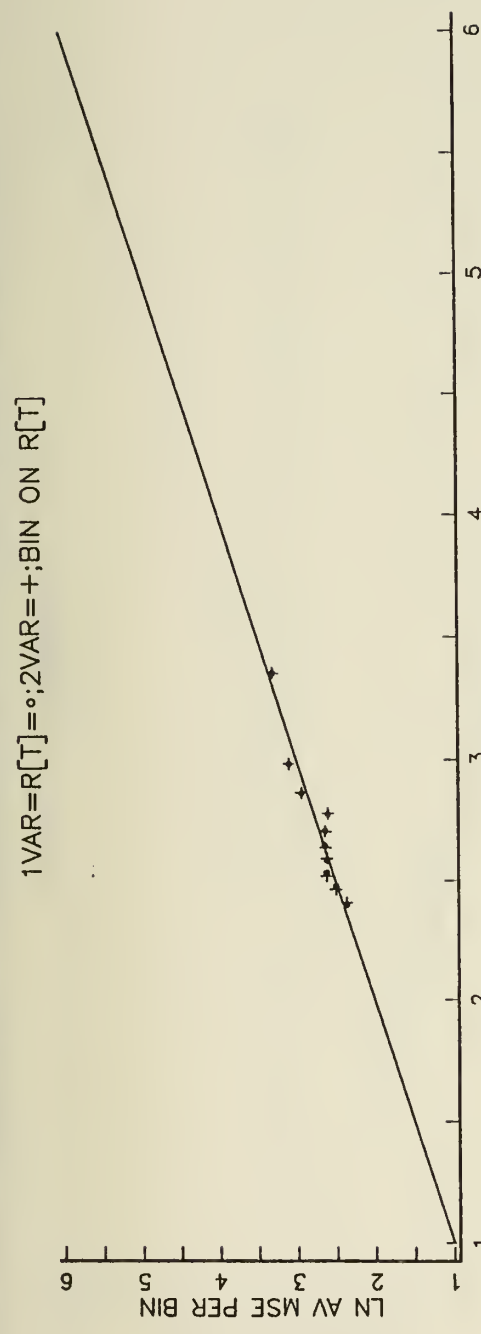
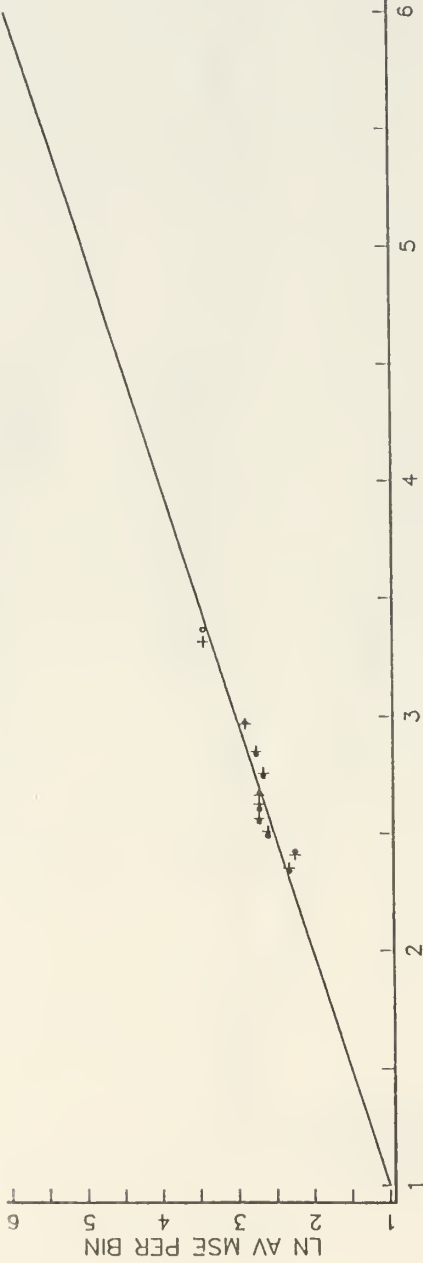


Figure 13

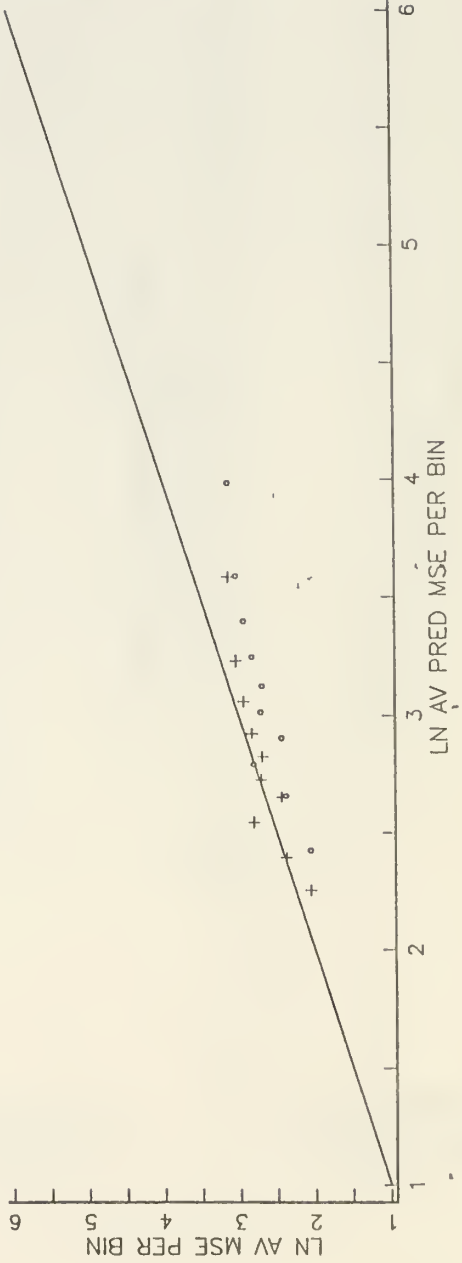
500 MB V WIND; MODEL B ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 14

500 MB V WIND; MODEL B ON DATA A; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



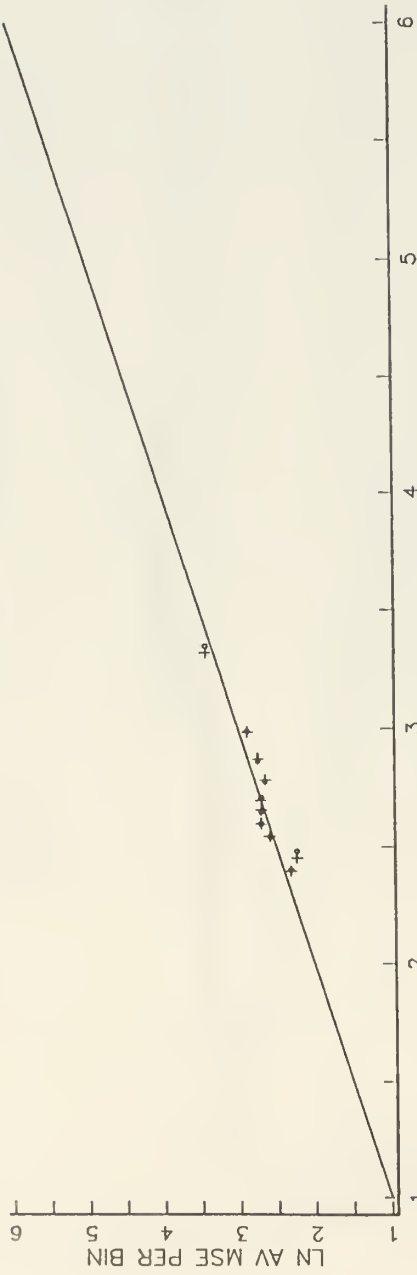
1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



Figure 15

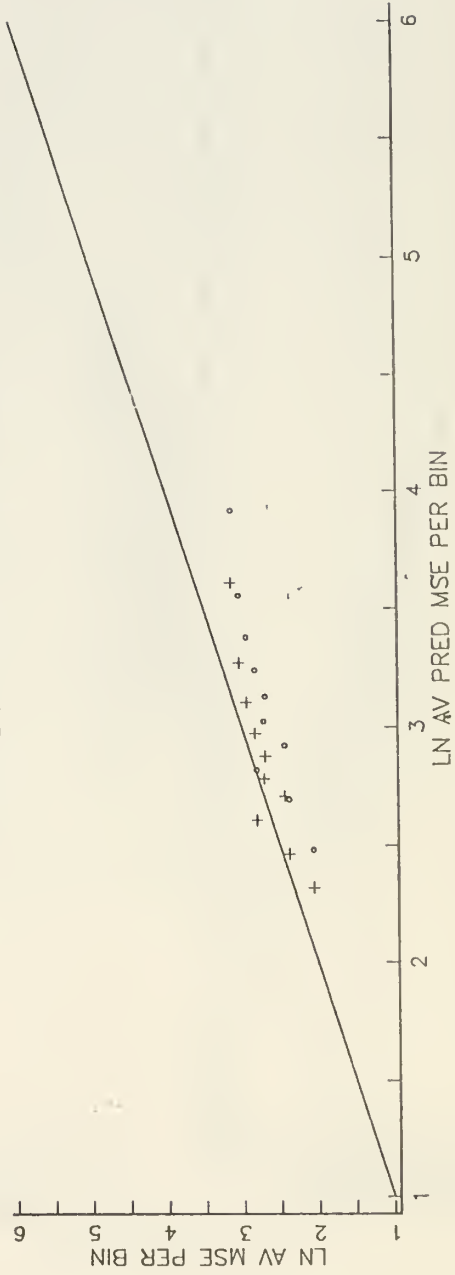
500 MB V WIND; MODEL A ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

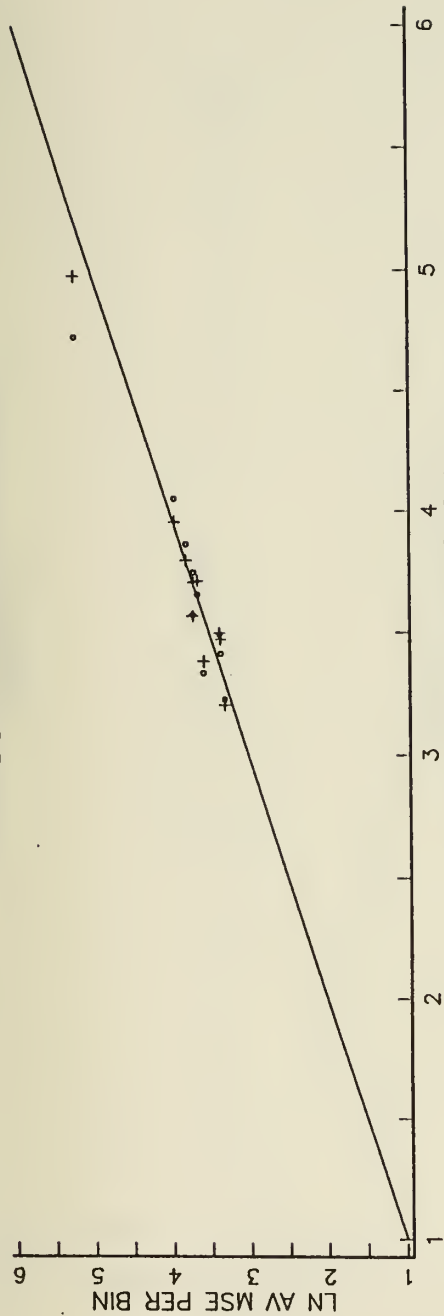


LN AV PRED MSE PER BIN

Figure 16

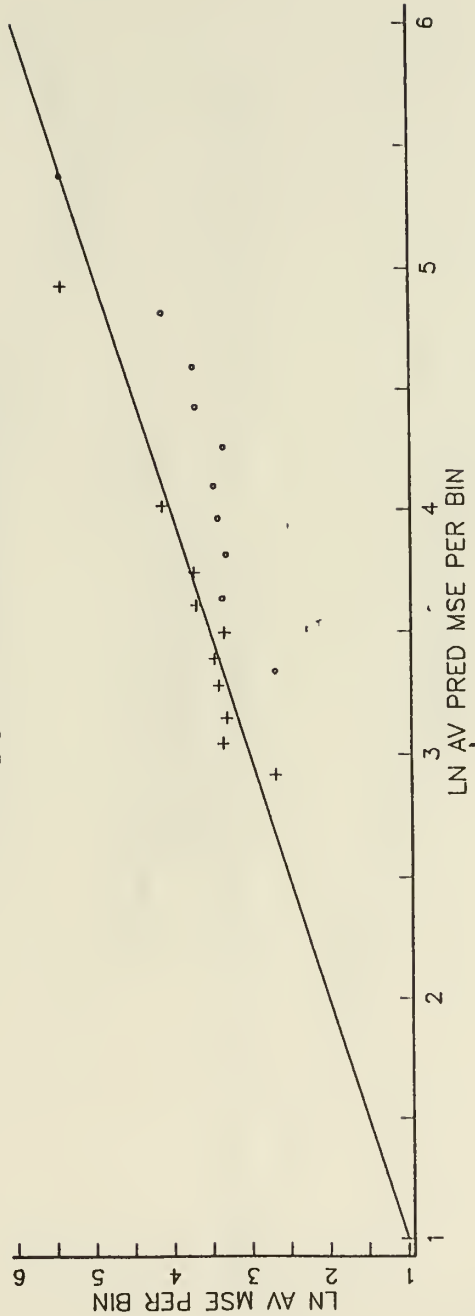
250 MB U WIND; MODEL A ON DATA A; FEB OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 17

250 MB U WIND; MODEL B ON DATA B; FEB OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

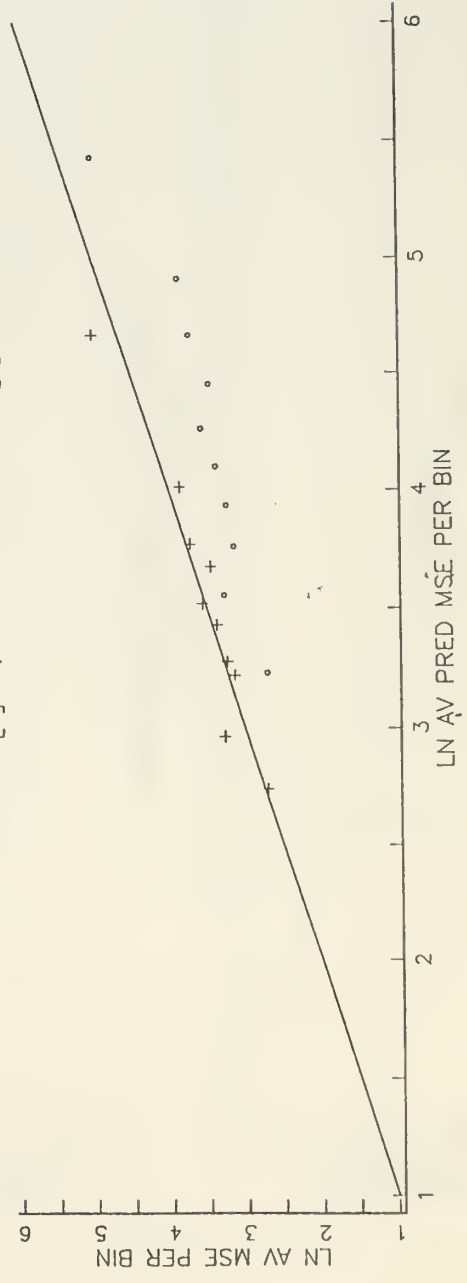


Figure 18



1VAR=R[T]=0;2VAR=+;BIN ON R[T]



1VAR=WS[T]=0;2VAR=+;BIN ON WS[T]

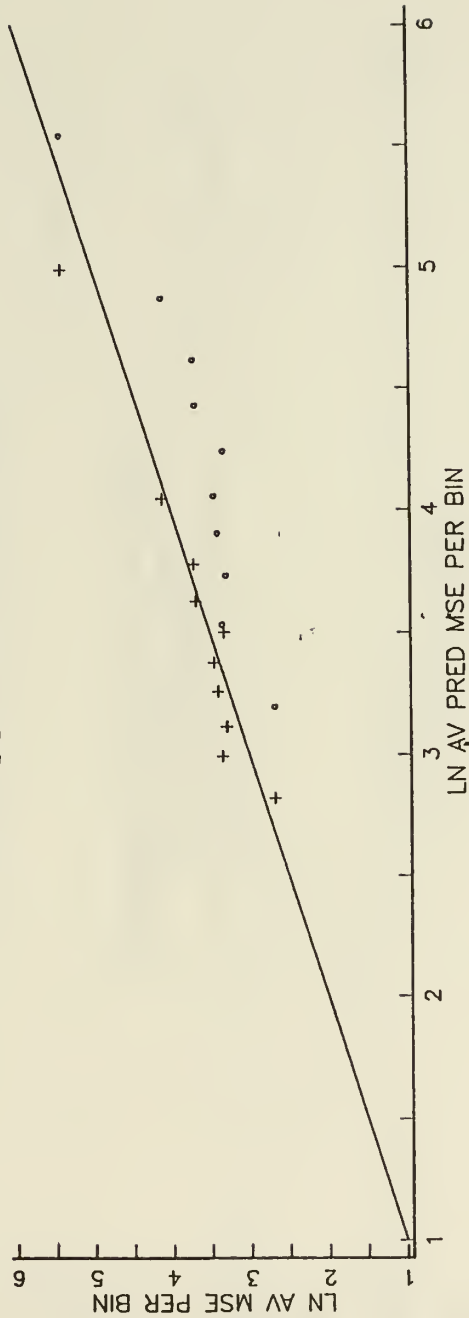
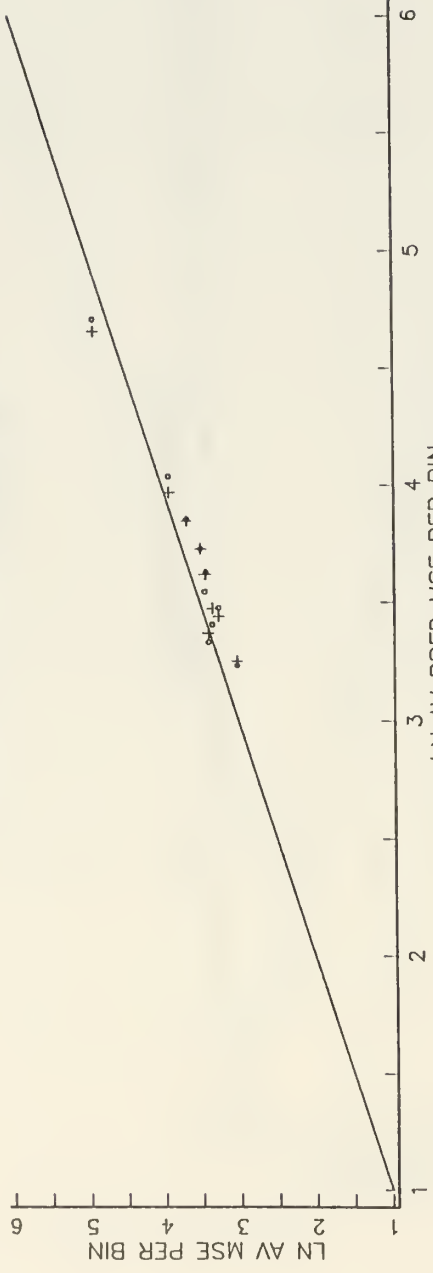


Figure 19

250 MB U WIND; MODEL A ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

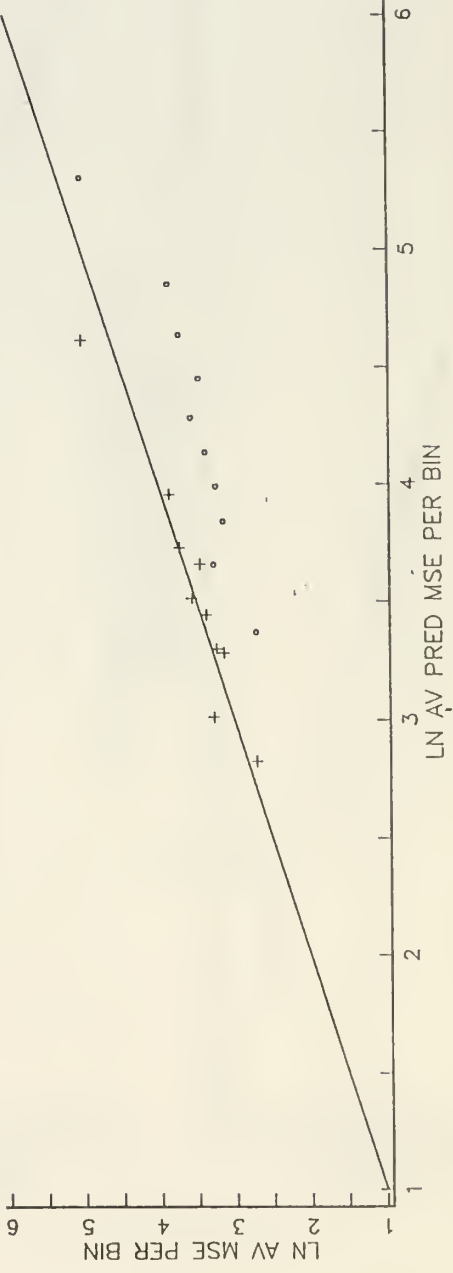
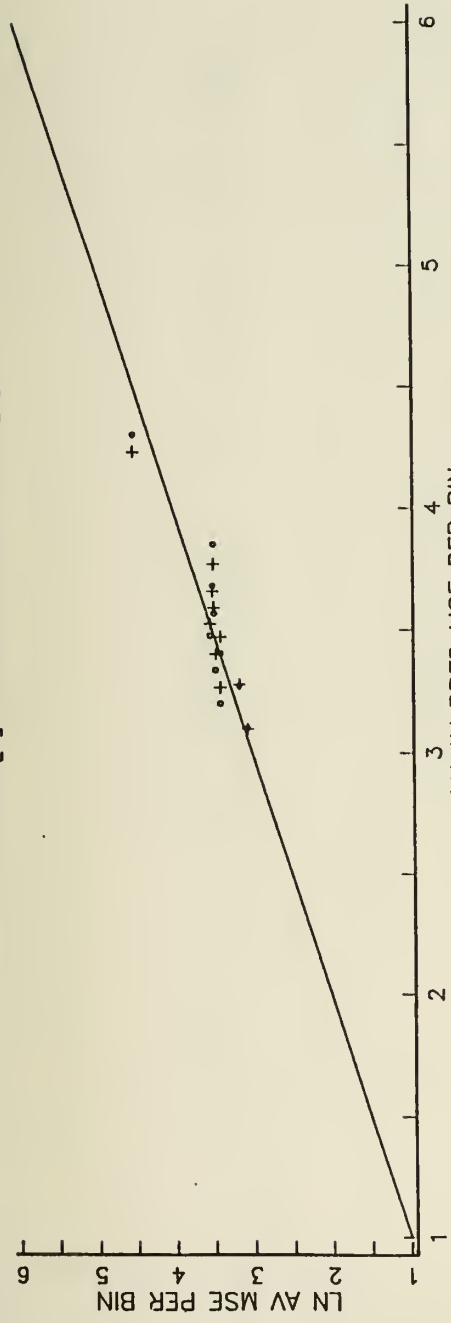


Figure 20

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

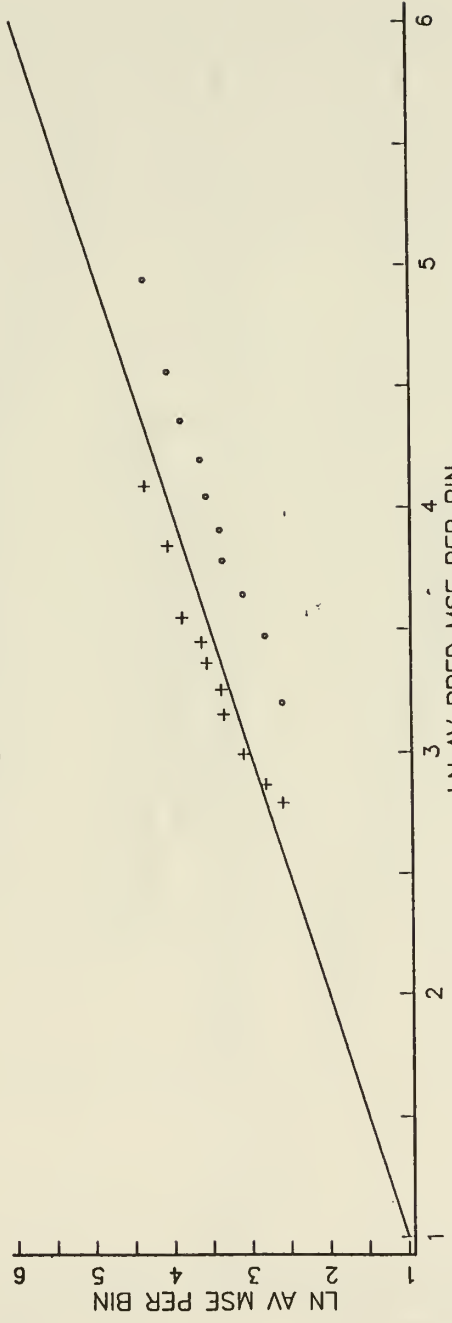


Figure 21

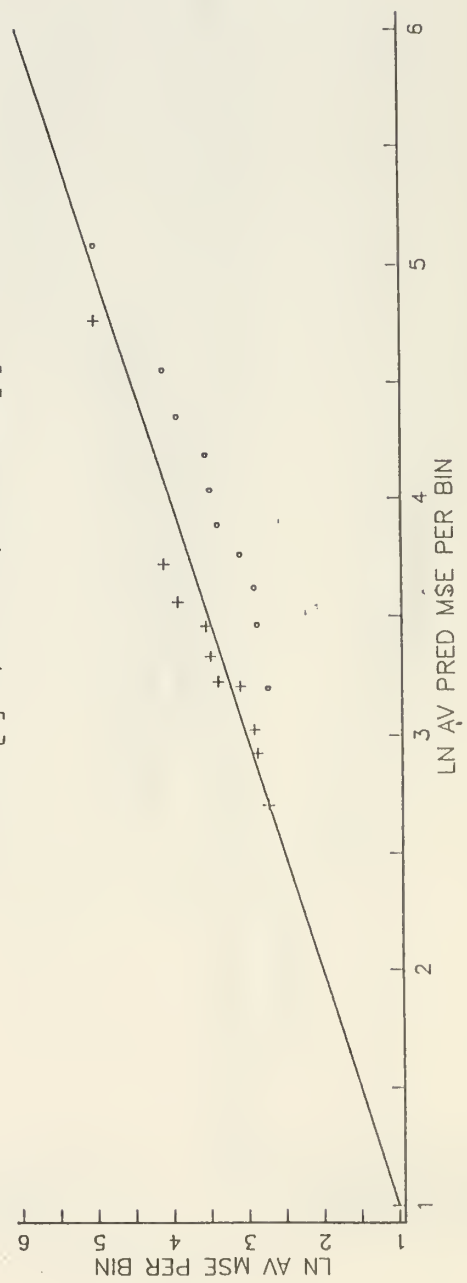
250 MB V WIND; MODEL B ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

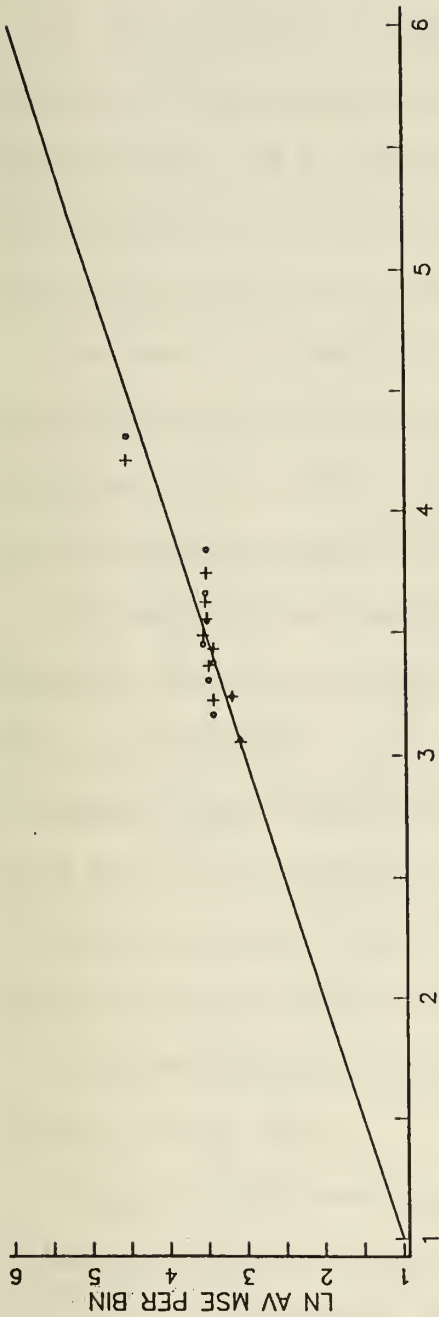
1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 22

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

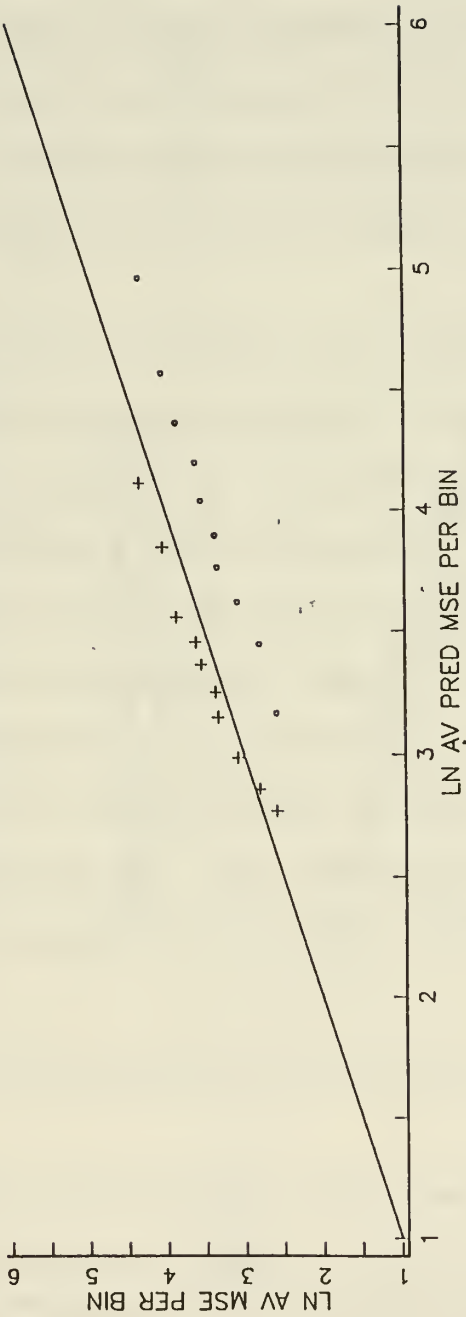


Figure 23

250 MB V WIND; MODEL A ON DATA B; FEB OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 24

## APPENDIX B. A GRAPHICAL ASSESSMENT OF GOODNESS OF FIT AND CROSS-VALIDATION OF MODELS OF FEBRUARY WIND COMPONENT MEAN SQUARE ERROR USING FIRST-GUESS WIND COVARIATES

In this appendix we present figures assessing goodness of fit and cross-validation of the normal models (1)–(3) with first-guess wind covariates fit to February data. As in subsection (3.2) the data is randomly divided into two sets called DA and DB without regard to the values of the data; these sets are the same as those in that section.

The maximum likelihood parameter estimates for each model (1)–(3) are obtained for each set DA and DB and appear in Table 3. The estimated variances  $\sigma_1^2(1,t)$ ,  $\sigma_1^2(2,t)$ ,  $\sigma_2^2(t)$  are computed for the parameters estimated from DA and DB using (1)–(3) for each data point in DA and DB.

To assess models (1) and (3) the data  $(y(t), r(t), s(t))$  are binned into 10 bins based on ordering the values of  $r(t)$  from smallest to largest. The data in the first bin correspond to the smaller values of  $r(t)$ ; the data in the 10<sup>th</sup> bin correspond to the larger values of  $r(t)$ . Each bin contains about  $\frac{1}{10}$  of the data with the 10<sup>th</sup> bin containing a few more data. The averages of the estimated variances for models (1) and (3) are computed for each bin. The average  $y(t)^2$  is also computed for each bin.

To assess models (2) and (3) the same procedure is used but the binning is based on values of  $s(t)$ .

Figures 1B-24B present graphs of the  $\ln[\text{average } y(t)^2]$  in each bin versus  $\ln[\text{average estimated variance}]$  in each bin for models (1) and (3) and models (2) and (3). Figures 1B, 5B, 9B, 13B, 17B, 21B (respectively 2B, 6B, 10B, 14B, 18B

22B) show the logarithm of the average of the  $y(t)^2$  values of DA (respectively DB) versus the logarithm of the average of the estimated variances for each bin using the estimated parameters from DA (respectively DB). If a model were perfect, a point should be close to the 45° line shown. These figures assess goodness of fit.

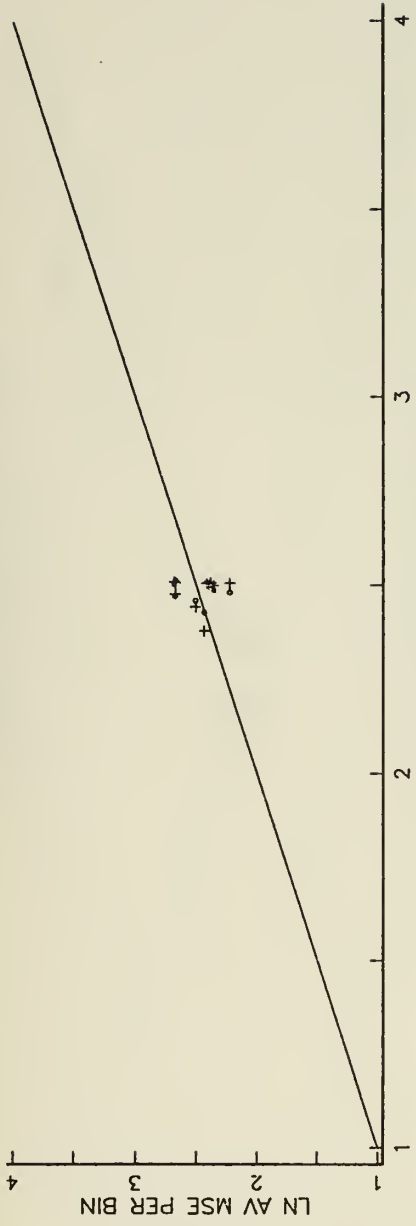
Figures 3B, 7B, 11B, 15B, 19B, 23B (respectively 4B, 8B, 12B, 16B, 20B, 24B) present graphs of  $\ln$  average  $y(t)^2$  of DA (respectively DB) versus  $\ln$  average estimated variances using parameters estimated using data DB (respectively DA). Once again if the model were perfect, the points would be close to the 45° line.

As suggested by the values of the  $\ln$ -likelihood  $\tilde{\ell}$  in Tables 2 and 4, the figures for models using first guess covariates indicate weaker goodness of fit and weaker cross-validation than Figures 1-24 for models with observed wind speed covariates. Both goodness-of-fit and cross-validation appear to improve somewhat for higher pressure height levels; Figures 17B-24B. This suggests that models using first guess covariates have greater predictive and descriptive value at 250mb height levels. However, they appear to be not as good as models using observed wind speed as covariates.



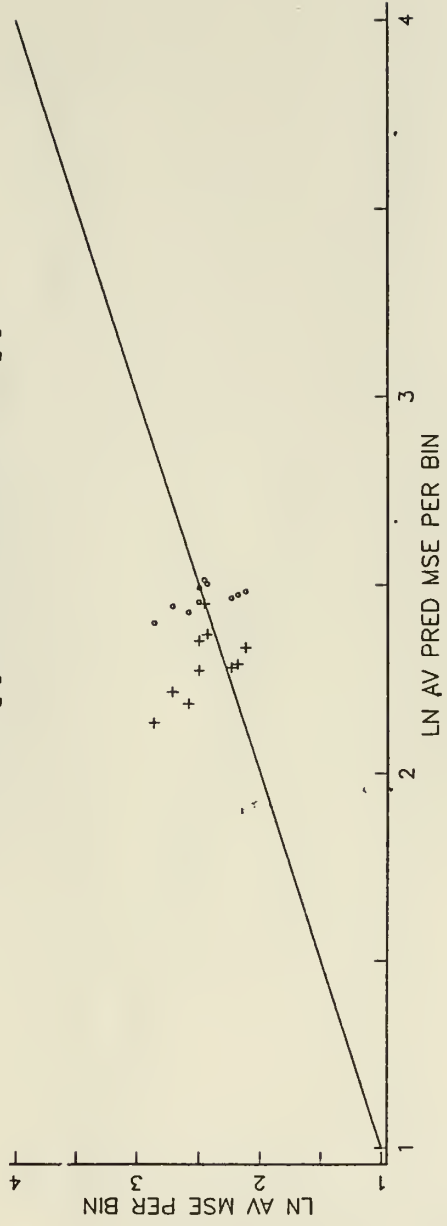
850 MB U WIND; MODEL A ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=0; 2 VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=0; 2 VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 1B

850 MB U WIND;MODEL B ON DATA B;FIRST-GUESS WINDS

1 VAR=R[T]=0;2 VAR=+;BIN ON R[T]



1 VAR=WS[T]=0;2 VAR=+;BIN ON WS[T]

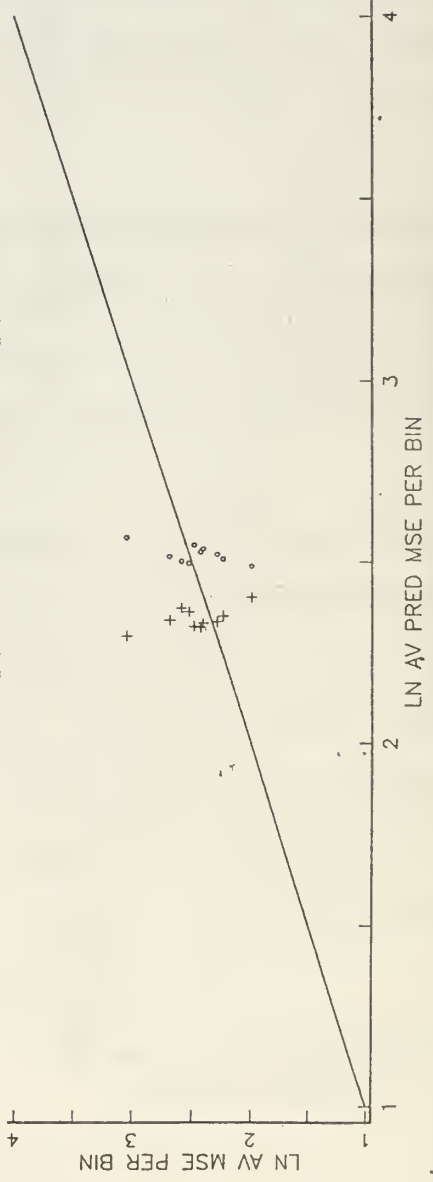
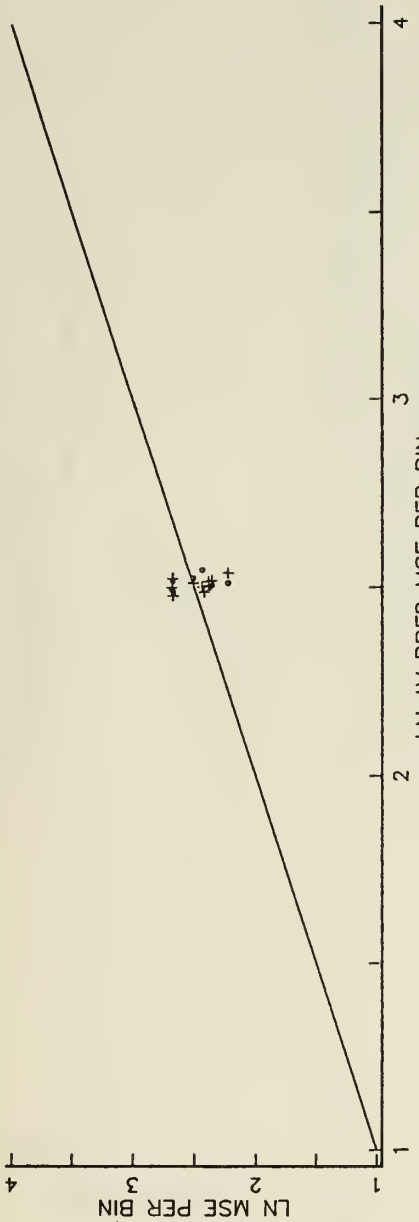


Figure 2B

850 MB U WIND; MODEL B ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=o; 2 VAR=+; BIN ON R[T]



1 VAR=WS[T]=o; 2 VAR=+; BIN ON WS[T]

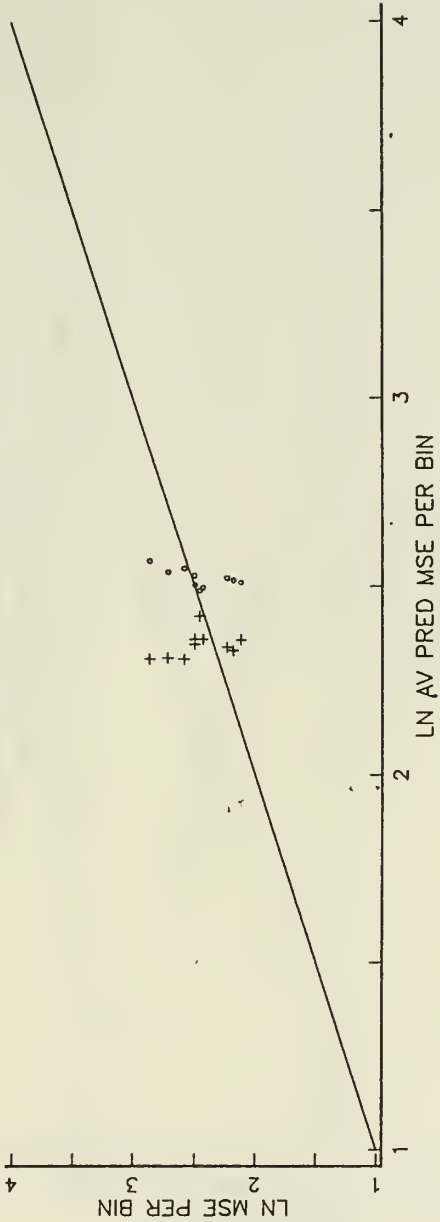
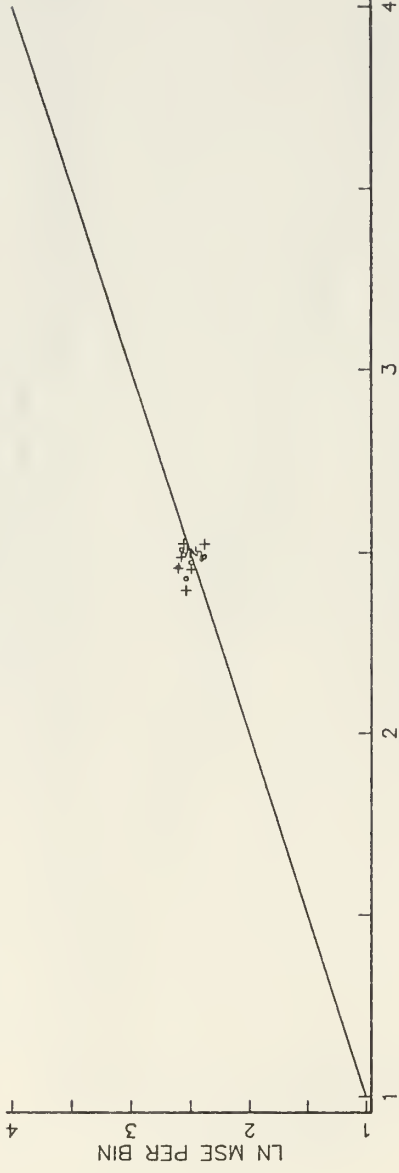


Figure 3B

850. MB U WIND; MODEL A ON DATA B; FIRST--GUESS WINDS

1 VAR=R[T]=o;2 VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=o;2 VAR=+;BIN ON WS[T]

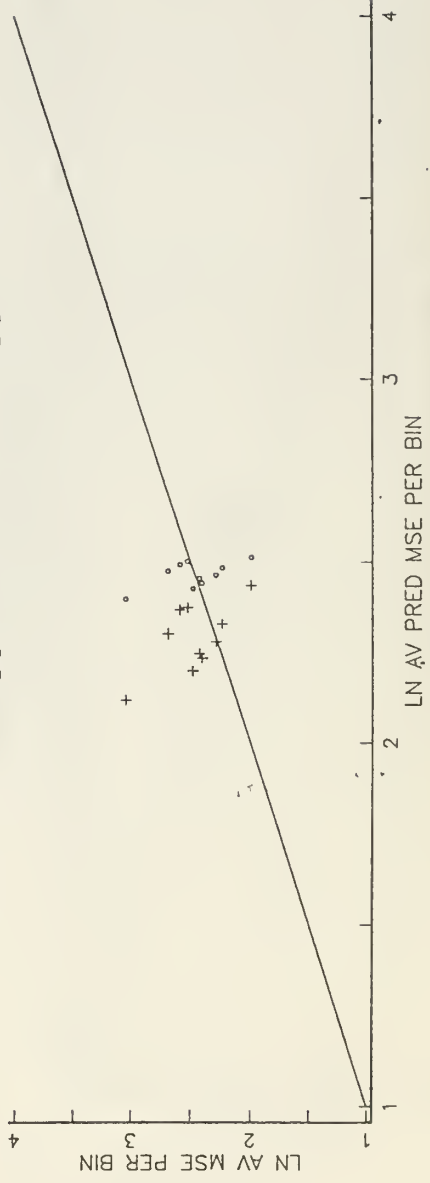
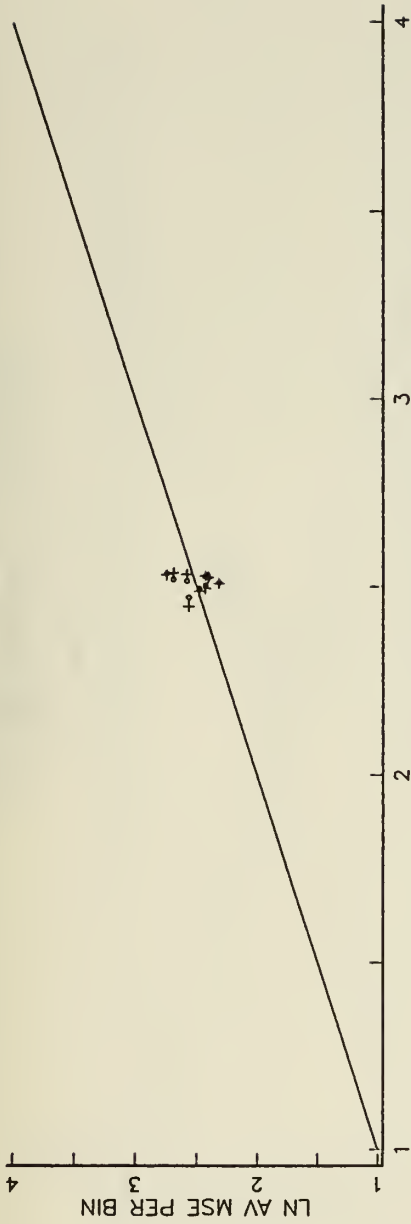


Figure 4B

850 MB V WIND; MODEL A ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=0; 2 VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN  
1 VAR=WS[T]=0; 2 VAR=+; BIN ON WS[T]

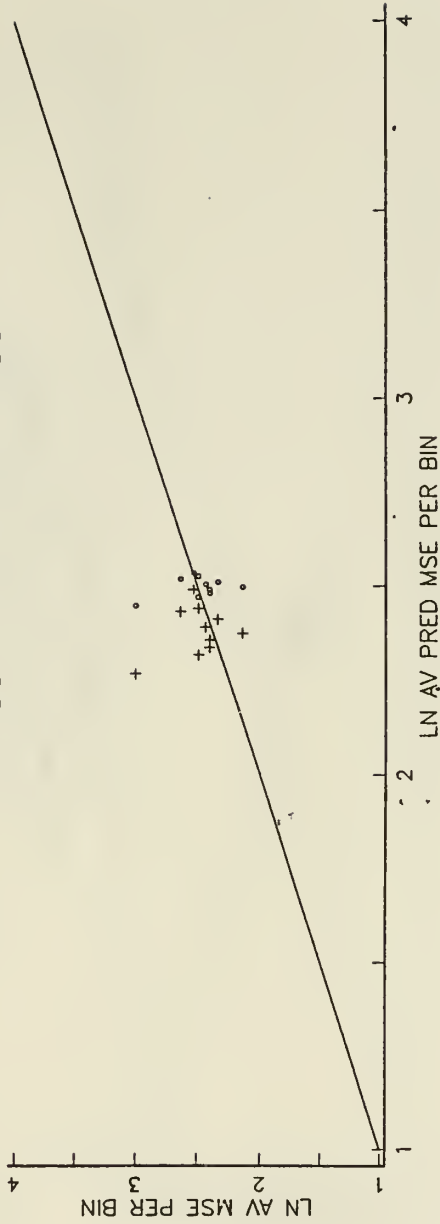
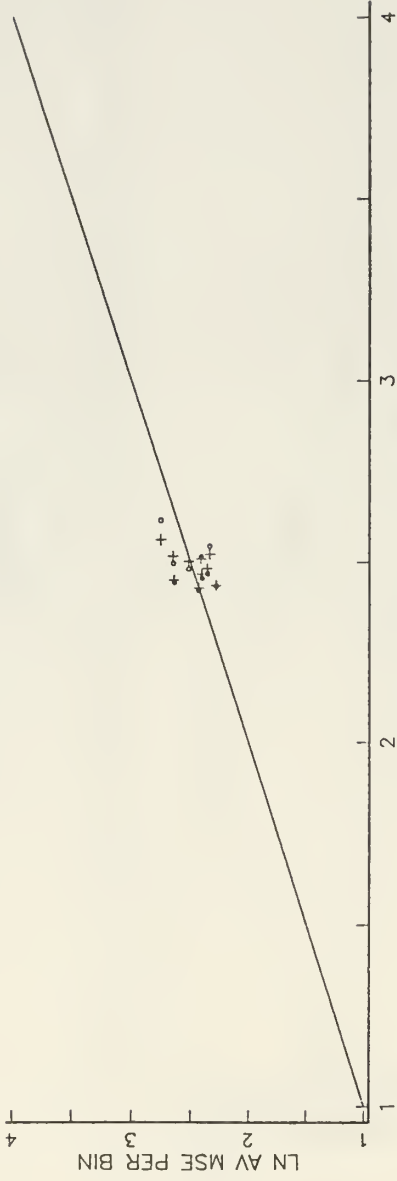


Figure 5B

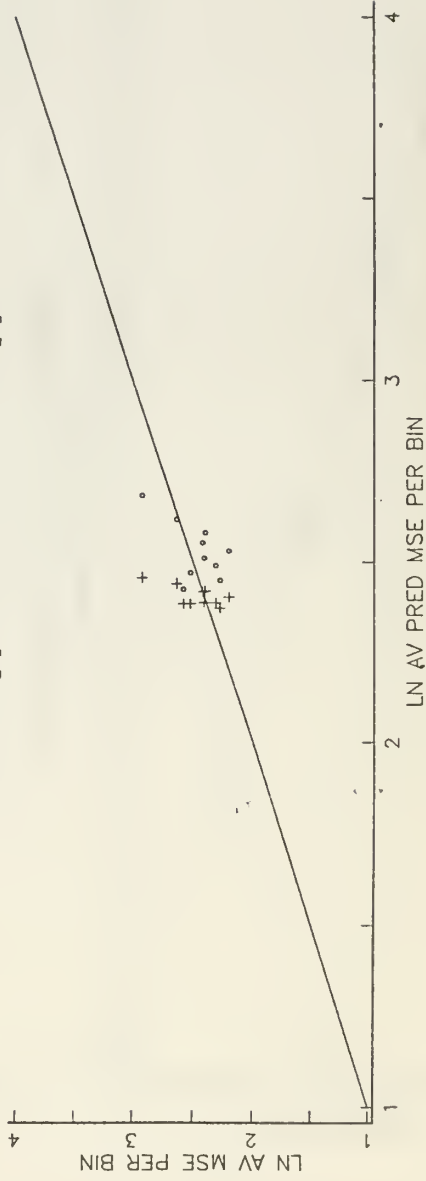
850 MB V WIND; MODEL B ON DATA B; FIRST-GUESS WINDS

1 VAR=R[T]=o;2 VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=o;2 VAR=+;BIN ON WS[T]

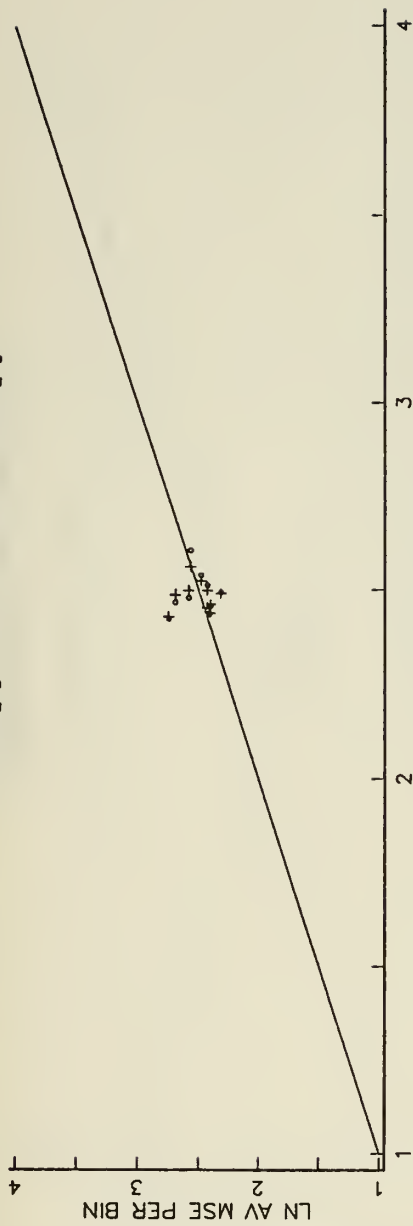


LN AV PRED MSE PER BIN

Figure 6B

850 MB V WIND; MODEL B ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=0; 2 VAR=+; BIN ON R[T]



1 VAR=WS[T]=0; 2 VAR=+; BIN ON WS[T]

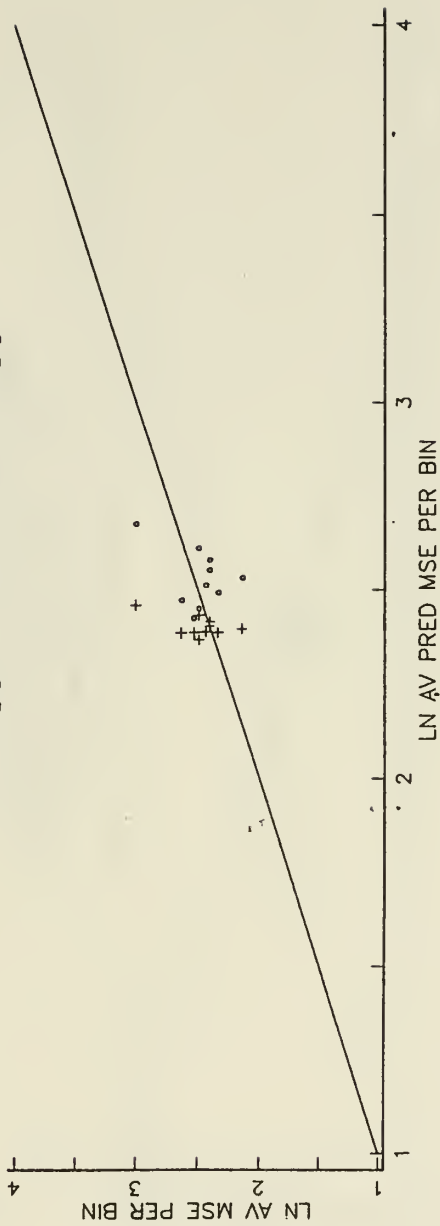


Figure 7B

850 MB V WIND; MODEL A ON DATA B; FIRST-GUESS WINDS

1 VAR=R[T]=0; 2 VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=0; 2 VAR=+; BIN ON WS[T]

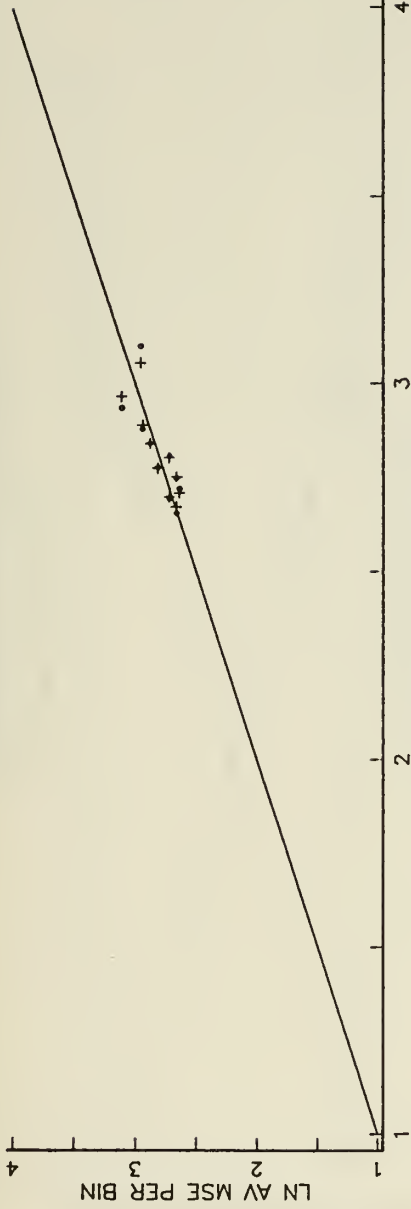


Figure 8B



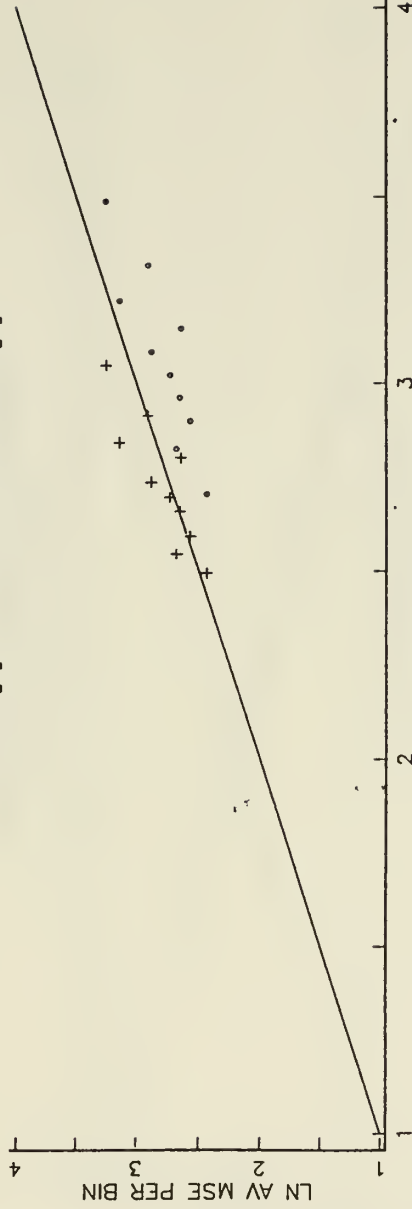
500 MB U WIND; MODEL A ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=0;2 VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=0;2 VAR=+;BIN ON WS[T]

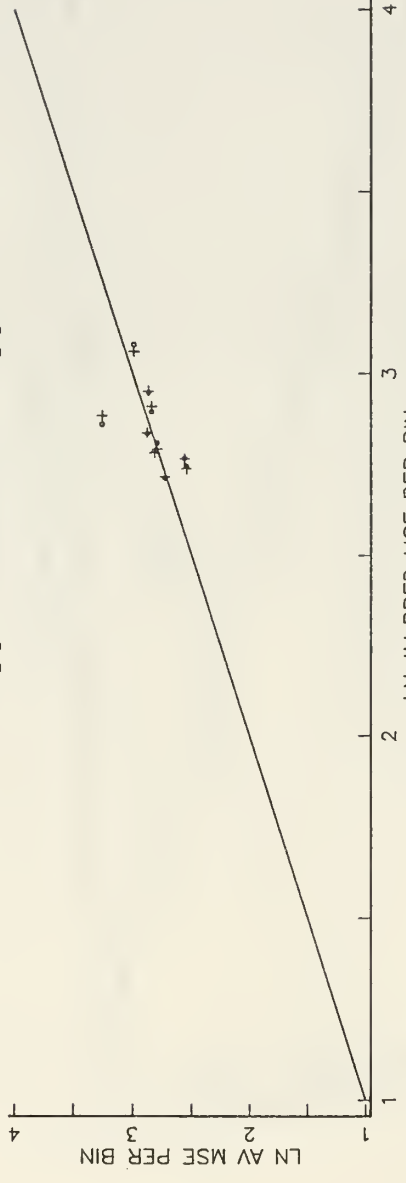


LN AV PRED MSE PER BIN

Figure 9B

500 MB U WIND; MODEL B ON DATA B; FIRST-GUESS WINDS

1 VAR=R[T]=0;2 VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=0;2 VAR=+;BIN ON WS[T]

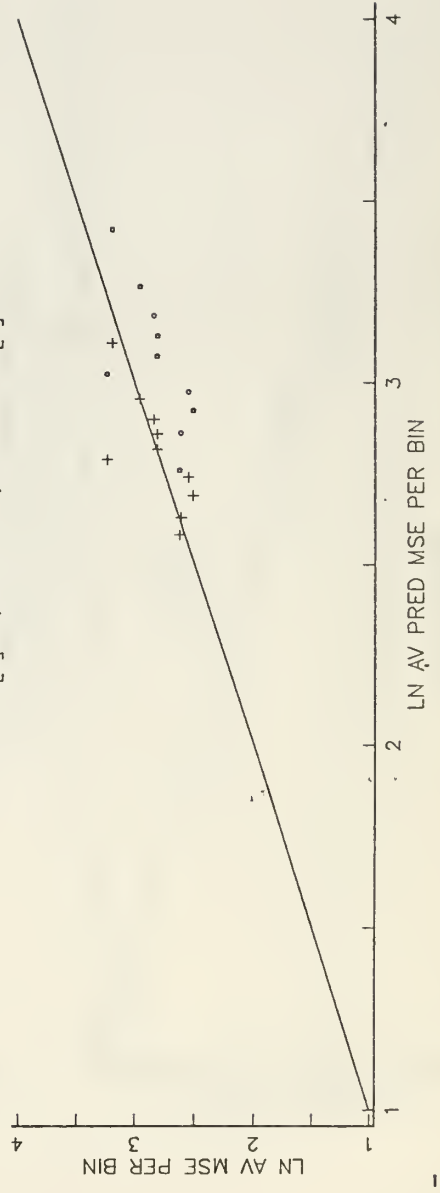
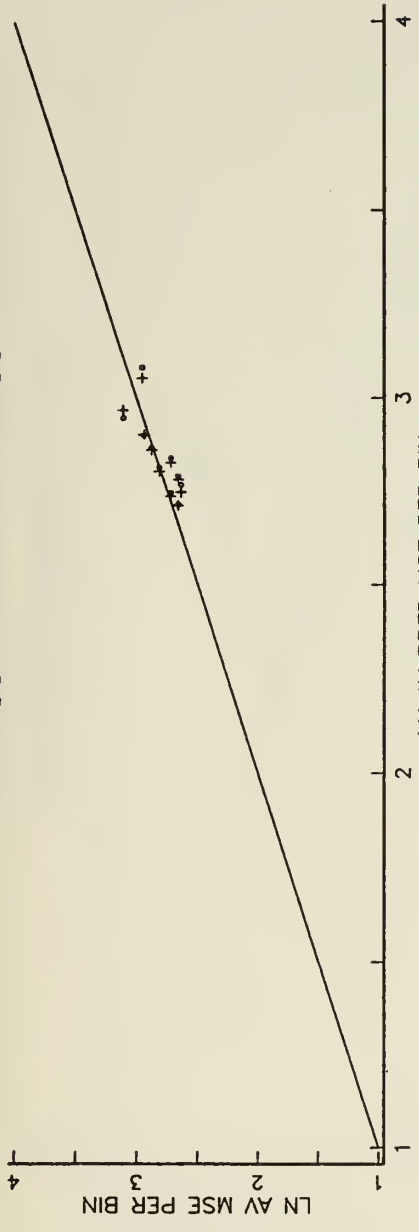


Figure 10B

500 MB U WIND; MODEL B ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=0; 2 VAR=+; BIN ON R[T]



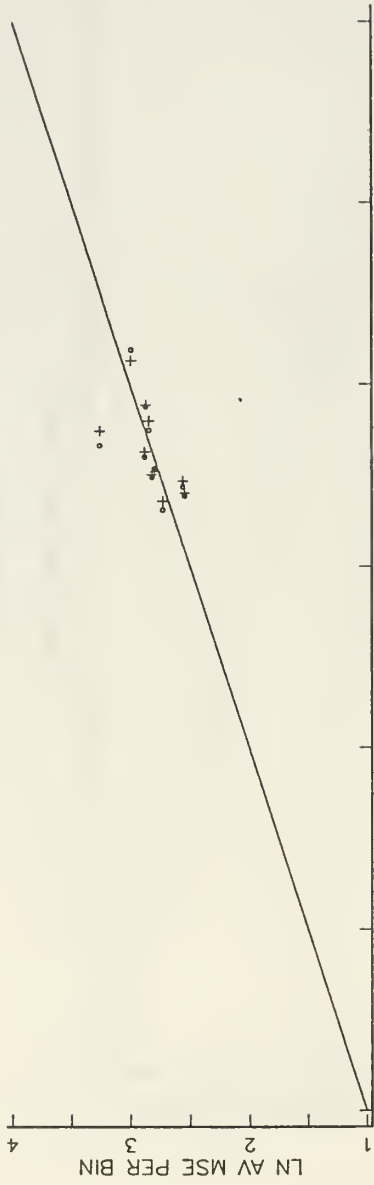
1 VAR=WS[T]=0; 2 VAR=+; BIN ON WS[T]



Figure 11B

500 MB U WIND; MODEL A ON DATA B; FIRST-GUESS WINDS

1 VAR=R[T]=o; 2 VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=o; 2 VAR=+; BIN ON WS[T]

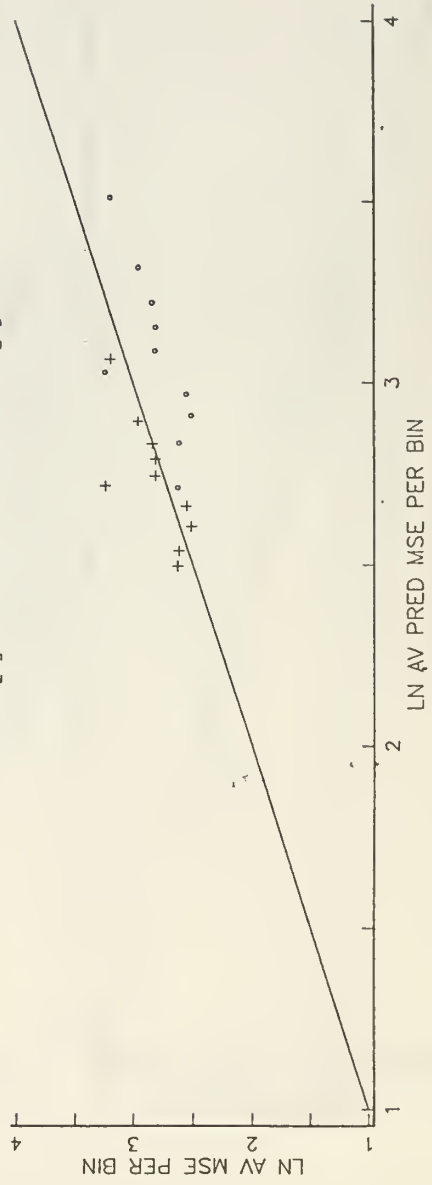
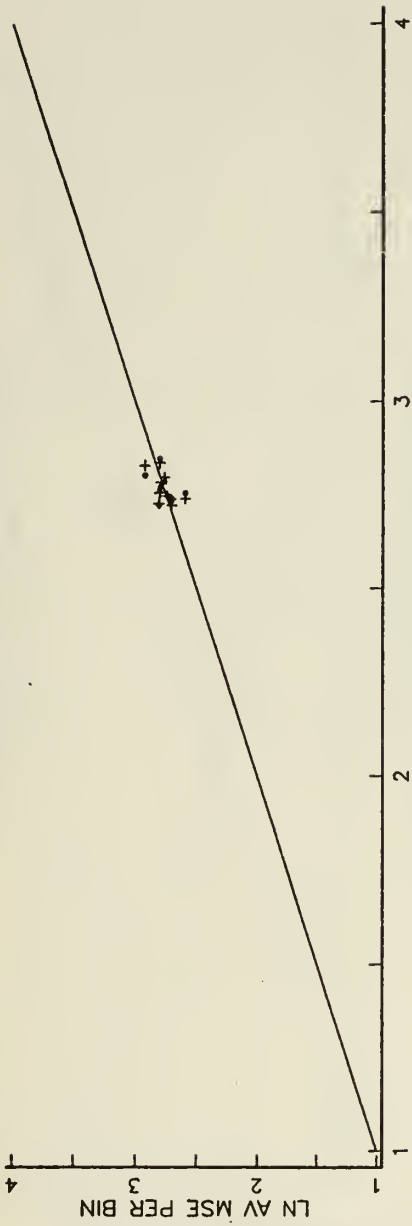


Figure 12B

500 MB V WIND; MODEL A; FIRST-GUESS WINDS

1 VAR=R[T]=0:2 VAR=+;BIN ON R[T]



1 VAR=WS[T]=0:2 VAR=+;BIN ON WS[T]

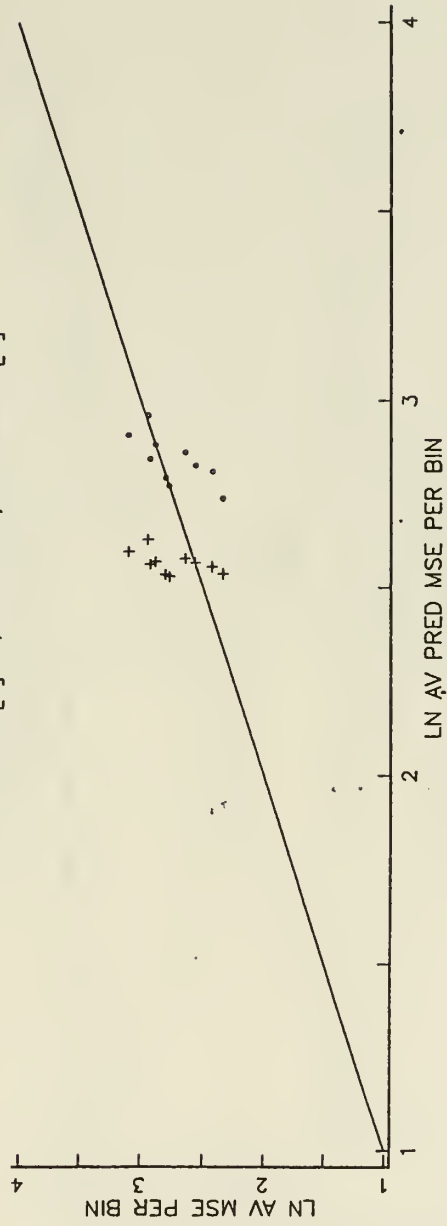


Figure 13B

500 MB V WIND; MODEL B ON DATA B; FIRST-GUESS WINDS

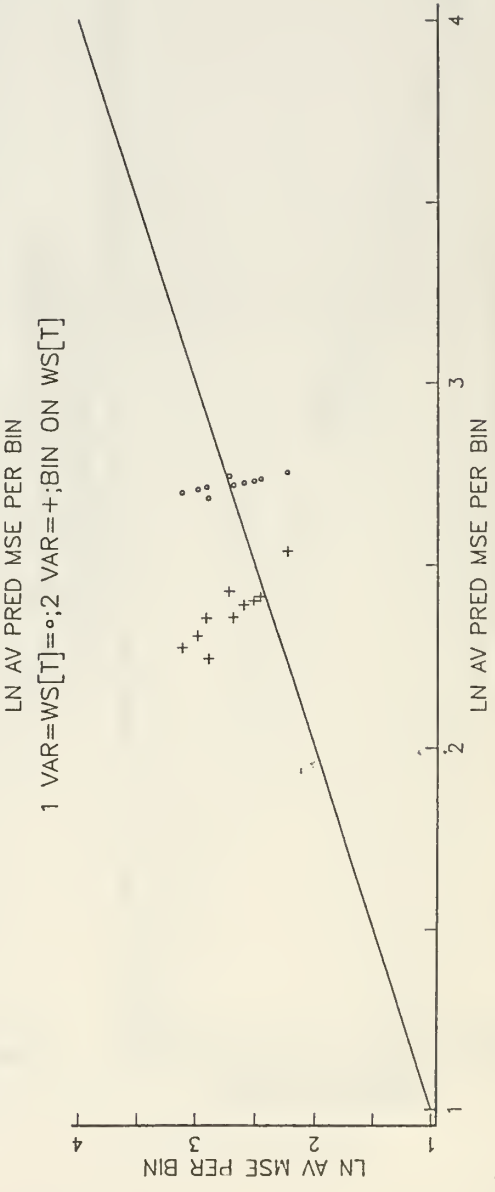
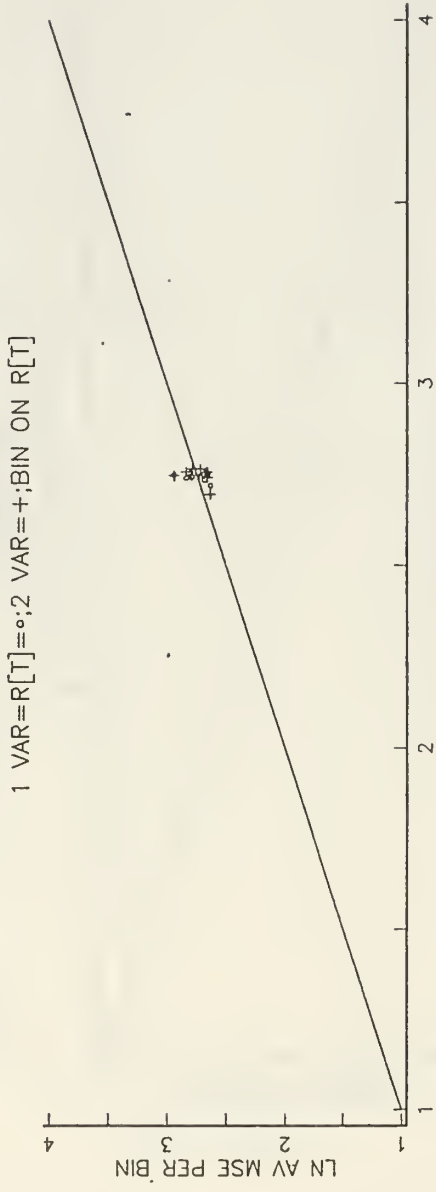
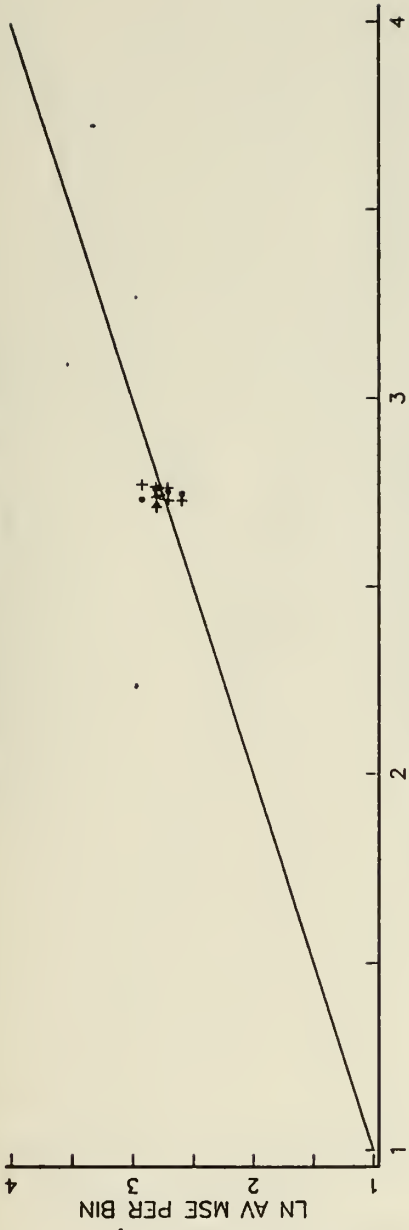


Figure 14B

500 .MB V WIND;MODEL B ON DATA A;FIRST--GUESS WINDS

1 VAR=R[T]=o;2 VAR=+;BIN ON R[T]



1 VAR=WS[T]=o;2 VAR=+;BIN ON WS[T]

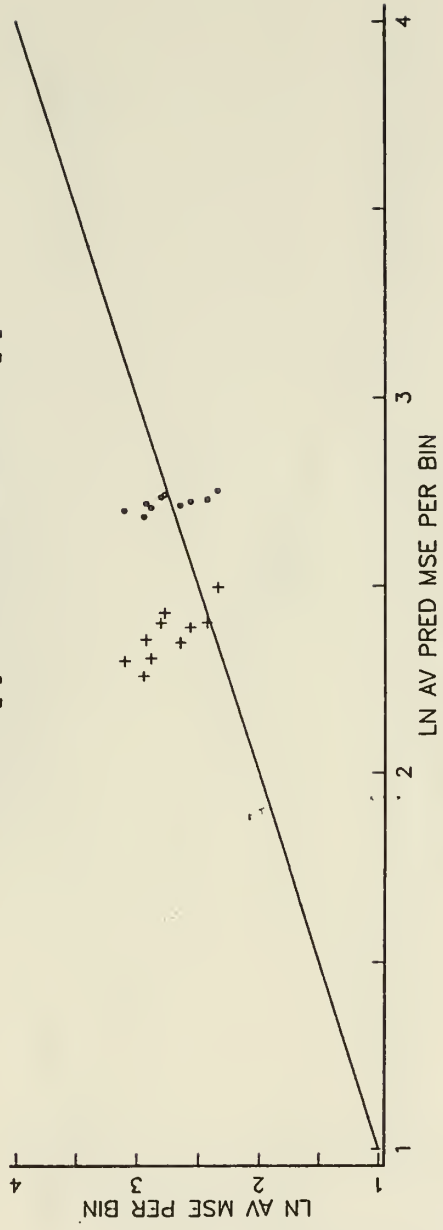


Figure 15B

500 MB V WIND; MODEL A ON DATA B; FIRST-GUESS WINDS

1 VAR=R[T]=o; 2 VAR=+; BIN ON R[T]



1 VAR=WS[T]=o; 2 VAR=+; BIN ON WS[T]

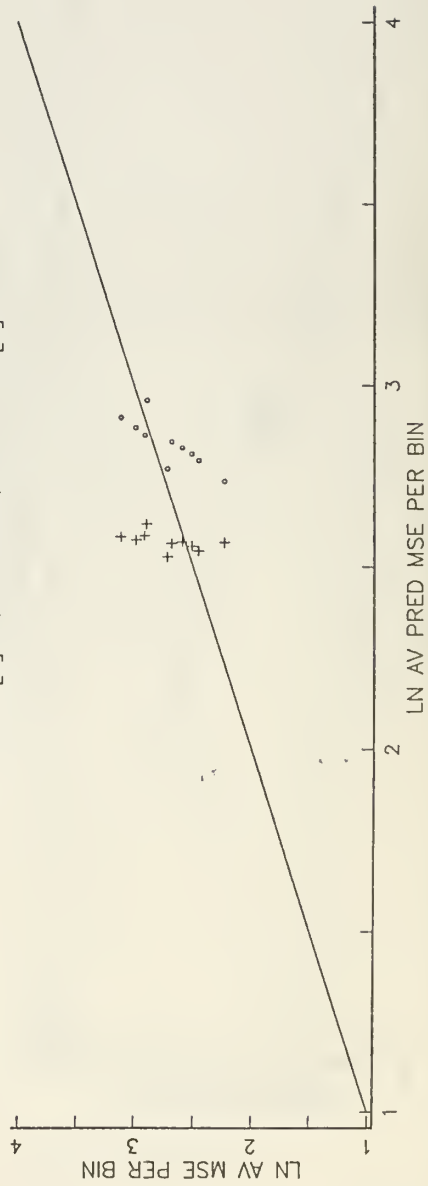


Figure 16B



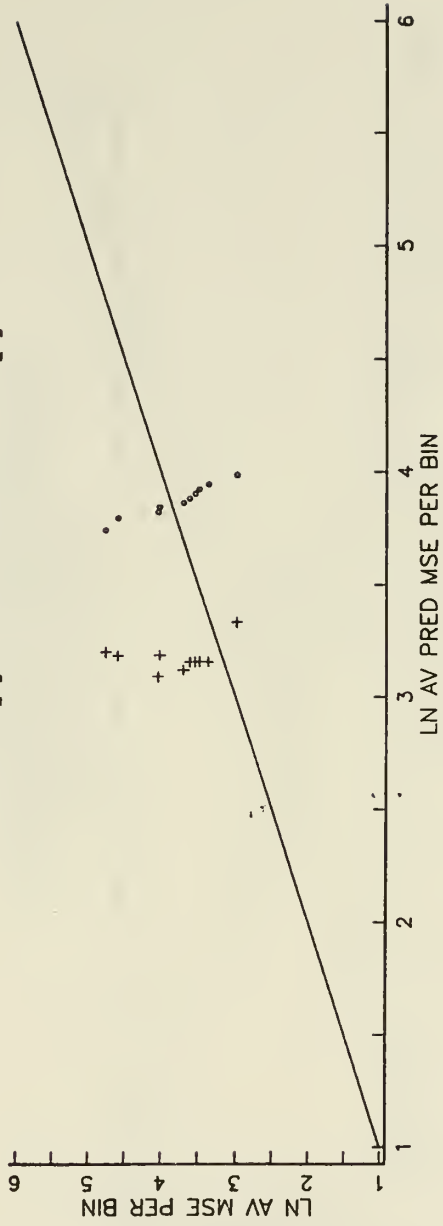
250 MB U WIND; MODEL A ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=0;2 VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=0;2 VAR=+;BIN ON WS[T]

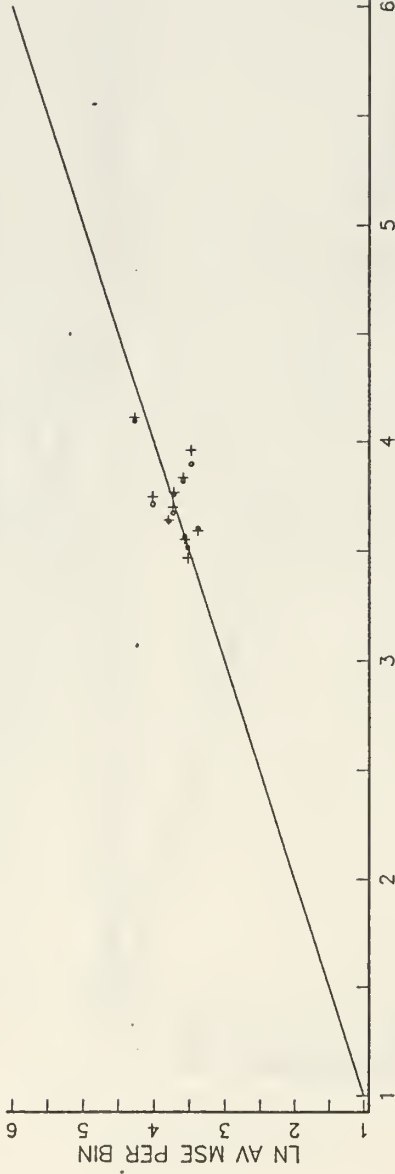


LN AV PRED MSE PER BIN

Figure 17B

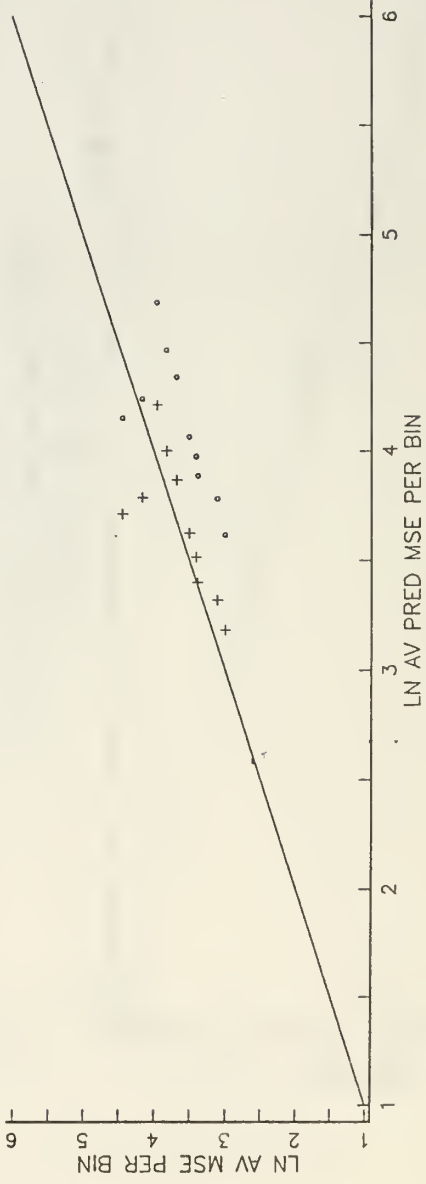
250 MB U WIND; MODEL B ON DATA B; FIRST-GUESS WINDS

1 VAR=R[T]=0;2 VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=0;2 VAR=+;BIN ON WS[T]

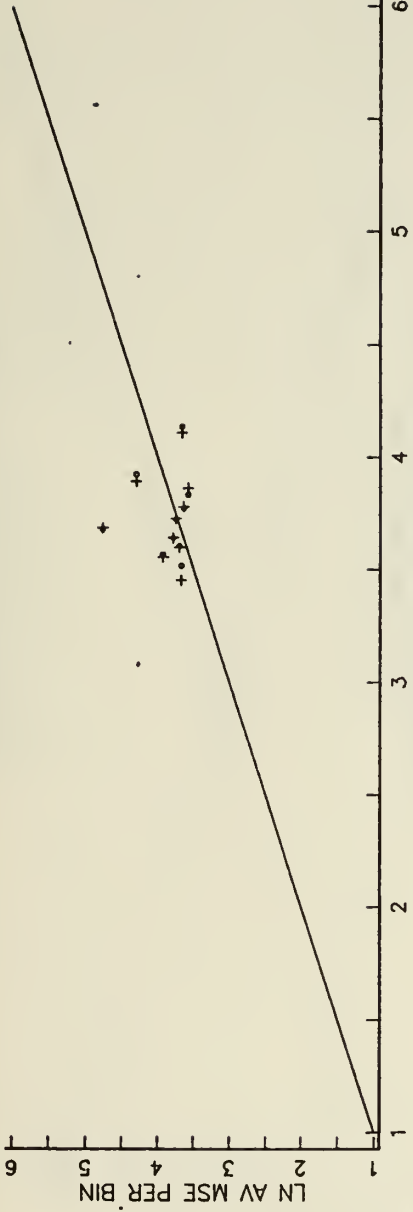


LN AV PRED MSE PER BIN

Figure 18B

250 MB U WIND; MODEL B ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=0;2 VAR=+;BIN ON R[T]



1 VAR=WS[T]=0;2 VAR=+;BIN ON WS[T]

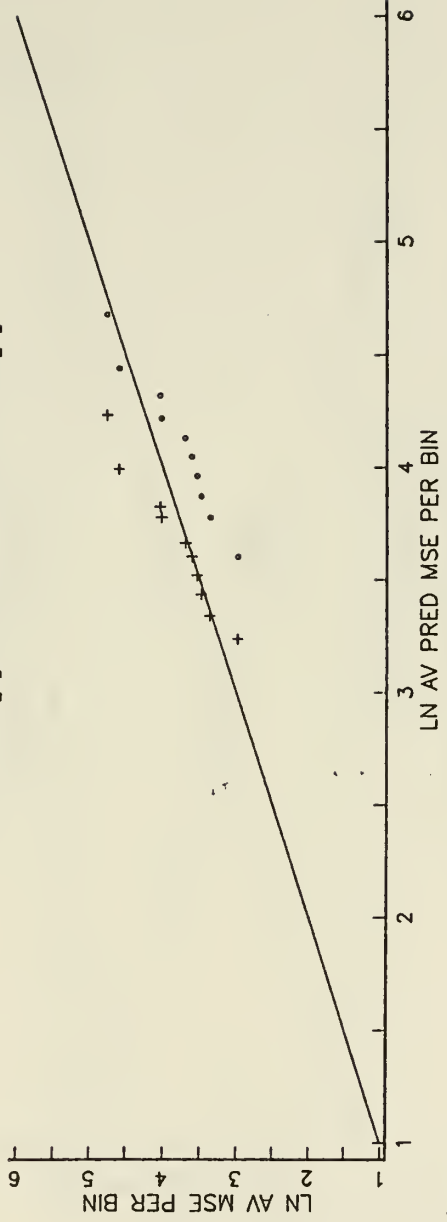
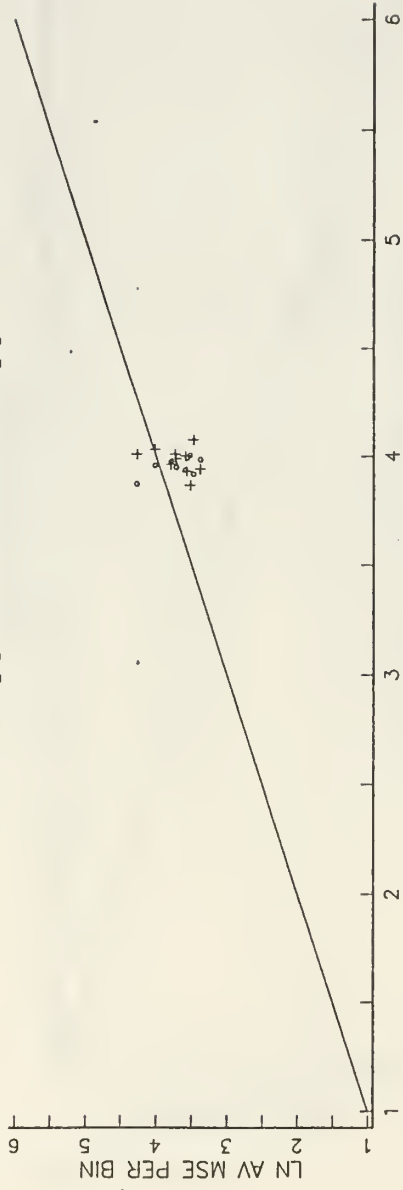


Figure 19B

250 MB U WIND; MODEL A ON DATA B; FIRST-GUESS WINDS

1 VAR=R[T]=0;2 VAR=+;BIN ON R[T]



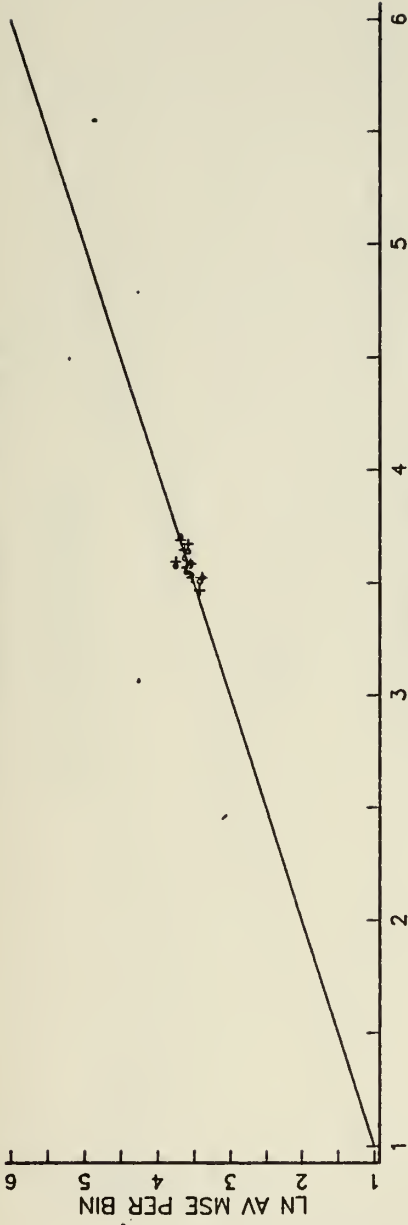
1 VAR=WS[T]=0;2 VAR=+;BIN ON WS[T]



Figure 20B

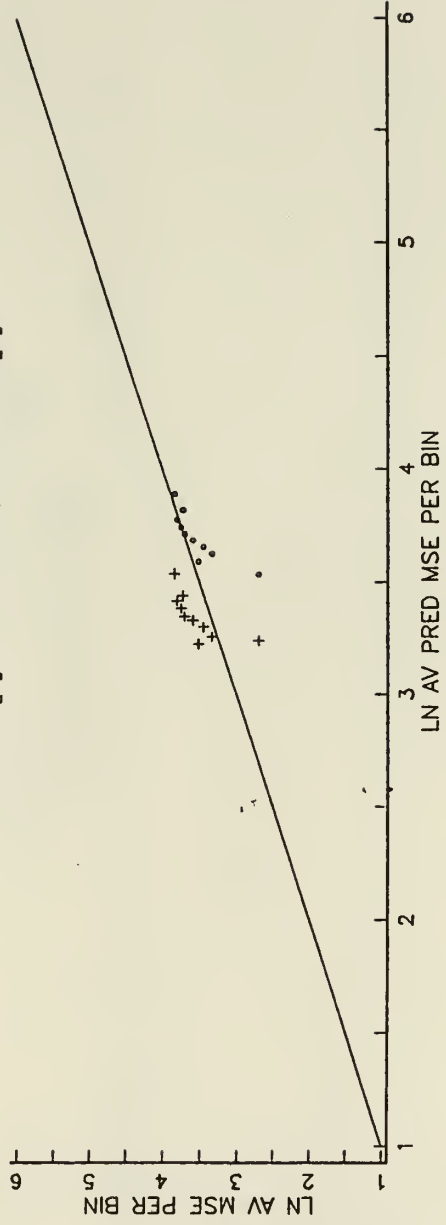
250 MB V WIND; MODEL A ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=0;2 VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=0;2 VAR=+;BIN ON WS[T]

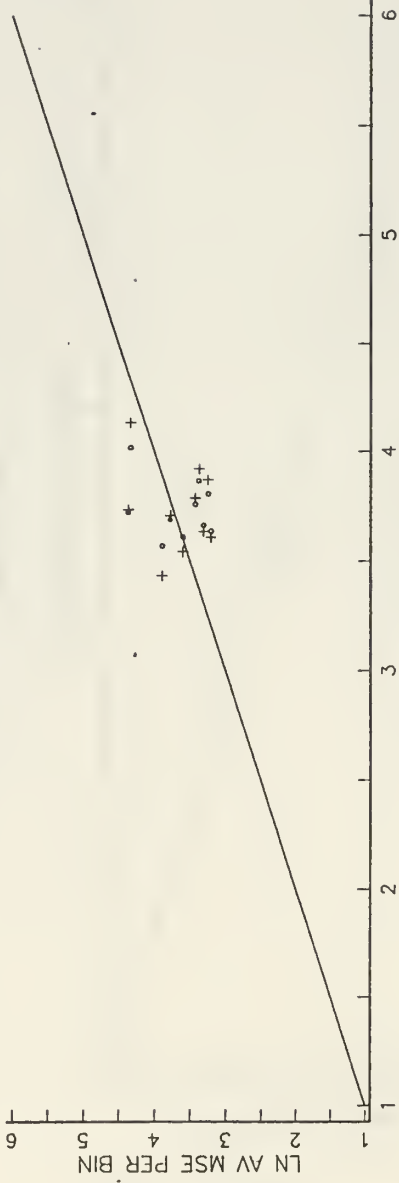


LN AV PRED MSE PER BIN

Figure 21B

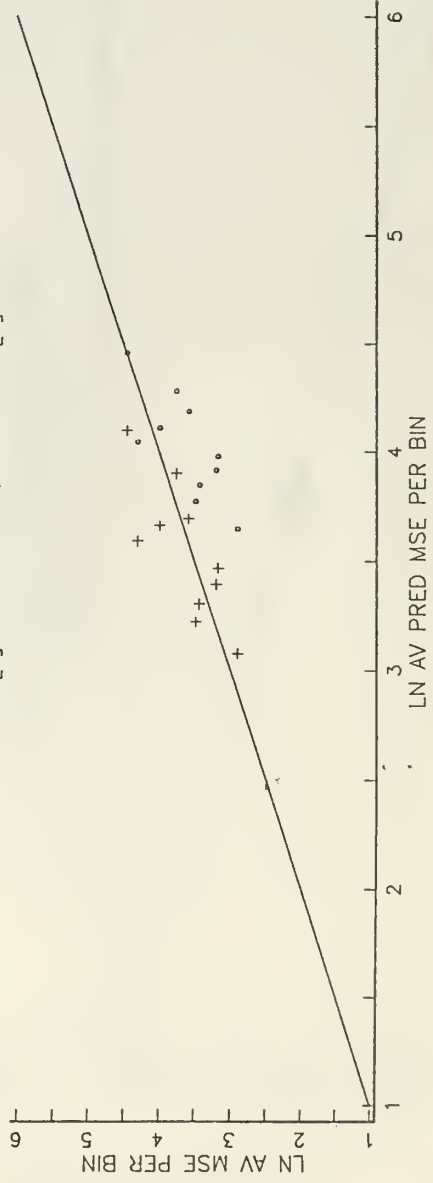
250 MB V WIND; MODEL B ON DATA B; FIRST-GUESS WINDS

1 VAR=R[T]=o; 2 VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=o; 2 VAR=+; BIN ON WS[T]

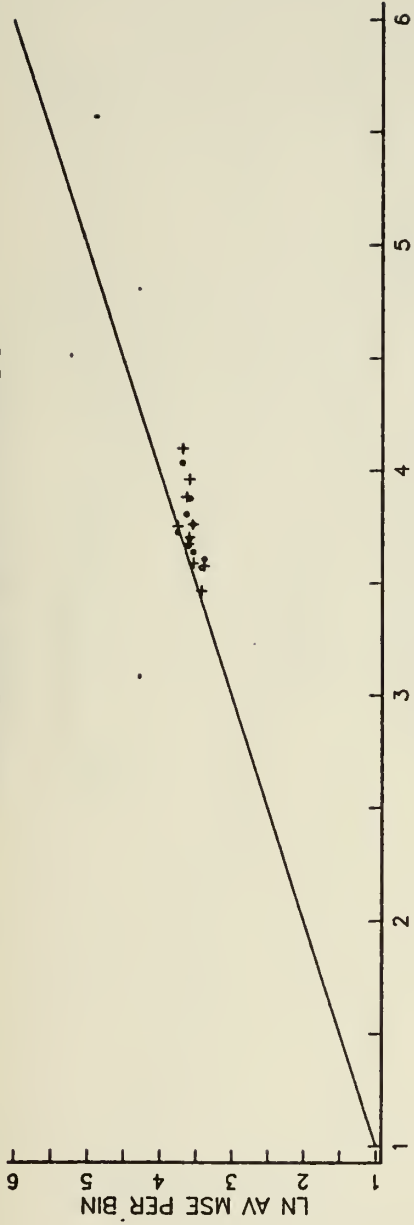


LN AV PRED MSE PER BIN

Figure 22B

250 MB V WIND; MODEL B ON DATA A; FIRST-GUESS WINDS

1 VAR=R[T]=0;2 VAR=+;BIN ON R[T]



1 VAR=WS[T]=0;2 VAR=+;BIN ON WS[T]

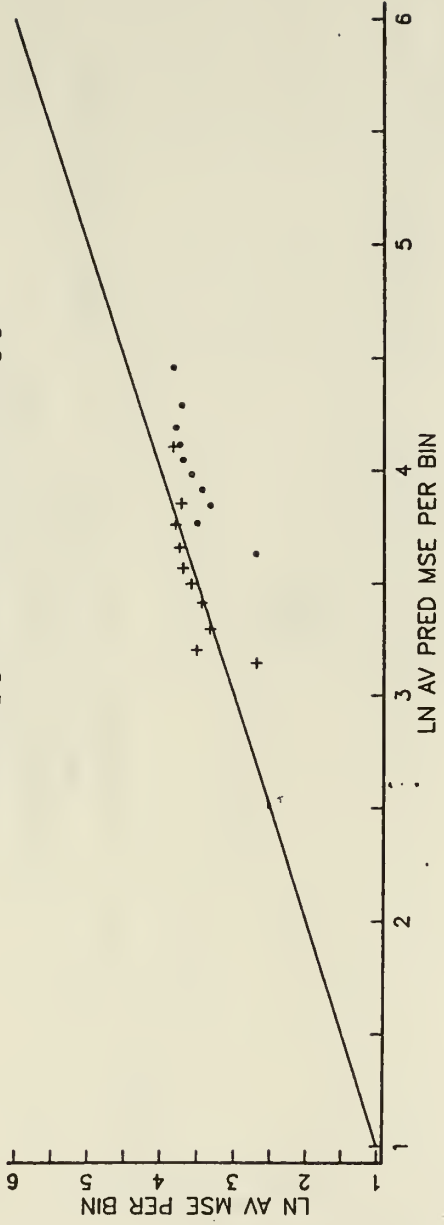
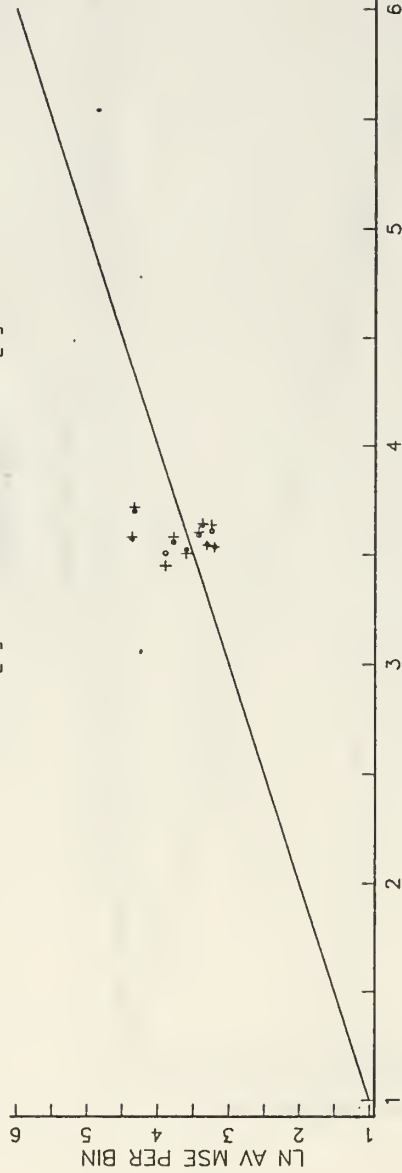


Figure 23B

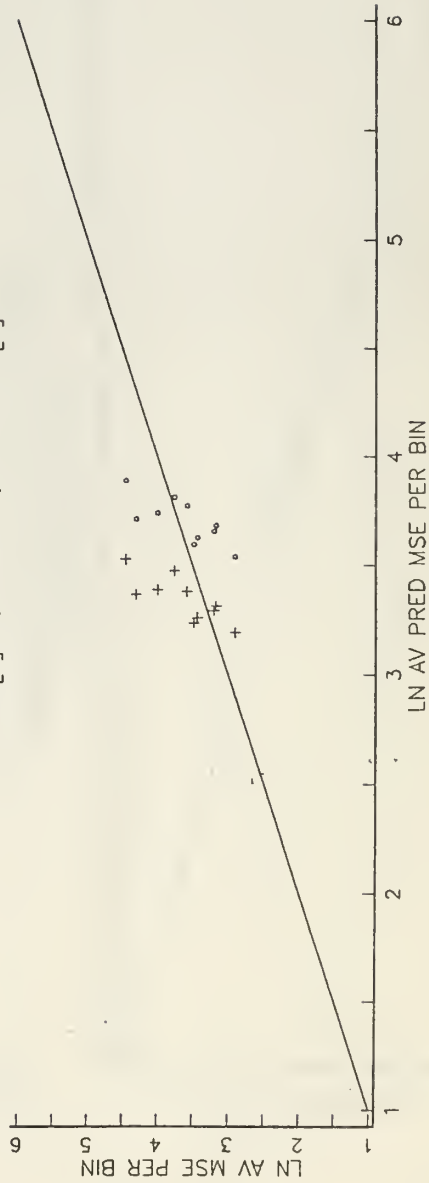
250 MB V WIND; MODEL A ON DATA B; FIRST-GUESS WINDS

1 VAR=R[T]=.2 VAR=+.BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=.2 VAR=+.BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 24B



**APPENDIX C. GRAPHICAL ASSESSMENT OF GOODNESS OF FIT AND  
CROSS-VALIDATION OF MODELS FOR FEBRUARY AND APRIL WIND  
COMPONENT MEAN SQUARE ERROR USING OBSERVED WIND  
COVARIATES**

In this appendix we present graphs assessing goodness of fit and predictive ability of the normal models (1)–(3) with observed wind covariates fit to April and February data.

The maximum likelihood parameter estimates for each model (1)–(3) are obtained for both February and April data and are displayed in Table 5. The estimated variances  $\sigma_1^2(1,t)$ ,  $\sigma_1^2(2,t)$ ,  $\sigma_2^2(t)$  are computed for the parameters estimated from February and April data using (1)–(3) for each data point in February and April.

To assess models (1) and (3) the data  $(y(t), r(t), s(t))$  for each data set are binned into 10 bins based on ordering the values of  $r(t)$  from smallest to largest. The data in the first bin correspond to the smaller values of  $r(t)$ ; the data in the 10<sup>th</sup> bin correspond to the larger values of  $r(t)$ . Each bin contains about  $\frac{1}{10}$ <sup>th</sup> of the data with the 10<sup>th</sup> bin containing a few more. The averages of the estimated variances for models (1) and (3) are computed for each bin. The average  $y(t)^2$  is also computed for each bin.

To assess models (2) and (3) the same procedure is used but the binning is done using  $s(t)$ .

Figures 1C-24C present graphs of the  $\ln[\text{average } y(t)^2]$  in each bin versus  $\ln[\text{average estimated variance}]$  in each bin for models (1) and (3) and models (2) and (3). Figures 1C, 5C, 9C, 13C, 17C, 21C (respectively 2C, 6C, 10C, 14C, 18C

22C) show the logarithm of the average of the  $y(t)^2$  values for February (respectively April) versus the logarithm of the average of the estimated variances for each bin using the estimated parameters from February (respectively April). If a model were perfect, a point should be close to the 45° line shown. These figures assess goodness of fit.

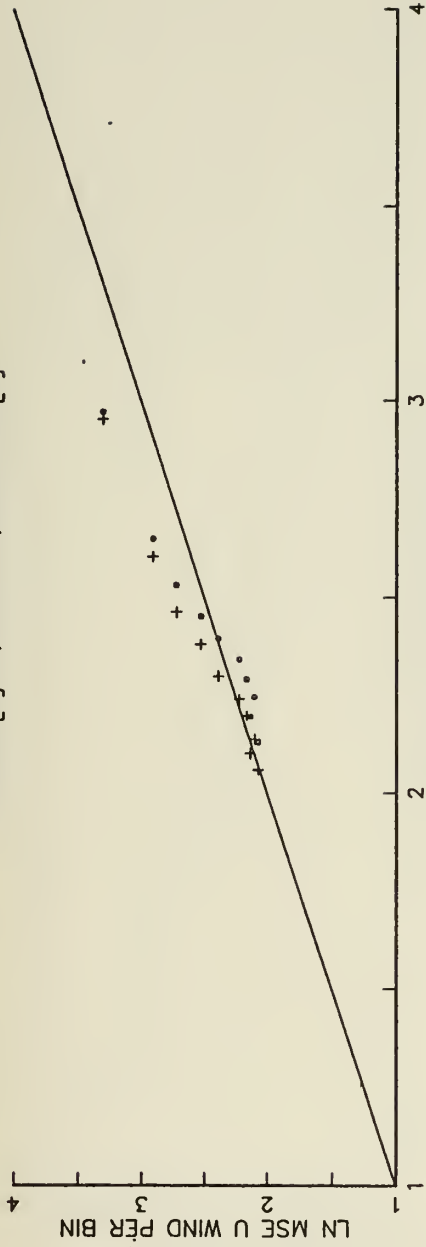
Figures 3C, 7C, 11C, 15C, 19C, 23C (respectively 4C, 8C, 12C, 16C, 20C, 24C) present graphs of  $\ln$  average  $y(t)^2$  of February (respectively April) versus  $\ln$  average estimated variances using parameters estimated using April (respectively February) data. Once again if the model were perfect, the points would be close to the 45° line. These figures assess the ability of models fit using February (respectively April) observed data to predict April (respectively February) wind component mean square error.

The figures indicate once again that the display of  $\ln$  averages can be quite sensitive to which variate is used to do the binning.

Keeping this binning sensitivity in mind, the figures suggest the following. The two-variate model (3) appears to best describe and predict the mean square component wind error. Of the two one-variable models, model (1) which uses  $r(t)$  as the covariate appears to be better. The one-variate model using  $s(t)$  appears to tend to overstate the predicted mean square error. The addition of the second covariate  $r(t)$  to the one-variate model using  $s(t)$  appears to tend to decrease the predicted mean square error.

# 850 MB FEB MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

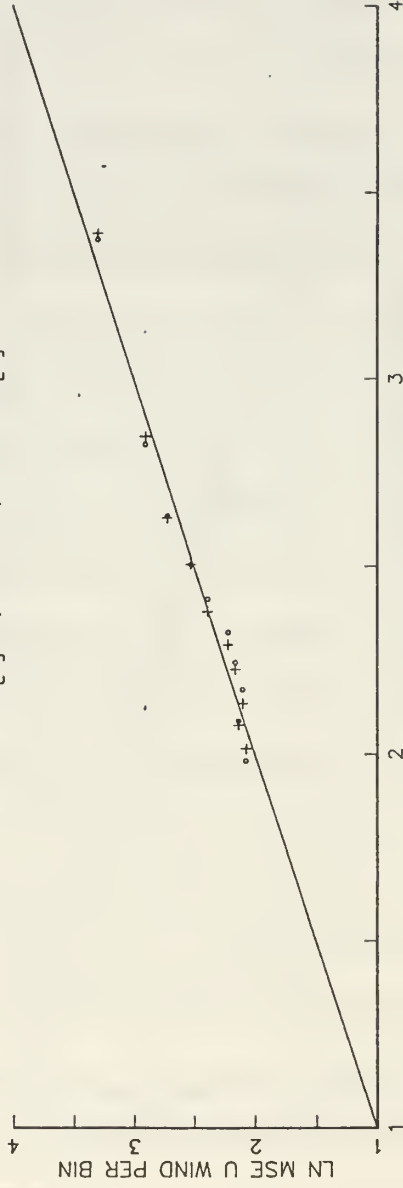


LN AV PRED MSE PER BIN

Figure 1C

850 MB APR MODEL ON APR DATA; OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

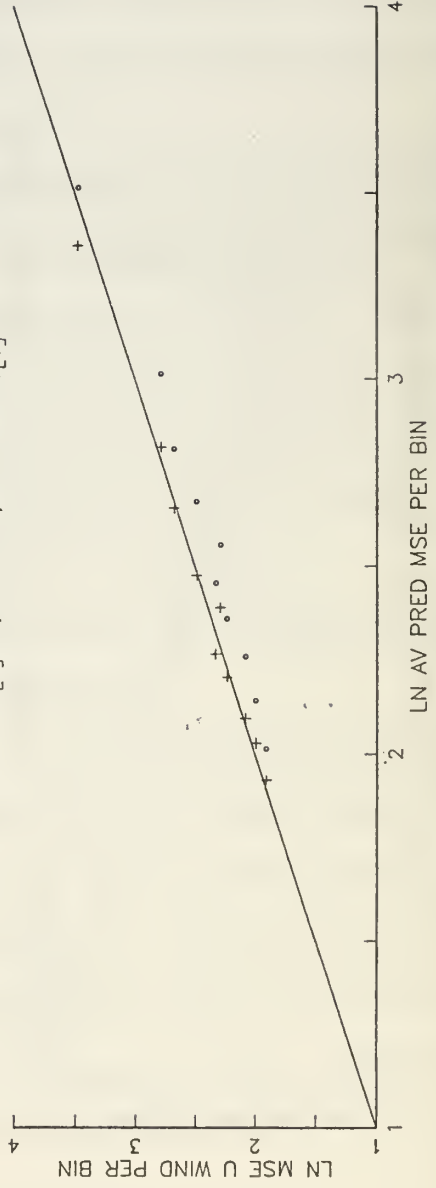
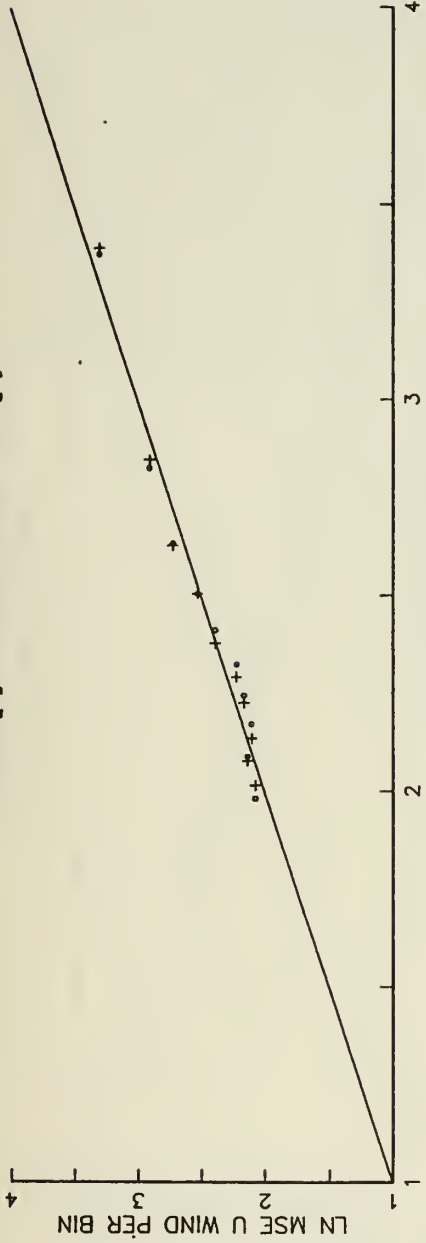


Figure 2C

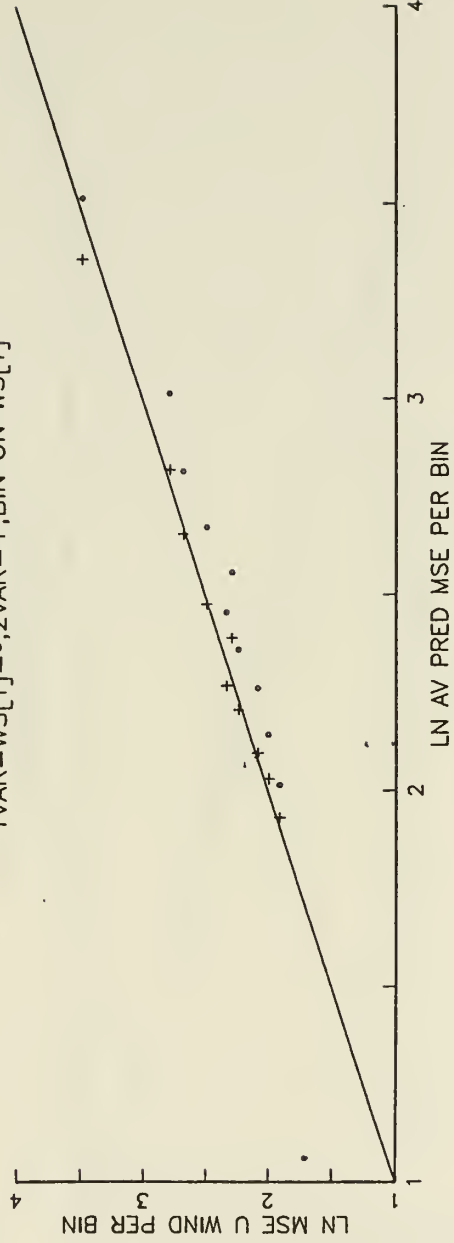
850 MB APR MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=0;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=0;2VAR=+;BIN ON WS[T]

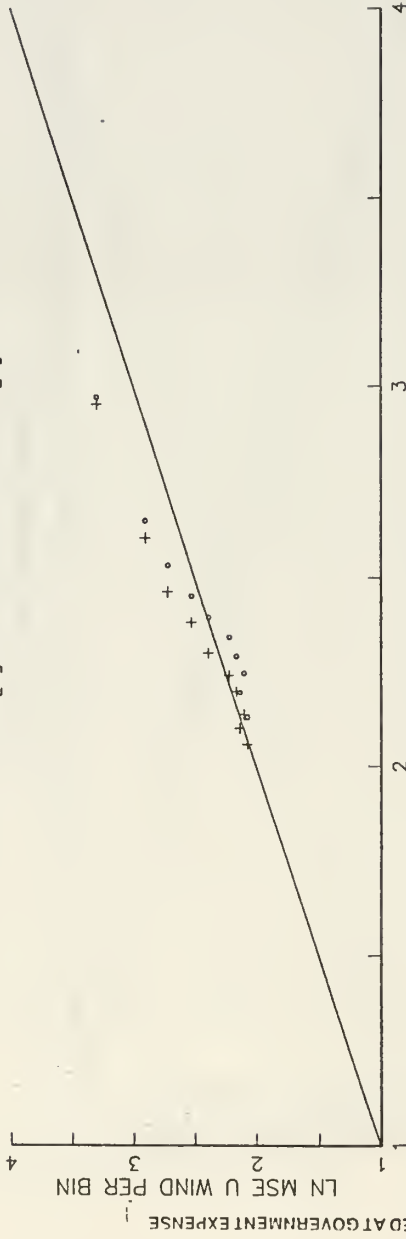


LN AV PRED MSE PER BIN

Figure 3C

850 MB FEB MODEL ON APR DATA; OBS WIND

$$1VAR=R[T]=\circ; 2VAR=+; BIN \text{ ON } R[T]$$



LN AV PRED MSE PER BIN

$$1VAR=WS[T]=\circ; 2VAR=+; BIN \text{ ON } WS[T]$$

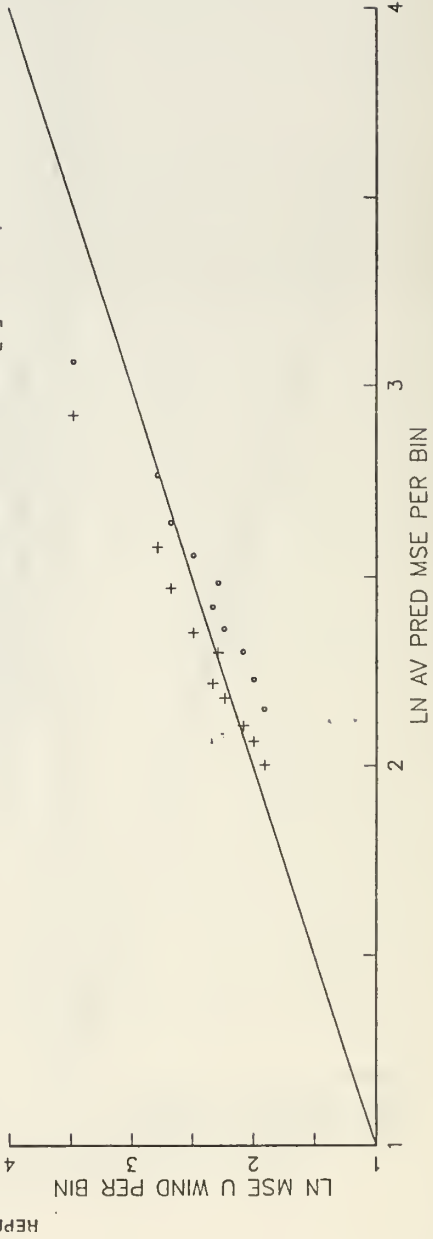


Figure 4C

REPRODUCED AT GOVERNMENT EXPENSE

850 MB V WIND; FEB MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

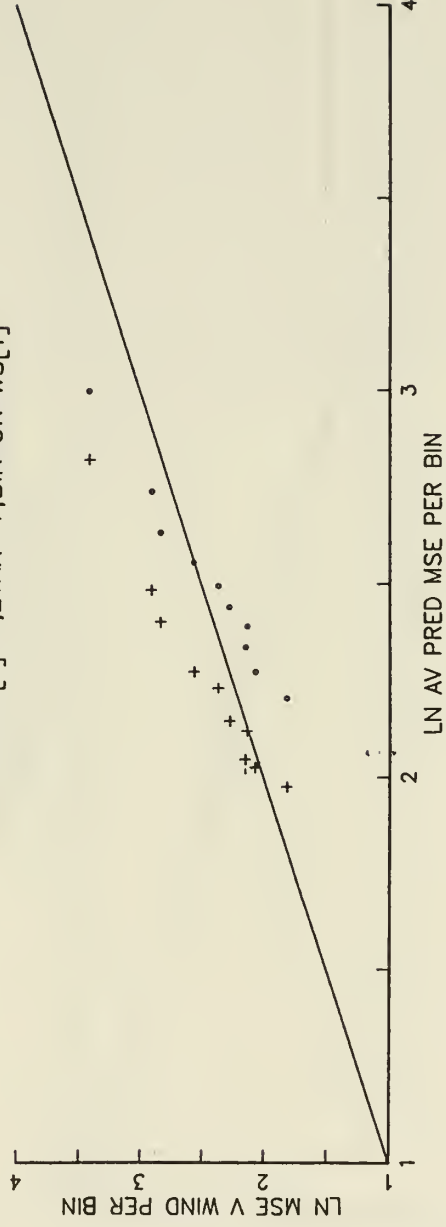
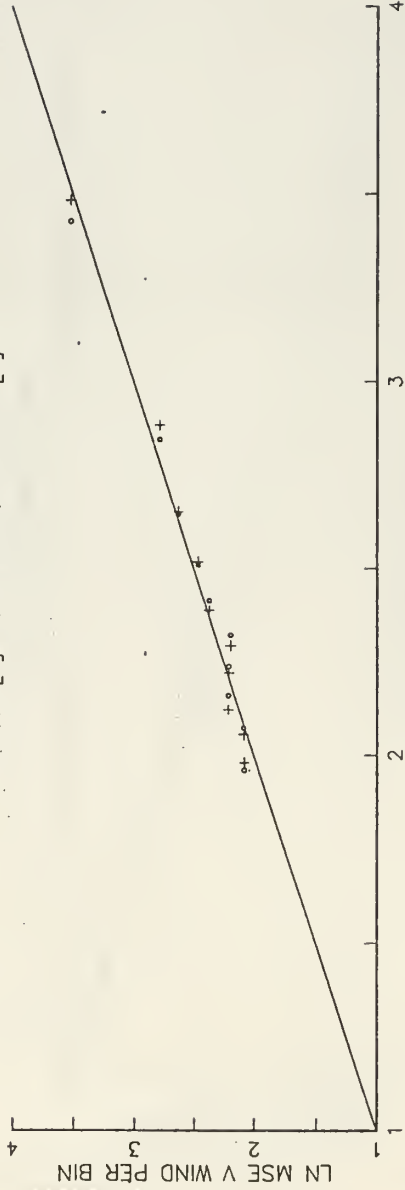


Figure 5C

850 MB V WIND; APR MODEL ON APR DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

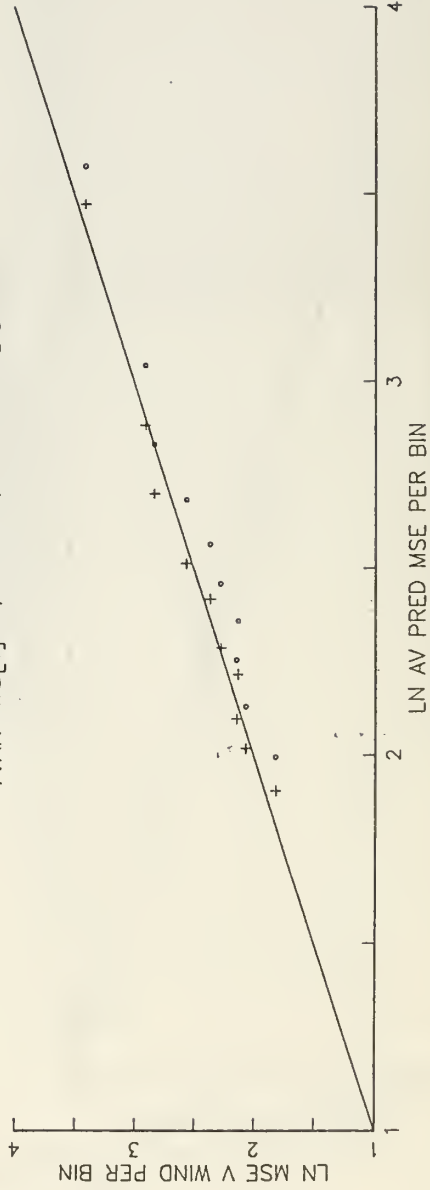
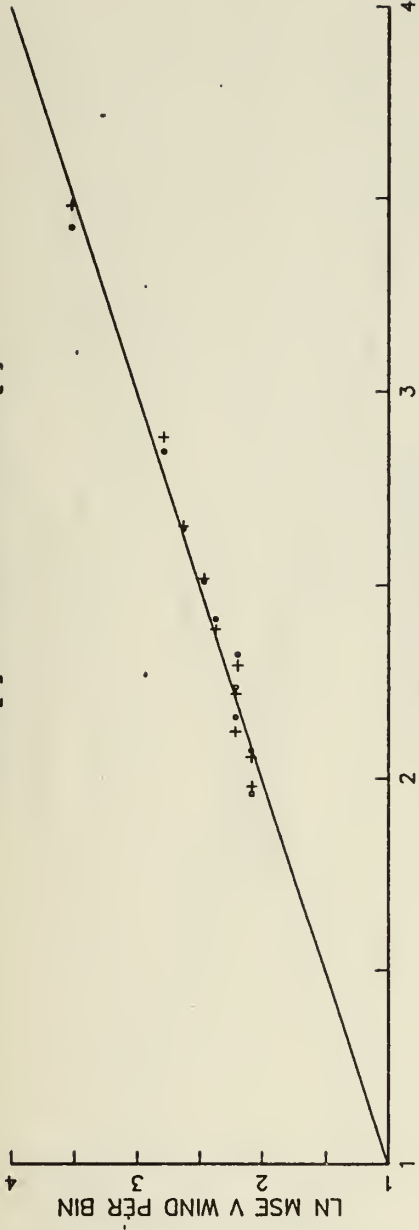


Figure 6C



850 MB V WIND; APR MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN  
1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

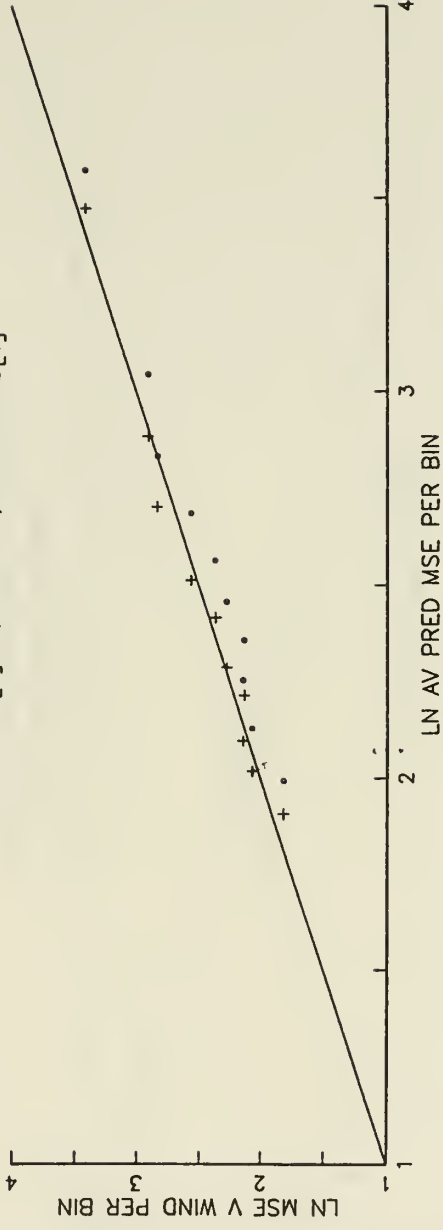
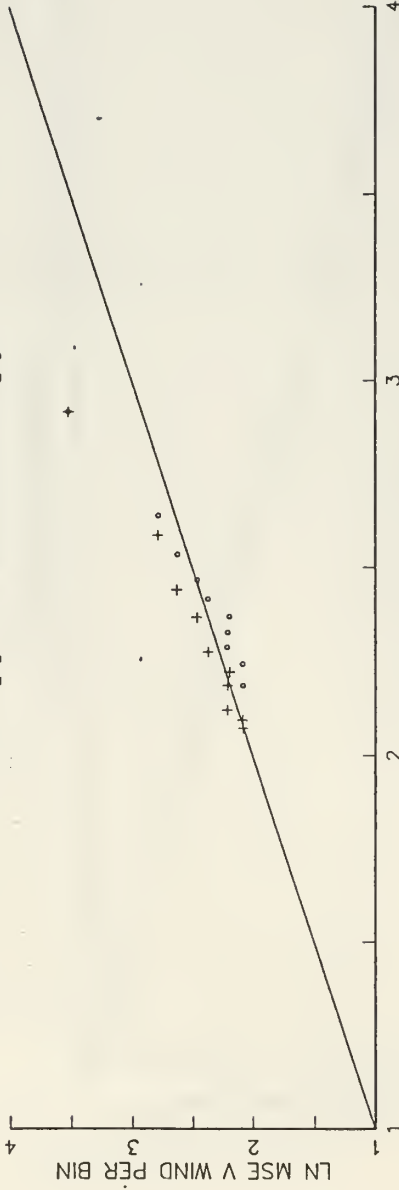


Figure 7C

850 MB V WIND; FEB MODEL ON APR DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

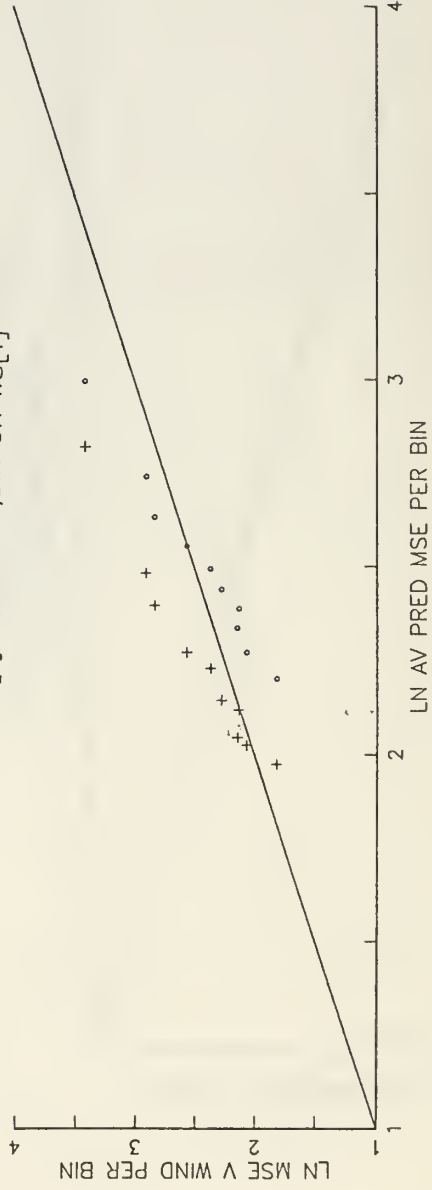
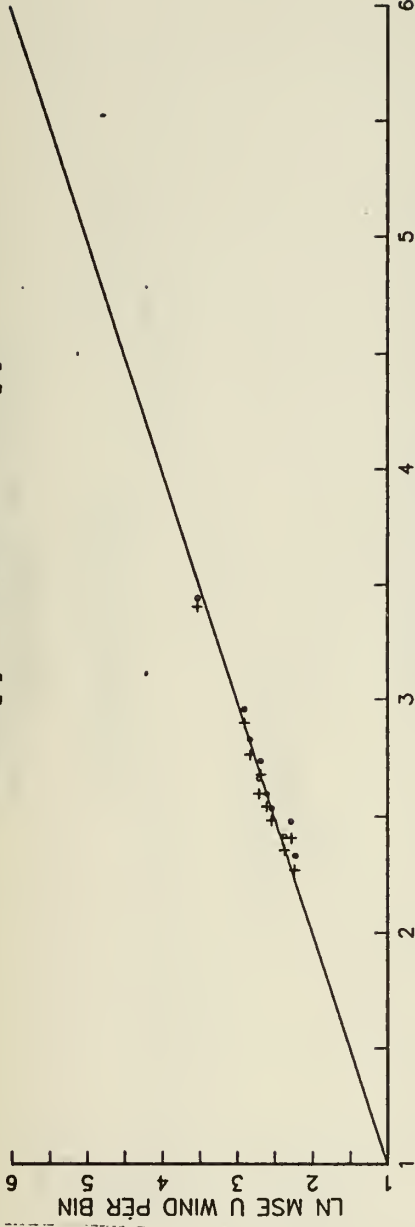


Figure 8C

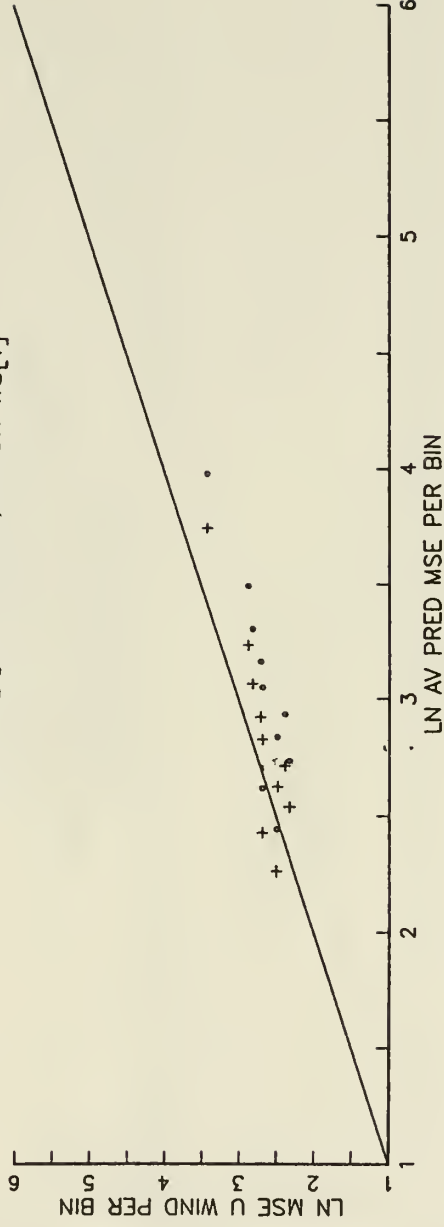
500 MB U WIND; FEB MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

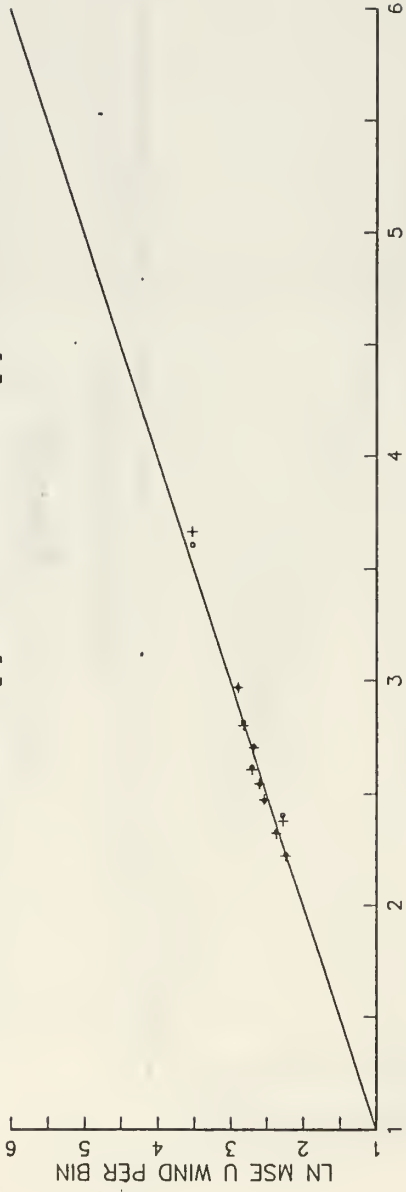


LN AV PRED MSE PER BIN

Figure 9C

500 MB U WIND; APR MODEL ON APR DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

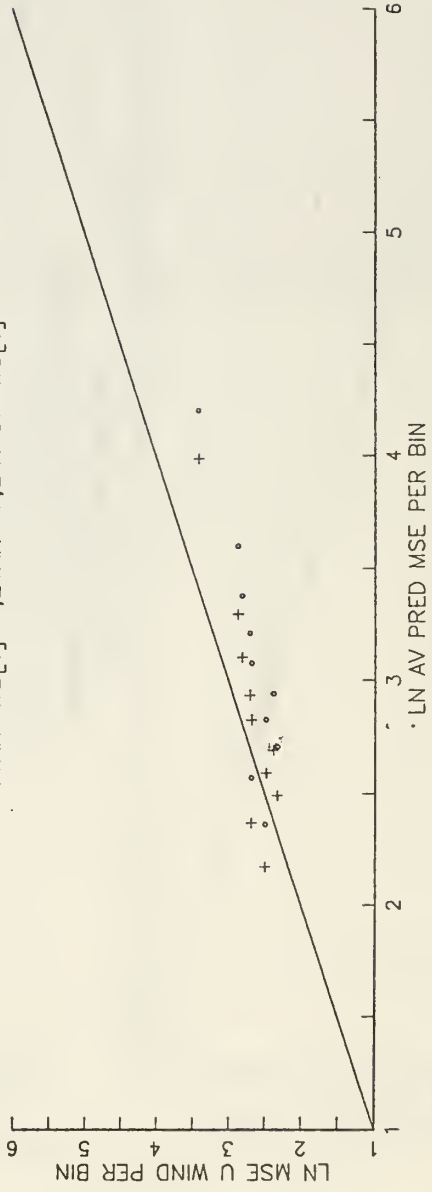
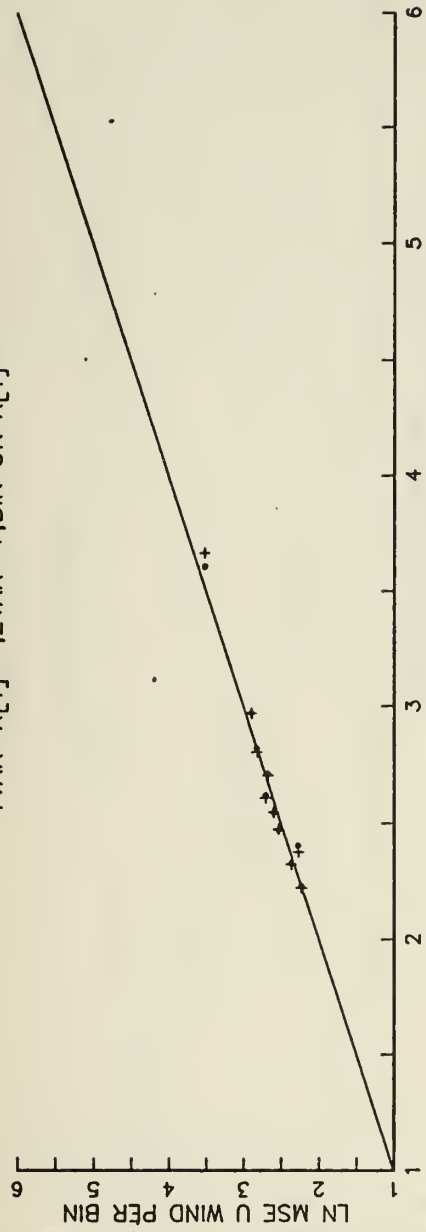


Figure 10C

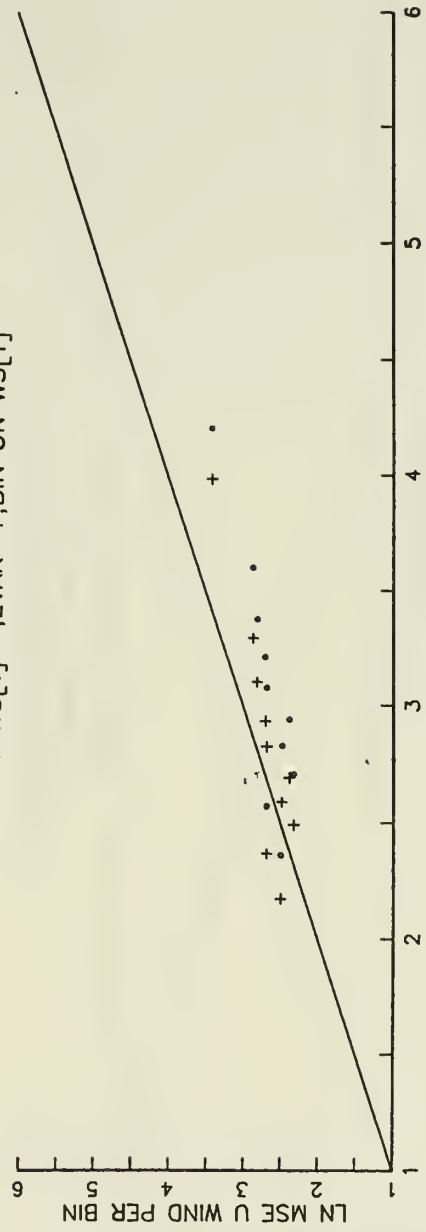
500 MB U WIND; APR MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=°;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=°;2VAR=+;BIN ON WS[T]

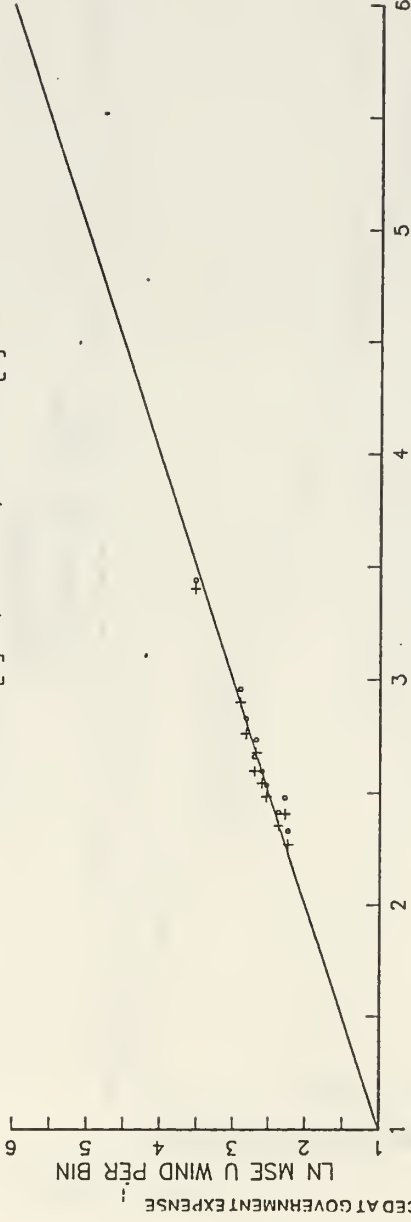


LN AV PRED MSE PER BIN

Figure 11C

500 MB U WIND; FEB MODEL ON APR DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

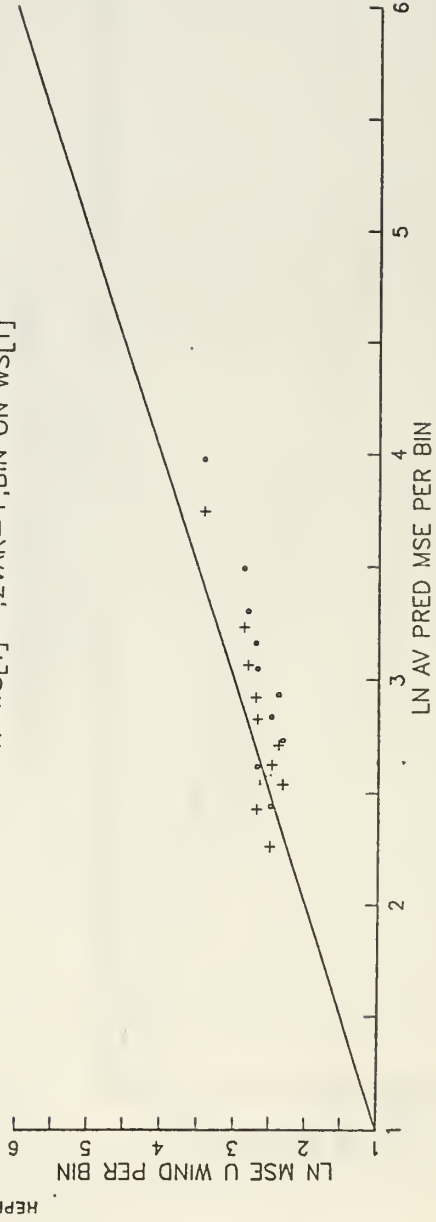
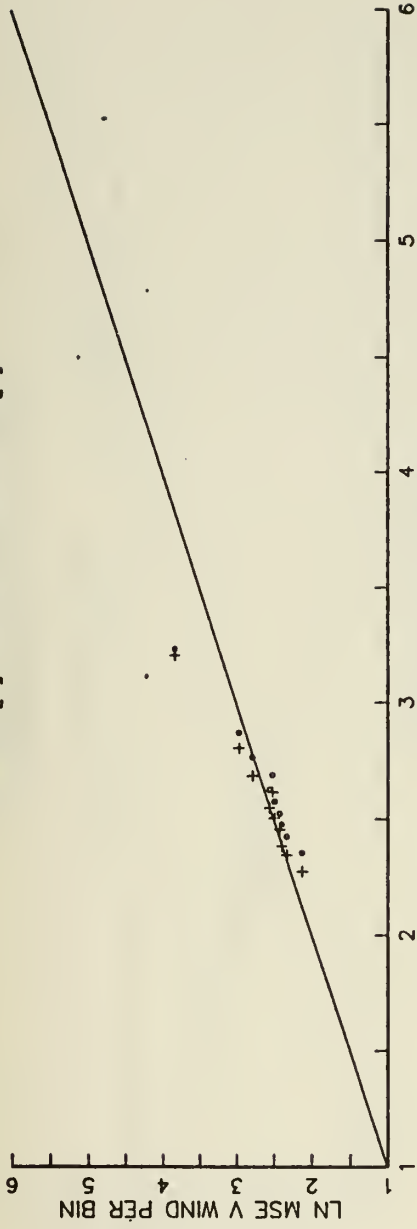


Figure 12C

500 MB V WIND; FEB MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

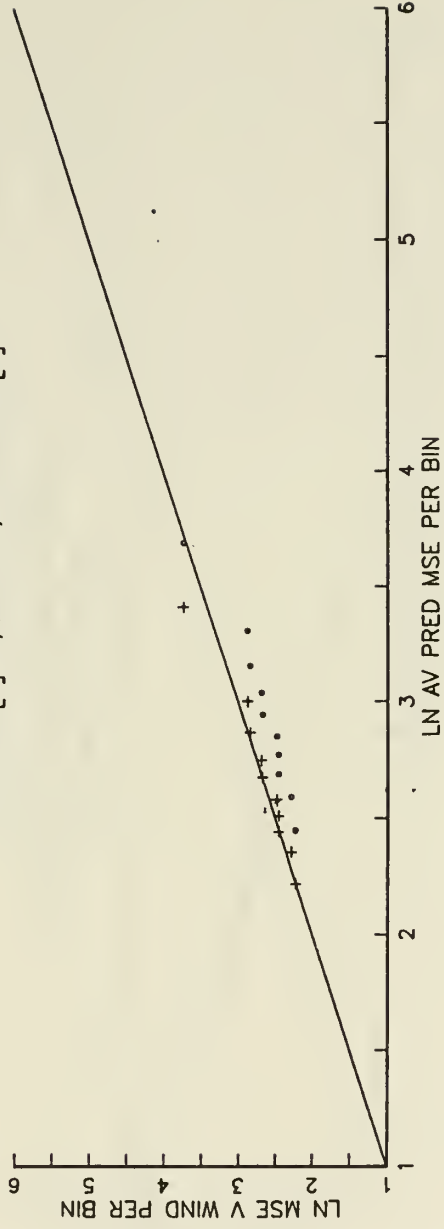
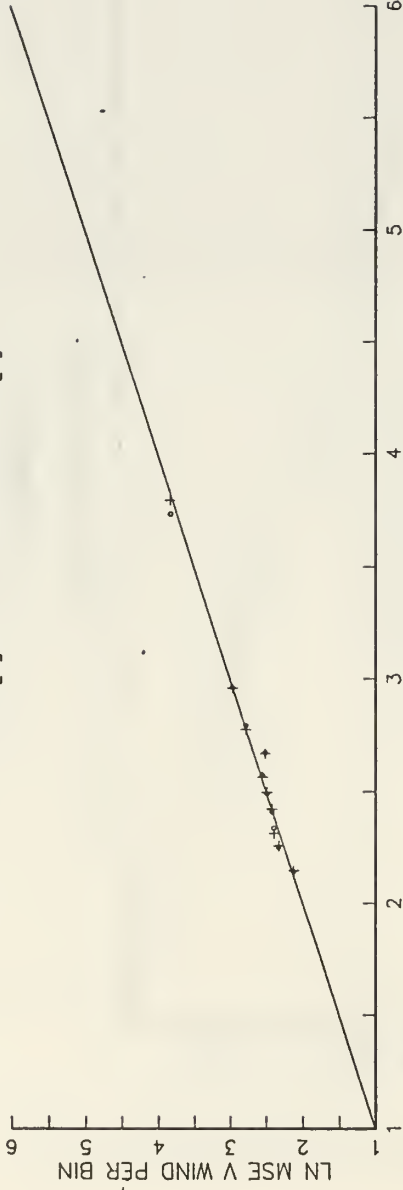


Figure 13C

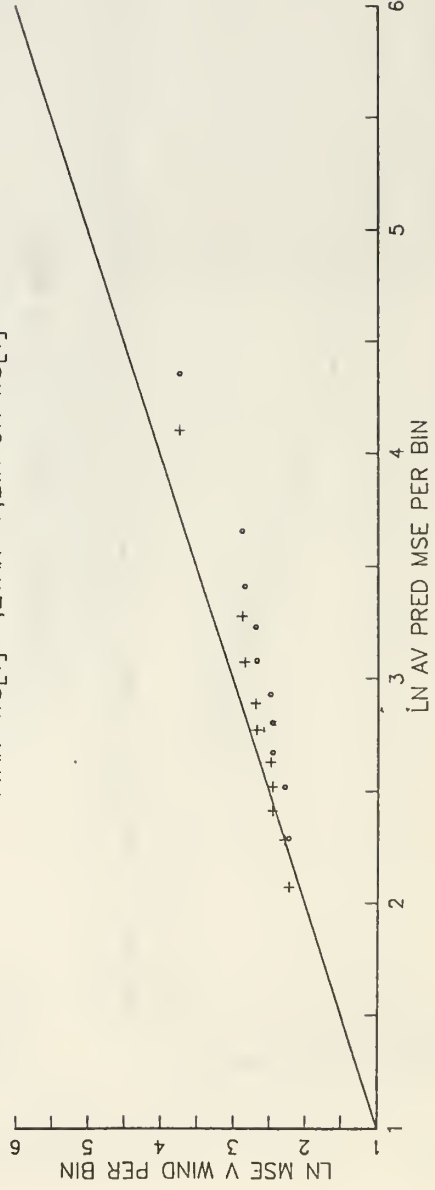
500 MB V WIND; APR MODEL ON APR DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T].



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]



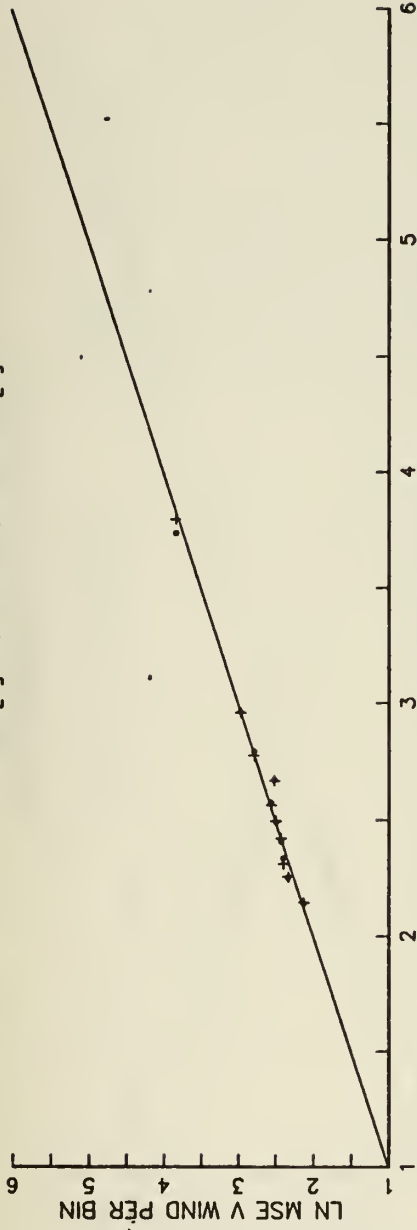
LN AV PRED MSE PER BIN

Figure 14C



500 MB V WIND; APR MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

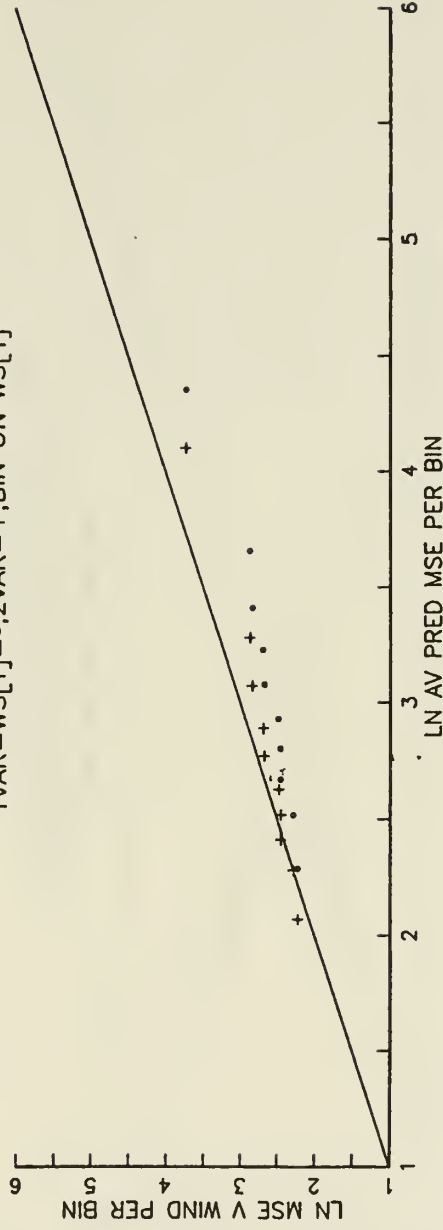
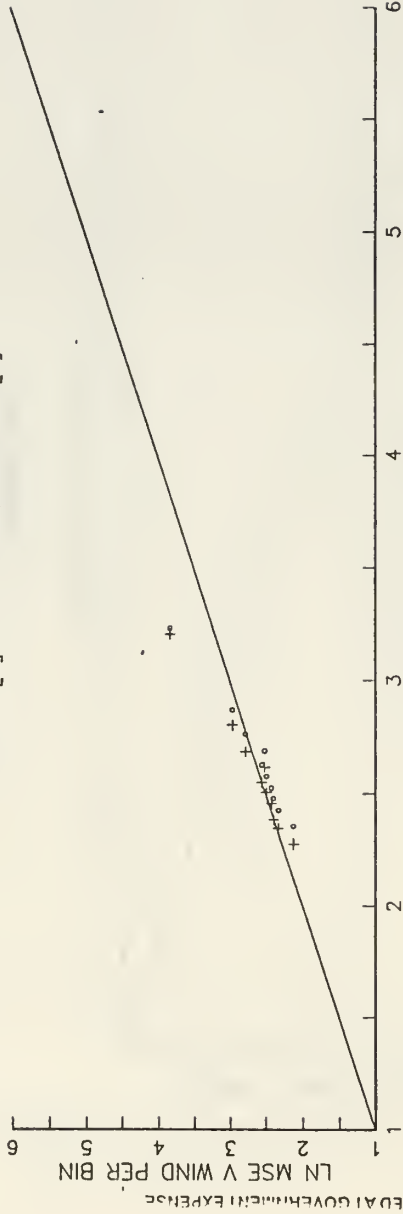


Figure 15C

500 MB V WIND; FEB MODEL ON APR DATA; OBS WIND

$$1\text{VAR}=\text{R}[T]=\circ; 2\text{VAR}=\text{+}; \text{BIN ON R}[T]$$



$$1\text{VAR}=\text{WS}[T]=\circ; 2\text{VAR}=\text{+}; \text{BIN ON WS}[T]$$

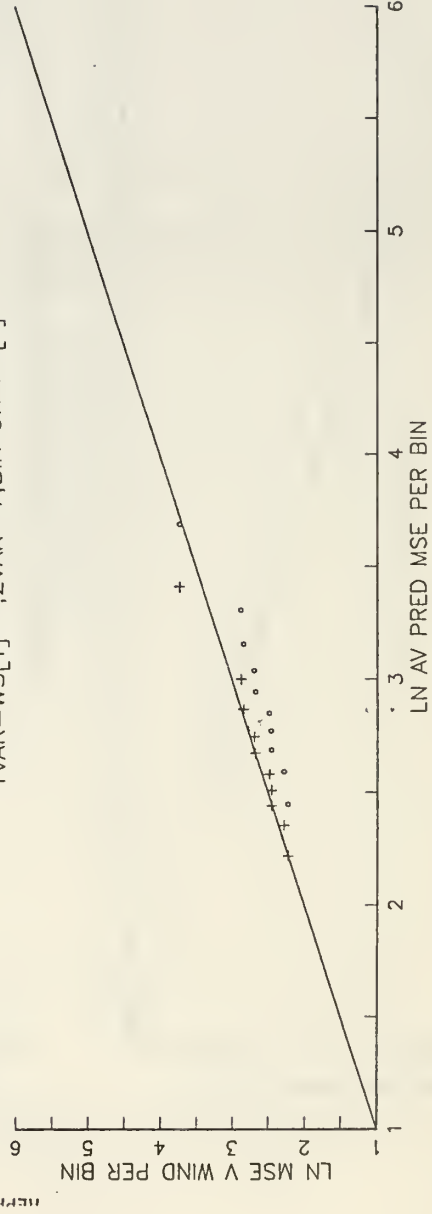
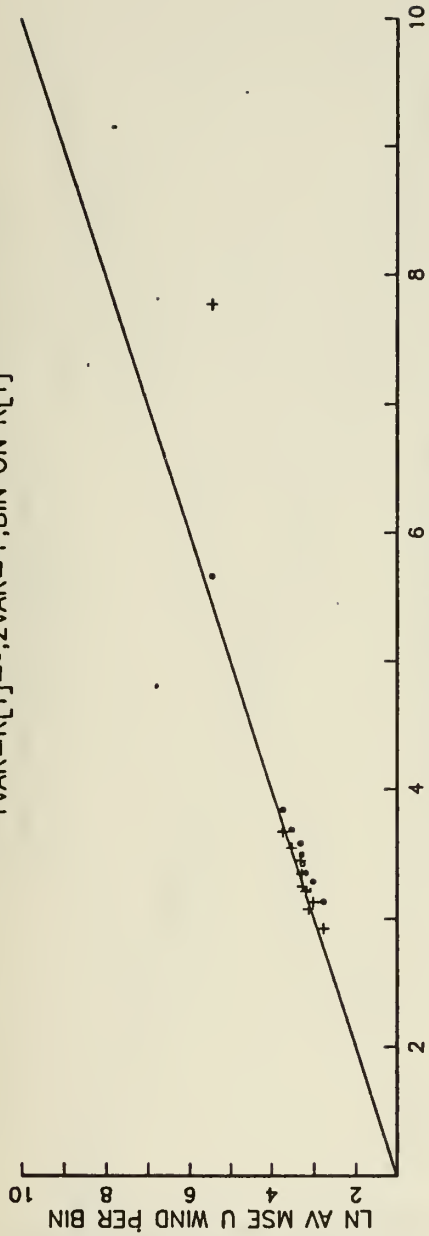


Figure 16C

250-MB U WIND; FEB MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

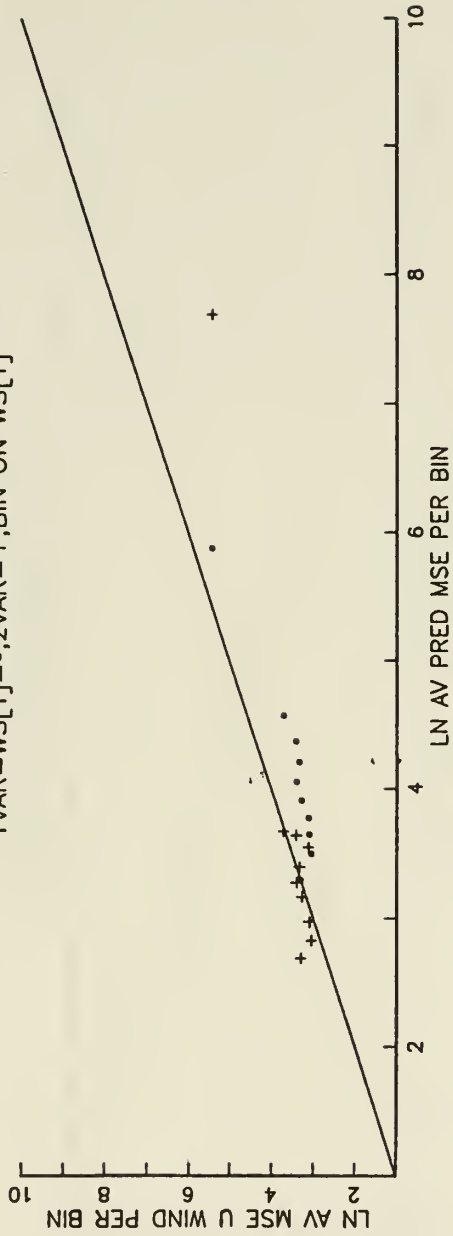
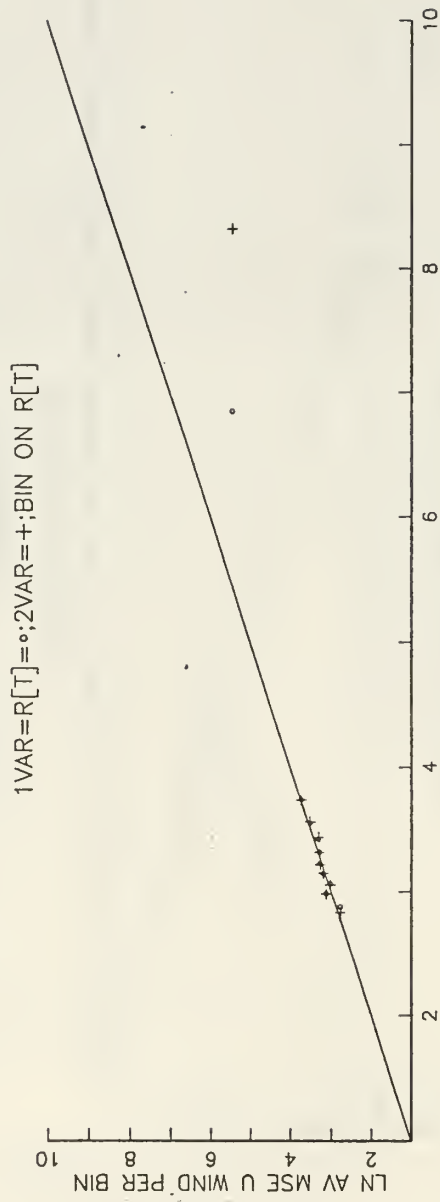


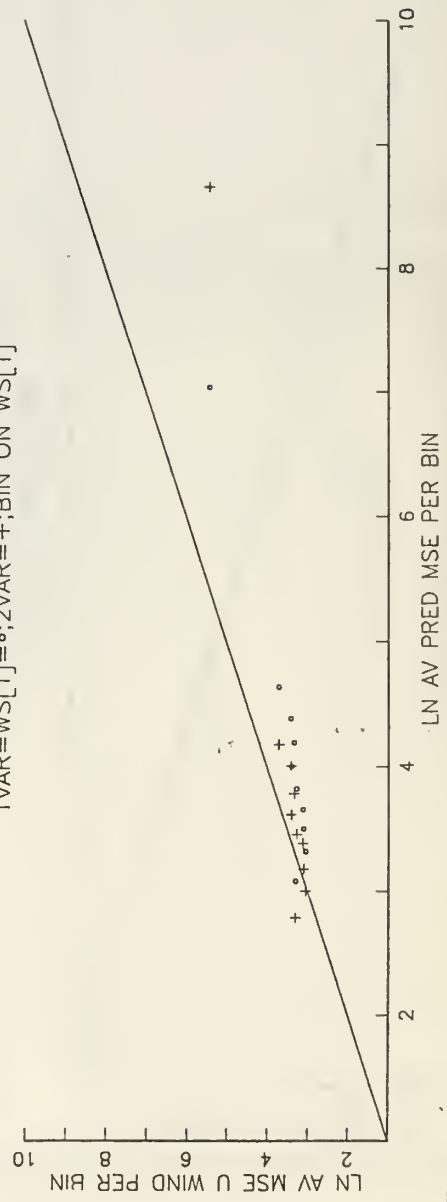
Figure 17C

250 MB U WIND; APR MODEL ON APR DATA; OBS WIND



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

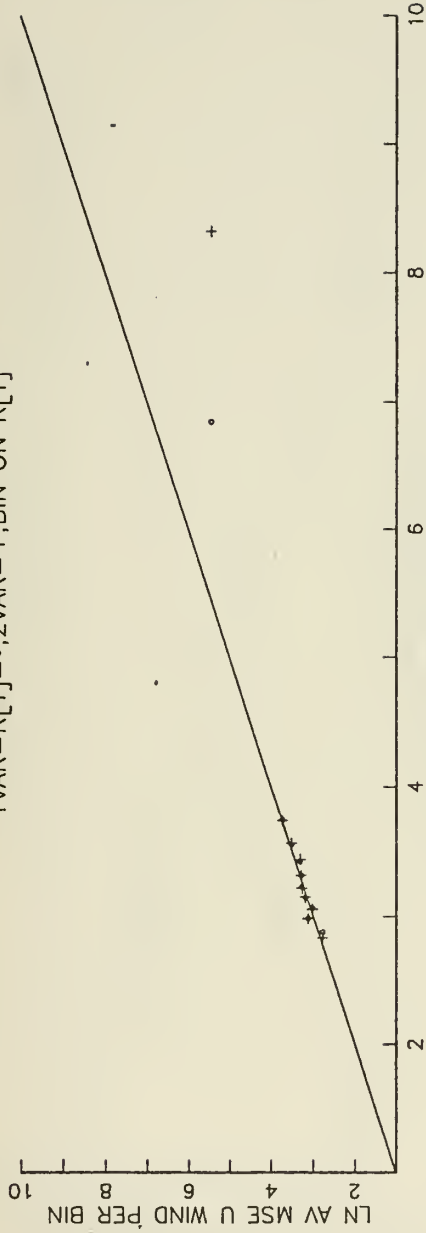


LN AV PRED MSE PER BIN

Figure 18C

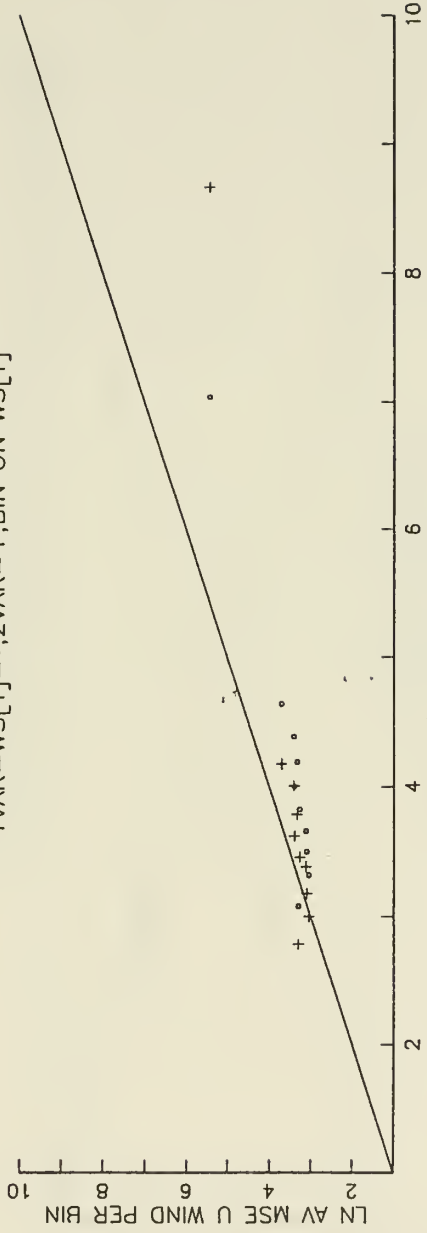
250 MB U WIND; APR MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=o;2VAR=+;BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]

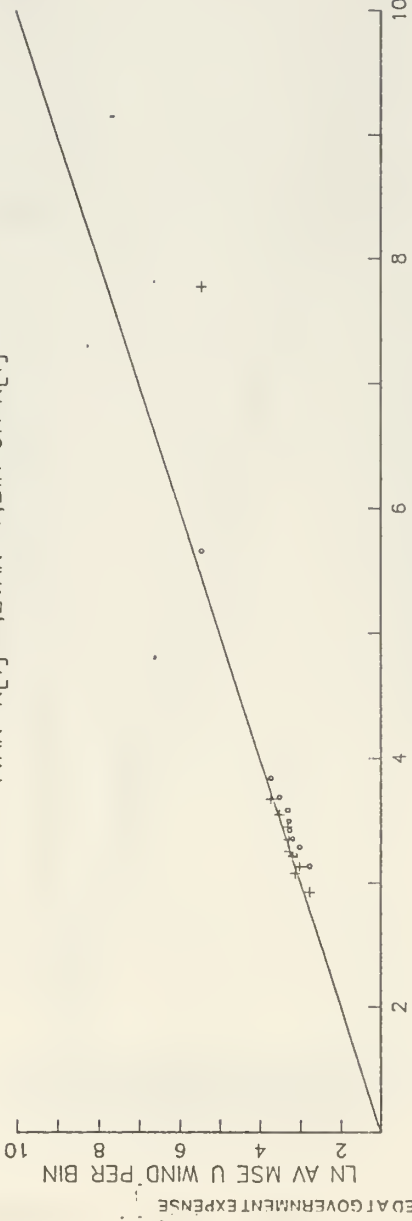


LN AV PRED MSE PER BIN

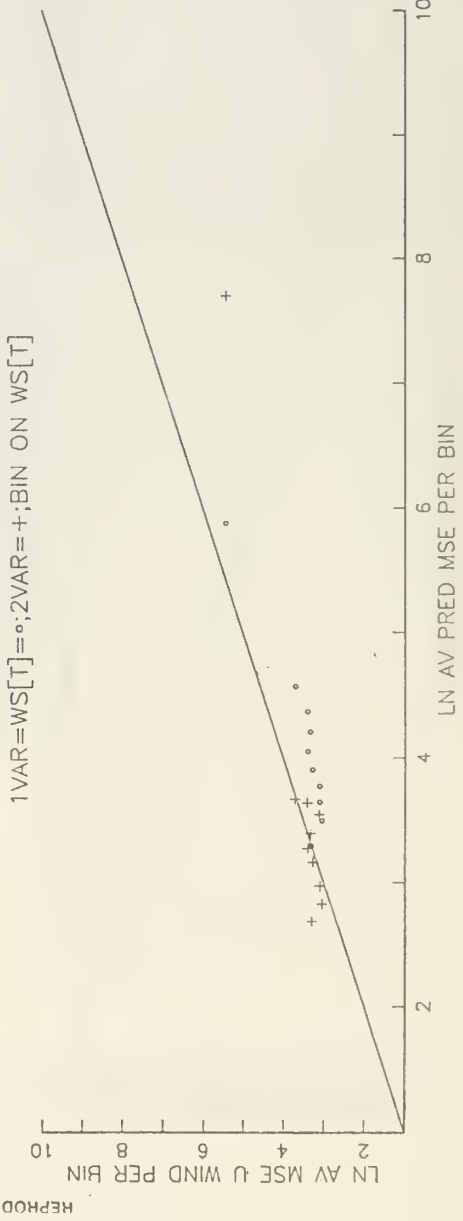
Figure 19C

250 MB U WIND; FEB MODEL ON APR DATA; OBS WIND

$$1 \text{ VAR} = R[T] = \circ; 2 \text{ VAR} = +; \text{BIN ON } R[T]$$



$$1 \text{ VAR} = WS[T] = \circ; 2 \text{ VAR} = +; \text{BIN ON } WS[T]$$

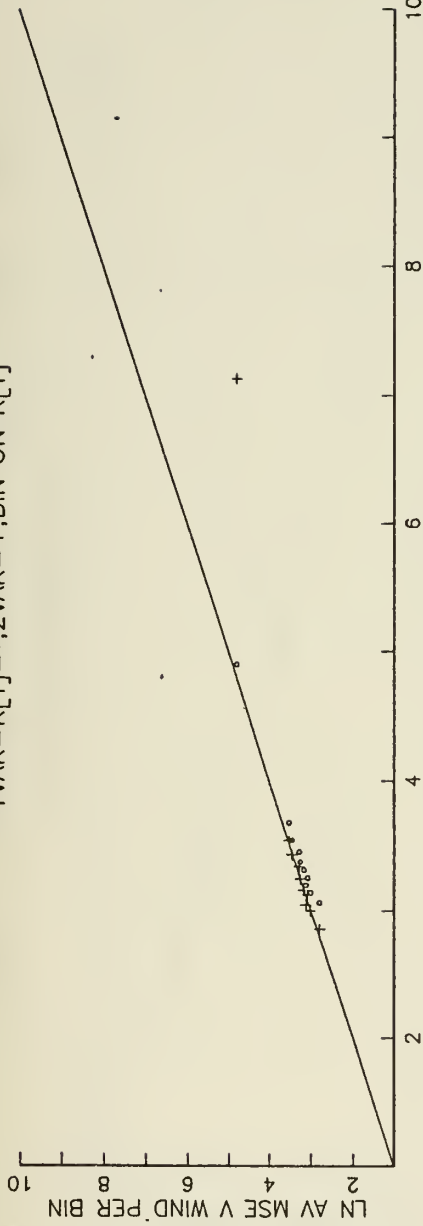


REPRODUCED AT GOVERNMENT EXPENSE

Figure 20C

250 MB V WIND; FEB MODEL ON FEB DATA; OBS WIND

1 VAR=R[T]=0; 2 VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1 VAR=WS[T]=0; 2 VAR=+; BIN ON WS[T]

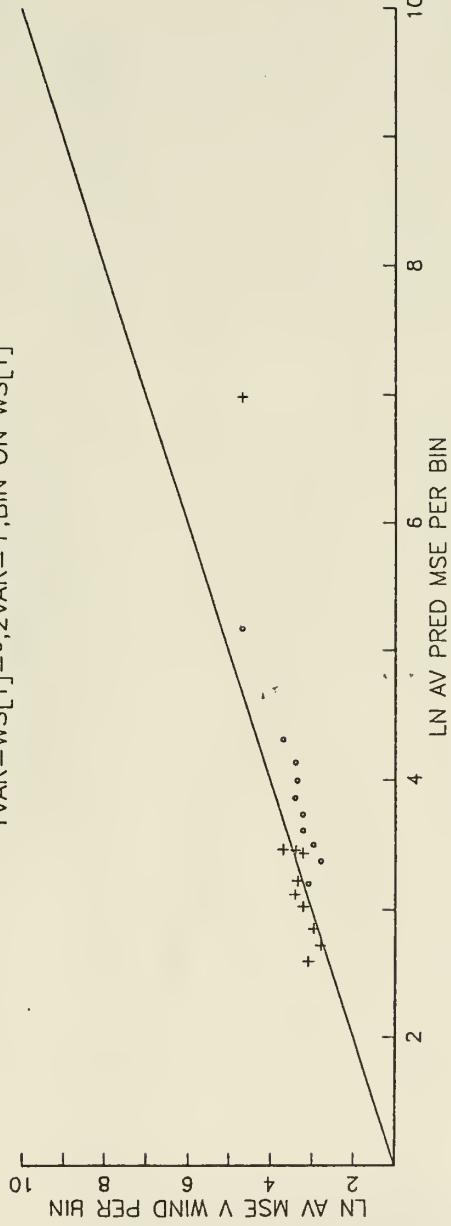
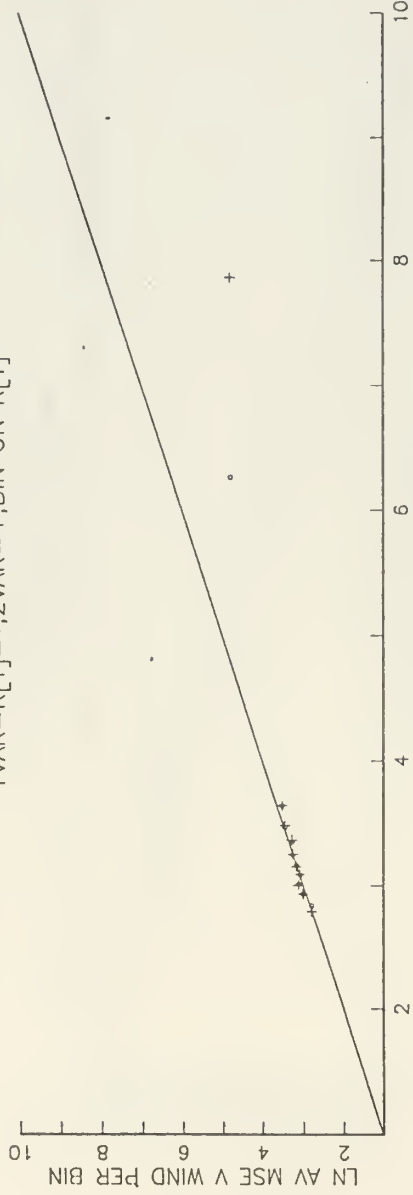


Figure 21C

250 MB V WIND; APR MODEL ON APR DATA; OBS WIND

$$1\text{VAR}=R[T]=\circ; 2\text{VAR}=\circ; \text{BIN ON } R[T]$$



$$1\text{VAR}=WS[T]=\circ; 2\text{VAR}=\circ; \text{BIN ON } WS[T]$$

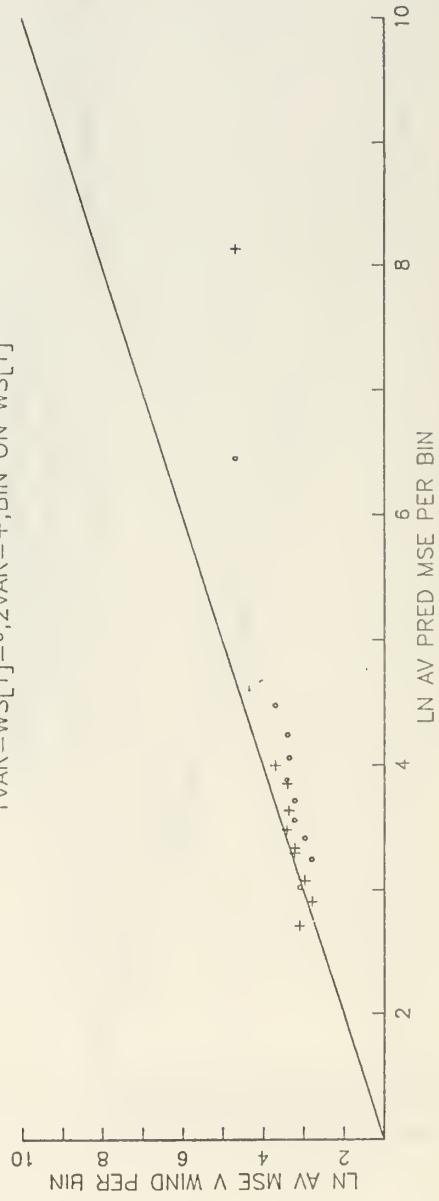
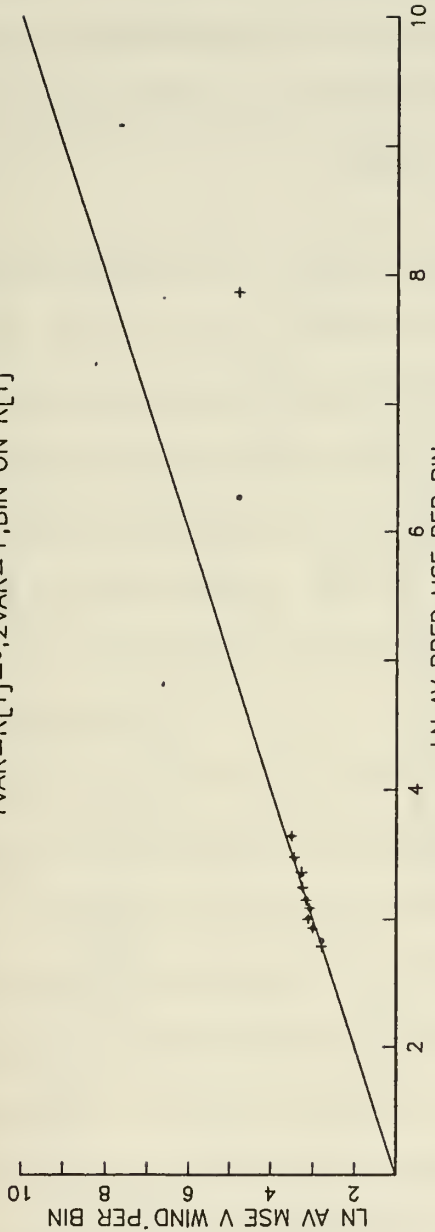


Figure 22C



250 MB V WIND; APR MODEL ON FEB DATA; OBS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

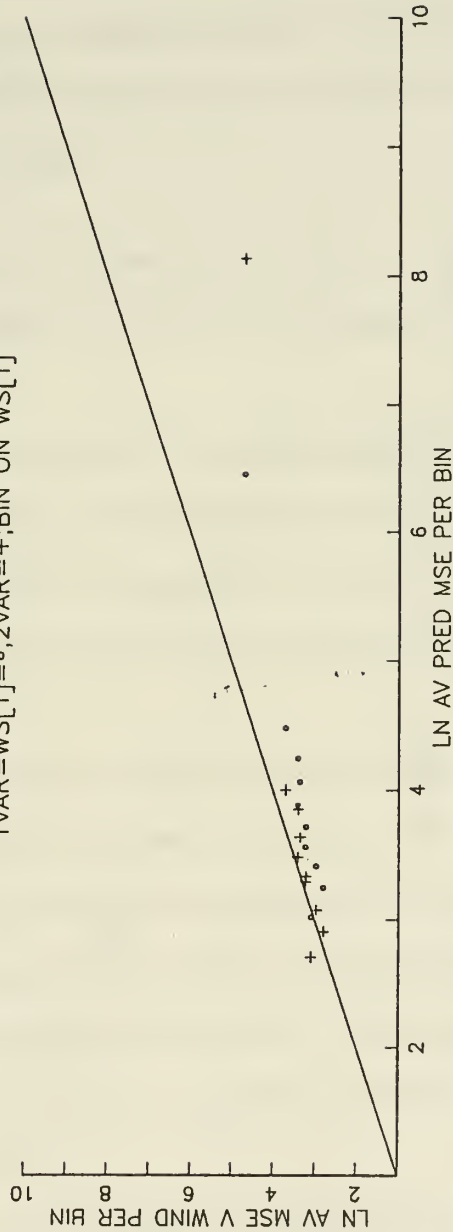
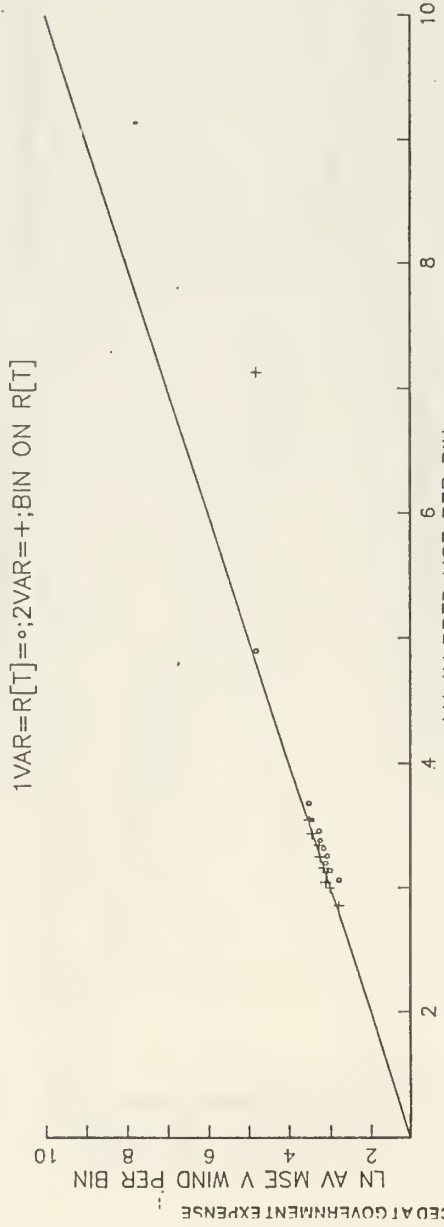


Figure 23C

250 MB V WIND; FEB MODEL ON APR DATA; OBS WIND



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

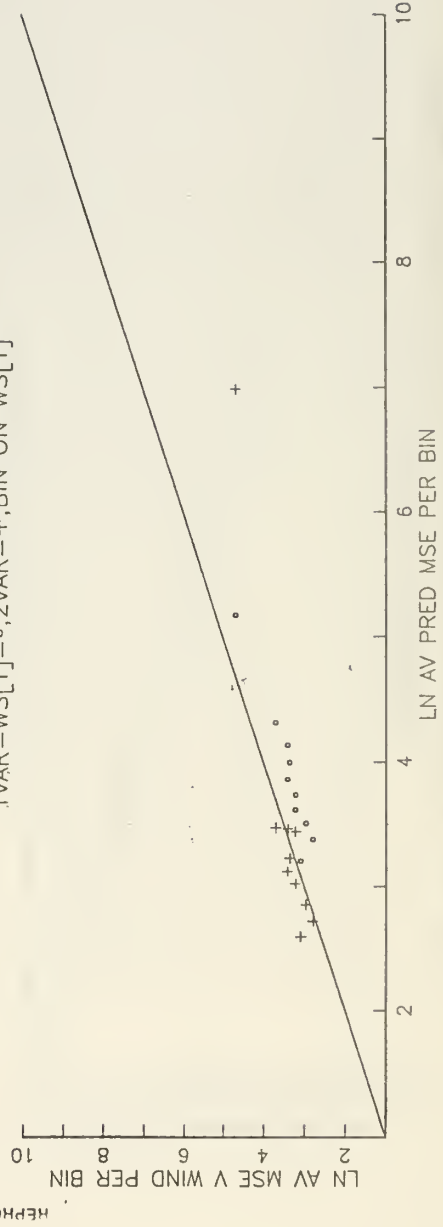


Figure 24C

**APPENDIX D. GRAPHICAL ASSESSMENT OF GOODNESS OF FIT AND  
CROSS-VALIDATION OF MODELS FOR FEBRUARY AND APRIL WIND  
COMPONENT MEAN SQUARE ERROR USING FIRST-GUESS WIND  
COVARIATES**

In this appendix we present graphs assessing goodness of fit and predictive ability of the normal models (1)–(3) with first-guess wind covariates fit to April and February data.

The maximum likelihood parameter estimates for each model (1)–(3) are obtained for both February and April data and are displayed in Table 7. The estimated variances  $\sigma_1^2(1,t)$ ,  $\sigma_1^2(2,t)$ ,  $\sigma_2^2(t)$  are computed for the parameters estimated from February and April data using (1)–(3) for each data point in February and April.

To assess models (1) and (3) the data  $(y(t), r(t), s(t))$  for each data set are binned into 10 bins based on ordering the values of  $r(t)$  from smallest to largest. The data in the first bin correspond to the smaller values of  $r(t)$ ; the data in the 10<sup>th</sup> bin correspond to the larger values of  $r(t)$ . Each bin contains about  $\frac{1}{10}$ <sup>th</sup> of the data with the 10<sup>th</sup> bin containing a few more. The averages of the estimated variances for models (1) and (3) are computed for each bin. The average  $y(t)^2$  is also computed for each bin.

To assess models (2) and (3) the same procedure is used but the binning is done using  $s(t)$ .

Figures 1D-24D present graphs of the  $\ln[\text{average } y(t)^2]$  in each bin versus  $\ln[\text{average estimated variance}]$  in each bin for models (1) and (3) and models (2) and (3). Figures 1D, 5D, 9D, 13D, 17D, 21D (respectively 2D, 6D, 10D, 14D,

18D 22D) show the logarithm of the average of the  $y(t)^2$  values for February (respectively April) versus the logarithm of the average of the estimated variance for each bin using the estimated parameters from February (respectively April). If a model were perfect, a point should be close to the 45° line shown. These figures are an indication of goodness of fit.

Figures 3D, 7D, 11D, 15D, 19D, 23D (respectively 4D, 8D, 12D, 16D, 20D, 24D) present graphs of  $\ln$  average  $y(t)^2$  of February (respectively April) versus  $\ln$  average estimated variances using parameters estimated using April (respectively February) data. Once again if the model were perfect, the points would be close to the 45° line. These figures assess the ability of models fit using February (respectively April) first-guess data to predict April (respectively February) wind component mean square error.

The figures indicate once again that the display of  $\ln$  averages can be quite sensitive to which variate is used to do the binning.

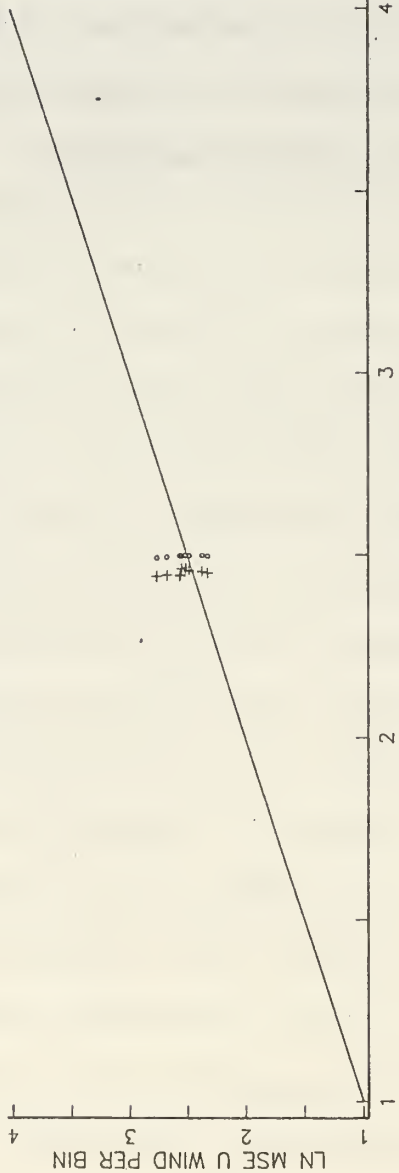
The figures indicate the following. As suggested by comparison of the  $\ln$  likelihood values,  $\tilde{\ell}$ , of Tables 6 and 8 for models with observed wind covariates and first guess wind covariates, the figures suggest that models using first guess wind covariates do not describe or predict mean square error for wind components as well as models using observed wind components. The two-variate model appears to tend to produce smaller mean square errors than the one-variate models; this tendency is most striking in the figure with first guess wind speed being used as the single covariate.

The models fit using April first guess data appear to tend to be better descriptive and predictive models than those fit using February first guess data.

The figures indicating predictive ability (3D, 4D, 7D, 8D, 11D, 15D, 19D, 20D, 23D and 24D) correspond fairly well to the differences between the minimizing value of  $\tilde{\ell}$  for the models with covariates and the value of  $\tilde{\ell}$  for the constant model (no covariates) in the corresponding rows of Table 8. If the value of  $\tilde{\ell}$  for the constant model is larger than any other values in the row, the corresponding figure for that row shows no association.

850 MB U WIND; FEB MODEL ON FEB DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

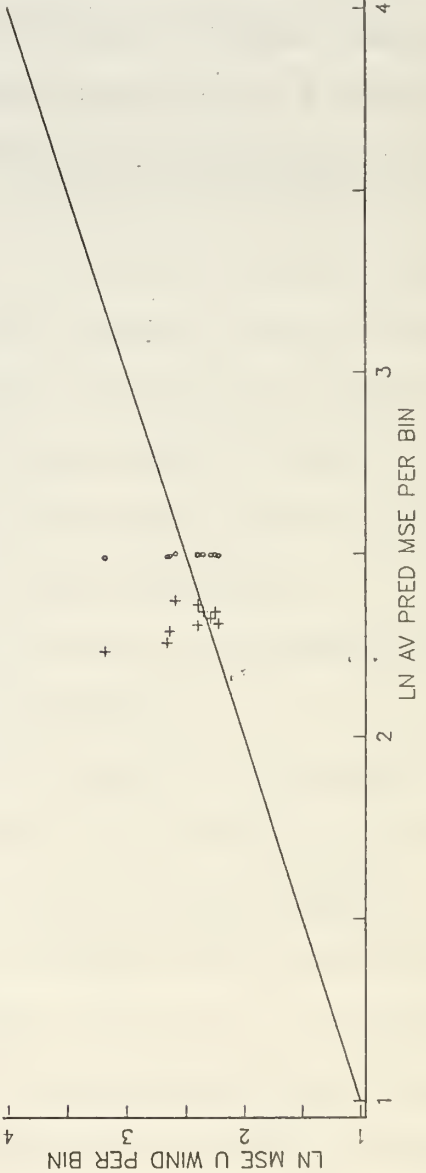
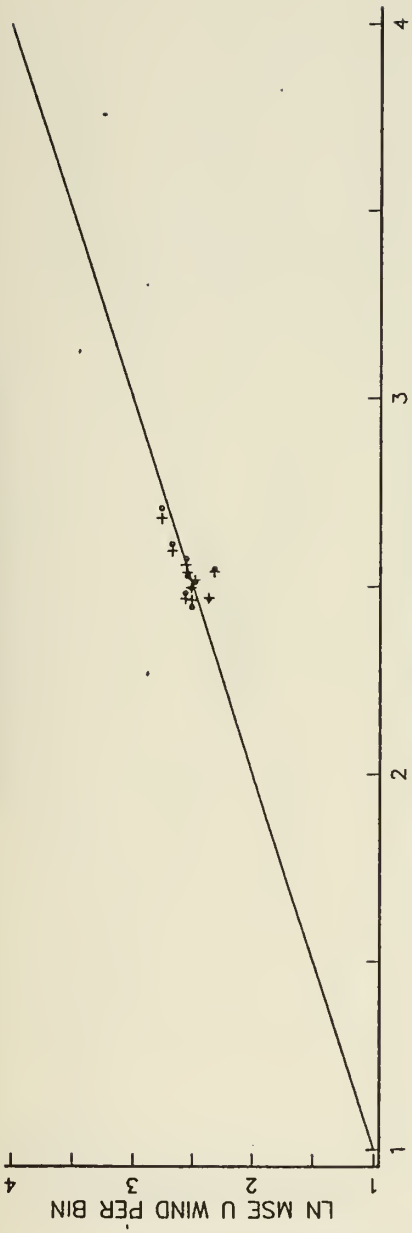


Figure 1D

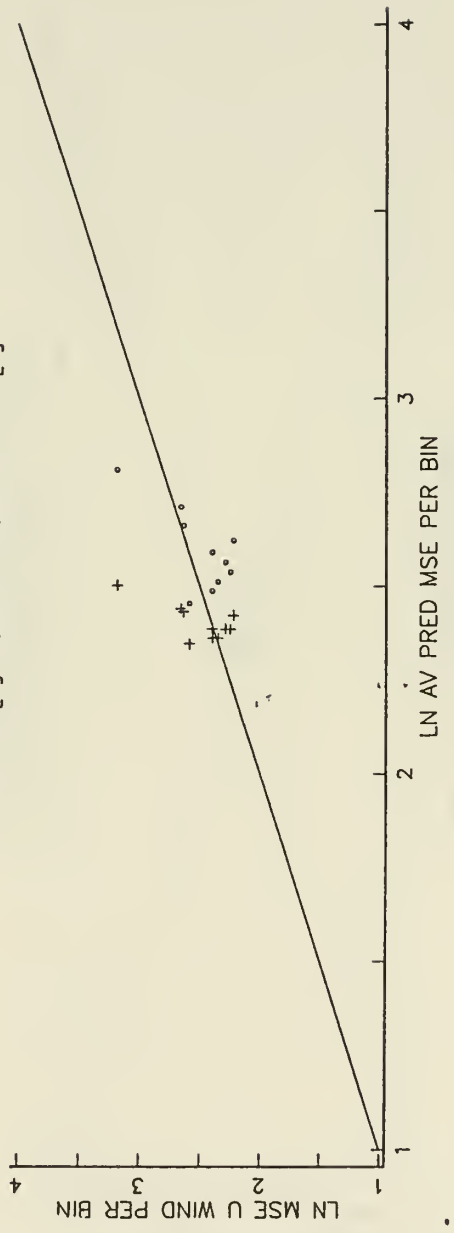
850 MB U WIND; APR MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 2D

850 MB U WIND; APR MODEL ON FEB DATA; 1ST GUESS WIND

$$1\text{VAR} = R[T] = \sigma; 2\text{VAR} = +; \text{BIN ON } R[T]$$



LN AV PRED MSE PER BIN

$$1\text{VAR} = WS[T] = \sigma; 2\text{VAR} = +; \text{BIN ON } WS[T]$$

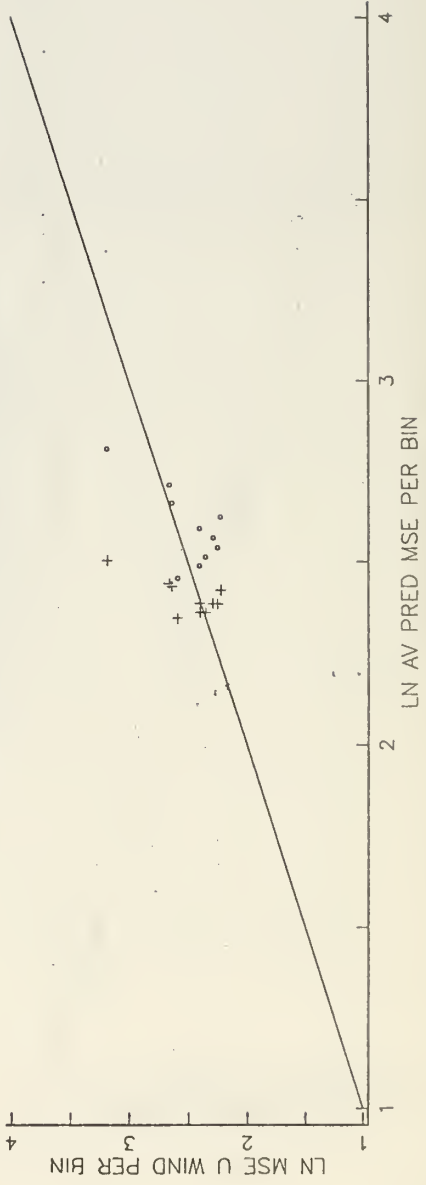


Figure 3D



850 MB U WIND; FEB MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=0; 2VAR=+; BIN ON R[T]



1VAR=WS[T]=0; 2VAR=+; BIN ON WS[T]

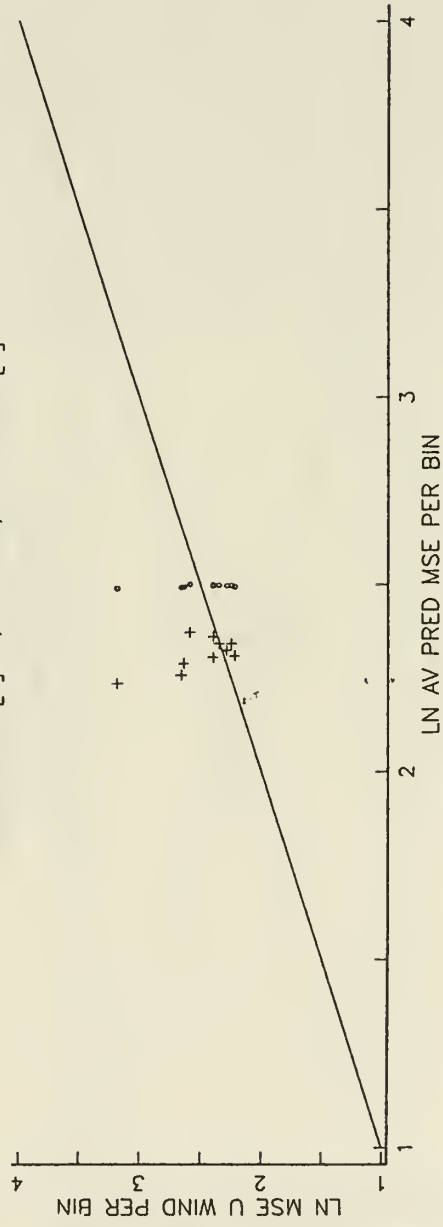
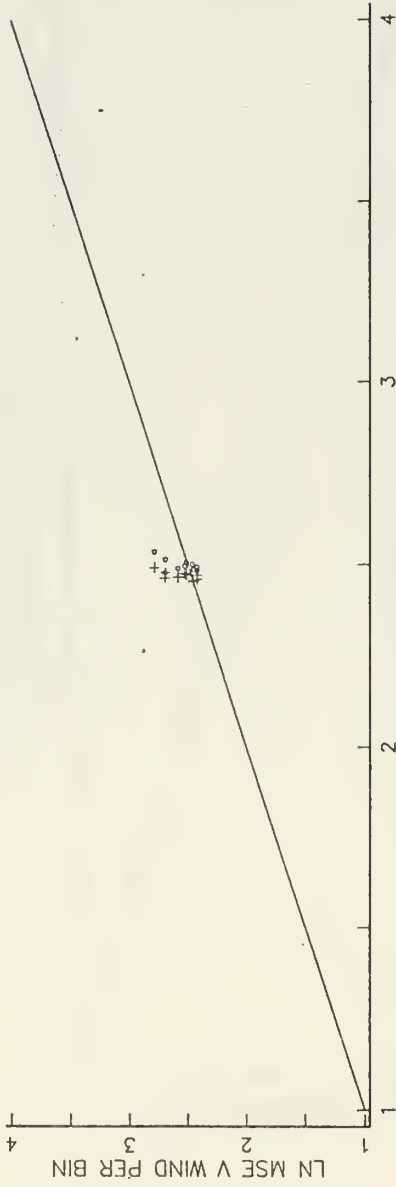


Figure 4D

850 MB V WIND; FEB MODEL ON FEB DATA; 1ST GUESS WIND

1VAR=R[T]=0; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=0; 2VAR=+; BIN ON WS[T]

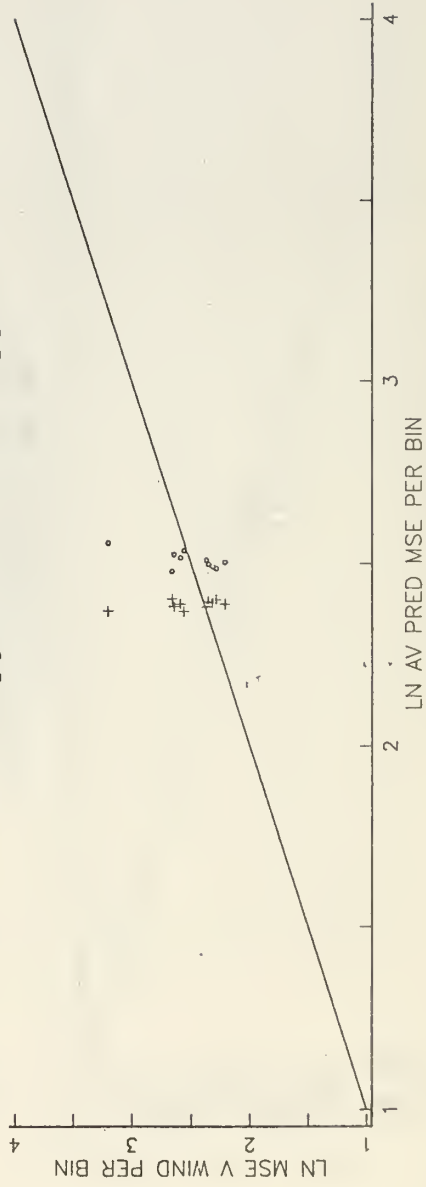


Figure 5D

850 MB V WIND; APR MODEL ON APR DATA; 1ST GUESS WIND

$$1 \text{VAR} = R[T] = \circ; 2 \text{VAR} = +; \text{BIN ON } R[T]$$



$$1 \text{VAR} = WS[T] = \circ; 2 \text{VAR} = +; \text{BIN ON } WS[T]$$

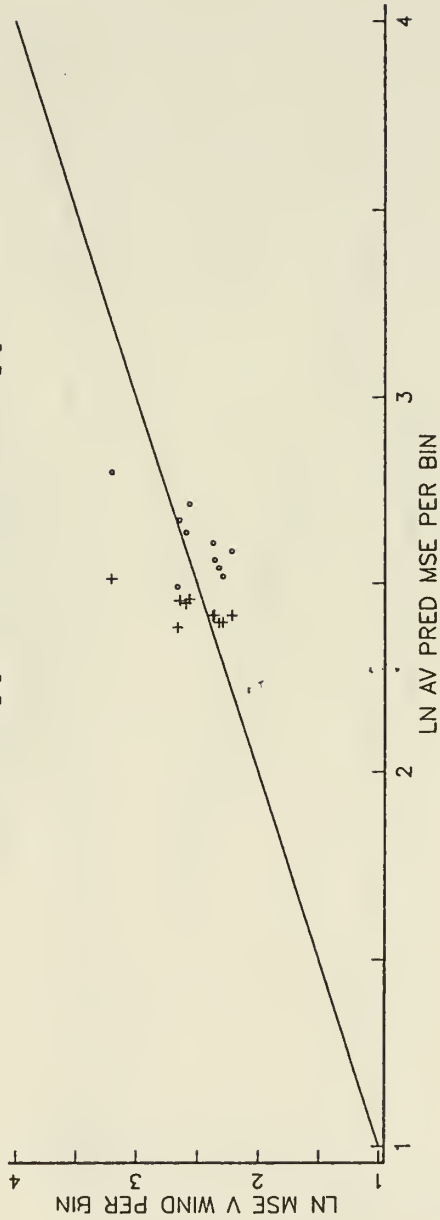
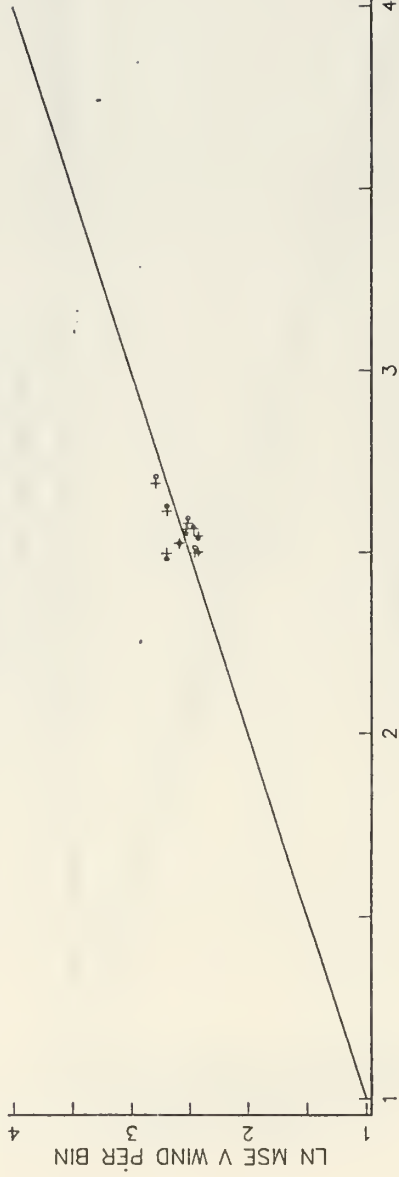


Figure 6D

850 MB V WIND; APR MODEL ON FEB DATA; 1ST GUESS WIND

$$1\text{VAR}=\text{R}[T]=0; 2\text{VAR}=\text{+}; \text{BIN ON R}[T]$$



$$1\text{VAR}=\text{WS}[T]=0; 2\text{VAR}=\text{+}; \text{BIN ON WS}[T]$$

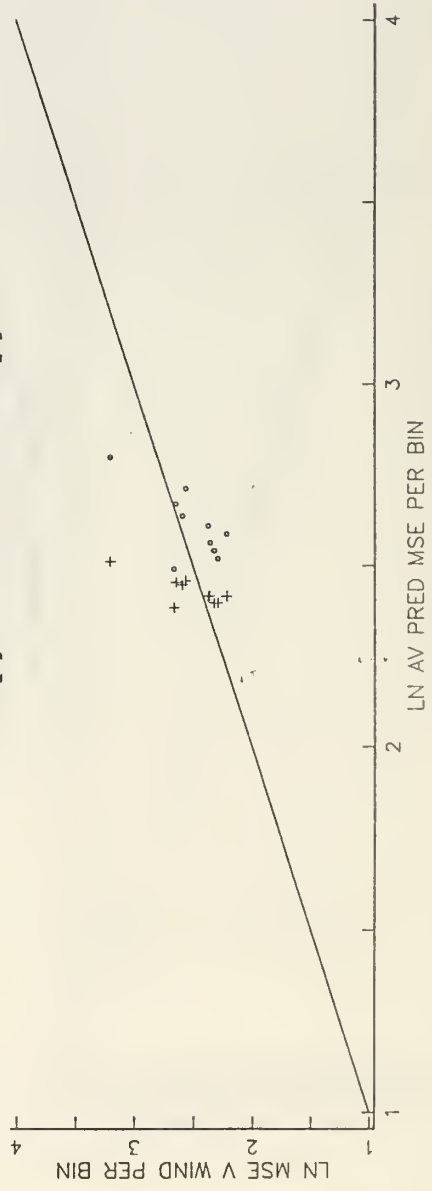


Figure 7D

850 MB V WIND; FEB MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=0; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=0; 2VAR=+; BIN ON WS[T]

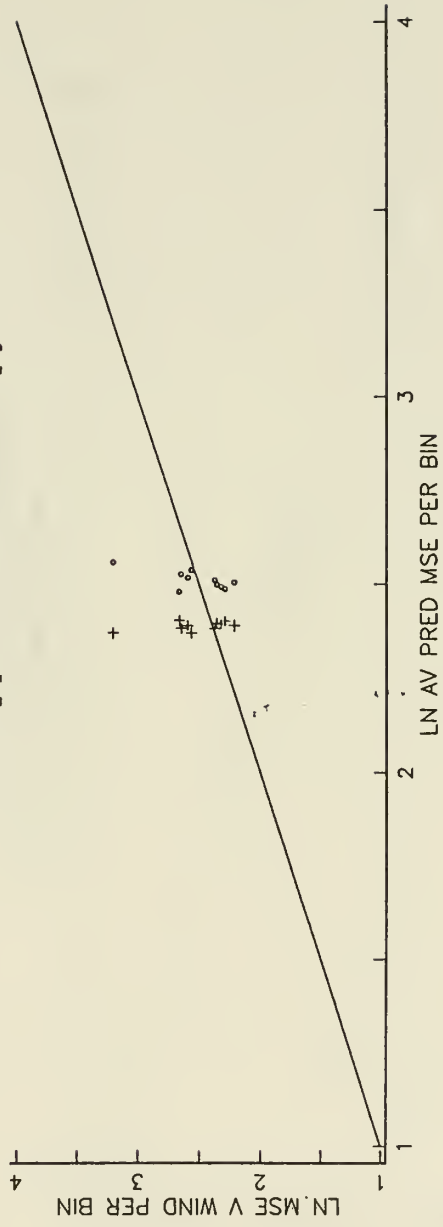


Figure 8D

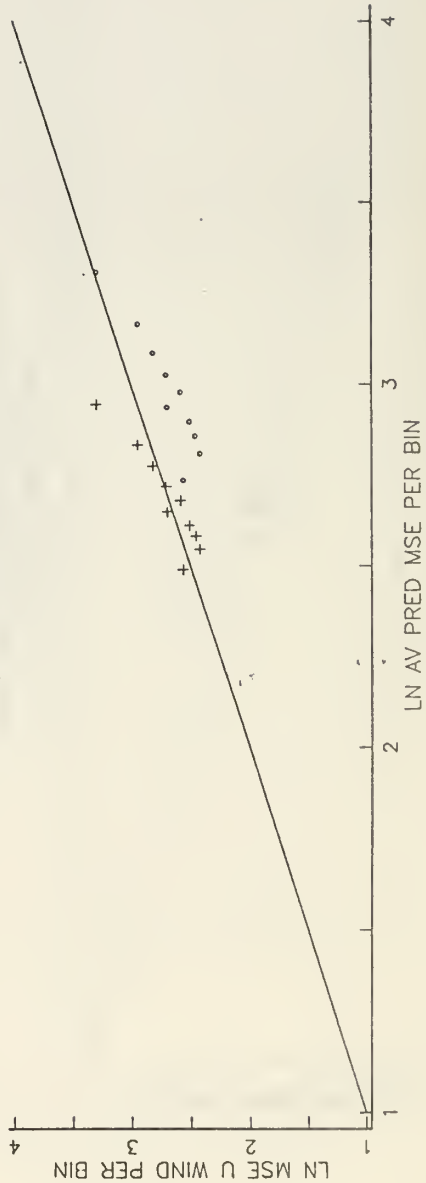
500 MB U WIND; FEB MODEL ON FEB DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

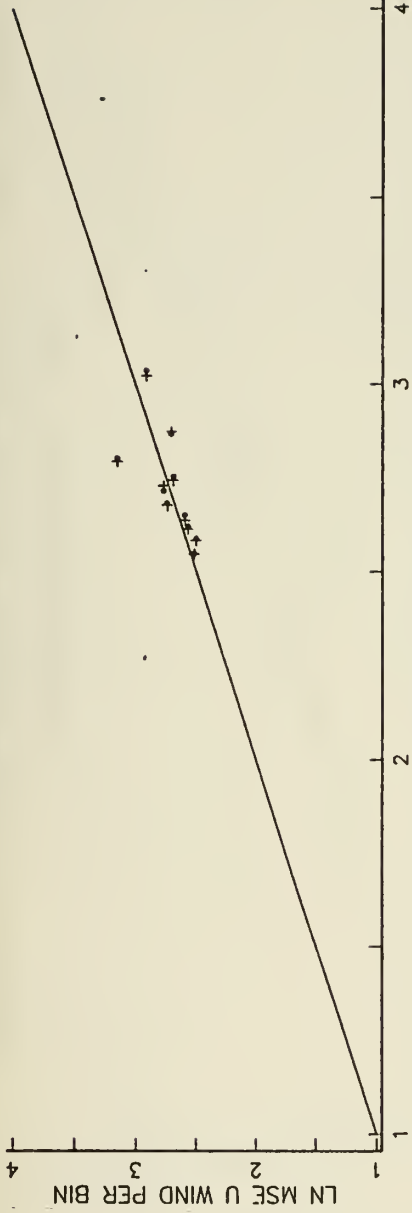


LN AV PRED MSE PER BIN

Figure 9D

500 MB U WIND; APR MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

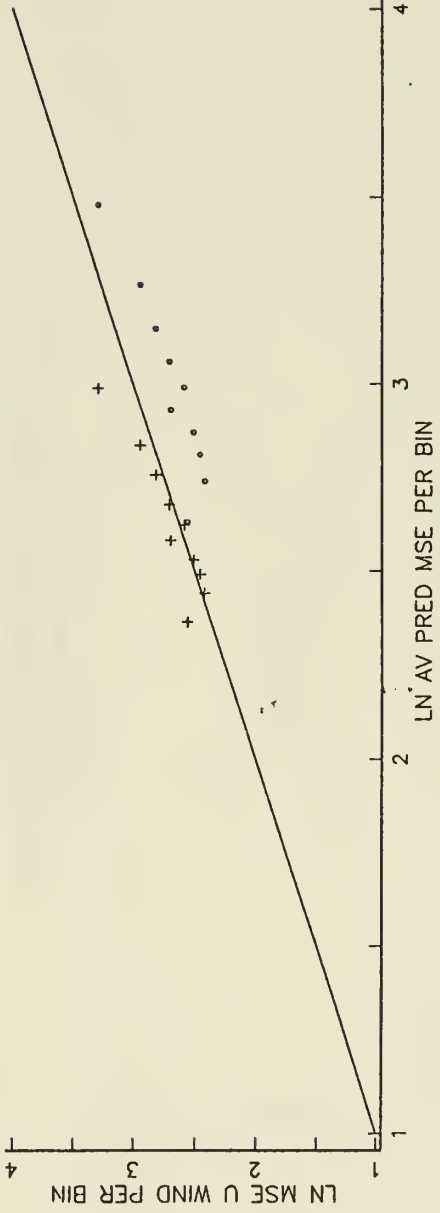
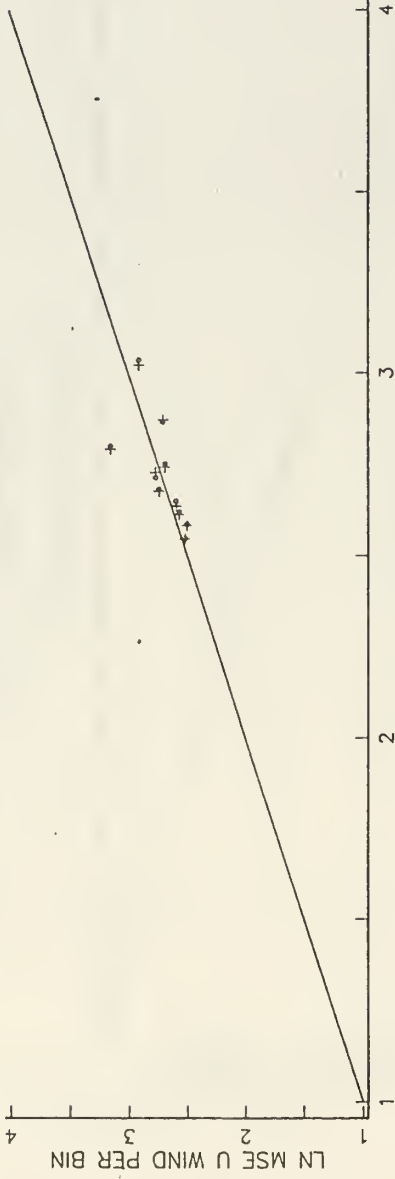


Figure 10D

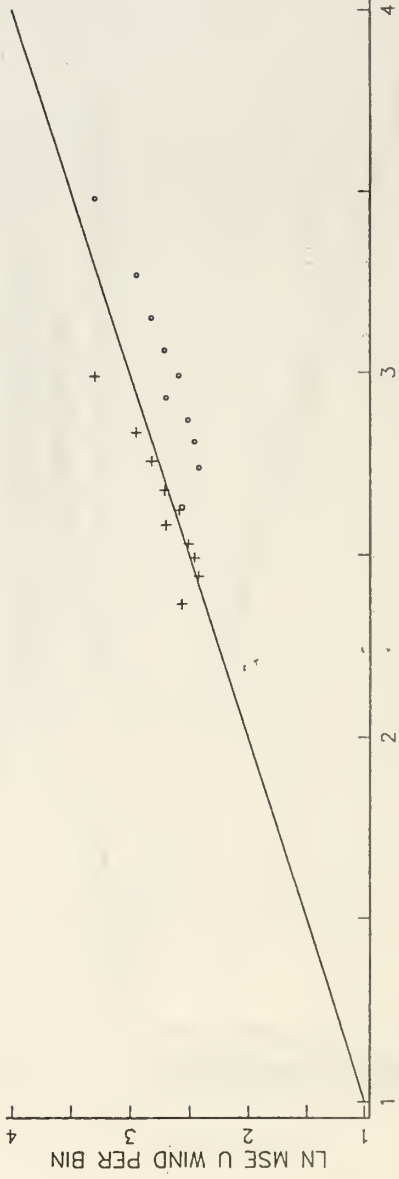
500 MB U WIND; APR MODEL ON FEB DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



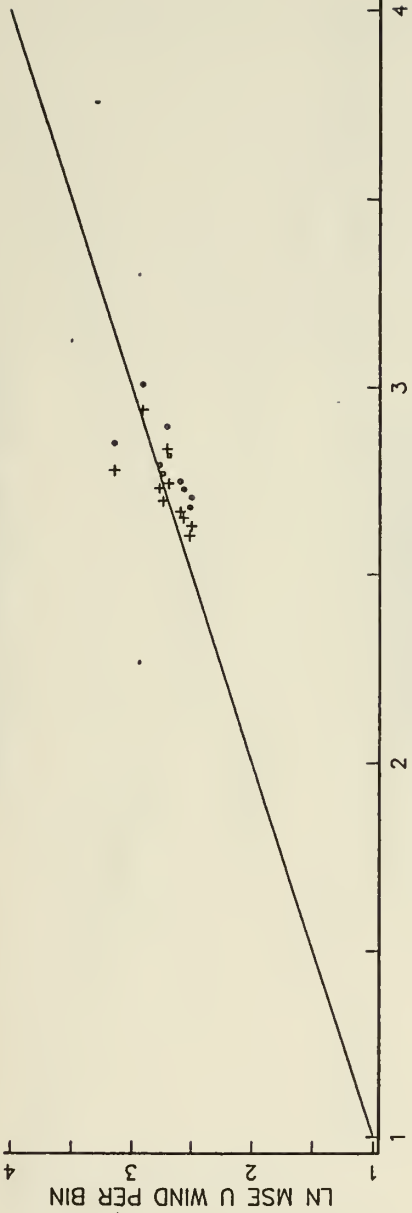
LN AV PRED MSE PER BIN

Figure 11D



500 MB U WIND; FEB MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

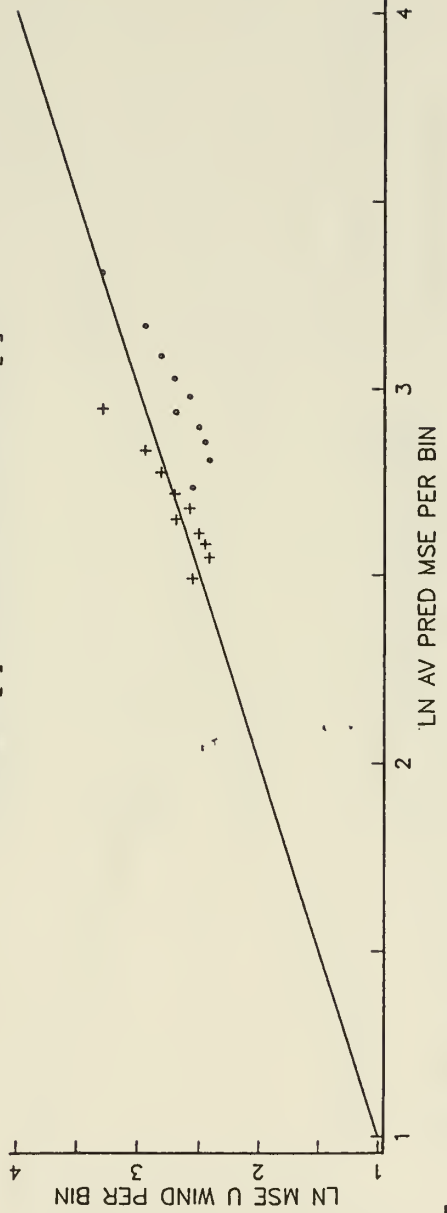


Figure 12D

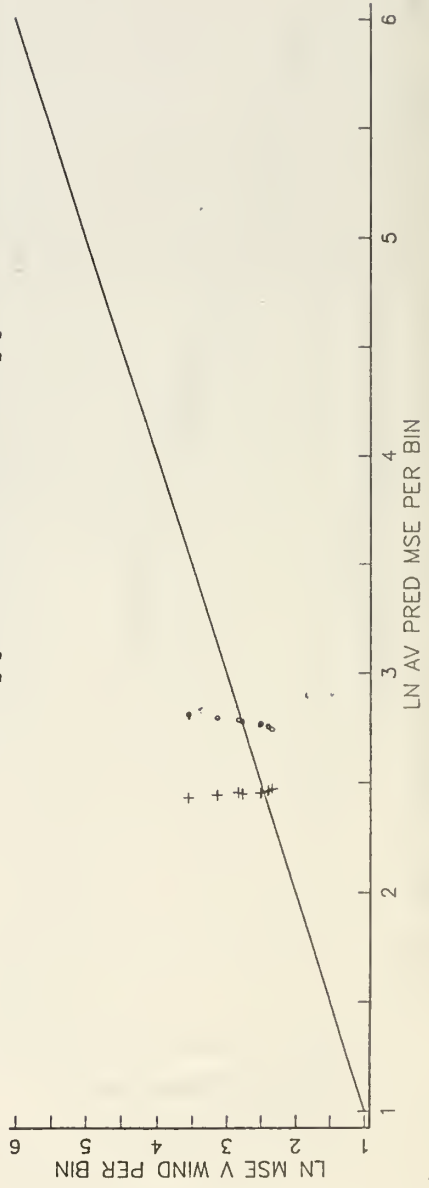
500 MB V WIND; FEB MODEL ON FEB DATA; 1ST GUESS WIND

1VAR=R[T]=°; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=°; 2VAR=+; BIN ON WS[T]

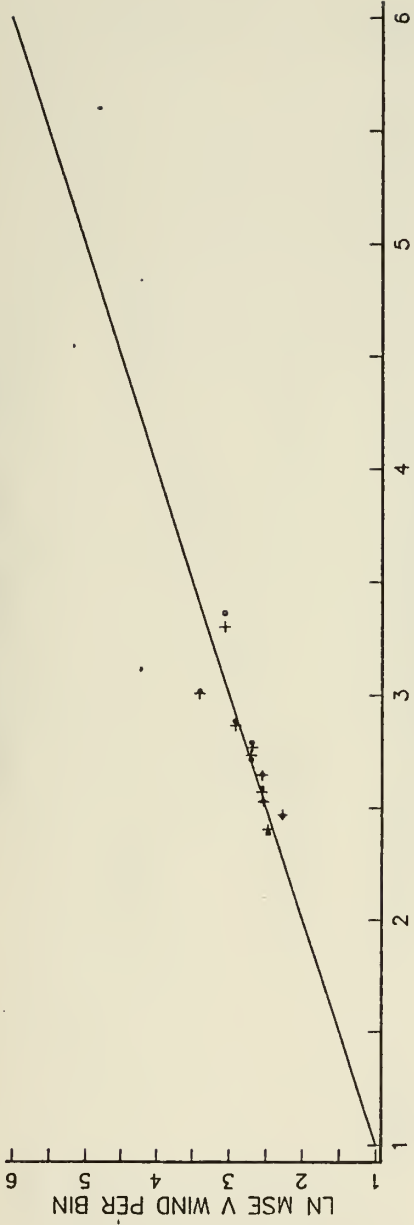


LN AV PRED MSE PER BIN

Figure 13D

500 MB V WIND; APR MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN  
1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

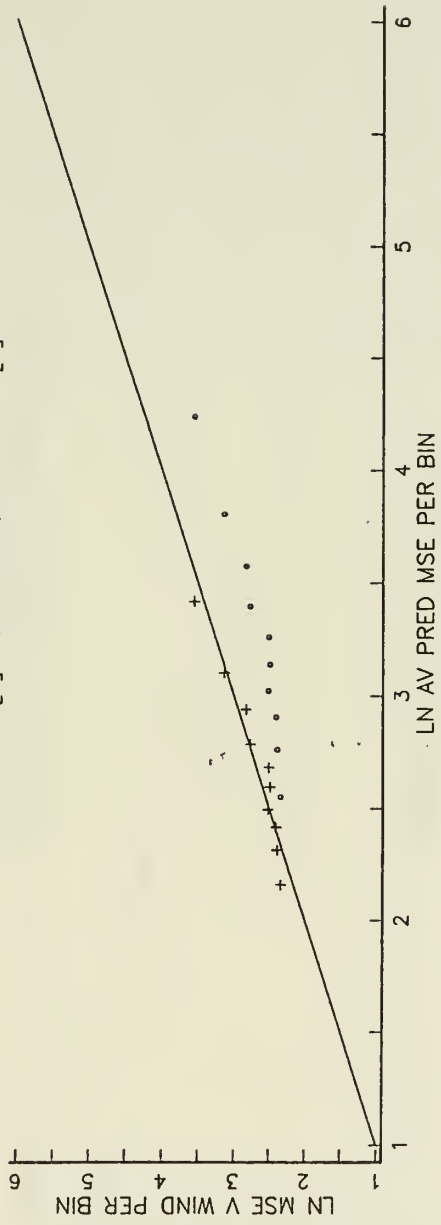
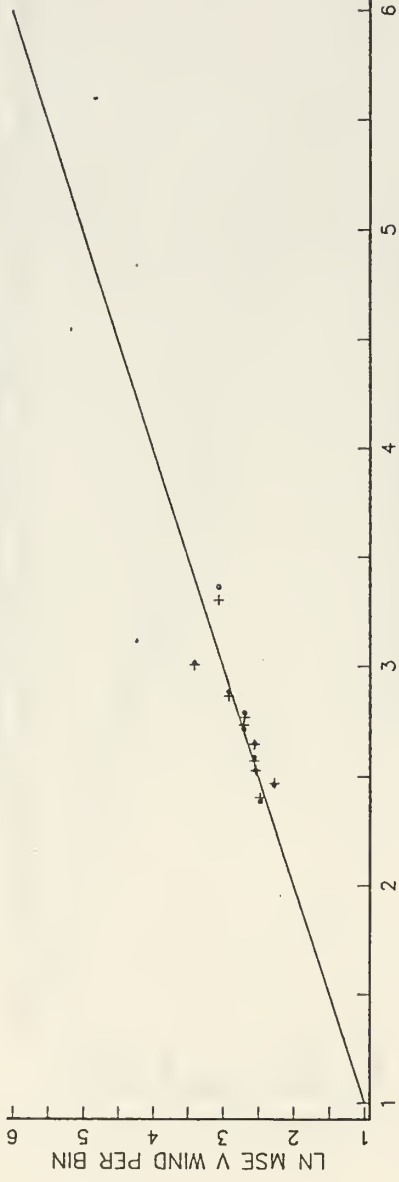


Figure 14D

500 MB V WIND; APR MODEL ON FEB DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

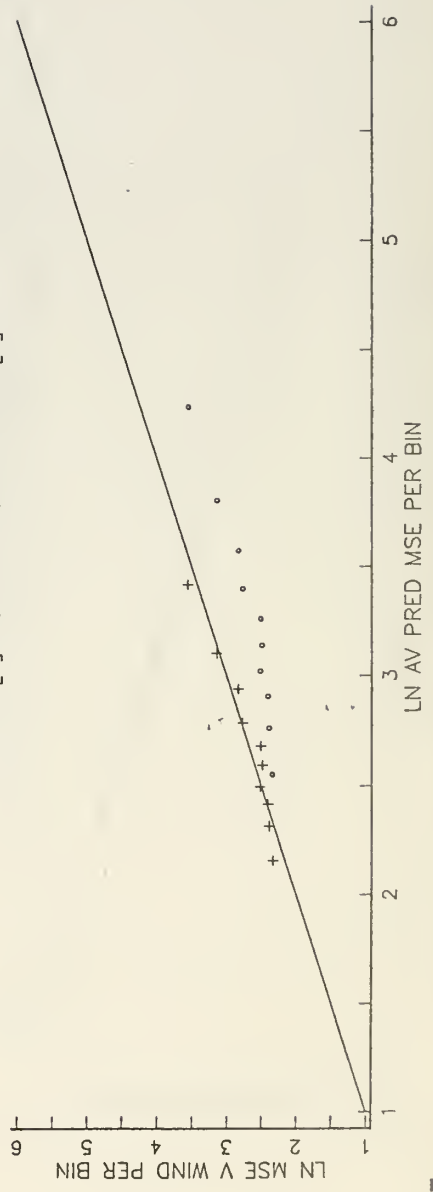


Figure 15D

500 MB V WIND; FEB MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=0; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN  
1VAR=WS[T]=0; 2VAR=+; BIN ON WS[T]

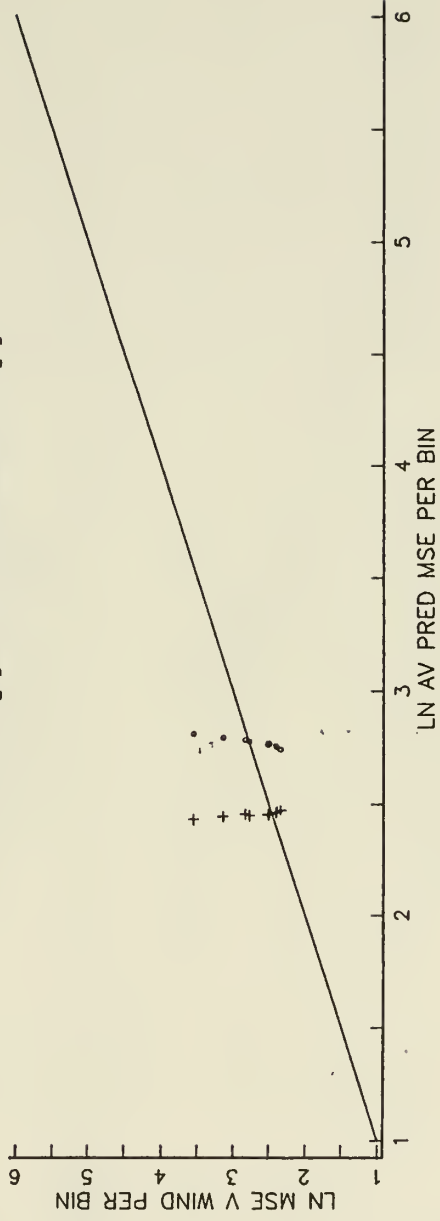


Figure 16D

250 MB U WIND; FEB MODEL ON FEB DATA; 1ST GUESS WIND

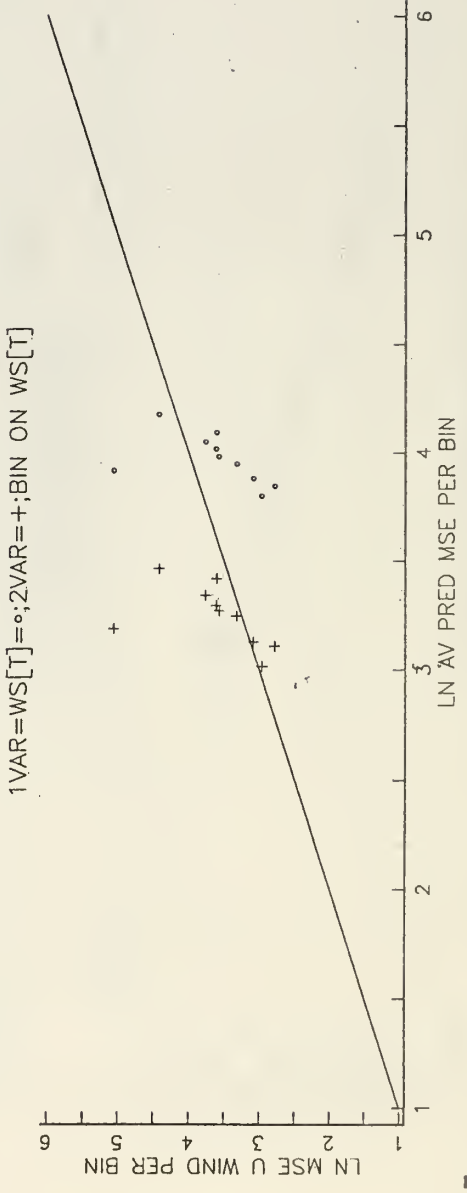
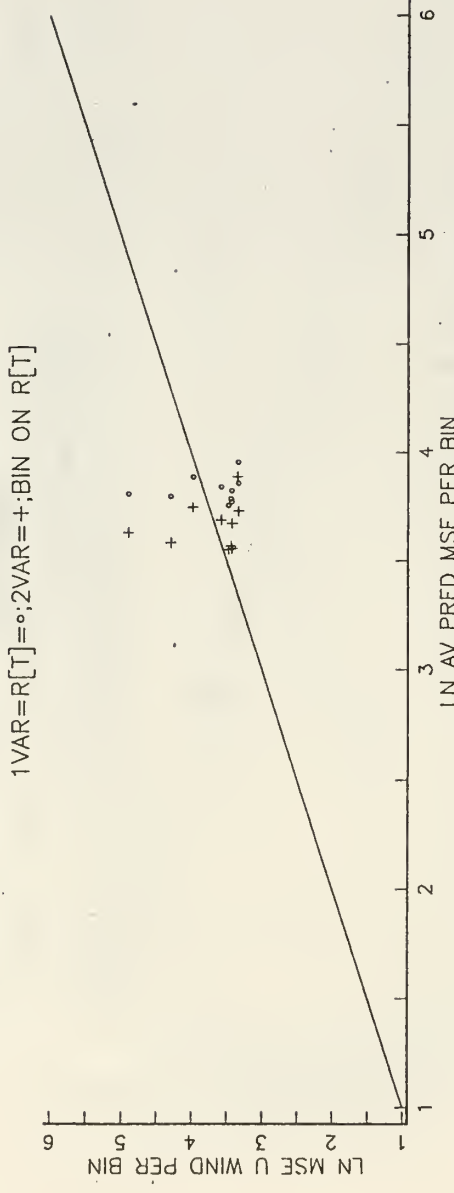


Figure 17D

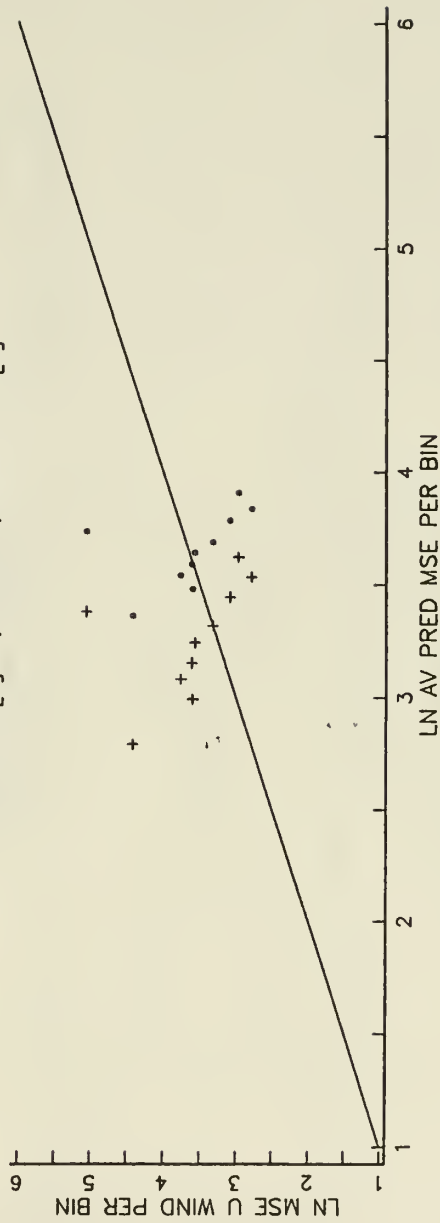
250 MB U WIND; APR MODEL ON APR DATA; 1ST GUESS WIND

$$1\text{VAR} = R[T] = 0; 2\text{VAR} = +; \text{BIN ON } R[T]$$



LN AV PRED MSE PER BIN

$$1\text{VAR} = WS[T] = 0; 2\text{VAR} = +; \text{BIN ON } WS[T]$$



LN AV PRED MSE PER BIN

Figure 18D

250 MB U WIND; APR MODEL ON FEB DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T].



1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T].

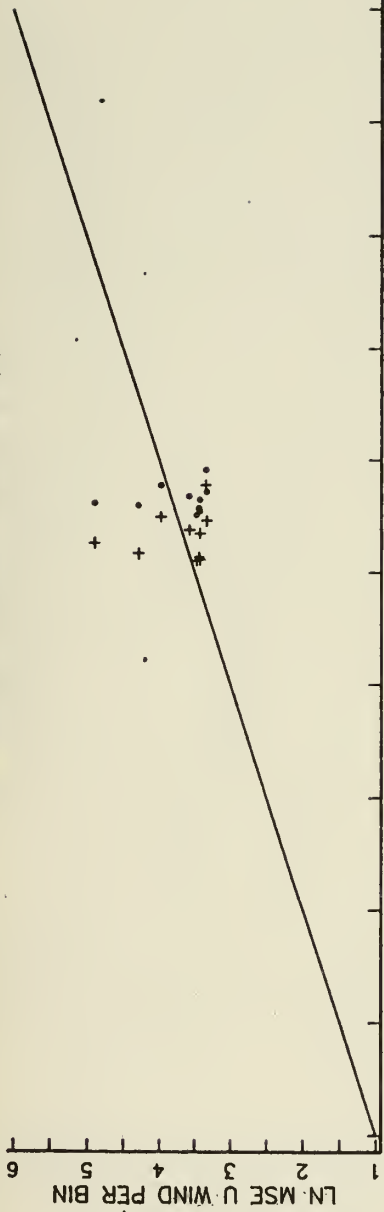


Figure 19D



250 MB U WIND; FEB MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

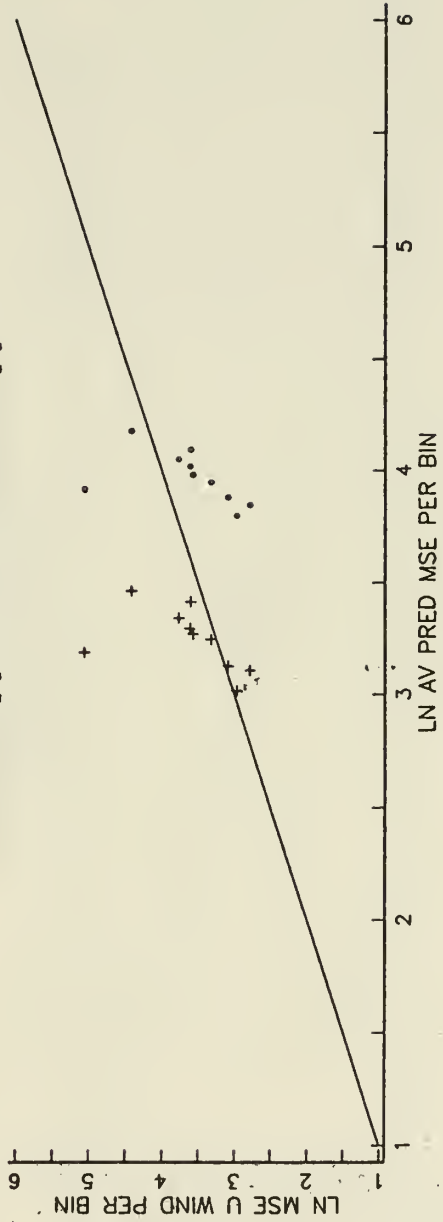
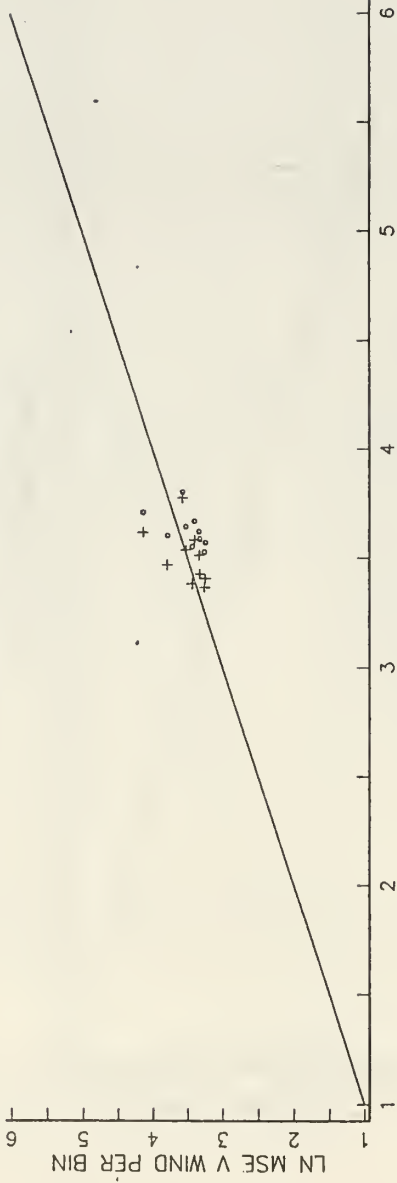


Figure 20D

250 MB V WIND; FEB MODEL ON FEB DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN  
1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]

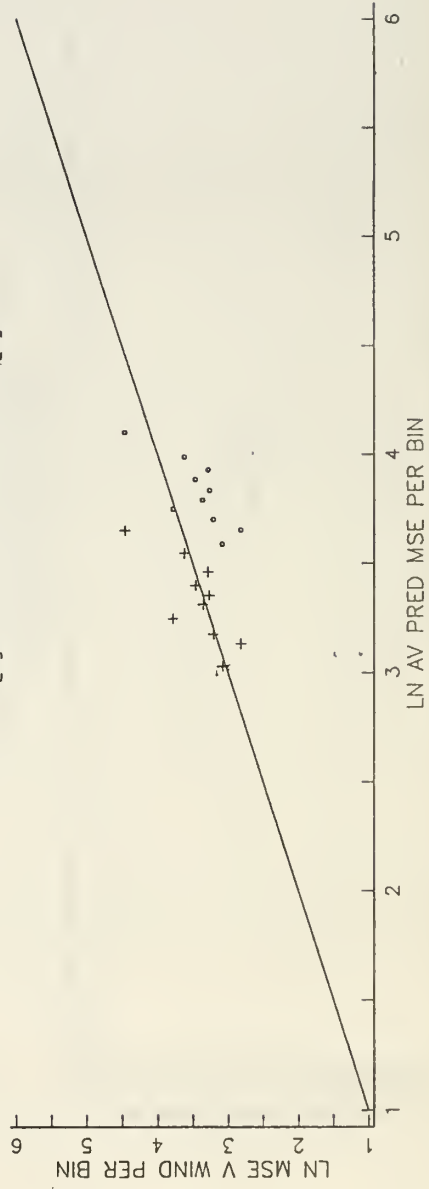


Figure 21D

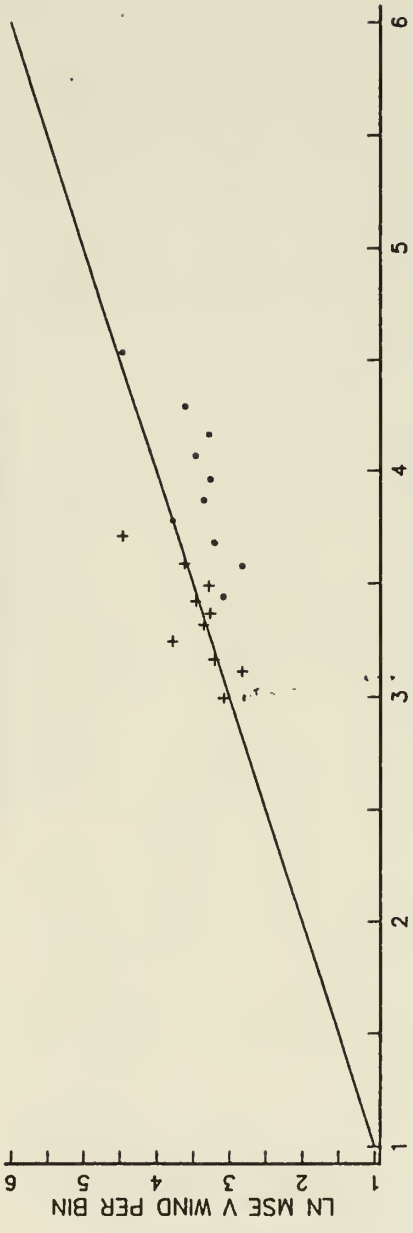
250 MB V WIND; APR MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=0; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=0; 2VAR=+; BIN ON WS[T]

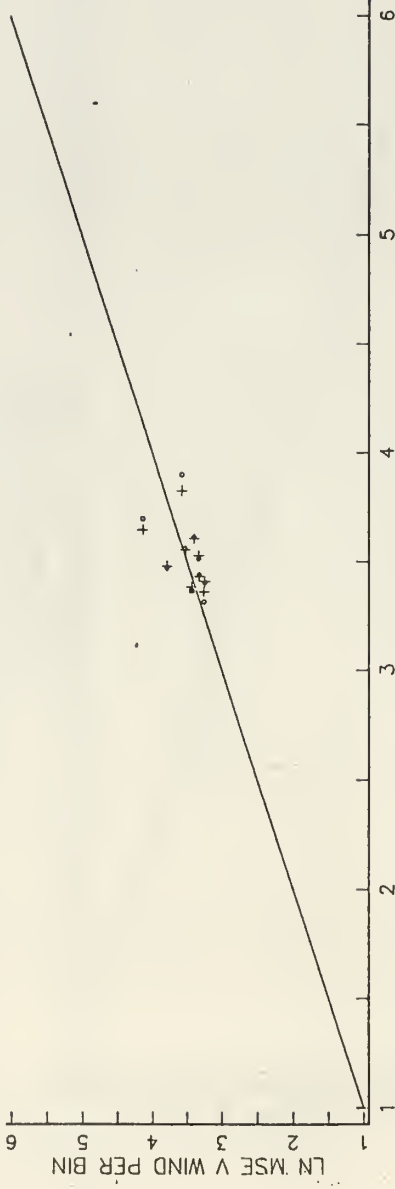


LN AV PRED MSE PER BIN

Figure 22D

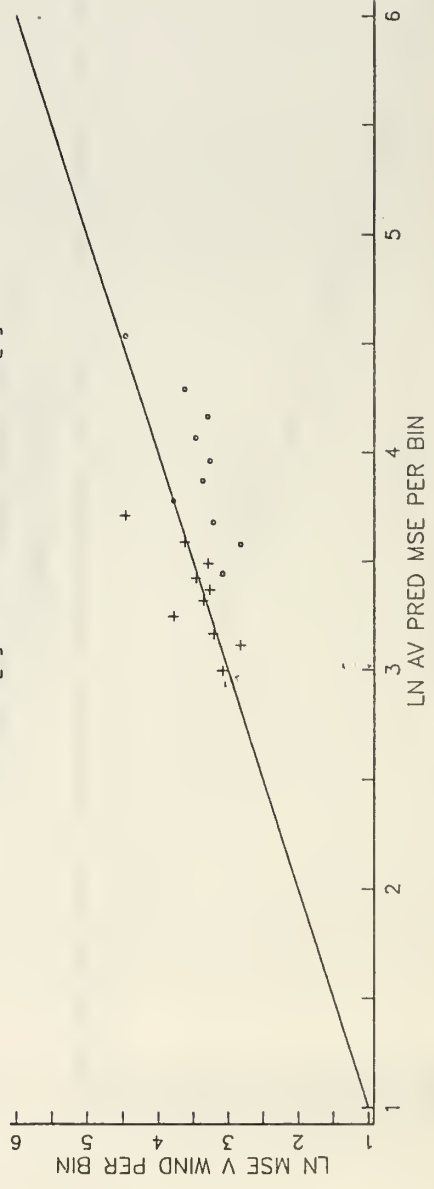
250 MB V WIND; APR MODEL ON FEB DATA; 1ST GUESS WIND

1VAR=R[T]=o; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=o; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 23D

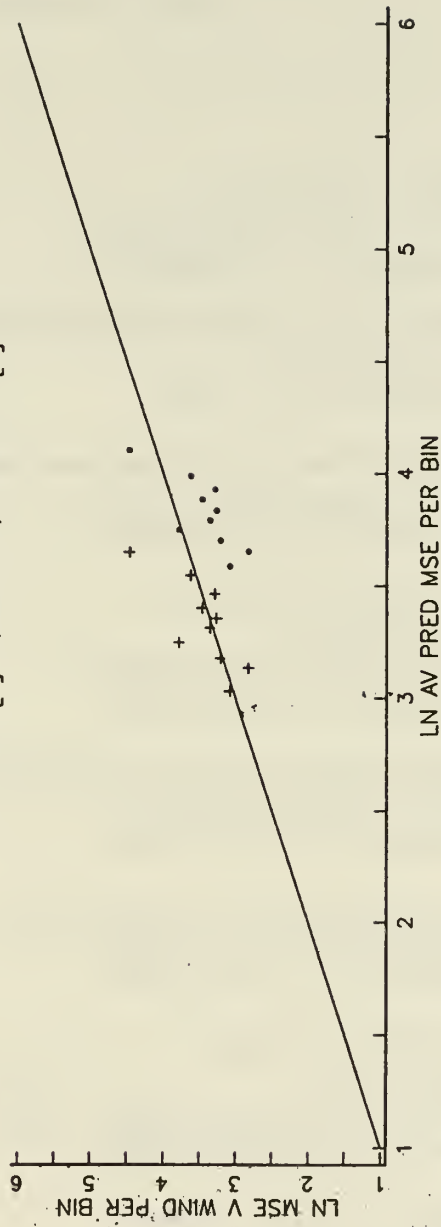
250 MB V. WIND; FEB MODEL ON APR DATA; 1ST GUESS WIND

1VAR=R[T]=0; 2VAR=+; BIN ON R[T]



LN AV PRED MSE PER BIN

1VAR=WS[T]=0; 2VAR=+; BIN ON WS[T]



LN AV PRED MSE PER BIN

Figure 24D

## INITIAL DISTRIBUTION LIST

1. Library (Code 0142).....2  
Naval Postgraduate School  
Monterey, CA 93943-5000
2. Defense Technical Information Center.....2  
Cameron Station  
Alexandria, VA 22314
3. Office of Research Administration (Code 012).....1  
Naval Postgraduate School  
Monterey, CA 93943-5000
4. Prof. Peter Purdue.....1  
Code OR-Pd  
Naval Postgraduate School  
Monterey, CA 93943-5000
5. Department of Operations Research (Code 55).....1  
Naval Postgraduate School  
Monterey, CA 93943-5000
6. Prof. Donald Gaver, Code OR-Gv .....5  
Naval Postgraduate School  
Monterey, CA 93943-5000
7. Prof. Patricia Jacobs.....5  
Code OR/Jc  
Naval Postgraduate School  
Monterey, CA 93943-5000
8. Center for Naval Analyses .....1  
4401 Ford Avenue  
Alexandria, VA 22302-0268
9. Dr. David Brillinger.....1  
Statistics Department  
University of California  
Berkeley, CA 94720

10. Prof. W. R. Schucany .....1  
 Dept. of Statistics  
 Southern Methodist University  
 Dallas, TX 75222
  
11. Prof. D. C. Siegmund .....1  
 Dept. of Statistics  
 Sequoia Hall  
 Stanford University  
 Stanford, CA 94305
  
12. Prof. H. Solomon .....1  
 Department of Statistics  
 Sequoia Hall  
 Stanford University  
 Stanford, CA 94305
  
13. Dr. Ed Wegman .....1  
 George Mason University  
 Fairfax, VA 22030
  
14. Dr. P. Welch .....1  
 IBM Research Laboratory  
 Yorktown Heights, NY 10598
  
15. Dr. Neil Gerr .....1  
 Office of Naval Research  
 Arlington, VA 22217
  
16. Prof. Roy Welsch .....1  
 Sloan School  
 M.I.T.  
 Cambridge, MA 02139
  
17. Dr. J. Abrahams, Code 1111, Room 607 .....1  
 Mathematical Sciences Division, Office of Naval Research  
 800 North Quincy Street  
 Arlington, VA 22217-5000
  
18. Prof. J. R. Thompson .....1  
 Dept. of Mathematical Science  
 Rice University  
 Houston, TX 77001

19. Dr. P. Heidelberger .....1  
 IBM Research Laboratory  
 Yorktown Heights  
 New York, NY 10598
20. Prof. J. B. Kadane.....1  
 Dept. of Statistics  
 Carnegie-Mellon University  
 Pittsburgh, PA 15213
21. Prof. J. Lehoczky .....1  
 Department of Statistics  
 Carnegie-Mellon University  
 Pittsburgh, PA 15213
22. Prof. M. Mazumdar.....1  
 Dept. of Industrial Engineering  
 University of Pittsburgh  
 Pittsburgh, PA 15235
23. Prof. H. Chernoff.....1  
 Department of Statistics  
 Harvard University  
 1 Oxford Street  
 Cambridge, MA 02138
24. Prof. Frank Samaniego.....1  
 Statistics Department  
 University of California  
 Davis, CA 95616
25. Dr. Shlomo Halfin.....1  
 Bellcore,  
 445 South Street  
 Morristown, NJ 07962-1910  
 (MRE 2L309)
26. Prof. Tom A. Louis.....1  
 School of Public Health  
 University of Minnesota  
 Mayo Bldg. A460  
 Minneapolis, MN 55455



27. Dr. Nan Laird.....1  
 Biostatistics Dept.  
 Harvard School of Public Health  
 677 Huntington Ave.  
 Boston, MA 02115
28. Dr. Marvin Zelen.....1  
 Biostatistics Department  
 Harvard School of Public Health  
 677 Huntington Ave.  
 Boston, MA 02115
29. Dr. John Orav .....1  
 Biostatistics Department  
 Harvard School of Public Health  
 677 Huntington Ave.  
 Boston, MA 02115
30. Prof. W. Stuetzle.....1  
 Department of Statistics  
 University of Washington  
 Seattle, WA 98195
31. Prof. F. W. Mosteller.....1  
 Department of Statistics  
 Harvard University  
 1 Oxford St.  
 Cambridge, MA 02138
32. Dr. D. C. Hoaglin .....1  
 Department of Statistics  
 Harvard University  
 1 Oxford Street  
 Cambridge, MA 02138
33. Prof. N. D. Singpurwalla.....1  
 George Washington University  
 Washington, DC 20052
34. Prof. George S. Fishman.....1  
 Curr. in OR & Systems Analysis  
 University of North Carolina  
 Chapel Hill, NC 20742

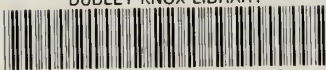
35. Dr. Alan F. Petty.....1  
Code 7930  
Naval Research Laboratory  
Washington, DC 20375
36. Prof. Bradley Efron.....1  
Statistics Dept.  
Sequoia Hall  
Stanford University  
Stanford, CA 94305
37. Prof. Carl N. Morris.....1  
Statistics Department  
Harvard University  
1 Oxford St.  
Cambridge, MA 02138
38. Dr. John E. Rolph .....1  
RAND Corporation  
1700 Main St.  
Santa Monica, CA 90406
39. Prof. D. C. Siegmund.....1  
Dept. of Statistics, Sequoia Hall  
Stanford University  
Stanford, CA 94305
40. Operations Research Center, Rm E40-164.....1  
Massachusetts Institute of Technology  
Attn: R. C. Larson and J. F. Shapiro  
Cambridge, MA 02139
41. Arthur P. Hurter, Jr. ....1  
Professor and Chairman  
Dept. of Industrial Engineering and Management Sciences  
Northwestern University  
Evanston, IL 60201-9990
42. Institute for Defense Analysis .....1  
1800 North Beauregard  
Alexandria, VA 22311
43. Prof. J. W. Tukey .....1  
Statistics Dept., Fine Hall  
Princeton University  
Princeton, NJ 08540

44. Dr. Daniel H. Wagner.....1  
 Station Square One  
 Paoli, PA 19301
45. Dr. Colin Mallows .....1  
 AT&T Bell Telephone Laboratories  
 600 Mountain Avenue  
 Murray Hill, NJ 07974
46. Dr. D. Pregibon .....1  
 AT&T Bell Telephone Laboratories  
 Mountain Avenue  
 Murray Hill, NJ 07974
47. Dr. Jon Kettenring .....1  
 Bellcore  
 445 South Street  
 Morris Township, NJ 07962-1910
48. Prof. David L. Wallace.....1  
 Statistics Dept., University of Chicago  
 5734 S. University Ave.  
 Chicago, IL 60637
49. Dr. S. R. Dalal.....1  
 Bellcore  
 445 South Street  
 Morristown, NJ 07962-1910
50. Dr. D. F. Daley .....1  
 Statistic Dept. (I.A.S.)  
 Australian National University  
 Canberra, A.C.T. 2606  
 AUSTRALIA
51. Koh Peng Kong.....1  
 OA Branch, DSO  
 Ministry of Defense  
 Blk 29 Middlesex Road  
 SINGAPORE 1024
52. Professor Sir David Cox.....1  
 Nuffield College  
 Oxford, OXI INF  
 ENGLAND

53. Dr. A. J. Lawrence .....1  
 Dept. of Mathematics,  
 University of Birmingham  
 P. O. Box 363  
 Birmingham B15 2TT  
 ENGLAND
54. Dr. John Copas.....1  
 Dept. of Mathematics,  
 University of Birmingham  
 P. O. Box 363  
 Birmingham B15 2TT  
 ENGLAND
55. Dr. D. Vere-Jones.....1  
 Dept. of Math, Victoria Univ. of Wellington  
 P. O. Box 196  
 Wellington  
 NEW ZEALAND
56. Dr. J. Goerss.....50  
 Naval Oceanographic and Atmospheric Laboratory  
 Monterey, CA 93943-5006



DUDLEY KNOX LIBRARY



3 2768 00330055 9