

AD-A259 585



NPSOR-93-004

2

J

# NAVAL POSTGRADUATE SCHOOL

Monterey, California



DTIC  
ELECTE  
JAN 29 1993  
S C D

**BAYESIAN PREDICTION OF MEAN SQUARE ERRORS  
WITH COVARIATES**

Donald P. Gaver  
Patricia A. Jacobs

November 1992

93-01614  
3728  
422547

Approved for public release; distribution is unlimited.

Prepared for:  
Naval Research Laboratory-West  
Monterey, CA 93943-5006


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 Research Laboratory-West, Monterey, California.


This report was prepared by:

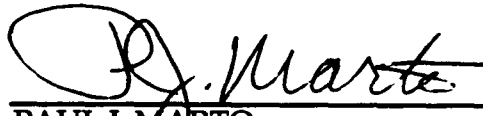
  
DONALD P. GAVER, JR.  
Professor of Operations Research

  
PATRICIA A. JACOBS  
Professor of Operations Research

Reviewed by:

Released by:

  
PETER PURDUE  
Professor and Chairman  
Department of Operations Research

  
PAUL J. MARTO  
Dean of Research

**REPORT DOCUMENTATION PAGE**

1a. REPORT SECURITY CLASSIFICATION <b>UNCLASSIFIED</b>		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPSOR-93-004		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School	6b. OFFICE SYMBOL (If applicable) OR	7a. NAME OF MONITORING ORGANIZATION Naval Research Laboratory-West	
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943		7b. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5006	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Naval Research Laboratory-West	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. 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.
11. TITLE (Include Security Classification) <i>Bayesian Prediction of Mean Square Errors with Covariates</i>			
12. PERSONAL AUTHOR(S) Donald P. Gaver and Patricia A. Jacobs			
13a. TYPE OF REPORT Technical Report	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, month day) November 1992	15. PAGE COUNT 37
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Hierarchical Model, Gaussian model with log-linear scale parameters	
FIELD	GROUP SUB-GROUP		
19. 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. Previous work has suggested that statistical models with log-linear scale parameters which include covariates can be used to predict mean square prediction errors. However, the parameters of the statistical relationships appear to change over time. A procedure is described to recursively update the estimated parameters. Data from July of 1991 are used to fit the model parameters and to study the predictive ability of the recursive procedure. This preliminary investigation indicates that observational and first guess wind components can be helpful in predicting mean square prediction error for wind components.			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION <b>UNCLASSIFIED</b>	
22a. NAME OF RESPONSIBLE INDIVIDUAL D. P. Gaver		22b. TELEPHONE (Include Area Code) (408) 646-2605	2c. OFFICE SYMBOL OR/Gv

DTIC QUALITY INSPECTED 1

Accession For	
NTIS	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	

# BAYESIAN PREDICTION OF MEAN SQUARE ERRORS WITH COVARIATES

by P. A. Jacobs and D. P. Gaver

By	
Distribution/	
Availability Codes	
Dist	Special
A-1	

## 1. INTRODUCTION AND SUMMARY

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 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 forecasted variables become available. 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 mean squared error of the first-guess value and the mean 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 Davidian and Carroll (1987), Nelder and Lee (1992), Aitken (1987), McCullagh and Nelder (1983).

In Jacobs and Gaver (1991, 1992) statistical models for the error of the first-guess are used to predict mean square error for first-guess wind components. The models assume that the error of the first-guess is normal with mean 0

and variance which is a function that is log-linear with suitable covariates. The cross-validation results of those papers suggest that covariates do have some predictive ability for the mean square errors. However, the relations change over time.

In this paper we introduce a procedure for recursively updating the estimated parameters of the variance function. The approach is Bayesian with recursive updating using an approximation based on the Laplace method; cf. deBruijn (1958).

In the next section the model is introduced. Details of the updating procedure are also given.

The third section presents results of using the procedure to predict mean square first-guess wind component errors. The data consist of measurement and 12 hour forecasts (first-guess values) of  $u$  and  $v$  wind components at 850 mb, 500 mb, and 250 mb pressure levels from 93 stations in North America, 25N-75N. The measurement values (if available) are subtracted from the first-guess values to obtain observations of the first guess error. The covariates considered are wind speed and resultant wind, (the sum of the squared difference of the  $u$ -wind component at two consecutive 12 hour periods and the squared difference of the  $v$ -wind component at the same times). The resultant wind is a measure of the change in the atmosphere. Higher wind speeds suggest more activity in the atmosphere.

The results of the data analysis suggest that the covariates do have predictive ability. The models using observed wind speed and resultant wind have more predictive ability than those using the first-guess values of wind speed and resultant wind. Further, models that use both wind speed and

resultant wind have more predictive ability than those using either one by itself. The change of the model parameters with time appears to be slow. This suggests that while the relationship of the mean square error and the resultant wind and wind speed is changing, it may not be necessary to update the model parameters in every period.

## 2. THE MODEL AND STATISTICAL PROCEDURE

Let  $Y_i(t)$  denote the first-guess error at location  $i$  at time  $t$ ;  $i = 1, \dots, L$ . Let  $x_i(t) = (x_{ij}(t); j = 1, \dots, p)$  denote the covariates at location  $i$  at time  $t$ .

Consider the following model for the first-guess errors.

$$P\{Y_i(t) \in dy_i(t); i = 1, \dots, L | x_i(t), \beta(t), y_j(s), x_j(s), s < t, j = 1, \dots, L\} \quad (2.1)$$

$$= \prod_{i=1}^L \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} x_i(t) \beta(t)\right\} \exp\left\{-\frac{1}{2} y_i(t)^2 \exp\{-x_i(t) \beta(t)\}\right\} dy_i(t)$$

where  $x_i(t) \beta(t) = \sum_{j=1}^p x_{ij}(t) \beta_j(t)$ ; that is, given  $\beta(t)$ , the first-guess errors are conditionally independent normally distributed with mean 0 and log linear variances

$$\sigma_i^2(t) = \exp\{x_i(t) \beta(t)\} \quad (2.2)$$

independent of everything else.

The coefficients  $\beta(t)$  are modeled as changing according to a random walk

$$\beta(t+1) = \beta(t) + \omega(t+1) \quad (2.3)$$

where  $\{\omega(t)\}$  are independent multivariate normal random variables with variance-covariance matrices  $\{W(t)\}$ . The matrix  $W(t+1)$  is independent of  $\{Y_i(s), x_i(s), \beta(s), s \leq t, i = 1, \dots, L\}$

In the next subsection we suggest a Kalman filter-like procedure to produce successive estimates of  $\beta(t)$  as new data become available. The procedure is based on a Laplace approximation to an integral.

### **2.1 An approximate Updating Procedure**

Assume the posterior distribution of  $\beta(t)$  given  $\{y_i(s), i = 1, \dots, L, s \leq t\}$  is multivariate normal with mean  $\mathbf{m}(t)$  and variance-covariance matrix  $\Sigma(t)$ .

Since it is known that

$$\beta(t+1) = \beta(t) + \omega(t+1),$$

the prior distribution of  $\beta(t+1)$  is multivariate normal with mean  $\mathbf{m}(t)$  and variance-covariance matrix

$$\mathbf{R}(t) = \Sigma(t) + \mathbf{W}(t+1). \quad (2.4)$$

A description of the procedure used to determine  $\mathbf{W}(t+1)$  appears in Section 3.

The forecast/prediction distribution of  $\{Y_i(t+1); i = 1, \dots, L\}$  in terms of data up to time  $t$  and covariate values at time  $t+1$  is

$$\begin{aligned} & P\{Y_i(t+1) \in dy_i; i = 1, \dots, L | Y_i(s), x_i(u), i = 1, \dots, L, s \leq t, u \leq t+1\} \\ &= \int \prod_{i=1}^L \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} x_i(t+1)\mathbf{b}\right\} \exp\left\{-\frac{1}{2} y_i^2 \exp\{-x_i(t+1)\mathbf{b}\}\right\} f_{\beta(t+1)}(\mathbf{b}) d\mathbf{b} dy_i \end{aligned} \quad (2.5)$$

where

$$f_{\beta(t+1)}(\mathbf{b}) = \left((2\pi)^p \det R(t)\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{b} - \mathbf{m}(t))R^{-1}(t)(\mathbf{b} - \mathbf{m}(t))'\right\}. \quad (2.6)$$

We now approximate the integral by the Laplace method; cf. Easton (1991), Cox and Hinkley (1974), de Bruijn (1958). Let the exponent of the integrand be

$$g(\mathbf{b}) = -\frac{1}{2} \left[ \sum_{i=1}^L (\mathbf{x}_i \mathbf{b}) + y_i^2 \exp\{-\mathbf{x}_i \mathbf{b}\} + (\mathbf{b} - \mathbf{m}(t)) \mathbf{R}^{-1}(t) (\mathbf{b} - \mathbf{m}(t))' \right] + K \quad (2.7)$$

where  $K$  is a constant and we let  $\mathbf{x}_i = \mathbf{x}_i(t+1)$ . Differentiating, we obtain

$$\frac{\partial}{\partial b_j} g(\mathbf{b}) = -\frac{1}{2} \left[ \sum_{i=1}^L x_{ij} [1 - y_i^2 \exp\{-\mathbf{x}_i \mathbf{b}\}] + 2\mathbf{R}_{j \cdot}^{-1}(t) (\mathbf{b} - \mathbf{m}(t)) \right] \quad (2.8)$$

where  $\mathbf{R}_{j \cdot}^{-1}(t)$  denotes the  $j$ th row of  $\mathbf{R}^{-1}(t)$  and

$$\frac{\partial^2}{\partial b_j \partial b_k} g(\mathbf{b}) = -\frac{1}{2} \left[ \sum_{i=1}^L y_i^2 \exp\{-\mathbf{x}_i \mathbf{b}\} x_{ij} x_{ik} + 2\mathbf{R}_{jk}^{-1} \right]. \quad (2.9)$$

Use a Newton procedure to solve the system of equations

$$0 = \frac{\partial}{\partial b_j} g(\mathbf{b}) \quad j = 1, \dots, p. \quad (2.10)$$

for  $\mathbf{m}(t+1)$ . Solve for  $\Sigma(t+1)$  using the second derivatives of  $g$  evaluated at  $\mathbf{m}(t+1)$ ; that is  $\Sigma(t+1)$  is minus the inverse of the matrix whose  $(j, k)$  entry is

$$\frac{\partial^2}{\partial b_j \partial b_k} g(\mathbf{m}(t+1)). \quad (2.11)$$

The posterior distribution of  $\beta(t+1)$  given  $\{Y_i(s), \mathbf{x}_i(s), i = 1, \dots, L, s = 1, \dots, t+1\}$  is approximated by a multivariate normal distribution with mean  $\mathbf{m}(t+1)$  and variance  $\Sigma(t+1)$ . The estimate of  $\beta(t+1)$  is  $\mathbf{m}(t+1)$ .

The predicted mean square error for the  $i$ th location at time  $t+2$  is

$$\exp\left\{\sum_{j=1}^p x_{ij}(t+2)m_j(t+1)\right\}.$$

### 3. DATA ANALYSIS

In this subsection we report results concerning using regression-like models for the mean square error of the first-guess with recursively updated parameter estimates to predict future mean-square errors of the first-guess.

The data consist of measurement and 12 hour forecasts (first-guess values) of  $u$  and  $v$  wind components at the 850 mb, 500 mb and 250 mb pressure levels from 93 stations in North America 25N–75N for the month of July 1991. The forecasts are produced using the NOGAPS Spectral Forecast Model; cf. Hogan et al., (1991). Each station has measurement and first-guess values for every 12 hours; there are some missing observations and suspicious values of wind components equal to 0. These missing and questionable values are deleted from the data set. The measurement values (if available) are subtracted from the first-guess values to obtain observations of the error of the first-guess value.

Let  $U(0;t)$ , (respectively  $V(0;t)$ ), be the observed  $u$ -wind, (respectively  $v$ -wind) component at time  $t$ . Let  $U(f;t)$ , (respectively  $V(f;t)$ ), be the first-guess  $u$ -wind (respectively  $v$ -wind) component at time  $t$ ;  $U(f;t)$  is the forecasted value of  $U(t)$  made 12 hours previously. The first-guess error for the  $u$ -wind component is

$$Y(t) = U(f;t) - U(0;t). \quad (3.1)$$

The following covariates are considered in the log-linear model for the mean square error of the first-guess.

$$r(0;t) = \left[ (U(0;t) - U(0;t-1))^2 + (V(0;t) - V(0;t-1))^2 \right]^{\frac{1}{2}} \quad (3.2)$$

$$s(0;t) = \left[ U(0;t)^2 + V(0;t)^2 \right]^{\frac{1}{2}} \quad (3.3)$$

$$r(f;t) = \left[ (U(f;t) - U(f;t-1))^2 + (V(f;t) - V(f;t-1))^2 \right]^{\frac{1}{2}} \quad (3.4)$$

$$s(f;t) = \left[ U(f;t)^2 + V(f;t)^2 \right]^{\frac{1}{2}}. \quad (3.5)$$

The resultant wind  $r(0;t)$ , (respectively  $r(f;t)$ ), is a measure of the observed (respectively forecasted), change in the wind. The variable  $s(0;t)$ , (respectively  $s(f;t)$ ), is the observed, (respectively forecasted), wind speed. Higher wind speeds suggest more activity in the atmosphere.

### 3.1 The Models

The following models for the mean square error are considered

#### *One Variable Models: Observed Covariates*

1. Given  $\beta_0(t)$ ,  $\beta_1(t)$ , the first-guess errors at each location  $\{Y_i(t); i = 1, \dots, L\}$  are independent normally distributed with mean 0. The variance of  $Y_i(t)$  is the following function of the observed resultant wind at location  $i$  at time  $t$ ,  $r_i(0;t)$

$$\sigma_1^2(1;t;r_i(0;t)) = \exp\{\beta_0(t) + \beta_1(t)r_i(0;t)\} \quad (3.6)$$

where  $r_i(0;t)$  is the observed resultant wind at location  $i$  at time  $t$  and  $Y_i(t)$  is the first-guess error at location  $i$  at time  $t$ .

2. Given  $\beta_0(t)$ ,  $\beta_1(t)$ ,  $\{Y_i(t); i = 1, \dots, L\}$  are independent normally distributed with mean 0. The variance of  $Y_i(t)$  is the following function of the observed wind speed at location  $i$  at time  $t$ ,  $s_i(0;t)$

$$\sigma_2^2(1;t;s_i(0;t)) = \exp\{\beta_0(t) + \beta_1(t)s_i(0;t)\} \quad (3.7)$$

where  $s_i(0;t)$  is the observed wind speed at location  $i$  at time  $t$ .

### Two-variable Model, Observed Covariates

3. Given  $\beta_0(t)$ ,  $\beta_1(t)$ ,  $\beta_2(t)$ ,  $\{Y_i(t); i = 1, \dots, L\}$  are independent normally distributed with mean 0. The variance of  $Y_i(t)$  is the following function of both the resultant wind and wind speed at location  $i$  at time  $t$

$$\sigma_2^2(t; r_i(0;t); s_i(0;t)) = \exp\{\beta_0(t) + \beta_1(t)r_i(0;t) + \beta_2(t)s_i(0;t)\}. \quad (3.8)$$

Similar one-variable and two-variable models but using first-guess values of the covariates are also considered. In all cases the first-guess error and the covariates are all evaluated at the same pressure level.

The regression parameters  $\beta(t)$  are assumed to evolve according to the random walk given by (2.3)

### 3.2 The Data Analysis

The results of Jacobs and Gaver (1992) suggest that of the models using observed values for covariates, the models for the 850 mb pressure level have the most predictive value. It is also suggested that of the models using first-guess values for covariates, the models for the 250 mb pressure level have the most predictive ability. As a result in what follows we will restrict our attention to these two cases.

#### a. Estimation and Prediction of 850 mb First-guess Mean Square Errors using Observed Wind Covariates.

The estimation procedure described in Section 2 was used to recursively estimate the regression parameters  $\beta(t)$  for each of models (3.6) – (3.8) for 850 mb first-guess errors using observed 850 mb wind covariates. The initializing estimates of  $\beta(t)$  are the estimates obtained using all April data recorded in

Jacobs and Gaver (1992); the initial variance-covariance matrix is the identity matrix. The estimates from April are used since April appeared to have more predictive ability for July than February, cf. Jacobs and Gaver (1992).

The variance-covariance matrix of the innovation  $W_t$  of the random walk (2.3) was taken to be a constant times the identity matrix. Preliminary explorations based on values of the predictive log-likelihood using different values of the constant suggest that for purposes of prediction, the constant should be very small. In what follows we set the constant equal to 0.

Figure 1, (respectively Figure 2) presents plots of the estimates of the slopes, e.g.,  $\beta_1(t)$  and  $\beta_2(t)$ , as a function of time which is labeled 1, 2, .... for  $u$ -wind (respectively  $v$ -wind) component error.

Figure 1 presents the values of estimates of the parameter multiplying  $r_i(0;t)$  (respectively  $s_i(0;t)$ ) for the one parameter models (3.6) and (3.7) in the upper graph. The lower graph presents the values of the estimates of the parameter multiplying  $r_i(0;t)$ , (respectively  $s_i(0;t)$ ) for the two-variate model. There appears to be a slight trend in the estimates.

Figure 2 presents the values of the estimate of the parameters multiplying  $r_i(0;t)$ , (respectively  $s_i(0;t)$ ) as  $o$ , (respectively  $+$ ), for the respective one-variate models. The values of the estimates of the parameters multiplying  $r_i(0;t)$ , (respectively  $s_i(0;t)$ ) for the two-variate model are presented as  $\times$  (respectively  $\nabla$ ) for each time  $t$ . These graphs suggest more evidence of a trend in the estimates. Note that the estimates of the slopes are positive. Hence increased values of the resultant wind  $r_i(0;t)$  and/or wind speed  $s_i(0;t)$  are associated with increased variance of the first-guess value. This is plausible physically,

since a large value of  $r_i(0;t)$  is indicative of a change in the atmosphere and a large value of  $s_i(0;t)$  is indicative of greater activity in the atmosphere.

To assess the predictive ability of the models, the models with parameters estimated at time  $t$  are used to forecast the variances of the first-guess errors at time  $t+1$ .

One procedure to informally assess the predictive ability of the models is by binning the data. To assess models (3.6) and (3.8) the data ( $y_i(t)$ ,  $r_i(0;t)$ ,  $s_i(0;t)$ ) are binned into 10 bins based on ordering the values of  $r_i(0;t)$  for all time  $t$  from smallest to largest. The data in the first bin correspond to the smallest values of  $r_i(0;t)$ ; the data in the 10th bin correspond to the largest values of  $r_i(0;t)$ . Each bin contains about 1/10th of the data with the 10th bin containing a few more data. The averages of the predictive variances for models (3.6) and (3.8) are computed for each bin. The average  $y_i(t)^2$  is also computed for each bin.

To assess models (3.7) and (3.8) the same procedure is used but the binning is based on the values of  $s_i(0;t)$ .

Figures 3 and 4 present graphs of the  $\log[\text{average } y_i(t)^2]$  in each bin versus  $\log[\text{average predictive variance}]$  in each bin for models (3.6) and (3.8) and models (3.7) and (3.8). If a model were perfect, the points should be close to the 45° line shown.

The figures suggest that of the two one-variate models, the one using the resultant wind  $r_i(0;t)$  has the better predictive ability. The two-variate model appears to have similar predictive ability to the model using only  $r_i(0;t)$ .

Another procedure to assess predictive ability is to compute the log-likelihood using estimated values of  $\beta(t)$  (eg.  $m(t)$ ) in the term for the first

guess errors at time  $t+1$ . Larger values of the (predictive) log-likelihood indicate better predictive ability. The (predictive) log-likelihood up to addition and multiplication by constants is

$$\bar{l} = -\sum_{t=1}^{T-1} \left[ \sum_{i=1}^L x_i(t+1)\mathbf{m}(t) + \sum_{i=1}^L y_i^2(t+1)\exp\{-x_i(t+1)\mathbf{m}(t)\} \right]. \quad (3.9)$$

Table 1 presents values of  $\bar{l}$  for the one-time step ahead predictions of the variance; these values appear in the column *Iterative*. Also displayed are the values of  $\bar{l}$  obtained by estimating the parameters once using all the data; this value of  $\bar{l}$  is a goodness-of-fit value and appears in the column labeled *All*; the estimates used to obtain the goodness-of-fit value of  $\bar{l}$  are those appearing for July in Table 7 of Jacobs and Gaver (1992).

Four models are considered: constant variance (no dependence on variables), two one-variate models (3.6) and (3.7) and the two-variate model (3.8). The parameter of the constant variance model using all the data is estimated using maximum likelihood; this estimate is used to calculate the goodness-of-fit value of  $\bar{l}$  for the constant variance model.

The goodness-of-fit constant variance value of  $\bar{l}$  is smaller than the prediction values of  $\bar{l}$  using models with covariates. This behavior indicates that the covariates do have predictive ability. The prediction value of  $\bar{l}$  for the two-variate prediction model is larger than that for either one-variate model indicating that both covariates have some predictive ability. Of the two one-variate models, the one using  $r_i(0;t)$  has the larger prediction value of  $\bar{l}$ . The closeness of the prediction values of  $\bar{l}$  obtained by iteratively updating the estimates and using them to predict variance of the next time period and the goodness-of-fit values of  $\bar{l}$  obtained by using model

parameters estimated from all the data suggest that the updating procedure is doing very well. Note that for the one-variate model using  $r_i(0;t)$ , the prediction values of  $\bar{l}$  are larger than the goodness-of-fit values of  $\bar{l}$ ; this suggests that there is systematic change in the values of  $\beta$  over time for this model.

**TABLE 1**  
**VALUE OF LOG-LIKELIHOOD OBSERVED COVARIATES, 850 mb**

Data	Model	Iterative (Prediction)	All (Goodness-of-Fit)
u-wind	Constant Variance	-9610.7	-9510.0
	One-variate		
	$r(0;t)$	-8544.1	-9098.3
	$s(0;t)$	-8607.9	-8596.2
	Two-variate	-8489.2	-8467.5
v-wind	Constant Variance	-9692.7	-9620.2
	One-variate		
	$r(0;t)$	-8730.9	-9294.1
	$s(0;t)$	-8769.2	-8756.5
	Two-variate	-8663.8	-8652.4

**b. Estimation and Prediction of First-guess Mean Square Errors using First-guess Wind Covariates**

The recursive estimation procedure in Section 2 was used to estimate the regression parameters  $\beta(t)$  for each of models (3.6) – (3.8) for 250 mb first-guess errors using first-guess 250 mb wind covariates; that is, the first-guess wind speed at location  $i$  at time  $t$  at the 250 mb level,  $s_i(f;t)$  replaces  $s_i(0;t)$ , etc. The initializing estimates of  $\beta(t)$  are the estimates obtained using all April data recorded in Jacobs and Gaver (1992); the initial variance-covariance matrix is the identity matrix. The estimates from April are used since April appears to have somewhat more predictive ability than February for July; Jacobs and Gaver (1992).

Once again, preliminary exploratory work using the resulting value of the predictive log-likelihood indicates that setting the variance-covariance matrix of the innovation of the random walk equal to 0 gives the best predictions. This suggests that the change in the relationship of the mean square error and the covariates is slow.

Figure 4, (respectively Figure 5) presents plots of the values of the estimates for the 250 mb  $u$ -wind component errors and the 250 mb  $v$ -wind component errors. The values of the estimates multiplying  $r(f;t)$ , (respectively  $s(f;t)$ ) are represented by  $\circ$ , (respectively  $+$ ) for the one-variate models. For the two-variate model, the estimates multiplying  $r(f;t)$ , (respectively  $s(f;t)$ ) are presented as  $\times$ , (respectively  $\nabla$ ). The figures suggest evidence of a trend in the estimates. Note that once again all the estimates are positive. Thus, increased first-guess resultant wind and/or wind speed tends to increase the mean square error

To assess the predictive ability of the models, the models with parameters estimated at time  $t$  are used to forecast the variances of the first-guess errors at time  $t+1$ .

One procedure to informally assess the predictive ability of the models is by binning the data. The data are binned as in the previous subsection. Figures 7 and 8 present graphs of the log [average  $y_i(t)^2$ ] in each bin versus log [average predicted variance] in each bin for models (3.6) and (3.8) and models (3.7) and (3.8). If a model were perfect, the points should be close to the 45° line shown.

Table 2 presents values of  $\bar{l}$ , given by (3.9), for the one-time step ahead prediction of variance; these values appear in the column iterative. Also

displayed are the values of  $\bar{l}$  obtained by estimating the model parameters once using all the data; this value of  $\bar{l}$  is a measure of goodness-of-fit and appears in the column labeled All; the estimates used in regressions with covariates for the goodness-of-fit evaluation of  $\bar{l}$  are those appearing for July in Table 11 of Jacobs and Gaver (1992). The constant variance estimate using all the data is the maximum likelihood estimate. The results for four models are presented; constant variance models, one-variate models (3.6) and (3.7), and the two-variate model (3.8).

Note that all of the iterative prediction values of  $\bar{l}$  for the models with covariates are larger than those for the constant variance goodness-of-fit value; this suggests that the covariates have some predictive value.

The iterative prediction value of  $\bar{l}$  for the one-variate model using  $r(f;t)$  is greater than the goodness-of-fit value obtained by using parameters estimated using all the data; this suggests that there is a systematic change in the model parameter values over time. For the other regressions using covariates the prediction values of  $\bar{l}$  are smaller than their corresponding goodness-of-fit values but not by much. The model that maximizes the prediction values of  $\bar{l}$  is the two-variate model suggesting that both covariates have some predictive value..

Figures 9 and 10 and Table 3 present results for models of the variance of the first-guess error at 500 mb level using 500 mb first-guess covariates. The implications of the results are similar to those of the 250 mb results. The predictive ability of the recursively estimated two-variate model appears somewhat greater at the 250 mb level than the 500 mb level; this conclusion is based on the values of  $(\bar{l}_2 - \bar{l}_c)/\bar{l}_c$  where  $\bar{l}_2$  is the prediction value of  $\bar{l}$  for the two-variate model and  $\bar{l}_c$  is the goodness-of-fit value of  $\bar{l}$  for the constant variance model; the value of this fraction is larger for 250 mb than for 500 mb.

**TABLE 2**  
**VALUE OF LOG-LIKELIHOOD FIRST-GUESS COVARIATES, 250 mb**

<b>Data</b>	<b>Model</b>	<b>Iterative (Prediction)</b>	<b>All (Goodness-of-Fit)</b>
<i>u</i> -wind	<i>Constant Variance</i>	-14,589.4	-14,516.5
	<i>One-variate</i>		
	<i>r(f;t)</i>	-13,640.6	-14,474.7
	<i>s(f;t)</i>	-13,598.3	-13,575.5
	<i>Two-variate</i>	-13,599.5	-13,579.4
<i>v</i> -wind	<i>Constant Variance</i>	-14,429.1	-14,363.1
	<i>One-variate</i>		
	<i>r(f;t)</i>	-13,389.2	-14,270.1
	<i>s(f;t)</i>	-13,388.4	-13,363.8
	<i>Two-variate</i>	-13,351.1	-13,329.6

**TABLE 3**  
**VALUE OF LOG-LIKELIHOOD FIRST-GUESS COVARIATES, 500 mb**

<b>Data</b>	<b>Model</b>	<b>Iterative (Prediction)</b>	<b>All (Goodness-of-Fit)</b>
<i>u</i> -wind	<i>Constant Variance</i>	-11,237.3	-11,213.5
	<i>One-variate</i>		
	<i>r(f;t)</i>	-10,561.2	-11,176.8
	<i>s(f;t)</i>	-10,518.7	-10,499.6
	<i>Two-variate</i>	-10,520.0	-10,489.0
<i>v</i> -wind	<i>Constant Variance</i>	-11,204.0	-11,160.9
	<i>One-variate</i>		
	<i>r(f;t)</i>	-10,431.0	-11,118.2
	<i>s(f;t)</i>	-10,423.0	-10,411.9
	<i>Two-variate</i>	-10,411.8	-10,393.3

## REFERENCES

- Aitken, M., (1987). "Modeling variance heterogeneity in normal regression using GLIM." *Appl Stat.*, 36, pp. 332-339.
- Cox, D. R. and D. V. Hinkley, *Theoretical Statistics*, Chapman and Hall, London, 1974.
- Davidian, M., and R. J. Carroll (1987). "Variance function estimation," *J. Am Statist. Ass.*, 82, pp. 1079-1091.
- Goerss, J. S. and P. A. Phoebus (1991a). *The Multivariate Optimum Interpolation Analysis of Meteorological Data at FNOC. NOARL Report Number 31*, Naval Oceanographic and Atmospheric Research Laboratory, Stennis Space Center, MS.
- Goerss, J. S. and P. A. Phoebus (1991b). "The Navy's operational atmospheric analysis." to appear in *Weather and Forecasting*.
- Hogan, T. F., 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.
- Jacobs, P. A. and D. P. Gaver, "Preliminary results from the analysis of wind component error." *Naval Postgraduate School Technical Report NPSOR-91-029*, September, 1991.
- Jacobs, P. A. and D. P. Gaver, "Preliminary results from the analysis of wind component error—July Data." *Naval Postgraduate School Technical Report*, to appear.
- McCullagh and J. A. Nelder, *Generalized Linear Regression*, Chapman and Hall, New York, 1983.
- Nelder, J. A. and Y. Lee (1992), "Likelihood, quasi-likelihood and pseudolikelihood: some comparisons." *J. R. Statist. Soc., B*, 54, No. 1, pp. 273-284.
- Carlin, B. P., N. G. Polson, D. S. Stoffer. "A Monte Carlo approach to nonnormal and nonlinear state space modeling." *J. Amer. Statist. Ass.* 87, No. 418, pp. 493-500.

de Bruijn, N. G., *Asymptotic Methods in Analysis*, Interscience, New York, 1958.

Easton, G. S., "Location compromise maximum likelihood estimators," in *Configural Polysampling*, ed. S. Morgenthaler and J. W. Tukey, Wiley, New York, 1991, pp. 157-192.

West, M. and J. Harrison, *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York, 1989.

# 850 MB U WIND

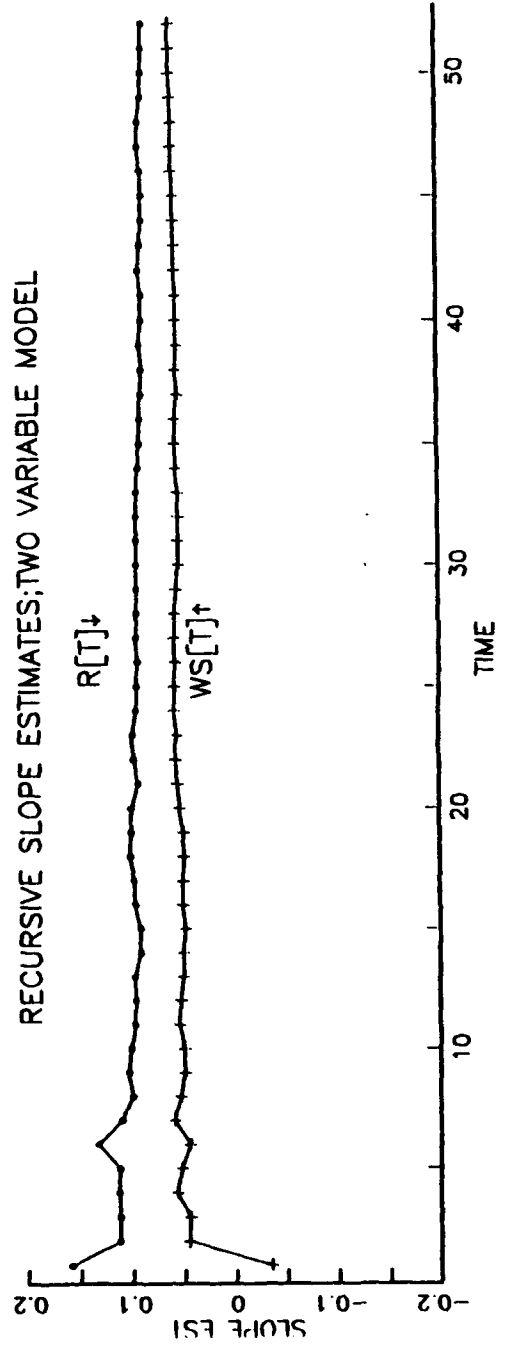
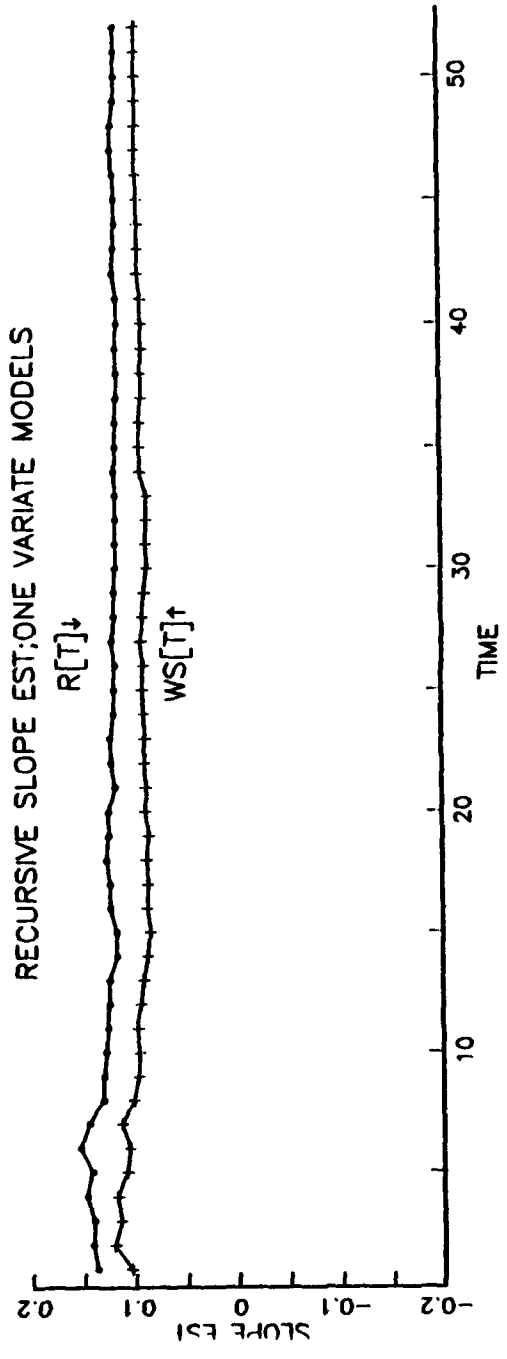


Figure 1

850 MB V WIND; RECURSIVE ESTIMATES; OBS WINDS

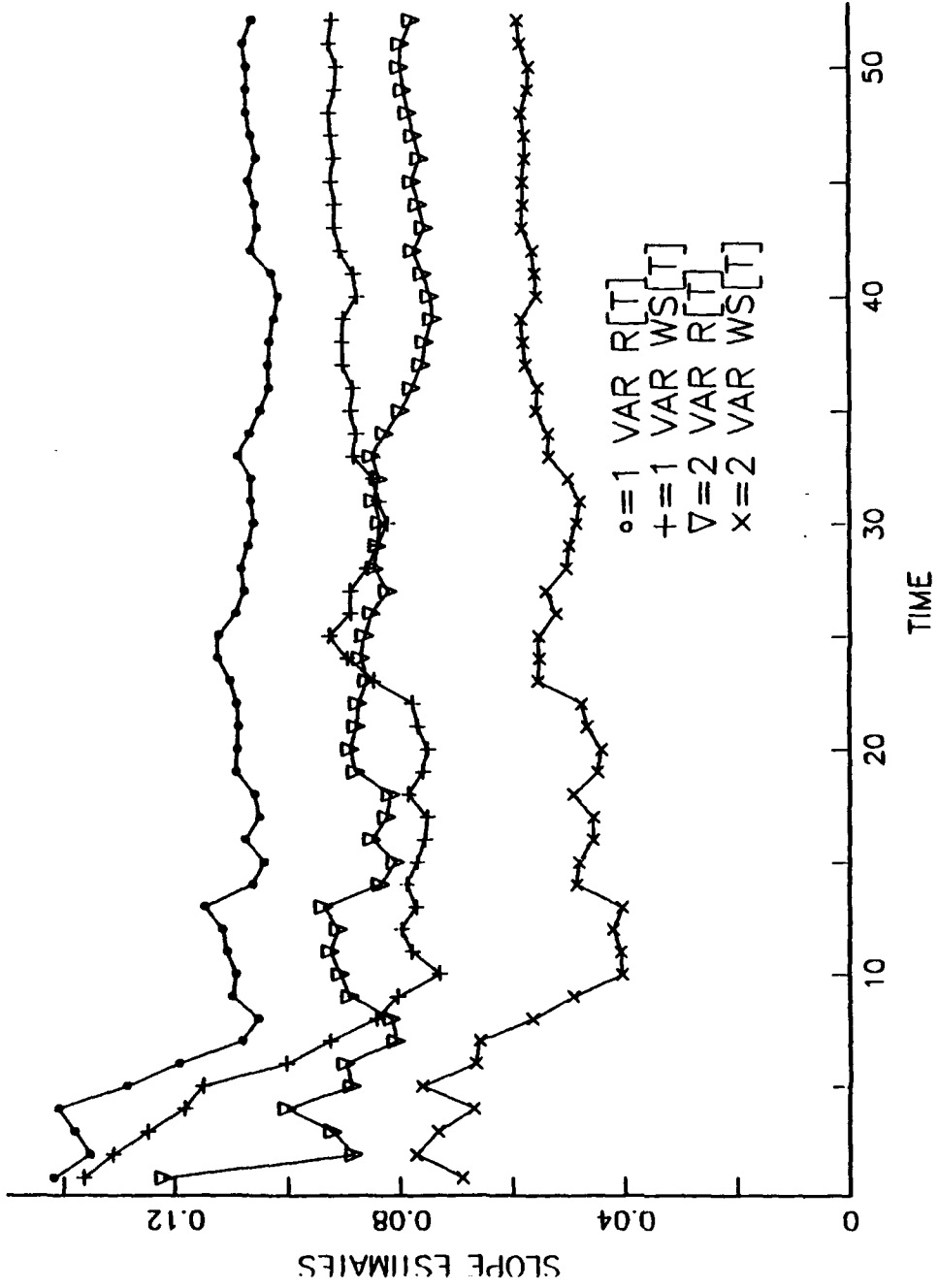
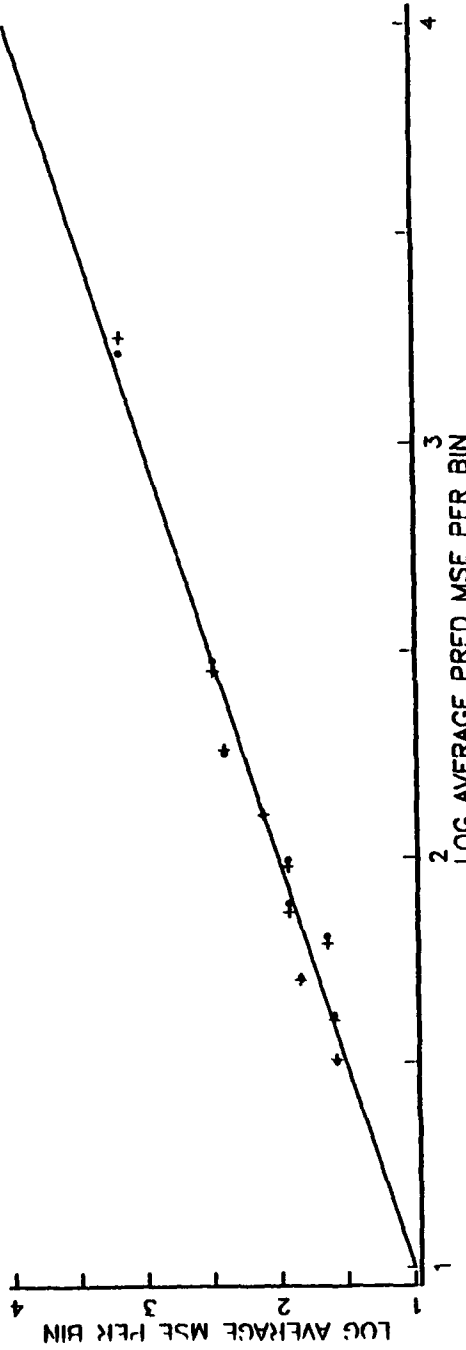


Figure 2

850 MB U WIND; RECURSIVE ESTIMATES; OBS WIND

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



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

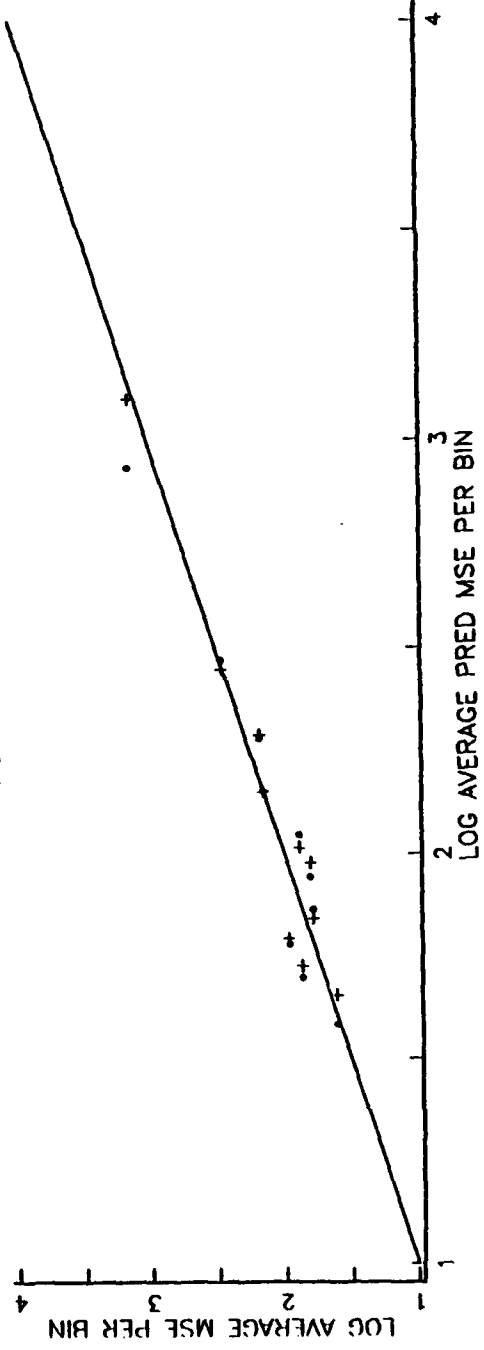
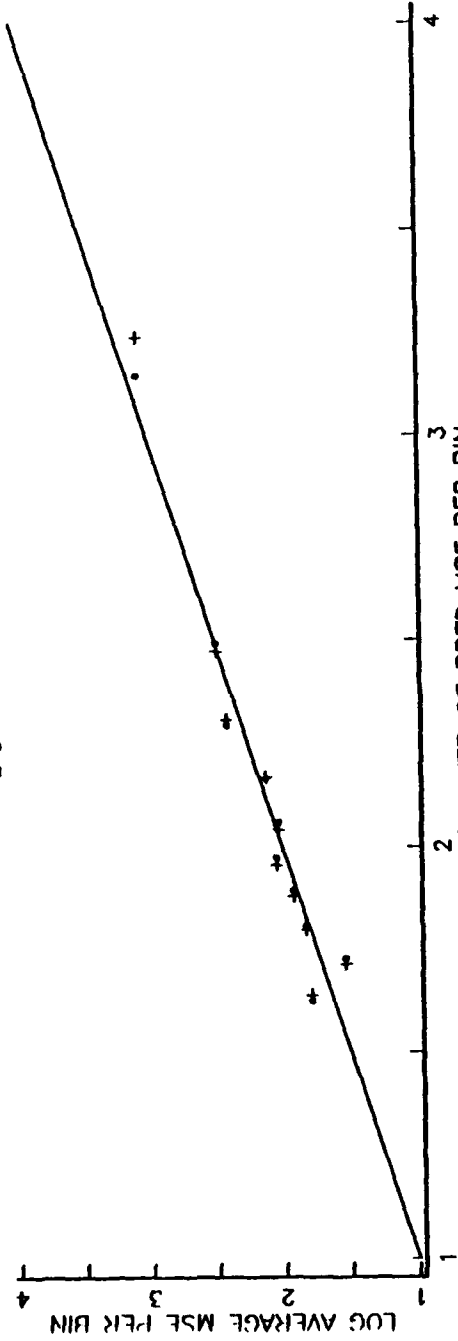


Figure 3

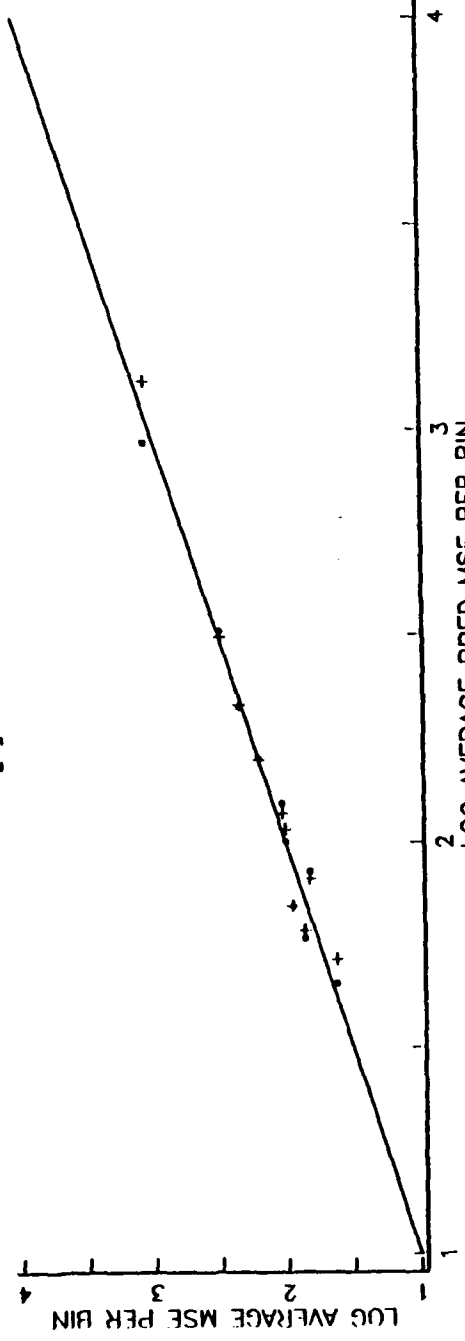
850 MB V WIND; RECURSIVE ESTIMATES; OBS WIND

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



2 LOG AVERAGE PRED MSE PER BIN

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



2 LOG AVERAGE PRED MSE PER BIN

Figure 4

250 MB U WIND; RECURSIVE ESTIMATES; 1ST GUESS

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

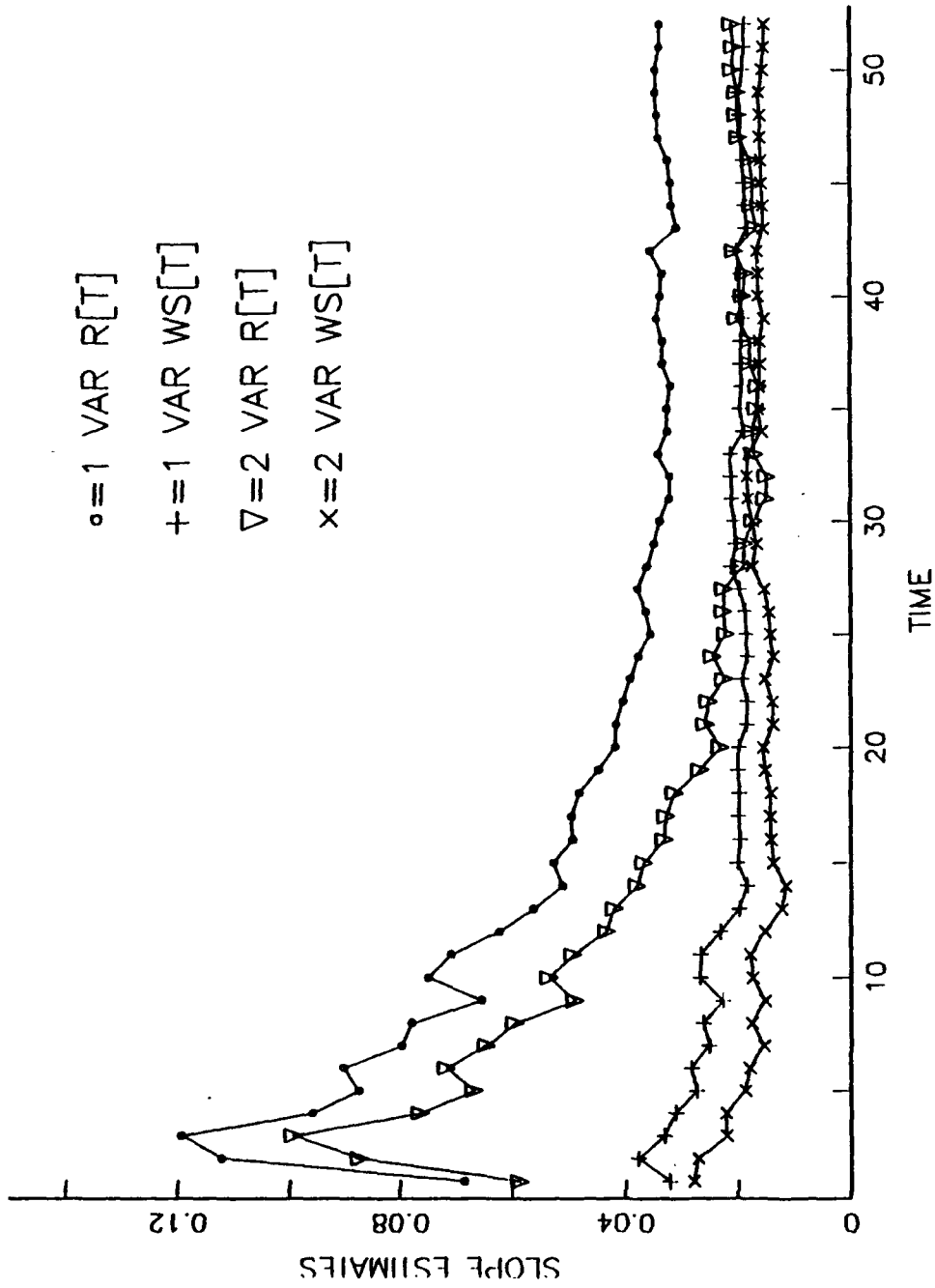


Figure 5

250 MB V WIND; RECURSIVE ESTIMATES; 1ST GUESS

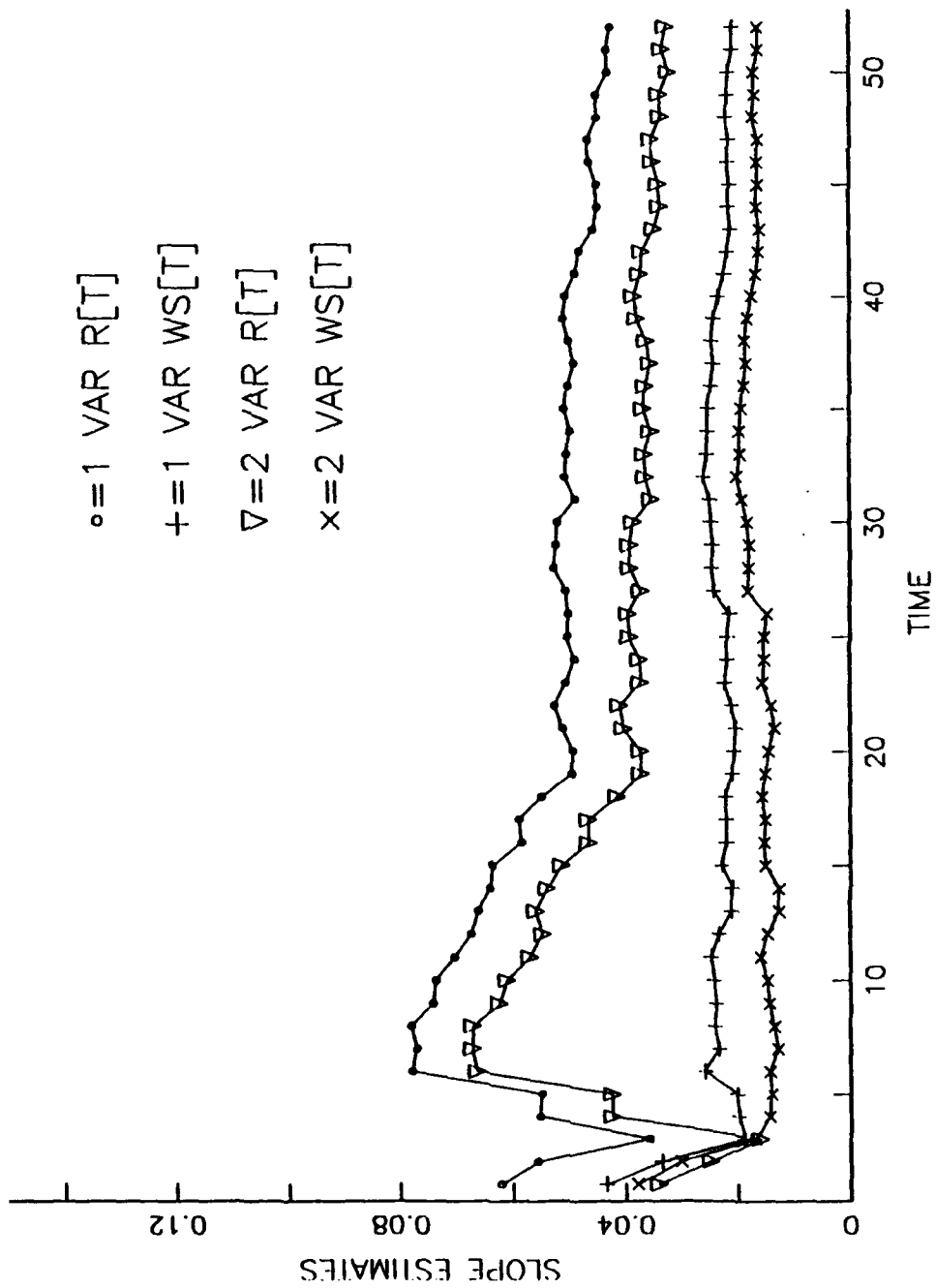
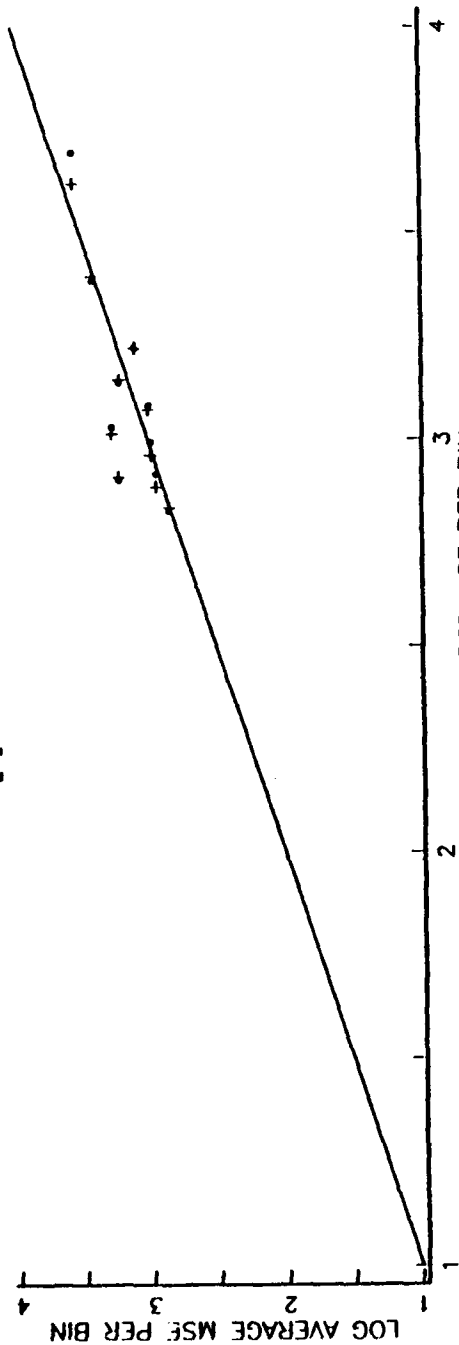


Figure 6

250 MB U WIND; RECURSIVE ESTIMATES; 1ST GUESS

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



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

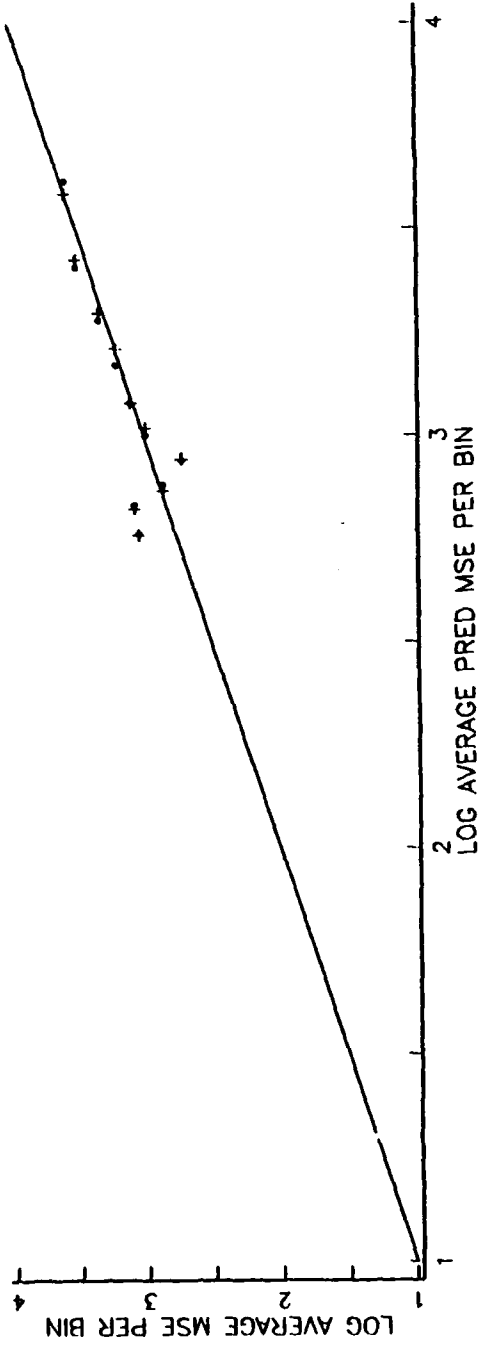
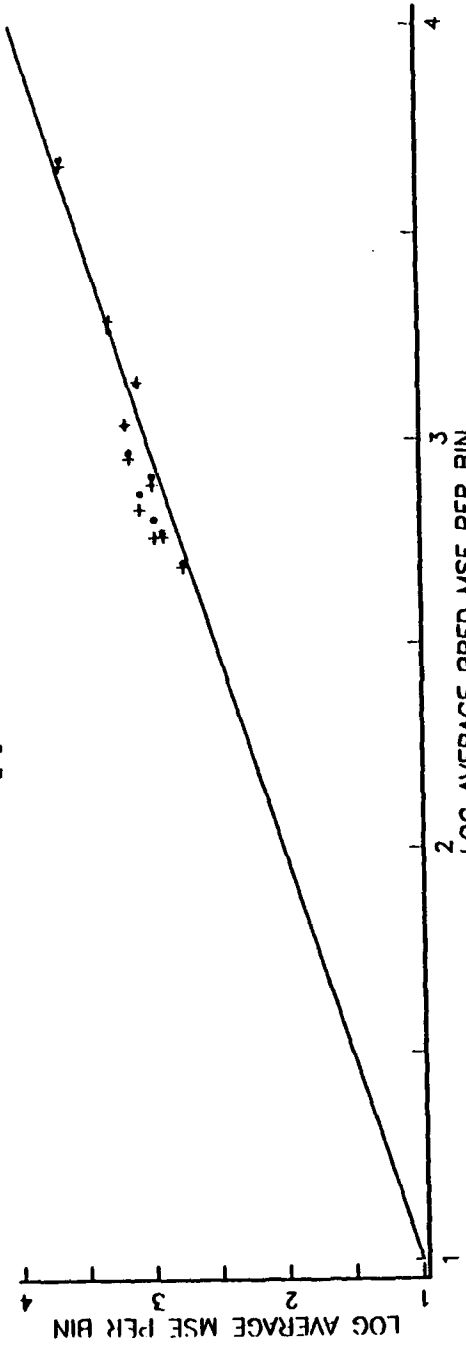


Figure 7

250 MB V WIND; RECURSIVE ESTIMATES; 1ST GUESS

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



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

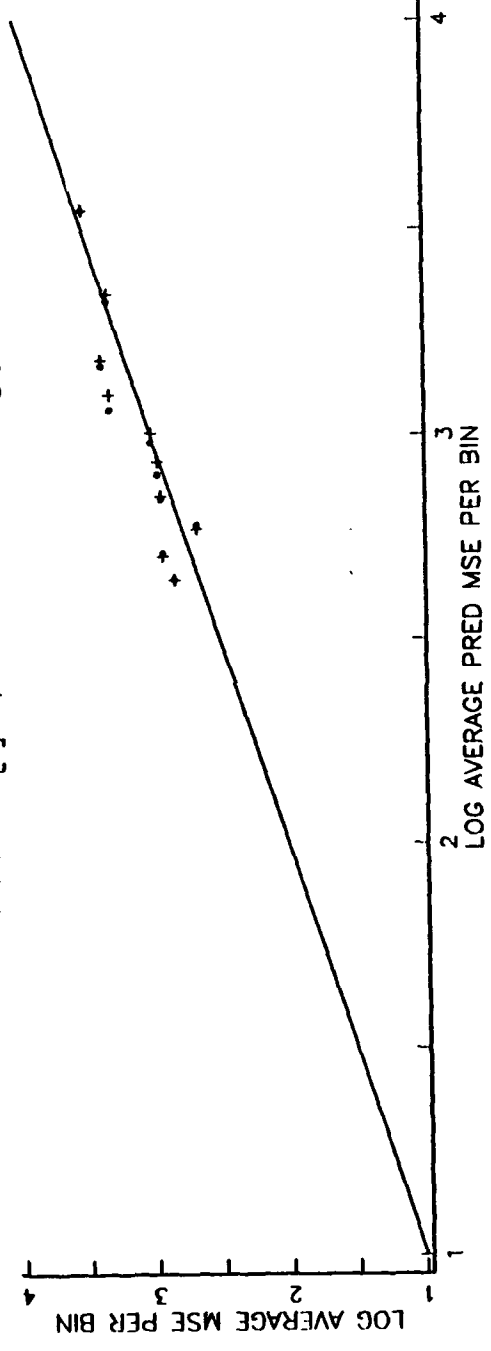
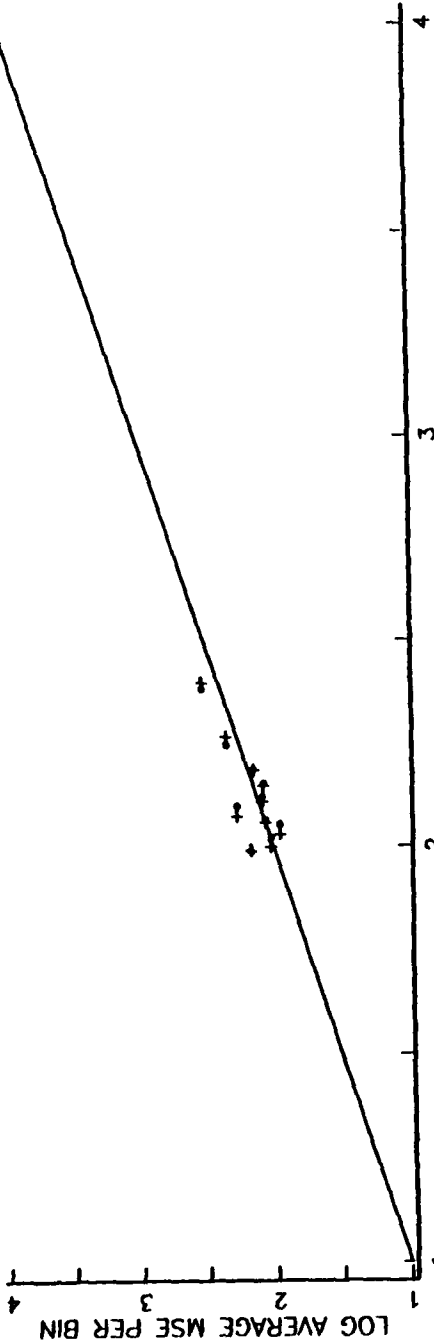


Figure 8

500 MB U WIND; RECURSIVE ESTIMATES; 1ST GUESS

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



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

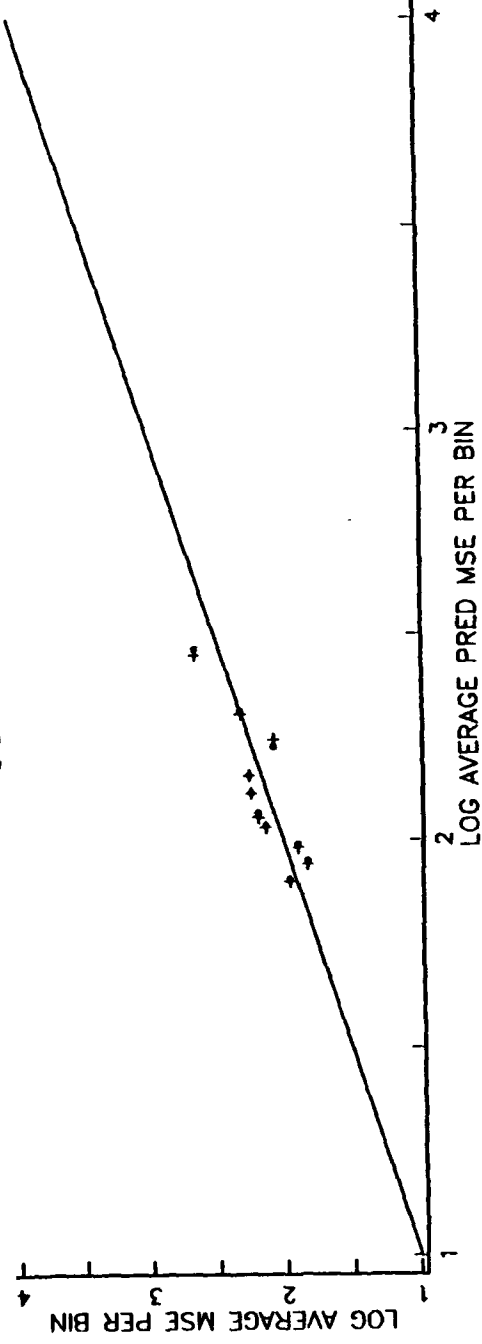
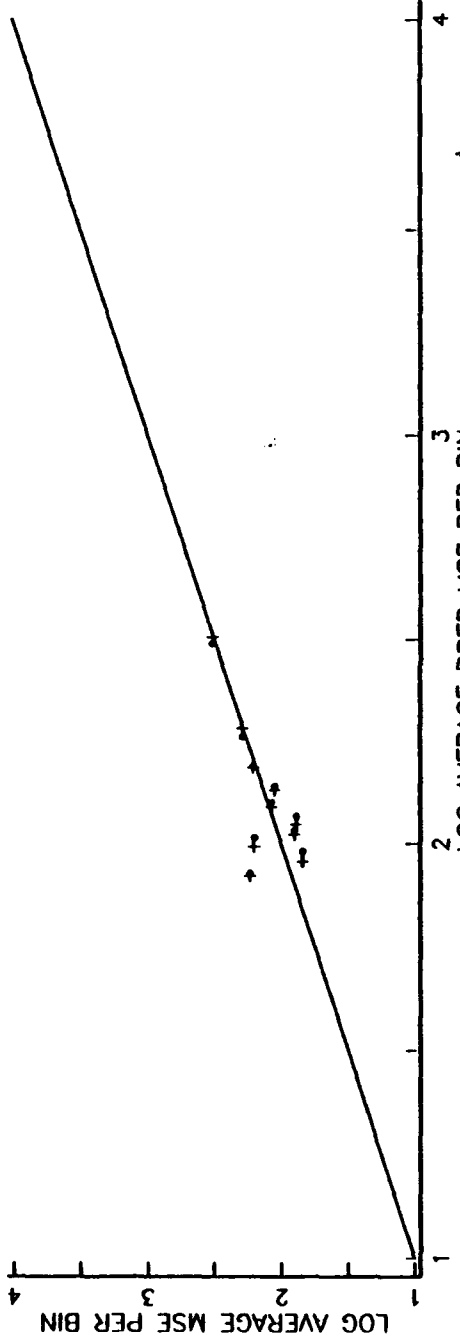


Figure 9

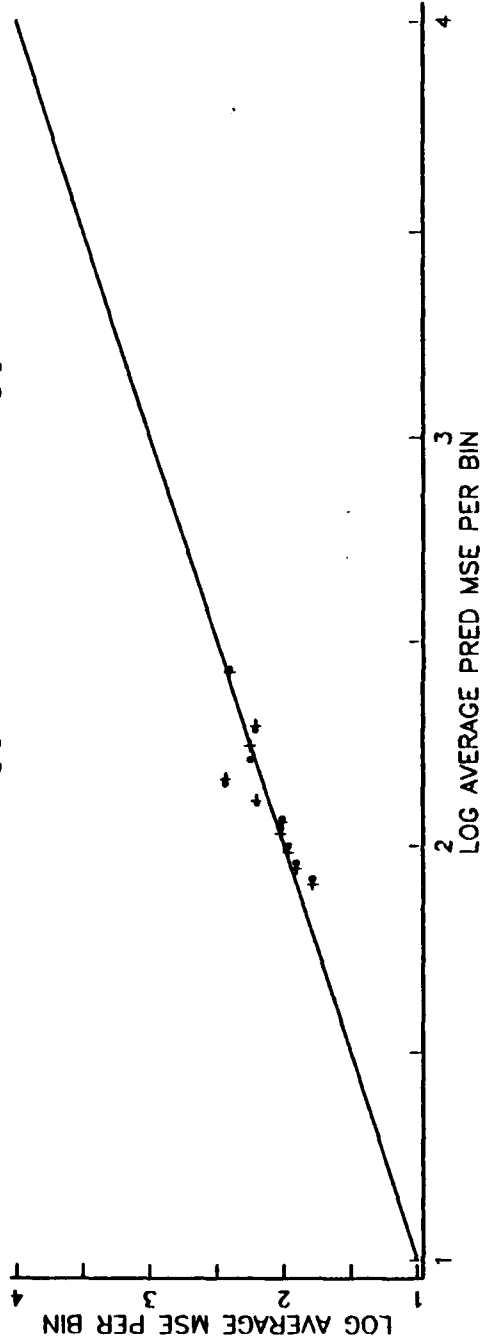
500 MB V WIND; RECURSIVE ESTIMATES; 1ST GUESS

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



2 LOG AVERAGE PRED MSE PER BIN

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



2 LOG AVERAGE PRED MSE PER BIN

Figure 10

## INITIAL DISTRIBUTION LIST

1. Library (Code 52).....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 81) .....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. Bernard Harris .....1  
 Dept of Statistics  
 University of Wisconsin  
 610 Walnut Street  
 Madison, WI 53706
  
11. Prof. W. R. Schucany.....1  
 Dept. of Statistics  
 Southern Methodist University  
 Dallas, TX 75222
  
12. Prof. D. C. Siegmund.....1  
 Dept. of Statistics  
 Sequoia Hall  
 Stanford University  
 Stanford, CA 94305
  
13. Prof. H. Solomon.....1  
 Department of Statistics  
 Sequoia Hall  
 Stanford University  
 Stanford, CA 94305
  
14. Dr. Ed Wegman.....1  
 George Mason University  
 Fairfax, VA 22030
  
15. Dr. P. Welch .....1  
 IBM Research Laboratory  
 Yorktown Heights, NY 10598
  
16. Dr. Neil Gerr .....1  
 Office of Naval Research  
 Arlington, VA 22217
  
17. Prof. Roy Welsch.....1  
 Sloan School  
 M.I.T.  
 Cambridge, MA 02139
  
18. Dr. J. Abrahams, Code 1111, Room 607 .....1  
 Mathematical Sciences Division, Office of Naval Research  
 800 North Quincy Street  
 Arlington, VA 22217-5000

19. Prof. J. R. Thompson.....1  
 Dept. of Mathematical Science  
 Rice University  
 Houston, TX 77001
20. Prof. M. Leadbetter.....1  
 Department of Statistics  
 University of North Carolina  
 Chapel Hill, NC 27514
21. Prof. J. B. Kadane.....1  
 Dept. of Statistics  
 Carnegie-Mellon University  
 Pittsburgh, PA 15213
22. Prof. J. Lehoczky.....1  
 Department of Statistics  
 Carnegie-Mellon University  
 Pittsburgh, PA 15213
23. Prof. M. Mazumdar.....1  
 Dept. of Industrial Engineering  
 University of Pittsburgh  
 Pittsburgh, PA 15235
24. Prof. M. Rosenblatt.....1  
 Department of Mathematics  
 University of California, San Diego  
 La Jolla, CA 92093
25. Prof. H. Chernoff.....1  
 Department of Statistics  
 Harvard University  
 1 Oxford Street  
 Cambridge, MA 02138
26. Prof. Frank Samaniego.....1  
 Statistics Department  
 University of California  
 Davis, CA 95616

27. Dr. James McKenna.....1  
 Bell Communications Research  
 445 South Street  
 Morristown, NJ 07960-1910
28. Prof. Brad Carlin.....1  
 School of Public Health  
 University of Minnesota  
 Mayo Bldg. A460  
 Minneapolis, MN 55455
29. Prof. W. Stuetzle.....1  
 Department of Statistics  
 University of Washington  
 Seattle, WA 98195
30. Prof. F. W. Mosteller.....1  
 Department of Statistics  
 Harvard University  
 1 Oxford St.  
 Cambridge, MA 02138
31. Dr. D. C. Hoaglin.....1  
 Department of Statistics  
 Harvard University  
 1 Oxford Street  
 Cambridge, MA 02138
32. Prof. N. D. Singpurwalla.....1  
 George Washington University  
 Washington, DC 20052
33. Prof. George S. Fishman.....1  
 Curr. in OR & Systems Analysis  
 University of North Carolina  
 Chapel Hill, NC 20742
34. Prof. Bradley Efron.....1  
 Statistics Dept.  
 Sequoia Hall  
 Stanford University  
 Stanford, CA 94305

35. Prof. Carl N. Morris.....1  
 Statistics Department  
 Harvard University  
 1 Oxford St.  
 Cambridge, MA 02138
36. Dr. John E. Rolph.....1  
 RAND Corporation  
 1700 Main St.  
 Santa Monica, CA 90406
37. Dr. Edward G. Coffman, Jr.....1  
 AT&T Bell Telephone Laboratories  
 600 Mountain Avenue  
 Murray Hill, NJ 07974
38. Operations Research Center, Rm E40-164.....1  
 Massachusetts Institute of Technology  
 Attn: R. C. Larson and J. F. Shapiro  
 Cambridge, MA 02139
39. Arthur P. Hurter, Jr.....1  
 Professor and Chairman  
 Dept. of Industrial Engineering and Management Sciences  
 Northwestern University  
 Evanston, IL 60201-9990
40. Institute for Defense Analysis.....1  
 1800 North Beauregard  
 Alexandria, VA 22311
41. Prof. J. W. Tukey.....1  
 Statistics Dept., Fine Hall  
 Princeton University  
 Princeton, NJ 08540
42. Dr. Colin Mallows.....1  
 AT&T Bell Telephone Laboratories  
 600 Mountain Avenue  
 Murray Hill, NJ 07974

43. Dr. Jon Kettenring.....1  
 Bellcore  
 445 South Street  
 Morris Township, NJ 07962-1910
44. Prof. David L. Wallace.....1  
 Statistics Dept., University of Chicago  
 5734 S. University Ave.  
 Chicago, IL 60637
45. Dr. S. R. Dalal.....1  
 Bellcore  
 445 South Street  
 Morristown, NJ 07962-1910
46. Dr. D. F. Daley.....1  
 Statistic Dept. (I.A.S.)  
 Australian National University  
 Canberra, A.C.T. 2606  
 AUSTRALIA
47. Dr. Guy Fayolle.....1  
 I.N.R.I.A.  
 Dom de Voluceau-Rocquencourt  
 78150 Le Chesnay Cedex  
 FRANCE
48. Professor H. G. Daellenbach.....1  
 Department of Operations Research  
 University of Canterbury  
 Christchurch, NEW ZEALAND
49. Koh Peng Kong.....1  
 OA Branch, DSO  
 Ministry of Defense  
 Blk 29 Middlesex Road  
 SINGAPORE 1024
50. Professor Sir David Cox.....1  
 Nuffield College  
 Oxford, OXI INF  
 ENGLAND

51. Dr. A. J. Lawrence.....1  
 Dept. of Mathematics,  
 University of Birmingham  
 P. O. Box 363  
 Birmingham B15 2TT  
 ENGLAND
52. Dr. F. P. Kelly.....1  
 Statistics Laboratory  
 16 Mill Lane  
 Cambridge  
 ENGLAND
53. Dr. John Copas.....1  
 Dept. of Mathematics,  
 University of Birmingham  
 P. O. Box 363  
 Birmingham B15 2TT  
 ENGLAND
54. Dr. D. Vere-Jones.....1  
 Dept. of Math, Victoria Univ. of Wellington  
 P. O. Box 196  
 Wellington  
 NEW ZEALAND
55. Prof. Guy Latouche.....1  
 University Libre Bruxelles  
 C.P. 212, Blvd. De Triomphe  
 B-1050 Bruxelles  
 BELGIUM
56. Dr. J. Goerss.....10  
 Naval Oceanographic and Atmospheric Laboratory  
 Monterey, CA 93943-5006