PL-TR-93-2205

SHORT-RANGE CLOUD AMOUNT FORECASTING WITH MODEL OUTPUT STATISTICS

M.E. Cianciolo

TASC
55 Walkers Brook Drive
Reading, MA 01867

September 1993

Final Report
February 1992 – September 1993

PHILLIPS LABORATORY
Directorate of Geophysics
AIR FORCE MATERIEL COMMAND
HANSCOM AIR FORCE BASE, MA 01731-3010

This technical report has been reviewed and is approved for publication.

H. STUART MUENCH
Contract Manager

for DONALD A. CHISHOLM
Chief, Atmospheric Prediction Branch
Atmospheric Sciences Division

FOR ROBERT A. McCLATCHEY
Director, Atmospheric Sciences Division

This report has been reviewed by the ESC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center (DTIC). All others should apply to the National Technical Information Service (NTIS).

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify PL/TSI, Hanscom AFB MA 01731-3010. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1993	Final (February 1992 – September 1993)

4. TITLE AND SUBTITLE
Short-Range Cloud Amount Forecasting With Model Output Statistics

5. FUNDING NUMBERS
PE 62101F
PR 6670 TA 10 WU EA
Contract F19628-91-C-0185

6. AUTHOR(S)
M.E. Cianciolo

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
TASC
55 Walkers Brook Drive
Reading, MA 01867

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Phillips Laboratory
29 Randolph Road
Hanscom AFB, MA 01731-3010

Contract Manager: H. Stuart Muench/GPAP

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
PL-TR-93-2205

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

TASC conducted a feasibility study of the utility of short-range (3-, 6-, and 9-hour) forecasts of total cloud amount and probability of cloud cover categories using the Model Output Statistics (MOS) approach. The MOS models were built using output from the Air Force Global Spectral Model (GSM) and the Real-Time NEPHanalysis (RTNEPH) model. The RTNEPH total cloud amount fields were also used as "ground truth."

Models were developed on the eighth-mesh RTNEPH grid (approximately 40 km resolution) for 71 data sets covering a range of locations (tropics and midlatitudes), seasons, and times of day. Results were compared with persistence forecasts using the Brier score, 20/20 score, and sharpness measure. The MOS-based forecasts outperformed persistence with respect to the overall reduction in mean square error, but forecast sharpness was sacrificed.

Estimates for future model performance for total cloud cover and probabilities of cloud cover categories were determined for two sample cases. Considering the short data records used to develop the equations, the MOS-based forecast models were remarkably robust in most cases.

Recommendations for model improvements and additional performance assessment included: using higher spatial resolution model data that better match the resolution of the eighth-mesh forecasts, using cloud layer variables in model development, developing regression coefficients using additional seasons of data, using non-linear models, and continuing model evaluation using independent cloud data.

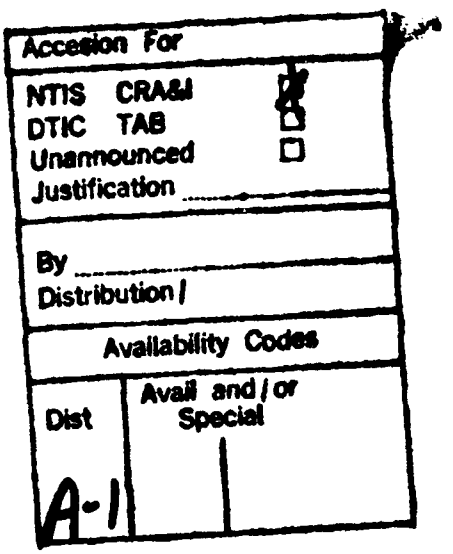
14. SUBJECT TERMS
Model Output Statistics
Cloud Forecasting
Statistical Forecasting

15. NUMBER OF PAGES
166

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

TABLE OF CONTENTS



LIST OF FIGURES

LIST OF TABLES

FOREWORD

The Model Output Statistics (MOS) approach has been used extensively to predict a variety of meteorological phenomena. The approach consists of extracting statistical relationships between a number of model-generated variables (the predictors) and the forecast variable of interest (the response — in our case, cloud cover fraction). The model data are typically output from a physically based (i.e., numerical) forecast model such as those used at the U.S. National Weather Service.

In order to develop a MOS-based forecasting procedure with sufficient skill to be of value to users, a tremendous amount of data must be acquired and analyzed. Furthermore, those data must be collected from the same physical environment in which the model will be applied. Thus, a large collection of winter data will, almost surely, be of little value for forecasting weather variables during the summer. Similarly, if the large physical model whose outputs include some of the predictors is changed for any reason, then MOS relationships must be re-derived, or at least adjusted and verified, to accommodate the changes.

In spite of the above complexities, the possibility of large-scale applications of the MOS approach to cloud-cover fraction forecasting is attractive to the U.S. Air Force because of its extreme efficiency once the MOS functional relationship has been established. This project, therefore, was funded to perform a carefully designed feasibility study of the MOS approach to cloud-cover fraction forecasting. The study considered forecasts of two response variables: cloud-cover *fraction* and *probability of cloud-cover fraction* less than a specified threshold value. The latter response variable is of special interest to one part of the customer community.

The one year of data made available to TASC by the government for this feasibility study would not be adequate to support full development of an operational MOS forecasting system valid over future years. It was, however, adequate to judge feasibility in terms of potential accuracy of the forecasts, variability of the forecast equations with seasons and with physical locations, and similar issues.

ACKNOWLEDGMENTS

I would like to acknowledge the efforts of the other members of the MOS team at TASC: Drs. Paul Janota (consultant), David Marcus, Charles Medler, and David Whitney. A special acknowledgment to Dr. Gary Rasmussen who was heavily involved with the early phases of the project and who wrote the technical proposal (Ref. 26) for this effort from which much of the background material in this report was taken.

1. INTRODUCTION

1.1 OVERVIEW OF OUR GOALS AND METHODOLOGY

The goal of this project is to *"demonstrate whether or not the model output statistics (MOS) approach should be considered as a serious contender for a low-cost, low-risk improvement to the current Air Force Global Weather Central (GWC) cloud amount forecasting system"* (Ref. 26). The Department of Defense has assigned GWC the task of providing accurate, timely, global cloud forecasts at three-hour intervals out to 48 hours (Ref. 9). Reduced visibility due to clouds in the fields of view of electro-optical sensor systems adversely impacts military operations such as reconnaissance missions, tactical air strikes, and mid-air refueling. To maximize the chance of mission success, Army and Air Force decision makers require accurate, timely forecasts of cloud amount. In addition, mission planning may require forecasts of the probability of exceeding some mission-specific cloud coverage.

During this project, TASC developed model output statistics regression models for 3-, 6-, and 9-hour cloud amount forecasts and compared the results to persistence forecasts to determine the utility of the MOS approach for short-range cloud prediction. Regression models were developed for up to eight different times of day on the eighth-mesh grid (approximately 40 km grid spacing at midlatitudes) within climatologically distinct regions of the northern hemisphere during all seasons. As data sources for the MOS forecasts, we used total cloud amount fields from the Real-Time NEPHanalysis (RTNEPH) model and meteorological variables from the Air Force Global Spectral Model (GSM). We also used terrain type and elevation variables from a terrain database (on the eighth-mesh grid) compiled at Phillips Laboratory. The RTNEPH cloud amount fields were also used as "ground truth."

The modeling process consisted of the following major components. A more complete description of the process can be found in Section 3.

STEP 1 Construct clusters of RTNEPH grid points based on cloud climatology and data availability. Select clusters for which to develop regression models.

STEP 2 Develop a pool of potential predictors from RTNEPH and GSM model data. This pool includes raw and derived variables along with averaged, differenced, and transformed variables.

STEP 3 Develop regression models for total cloud amount using variables selected from the pool of predictors. Determine conditional probabilities of cloud amount thresholds.

STEP 4 Predict MOS model performance for future data sets using the jackknife resampling technique to analyze total cloud cover forecasts and a Bayesian approach to analyze probabilistic forecasts.

Although the overall process is shown as a linear sequence, we continued to iterate on the sequence as the modeling effort proceeded. Keeping in mind the goal of the project — a feasibility study — we emphasized research and exploration and a dynamic modeling process in which findings from later steps led us to modify earlier ones.

Although not tasked to develop exhaustive operational MOS-based forecast models, we were acutely aware of the operational constraints of the RTNEPH and GSM model runs. Therefore we were careful to include only those data sources that are operationally available at any given time as MOS inputs. Our development data set consisted of one year (1989) of Global Spectral Model output mapped to RTNEPH coordinates within 20 RTNEPH boxes in the northern hemisphere and the RTNEPH cloud analysis grids for the same time period and spatial regions. Both of these data sets are described further in Section 2. Jacks et al., in Ref.17, point out that at least two years of model data are needed to develop meaningful MOS equations based on the Nested Grid Model. We use one year of data in this feasibility study with the understanding that more robust models that better account for year-to-year meteorological variability would be necessary in an operational setting.

1.2 OVERVIEW OF THIS REPORT

The remainder of this section discusses the current forecasting capabilities at GWC and the potential improvements to be made by using a MOS approach. Section 2 provides descriptions of the RTNEPH and GSM data sources used in model development. Section 3 contains a description of our modeling process along with the modeling and validation results. Finally, we summarize and make recommendations for further research in Section 4.

1.3 THE CURRENT GWC CLOUD FORECASTING SYSTEM

At present, GWC operates a Global Spectral numerical prediction Model (GSM), a global cloud analysis model (RTNEPH), a global High Resolution Analysis System (HIRAS) to initialize the GSM, and three cloud forecasting models: the Tropical Cloud Forecast Model (TRONEW), the Five-Layer Cloud Forecast Model (5LAYER), and the High-Resolution Cloud Prognosis Model (HRCP). Figure 1 depicts the system's component models and the data flow among them. The RTNEPH cloud analysis model provides initial data for the

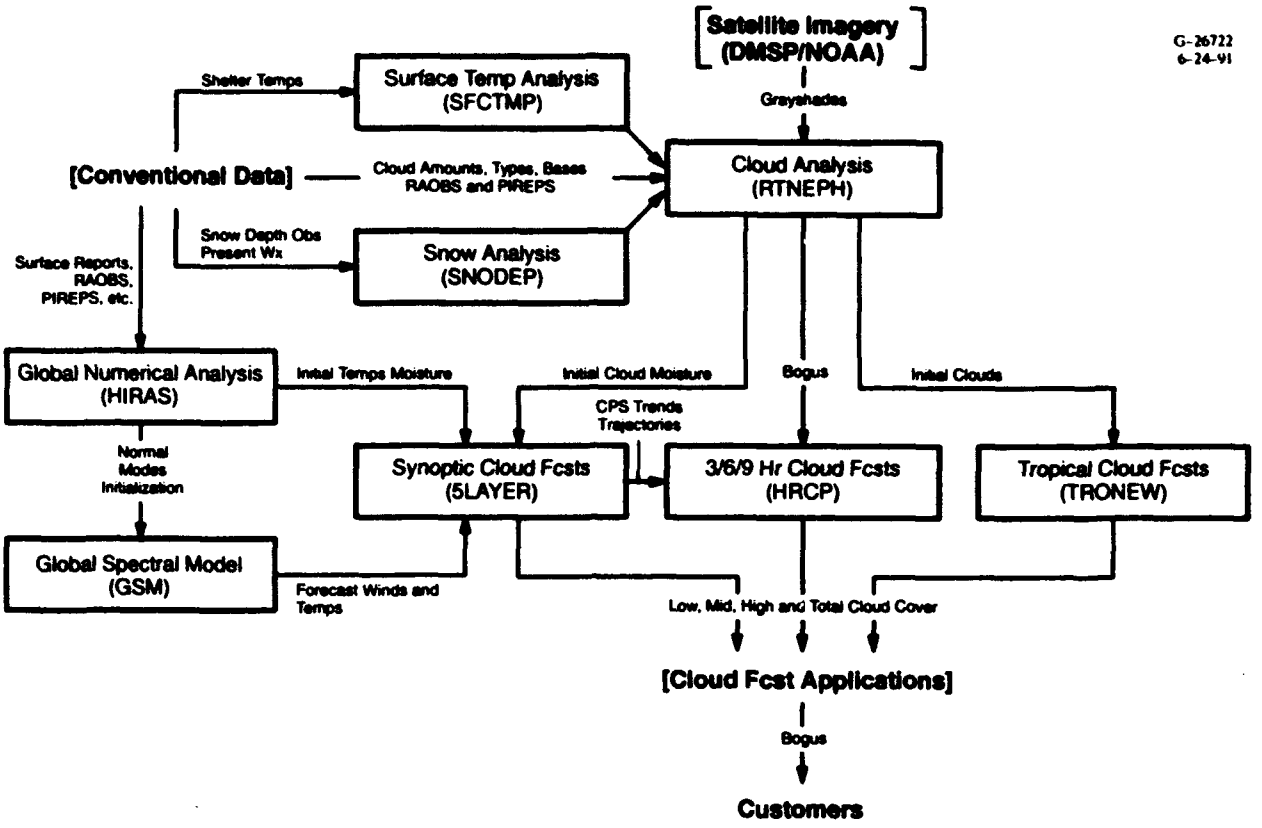


Figure 1 Global Weather Central Cloud Analysis and Forecast System

three cloud forecast models. It and the Global Spectral Model are discussed in some detail in Section 2. Here we summarize the three cloud forecast models.

Tropical Cloud Forecast Model (TRONEW)—The TRONEW model (Ref. 9) is based on RTNEPH cloud analyses. It persists the diurnal cycle for 24 hours, and outputs total cloud amounts and layer cloud amounts and types at three-hour intervals out to 21 hours. The complete half-mesh grid (approximately 100 km grid spacing at the equator) is used in both hemispheres to provide global coverage. Known problems include: the simple forecast technique (no dynamics), poor horizontal resolution, and poor temporal resolution in the tropics due to inadequate DMSP satellite coverage in that area.

Five-Layer Cloud Forecast Model (5LAYER)—The 5LAYER model (Ref. 9) inputs are the RTNEPH cloud analyses and the HIRAS/GSM analyses and forecasts of wind, temperature, and moisture. It uses a quasi-Lagrangian trajectory scheme to advect temperature and condensation pressure spread (CPS). A subset of the half-mesh grid (an octagon which excludes latitudes between 0 and approximately 13 degrees) is used in each hemisphere. Forecasts are made on five vertical levels: gradient (60 mb above the surface), 850, 700, 500, and 300 mb. Output is generated every three hours to 48 hours in the

northern hemisphere and to 24 hours in the southern hemisphere. Output parameters include: total cloud amounts, layered cloud amounts, precipitation amounts and types, accumulated precipitation, icing type, dew point depression, trajectories, Showalter stability, temperature, and CPS. Known problems include: lack of coverage near the equator, poor performance in the boundary layer, lack of entrainment and cumulus development, poor handling of cirrus and fog/stratus discrimination, and the simplicity of the empirical relationship between CPS and cloud amount.

High-Resolution Cloud Prognosis Model (HRCP)—The HRCP model (Ref. 9) inputs are the RTNEPH analyses, the 5LAYER CPS and trajectory fields, and terrain heights. It can use either persistence or forecast trends in CPS to project CPS out to three, six, and nine hours. CPS is then related to cloud amount using empirical curves originally developed in the 1960's. The HRCP model operates in a window on the eighth-mesh grid. The window is movable in the sense that up to a maximum of 13 of the 120 RTNEPH boxes (60 per hemisphere) may be specified for a given run. Each box contains 4096 (64 by 64) points with a nominal spacing of 40 km. Outputs include total cloud amounts and cloud amounts and types for low, middle, and high layers. Known problems include: low quality forecasts for the boundary layer and the middle layers, and inadequate parameterization of cloud amount as a function of CPS.

1.4 STATISTICAL CLOUD FORECASTING

1.4.1 Terminology

The terms *forecast* and *predict* (and their derivatives) are commonly used interchangeably in meteorology to imply a projection from an *initial time* to a later *valid time.* This is a *prognostic* process. It may employ dynamical, statistical, subjective, or other techniques. Unfortunately, both terms are also commonly used interchangeably in statistics with an entirely different connotation. In statistics they imply existence of an association between a dependent variable and a set of independent (or explanatory) variables. This relationship may be *prognostic*, or it may be *diagnostic* (i.e., without a time projection).

In statistical models such as regression, the dependent variable is called the *response* while the independent variables are called *predictors*. The terms *dependent* and *independent* are unfortunate. While meaningful in the vocabulary of mathematical functions, they lead to confusion in this context since the independent variables are usually not statistically independent. Thus, we choose to use the terms *predictor* and *response* rather than *dependent* and *independent variables*.

To minimize likely confusion resulting from a dual use of the term *predict*, we adopt the following conventions. The term *predict* (and its derivatives) will always be used to imply application of a statistical model. As noted above, the statistical association may be either prognostic or diagnostic. The term *forecast* (and its derivatives) will always be used to imply a prognostic process, whether by statistical, dynamical, or other means. With these conventions, a forecast may or may not be a prediction, and vice versa.

1.4.2 Three Basic Approaches

The three basic approaches to the statistical forecasting of weather parameters such as cloud amount for a specified location, at a specified valid time are: the classical approach, the perfect prog approach, and the model output statistics approach. In addition, various combinations and extensions are possible. The three approaches are described below.

Classical Approach—Predictors are specified at the initial time. Cloud amount is forecast for the valid time using a statistical method. Relevant techniques include correlation, regression, conditional probability, and autoregressive moving average (ARMA) time series models. A major shortcoming of the classical approach is that it ignores the benefits of available dynamical forecast models.

Perfect Prog Approach—Predictors are specified at the valid time by a prognostic process, usually numerical. Cloud amount is then diagnosed by a statistical technique. Parameters of the diagnostic model (e.g., the regression coefficients) are derived from observations, not from forecast model outputs. Since the statistical model is unaware of the forecast model biases, it cannot correct for them. From one point of view, however, this shortcoming is an advantage: the diagnostic model will not require revision when the forecast model undergoes a significant upgrade.

Model Output Statistics Approach—Predictors are specified at the valid time by a numerical forecast model. Cloud amount is diagnosed by a statistical model. Parameters of the diagnostic model are based on *both* observations of cloud amount and on numerical model forecasts of the predictors. Since the statistical model coefficients are derived from forecast model outputs, the MOS forecast can correct inherent numerical model biases. The model output statistics approach usually outperforms both the classical approach and perfect prog approach because it uses more information than its competitors. This assumes, of course, that an adequate development database is available.

A key point to keep in mind concerning the MOS approach is that although it can account for systematic errors in the driving model, it cannot account for random model errors. In addition, any significant changes to the underlying forecast model will require that a new set of model equations be developed based on at least one or two years of model output. In the end, of course, the quality of the MOS model is strongly dependent on both the consistent performance and accuracy of the driving forecast model.

Various extensions to and combinations of the three basic approaches to statistical cloud forecasting are possible. We used a combination of the classical and model output statistics approaches in this effort. That is, we included both observations at the initial time and forecast variables at the valid time in the pool of potential predictors.

1.4.3 Recent Work Using the Model Output Statistics Approach

The body of literature on model output statistics is very large. Here we outline some of the developments in MOS cloud amount forecasting.

Model output statistics methods for cloud amount forecasting date back to the early 1970s and the pioneering work by Glahn at the National Weather Service's Technique Development Laboratory (Ref. 13). With Lowry, he developed multiple linear regression equations for coded cloud amount (0=clear, 1=partial obscuration, 2=thin scattered, . . ., 7=overcast, 8=obscured) at four stations for four times of day. They reported success using observed cloud amount, and forecasts of mean relative humidity, saturation deficit, and 1000 mb zonal and meridional wind components. Interestingly, in all cases, the forecast time that was selected for humidity variables by their screening procedure was *later* than the MOS forecast valid time. By using the MOS methodology, they were able to detect a phase error in the model humidity field.

A defect in their scheme was use of a coded cloud amount rather than a categorical cloud amount. By 1974, Glahn realized this error (Refs. 14 and 15) and switched to a REEP (Regression Estimation of Event Probability) forecast of categorical cloud amount (see Section 3.1 for a brief description of the REEP technique). He also dropped the surface observation as a predictor and reported on a problem with choosing the category with highest predicted probability, namely, biased forecasts. By 1976 (Ref. 7), Carter and Glahn reported a solution to the bias problem: inflation of variance followed by multiplication by the "minimum bias matrix" adjustment factor.

In a 1978 AFGL technical report (Ref. 16), Glahn reported on a design of an automated MOS system for the Air Weather Service. In 1985, Perrone and Miller (Ref. 21) reported that the generalized exponential Markov (GEM) model compared favorably with MOS for cloud amount forecasts. In 1988, Brunet et al. (Ref. 4) reported that for short-range cloud amount forecasts in Canada, perfect prog was superior to model output statistics when the surface observations were used as predictors. In 1989, the opposite conclusion was reached by Carter et al. (Ref. 8) for U.S. forecasts.

Finally, in 1990, Jacks et al. (Ref. 17) reported on a new MOS-based cloud amount forecasting system based on output from the NWS Nested Grid Model. Again, categorical cloud amount was forecast, but a new threshold technique was used to remove bias. Significant predictors included forecasts of mean relative humidity, relative humidity at various levels, vertical velocity, and 850 mb moisture convergence. In addition, Jacks et al. used a variety of "binary" predictors, i.e., predictors whose value is one if a certain criterion is met and zero otherwise. An example of a binary predictor is the relative humidity at the surface tested against various percentage thresholds.

2. DATA SOURCES

Forecasts from the Air Force Global Spectral Model and Real-Time Nephanalysis data were selected by the Government for use during model development in this effort. The RTNEPH data were also used as "ground truth" because they represent a unique, uniformly-derived representation of global cloud conditions. In addition to those two meteorological data sources, we used a terrain database compiled at Phillips Laboratory on the eighth-mesh grid that contains terrain type (land, water, coast) and elevation in meters at each grid point.

A large part of our MOS effort was spent obtaining, preparing, and validating our year-long developmental database. GSM data were first generated and then remapped to the RTNEPH grid system by government personnel at Phillips Laboratory. RTNEPH data were obtained directly through the Air Force Environmental Technical Applications Center (USAFETAC). Descriptions of the RTNEPH and GSM data sources are provided in the following sub-sections. Along with those descriptions, we also document the data processing and validation tasks that were an early part of the effort. All the figures for this section are included at the end of the section, starting on page 19.

2.1 RTNEPH GRIDDED CLOUD COVER

2.1.1 Data Description

The RTNEPH data archive has been maintained by the U.S. Air Force meteorologists since 1983 when the RTNEPH format replaced that of the previous 3DNEPH archive. Formats and procedures of RTNEPH are summarized thoroughly in Ref. 23, but documentation cannot capture the extreme complexity of the procedures and the degree of human interaction required to produce this massive data set operationally on a 3-hour cycle without interruption. The RTNEPH archive is produced by Air Force meteorologists at GWC and at USAFETAC's Operating Location A at Asheville, NC. A complete global RTNEPH analysis is provided every three hours in the northern hemisphere and every six hours in the southern hemisphere using data from two sources: the DMSP satellites and the Automated Weather Network (AWN).

AWN data include surface, upper air, and aircraft pilot report information. These point data are of great value when and where available. The fractional cloud cover estimates by trained surface observers can be especially valuable in establishing the output

cloud cover information described below. Unfortunately, in the U.S. even these data are less valuable than they might be because Airways codes (clear, scattered, broken, and overcast) are used by surface observers instead of the "eighths" categories used in most locations world-wide. Since surface observers and reporting aircraft are not available around the clock or around the world to support RTNEPH requirements, it is on the satellite data that most of the data archive depends.

The Operational Linescan Scanners (OLS) on two DMSP satellites (F-8 and F-9) currently provide reflectance and radiance data with approximately 0.6 km visible and thermal IR resolution as they trace 101 minute, sun-synchronous orbits around the earth. A third satellite (F-10) is in a non-circular orbit and is used only occasionally at GWC. Most points on the earth are scanned four or more times per day by the two OLS instruments. Using thresholding techniques, cloud/no-cloud analyses are derived from the DMSP data and the fraction of cloudy pixels within the area represented by each grid point becomes the satellite-based estimate of cloud-cover fraction. To the extent possible, layered cloud information is extracted from the IR signals.

Figure 2 shows the importance of the time flag reported in the RTNEPH data. Due to the geometry of the satellite paths, certain areas of the globe are updated more often than others; the polar regions receive much more coverage than the equatorial regions. The upper curve in Figure 2 represents the time between consecutive DMSP observations as a function of latitude along a fixed meridian (30° E longitude) by the current constellation. As we will see in Section 2.1.3, the age of the data played an important role in the selection of model regions and times for which to perform the model development as well as in the predictor selection itself.

Probably the most critical step in the RTNEPH procedure is that of merging the two data sources to produce the best possible assessment at each grid point. For many grid points the satellite-derived values are the only data available. In some cases, human intervention, called "bogusing," can occur during RTNEPH generation especially at earth locations of particular interest or known to exhibit particular systematic problems (e.g., coastal areas with persistent low clouds, snow-covered ground, etc.).

Each eighth-mesh RTNEPH analysis produces up to 4096 grid points within each of 120 RTNEPH boxes (60 in each of the northern and southern hemispheres). Grid point spacing varies as a function of latitude from 51 km near the poles to 26 km near the equator. Figure 3 shows the 60 RTNEPH boxes that make up the northern hemisphere. In the figure, we highlighted the 20 boxes of interest in this effort as selected by government personnel.

Data items at each grid point include both quantitative information regarding cloud cover and qualitative information on weather conditions and data quality (see Table 1). We extracted total cloud amount information, the time flag, and eight diagnostic flags for our MOS development as described below.

2.1.2 Data Preparation

All data processing was performed on the computers at Phillips Laboratory. Complete RTNEPH analysis fields were received from USAFETAC on 9-track tape in box-time order. Each tape holds four months of data (reported every three hours: 0, 3, 6, 9, 12, 15 18, 21Z) for one box. The complete data set fits on sixty 9-track tapes (20 boxes, for 12 months). By selecting only certain fields from the total data set, we were able to reduce the number of tapes required for this feasibility study to seven.

Table 1 RTNEPH Gridpoint Information

Cloud Data
Total Cloud Amount (in percent of grid area covered)
For up to four cloud layers:
Cloud Amount (rounded up to next higher 5 percent)
Cloud Type
Cloud Base Height
Cloud Top Height
Weather Data
Present Weather
Surface Visibility
Data Quality Information
Time Flag for Data Age
Source Flag for Layer Data
Diagnostic Information ("best report" flags, viewing conditions, bogus flags, . . .)

The exact format of the USAFETAC tapes is described in the RTNEPH User's Handbook (Ref. 23). Briefly, the RTNEPH database consists of a number of variables pertaining to cloud cover (layer amount, type, and base and top heights as well as total cloud amount), qualitative information on weather conditions, and data quality indicators. For this effort we were only concerned with total cloud amount, to be used both as a predictor and as ground truth. We did not take layer information into account for a variety of reasons going beyond our desire to control costs and computational complexity. For example, it is

known that the RTNEPH analysis model is less accurate for middle layer clouds than for the highest and lowest layers which can be observed by satellites and human observers. Furthermore, many critical operational missions can be served with total cloud amount forecasts alone.

Along with header information (including box number, year, month, day, and hour), we extracted three single-byte variables for each eighth-mesh grid point. They are:

- total cloud cover amount (0–100%)
- time in hours of the newest data at the point (0–255)
- 8-bit diagnostic value, where each of the 8 bits represents one diagnostic flag as follows:

  bit 1 — point was bogused
  bit 2 — conventional report was included
  bit 3 — data from a conventional report was spread to this point
  bit 4 — visual satellite data were included in this analysis
  bit 5 — IR satellite data were included in this analysis
  bit 6 — low level cloud has been persisted past normal data cutoff time
  bit 7 — although the visible satellite observed cloudy, point was marked clear for lack of supporting data
  bit 8 — fog/haze superseded by other weather elements in present weather.

As we will discuss later in Section 3, we used the total cloud amount directly as a predictor and as ground truth. The data flags were used to eliminate old and/or questionable data points from the regression analysis.

2.1.3 Data Validation

RTNEPH data validation occurred during tape processing at Phillips Laboratory and later on TASC computers. Our goals during this task went beyond simply validating the data, to include data analysis. In validating the RTNEPH data we were able to detect patterns in the data and consider the effects of data timeliness, cloud cover distributions, and artifacts on the regression model development. Data validation and analysis consisted of the following components:

- check headers for correctness
- verify that data fields in adjacent boxes are continuous across box boundaries
- compare total cloud amount fields with daily weather maps

- visually analyze concurrent cloud cover, time, and diagnostic flags
- generate histograms of cloud cover, time, and diagnostic flags.

The last three components deserve further description.

Comparison with weather data—We compared cloud cover fields in box 45, which includes the U.S. east coast, with daily weather maps produced by the NOAA Climate Analysis Center for the same region over a three-month period (February-April 1989). Figure 4 shows an example in which a strong front is present in the daily weather map over the east coast stretching from Virginia to Nova Scotia. The clouds associated with that front are clearly visible in the corresponding RTNEPH data field. Over the three-month period we were able to qualitatively verify the RTNEPH analysis fields on days that had strong cloud signals.

Visual analysis—In this task we simultaneously displayed (for a selected box/time) gridded cloud cover amount, data age, and eight diagnostic flags. By viewing all ten data fields at one time, we were able not only to see artifacts in the cloud amount fields, but also to use the ancillary diagnostic data to determine the source of the artifacts. For example, linear artifacts are commonly found in RTNEPH cloud amount fields. By analyzing the time and diagnostic flags, it is clear that those linear artifacts frequently separate areas of distinct "age" and form the boundaries of satellite swaths through the region. Figure 5 shows an example of just this type of artifact.

The upper-left image in Figure 5 shows fractional cloud cover (black represents clear, white is used for overcast, and a gray scale is used to represent partially cloudy grid points). The next image to the right shows the relative age of the data used in the RTNEPH analysis. Here the gray scale changes from dark for newer data to light for older data. The next eight binary images are the eight diagnostic flags. In those images white represents 0 (the flag is not set) and black represents 1 (the flag is set). The 8 diagnostic flags are described briefly in Section 2.1.2.

Other common features found in the RTNEPH cloud amount fields are the star- and cross-shaped artifacts that are a sign of "spread" data. That is, cloud cover at one grid point is "spread" into adjacent grid points. Again, by viewing the diagnostic flags, data age, and cloud cover fields simultaneously, we were able to determine that this type of artifact is a characteristic of the RTNEPH analysis algorithms, not a data processing error. Figure 6 shows a case where spread data are clearly visible. The format of the ten images is the same as described above. We decided to remove all spread data points from our model development data set.

The visual analysis task was important not only to verify our data processing, but to understand the strengths and weaknesses of the RTNEPH data fields and the effects of both on regression model development. By viewing a large number of data fields we were able to qualitatively compare data from different regions, seasons, and time-of-day. We looked for the number and type of artifacts present, the relative percentage of surface observations and satellite data used, the age of the data and how it changed with time of day, the frequency of bogused data points and any missing data. From this analysis, we were able to estimate the potential limitations of the RTNEPH data both as a predictor and ground-truth data source, especially with respect to the age of the data in certain regions.

Histogram generation—This was the most quantitative and telling of the data validation and analysis tasks. Following directly from the qualitative visual analysis described above, we generated histograms for the cloud cover, time flag, and two of the diagnostic flags to better understand these data and their potential effects on model development. Figures 7 through 15 include histograms of fractional cloud cover and the data time flag for several boxes, seasons, and times that are representative of the many cases we looked at. These figures are discussed below (refer to Figure 3 for the geographic locations of each box).

We begin with a discussion about the time flag and its significance. A key element of short-range (temporal) forecasting is up-to-date *initialization data*. A key element of forecast validation and regression modeling is up-to-date *validation data*. Because we used the RTNEPH cloud cover fields in both roles, its temporal characteristics were especially important for model development.

In many regions of the globe, especially in the tropics and over water, the RTNEPH is almost solely driven by satellite observations at sparse intervals, thus there are large gaps in the data. These data gaps limited the number of regions and times for which we could develop meaningful models and from which we could draw meaningful conclusions regarding forecast quality based on comparisons with persistence.

Figure 7 shows the fraction of cloud data in the RTNEPH that is less than three hours old for box 34 (mid Pacific). This plot is representative of the problems in ocean regions: very few, if any, automated weather sites and scanty satellite coverage, resulting in large time gaps between cloud analysis updates. A plot of the percentage of new (i.e., less than three hours old) data in box 61 (including Colombia and Venezuela), shown in Figure 8, shows similar characteristics. Figure 9 shows a histogram of the data age for box 45 (east coast U.S. and western Atlantic) at 21Z. In this example, *less than 20% of the data are less than 7 hours old.*

On the other hand, box 30, which includes eastern Europe and western Russia, has a large automated weather network and more frequent satellite coverage. Thus, a large percentage of the grid points are "new," as shown in Figure 10. A histogram of box 44 data (central U.S. and Canada) shows similar results (see Figure 11). Those cases with the most up-to-date data will provide the best evaluation of the potential value of the MOS approach.

We found a wide variety in cloud cover means and distributions across the 20-box data set. Some of that variety can be seen in Figures 12 through 15. For example, Figures 12 and 13 (containing data for box 61 afternoon and evening, respectively) show the diurnal differences in cloud cover characteristics of the tropics. Figure 12 has almost 15% fewer grid points in the two end categories, clear and overcast, than Figure 13. Thus, a greater percentage of the grid points in the afternoon are partly cloudy which is characteristic of popcorn cumulus which depend strongly on diurnal heating.

Cloud patterns in the winter at midlatitudes (Figures 14 and 15) show a different diurnal signature whose dominant feature is the pronounced clearing at night. A unique artifact present only in the those boxes containing regions of the U.S. is the peaks in cloud cover at 25 and 75%. These two values correspond to the U.S. Airways codes for scattered and broken, respectively, and can be seen clearly in these two figures corresponding to cloud cover in box 44.

By analyzing these histograms and many others we were able to *quantitatively* verify the RTNEPH data. In addition, this analysis led us to concentrate our model development efforts on those regions and times for which we would best be able to evaluate the MOS approach with respect to persistence for 3-, 6-, and 9-hour forecasts. This precluded those regions and times for which little up-to-date cloud cover data were available, because in those cases, persistence dominated. As an extreme example, if all the grid points for a given valid time in a development data set had cloud amounts that had not been updated in over 9 hours, a comparison of the valid-time RTNEPH cloud cover to the initial-time (persisted) cloud cover would result in 100% skill for 3-, 6-, and 9-hour forecasts. Similarly, the MOS-produced forecasts would perform 100% because they also use persisted cloud cover in the predictor pool. Evaluating these forecasts would hardly provide a worthwhile or meaningful assessment of the MOS approach.

On the other hand, though evaluation requires only the most current data, operational forecast models must use whatever data are available at initialization time. Keeping both of these requirements in mind, we included only those data points whose valid-time RTNEPH age was less than three hours in model development and validation, but we had no cutoff age for the persisted RTNEPH cloud cover used as a predictor in the MOS model.

2.2 GLOBAL SPECTRAL MODEL OUTPUT

2.2.1 Data Description

The Air Force Phillips Laboratory maintains and runs a copy of the twelve-layer Global Spectral Model that was the primary source of model data for this effort. The GSM (Ref. 24) is based on integration of equations representing five prognostic variables:

absolute vorticity

divergence

temperature

surface pressure (logarithm)

specific humidity.

From these, four additional diagnostic variables are computed routinely: geopotential height, vertical velocity, and the u (zonal) and v (meridional) wind components.

The vertical coordinate used within the model is the normalized pressure or "sigma" coordinate. Typically, model variables are interpolated to constant pressure mandatory levels for output as shown in Table 2. However, in this effort we were provided with wind, temperature, and humidity variables for sigma layers as used in the model, thus avoiding errors introduced by interpolating from sigma to mandatory levels. The thirteen sigma levels that define the twelve sigma layers used in the GSM are shown in Figure 16.

Moisture is represented in the GSM by the mixing ratio, or equivalently, specific humidity (one of the prognostic model variables). Evaporation from the oceans is represented in the model including the impact of the surface wind speed on the evaporation rate. Evaporation from land, however, is not represented and is a known shortcoming of the present model.

Spatial resolution of the GSM is limited by the 40-wave truncation of the spherical harmonics used to represent the meteorological fields at each pressure level. The 40-wave model can support resolutions down to 250 km near the equator, but it is often evaluated on a 2.5° by 2.5° grid (about 280 km x 280 km at the equator). For this effort we obtained GSM outputs on the half-mesh grid (approximately 100 km grid spacing near the equator) that were produced by interpolating from the 2.5° by 2.5° grid to the half-mesh grid. That interpolation was performed by personnel at Phillips Laboratory.

Table 2 Air Force GSM Forecast Fields

FIELD	SURFACE	Gradient Level	1000 mb	850 mb	700 mb	500 mb	400 mb	300 mb	250 mb	200 mb	150 mb	100 mb	1000–850 mb	Tropopause
Sea-level Pressure	X													
Heights (D-Value)			X	X	X	X	X	X	X	X	X	X		
Temperature	X		X	X	X	X	X	X	X	X	X	X		
Temperature Advection													X	
U-component Wind		X	X	X	X	X	X	X	X	X	X	X		
V-component Wind		X	X	X	X	X	X	X	X	X	X	X		
Vertical Velocity			X	X	X	X	X	X	X	X	X			
Relative Humidity			X	X	X	X	X							
Specific Humidity	X		X	X	X	X	X							
Dewpoint Depression			X	X	X	X	X							
Vorticity				X		X		X						
Vorticity Advection						X								
Stream Function			X	X	X	X	X	X	X	X	X	X		
Tropopause Pressure														X
Tropopause Temperature														X
Tropopause Height														X

The GSM generates forecasts every 15 minutes out to 24 hours and is initialized twice a day at 0Z and 12Z. For this effort we obtained forecasts every 3 hours to match the temporal resolution of the RTNEPH data. We obtained both the 0Z- and the 12Z-initialized model runs every day of 1989 for the twenty boxes shown in Figure 3. For this study GSM model runs were initialized with historical HIRAS data and thus were identical to operational model runs at Global Weather Central for that same period.

2.2.2 Data Preparation

The Global Spectral Model runs were performed on the Cray2 at the Weapons Laboratory in Albuquerque, NM by Phillips Laboratory personnel. In a post-processing step, spectral outputs of the GSM were converted to latitude-longitude coordinates, vorticity and divergence values were translated to u and v wind components, and specific humidity was transformed to relative humidity. Finally, a nine-point quadratic interpolation was used to interpolate from latitude-longitude coordinates to the half-mesh RTNEPH grid. These data were then passed on to TASC for further processing.

As part of our data processing task, we sorted the processed GSM outputs from synoptic order to box/time order using the computers at Phillips Laboratory. We used box/time arrangement throughout this effort because it was the most convenient for developing regional models. During the sorting process we also scaled the 32-bit data values at each grid point to 16-bit integer values to save space and speed up processing, and yet retain the desired accuracy.

The resulting data files contained a header for each forecast time that included: box number, year, month, day, model initialization time (0Z or 12Z), forecast time (3-, 6-, . . ., 24-hour forecasts), and sigma layer heights. The remainder of the data files contained the following gridded variables on the half-mesh grid at up to ten vertical levels: u wind (10 sigma levels), v wind (10 sigma levels), temperature (10 sigma levels), relative humidity (lowest six sigma levels), surface pressure, and surface height.

Because our model development data set contains model runs initialized both at 0Z and 12Z, and because each run produces forecasts out to 24 hours, we have two forecasts available at any given valid time. For example, for a 3Z valid time, we have the 3-hour forecast from the 0Z model run of the same day and the 15-hour forecast from the 12Z model run of the previous day. However, both of these forecasts are not always available due to operational schedules. In our example of a 3Z valid time, output from the model run initialized at 0Z would not be immediately available at 0Z in order to make a 3-hour forecast valid at 3Z. Similarly, there would be no 0Z-initialized data available for the 6- or 9-hour forecasts valid at 3Z. Figures 17 through 20 show time lines for each of four forecast valid times (0, 3, 6, and 9Z) and the GSM data available for the 3-, 6-, and 9-hour forecasts at each time. Similar timelines can be made for forecast times 12Z through 21Z in which the 0Z and 12Z GSM data sources are switched.

For any given box, season, and valid time, a single model development data file was built that included the GSM data for the valid time, the corresponding RTNEPH cloud amount and diagnostic fields, along with RTNEPH cloud cover from the forecast initialization time back to 24 hours before the valid time to use for persistence.

2.2.3 Data Validation

Data validation tasks for the Clobal Spectral Model data fields were similar to those we performed for the RTNEPH data. Briefly they were:

- check headers for correctness
- verify that data fields in adjacent boxes are continuous across box boundaries

- visually analyze data fields
- generate histograms of model variables.

The last two tasks are described in more detail below.

Visual analysis—In this task we displayed the u, v, temperature, relative humidity, surface pressure, and surface height fields for many selected boxes and times. We displayed all layer data at one time to verify continuity in the vertical dimension. All data were displayed as gray-scale images. We analyzed the images for potential artifacts and overall structure. We did not find any questionable data fields.

Figures 21 through 23 show examples of the images analyzed. The first of those figures shows the u field for a specific day and time in box 30. The upper-left image in the sequence is the surface-level u field and each consecutive image shows the u field at the next higher sigma layer. We use a gray scale to show the relative magnitude of the field value, where black is low and white is high. Figure 22 shows the similar images for the corresponding temperature field. Figure 23 continues with the same box and time and includes the relative humidity at the lower six sigma layers of the atmosphere and the surface pressure and height fields. (Note how the surface height field follows the elevation contours we expect to see for the region in Europe corresponding to box 30.)

Histogram generation—As a means to validate and better understand the MOS data, we generated histograms to show the overall range, mean, and distribution of each model variable in different locations and for different seasons and times. This activity naturally led to generating histograms for meteorological quantities derived from the MOS data as well, such as vertical velocity, divergence, etc. Example histograms of derived quantities are shown in Section 3.3. Here, in Figures 24 and 25, we provide examples of histograms of the GSM variables: u, v, temperature, relative humidity, and surface pressure for two different box seasons.

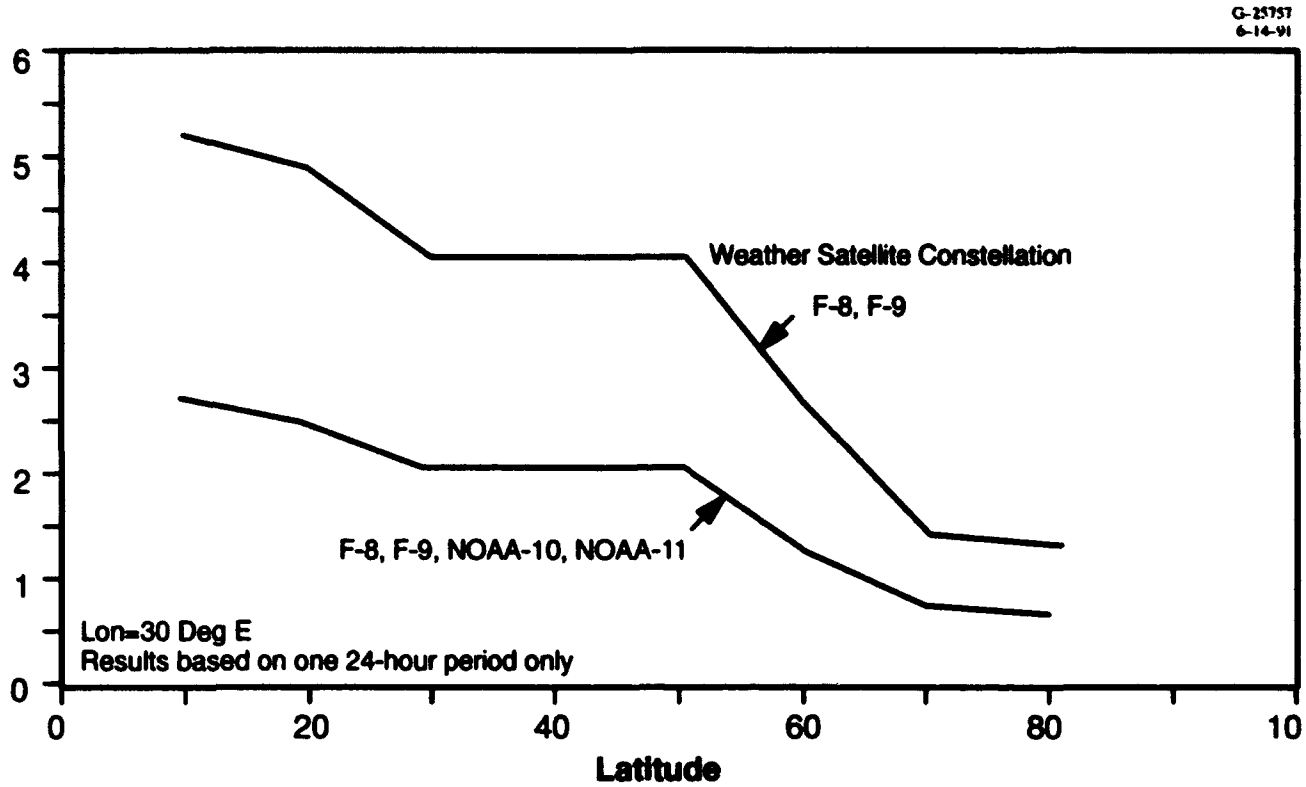


Figure 2 Satellite Weather Data Access Performance — Average Case Sample

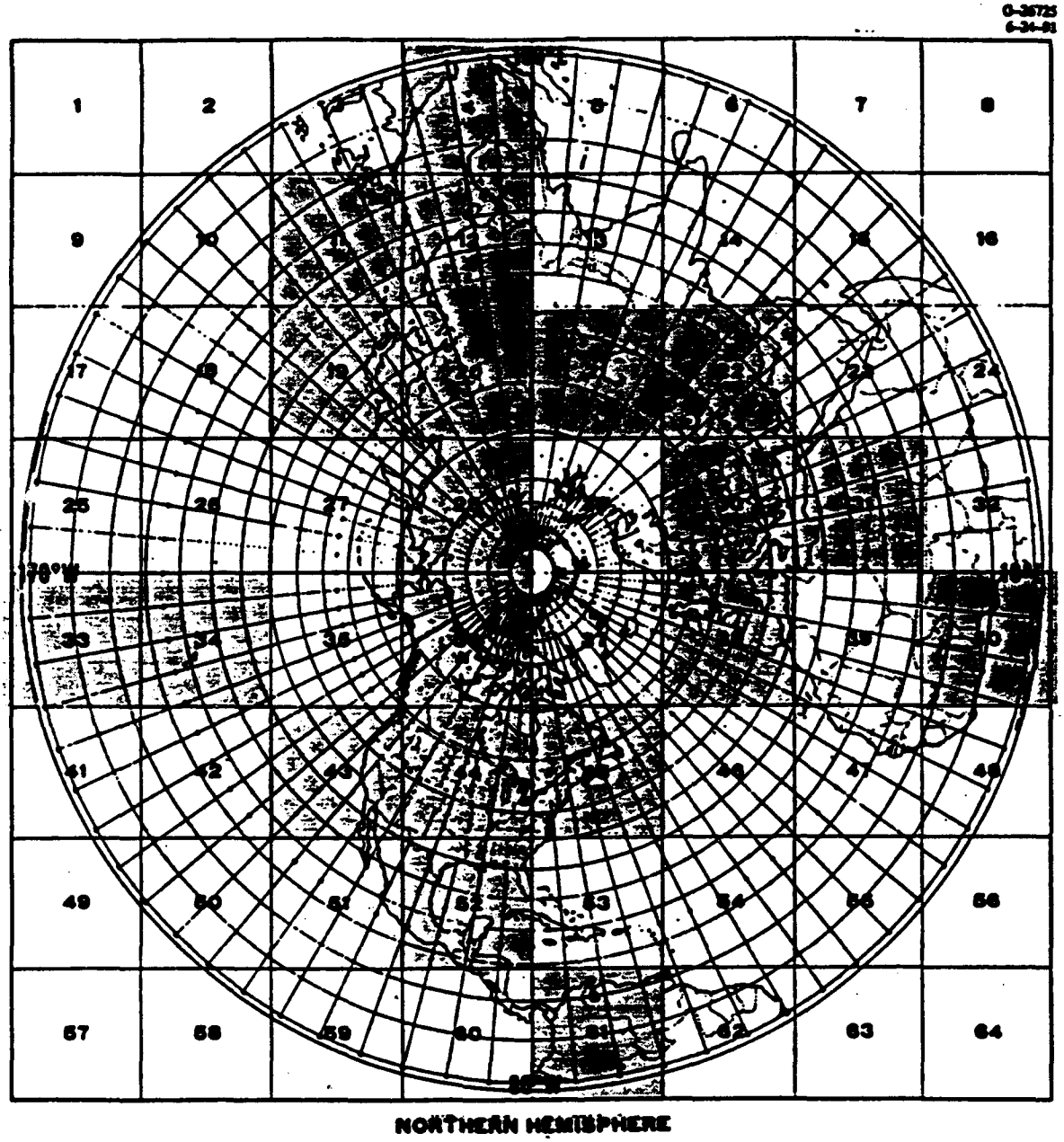


Figure 3 RTNEPH Boxes Selected for This Study

Figure 4 Coincident Surface Weather and RTNEPH Analysis for Morning of March 29, 1989 Showing Similar Cloud Cover Over the East Coast in the Common Region Outlined in Black

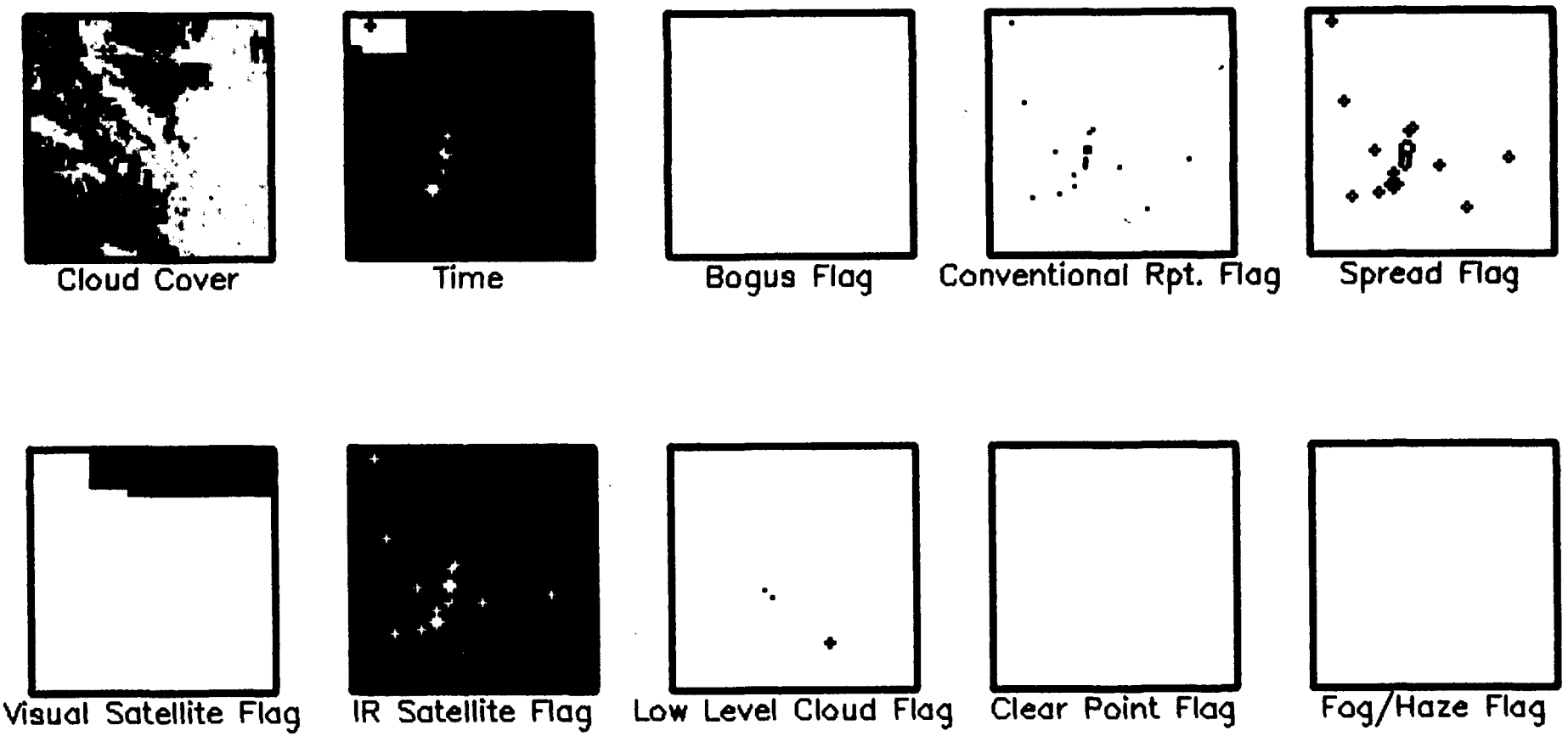


Figure 5 Gray-scale Images of Cloud Cover and Time Flag Along with Corresponding Binary Diagnostic Flags that Clearly Show the Presence of a Linear Artifact Due to Time Differences in the Satellite Data (Box 34, March 4, 6Z)

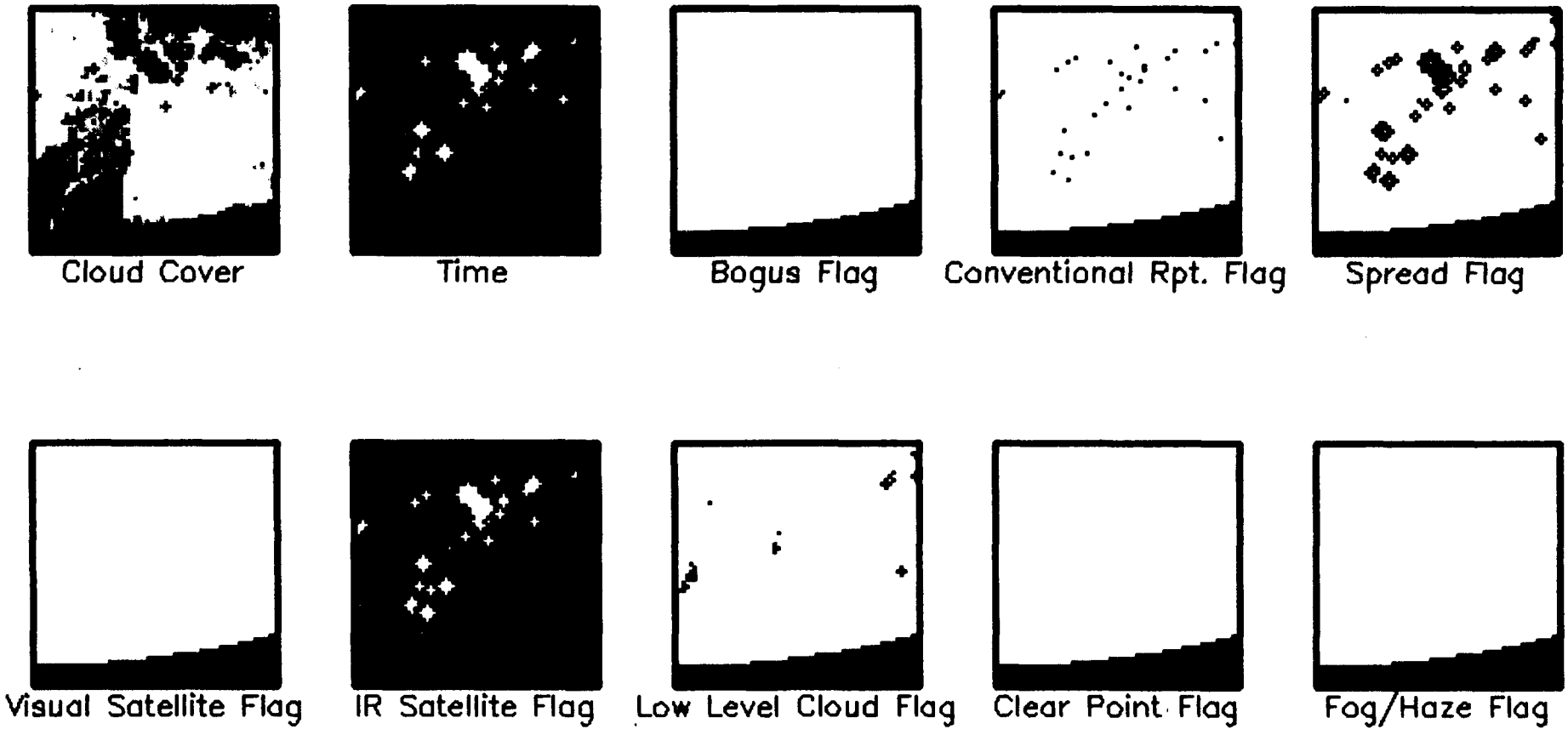


Figure 6 Gray-Scale Images of Cloud Cover, Time, and Diagnostic Flags for Dec. 1, 00Z in Box 61. The Black Strip of Data Along the Bottom of the Images Corresponds to "Off World" Data. A strong Linear Artifact is Present in the Cloud Cover Field Due to Time Differences. Spread Data Can Also be Seen

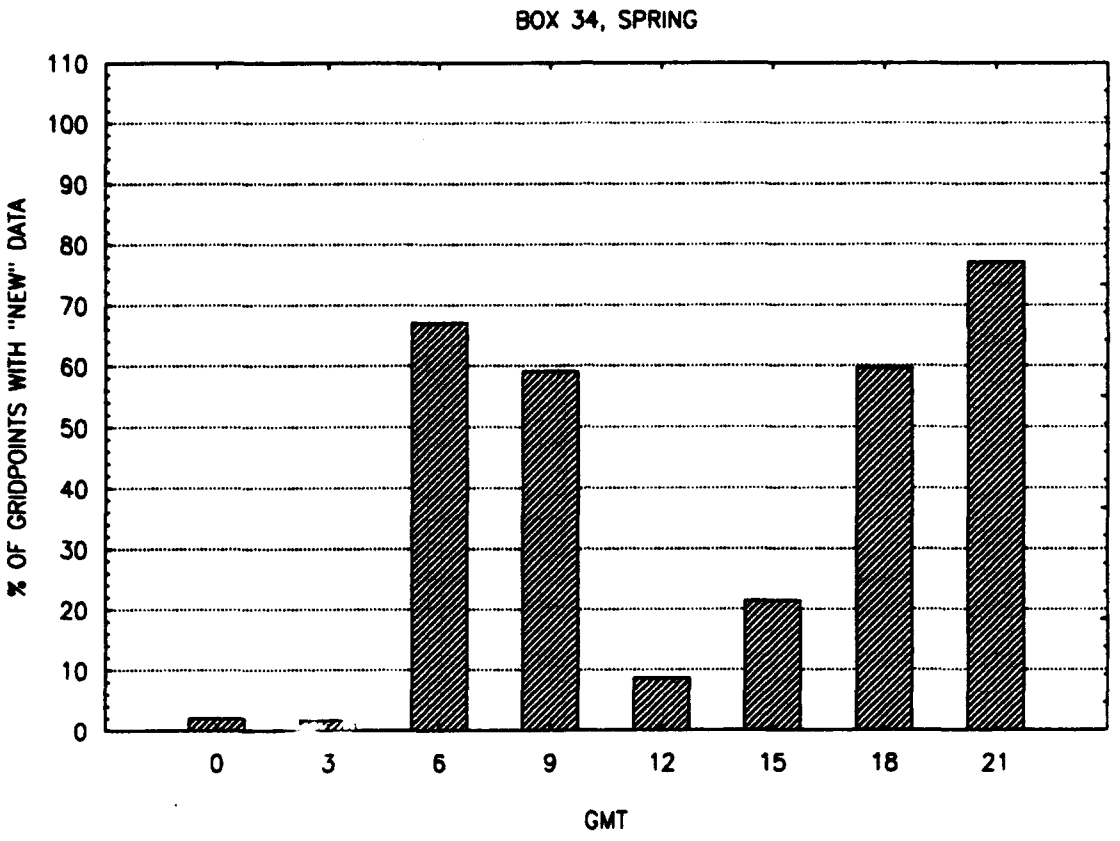


Figure 7 Histogram of the Percentage of Grid Points That are Less Than 3 Hours Old as a Function of Time of Day in Box 34, Spring

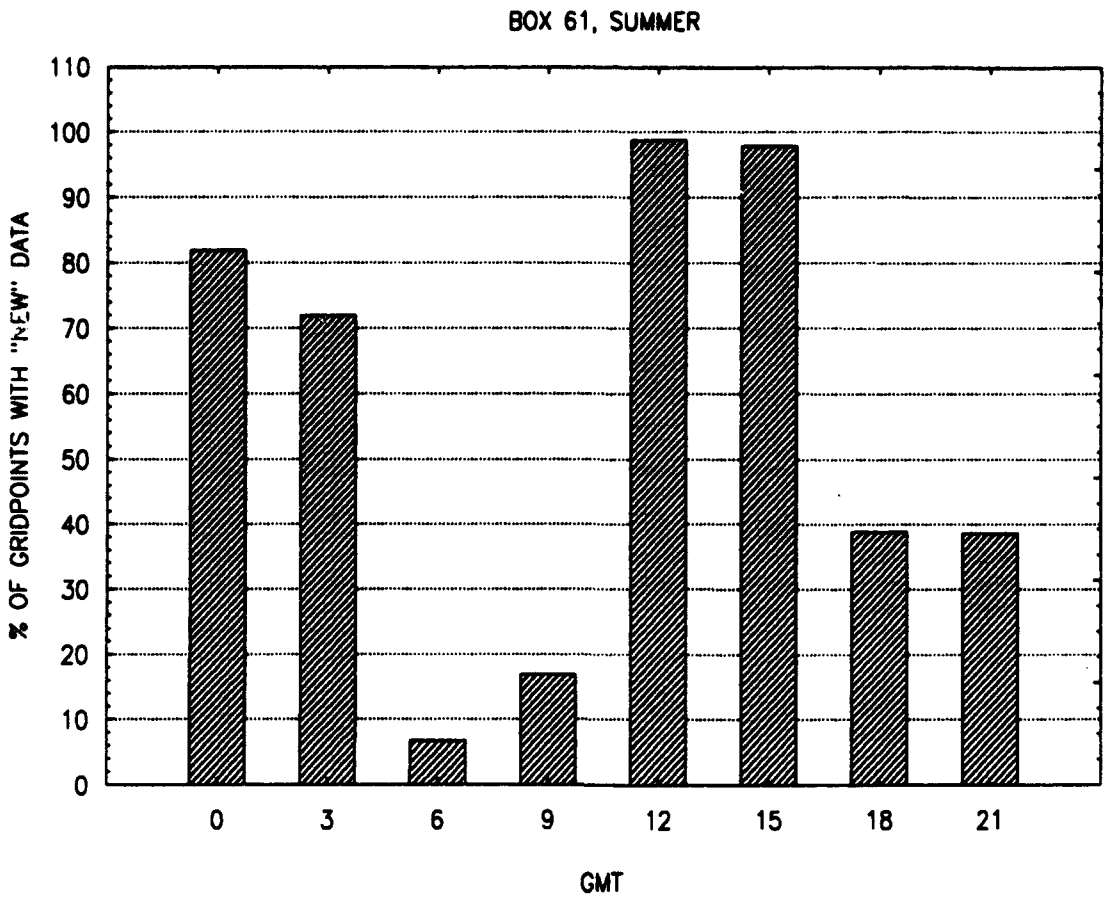


Figure 8 Histogram of the Percentage of Grid Points That are Less Than 3 Hours Old as a Function of Time of Day in Box 61, Summer

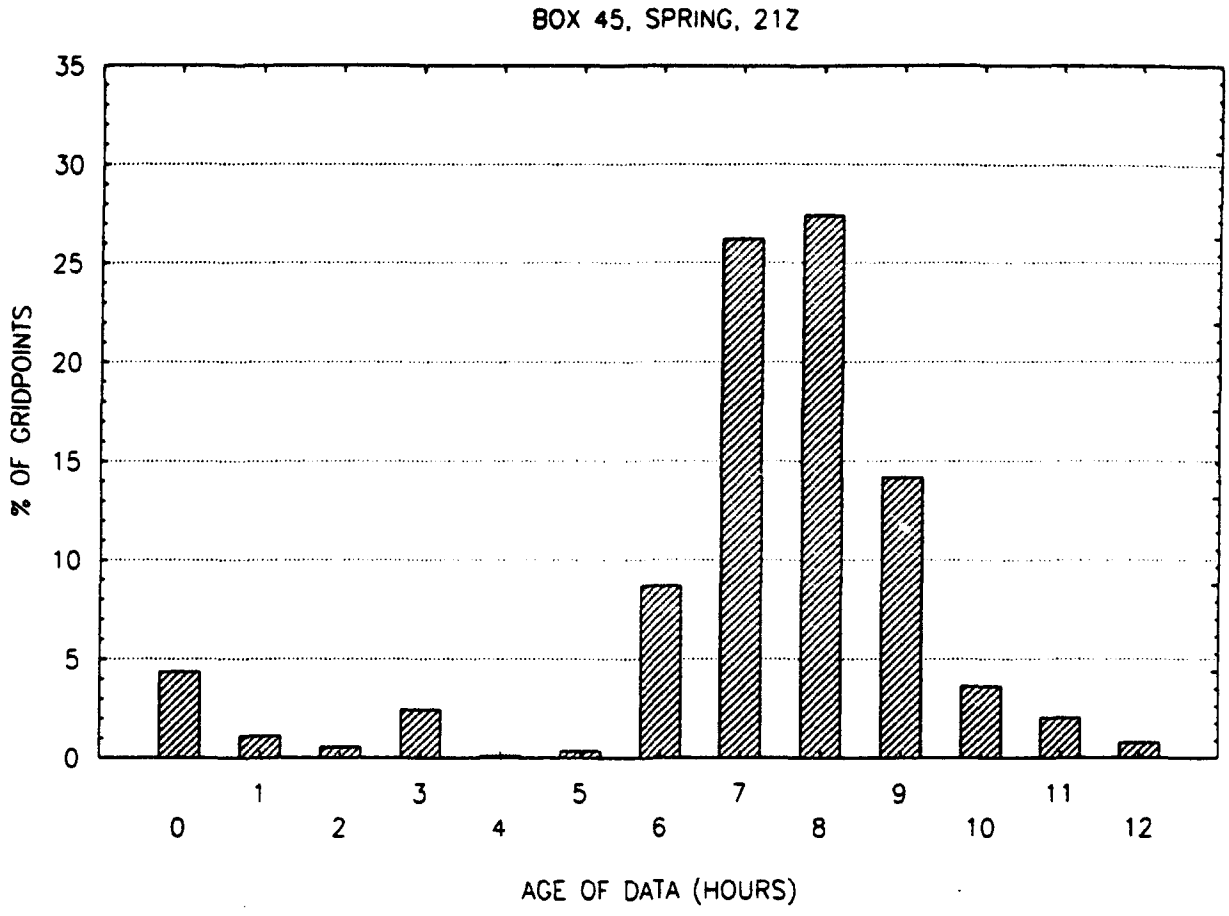


Figure 9 Histogram Showing the Percentage of Grid Points of Various Ages in Box 45 Over the Spring Season at 21Z

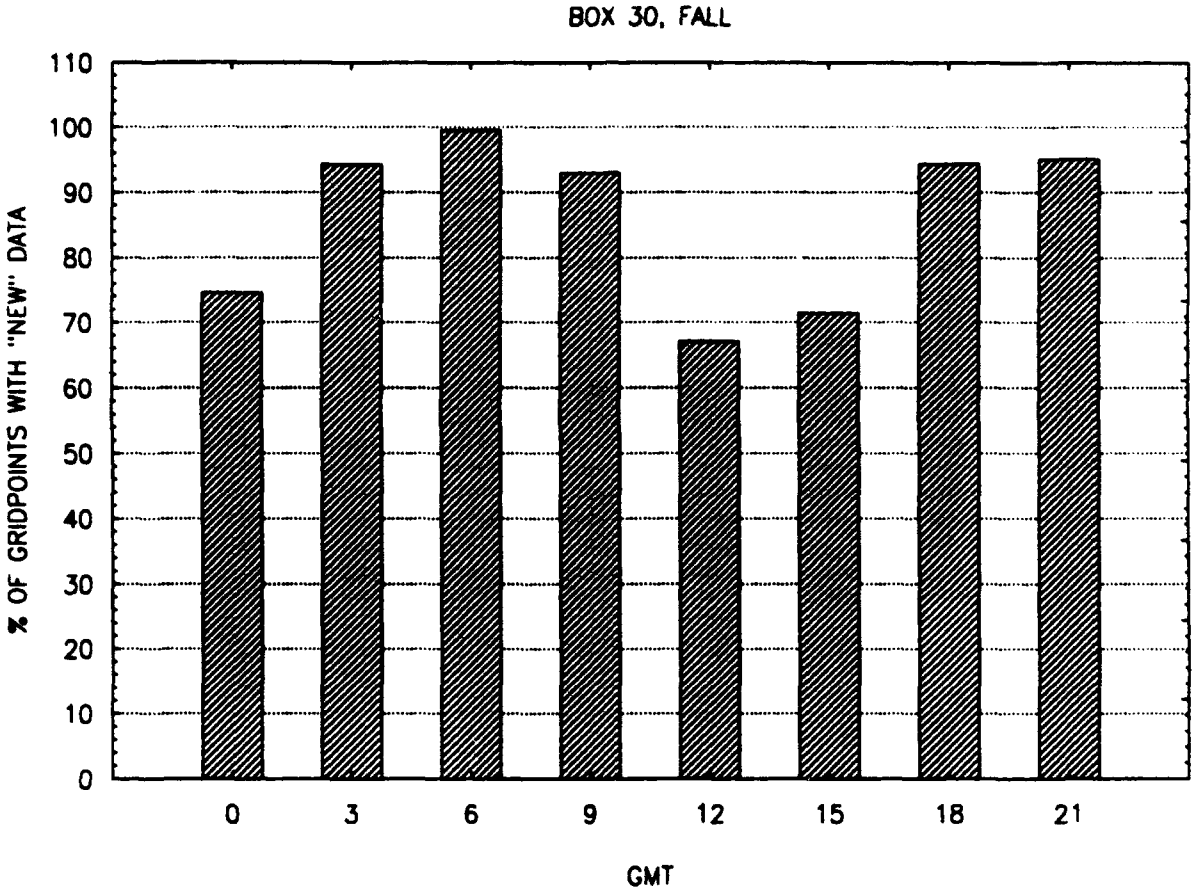


Figure 10 Histogram of the Percentage of Grid Points That are Less Than 3 Hours Old as a Function of Time of Day in Box 30, Fall

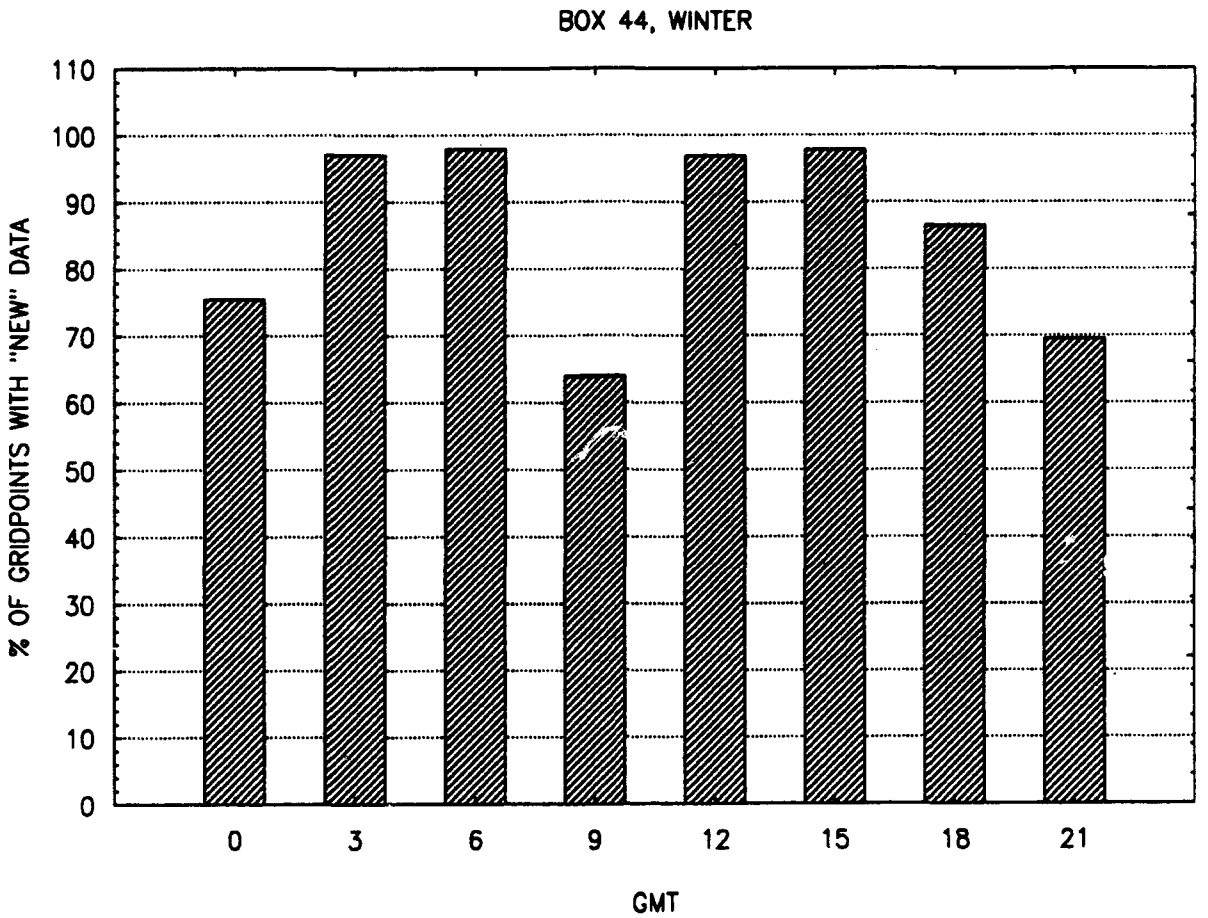


Figure 11 Histogram of the Percentage of Grid Points That are Less Than 3 Hours Old as a Function of Time of Day in Box 44, Winter

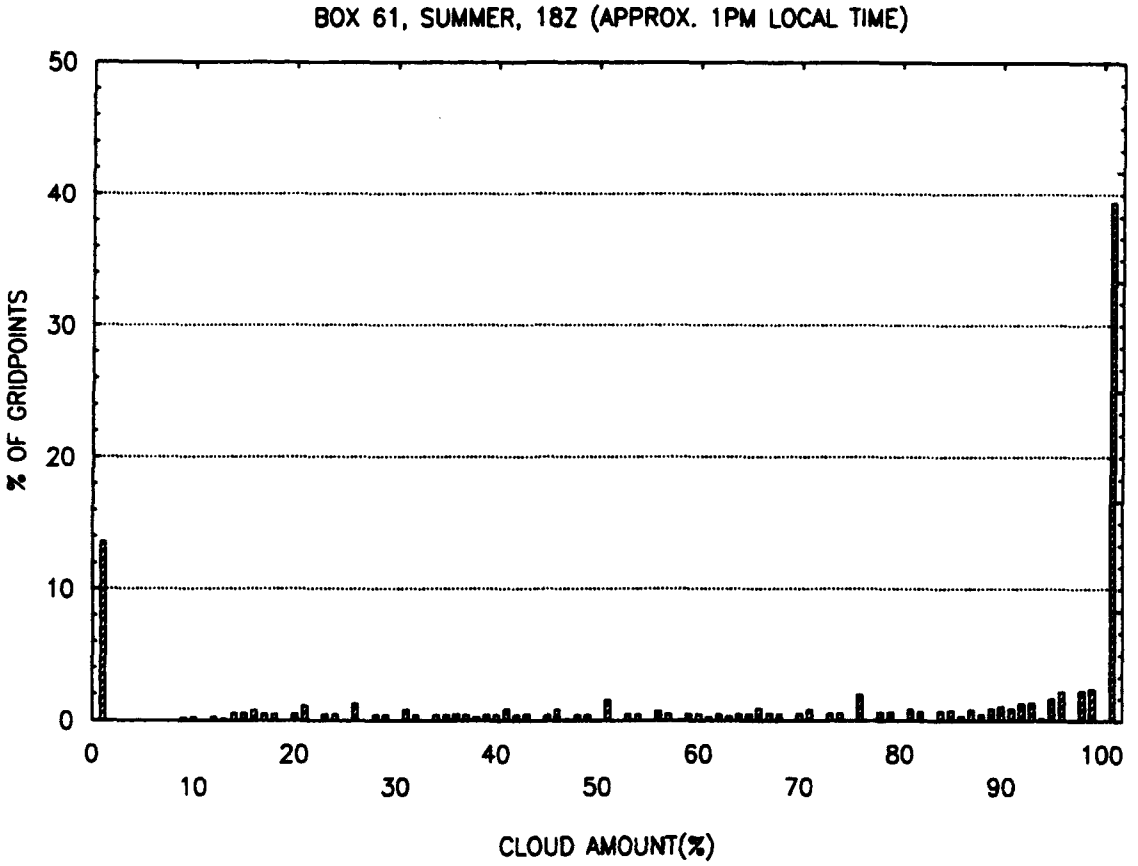


Figure 12 Cloud Cover Distribution for Box 61, Summer at Approximately 1 p.m. Local Time

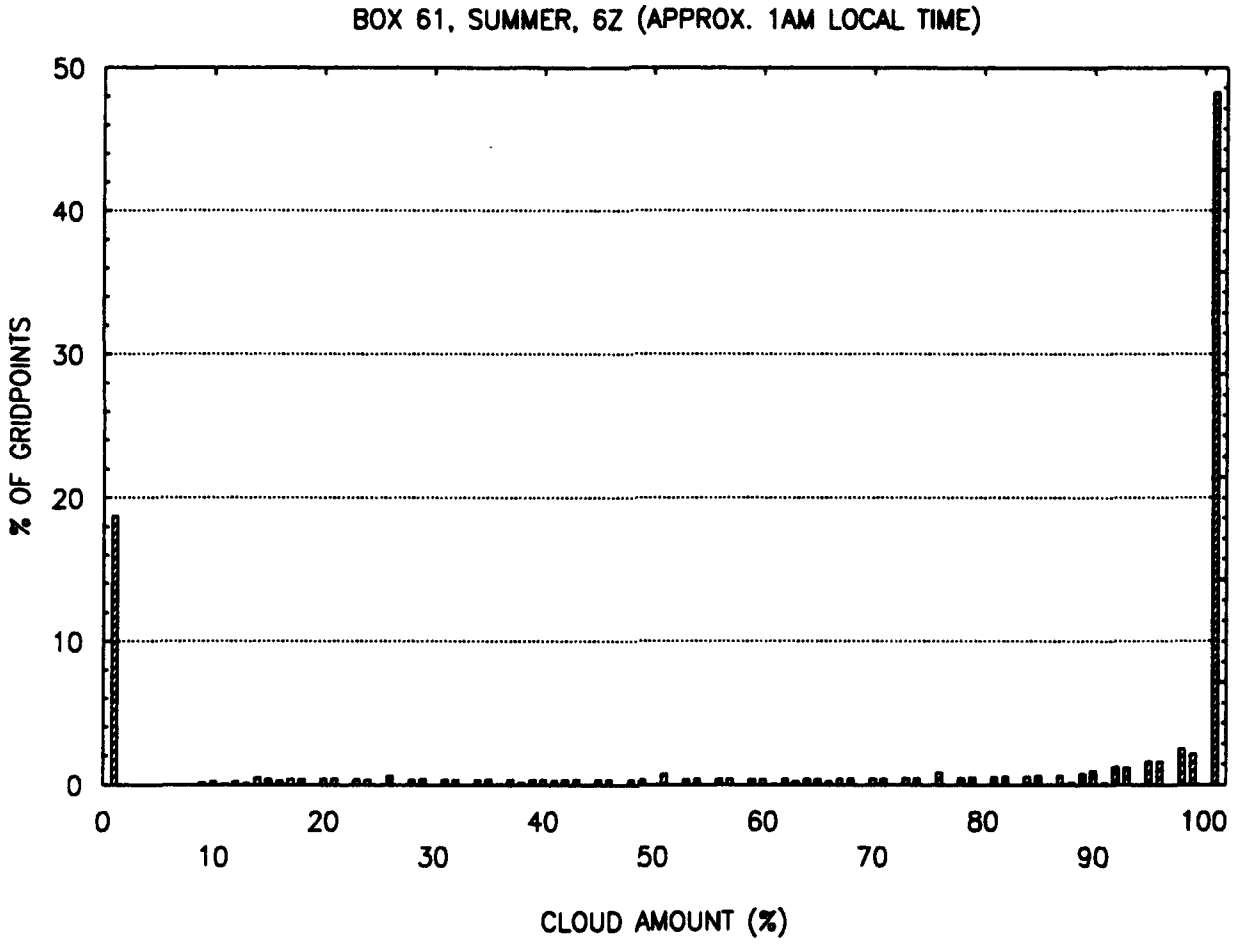


Figure 13 Cloud Cover Distribution for Box 61, Summer at Approximately 1 a.m. Local Time

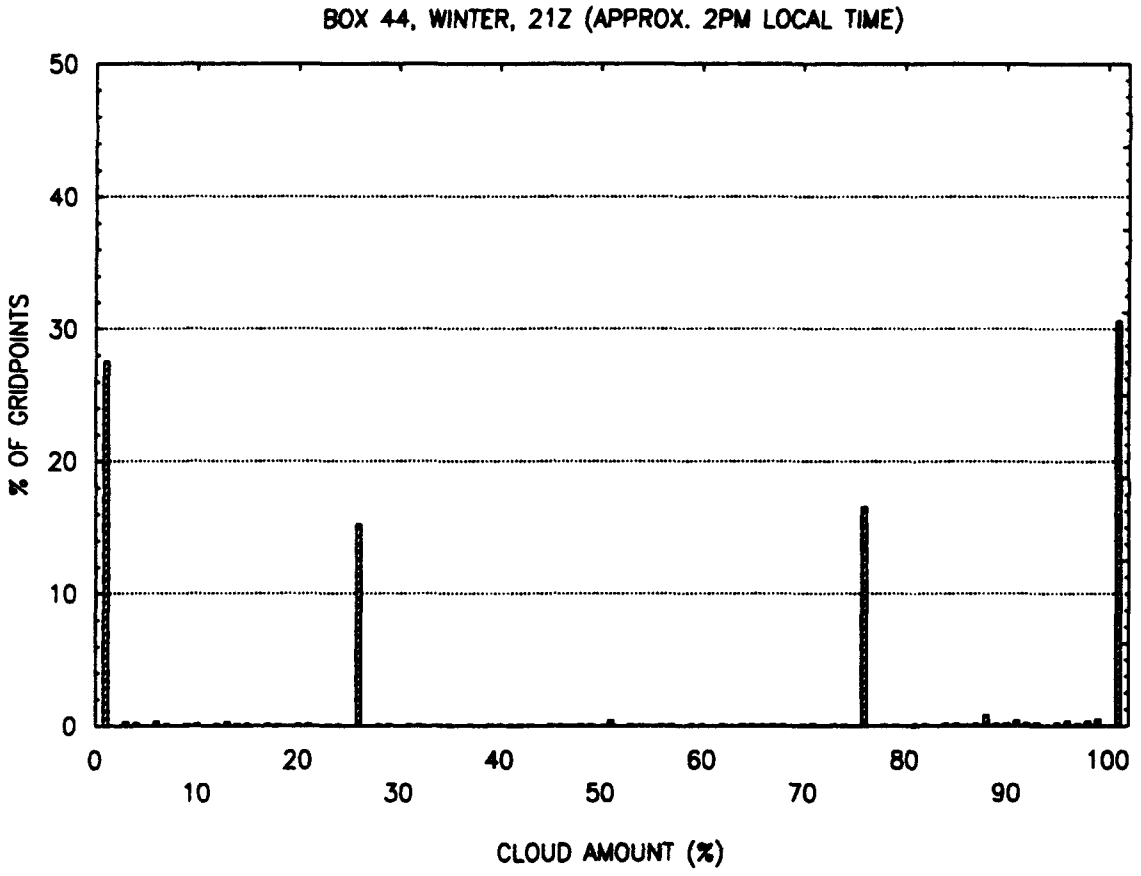


Figure 14 Cloud Cover Distribution for Box 44, Winter at Approximately 2 p.m. Local Time

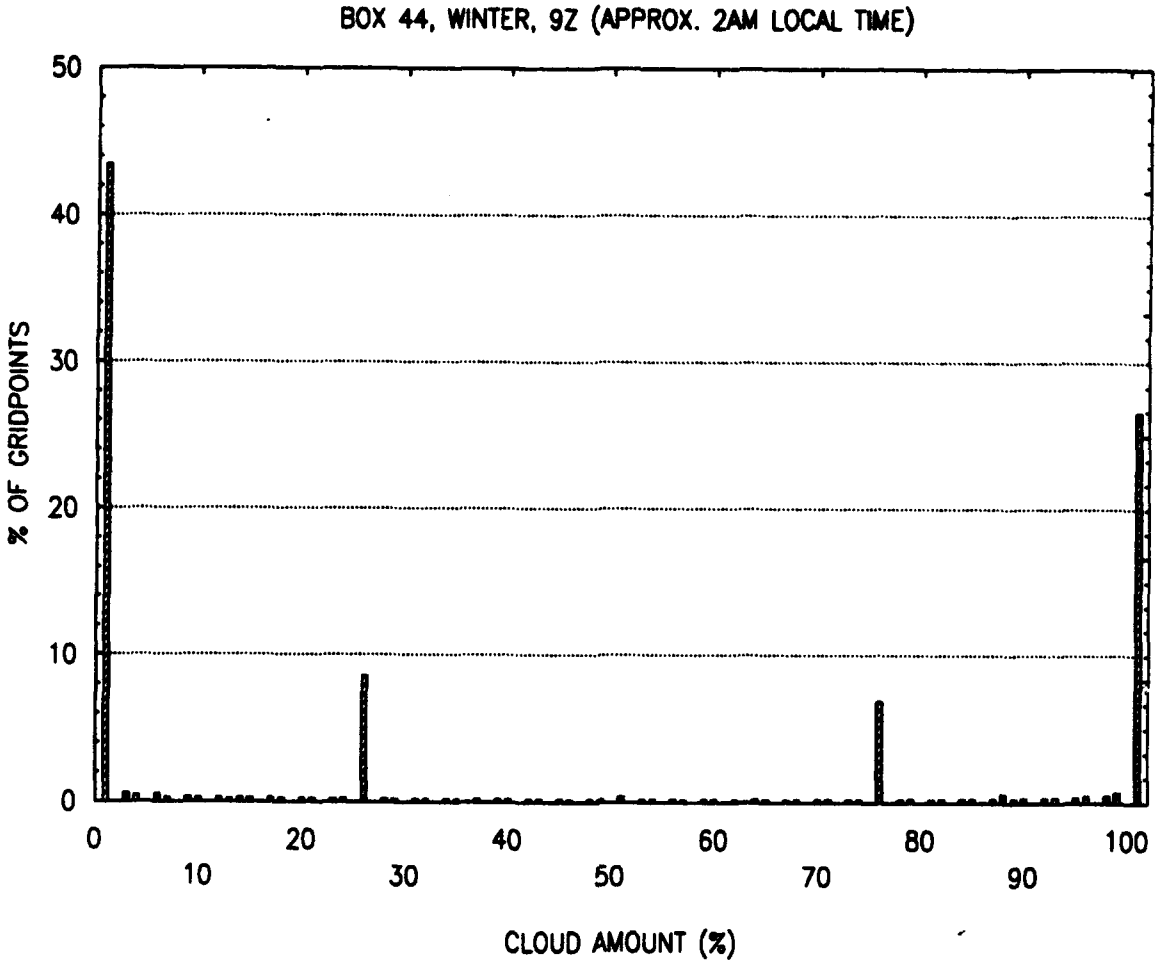


Figure 15 Cloud Cover Distribution for Box 44, Winter at Approximately 2 a.m. Local Time

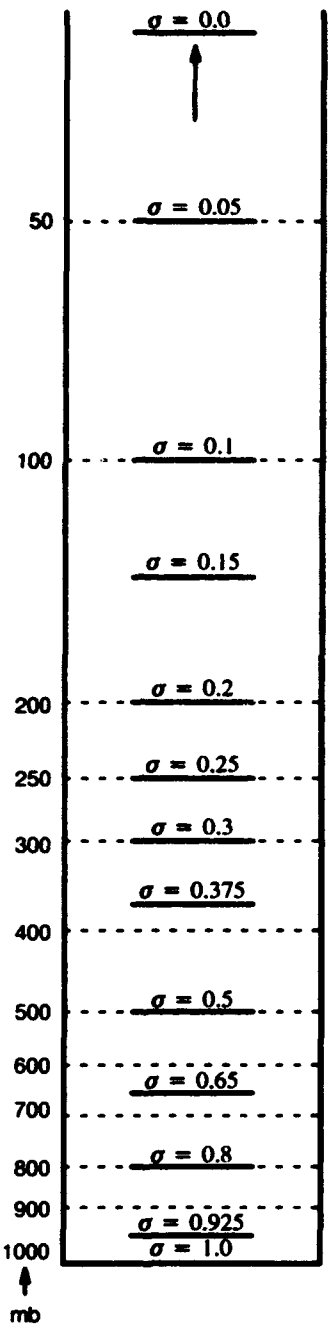


Figure 16 GSM Sigma Surfaces (Ref. 24)

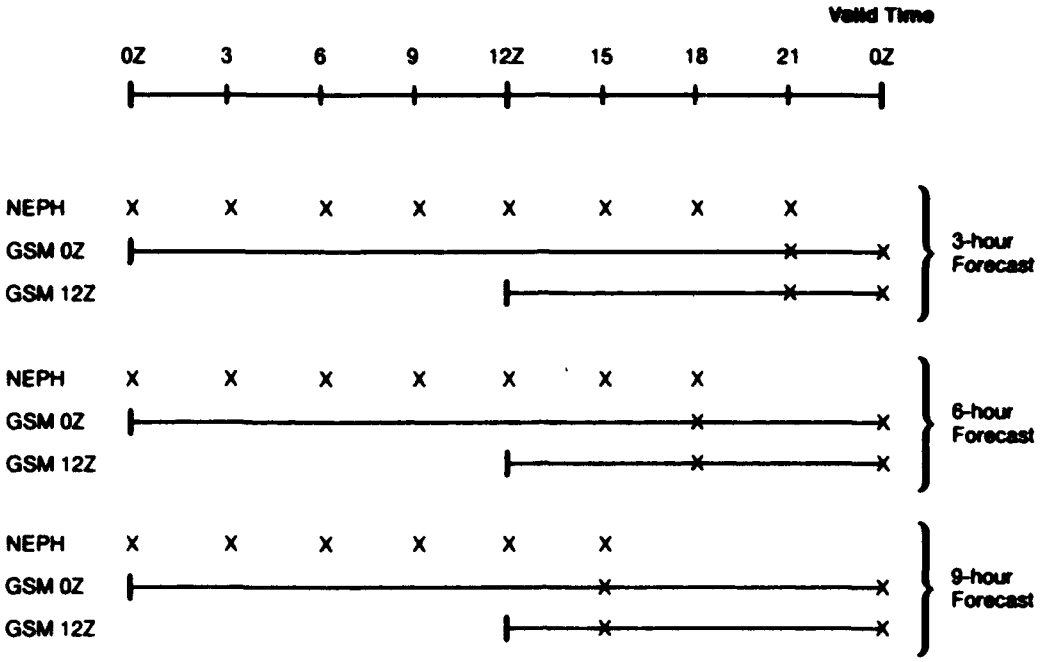


Figure 17 Schematic Showing Which RTNEPH and GSM Data Are Available for 3-, 6-, and 9-hour Forecasts Valid at 0Z ('x' Marks the Data Sources Used in Model Development)

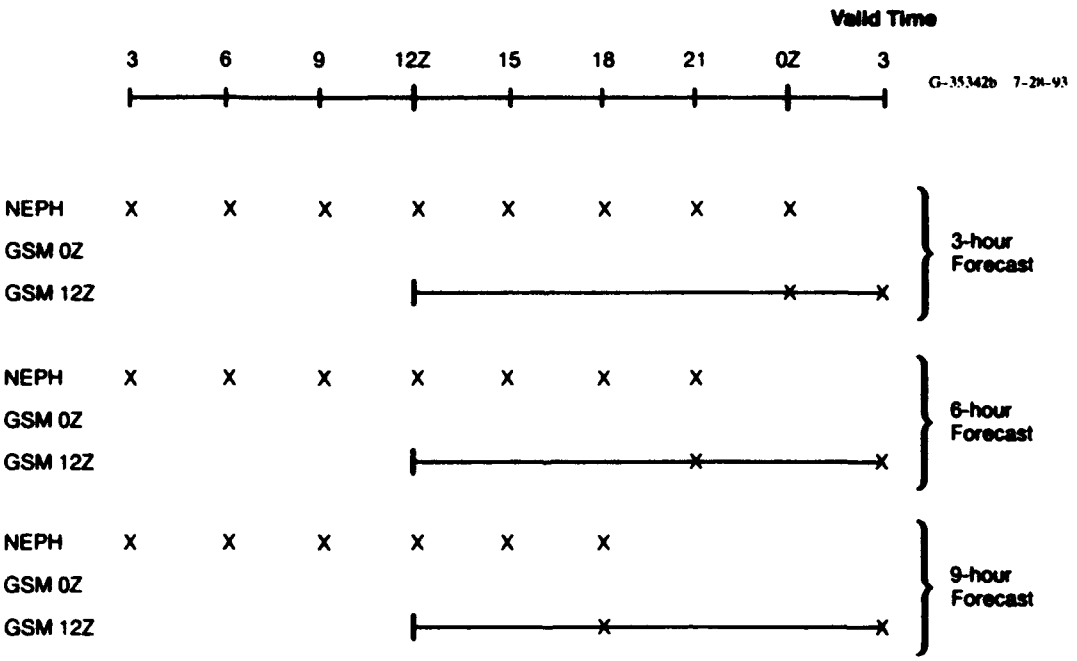


Figure 18 Schematic Showing Which RTNEPH and GSM Data Are Available for 3-, 6-, and 9-hour Forecasts Valid at 3Z ('x' Marks the Data Sources Used in Model Development)

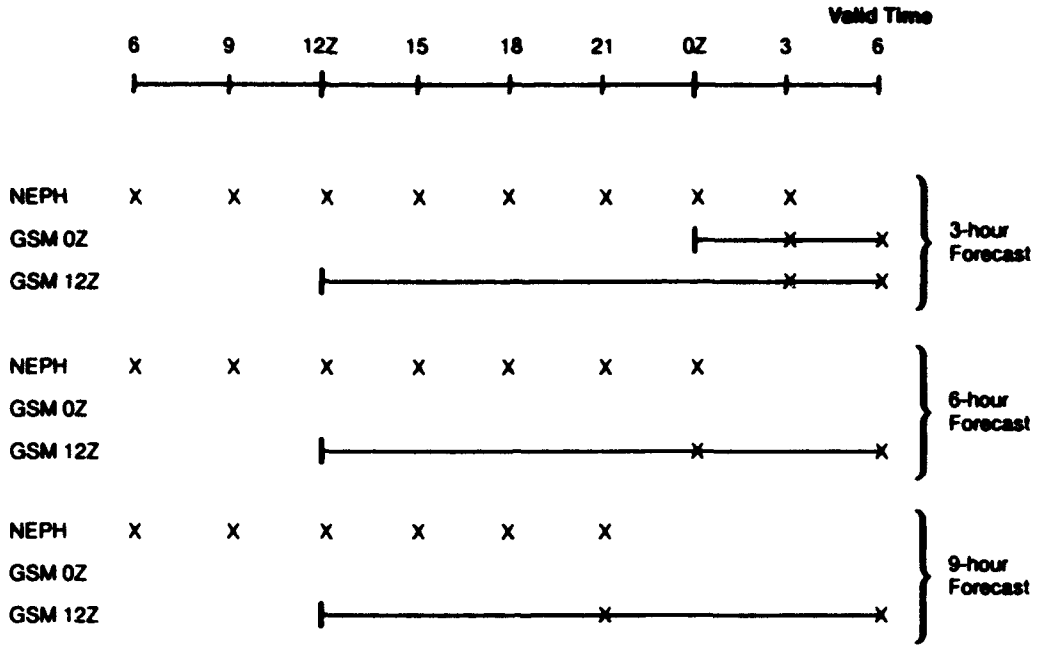


Figure 19 Schematic Showing Which RTNEPH and GSM Data Are Available for 3-, 6-, and 9-hour Forecasts Valid at 6Z ('x' Marks the Data Sources Used in Model Development)

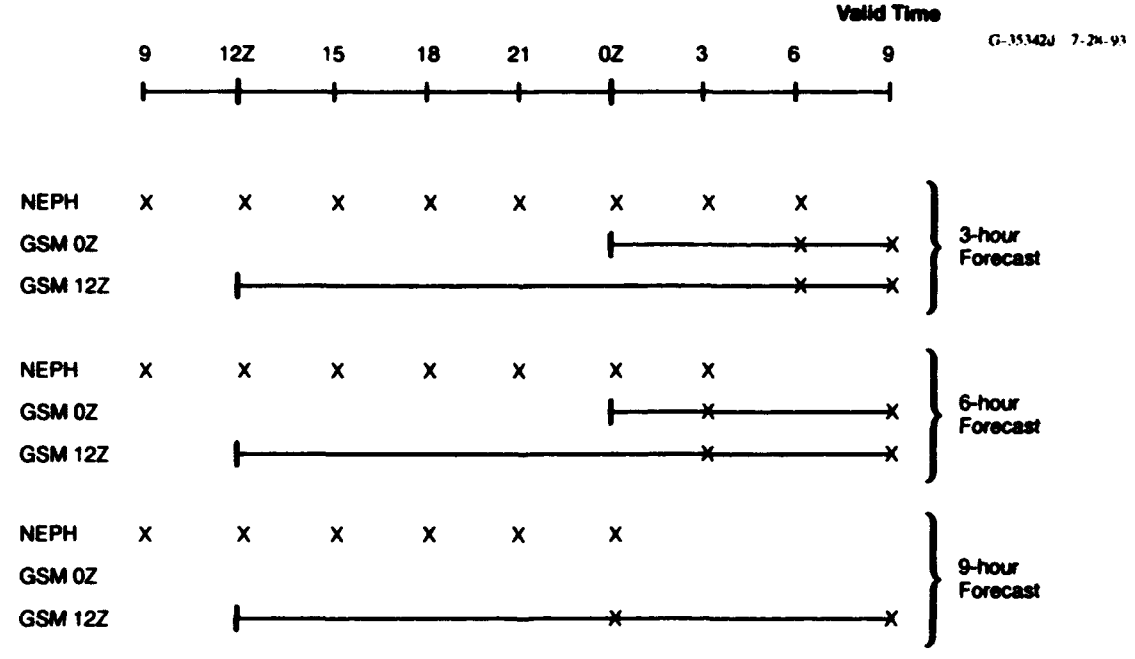


Figure 20 Schematic Showing Which RTNEPH and GSM Data Are Available for 3-, 6-, and 9-hour Forecasts Valid at 9Z ('x' Marks the Data Sources Used in Model Development)

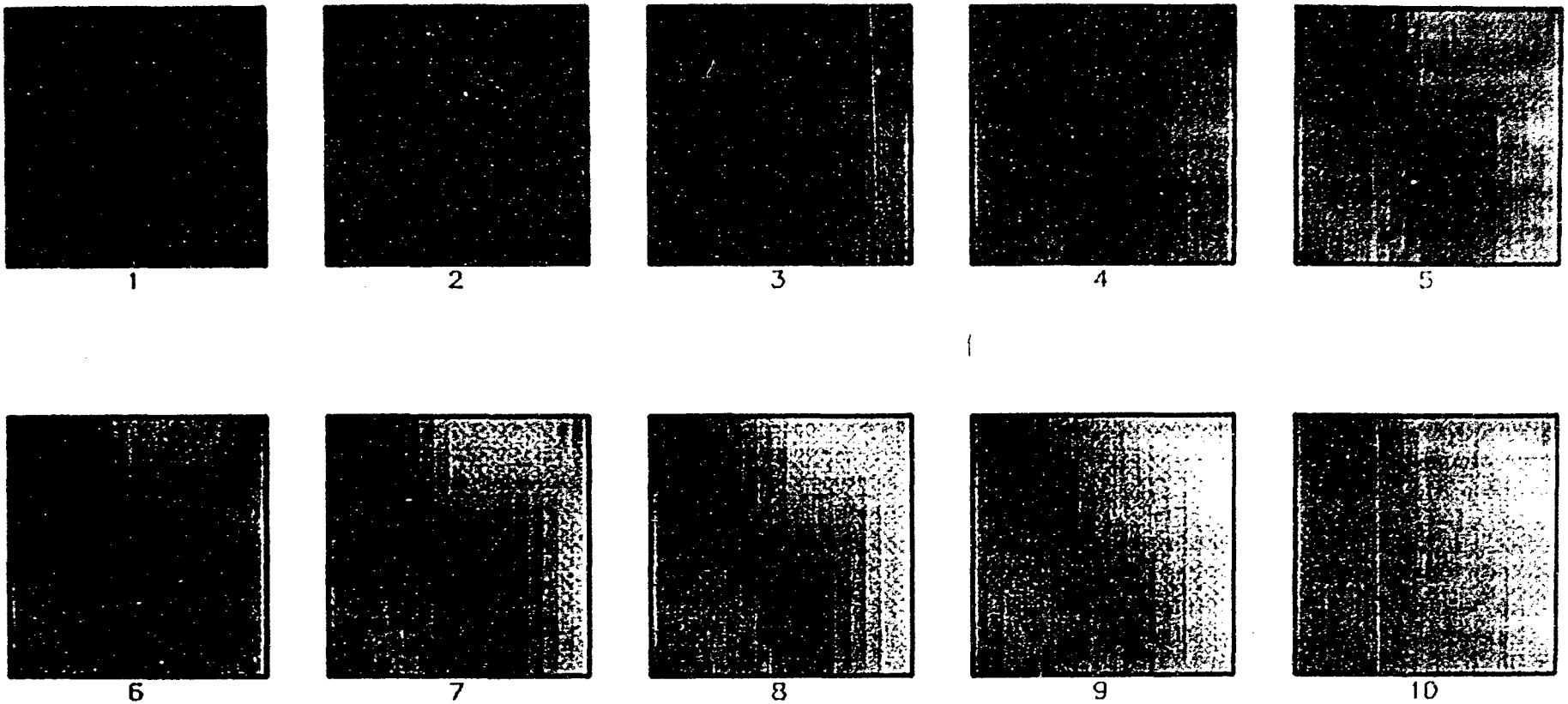


Figure 21 Gray-Scale Images Showing GSM Forecast u Wind Component at 10 Consecutive Sigma Layers (Layer 1 is the Surface and Layer 10 Corresponds to Approximately 100 mb) for Box 30, June 1, 15Z

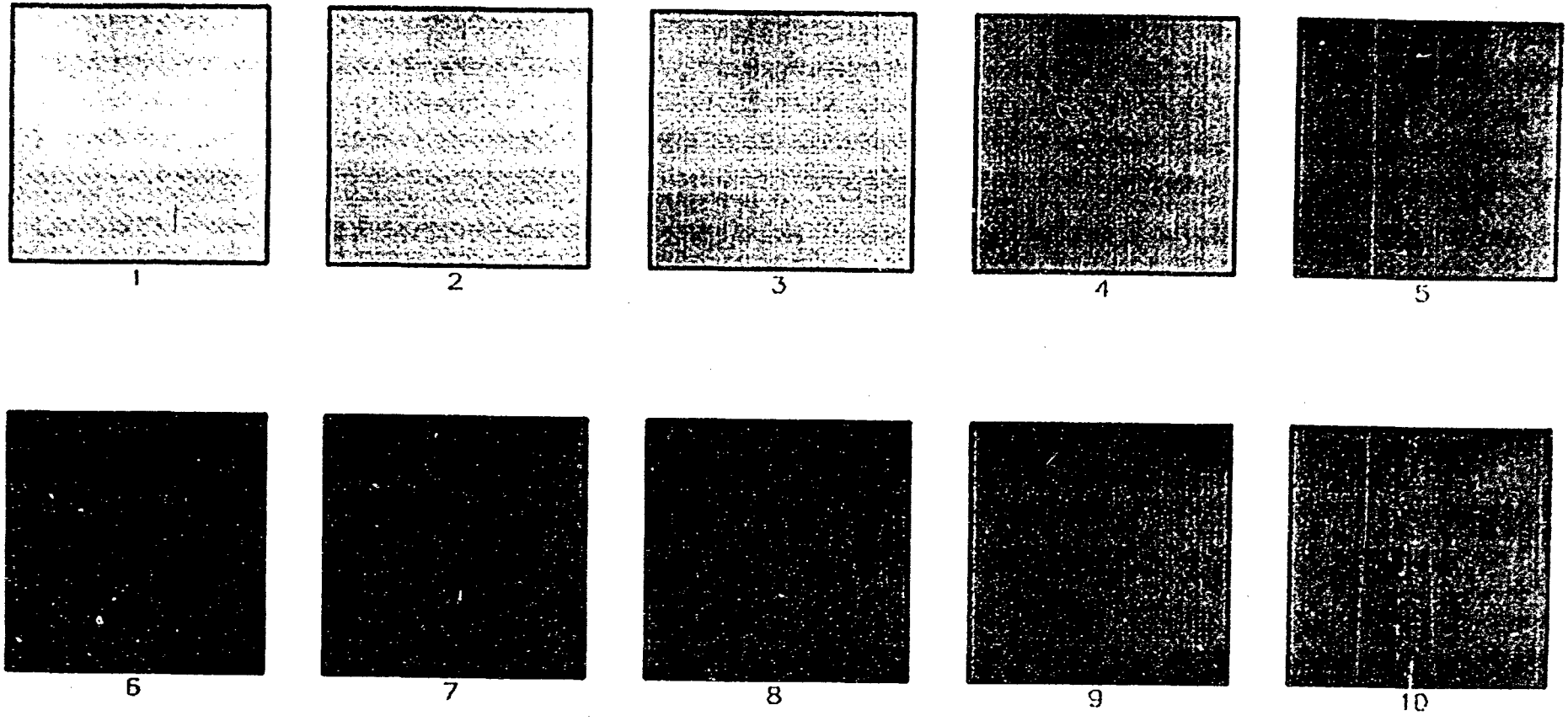


Figure 22 Gray-Scale Images Showing GSM Forecast Temperature at 10 Consecutive Sigma Layers (Layer 1 is the Surface and Layer 10 Corresponds to Approximately 100 mb) for Box 30, June 1, 15Z

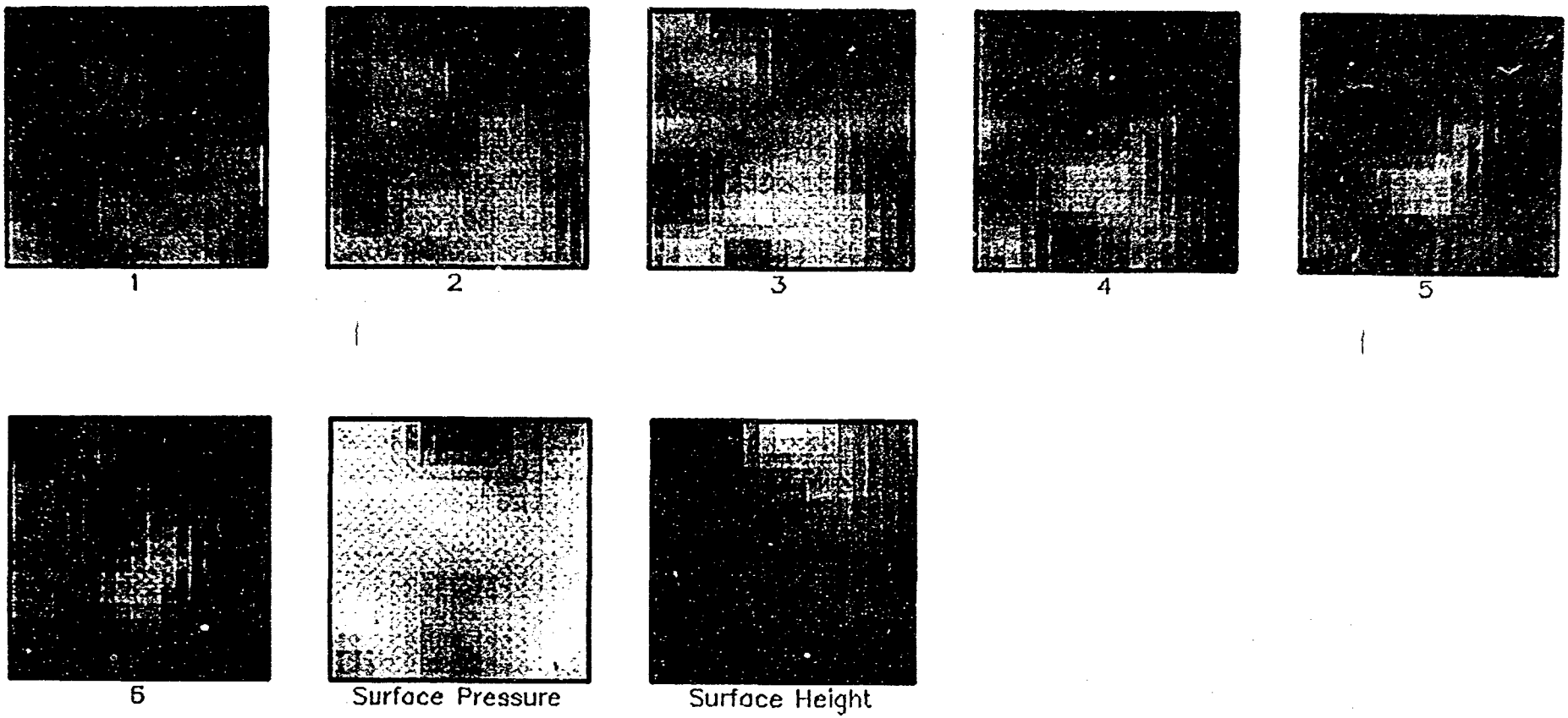


Figure 23 Gray-Scale Images Showing GSM Forecast Relative Humidity at 6 Consecutive Sigma Layers (Layer 1 is the Surface Layer 6 Corresponds to Approximately 300 mb), Surface Pressure, and Surface Height Fields for Box 30, June 1, 15Z

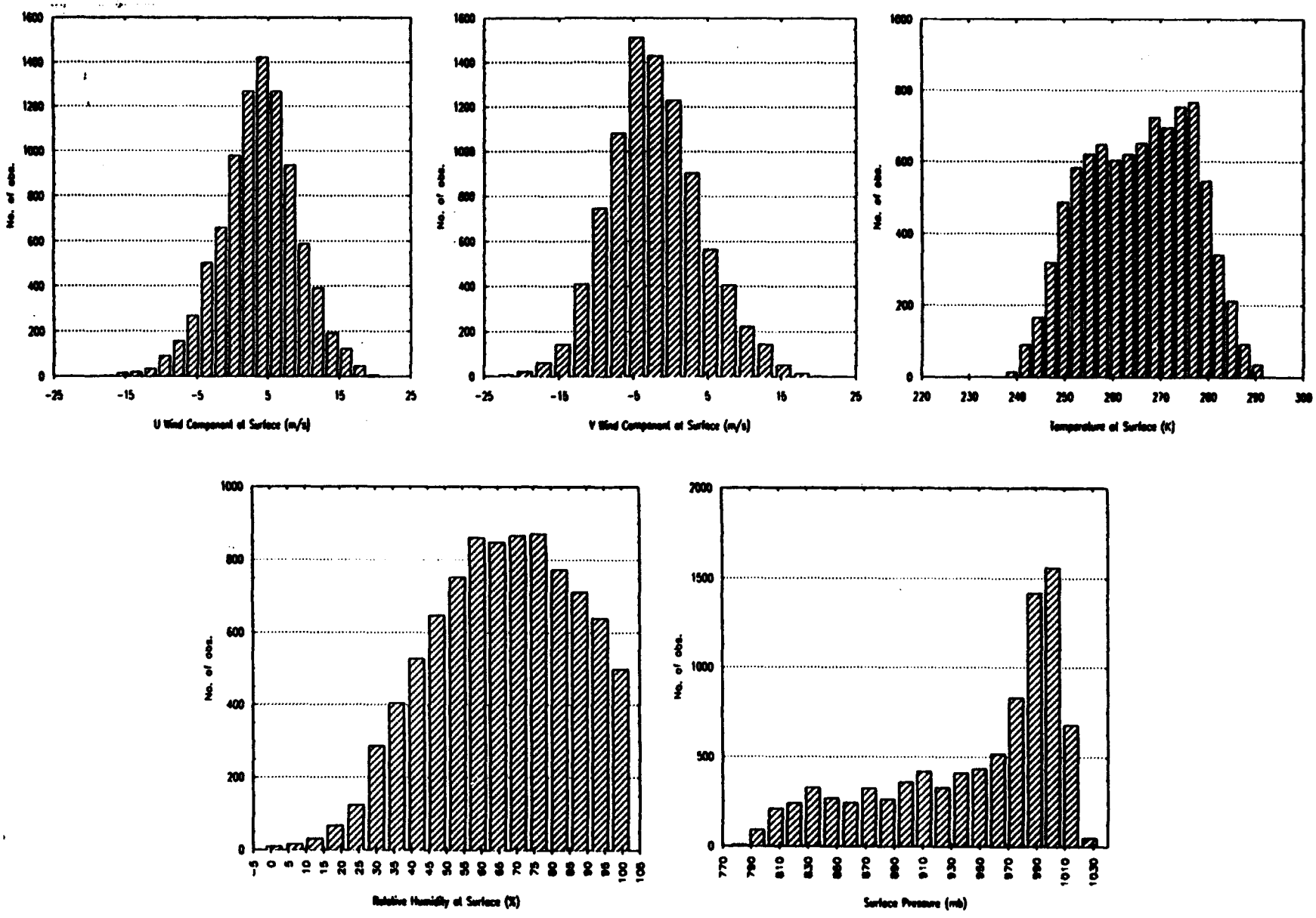


Figure 24 Histograms of Raw GSM Variables: u,v, Temperature, Relative Humidity, and Pressure at the Surface for Box 44, December, 0Z

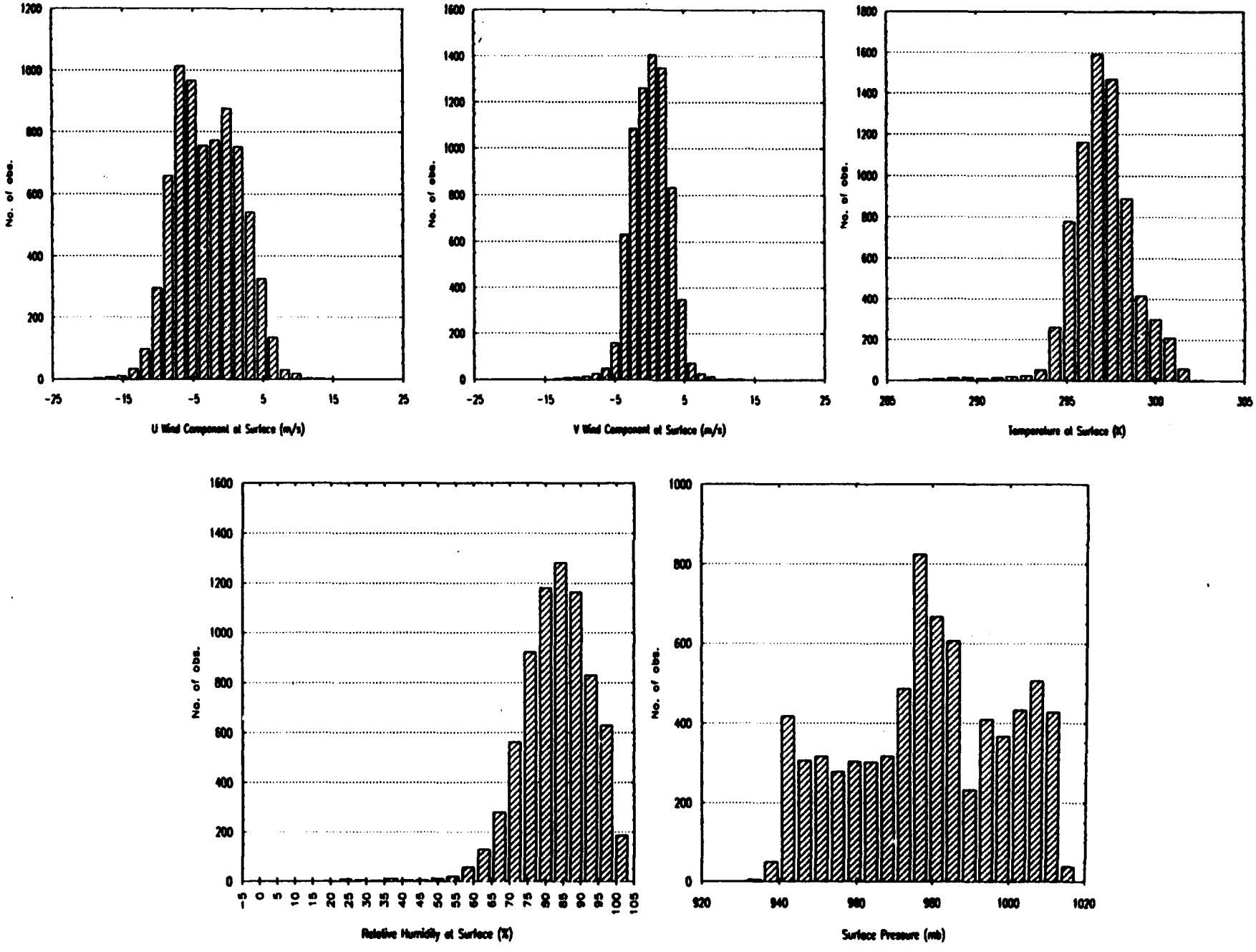


Figure 25 Histograms of Raw GSM Variables: u,v, Temperature, Relative Humidity, and Pressure at the Surface for Box 61, June, 12Z

3. MODEL DEVELOPMENT AND RESULTS

In this section we present the results of our modeling effort. A general introduction of various regression techniques and the mathematics behind developing linear regression models can be found in most introductory statistics texts (Refs. 10 and 19 are good examples) and will therefore not be included here. We will introduce those concepts unique to our model implementation and development data set, but will otherwise assume that the reader has a basic knowledge of linear regression for the following discussion and the remainder of this report.

Throughout the model development process, we employed a commercial statistical and graphical analysis software package called STATISTICA™. It served as a database management system for the model development data sets and provided a variety of regression techniques and corresponding significance tests, residual analysis, and graphing that were critical to efficient modeling and analysis during this project. The STATISTICA™ software package provided modules to perform multiple linear regression, using least squares estimation, and nonlinear regression in which the user can estimate any arbitrary regression function and any arbitrary loss function. It also enabled us to save the forecast cloud amounts and residuals to compute probabilistic forecasts off-line.

All of the figures discussed in this section are included at the end of the section, starting on page 96.

3.1 OVERVIEW OF THE MODELING PROCESS

Model development consisted of the following four major steps. These four steps are described at length in subsections 3.2, 3.3, 3.4, and 3.5, respectively.

STEP 1 **Build clusters based on cloud climatology** — The first step in the model development sequence was to reduce the number of regression models using principal components analysis. Regression models were built for each cluster rather than for individual grid points. We performed cluster analysis using principal components for four box/season combinations including one polar, one tropical, and two midlatitude boxes. Variables in the analysis included: cloud cover mean and scale distance, terrain type, elevation, and grid point location.

STEP 2 **Build the pool of potential predictors**—Using knowledge of meteorological processes in general, cloud formation in particular, and some trial-and-error, we selected variables from GSM outputs, RTNEPH analyses, and the RTNEPH terrain database to include in the pool of potential predictors. In addition, we derived other meaningful quantities from these outputs for inclusion. We used this common pool of predictors for all model development. We then selected a subset of these variables for each box/season/time using a forward regression technique that ranked the predictors in order of significance. Using only the most significant predictors, we estimated regression coefficients in Step 3.

STEP 3 **Develop regression models**—Many regression techniques are available. These include multiple linear regression, nonlinear regression, regression estimation of event probability, empirical orthogonal functions, regression trees, and many others. We used multiple linear regression to predict total cloud amount at eighth-mesh resolution with the predictors selected in the previous step. We developed equations for selected cases (i.e., a particular box, season, and time). We extended model development to predict probabilistic forecasts of cloud amount categories by analyzing the distributions of the residuals as described later in this section.

STEP 4 **Evaluate model performance**—It is virtually always the case that a statistical forecasting scheme performs better for the development data set than for a new, independent data set. This may be due to fitting noise in the development data set or to the incompleteness of the development data set. We used the jackknife technique to estimate the performance of our MOS approach and to determine confidence limits on that estimate. In addition, we also compared the MOS results to those obtained using simple persistence. The metrics we used in this comparison were forecast skill scores (the Brier score, sharpness, and the 20/20 score discussed later in this section). Lastly, we validated the probabilistic forecasts using Bayesian analysis of the probability distributions obtained using the model development data set and an independent set of data.

3.1.1 The Proposed Methodology Using REEP

In our original proposal (Ref. 26), we proposed developing MOS equations to forecast total cloud amount and probabilities of cloud amount categories using the regression estimation of event probability methodology (a linear regression technique) to support two different forecast formats required by Air Force and Army decision makers. It turned out that we were able *to produce both forecast output types using a more efficient and direct multiple linear regression technique* which avoided some of the pitfalls of the regression estimation of event probability (REEP) method.

The REEP method (Ref. 22) requires a categorical response variable. The first step is to transform the RTNEPH continuous cloud amount to a categorical variable by binning cloud cover into N non-overlapping categories. The single response used in multiple linear regression is then replaced with N binary response variables. For each observation, the response is one if the observed cloud amount falls within the corresponding category and zero otherwise. Multiple models are built, one for each of the categories, all with exactly the same predictor set and exactly the same database. The output of each model is an estimate of the probability that the observed cloud amount is within the corresponding range for a given realization of the predictors. A REEP model is constructed to guarantee that the sum of the N probabilities is one. The predicted total cloud amount can be determined from the N-vector of probabilities using a prescribed selection criterion (e.g., the most probable category, the median category, etc.).

REEP posed two problems in our modeling effort. First, it is well known that although the sum of the resulting probabilities is one, linear regression does not constrain any individual probability to lie between zero and one, as a true probability must. It is possible to deal with this problem by using a nonlinear response function such as the logistic function (see Ref. 1), but this increases the computational costs immensely and for our data set was therefore not a feasible solution. Alternatively, it is possible to scale the probabilities, or to simply set those probabilities less than zero to zero and those greater than one to one. This is rather an ad hoc and unsatisfying solution.

The second problem is the increased computational costs of developing the initial predictor data set. There is no reason to believe that the same variables that are strong predictors of overcast are also strong predictors of clear or partly cloudy. Therefore one must determine the strongest predictors for each of the cloud cover categories from the overall pool of potential predictors individually. That is a computationally intensive process since the number of potential predictors is large (see Section 3.3). After selecting the most significant predictors for each of the categories one can build a group that is common to all of the categories (a somewhat subjective process) and perform the regression using the common predictor data set. The actual work required then to perform the N multiple regressions from the same predictor set is little more than that to perform a single regression because the covariance matrix does not change.

In Section 3.4.3 we discuss the particulars of how we implemented REEP and show the results of a sample case. Now we continue with a brief description of the alternate methodology we adopted to replace REEP, and discuss how we generated both forecast output types (total cloud amount and probabilities of cloud amount categories) from a single linear regression model.

3.1.2 An Alternate Methodology Using Actual Residuals

Our goal was to find an empirical linear function in which the response variable, total cloud amount, is expressed in terms of k predictor variables selected from GSM output, RTNEPH analyses, and other candidate predictors. It is important to note that this model is linear in the predictors, but not necessarily in the meteorological variables, as predictors may be transformations of meteorological variables.

Following the notation in Ref. 6, let $x_1, x_2, \ldots, x_k$ be k predictors and y be the response variable, then the linear regression model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i \qquad i = 1, 2, \ldots, n \tag{3-1}$$

where

n is the number of sample observations in our data set

y_i is the i^{th} observation of the response

ε_i is the model error associated with y_i

$\beta_0, \beta_1, \ldots, \beta_k$ are the unknown linear parameters.

This general linear model can be expressed in matrix form

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{3-2}$$

where

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

In matrix form, the least squares estimate for the regression coefficient vector, $\hat{\mathbf{B}}$, is given by the following where the superscript "T" indicates matrix transpose

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\mathbf{Y}. \tag{3-3}$$

The estimated regression equation is then

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}} \tag{3-4}$$

and $\mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}}$ is the vector of model residuals.

To determine the probabilities of cloud amount categories, we used the results of the general linear regression model described above. Our goal was to be able to answer the following types of questions that are critical for many Air Force and Army mission operations:

1. Given a cloud amount forecast, what is the probability that the observed cloud amount will be less than a critical threshold?
2. Given a cloud amount forecast, what is the probability that the observed cloud amount will be within a critical range of values?

To answer these questions we must understand the characteristics of the error distribution in our model. To find the probability that the observed cloud amount will be less than or equal to a critical threshold, c (i.e., Pr $\{\mathbf{Y} \le c\}$), first calculate $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}$ where $\hat{\mathbf{B}}$ is the best estimate for the regression coefficients using least squares estimation. Then, the probability can be computed from the residual distribution as follows:

$$\begin{aligned} &\Pr\{\mathbf{Y} \le c \mid \hat{\mathbf{Y}}\} \\ &= \Pr\{\hat{\mathbf{Y}} + \mathbf{E} \le c \mid \hat{\mathbf{Y}}\} \\ &= \Pr\{\mathbf{E} \le c - \hat{\mathbf{Y}} \mid \hat{\mathbf{Y}}\}. \end{aligned} \tag{3-5}$$

Figure 26 shows graphically how to compute the probability in Eq. 3-5 from a sample error distribution; the shaded area under the curve is the probability referred to in Eq. 3-5 above. It is a trivial extension of that equation to compute the probability that the observed cloud cover is within a specified range by finding the difference between the probabilities of the two extreme thresholds that define the range.

Linear regression assumes that model residuals, or errors, are normally distributed over the entire sample: $\mathbf{E} \sim N(0, \sigma^2)$. Our results have verified that when taken as a whole, they are indeed approximately normal (see Figure 42 in Section 3.4.1). A naive approach to computing the probabilistic forecast would assume that the residuals are normally distributed across any *range* of cloud cover estimates. However, over any range of estimates, the associated residuals may be far from normal due to the bounded nature of the cloud cover variable. In most cases the error distributions (especially for clear and overcast forecasts) are in fact, U-, J-, or L-shaped. (See Figure 43 in Section 3.4.2)

Therefore, a better way to compute the probability in Eq. 3-5 is to use the *actual* error distributions for the development data set. We divided the cloud amount forecasts into 11 bins: 0–4%, 5–14%, . . . , 85–94%, 95–100% and then generated histograms of the model errors for each of the forecast categories. The probability that the observed cloud cover was less than a critical threshold, given the estimated cloud coverage, was computed by summing the error histograms corresponding to that estimated coverage.

3.2 SPATIAL CLUSTERING OF CLOUD COVER

The first step in the model development sequence was to reduce the number of regression models using principal components analysis. Regression models were built for each cluster rather than for individual grid points. This reduced the overall number of models to be developed and also increased the size of the development data sets. Cluster-based development did not reduce the specificity of the final models since parameters used to define the clusters were also included in the pool of potential predictors.

Clustering was constrained to individual RTNEPH boxes and used location, terrain, and cloud amount frequency distribution variables for each of the 4096 (i.e., 64 × 64) RTNEPH grid points in a box. Specifically, the cluster data sets included, for each eighth-mesh grid point: latitude, longitude, terrain elevation, terrain type, mean cloud cover, and scale distance. The last two variables in this list are computed from the RTNEPH analyses over one season (3 months) for 8 times of day.

Scale distance is a parameter of the Burger Area Algorithm (BAA) which was developed at the Air Force Geophysics Laboratory and is described in Ref. 5. The BAA is a model of the probability that the areal cloud coverage is less than or equal to a threshold coverage value given the area, scale distance, and mean clearness. The scale distance can be obtained by using the inverse BAA when provided with a cloud cover data set. It is a measure of the range of significant spatial autocorrelation of cloud cover.

We performed cluster analysis using principal components for four box/season combinations including one polar, one tropical, and two midlatitude boxes. Principal component analysis (PCA) has two main applications: 1) to reduce the number of variables in a data set and 2) to detect relationships between variables. We used both applications in our analysis.

Using the STATISTICA™ factor analysis module, we extracted principal components from our four development data sets, excluding the terrain type variable. Successive principal components are extracted from the data. The first component is a best fit through the data that maximizes the variance of that component, the second component is orthogonal to the first and maximizes the remaining variability, and so forth.

Consecutive principal components account for less and less variability, thus it is possible to reduce the number of variables by deciding to keep only the first few components. The decision to cut off some components can be somewhat subjective, but a rule of thumb is to keep those components which account for at least as much variance as what one individual variable would account for. The variance of the individual variables (once normalized) is one. Therefore if the variance of a principal component of the normalized data set is greater than or equal to one, we will retain it.

Figures 27 and 28 show the variance of the five principal components calculated for two winter cases: box 38 in the midlatitudes and box 61 in the tropics (refer to Figure 3 for the RTNEPH box definitions). The first figure, corresponding to box 38, shows a case in which the first two principal components have variance greater than one. We retained both components for further analysis. The second figure, corresponding to box 61, is an example of the subjectivity involved in deciding how many principal components to retain for analysis. In that case we chose to retain the first three components even though the variance of the third component was slightly less than one.

Principal components are linear combinations of the input variables. Figures 29 and 30 show the normalized breakdown of the components into their parts. In these two cases, the first principal component was strongly dependent on the mean cloud cover and latitude. The scale distance parameter weighed heavily in the first component for box 38 and in the second component for box 61.

Although all that we have discussed so far in this section is the use of PCA for data reduction, the goal of this clustering task was *not* to reduce the five terrain and cloud distribution parameters to two or three linear combinations, but to determine clusters of grid points with similar characteristics. We used the principal components to determine how

similar grid points were by looking for groups of grid points in a scatterplot of the principal components. If the points were close in principal component space they were also close in terms of cloud climatology.

We found that the grid points grouped strongly by terrain type (land and water), which makes intuitive sense since cloud climatology is quite distinct in water and land regions and the bulk of the grid points with the same terrain type are spatially close. Coastal grid points seemed to be mixed between the two groups as expected. Figures 31 and 32 show the scatterplots for boxes 38 and 61. The figures for box 38 are two-dimensional scatterplots because we retained only the first two principal components in our analysis, on the other hand, the scatterplots for box 61 are three-dimensional to accommodate the three components we retained for analysis.

The two other cases we studied showed similar results: both boxes 45 and 28 showed strong clustering by terrain type. Although not conclusive, this analysis led us to believe that we could generate MOS equations for distinct terrain types within RTNEPH boxes. Ideally, with much greater resources, it would be best to develop models for each grid point individually and only after comparing the models, group points with similar models into clusters. However, that type of development would require a much larger development data set and modeling budget.

Findings in this task led us to develop a number of terrain-specific regression models for analysis. However, for the bulk of the models that we developed, we combined land and water grid points to provide a greater number of data points for model development.

3.3 BUILDING THE POOL OF POTENTIAL PREDICTORS

The selection of variables to use as potential predictors was a critical task in the model development process. We selected parameters based on input from personnel from Phillips Laboratory, the Techniques Development Laboratory (TDL) of the National Weather Service, and our own cloud modeling experts.

Potential predictors were taken primarily from the GSM half-mesh output fields and included explanatory variables such as relative humidity, temperature, and surface pressure as well as derived quantities, transformed variables, and differenced variables (difference over 3-, 6-, and 9-hour periods). Other predictors included eighth-mesh cloud amount analyses from the RTNEPH, eighth-mesh elevation and type from the RTNEPH terrain database, latitude, longitude, date, and solar zenith angle.

Our challenge was to *limit* the total number of variables to include in the predictor pool without omitting good predictors of cloud cover at eighth-mesh resolution. Limiting the overall number of variables was important since our findings may lead to future development of an operational cloud-amount forecast model in which *accuracy* and *economy* are both major concerns.

Potential predictors included point variables (e.g., surface pressure and cloud amount), column variables (e.g., temperature and wind at vertical sigma levels), field variables (e.g., terrain elevation), and variables at times other than the valid time. Some of the data fields were highly correlated. It was our goal to reduce the number of potential predictors by taking advantage of this data redundancy using layer averaging, principal components, etc.

The set of potential predictors we settled on included approximately eighty variables. The fifteen most significant of these potential predictors were selected for each box/season/time using a "forward selection" scheme. These fifteen were then used as predictors in final model development. Forward selection methods (discussed in Ref. 10) represent one way to reduce the overall predictor pool. In those methods, variables are selected one at a time. Once a variable is selected, it is retained without chance of later being discarded. First the best single variable is selected. With k potential predictors, this requires k simple (one-variable) regressions. Next, the second-best variable, given the particular choice of the best, is selected by performing k-1 two-variable regressions. This process is continued, adding one variable at each step, until the selection of an additional variable is not warranted according to a pre-specified stopping rule or until all variables have been selected.

Our stopping rule was based on the reduction in variance. We found that the first five to ten variables accounted for the majority of the total reduction in variance achieved with all the predictors. After looking at the results from many regions, seasons, and times, we decided to select the strongest fifteen variables for use in model development. By keeping fifteen predictors, we accounted for the bulk of the reduction in variance and drastically reduced the computational time of performing the regression. In addition, we retained enough predictors as to not lose critical insights into the subtle differences in cloud processes in varying locations, at varying times of day, etc.

In Section 3.4, we present our modeling results. For each case, we show which fifteen predictors were selected using the forward selection method and what the relative

ranking of those variables was. It would probably not be necessary to use all fifteen predictors in an operational model, but the increased computations for this feasibility study were well worth the effort.

A complete list of the variables used in this study is included in Appendix A. We briefly describe the RTNEPH variables below and then devote the remainder of this section to describing the GSM-based predictors.

3.3.1 Candidate Predictors From the RTNEPH Cloud Analyses and Terrain Database

These predictors included total cloud amounts for the initial time and for times at three-hour intervals preceding the initial time back to 24 hours prior to the forecast *valid* time. The MOS equations used these variables to model persistence effects and trends in cloud amount. Terrain elevation and type were also included in the pool of predictors along with the terrain gradient in the direction of GSM winds.

A key predictor based on the RTNEPH gridded analyses turned out to be advected cloud cover. Intuition tells us that to forecast cloud cover at a point we should look upwind to see what weather (i.e., cloud cover) will move into the area around that point over the forecast time period.

We defined a relatively simple advected cloud cover predictor driven by forecast winds at the valid time for the point of interest. We computed average low, middle, and high layer winds (see below for how we defined the three vertical regions) over a horizontal region including the point of interest and the eight surrounding half-mesh grid points. We then assumed the winds acted constantly over the forecast period (3, 6, or 9 hours) to find the upwind cloud cover. We determined the upwind cloud amount over a region the size of a half-mesh grid point by averaging the nearest 16 eighth-mesh cloud amount values. We produced advected cloud amounts for low, middle, and high layer winds over 3-, 6-, and 9-hour periods. Further discussion of these predictors and their frequency of occurrence in the strongest fifteen predictors can be found in Section 3.4.

3.3.2 Candidate Predictors From the Global Spectral Model

Raw Variables

The GSM outputs available for use in this modeling effort included: u and v wind components and temperature at the ten lowest vertical (sigma) levels, relative humidity

at the six lowest sigma levels, surface pressure and model surface height for each grid point on the half-mesh grid. Output variables from the 0Z- and 12Z-initialized model runs were both available, in general, though frequently data from only one model run were used to develop the MOS equations to simulate operational constraints (see Section 2.2.2).

Two other variables were obtained directly from the GSM layer data. They are: the maximum relative humidity, and the humidity at the layer above the maximum (this last variable was suggested in Ref. 20).

Derived Variables

It is possible to derive a number of meteorologically important variables from the "raw" u, v, temperature, and humidity variables mentioned above. An obvious predictor for cloud amount that can be derived from these quantities is vertical velocity. Others include: divergence, vorticity, temperature advection, layer thickness, wind speed, vertical wind shear, and a number of stability parameters. Each of these derived quantities is described below.

Divergence — is a measure of how a fluid is expanding or contracting. In this effort, it is a measure of mesoscale motion in the atmosphere. In a simple way, convergence (negative divergence) can be associated with colliding air masses, and positive vertical motion which are both indicative of increased cloud cover. Positive divergence, on the other hand, can be a sign of clearing. Horizontal divergence is computed as the sum of the partial derivatives of the u and v wind components, where u corresponds to the component of the wind in the x-direction (zonal flow), and v corresponds to the component of the wind in the y-direction (meridional flow)

$$\text{Divergence} = \nabla_h \cdot \vec{v}_h = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \tag{3-6}$$

where the subscript "h" specifies the horizontal components only. In practice, we can only *approximate* this theoretical definition by computing the divergence at each half-mesh grid point from winds at surrounding grid points. Figure 33 shows the geometry we used to implement the horizontal derivatives. The divergence at the center-point of the grid is given by

$$\text{Div}_h = \frac{(u_a - u_c)}{2\Delta x} + \frac{(v_d - v_b)}{2\Delta y} \tag{3-7}$$

where Δx and Δy are the distances between adjacent grid points in the x and y directions, respectively. Divergence is very sensitive to errors in calculation because the two terms

in Eq. 3-7 are usually small, often of opposite sign, and often very close in magnitude. Thus we need to know the terms very accurately in order to make a reasonable estimate of the divergence.

Vorticity — is a measure of the local rotation of a fluid. It is intimately connected to divergence because as fluids converge or diverge their rate of rotation changes to conserve angular momentum. Mathematically, vorticity is given by

$$\text{Vorticity} = \hat{k} \cdot (\nabla_h \times \vec{v}_h) = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}. \tag{3-8}$$

Implementation used approximate derivatives to determine the vorticity at the center of the grid (refer again to Figure 33):

$$\text{Vort}_h = \frac{(v_a - v_c)}{2\Delta x} - \frac{(u_d - u_b)}{2\Delta y}. \tag{3-9}$$

Vertical Velocity — is a measure of the vertical motion in the atmosphere. It is usually a good measure of convective activity and precipitation and thus, indirectly, cloud cover. It is usually derived from divergence, and thus is sensitive to errors in the divergence computation. There are many methods to compute vertical velocity. We used the method followed in the Air Force version of the Global Spectral Model (Ref. 2). The vertical velocity, $\dot{\sigma}$, is the total derivative of the sigma coordinate as follows:

$$\dot{\sigma} = \sigma(\overline{C} + \overline{D}) - (\overline{C}^{\sigma} + \overline{D}^{\sigma}) \tag{3-10}$$

where

$$(\bar{\ })^{\sigma} = \int_0^{\sigma} (\ )d\sigma$$

$$(\bar{\ }) = \int_0^{1} (\ )d\sigma$$

$C = \vec{v} \cdot \nabla q$, $\quad q = \log(\text{pressure})$

$D = \text{Divergence}$

and where $\dot{\sigma} = 0$ at $\sigma = 0$ and $\sigma = 1$.

In our implementation of this equation, we limited the vertical range over which we performed the integrations to the ten reported sigma levels available to us. That is, the part of the atmosphere above the top sigma level of our data (approximately 120 mb) was ignored. Due to the nature of the MOS approach, this approximation was most likely of little consequence, because MOS can use the trend in vertical velocity rather than the absolute value as a predictor of cloud amount.

Temperature Advection — is a measure of how fast the temperature gradient is changing. It can be a measure of baroclinic development and thus of the changes in cloud cover. Temperature advection is given by the following equation:

$$\text{Temperature Advection} = \vec{v}_h \cdot \nabla_h T = -u\frac{\partial T}{\partial x} - v\frac{\partial T}{\partial y}. \qquad (3\text{-}11)$$

Again, we approximate the partial derivatives using the gridded data available to us and implement temperature advection at the center of the grid using the following (refer again to Figure 33):

$$\text{Tadv}_h = -u_{avg}\frac{(T_a - T_c)}{2\Delta x} - v_{avg}\frac{(T_d - T_b)}{2\Delta y} \qquad (3\text{-}12)$$

where u_{avg} and v_{avg} are the u and v wind components averaged over the center point o and its two neighbors along each direction (we weight the center point twice as heavily as the other grid points).

Wind Speed — is simply the resultant speed computed using the u and v wind components at the surface as follows:

$$\text{Wind Speed} = \sqrt{(u^2_{surface} + v^2_{surface})}. \qquad (3\text{-}13)$$

Vertical Wind Shear — is a measure of the magnitude of the wind shear between high and low layers in the atmosphere. We computed wind shear as

$$\text{Wind Shear} = \sqrt{(u_{high} - u_{low})^2 + (v_{high} - v_{low})^2} \qquad (3\text{-}14)$$

where the high layer winds are averages over the top five sigma levels (approximately 375 to 100 mb) and the low layer winds are averages of the lower two sigma levels (surface to approximately 800 mb).

Stability Measure — is a simple measure of the temperature change with height in the atmosphere. We computed two stability measures as follows:

$$\text{Stability Measure 1} = T_{850\ mb} - T_{surface} \quad (3\text{-}15)$$

$$\text{Stability Measure 2} = T_{500\ mb} - T_{850\ mb} \quad (3\text{-}16)$$

where all pressures are approximate because the temperatures are actually evaluated on sigma coordinates which are defined relative to surface pressure.

Layer Thickness — is the geometric thickness of the layer bounded at the bottom by pressure P_b and at the top by pressure P_t. It is a type of stability measure in the atmosphere. The algorithm to compute the thickness involves numerical integration of the hydrostatic equation. It is described in Ref. 25. We used $P_b=P_{surface}$ and $P_t=850$ mb.

Transformed Variables

Another group of potential predictors included transformed variables. During preliminary data analysis we studied scatterplots of individual predictors versus cloud amount and residual plots which led us to try non-linear transformations of key variables, especially relative humidity (RH). Through trial and error and analysis of previous results we settled on the following group of transformed predictors:

- log(RH)
- RH^2
- RH^4
- vertical velocity $\times$ RH
- RH > 50%, RH > 70%, RH > 90%.

The first three of these predictors emphasize extreme relative humidity values. In the same way, the product of vertical velocity and relative humidity emphasizes those situations when both variables are extremely low or high. This predictor was suggested by personnel at the TDL (Ref. 11). The same group has also found thresholded predictors useful in building MOS equations using output from the Nested Grid Model. The last predictors are examples of thresholded or binary variables, equal to one if the humidity is greater than the threshold, and zero otherwise.

Differenced Variables

The Global Spectral Model generates forecasts every twelve hours in three hour increments. In some cases it may be desirable to use model output from three hours before or after the actual valid time in the predictor pool to account for possible spin-up problems in the model. Due to computational limitations we were not able to add data from both before and after the valid time in our pool of potential predictors. Instead, we used 3-, 6-, and 9-hour differences in key variables to capture the temporal component of the model. Using time differenced predictors in the MOS equations allowed us to avoid any absolute errors in the model forecasts (e.g., moisture values are known to be biased at high sigma levels, Ref. 20) and instead rely on *trends* in meteorological variables. For example, when making a 6-hour forecast valid at 15Z, we use the GSM forecast variables valid at 15Z and the change in forecast variables over the 9Z to 15Z period. The difference predictors can be used to predict changes in cloud amount from the amount at initial time. This is the strength of our MOS approach: combining observations at initial time, forecast variables at the valid time, and differences in model variables from initial to valid time, so as to accommodate potential model biases and spin-up problems.

Reduction of the Number of Variables

If we considered all of the raw, derived, transformed, and differenced variables mentioned above for each vertical level, our pool of potential predictors would contain over 200 variables for each grid point. Because computations involved in forward variable selection and regression in general are so costly, we investigated the tradeoffs associated with reducing the overall number of potential predictors.

Based on our preliminary analyses, we concluded that we could reduce the ten sigma layers to averaged low, middle, and high layers with little loss of fidelity. In fact, we found that the three-layer variables were more stable predictors of total cloud amount than their ten individual layer counterparts. This enabled us to cut by over two-thirds the number of column variables in the pool of potential predictors. The definition of the low, middle, and high layers we used in this effort with respect to the ten GSM sigma layers is shown in Figure 34.

The resulting list of potential predictors used in our final regression model development is found in Appendix A.

3.4 REGRESSION MODEL RESULTS

Modeling consisted of two phases: exploratory and final. In the exploratory phase discussed in previous sections, we built up the predictor list, studied the effects of time of day, length of forecast, and location on data availability, and defined the general modeling process. During that phase we mainly studied three regions: RTNEPH boxes 44, 45, and 61. Most of our results from that phase were qualitative.

In the final modeling phase we chose a number of box/season/time combinations based on availability of timely data, variety of terrain types and cloud climatologies, and strategic importance. A base set of 59 combinations was selected for analysis. To that set we added 12 more cases that were terrain specific; that is, they only included grid points of a single terrain type: land or water as suggested by our cluster analysis. All 71 analysis cases are listed in Table 3.

We defined a standard set of predictors (listed in Appendix A) and modeling operations including:

- selecting the strongest predictors from the pool
- determining regression coefficients
- computing partial correlations
- determining the reduction in variance
- analyzing model residuals.

From model results we then determined categorical cloud amount probabilities. Finally, we compared the MOS results to persistence using meteorological skill scores and predicted future model performance.

Regression models were developed from data sets containing one season of data and 225 eighth-mesh grid points (fewer for terrain-specific models) one grid point for each of 15×15 half-mesh cells. (The original GSM outputs were on 17×17 half-mesh grids after processing at Phillips Laboratory. We used only the inner 15×15 cells so as to be able to compute the horizontal gradients used in divergence, vorticity, and temperature advection computations). The four seasons were defined as follows:

Winter	= {December, January, February}
Spring	= {March, April, May}
Summer	= {June, July, August}
Fall	= {September, October, November}.

Table 3 List of 71 Model Development Data Sets

CASE #	CASE NAME	BOX	SEASON	TIME OF DAY (Z)	FORECAST LENGTH (HOURS)	GSM INITIAL TIME (Z)	TERRAIN TYPE
1	30/SP/3/3/12	30	SPRING	3	3	12	ALL
2	30/SP/3/6/12	30	SPRING	3	6	12	ALL
3	30/SP/3/9/12	30	SPRING	3	9	12	ALL
4	30/SP/9/3/0	30	SPRING	9	3	0	ALL
5	30/SP/9/6/0	30	SPRING	9	6	0	ALL
6	30/SP/9/6/12	30	SPRING	9	6	12	ALL
7	30/SP/9/9/12	30	SPRING	9	9	12	ALL
8	30/SP/21/3/12	30	SPRING	21	3	12	ALL
9	30/SP/21/6/12	30	SPRING	21	6	12	ALL
10	30/SP/21/6/0	30	SPRING	21	6	0	ALL
11	30/SP/21/9/0	30	SPRING	21	9	0	ALL
12	30/SU/9/3/0	30	SUMMER	9	3	0	ALL
13	30/SU/9/6/0	30	SUMMER	9	6	0	ALL
14	30/SU/9/6/12	30	SUMMER	9	6	12	ALL
15	30/SU/9/9/12	30	SUMMER	9	9	12	ALL
16	30/SU/21/3/12	30	SUMMER	21	3	12	ALL
17	30/SU/21/6/12	30	SUMMER	21	6	12	ALL
18	30/SU/21/6/0	30	SUMMER	21	6	0	ALL
19	30/SU/21/9/0	30	SUMMER	21	9	0	ALL
20	30/FA/9/3/0	30	FALL	9	3	0	ALL
21	30/FA/9/6/0	30	FALL	9	6	0	ALL
22	30/FA/9/6/12	30	FALL	9	6	12	ALL
23	30/FA/9/9/12	30	FALL	9	9	12	ALL
24	30/FA/21/3/12	30	FALL	21	3	12	ALL
25	30/FA/21/6/12	30	FALL	21	6	12	ALL
26	30/FA/21/6/0	30	FALL	21	6	0	ALL
27	30/FA/21/9/0	30	FALL	21	9	0	ALL
28	30/WI/9/3/0	30	WINTER	9	3	0	ALL
29	30/WI/9/6/0	30	WINTER	9	6	0	ALL
30	30/WI/9/6/12	30	WINTER	9	6	12	ALL
31	30/WI/9/9/12	30	WINTER	9	9	12	ALL
32	30/WI/21/3/12	30	WINTER	21	3	12	ALL
33	30/WI/21/6/12	30	WINTER	21	6	12	ALL
34	30/WI/21/6/0	30	WINTER	21	6	0	ALL
35	30/WI/21/9/0	30	WINTER	21	9	0	ALL
36	44/SU/0/3/12	44	SUMMER	0	3	12	ALL
37	44/SU/0/6/12	44	SUMMER	0	6	12	ALL
38	44/SU/0/9/12	44	SUMMER	0	9	12	ALL
39	44/SU/15/3/0	44	SUMMER	15	3	0	ALL
40	44/SU/15/6/0	44	SUMMER	15	6	0	ALL

CASE #	CASE NAME	BOX	SEASON	TIME OF DAY (Z)	FORECAST LENGTH (HOURS)	GSM INITIAL TIME (Z)	TERRAIN TYPE
41	44/SU/15/9/0	44	SUMMER	15	9	0	ALL
42	44/WI/0/3/12	44	WINTER	0	3	12	ALL
43	44/WI/0/6/12	44	WINTER	0	6	12	ALL
44	44/WI/0/9/12	44	WINTER	0	9	12	ALL
45	44/WI/15/3/0	44	WINTER	15	3	0	ALL
46	44/WI/15/6/0	44	WINTER	15	6	0	ALL
47	44/WI/15/9/0	44	WINTER	15	9	0	ALL
48	61/SU/3/3/12	61	SUMMER	3	3	12	ALL
49	61/SU/3/6/12	61	SUMMER	3	6	12	ALL
50	61/SU/3/9/12	61	SUMMER	3	9	12	ALL
51	61/SU/15/3/0	61	SUMMER	15	3	0	ALL
52	61/SU/15/6/0	61	SUMMER	15	6	0	ALL
53	61/SU/15/9/0	61	SUMMER	15	9	0	ALL
54	61/WI/3/3/12	61	WINTER	3	3	12	ALL
55	61/WI/3/6/12	61	WINTER	3	6	12	ALL
56	61/WI/3/9/12	61	WINTER	3	9	12	ALL
57	61/WI/15/3/0	61	WINTER	15	3	0	ALL
58	61/WI/15/6/0	61	WINTER	15	6	0	ALL
59	61/WI/15/9/0	61	WINTER	15	9	0	ALL
60	30/SU/21/3/12L	30	SUMMER	21	3	12	LAND
61	30/SU/21/6/0L	30	SUMMER	21	6	0	LAND
62	30/SU/21/9/0L	30	SUMMER	21	9	0	LAND
63	30/SU/21/3/12W	30	SUMMER	21	3	12	WATER
64	30/SU/21/6/0W	30	SUMMER	21	6	0	WATER
65	30/SU/21/9/0W	30	SUMMER	21	9	0	WATER
66	61/WI/3/3/12L	61	WINTER	3	3	12	LAND
67	61/WI/3/6/12L	61	WINTER	3	6	12	LAND
68	61/WI/3/9/12L	61	WINTER	3	9	12	LAND
69	61/WI/3/3/12W	61	WINTER	3	3	12	WATER
70	61/WI/3/6/12W	61	WINTER	3	6	12	WATER
71	61/WI/3/9/12W	61	WINTER	3	9	12	WATER

3.4.1 Predicting Cloud Cover Fraction

Strongest Predictors

We generated regression models for all 71 cases listed in Table 3. The first step in the process was to determine the strongest predictors in each case. The fifteen strongest predictors, as selected by the forward stepwise regression module in the STATISTICA™ software, are listed for each of the 71 model cases in Table 4. Brief descriptions of the variables, listed by their abbreviated names in the table, can be found in Appendix A.

Table 4 List of Strongest Fifteen Predictors Selected Using Forward Stepwise Regression for the 71 Development Data Sets (Coded Variable Names Listed Here are Described in Appendix A)

#	CASE NAME	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	30/SP/3/3/12	ACCM3	NEPH3	ACCH3	UM	LNRHM	NEPH24	NEPH6	D3RHM	LAT	ELEV	NEPH12	D3P	D3WL	ACCL3	LON
2	30/SP/3/6/12	ACCM6	NEPH6	RHMX	UM	ACCH6	NEPH24	LAT	D6P	ACCL6	D6RHM	NEPH12	THICK	VTL	VL	RHM4
3	30/SP/3/9/12	RHMX	ACCM9	UM	NEPH9	D9P	NEPH24	ACCH9	TL	VL	ACCL9	ZENITH	D9VTM	TAH	VTL	D9RHM
4	30/SP/9/3/0	ACCM9	NEPH3	RHMX	ACCH3	NEPH6	LAT	ACCL3	UM	NEPH21	LON	SPEED	VTM	D3RHM	DELEV	ZENITH
5	30/SP/9/6/0	ACCM6	NEPH6	RHMX	UM	ACCH6	NEPH21	ACCL6	SPEED	VTL	D6VL	DELEV	LAT	D6RHM	RHM4	NEPH9
6	30/SP/9/6/12	ACCM6	NEPH6	RHMX	NEPH21	LAT	UM	ACCH6	ACCL6	VTH	SPEED	DELEV	D6VM	D6UM	VTL	RH90
7	30/SP/9/9/12	RHMX	ACCM9	NEPH9	LAT	UM	NEPH21	ACCH9	VTL	D9VL	ACCL9	RH90	D9RHM	SPEED	DELEV	RHM4
8	30/SP/21/3/12	ACCM3	NEPH3	RHM2	ACCH3	ZENITH	D3P	NEPH18	ACCL3	RHL4	D3RHM	VL	LON	DAY	TAH	NEPH21
9	30/SP/21/6/12	ACCM6	RHM2	NEPH6	D6P	ACCH6	ELEV	RHL4	VL	UM	NEPH21	VM	ACCL6	WM	DELEV	D6RHM
10	30/SP/21/6/0	ACCM6	NEPH6	RHM2	ACCH6	D6P	ELEV	NEPH21	ACCL6	VL	TAH	THICK	D6TAL	D6RHM	RHL4	VM
11	30/SP/21/9/0	ACCM9	RHM2	NEPH9	D9P	ELEV	ACCH9	NEPH21	VL	ACCL9	THICK	WRHL	D9VL	D9TAL	D9RHM	NEPH18
12	30/SU/9/3/0	ACCM3	NEPH3	LAT	TM	NEPH21	RHM	NEPH6	ACCH3	D3RHM	VTL	D3VL	VM	UM	RHMX	NEPH15
13	30/SU/9/6/0	ACCM6	RHMX	TM	NEPH6	ACCH6	NEPH24	D6RHM	VTL	NEPH9	UM	WL	ACCL6	DELEV	D6TAH	SFCH
14	30/SU/9/6/12	ACCM6	RHMX	TM	NEPH6	ACCH6	NEPH24	DVH	NEPH9	ACCL6	SFCH	D6RHM	VTL	UM	VM	RHM4
15	30/SU/9/9/12	RHMX	ACCM9	TM	NEPH9	DVH	ACCH9	NEPH24	UM	WL	VTL	D9RHM	LAT	RHM4	ACCL9	NEPH15
16	30/SU/21/3/12	ACCM3	NEPH3	RHM2	THICK	ACCH3	D3P	D3RHM	NEPH21	D3VTM	RHL4	VL	ELEV	NEPH15	VTH	SHEAR
17	30/SU/21/6/12	ACCM6	RHM2	SFCP	NEPH6	ACCH6	DVL	ELEV	DVH	D6P	NEPH24	D6DVL	D6RHM	RHL4	D6VTH	ZENITH
18	30/SU/21/6/0	ACCM6	RHM2	NEPH6	SFCP	ACCH6	DVH	ELEV	DVL	ZENITH	NEPH18	VL	UM	VTL	LON	D6P
19	30/SU/21/9/0	RHM	ACCM9	THICK	NEPH9	ACCH9	DVL	ZENITH	VL	NEPH18	ELEV	DVH	VTL	D9WM	D9RHM	TAM
20	30/FA/9/3/0	ACCL3	NEPH3	ACCH3	NEPH9	SFCH	WRHM	NEPH6	NEPH21	D3RHM	VTL	TAH	SPEED	DELEV	D3VL	DAY
21	30/FA/9/6/0	ACCL6	ACCH6	NEPH6	RHMX	NEPH9	D6RHM	ELEV	ACCM6	NEPH21	TAH	SFCH	THICK	DELEV	SPEED	NEPH15
22	30/FA/9/6/12	ACCL6	NEPH6	ACCH6	RHMX	SFCH	NEPH9	ACCM6	UM	D6RHM	NEPH24	DELEV	VTL	TAM	LON	SPEED
23	30/FA/9/9/12	RHMX	NEPH9	ACCM9	ACCL9	UM	SFCH	D9RHM	NEPH18	ACCH9	SPEED	DELEV	VTL	D9VL	RH90	NEPH15
24	30/FA/21/3/12	ACCL3	NEPH3	ACCH3	NEPH24	ZENITH	RHM	NEPH6	ACCM3	D3RHM	D3P	RHL4	D3VL	DELEV	LON	NEPH15
25	30/FA/21/6/12	ACCL6	RHM	NEPH6	ACCH6	ZENITH	NEPH21	D6RHM	RH70	D6P	ACCM6	ELEV	DELEV	UM	RHL4	D6VL
26	30/FA/21/6/0	ACCL6	NEPH6	RHM2	ACCH6	NEPH21	UM	ZENITH	ACCM6	ELEV	TL	DELEV	DVL	D6RHM	NEPH15	RH90
27	30/FA/21/9/0	ACCL9	RHM	NEPH9	NEPH12	ZENITH	UM	ACCM9	NEPH18	D9VTL	DELEV	ACCH9	ELEV	RH90	D9RHM	NEPH12
28	30/WI/9/3/0	ACCL3	NEPH3	ACCH3	NEPH9	SFCH	D3RHL	NEPH24	NEPH6	ELEV	TM	RHMX	D3P	WH	SFCP	D3WL
29	30/WI/9/6/0	ACCL6	NEPH6	RHMX	ELEV	NEPH9	D6RHM	TM	ACCM6	NEPH24	D6P	D6WL	WM	DELEV	TYPE	D6RHL
30	30/WI/9/6/12	ACCL6	NEPH6	ELEV	NEPH9	ACCH6	D6RHL	NEPH24	TL	WM	D6RHM	DELEV	TYPE	SFCP	LNRHL	TAM
31	30/WI/9/9/12	NEPH9	RHMX	ACCL9	UM	NEPH24	ACCH9	ELEV	D9RHM	RM	WRHL	DELEV	VTL	TYPE	SPEED	D9RHL
32	30/WI/21/3/12	NEPH3	ACCM3	NEPH6	LON	ACCH3	ACCL3	NEPH24	D3RHM	RHL	D3P	WM	WRHL	NEPH15	TAM	VL
33	30/WI/21/6/12	ACCL6	NEPH6	RHMX	NEPH24	D6RHM	LON	ACCH6	TAH	VL	D6P	VM	NEPH15	DELEV	RHL2	RH70
34	30/WI/21/6/0	ACCL6	NEPH6	RHM	NEPH24	ACCH6	WM	D6RHL	NEPH15	TL	D6TAH	VL	TAM	D6VTH	RHMX	RH50
35	30/WI/21/9/0	ACCL9	NEPH9	RHM	NEPH24	WRHM	D9RHL	ACCM9	D9TAH	TL	TAM	VL	NEPH15	RH70	D9P	NEPH21
36	44/SU/0/3/12	ACCL3	NEPH3	ACCH3	TDIF2	RHM	ELEV	D3VTH	NEPH24	ACCM3	VM	TAM	SPEED	DVL	NEPH9	NEPH12
37	44/SU/0/6/12	ACCM6	RHM	NEPH6	ELEV	ACCH6	VTL	NEPH24	D6VTH	ACCL6	TDIF2	TAM	D6UL	DVL	DAY	SPEED
38	44/SU/0/9/12	ACCM9	RHM	ELEV	NEPH9	ACCH9	VTL	NEPH24	D9VTH	ACCL9	TDIF1	D9RHL	NEPH12	TAM	DAY	TM
39	44/SU/15/3/0	ACCL3	RHMX	ACCM3	NEPH3	SFCH	VTM	NEPH9	NEPH24	D3VTH	ACCH3	UL	SPEED	TYPE	VL	D3P

#	CASE NAME	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
40	44/SU/15/6/0	ACCM6	RHMX	NEPH6	SFCH	ACCL6	NEPH24	ACCH6	D6VTH	D6UL	VTM	SPEED	D6RHM	SHEAR	RHL2	TYPE
41	44/SU/15/9/0	RHMX	ACCM9	ACCL9	SFCH	NEPH9	NEPH24	D9WL	TDIF1	VTM	ACCH9	TYPE	D9UL	D9VTH	SPEED	TAM
42	44/WI/0/3/12	ACCL3	NEPH3	ACCH3	RHM2	ACCM3	TM	VTL	NEPH18	TAH	RH70	D3P	NEPH24	D3RHL	TYPE	LON
43	44/WI/0/6/12	ACCL6	ACCM6	NEPH6	RHM	TM	RH70	DVH	NEPH24	VTL	TAM	NEPH18	WH	TAH	TYPE	D6RHM
44	44/WI/0/9/12	ACCL9	RHM	NEPH9	ACCM9	VTL	TM	RH70	DVH	NEPH18	NEPH24	TYPE	TAH	WH	D9DVH	D9UM
45	44/WI/15/3/0	ACCL3	NEPH3	ACCM3	RHMX	NEPH6	TAH	NEPH18	TM	RHM2	SHEAR	ACCH3	NEPH12	D3VTH	D3TAL	NEPH24
46	44/WI/15/6/0	ACCL6	RHMX	NEPH6	ACCM6	UM	TM	NEPH12	TAH	TAM	D6RHM	NEPH24	D6VTH	D6TAM	SPEED	ACCH6
47	44/WI/15/9/0	ACCL9	RHMX	NEPH9	TM	ACCM9	TAH	NEPH18	LON	RHM2	SPEED	TAM	NEPH12	UM	TDIF1	D9VTH
48	61/SU/3/3/12	ACCM3	ACCH3	LAT	ACCL3	NEPH21	NEPH15	NEPH3	LNRHM	D3VTL	TDIF1	NEPH12	TYPE	D3TAH	D3VTH	WRHM
49	61/SU/3/6/12	ACCM6	LNRHM	NEPH18	ACCL6	ACCH6	NEPH15	TAM	ZENITH	NEPH12	TYPE	RHM	TDIF2	VM	VL	WL
50	61/SU/3/9/12	ACCL9	LNRHM	ACCH9	NEPH18	NEPH12	ACCM9	RHM	TYPE	UL	D9TAH	WRHM	D9VTH	TDIF1	NEPH15	D9TAL
51	61/SU/15/3/0	ACCM3	LAT	ACCH3	NEPH24	ACCL3	WM	TAM	VTL	NEPH3	UL	ELEV	SFCP	ZENITH	NEPH9	NEPH6
52	61/SU/15/6/0	ACCH6	LAT	NEPH24	ACCM6	WM	VTL	LNRHM	D6WH	D6UL	D6P	RHAB	NEPH6	DELEV	VM	UL
53	61/SU/15/9/0	ACCM9	NEPH24	LAT	ACCH9	WM	VTL	LNRHM	D9WH	UL	NEPH9	TAM	ZENITH	LON	UM	D9P
54	61/WI/3/3/12	ACCM3	NEPH21	NEPH3	LAT	ACCL3	VTM	ACCH3	RHM2	DELEV	NEPH6	WH	NEPH15	VTH	TL	D3RHM
55	61/WI/3/6/12	ACCM6	NEPH21	LAT	ACCH6	ACCL6	D6VM	VM	NEPH6	DELEV	WH	VTH	NEPH15	VL	D6VL	D6UM
56	61/WI/3/9/12	NEPH21	ACCM9	LAT	ACCL9	ACCH9	D9VM	VM	DELEV	NEPH12	WH	VTH	NEPH15	LON	D9WL	D9TAM
57	61/WI/15/3/0	ACCM3	NEPH24	TL	ACCL3	ACCH3	RHM2	NEPH3	NEPH9	D3VM	D3RHL	RHL2	VTM	LAT	SFCH	D3WH
58	61/WI/15/6/0	ACCM6	NEPH24	RHM	D6VM	ACCH6	ACCL6	NEPH12	VTM	D6DVL	SFCH	UL	D6WL	D6VTM	ELEV	WL
59	61/WI/15/9/0	ACCM9	NEPH24	ACCL9	VTM	NEPH21	RHL4	D9DVL	ACCH9	D9VM	SPEED	SHEAR	TL	D9WL	D9VTM	RHM
60	30/SU/21/3/12L	ACCM3	NEPH3	RHM4	ACCH3	ELEV	DVH	NEPH21	D3P	D3RHM	RHL4	VTH	NEPH6	VTL	DELEV	D3VTM
61	30/SU/21/6/0L	ACCM6	RHM2	NEPH6	ELEV	ACCH6	DVH	D6VTH	DVL	NEPH24	VTH	RHM4	D6VL	VTL	D6P	TM
62	30/SU/21/9/0L	RHM	ACCM9	THICK	NEPH9	ACCH9	VL	ZENITH	VTL	ELEV	D9WM	D9VM	D9P	NEPH24	D9VL	D9VTH
63	30/SU/21/3/12W	ACCM3	ACCH3	NEPH3	DVH	D3P	D3VM	UM	TAM	VTL	TL	D3RHM	TAH	D3UL	DAY	RH90
64	30/SU/21/6/0W	ACCM6	DVH	RHM4	ACCH6	UM	NEPH18	VTL	D6VM	DAY	TAM	TL	D6TAM	D6P	WL	D6VL
65	30/SU/21/9/0W	ACCM9	DVH	ACCH9	VTL	NEPH18	RHM4	UM	DAY	D9VM	TAM	D9TAM	D9P	TL	NEPH9	LON
66	61/WI/3/3/12L	ACCM3	NEPH3	NEPH21	RHM	ACCL3	ACCH3	LAT	VTM	TAL	NEPH6	TL	DELEV	LNRHM	LON	D3VM
67	61/WI/3/6/12L	ACCM6	NEPH21	ACCH6	LAT	ACCL6	VTM	LON	NEPH6	DELEV	D6VM	SFCH	WH	D6UM	VL	D6VL
68	61/WI/3/9/12L	ACCM9	NEPH21	ACCH9	LAT	ACCL9	LON	D9WL	DELEV	NEPH12	D9VM	VM	TL	D9TAM	WH	VL
69	61/WI/3/3/12W	ACCL3	NEPH21	UL	NEPH15	NEPH3	ACCH3	D3WH	WRHL	VTH	DELEV	D3RHM	SFCP	RHL4	RH90	ZENITH
70	61/WI/3/6/12W	NEPH21	ACCL6	LAT	NEPH15	WH	VM	ZENITH	ACCH6	VTH	DELEV	D6VL	LON	RHAB	D6UM	DVL
71	61/WI/3/9/12W	NEPH15	ACCL9	NEPH21	LAT	WH	VM	VTH	D9VL	NEPH9	DELEV	RHAB	D9UM	SPEED	DVH	ELEV

Figure 35 shows a histogram of those predictors that were among the fifteen strongest most often over all 71 development data sets. In that figure we normalized the number of occurrences to the total number of possible occurrences for each variable. Recall that some predictors (e.g., wind speed) are common to all 3-, 6-, and 9-hour forecasts, whereas others (e.g., 6-hour change in surface pressure) are unique to only one forecast length.

While Figure 35 reflects the frequency with which specific predictors appear in the strongest fifteen, it does not contain information about whether the variables were *very strong* predictors of cloud cover (occurring always in the top half of the group) or simply *supporting* predictors (occurring in the lower half of the group of fifteen). For this reason it was necessary to analyze this figure in conjunction with the ordered predictor lists of Table 4 to make more specific conclusions about the meaning and utility of various predictors.

It is important to note however, that especially for those predictors that occur in supporting roles, the *absolute* utility of any individual predictor cannot be ascertained by its rank position in the top predictors, or even by its contribution to the overall reduction in variance. This is due to the high correlation between some predictors. The forward regression technique finds the set of fifteen predictors that together reduce the variance by the greatest amount. There may be alternate sets of predictors that perform nearly as well with different supporting predictors.

Similarly, for a different development data set we may find that the group of supporting predictors changes. We performed a limited analysis in which we split a development data set in half and found the strongest predictors from the two halves individually and the whole data set. We found the very strongest predictors remained constant across all three data sets. Some of the remaining variables changes rank position or dropped out altogether to be replaced by others.

Keeping these caveats in mind, and realizing that our conclusions are based only on one year of data, a number of interesting observations drawn from Figure 35 and Table 4 along with results from our exploratory analysis phase are grouped loosely by predictor type and listed below:

Advected Cloud Cover Terms

- The advected cloud cover terms (ACCL3, ACCM3, ACCH3, ACCL6, etc.) were very strong predictors (frequently in the strongest five) for all boxes, times of day, seasons, and forecast lengths we studied.
- The advection terms had their greatest influence in 6-hour forecasts, especially in box 30.

- The high-layer advection terms did not perform as strongly as low- and middle-layer advection in all box regions. That is, although they were frequently in the group of the strongest predictors, they often played supporting roles.

Persisted Cloud Cover Amount

- The cloud cover at initial time (i.e., the persistence forecast) was, as expected, a strong predictor along with the advection terms in boxes 30 and 44.
- The cloud cover at initial time was rarely among the very strongest predictors for box 61, although it frequently turned up in a supporting role.
- The cloud cover at initial time generally dropped in significance with forecast length for any given box/season/time combination.
- The 24-hour persisted cloud cover turned up more often in the very strongest predictors for box 61 than in boxes 30 and 44. However, 24-hour persistence occurred more often overall in box 44.
- Cloud cover 21 hours before valid time was a frequent contributor in box 30, probably representing 24-hour persistence that was three hours old at 21 hours before forecast valid time.
- In box 61, the 24-hour persisted cloud cover was more prevalent at 15Z (approximately 10 a.m. local time) than at 3Z (approximately 10 p.m.).

Relative Humidity Variables

- Humidity terms in general were very significant predictors.
- Maximum layer relative humidity occurred frequently in the strongest fifteen for all box/season/time combinations.
- Mid-level relative humidity and transformations of that variable (ln(RH), RH^2, and RH^4) were also present frequently.
- Binary relative humidity variables were also in the strongest predictors for box 30.
- 3-, 6-, and 9-hour humidity change variables occurred with some regularity in box 30.

Other Variables

- Change in elevation along the local wind direction was a common predictor.
- Latitude was an important predictor for box 61, but less so for boxes 30 and 44.
- Mid-level u wind and low-level v wind components were common for box 30.

- 3-, 6-, and 9-hour changes in surface pressure were significant predictors in box 30.
- Low-level vorticity was a commonly-selected predictor across all three regions.
- Mid-level temperature occurred with some frequency in box 44.
- Mid- and high-level temperature advection turned up frequently in box 44.
- Vertical velocity was a stronger predictor in box 61 than in boxes 30 or 44.
- In box 61, the vertical velocity was more prevalent at 15Z (approximately 10 a.m. local time) than at 3Z (approximately 10 p.m.).
- Elevation and stability parameters occurred more often in the strongest fifteen predictors for box 44 at 0Z (approximately 6 p.m.) than at 15Z (approximately 9 a.m.), characteristic of afternoon heating and lifting over elevated terrain.

The strength and ubiquitousness of the simple cloud cover advection predictors surprised us. There were two possibilities for their predictive strength: 1) they represented regional average cloud cover amounts (these variables were averaged to half-mesh resolution) or 2) they really represented advection of regional cloud systems. To determine whether their predictive effects were due to advection or to area averaging we substituted winds rotated by 180° in the computations. The result was that the advection terms dropped out dramatically from the strongest predictors. Not only did this imply that these predictors represented advection, but it also reflected the accuracy of the GSM winds.

The last twelve cases in Table 4 correspond to models developed for specific terrain types (i.e., land or water). The strongest predictors for these terrain-specific cases were distinct from those developed for the same regions and times including all terrain types. This conclusion supports our findings during the clustering task: grid points of similar terrain type which grouped tightly in principal component space represented unique model development cases with unique cloud distribution characteristics. (We did not develop models for coastal grid points because there were not enough data points to support meaningful results.)

Many of the same observations made previously and discussed above for the 59 all-terrain-type cases are also true for these twelve terrain-specific cases (e.g., the overall strength of the advection and persistence terms). However, a few observations specific to these cases can also be made:

- Box 61 land/water cases differed less than box 30 land/water cases from their all-terrain-type counterparts.
- We saw a greater dependence on wind and divergence predictors in the water model than in the land model for all forecast lengths in box 30/summer.

- Box 61/winter showed a similar emphasis on wind-derived predictors.
- Cloud cover at the initial time was a stronger predictor over land than over water in box 30/summer.
- Temperature and temperature advection terms occurred in all three water cases in box 30/summer.

Regression Coefficients

After determining the strongest fifteen predictors for each data set, we determined the model coefficients (or weights) using the standard linear regression technique offered in STATISTICA™. In Tables 5 through 9 we show the raw regression weights (B) and the standardized regression weights (BSTD) for 5 of the 71 models (cases 24, 25, 27, 42, and 52 in Table 3). We selected these five cases to highlight in this results section because they represented a good cross section of the data. Three of the cases are based on one box/season/time combination, but have different forecast lengths (3, 6, and 9 hours). One can detect the effects of forecast length on model development by comparing their results. The other two cases are from different boxes, seasons, and times.

The magnitude of the raw B weights is *not* a measure of the relative strength of the predictors, but rather a reflection of the predictors' units and magnitude. The standardized regression weights are those we would have obtained had we first standardized all the predictors to mean 0 and standard deviation 1. Thus, the standardized values give us a measure of the relative contribution of each predictor to the overall cloud amount prediction equation. Positive weights reflect the positive partial correlations between the predictor and the predictand; e.g., mid-level relative humidity (RHM) and cloud cover in Table 7. Negative weights imply a negative relationship; e.g., 3-hour change in surface pressure (D3P) and cloud cover in Table 5.

Table 5 Regression Results for Case #24 (Box 30, Fall, 21Z, 3-Hour Forecast) Showing Top 15 Predictors, Regression Weights (Raw and Standardized), and Partial Correlations

VARIABLE NAME	B	BSTD	PARTIAL CORRELATION
ACCL3	.171	.140	7.62e–2
NEPH3	.307	.300	.252
ACCH3	.151	.123	7.93e–2
NEPH24	4.26e–2	4.26e–2	5.65e–2
ZENITH	–.161	–5.29e–2	–8.10e–2
RHM	3.48e–2	1.71e–2	1.61e–2
NEPH6	6.08e–2	5.61e–2	5.57e–2
ACCM3	.157	.128	5.82e–2
D3RHM	.331	5.00e–2	6.85e–2
D3P	–2.04e–4	–4.01e–2	–5.77e–2
RHL4	.104	5.02e–2	–5.39e–2
D3VL	–1.19e–2	–3.24e–2	–5.08e–2
DELEV	4.36	2.70e–2	4.30e–2
LON	1.27e–3	2.75e–2	4.19e–2
NEPH15	3.62e–2	3.59e–2	4.20e–2

Table 6 Regression Results for Case #25 (Box 30, Fall, 21Z, 6-Hour Forecast) Showing Top 15 Predictors, Regression Weights (Raw and Standardized), and Partial Correlations

VARIABLE NAME	B	BSTD	PARTIAL CORRELATION
ACCL6	.248	.193	.127
RHM	8.74e–2	4.29e–2	3.16e–2
NEPH6	.210	.194	.172
ACCH6	.120	9.25e–2	7.31e–2
ZENITH	–.154	–4.98e–2	–6.90e–2
NEPH21	7.05e–2	7.06e–2	8.60e–2
D6RHM	.305	7,68e–2	8.54e–2
RH70	6.21e–2	7.05e–2	6.56e–2
D6P	–1.44e–4	–5.24e–2	–6.52e–2
ACCM6	6.17	.100	6.09e–2
ELEV	4.10e–5	4.11e–2	5.71e–2
DELEV	6.17	3.81e–2	5.49e–2
UM	–2.57e–3	–5.38e–2	–6.74e–2
RHL4	.131	6.38e–2	5.97e–2
D6VL	–8.23e–3	–3.87e–2	–5.39e–2

Table 7 Regression Results for Case #27 (Box 30, Fall, 21Z, 9-Hour Forecast) Showing Top 15 Predictors, Regression Weights (Raw and Standardized), and Partial Correlations

VARIABLE NAME	B	BSTD	PARTIAL CORRELATION
ACCL9	.229	.176	.133
RHM	.179	9.54e–2	7.27e–2
NEPH9	.150	.137	9.39e–2
NEPH21	6.50e–2	6.51e–2	6.70e–2
ZENITH	–.201	–6.46e–2	–8.68e–2
UM	–4.04e–3	–8.71e–2	–.107
ACCM9	.141	.108	7.61e–2
NEPH18	5.91e–2	5.97e–2	5.53e–2
D9VTL	1.63e3	4.59e–2	5.82e–2
DELEV	6.30	3.89e–2	5.39e–2
ACCH9	8.73e–2	6.67e–2	6.00e–2
ELEV	5.01e–5	3.91e–2	5.24e–2
RH90	5.62e–2	5.88e–2	5.99e–2
D9RHM	.145	5.10e–2	5.83e–2
NEPH12	6.81e–2	6.44e–2	4.65e–2

Table 8 Regression Results for Case #42 (Box 44, Winter, 0Z, 3-Hour Forecast) Showing Top 15 Predictors, Regression Weights (Raw and Standardized), and Partial Correlations

VARIABLE NAME	B	BSTD	PARTIAL CORRELATION
ACCL3	.288	.238	.167
NEPH3	.341	.336	.330
ACCH3	.101	8.46e–2	6.67e–2
RHM2	6.47e–2	3.81e–2	4.24e–2
ACCM3	.111	9.26e–2	5.65e–2
TM	2.90e–2	4.99e–2	7.22e–2
VTL	6.54e+2	3.40e–2	5.04e–2
NEPH18	2.87e–2	3.03e–2	3.91e–2
TAH	1.37e+2	3.97e–2	5.72e–2
RH70	3.38e–2	3.98e–2	4.93e–2
D3P	–1.00e–2	–3.34e–2	–4.80e–2
NEPH24	3.27e–2	3.31e–2	4.37e–2
D3RHL	.187	2.90e–2	4.19e–2
TYPE	1.43e–2	1.92e–2	3.00e–2
LON	–8.00e–2	–1.84e–2	–2.80e–2

Table 9 Regression Results for Case #52 (Box 61, Summer, 15Z, 6-Hour Forecast) Showing Top 15 Predictors, Regression Weights (Raw and Standardized), and Partial Correlations

VARIABLE NAME	B	BSTD	PARTIAL CORRELATION
ACCH6	.149	.133	9.25e–2
LAT	–2.51e–2	–.233	–.158
NEPH24	9.08e–2	8.97e–2	9.19e–2
ACCM6	.121	.107	7.22e–2
WM	9.75e3	.159	.118
VTL	2.27e3	6.66e–2	6.57e–2
LNRHM	4.61e–2	4.19e–2	3.64e–2
D6WH	–9.95e3	–6.44e–2	–6.50e–2
D6UL	1.17e–2	5.06e–2	5.05e–2
D6P	–1.49e–4	–3.85e–2	–4.03e–2
RHAB	9.84e–2	3.70e–2	3.44e–2
NEPH6	3.80e–2	3.81e–2	3.22e–2
DELEV	1.75	2.86e–2	3.17e–2
VM	6.45e–3	3.20e–2	3.21e–2
UL	3.30e–3	3.70e–2	2.61e–2

Partial Correlations

One way to analyze the results of regression modeling is to look at partial correlations. Partial correlations are related to the contribution of each individual independent variable to the prediction of the dependent variable (cloud cover) after predictability derived from previously selected independent variables has been accounted for. The squared partial correlation is a measure of that part of the residual variance accounted for by each variable.

The partial correlations for the five sample cases are listed along with the regression weights in Tables 5 through 9. Note that in the tables, variables are ordered according to their order of selection from the original set of approximately 80 predictors. Within the fifteen predictors in each table, selection order would have been from the largest to smallest (magnitude) partial correlation.

Reduction in Variance

Another way to study the results of regression modeling is to look at the reduction in variance due to the individual predictors and due to the complete regression model. We can evaluate the MOS approach to cloud amount prediction by comparing the reduction in variance, R^2, from the MOS equations to that obtained using only the persisted cloud cover amount. That comparison (organized into three tables by box number), for the 71 cases we modeled, is shown in Tables 10a through 10c.

Table 10a Comparison of the Reduction in Variance (R^2) Due to MOS and Persistence for Box 30 Cases — Values Marked With "L" and "W" Correspond to Land and Water Cases, Respectively

VALID TIME (Z)	FORECAST LENGTH (Hours)	SPRING		SUMMER		FALL		WINTER	
		MOS	Persist	MOS	Persist	MOS	Persist	MOS	Persist
3	3	.5370	.3569						
3	6	.4909	.2761						
3	9	.4487	.2012						
9	3	.5831	.4638	.6222	.4946	.6430	.5336	.5672	.4381
9	6	.4868	.2982	.5434	.3494	.5538	.3649	.4831	.3052
9	6	.5005	.2982	.5614	.3494	.5563	.3649	.4780	.3052
9	9	.4468	.1761	.5221	.2489	.5043	.2534	.4199	.2086
21	3	.5230	.4048	.5510	.4572	.6096	.5010	.5385	.4295
21	6	.4132	.2293	.4448	.2776	.5283	.3484	.4300	.2651
21	6	.4013	.2293	.4297	.2776	.5194	.3484	.4212	.2651
21	9	.3661	.1627	.4002	.2232	.4921	.2938	.4063	.2092
21	3			$.5378^{L}$	$.4315^{L}$				
21	6			$.4286^{L}$	$.2587^{L}$				
21	9			$.3916^{L}$	$.1972^{L}$				
21	3			$.5311^{W}$	$.3831^{W}$				
21	6			$.3829^{W}$	$.1519^{W}$				
21	9			$.3808^{W}$	$.1347^{W}$				

Table 10b Comparison of the Reduction in Variance (R^2) Due to MOS and Persistence for Box 44 Cases

VALID TIME (Z)	FORECAST LENGTH (Hours)	SPRING		SUMMER		FALL		WINTER	
		MOS	Persist	MOS	Persist	MOS	Persist	MOS	Persist
0	3			.5062	.3519			.6124	.4784
0	6			.4181	.1910			.5340	.2300
0	9			.3727	.1109			.4636	.1890
15	3			.5378	.3513			.5258	.3785
15	6			.4250	.1760			.4448	.2192
15	9			.3938	.1288			.4247	.1954

Table 10c Comparison of the Reduction in Variance (R^2) Due to MOS and Persistence for Box 61 Cases — Values Marked With "L" and "W" Correspond to Land and Water Cases, Respectively

VALID TIME (Z)	FORECAST LENGTH (Hours)	SPRING		SUMMER		FALL		WINTER	
		MOS	Persist	MOS	Persist	MOS	Persist	MOS	Persist
3	3			.4168	.2549			.5594	.4186
3	6			.4378	.0688			.4055	.1761
3	9			.1677	.0518			.3843	.1426
3	3							$.5324^{L}$	$.4018^{L}$
3	6							$.3525^{L}$	$.1807^{L}$
3	9							$.3377^{L}$	$.1628^{L}$
3	3							$.5693^{W}$	$.3796^{W}$
3	6							$.5030^{W}$	$.1472^{W}$
3	9							$.5263^{W}$	$.1050^{W}$
15	3			.3387	.2106			.3979	.2684
15	6			.1940	.0721			.2930	.1693
15	9			.1830	.0625			.2884	.1619

From the entire 71 cases we were able to draw a few conclusions.

- The reduction in variance due to MOS or persistence decreases with increasing forecast length. That is, R^2 for a 3-hour forecast is greater than R^2 for a 6-hour forecast which is in turn greater than R^2 for a 9-hour forecast for a given box, season, and time.
- The percentage increase in the reduction of variance using the MOS approach over that using persistence varied substantially from one region and season to the next. The greatest increases were found in box 61 summer 3Z and winter 3Z (most notable was the box 61, winter, 3Z, 9-hour forecast for water grid points).
- R^2 due to persistence was higher in the winter than the summer in boxes 44 and 61. Its highest values were in box 30, fall. Lowest values in box 61, summer.
- The most variability in R^2 values was found in the box 61 cases.

In all of these cases, the MOS approach improved upon persistence because it used more information. Each additional variable that is used in the regression equation adds to the overall reduction in variance, though each successive predictor adds less and less to the overall reduction, in general.

In addition to studying the overall reduction in variance due to all fifteen predictors in each case we also studied the incremental changes in R^2 due to individual predictors.

Table 11[1] shows the cumulative reduction in variance for two sample cases (numbers 42 and 52) as successive variables were added to the regression equations. From the results presented in Table 11 it is clear that the bulk of the variance reduction comes from the first few predictors. In terms of variance reduction, our decision to retain fifteen predictors in the models was very conservative.

Table 11 Incremental Changes to R^2 for Cases #42 (Box 44, Winter, 0Z, 3-Hour Forecast) and #52 (Box 61, Summer, 15Z, 6-Hour Forecast)

Case#42 Box 44		Case#52 Box 61	
Variable	**Cumulative R^2**	**Variable**	**Cumulative R^2**
ACCL3	.5187788	ACCH6	.1166509
NEPH3	.5781473	LAT	.1387272
ACCH3	.5920656	NEPH24	.1539613
RHM2	.5962372	ACCM6	.1637530
TM	.5979326	WM	.1706838
TAH	.5996365	VTL	.1780713
NEPH24	.6011909	LNRHM	.1818255
RH70	.6025454	D6WH	.1851151
ACCM3	.6037927	D6UL	.1853130
D3P	.6048484	D6P	.1890684
VTL	.6057851	RHAB	.1901213
D3RHL	.6064843	NEPH6	.1910302
NEPH18	.6071277	DELEV	.1918519
TYPE	.6073927	VM	.1926227
LON	.6076998	UL	.1931708

Residual Analysis

A final step in our analysis was to examine the residuals. Linear regression assumes *linear* relationships between each of the independent variables and the predictand.

[1] The R^2 values included in Tables 10a through 10c correspond to the reduction in variance due to the strongest fifteen predictors of the pool of approximately 80 predictors used during the forward regression step. The R^2 values reported in Table 11 were determined during the regression modeling step when *only* the strongest fifteen predictors were used. Slight discrepancies between the numbers are due to the influence of the remaining predictors not used in the final regression modeling step.

Linear regression also assumes that the residual population density is unimodal and symmetric around zero. By analyzing a "normal plot of the residuals" (one of the options available through STATISTICA™) we were able to verify whether our model residuals were normal. Testing for normality is a stronger test than required by the assumptions of linear regression, but it can point to potential problems in the data.

A normal probability plot of the residuals is generated by plotting the actual residuals on the x axis and the z-values (standardized *assuming* a normal distribution) of the rank-ordered residuals on the y axis. If the residuals follow a normal distribution, the data lie along a straight line (the "normal" line). Gross deviations from normality show up as deviations from the normal line. In certain cases some non-linearities can be removed by transforming the data. Normal plots of the residuals for the five sample cases mentioned above are included in Figures 36 through 40.

In four of the cases that we looked at, the data did not show any large deviations from the ideal normal line. Significant deviations were found in box 61, shown in Figure 40. The most negative residual points that fall well below the normal line indicate a heavy-tailed distribution. The flattening of the curve around residual values between -0.5 and 0 indicates a tendency toward a multi-modal density function. A histogram of the residuals for the box 61 case (see Figure 41) clearly supports both of those indications. In contrast, Figure 42 shows a histogram of the model residuals from the case presented in Figure 38 that fell nicely along the normal line. In this case, the residual distribution clearly satisfies the zero mean and unimodal requirements assumed in linear regression.

3.4.2 Predicting Probabilities of Cloud Cover Categories

An introduction to the process of predicting probability of cloud cover categories was given in Section 3.1. We used the actual error distributions resulting from the regression modeling to determine those probabilities. First we divided the predicted cloud amounts into 11 bins: 0–4%, 5–14%, . . . , 85–94%, 95–100%. Then we generated histograms of the errors (binned in 0.1 intervals) by forecast cloud category. In general, these residuals have values between –1 and +1. For any given cloud amount estimate, $\hat{y}$, however, the residual values can range from a low of $-\hat{y}$ to a high of $1-\hat{y}$. Due to the bounded nature of cloud cover observations, the error histograms frequently followed U-, J-, or L-shaped distributions. Figure 43 shows the residual distributions for each of the forecast categories for case #27 (box 30/Fall/21Z/9-hour forecast). A histogram of the entire residual data set for this case was shown in Figure 42.

As shown in Section 3.1, the probability that the observed cloud cover will be less than or equal to a threshold value, c, is given by

$$\Pr\{Y \le c \mid \hat{Y}\} = \Pr\{E \le c-\hat{Y} \mid \hat{Y}\} \qquad (3\text{-}17)$$

where **Y** is the observed cloud cover, $\hat{Y}$ is the estimated cloud cover using the MOS equations, and **E** is the error. For example, to find the probability that the observed cloud cover will be ≤ 50% given that the predicted cloud amount is 80%, we determined the cumulative density (i.e., probability) below –0.3 (i.e., 0.5–0.8) in the error histogram for cloud cover forecasts in the range 75–84%. The probability of a single cloud cover category (i.e., that the observed cloud cover is between two thresholds) is simply the difference between the probabilities computed for each of the thresholds individually.

These calculations do not make any assumptions about the behavior of the errors. There is some approximation involved, however, due to the finite widths of the bins used to categorize cloud cover forecasts and the errors themselves.

We generated MOS-based probabilities of cloud cover categories at 10% intervals for all 71 cases listed in Table 3. Figures 44 through 48 show the results for the five sample cases discussed previously.

3.4.3 Modeling with the REEP Methodology

In Section 3.1 we introduced the regression estimation of event probabilities methodology to estimate cloud cover categories. In this section we discuss how we implemented the REEP approach, show results of a sample case (box 44/winter/3Z/3-hour forecast), and discuss why we decided to implement an alternate methodology to predict both total cloud cover amount and probability of cloud cover categories.

First, we transformed the response variable, total cloud cover (cc), to 6 categorical response variables defined in consultation with Phillips Laboratory personnel as

Category 1: cc = 0%
Category 2: 0 < cc ≤ 25%
Category 3: 25 < cc ≤ 50%
Category 4: 50 < cc ≤ 75%
Category 5: 75 < cc < 100%
Category 6: cc = 100%.

The result was a binary response vector for each observation containing 5 zeroes and 1 one.

The REEP methodology uses a common predictor set to build a regression equation for each category; this ensures that the probabilities sum to one. Therefore, the next step was to determine a set of predictors common to all of our six categories. We found that this step was somewhat subjective. We used the forward stepwise regression module to select the strongest predictors for each category individually. Then we made a histogram of the strongest predictors across all six categories and selected those variables that occurred most frequently. When two or more variables occurred with the same frequency, we selected that one which was the "stronger" predictor (i.e., occurred closer to the top of the list of predictors). Table 12 contains a list of the fifteen strongest predictors for each of the six categories as selected using forward regression. Notice that in one of the categories only 8 predictors were selected (i.e., only those eight were found to be significant using the F test provided in the STATISTICA™ model software). The last column contains those variables we chose for the *common* predictor set in the sample case.

Next, we used REEP to develop six models, one for each of the six categories, using the fifteen common predictors. The regression coefficients for each category in the sample case are listed in Table 13 (we list both raw and standardized weights as in the previous section).

Table 12 List of Strongest Predictors Selected for Each of the Six Cloud Cover Categories and the Common Predictor Set Used in the REEP Case (Box 44, Winter, 3Z, 3-Hour Forecast)

NUMBER	CATEGORY 1	CATEGORY 2	CATEGORY 3	CATEGORY 4	CATEGORY 5	CATEGORY 6	COMMON SET OF PREDICTORS
1	ACCL3	TDIF1	LAT	NEPH3	ACCL3	NEPH3	NEPH3
2	NEPH3	VTL	DAY	ELEV	ELEV	ACCL3	ACCL3
3	RHM	RH50	RHL2	MCC3	LAT	RHM2	LON
4	NEPH21	RHL2	NEPH3	UL	LON	ZENITH	RHL2
5	TL	ZENITH	TYPE	LNRHM	DAY	NEPH21	ELEV
6	ACCH3	WH	UL	TDIF1	SFCP	TN	TYPE
7	TYPE	LNRHM	D3VL	TYPE	DELEV	WH	SFCP
8	ELEV	RHM2	SFCP	DELEV	TAH	TAH	RH50
9	RH50		LON	SFCP	D3WL	RHM	TDIF1
10	D3VTH		WRHM	LON	NEPH3	ACCM3	ZENITH
11	TAM		RH90	RH50	NEPH6	NEPH9	NEPH21
12	LON		DVM	RHL2	D3RHL	RHL2	UL
13	TDIF1		TM	ACCL3	D3VTL	LAT	RHM2
14	D3WL		NEPH18	RHL	TAL	NEPH15	RHM
15	D3RHM		VTL	RHAB	TM	TAM	LNRHM

Table 13 Results for REEP Case (Box 44, Winter, 3Z, 3-Hour Forecast) Showing the Top 15 Predictors Common to All Categories and Corresponding Regression Weights (Raw and Standardized)

VARIABLE NAME	CATEGORY 1		CATEGORY 2		CATEGORY 3		CATEGORY 4		CATEGORY 5		CATEGORY 6	
	B	BSTD	B	BSTD	B	BSTD	B	BSTD	B	BSTD	B	BSTD
NEPH3	-.329	-.282	-9.13e-3	-1.08e-2	-2.12e-2	-5.44e-2	6.59e-2	8.75e-2	6.22e-3	1.29e-2	.287	.283
ACCL3	-.373	-.263	-3.52e-2	-3.42e-2	7.34e-3	1.55e-2	7.63e-2	8.32e-2	3.73e-2	6.36e-2	.287	.233
ELEV	-9.2e-5	-.121	1.7e-5	3.16e-2	-8e-6	-3.02e-2	5.0e-5	.101	1.6e-5	5.14e-2	1.7e-5	2.50e-2
LON	-1.80e-3	-3.58e-2	-6.26e-4	-1.71e-2	9.24e-4	5.49e-2	-1.68e-3	-5.16e-2	2.19e-3	.105	9.98e-4	2.28e-2
TYPE	-4.68e-2	-5.11e-2	1.98e-2	2.98e-2	6.06e-3	1.98e-2	1.01e-2	1.70e-2	4.60e-4	1.22e-3	1.04e-2	1.30e-2
RHL2	7.79e-2	4.06e-2	-9.23e-2	-6.62e-2	2.45e-2	3.81e-2	-6.04e-2	-4.86e-2	1.43e-2	1.80e-2	3.59e-2	2.15e-2
SFCP	-2e-6	-2.87e-2	2e-6	3.10e-2	-2e-6	-5.77e-2	3e-6	6.60e-2	-2e-6	-5.65e-2	< 1e-6	7.13e-3
RHM2	.166	8.89e-2	-.447	-.330	1.41e-2	2.26e-2	-.309	-.256	-5.11e-2	-6.62e-2	.627	.386
RH50	-9.78e-2	-5.99e-2	6.66e-2	5.62e-2	1.02e-2	1.86e-2	4.12e-2	3.90e-2	6.49e-3	9.61e-3	-2.67e-2	-1.88e-2
ZENITH	5.21e-2	8.82e-3	7.59e-2	1.77e-2	9.71e-2	4.91e-2	-8.56e-2	-2.24e-2	.147	6.02e-2	-.287	-5.57e-2
RHM	-.552	-.272	.577	.392	-2.41e-2	-3.55e-2	.443	.338	.106	.127	-.551	-.312
TDIF1	6.60e-3	5.17e-2	-4.51e-3	-4.86e-2	7.58e-4	1.77e-2	-4.41e-3	-5.34e-2	3.75e-4	7.10e-3	1.19e-3	1.07e-2
NEPH21	-6.81e-2	-6.05e-2	7.03e-3	8.60e-3	-3.02e-3	-8.01e-3	-3.15e-3	-4.33e-3	7.19e-4	1.55e-3	6.65e-2	6.79e-2
LNRHM	4.14e-2	5.06e-2	-3.65e-2	-6.16e-2	9.81e-3	3.58e-2	-3.23e-2	-7.05e-2	-1.31e-2	-3.87e-2	3.57e-2	5.02e-2
UL	-2.00e-3	-2.39e-2	1.24e-4	2.05e-3	-5.78e-4	-2.07e-2	1.59e-3	2.95e-2	-4.98e-4	-1.44e-2	1.36e-3	1.87e-2

The equations predict probabilities of cloud cover categories, but unlike true probabilities, these individual values are not required to be between zero and one. For example, the REEP model produces the following probability estimates at a single grid point on the first day of the sample case (Dec. 1, box 44):

Category 1: -0.120664
Category 2: 0.100010
Category 3: 0.013345
Category 4: 0.221180
Category 5: 0.054954
Category 6: 0.731176.

The probabilities sum to one as required, but are not bounded by 0 and 1 as a result of using a non-bounded linear regression model. The predicted total cloud amount can be determined from the probabilities using a prescribed selection criterion such as the most probable category or the median category. In the example above, either of these criteria would result in a cloud cover estimate of category 6; 100%.

Besides the two disadvantages described in Section 3.1,

1. high computational cost to determine a common predictor set for all six categorical response variables
2. probabilities outside the range [0,1],

our trial test showed that the data would not support predicting middle (partly-cloudy) categories. This is evident when looking at the reduction in variance, R^2, for the six cloud cover categories shown in Table 14. The reduction in variance using the REEP approach is relatively high for the two outer (clear and overcast) categories and very low for intermediate categories. There are very few data in these categories and thus little to support model development.

Table 14 Results of REEP Case Showing Reduction in Variance (R^2) for the Six Categorical Cloud Cover Regression Models. Note the Low Values for the Four Middle Categories

CATEGORY	TOTAL REDUCTION IN VARIANCE (R^2)
1	.3858
2	.0148
3	.0064
4	.0417
5	.0170
6	.3236

3.5 ESTIMATING CURRENT AND FUTURE MODEL PERFORMANCE

How does the MOS approach compare to current technology? How will the equations developed under this effort perform in the future with independent weather data? This section seeks to answer both of those questions.

First, we compare the MOS approach to persistence. This is not an arbitrary reference point. Persistence is, in general, a good predictor of cloud cover and is relied on heavily in certain parts of the world as a forecasting tool. For example, the TRONEW and HRCP models both use a type of persistence to model cloud cover. *Second*, we estimate future performance using the jackknife technique which subsamples the model development data set, develops independent models for each subset, then assesses performance variability when each model is applied to data not used in its development. *Finally*, we validate the cloud cover category forecasts by developing a model based on a subset of the development data set and comparing those model results to the results computed by applying the same model to the remainder of the data.

3.5.1 Comparison of MOS Approach to Persistence

To answer the question "How does the MOS approach compare to persistence?" one can use any one of a number of statistical performance statistics. These include: the F statistic for significance of the regression, the residual mean square, the coefficient of determination, the total squared error, etc. To compare the two forecast approaches then, one compares performance statistics. We chose to use performance statistics more familiar to the weather/forecasting community than those listed above to facilitate comparison with other forecast models.

We selected, in conjunction with Phillips Laboratory personnel, three performance statistics: the Brier score, the 20/20 score, and sharpness. Sharpness is not strictly a performance statistic (i.e., it is not a measure of the bivariate forecast-observation distribution), rather it is a measure of a single cloud cover distribution (e.g., of forecast data), which can be compared to other distributions (e.g., of observed data). That is, sharpness is a measure of the forecast and observed marginal distributions.

We begin our discussion with a description of the 21×21 bin contingency tables used to contain the forecast and observed data from which we compute the performance measures.

The Contingency Table

A contingency table is simply a two-dimensional histogram with observed cloud cover along one axis and forecast cloud cover (generated by MOS or persistence) along the other axis. We used a standard binning convention that assigns cloud cover (cc) to 21 increments as follows:

bin 1: $cc < 2.5\%$

bin 2: $2.5\% \leq cc < 7.5\%$

bin 3: $7.5\% \leq cc < 12.5\%$

•
•
•

bin 19: $87.5\% \leq cc < 92.5\%$

bin 20: $92.5\% \leq cc < 97.5\%$

bin 21: $cc \geq 97.5\%$.

The resulting contingency table has 21×21 elements (or cells). A schematic of a 21×21 contingency table is provided in Figure 49. For each of the 71 cases we analyzed, we built contingency tables containing one season of data (approximately 90 days) for 225 eighth-mesh grid points (15×15 points), a maximum of approximately 20,250 counts (box 61 cases had fewer points because a portion of the box contains "off world" points, see Figure 3). The total number of counts in the table is denoted by N. In most cases, N was much less than the maximum after accounting for missing and "old" cloud cover data points. The horizontal (i) axis of the table in Figure 49 corresponds to the observed cloud cover. The vertical (j) axis of the table corresponds to the forecast cloud cover. For any (i, j) cell, the number of counts in the cell is denoted by n_{ij}.

The table is filled one count at a time by determining the cell that corresponds to the forecast value along the row axis and observed cloud cover along the column axis. If the forecast is perfect at every point over the entire season, all of the entries in the contingency table will lie along the diagonal where forecast cloud cover is equal to observed cloud cover.

The Brier score, 20/20 score, and sharpness are measures of how the counts are distributed throughout the contingency table. They are described below.

The Brier Score

The Brier score (Ref. 3), measures the mean squared difference between the forecast (F) and observed (O) cloud cover amounts (each ranges from 0.0 to 1.0) from the 21×21 contingency table. The Brier score, S_{Brier}, is computed by summing over all data points in a season as follows:

$$S_{Brier} = \frac{1}{N} \sum_{j=1}^{21} \sum_{i=1}^{21} n_{ij}(F_j - O_i)^2. \qquad (3\text{-}18)$$

This score is particularly sensitive to counts in the extreme off-diagonal cells due to the quadratic term. The Brier score ranges from a perfect (minimum) value of zero, corresponding to all data points lying along the diagonal, to a maximum of one, corresponding to all data points lying in the extreme bins.

We computed Brier scores for the MOS-based forecast and the persistence forecast for each of the model cases. Since the Brier score measures mean squared error and linear least squares regression minimizes squared error, we expected that the MOS-based forecasts would perform well by this metric. Indeed, MOS outperformed persistence in *all* of the 71 cases. The scores are listed in Table 15. Figures 50 through 52 show bar graphs of the Brier scores for each of the RTNEPH boxes that we studied: 30, 44, and 61, respectively.

The cyclic behavior evident in Figures 50 through 52 is caused by ordering the model cases by forecast length (refer to Table 3). The cases are listed in order of 3-, 6- (some cases have two 6-hour forecasts), and 9-hour forecasts. By looking at those cases in Box 30 for which we generated two 6-hour forecasts (one each for the 0Z and 12Z initialized GSM data), we see that there is no evidence to suggest that one initialization time is inherently better than another. This implies that the MOS equations compensated for any model spin-up problems that may have existed in the GSM.

If we explicitly average all the forecasts of a given length from one box and compare the MOS-based and persistence-based average scores for differing forecast lengths (see Figure 53), it is clear that Brier scores increase with increasing forecast length. That is, S_{Brier} for a 9-hour forecast is greater than S_{Brier} for a 6-hour forecast which is greater than S_{Brier} for a 3-hour forecast (for both persistence and MOS). This implies the obvious, that shorter forecasts are more skillful (i.e., have less error) than longer forecasts.

Table 15 Brier Scores, 20/20 Scores, and Sharpness for the 71 Model Cases Introduced in Table 3

#	CASE NAME	N	BRIER MOS	BRIER PERSISTED	20/20 MOS	20/20 PERSISTED	SHARPNESS MOS	SHARPNESS PERSISTED	SHARPNESS OBSERVED
1	30/SP/3/3/12	15596	0.0942	0.1684	0.5651	0.6238	0.4877	0.8007	0.8222
2	30/SP/3/6/12	14093	0.1017	0.1865	0.5365	0.5911	0.4632	0.7925	0.8251
3	30/SP/3/9/12	12604	0.1092	0.2011	0.4821	0.5346	0.4084	0.7111	0.8294
4	30/SP/9/3/0	15347	0.0679	0.1115	0.6767	0.6644	0.5499	0.7977	0.7076
5	30/SP/9/6/0	13779	0.0833	0.1566	0.5956	0.5852	0.4426	0.8222	0.7037
6	30/SP/9/6/12	13378	0.0839	0.1566	0.6008	0.5846	0.4531	0.8201	0.7031
7	30/SP/9/9/12	11842	0.0932	0.2200	0.5285	0.5016	0.3653	0.8035	0.7026
8	30/SP/21/3/12	16139	0.0909	0.1311	0.5867	0.6201	0.4988	0.7030	0.7896
9	30/SP/21/6/12	14572	0.1112	0.1835	0.4677	0.5149	0.3627	0.6330	0.7929
10	30/SP/21/6/0	15013	0.1143	0.1838	0.4635	0.5129	0.3602	0.6311	0.7956
11	30/SP/21/9/0	13367	0.1216	0.2090	0.4286	0.4828	0.3221	0.6301	0.7991
12	30/SU/9/3/0	14926	0.0589	0.1001	0.7129	0.6733	0.5794	0.8062	0.7028
13	30/SU/9/6/0	13232	0.0714	0.1484	0.6354	0.5859	0.4825	0.8246	0.6913
14	30/SU/9/6/12	13274	0.0712	0.1464	0.6418	0.5887	0.4925	0.8251	0.6927
15	30/SU/9/9/12	11480	0.0777	0.1801	0.5962	0.5374	0.4280	0.7992	0.6855
16	30/SU/21/3/12	16283	0.0723	0.1069	0.6549	0.6420	0.5637	0.6908	0.7835
17	30/SU/21/6/12	14448	0.0915	0.1588	0.5550	0.5263	0.4519	0.6071	0.7798
18	30/SU/21/6/0	14687	0.0939	0.1583	0.5431	0.5293	0.4424	0.6086	0.7809
19	30/SU/21/9/0	12860	0.1018	0.1808	0.5098	0.4932	0.4061	0.5953	0.7792
20	30/FA/9/3/0	15033	0.0620	0.0980	0.7131	0.7098	0.6100	0.8104	0.7560
21	30/FA/9/6/0	13226	0.0777	0.1401	0.6299	0.6373	0.5305	0.8480	0.7583
22	30/FA/9/6/12	13428	0.0785	0.1409	0.6302	0.6335	0.5308	0.8474	0.7569
23	30/FA/9/9/12	11705	0.0884	0.1901	0.5742	0.5800	0.4556	0.8421	0.7586
24	30/FA/21/3/12	16119	0.0745	0.1075	0.6739	0.6999	0.6088	0.7864	0.8200
25	30/FA/21/6/12	14349	0.0918	0.1438	0.5881	0.6037	0.5144	0.7051	0.8260
26	30/FA/21/6/0	14375	0.0937	0.1439	0.5797	0.6061	0.5087	0.7082	0.8269
27	30/FA/21/9/0	12556	0.0999	0.1591	0.5530	0.5804	0.4875	0.6984	0.8327
28	30/WI/9/3/0	14110	0.0741	0.1163	0.6698	0.7055	0.5972	0.8327	0.7996
29	30/WI/9/6/0	13759	0.0900	0.1568	0.5826	0.6429	0.4935	0.8592	0.7982
30	30/WI/9/6/12	12973	0.0929	0.1565	0.5813	0.6457	0.5044	0.8632	0.8011
31	30/WI/9/9/12	11269	0.1047	0.2038	0.5027	0.5920	0.4080	0.8562	0.8016
32	30/WI/21/3/12	15025	0.0921	0.1341	0.6045	0.6815	0.5504	0.8152	0.8384
33	30/WI/21/6/12	13293	0.1138	0.1758	0.4970	0.5946	0.4373	0.7560	0.8450
34	30/WI/21/6/0	13483	0.1157	0.1765	0.4932	0.5956	0.4309	0.7571	0.8451
35	30/WI/21/9/0	13580	0.1216	0.1949	0.4590	0.5682	0.3927	0.7555	0.8433
36	44/SU/0/3/12	13131	0.0631	0.1021	0.6438	0.5749	0.3588	0.3616	0.3917
37	44/SU/0/6/12	11605	0.0754	0.1505	0.5596	0.4525	0.2790	0.4401	0.3990
38	44/SU/0/9/12	10142	0.0827	0.2083	0.5193	0.3847	0.2367	0.6783	0.4077
39	44/SU/15/3/0	16411	0.0840	0.1500	0.6143	0.5976	0.5315	0.7466	0.7230
40	44/SU/15/6/0	14441	0.1055	0.2065	0.4964	0.4750	0.3971	0.6588	0.7231

Table 15 Brier Scores, 20/20 Scores, and Sharpness for the 71 Model Cases Introduced in Table 3 (Continued)

#	CASE NAME	N	BRIER MOS	BRIER PERSISTED	20/20 MOS	20/20 PERSISTED	SHARPNESS MOS	SHARPNESS PERSISTED	SHARPNESS OBSERVED
41	44/SU/15/9/0	12609	0.1118	0.2325	0.4644	0.4601	0.3620	0.7066	0.7274
42	44/WI/0/3/12	12450	0.0675	0.1044	0.6746	0.6225	0.5369	0.5815	0.6223
43	44/WI/0/6/12	11821	0.0810	0.1569	0.5982	0.5225	0.4733	0.6304	0.6277
44	44/WI/0/9/12	10288	0.0929	0.2148	0.5372	0.4791	0.4185	0.7822	0.6358
45	44/WI/15/3/0	14197	0.0935	0.1573	0.5974	0.6338	0.5566	0.8196	0.7988
46	44/WI/15/6/0	11353	0.1115	0.2164	0.4974	0.5538	0.4548	0.7973	0.8108
47	44/WI/15/9/0	11746	0.1173	0.2350	0.4631	0.5298	0.4133	0.7981	0.8101
48	61/SU/3/3/12	11025	0.0843	0.1557	0.6185	0.6461	0.5790	0.8107	0.8639
49	61/SU/3/6/12	10020	0.1155	0.2057	0.5250	0.5580	0.5247	0.7676	0.8651
50	61/SU/3/9/12	8977	0.1156	0.2070	0.5347	0.5579	0.5307	0.7716	0.8635
51	61/SU/15/3/0	13866	0.0949	0.1645	0.5692	0.6310	0.5107	0.8307	0.8095
52	61/SU/15/6/0	12354	0.1170	0.2130	0.4803	0.5717	0.4175	0.8621	0.8095
53	61/SU/15/9/0	12014	0.1165	0.2131	0.4797	0.5702	0.4148	0.8668	0.8114
54	61/WI/3/3/12	11043	0.0871	0.1381	0.6027	0.6653	0.5519	0.8183	0.8522
55	61/WI/3/6/12	10166	0.1168	0.2111	0.4777	0.5461	0.4339	0.7561	0.8506
56	61/WI/3/9/12	9000	0.1203	0.2287	0.4590	0.5253	0.4127	0.7541	0.8511
57	61/WI/15/3/0	12771	0.1062	0.1780	0.4973	0.5917	0.3999	0.8235	0.7691
58	61/WI/15/6/0	11827	0.1260	0.2220	0.4032	0.5383	0.2894	0.8484	0.7731
59	61/WI/15/9/0	10536	0.1275	0.2271	0.3942	0.5296	0.2812	0.8492	0.7738
60	30/SU/21/3/12L	11143	0.0783	0.1147	0.6178	0.6091	0.5147	0.6557	0.7629
61	30/SU/21/6/0L	10329	0.0985	0.1628	0.5183	0.4997	0.4069	0.5692	0.7633
62	30/SU/21/9/0L	9289	0.1061	0.1873	0.4875	0.4596	0.3701	0.5550	0.7615
63	30/SU/21/3/12W	2617	0.0487	0.0793	0.7746	0.7658	0.7318	0.8345	0.8785
64	30/SU/21/6/0W	2102	0.0701	0.1435	0.6974	0.6556	0.6637	0.7897	0.8739
65	30/SU/21/9/0W	1719	0.0738	0.1573	0.6766	0.6358	0.6492	0.7865	0.8802
66	61/WI/3/3/12L	8331	0.0883	0.1397	0.5942	0.6578	0.5337	0.8126	0.8548
67	61/WI/3/6/12L	7801	0.1212	0.2023	0.4583	0.5533	0.4073	0.7586	0.8523
68	61/WI/3/9/12L	7073	0.1241	0.2178	0.4383	0.5395	0.3863	0.7645	0.8517
69	61/WI/3/3/12W	1909	0.0763	0.1371	0.6469	0.6867	0.6097	0.8481	0.8523
70	61/WI/3/6/12W	1651	0.0887	0.2532	0.5942	0.5257	0.5584	0.7953	0.8474
71	61/WI/3/9/12W	1738	0.0908	0.3192	0.5978	0.4465	0.5748	0.7658	0.8550

In this effort we did not have an *absolute* standard by which to evaluate the MOS equations. Instead, a useful measure of performance was the *relative* improvement of MOS-based forecasts over persistence using the Brier score. In general, we found that Brier scores for MOS-based forecasts were 35–55% smaller than their persistence-based counterparts. A quick glance at Figures 50 through 52 shows the qualitative improvement of the MOS-based forecasts over persistence.

In Figure 54 we show the averaged percentage improvement of MOS over persistence as a function of forecast length for each box. That figure shows that the relative improvement of MOS over persistence increases with increasing forecast length as the influence of persistence in the forecasts decreases.

The 20/20 Score

The 20/20 score ($S_{20/20}$) measures the fraction of the counts in the contingency table that are within 20% (i.e., 4 bins) of the diagonal on either side. It is calculated from the contingency table from the following:

$$S_{20/20} = \frac{1}{N} \sum_{j=1}^{21} \sum_{i=\max(1,j-4)}^{\min(21,j+4)} n_{ij}. \qquad (3\text{-}19)$$

The 20/20 scores for all of the 71 cases (grouped by RTNEPH box number) are displayed in Figures 55 through 57. The scores are also listed by case number in Table 15. A perfect forecast would have a 20/20 score equal to one, since all the data points would lie along the diagonal. A score of zero (the minimum possible score) would correspond to all of the data points lying outside 20% of the diagonal of the contingency table.

Unlike the results from our analysis of the Brier score, MOS did not outperform persistence consistently with respect to the 20/20 score. Similar to the Brier score results however, we found a trend toward decreasing skill with increasing forecast length, as expected.

Results varied considerably as functions of the region, season, and time of day. For example, in the box 44 cases, MOS outperformed persistence (i.e., MOS had higher 20/20 scores than persistence) consistently for the 0Z forecast valid time (both winter and summer), but did not perform as well for the 15Z forecast valid time. In box 61, we found that MOS outperformed persistence only in two cases, both of which contained only water grid points (case numbers 70 and 71). In box 30, overall scores (both persistence and MOS) were higher for the cases with valid forecast time of 9Z than 21Z in all four seasons and were highest for the all water case (summer, 21Z, case number 63). $S_{20/20}$ for the MOS-based forecasts were higher than those corresponding to persistence for many of the 9Z cases in box 30.

We found that for most box/season/times, MOS forecasts were consistently higher or lower than persistence over all forecast lengths. For example, if MOS performed better than persistence for a 3-hour forecast for a given box, season, and valid time, it generally performed better for the 6- and 9-hour forecasts also.

As we saw in our analysis of the Brier score results of those Box 30 cases in which we developed two forecast models (one using the 0Z GSM data and the other using the 12Z data), there was no evidence to suggest that there was any measurable difference in performance. That is, the MOS-based forecast models accounted for any existing GSM spin-up problems as measured by the 20/20 skill score as well as the Brier score.

Sharpness

Sharpness is a measure of a single cloud cover distribution. It measures the weight of the distribution within 20% (i.e., 4 bins) of the outer bins. That is, it is the number of data points whose cloud cover is less than 22.5% or greater than 77.5%.

We computed the sharpness of the observed and forecast (MOS-based and persistence) marginal distributions directly from the contingency table using the following:

$$\text{Sharpness (Observed)} = \frac{1}{N} \sum_{j=1}^{21} \left\{ \sum_{i=1}^{5} n_{ij} + \sum_{i=17}^{21} n_{ij} \right\} \tag{3-20}$$

$$\text{Sharpness (Forecast)} = \frac{1}{N} \sum_{i=1}^{21} \left\{ \sum_{j=1}^{5} n_{ij} + \sum_{i=17}^{21} n_{ij} \right\}. \tag{3-21}$$

Sharpness ranges from a maximum of one, corresponding to a distribution in which all of the data points are in the outer 20% bins, to a minimum of zero, corresponding to all the data points falling in the interior 60% bins.

Because sharpness characterizes a cloud cover distribution, we gain meaning not from the value itself, but from comparing the sharpness of one distribution to another. We computed the sharpness for the ground truth (observed) cloud cover, the persisted cloud cover, and the MOS-based forecast cloud cover. Those values are included in Table 15. The bar graphs in Figure 58 through 60 show the sharpness values for observed and forecast cloud cover for the 71 cases grouped by RTNEPH box number.

Observations of cloud cover tend to have U-, J-, or L-shaped distributions with most values occurring in the outer (clear and overcast) bins. An example of this type of distribution was presented in Figure 12. Because persisted cloud cover is simply observed cloud cover that has been shifted in time, we expect similar sharpness values for those distributions. The MOS-based forecasts were less sharp than either the observed or persisted

cloud cover. That was to be expected since the MOS technique (based on least squares minimization of errors) tends to predict conservative (i.e., tending toward the mean) cloud cover amounts. Figures 61 and 62 show observed and forecast cloud cover distributions, respectively, with very different sharpness values.

In fact, MOS-based forecasts become increasingly conservative (i.e., less sharp) with forecast length, such that a 3-hour forecast is more sharp than a 6-hour forecast which is more sharp than a 9-hour forecast. This effect is shown clearly in Figure 63 and has been reported by others as well (Ref. 17).

One method of producing sharper forecasts is to replace the least squares minimization in the linear regression with a method that minimizes absolute error instead. Also, as discussed in Ref. 26, one could use the logistic response function in categorical cloud amount forecasts to better model the shape of the cloud cover distributions. Both of these methods are more difficult and expensive than linear regression because they require iteration and good initial estimates of model coefficients.

The technique that we employed in this feasibility study emphasized reduction in total squared error and thus performs very well with respect to the Brier score. In some operational settings, forecast sharpness may be crucial, and therefore a non-linear technique that yields sharper forecasts would be preferred.

3.5.2 Estimation of Performance Using the Jackknife Technique

To answer the question "How will the models developed under this effort perform in the future with independent weather data?" we used the jackknife resampling technique to estimate the performance of two models and determine the confidence limits on those estimates.

The jackknife technique is one of a number of resampling techniques that estimates model performance (Ref. 12). It improves upon earlier techniques such as cross validation which divides the data set in two (using one half for model development and the other half for validation) by using *more* of the data for development and computing *multiple* performance measures using different subsamples of the data. Due to the limited size of our data set, it was necessary to keep as much of the data as possible during evaluation; the jackknife technique allowed us to do that. In addition, the jackknife technique accounts for any bias that might be present when small samples are used in the estimation. The method consists of the following steps.

STEP 1 Split the model development data set into M groups of equal size. We used 9 groups of 10 contiguous days each over a season.

STEP 2 Leave out one group of data and develop the regression model (i.e., determine model coefficients). Repeat the process for each of the M groups, leaving out a different group of data in each model.

STEP 3 Validate each model with the omitted group of data. Compute a performance statistic. We used the Brier score.

STEP 4 Use the sample of M performance statistics along with the performance computed using the *whole* data set to compute average performance and the variance of the average.

We selected two model development data sets from the group of 71 for which to perform the jackknife analysis. The first case (#12 from box 30) had the lowest Brier score (i.e., the highest performance) of all the 71 cases. The second case (#59 from box 61) had the highest Brier score (i.e., lowest performance) of all the cases. It also contained fewer grid points due to the geometry of the box and thus the presence of off-world points. We selected these two cases to provide upper and lower bounds for the results of our jackknife analysis.

The jackknife estimates are computed from the M individual performance estimates as follows:

y_{all} is the estimate using all data

y_j is the estimate with the jth group omitted

y_j^* is a "pseudo" estimate defined by

$$y_j^* = M\, y_{all} - (M-1)\, y_j \tag{3-22}$$

y^* is the jackknife performance estimate

$$y^* = \frac{1}{M} \sum_{j=1}^{M} y_j^* \tag{3-23}$$

s_*^2 is the estimated variance of y^*

$$s_*^2 = \frac{1}{M(M-1)} \sum_{j=1}^{M} [y_j^* - y^*]^2. \tag{3-24}$$

The y_{all}, y_j, y^*, and s_*^2 values, computed for the two cases, are presented in Table 16. There is little change in estimated performance using the jackknife analysis (y^*) versus using the whole data set (y_{all}) in both cases. This implies that the MOS models are robust with respect to the variability in the data sets.

In a preliminary analysis we divided the data set into 9 groups by sampling every ninth grid point from the 15 × 15 grid points in the RTNEPH box over one season. (Recall that we used 15 × 15 grid points to allow us to compute horizontal gradients for divergence, vorticity, etc.) We found that the variance in the estimated performance was about a factor of five smaller than the numbers shown in Table 16 which were generated using the 9 10-day data sets. This was due to the fact that our preliminary sampling technique which included every ninth grid point had not accounted for the high spatial correlation in the cloud cover data. Thus our 9 model data sets were highly correlated. By sampling 9 10-day periods instead, as we did in our final analysis, the model development data sets were less correlated. However, due to the nature of cloud cover data, the resulting variances (those reported in Table 16) are most likely still overly optimistic. Ideally, one would evaluate the performance of the MOS-based models against completely independent data taken from another year.

Table 16 Results of Jackknife Analysis For Cases #12 and #59. The Average Performance (y^*) and Variance (s_*^2) for Both Cases are Highlighted

	Case #12	Case #59
y_{all}	.0589	.1275
y_1	.0614	.1342
y_2	.0595	.1308
y_3	.0643	.1326
y_4	.0658	.1134
y_5	.0551	.1266
y_6	.0577	.1419
y_7	.0544	.1369
y_8	.0549	.1313
y_9	.0586	.1263
y^*	.0575	.1039
s_*^2	5.332e-4	4.578e-4

3.5.3 Verifying MOS-based Probabilities of Cloud Cover Categories

We used a Bayesian approach to assess the likelihood that the MOS-based forecast probabilities are "correct". That is, that they accurately represent the frequency of occurrence, in nature, of the events of interest. For this evaluation, we selected the same two model development data sets used in the evaluation of the MOS-based total cloud amount forecasts presented in Section 3.5.: Case #12 from box 30 and Case #59 from box 61. These two cases represent the highest and lowest performance (as measured by the accuracy of the total cloud amount forecasts) out of the set of 71 listed in Table 3.

The following paragraphs describe our approach in general terms and the way in which the available data are used in its implementation. Following that description, a key formula based on the Bayesian viewpoint is presented. (Derivation of this formula is provided in Appendix B.) Finally, the results are presented in tabular and graphical forms accompanied by a discussion of their implications.

The Bayesian Viewpoint

To carry out the evaluation task, we must provide a careful definition of the "events of interest" mentioned above. MOS-based probability forecasts have been developed for 71 cases, each defined by a region (i.e., an RTNEPH box), a season, a forecast valid time, and a forecast length. In each of the two cases we analyzed, we considered 27 "events of interest" as follows:

> For each one of nine cloud-cover fraction (ccf) bins, B_i, (i.e., B_1 = "0.1" represents {5% < ccf <= 15%} , B_2 = "0.2" represents {15% < ccf <= 25%}, etc.)
>
> and
>
> For each one of three specified cloud-cover threshold values, T_j, {0.3, 0.5, 0.7},
>
> The event E(i, j) occurs at the specified valid time if the MOS cloud cover forecast for that time is in bin B_i and the actual ccf at that time is less than or equal to T_j.

For an ideal validation, one would observe several thousand occurrences of each ccf forecast bin for the region, season, valid time and forecast length of interest. One could then count the number of times that the actual ccf is less than each threshold of interest and accept or reject hypotheses regarding the validity of the MOS forecast probabilities based on the observed frequency of occurrence.

Since such vast amounts of data are not available for any one model case, we adopted a Bayesian viewpoint in which MOS-based probabilities of the 27 events, E(i, j), are generated using one-half of the available data (Ref. 18). These probabilities are then viewed as prior information, establishing a *prior probability measure* on the unknown, true probability that E(i, j) will occur. The other half of the original data is then used to update that prior probability measure, obtaining a *posterior probability measure* on the true probability. If the new data substantially increase the weight at the MOS-based value, we will have shown consistency among at least the limited data set available. Conversely, if the weight is drastically decreased at the MOS value, then applicability of the MOS model for the given forecast valid and lead times) is called into question.

A Formula for Posterior Probability

Results presented in this section are calculated using the following formula for the posterior probability (p^*) that the true probability of the event of interest (p), is equal to the MOS forecast probability ($\hat{p}$).

$$p^* = \frac{L(\hat{p}/D_2)}{L(\hat{p}/D_2) + I(n,k)} \tag{3-25}$$

where

$L(\hat{p}/D_2) = \hat{p}^k(1-\hat{p})^{n-k}$ = likelihood of $\hat{p}$ given data set D_2

n = total number of observations in data set D_2

k = number of times that E(i,j) occurred in D_2

and

$$I(n,k) = \int_0^1 x^k(1-x)^{n-k}\,dx. \tag{3-26}$$

To evaluate the integral, we used the following M-term approximation:

$$I(n,k) = \sum_{m=1}^{M} x_m^k(1-x_m)^{n-k}\Delta_m \tag{3-27}$$

where (Δ_m, x_m) are the width and midpoint of the m^{th} interval in the sum. All of the results presented here were calculated with M = 100 and Δ_m = 0.01. Typically, only fifteen terms of the sum surrounding the maximum of the integrand were required to evaluate the sum accurately to six significant digits. This is because of the integrand's extremely narrow peak caused by the large values of n and k for most cases (e.g., n = 729 and k = 568 for case #12, ccf category 0.3, with threshold = 0.5).

Results

We considered two cases, each defined by a region (i.e., an RTNEPH box), season, forecast valid time, and forecast length:

Case #12 — Box 30, summer, three hour forecast valid at 09Z

Case #59 — Box 61, winter, nine hour forecast valid at 15Z.

The data sets for each case were divided into the first and second halves of each season. Data set 1 (columns 2, 5 and 8 of Tables 17 and 18) was used to develop MOS-based models of probability as described in Section 3.4.2 of this report. The "Data Set 1" columns in the tables are the resulting MOS-forecast probabilities. Data set 2 was then used as "new" data for verifying the correctness of those forecasts. The "Data Set 2" columns in the tables are the frequencies of occurrence as described above.

If the MOS models are any good at all, we would expect that the MOS forecast probabilities and the frequencies of occurrence would be reasonably close to each other. In fact, looking at Tables 17 and 18, this is generally the case. But, are the differences significant? A Bayesian viewpoint will help to answer that question.

We will discuss two examples, forecast ccf categories 0.6 and 0.7 with threshold = 0.3 in each example, and use the data for case #59 (Table 18) to illustrate the approach. The events of interest are that the actual ccf is less than 30% when the MOS forecast ccf is between 55% and 65% or between 65% and 75%, respectively. We assume that the unknown, true probability is a random variable distributed on [0, 1] with 50% of the probability at the MOS-value (0.30 and 0.18 for our two examples) and the other 50% distributed uniformly on [0, 1]. We then calculate the *posterior probability* at the MOS-value based on processing the new (Data Set 2) data. These posterior probabilities are, from the columns labeled "Probability" in Table 18, 0.31 and 0.96 for our two examples.

Table 17 MOS Probability Verification (Case #12)

FORECAST CLOUD-COVER FRACTION	Threshold (Lambda) Values								
	0.3			0.5			0.7		
	Data Set 1	Data Set 2	Probability	Data Set 1	Data Set 2	Probability	Data Set 1	Data Set 2	Probability
0.1	0.93	0.93	0.98	0.98	0.97		0.98	0.98	0.97
0.2	0.75	0.79		0.87	0.9		0.95	0.96	0.96
0.3	0.58	0.63		0.76	0.78	0.92	0.92	0.9	0.88
0.4	0.41	0.46		0.63	0.64	0.94	0.84	0.86	0.92
0.5	0.32	0.32	0.95	0.49	0.5	0.94	0.76	0.73	0.88
0.6	0.23	0.24	0.94	0.4	0.4	0.95	0.68	0.65	0.87
0.7	0.1	0.12	0.92	0.23	0.24	0.95	0.53	0.5	0.85
0.8	0.04	0.04	0.98	0.13	0.11	0.88	0.38	0.3	
0.9	0.02	0.01	0.96	0.05	0.04	0.94	0.18	0.18	0.96

Values labeled "Probability" are posteriori probabilities that the prior probabilities labeled "Data Set 1" are correct

Indicates Strong rejection of the prior value

Indicates Neutral opinion regarding the prior value

Indicates Strong affirmation of the prior value

Table 18 MOS Probability Verification (Case #59)

FORECAST CLOUD-COVER FRACTION	Threshold (Lambda) Values								
	0.3			0.5			0.7		
	Data Set 1	Data Set 2	Probability	Data Set 1	Data Set 2	Probability	Data Set 1	Data Set 2	Probability
0.1	0.9	0.85		0.94	0.9		0.97	0.94	0.8
0.2	0.78	0.78	0.95	0.86	0.82		0.92	0.89	
0.3	0.63	0.65	0.92	0.72	0.71	0.95	0.83	0.81	0.87
0.4	0.52	0.55	0.81	0.62	0.61	0.95	0.76	0.73	0.85
0.5	0.39	0.37	0.94	0.49	0.46	0.85	0.61	0.58	0.85
0.6	0.3	0.25		0.41	0.33		0.53	0.45	
0.7	0.18	0.19	0.96	0.27	0.25	0.92	0.39	0.34	
0.8	0.15	0.13	0.92	0.2	0.17		0.29	0.25	
0.9	0.1	0.09	0.94	0.15	0.12	0.9	0.22	0.23	0.92

Values labeled "Probability" are posteriori probabilities that the prior probabilities labeled "Data Set 1" are correct

Indicates Strong rejection of the prior value

Indicates Neutral opinion regarding the prior value

Indicates Strong affirmation of the prior value

Thus, for the first example (MOS forecast category 0.6 and a 30% ccf threshold) considering the second data set reduced our confidence that the MOS forecast probability (0.30) is correct from 0.5 (the prior probability) down to 0.31. On the other hand, for a MOS forecast ccf in the 0.7 bin, the probability is increased substantially from 0.5 up to 0.96. In the former case, considerable variability in the data over the season of interest is indicated, casting doubt on the "correctness" of the resulting MOS model. In the latter case, however, the second data set considerably *strengthened* our confidence in the MOS forecast, indicating that *at least for the one season of data available, the MOS forecast probability is correct.*

In the two tables we have used shading to identify cases in which the MOS forecast probability is "strongly rejected" and "marginally accepted" (or "neutral"). Somewhat arbitrarily, we used posterior probability ranges [0.0, 0.3] and [0.3, 0.8] to define these two designations.

Figures 64 through 69 present cross-sections of the data from the two tables which highlight the sensitivity of our results to variations between the two parts of the partitioned data sets. In each figure, the line graphs show the MOS forecast probabilities (solid squares, Data Set 1) and frequencies of occurrence (hollow squares, Data Set 2) as functions of the forecast ccf category. The superimposed bars represent the posterior probabilities for each case and clearly show those cases for which the data set partitions are most inconsistent.

While it is difficult to generalize regarding why certain of the MOS forecasting models are rejected, some understanding of the causes can be obtained by examining a few specific examples. We will consider the case #12 (box 30) results found in Table 17 and plotted in Figures 64 through 66 and examine the evaluation results for three specific ccf forecast categories in more detail.

For ccf forecast category 0.7, Table 17 reveals that the MOS model was strongly affirmed for all three thresholds (0.3, 0.5, and 0.7). Figure 70 contains plots of the experimental distribution functions of the true ccf (i.e., the RTNEPH analysis values) for only those occasions at which the MOS-based ccf forecast category was 0.7. In the figure, arrows indicate the threshold locations; the intercepts of the Data Set 1 and 2 curves with those thresholds define the probabilities found in Table 17 under Forecast Cloud-Cover Fraction 0.7. For this forecast category, the two distribution functions are close to another at the threshold values, consistent with the strong affirmation of the MOS-based probability.

Contrast the above case with that for ccf forecast category = 0.2 whose experimental distribution functions are plotted in Figure 71. In this case, because of the substantially larger percentage of occurrences of clear conditions (ccf < 0.15) in Data Set 2 (66%) than Data Set 1 (60%), the distribution intercepts with the 0.3 threshold line are farther apart (0.752 and 0.794 are the actual values). Thus, for the threshold = 0.3 case, the MOS model was strongly rejected by the evaluation procedure, while for the two higher threshold, no inconsistency was noted.

Similarly, for ccf forecast category = 0.8, shown in Figure 72, the distributions are rather far apart at the 0.7 threshold (and the MOS model is strongly rejected) while the models are supported for the two lower threshold values. In this case, the discrepancies arise because Data Set 2 contains a significantly larger percentage of cloudy conditions (69%) than Data Set 1 (59%).

We can summarize the above as follows: the MOS probability forecasts were rejected when the differences between the cloud-cover fraction distributions of the original and verifying data sets occurred at ccf values near the threshold value of interest. (I.e., differences in the fraction of clear cases caused the threshold = 0.3 model to be rejected, and differences in the fraction of very cloudy cases cause the threshold = 0.7 model to be rejected.) Basing the MOS model on the complete data set (instead of on the earliest 50% as in this evaluation) should improve the model robustness with respect to seasonal variations. Whether or not that improvement will be sufficient must be evaluated using the multi-year data sets.

OVERALL CONCLUSIONS

1. For most events (39 out of the 54 examined), the MOS forecast values are strongly affirmed by the second data set. This indicates, at least, that within the one season for which we have data, the MOS-models for these two cases are reasonably good.

2. For 5 out of the 54 cases, the MOS forecast values were strongly rejected. However, comparing the line plots in Figures 64 through 69 shows that, even for those five cases, the differences between the frequency of occurrence and the MOS forecast are not very large. Some variation within the season is indicated, but the MOS models are still not drastically in error. (For example: for case #12, box 30, with MOS forecast ccf = 0.2, the forecast probability that the actual ccf would be less than 0.3 was 0.75, but in fact, in Data Set 2, 79% of the cases had actual ccf less than 0.3.)

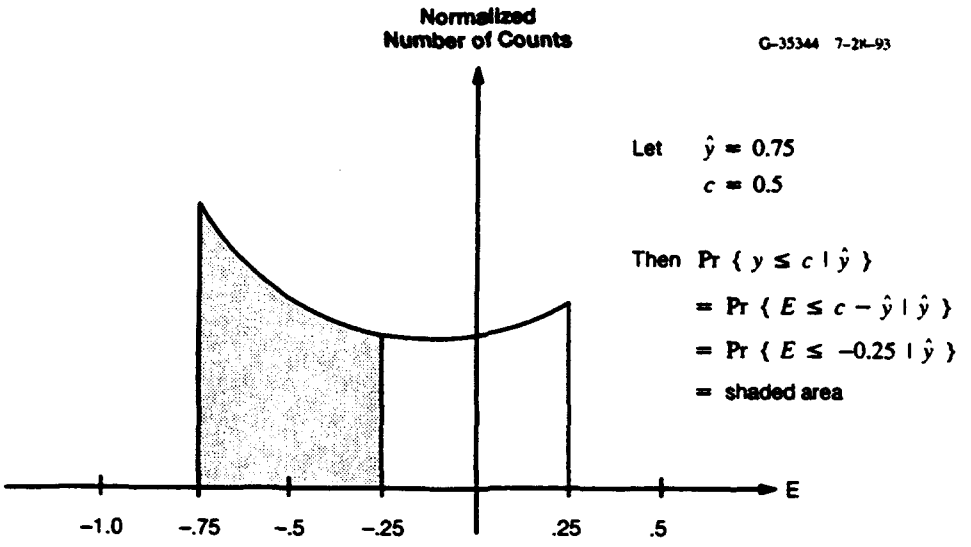


Figure 26 Example of an Error Histogram Showing How to Compute Conditional Cloud Cover Probability

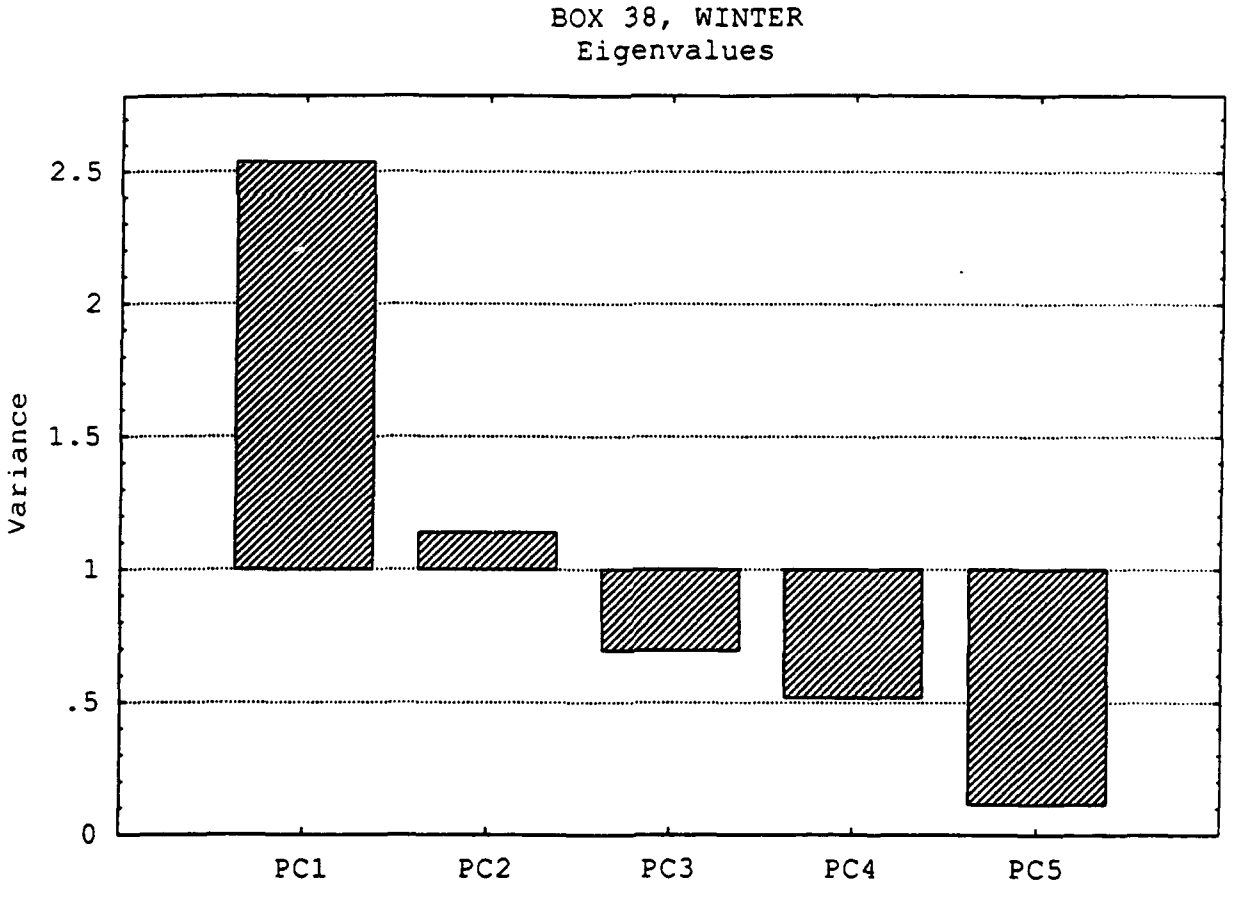


Figure 27 Variance of Principal Components (PC) Computed For Box 38, Winter

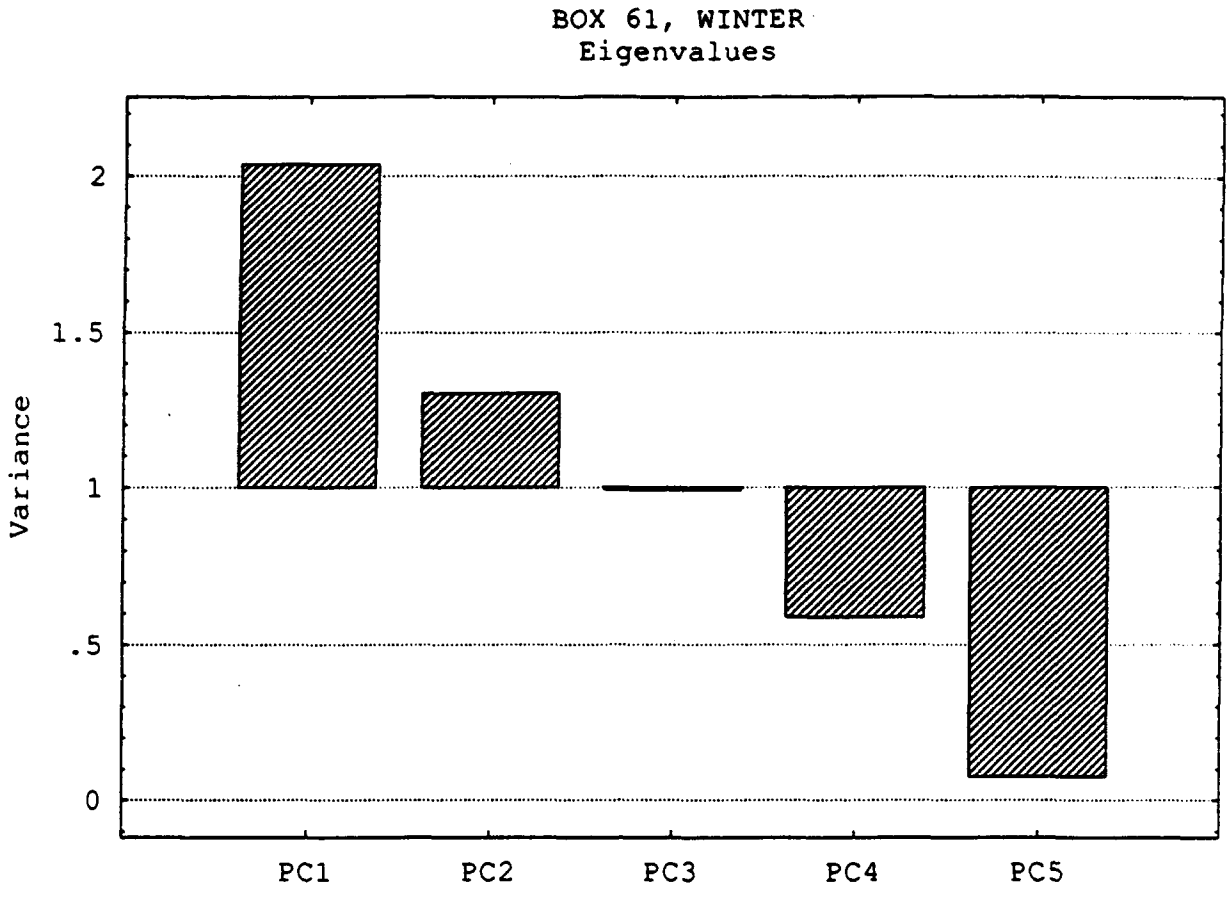
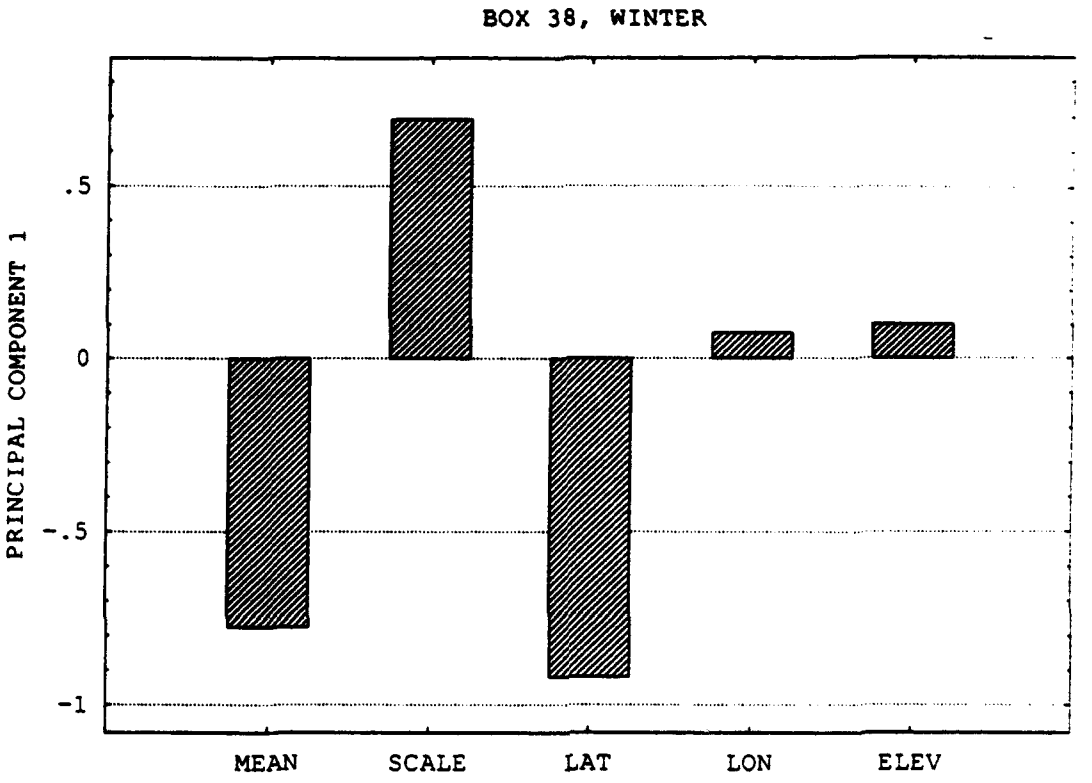
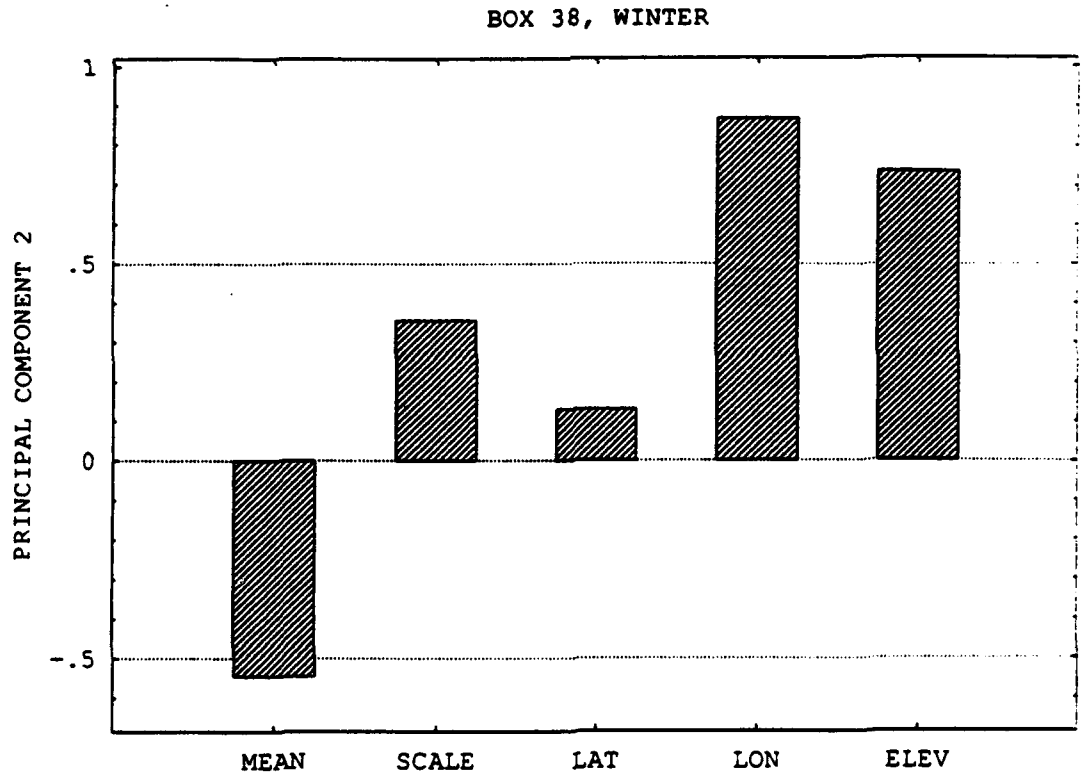


Figure 28 Variance of Principal Components Computed For Box 61, Winter



(a)



(b)

Figure 29 **Breakdown of the a) First and b) Second Principal Component Into its Component Parts, Box 38, Winter**

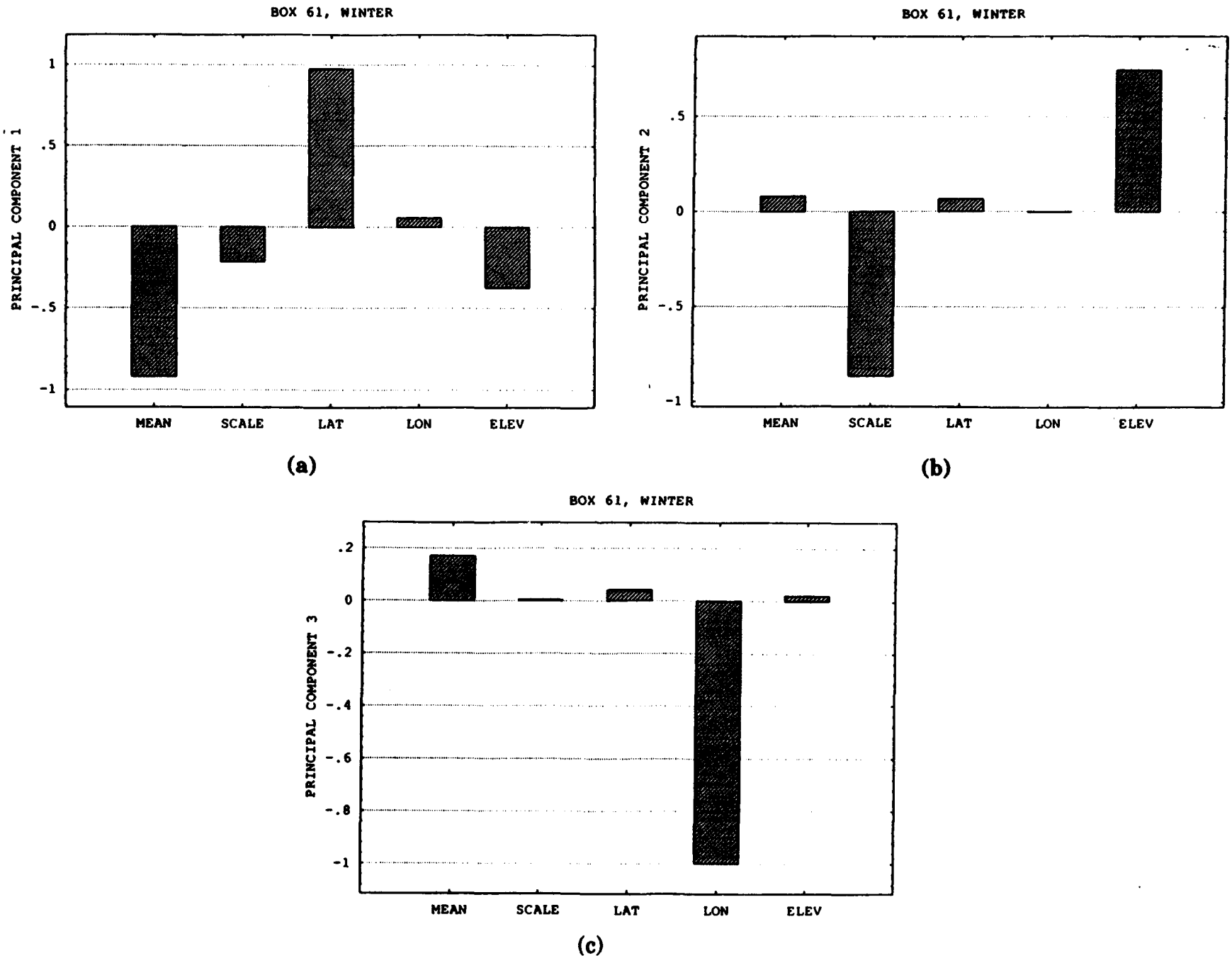


Figure 30 Breakdown of the a) First, b) Second, and c) Third Principal Component Into its Component Parts, Box 61, Winter

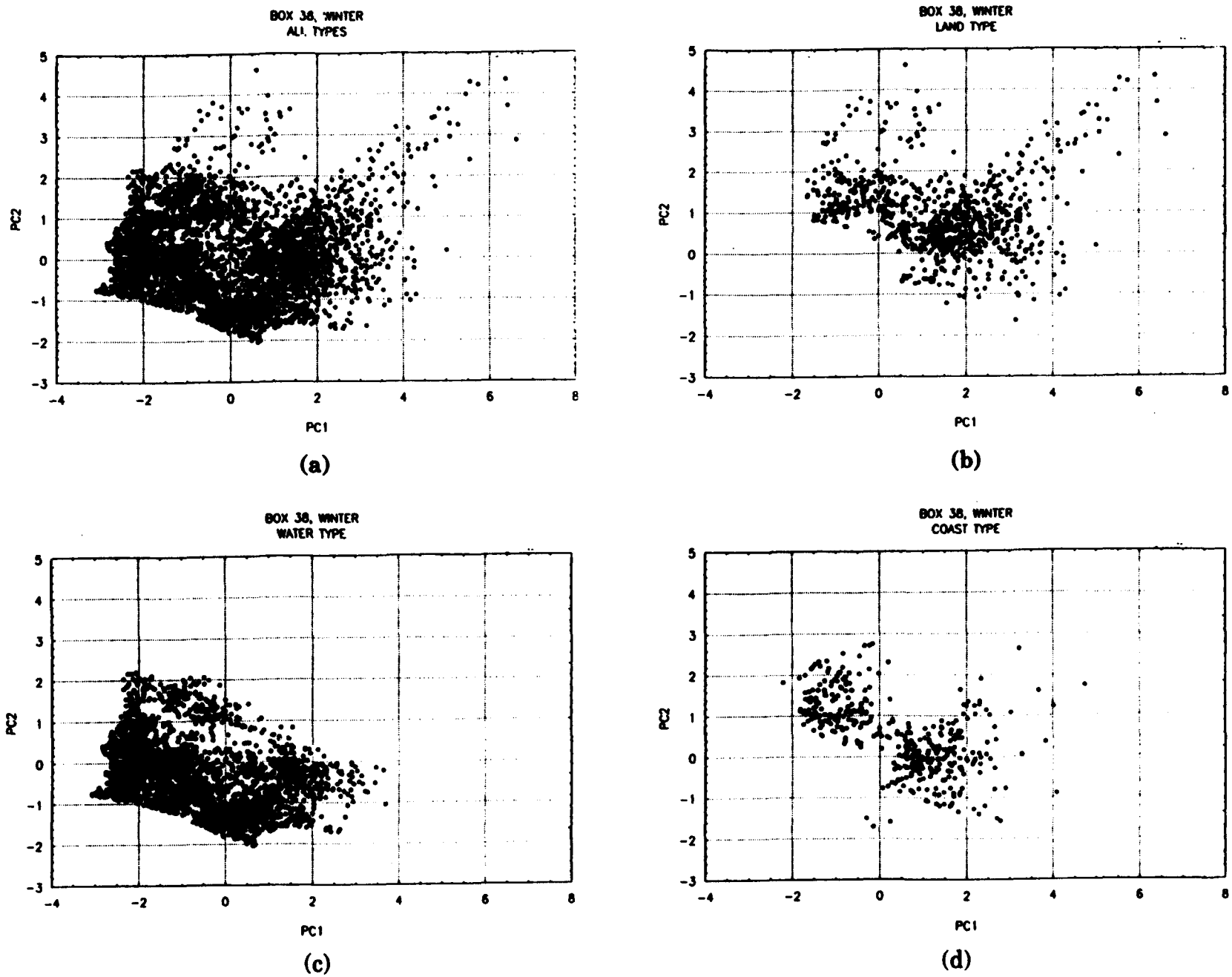


Figure 31 Scatterplots in Principal Component Space. Each point in the Plots Corresponds to One Eighth-Mesh Grid Point in Box 38. Figure a) Contains All Points, b) Contains Only Land Grid Fonts, c) Has Only Water Grid Points, and d) Only Those Grid Points On the Coast

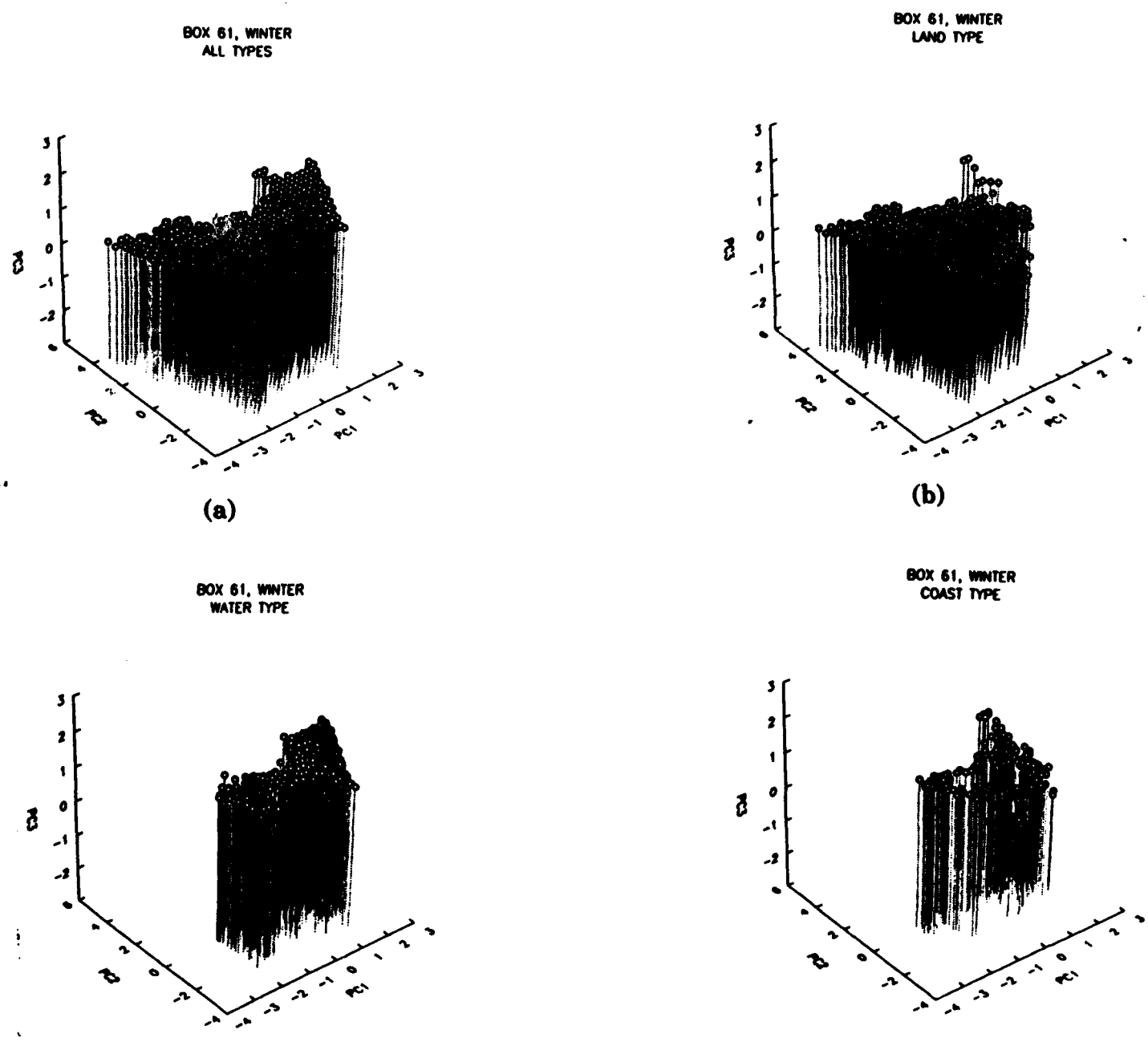


Figure 32 Scatterplots in Principal Component Space. Each point in the Plots Corresponds to One Eighth-Mesh Grid Point in Box 61. Figure a) Contains All Points, b) Contains Only Land Grid Fonts, c) Has Only Water Grid Points, and d) Only Those Grid Points On the Coast

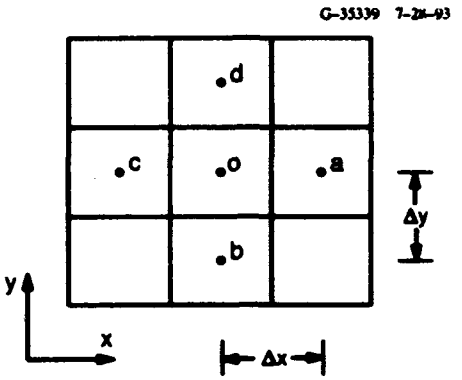


Figure 33 Schematic Showing the Grid Geometry Used to Compute Derived Variables Such as Divergence and Vorticity. The (x,y) Coordinate System Used in this Figure Corresponds to the (i,j) Coordinate System Used in the Nephanalysis Fields

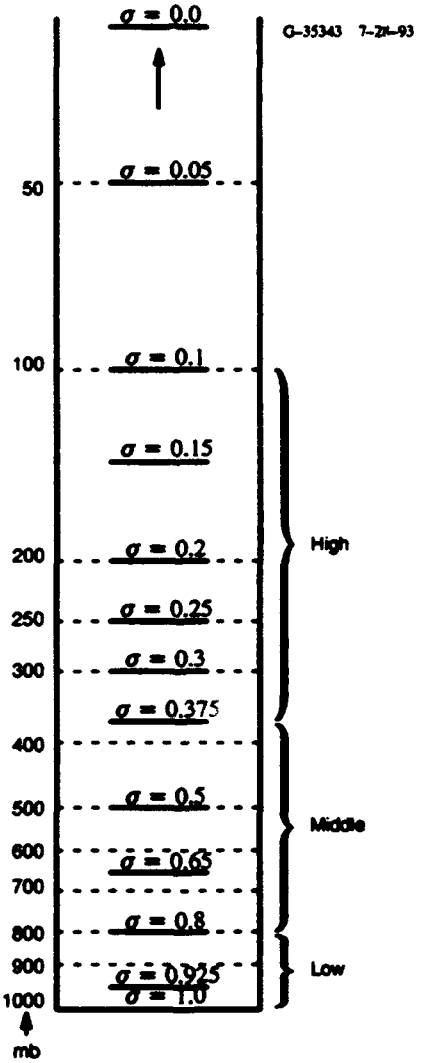


Figure 34 Definition of Low, Middle, and High Layers in Terms of Sigma Surfaces Used in Model Development (Ref. 24)

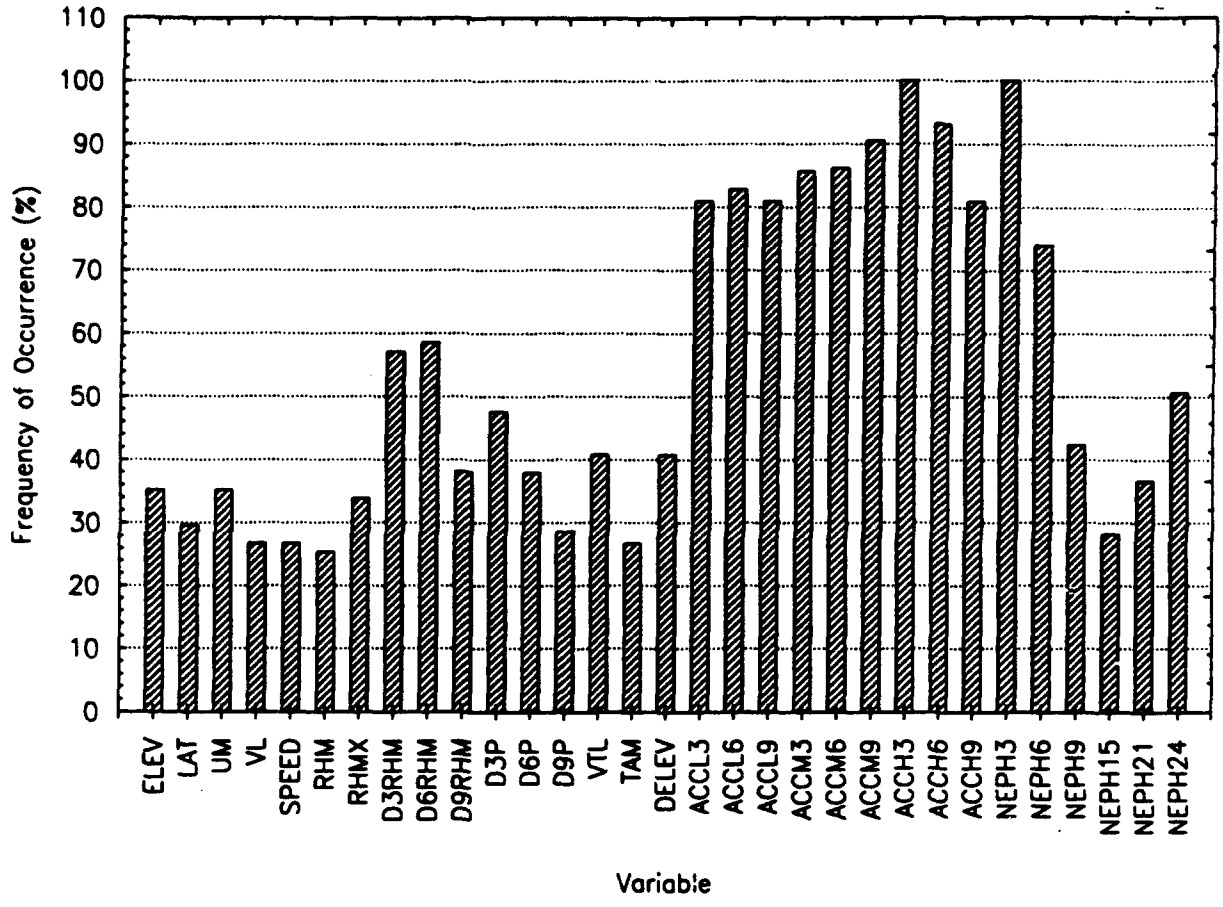


Figure 35 Histogram Showing All Those Predictors That Were Selected in the Top 15 More Than 25% of the Time Over all 71 Models

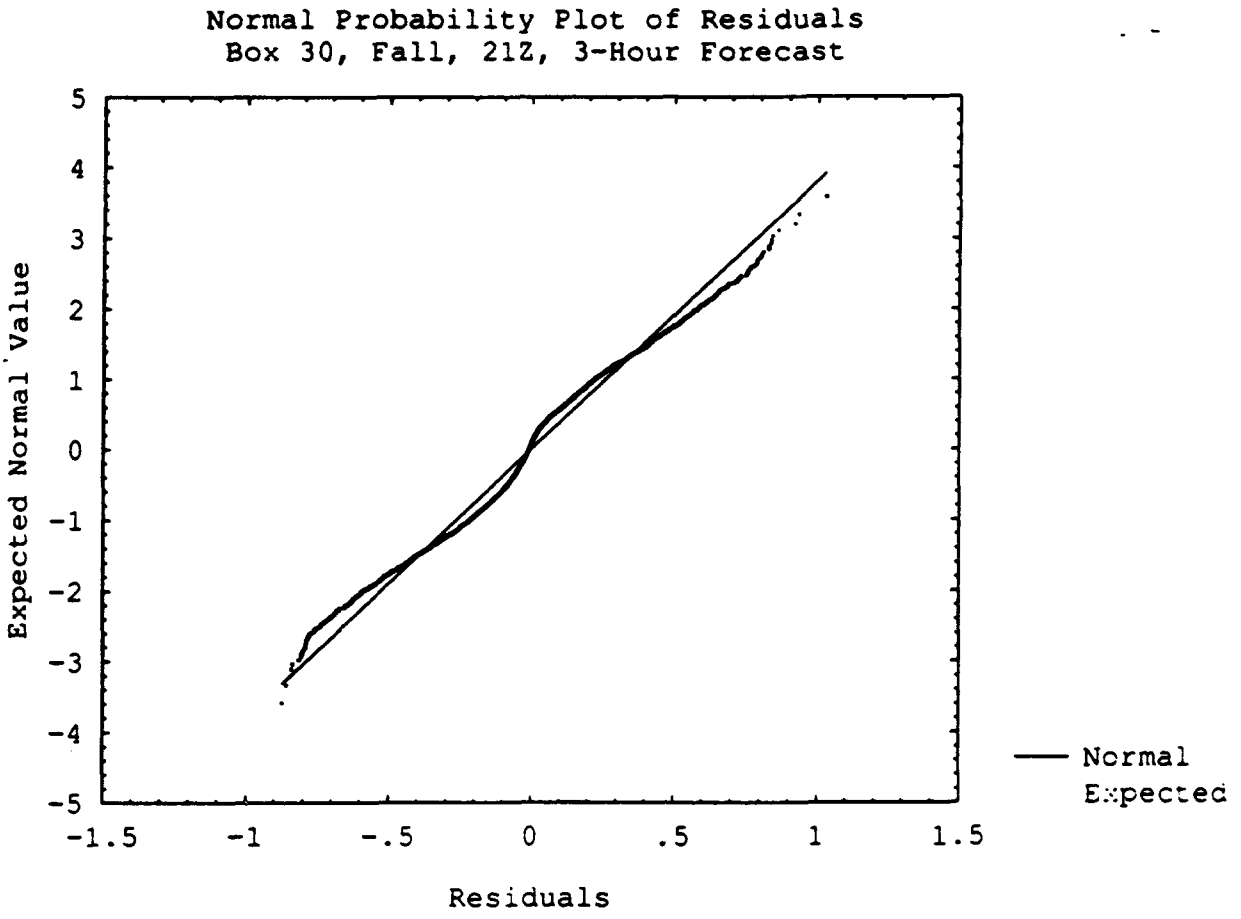


Figure 36 Normal Plot of the Residuals for Case #24

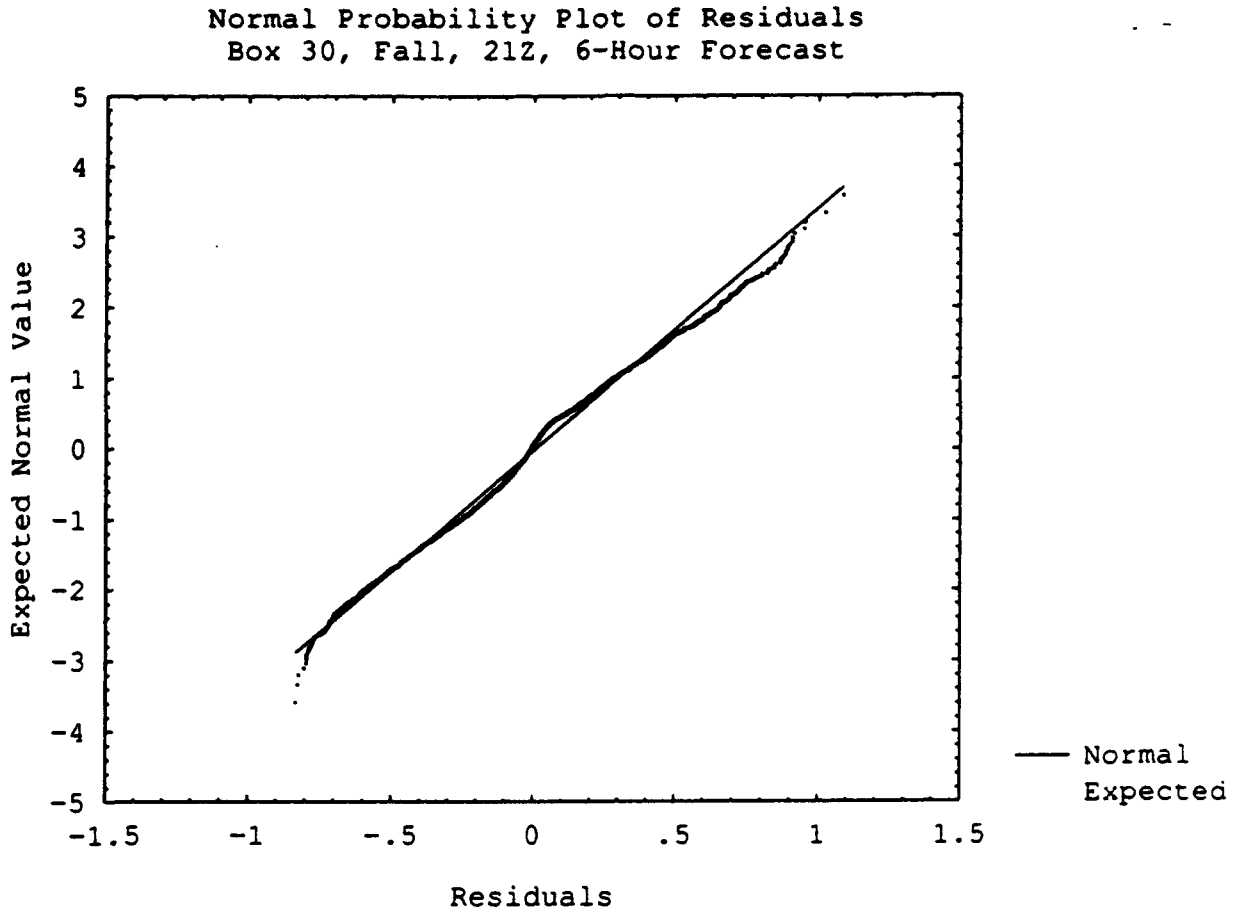


Figure 37 **Normal Plot of the Residuals for Case #25**

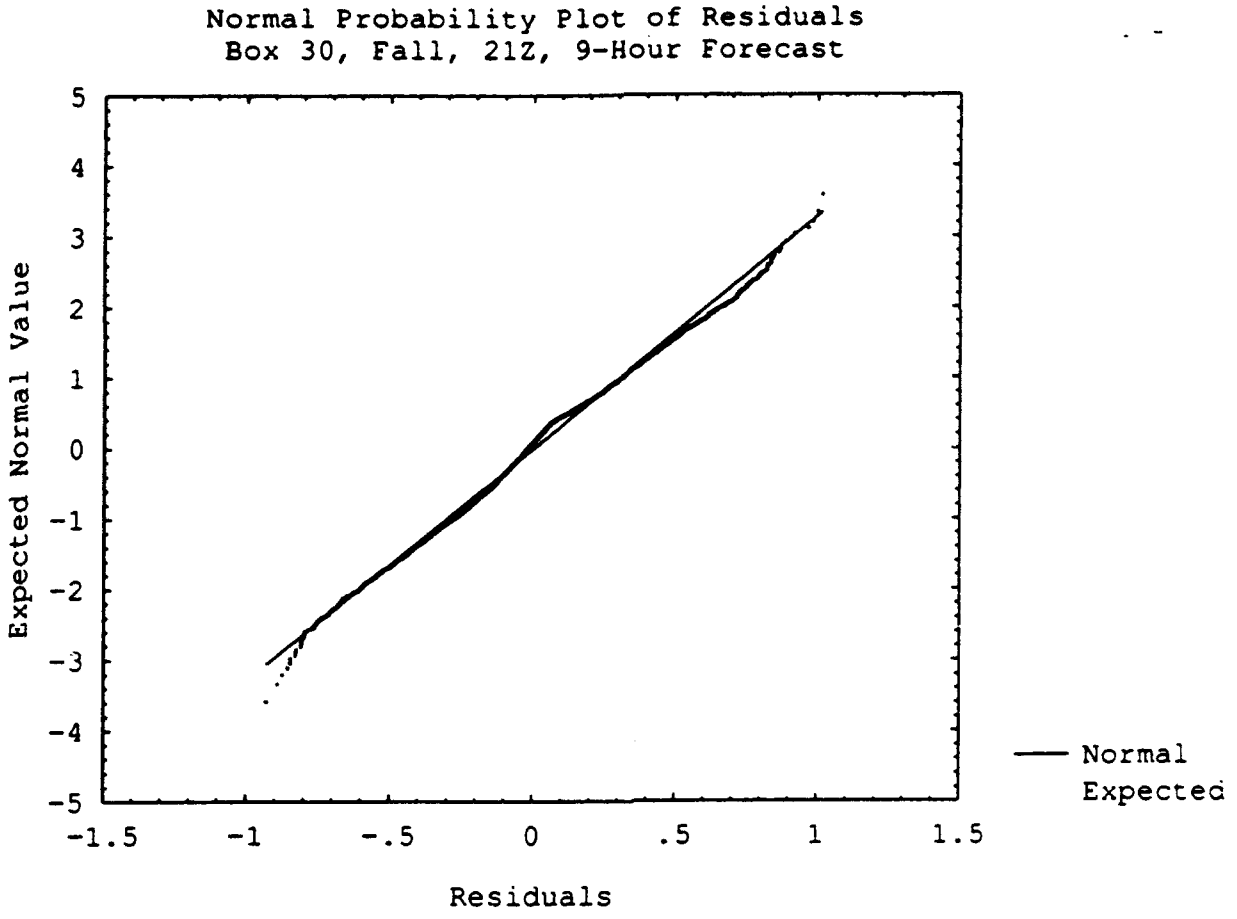


Figure 38 **Normal Plot of the Residuals for Case #27**

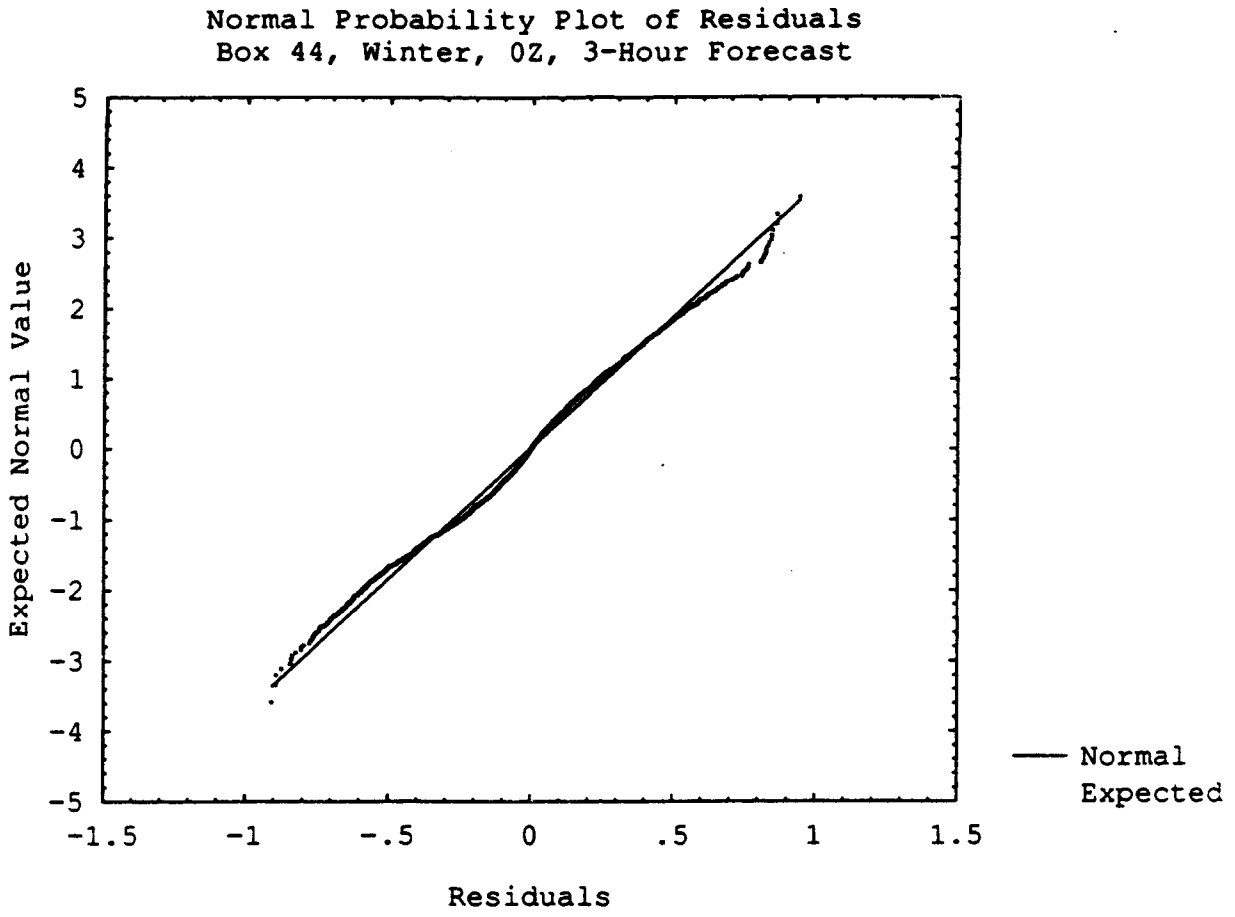


Figure 39 Normal Plot of the Residuals for Case #42

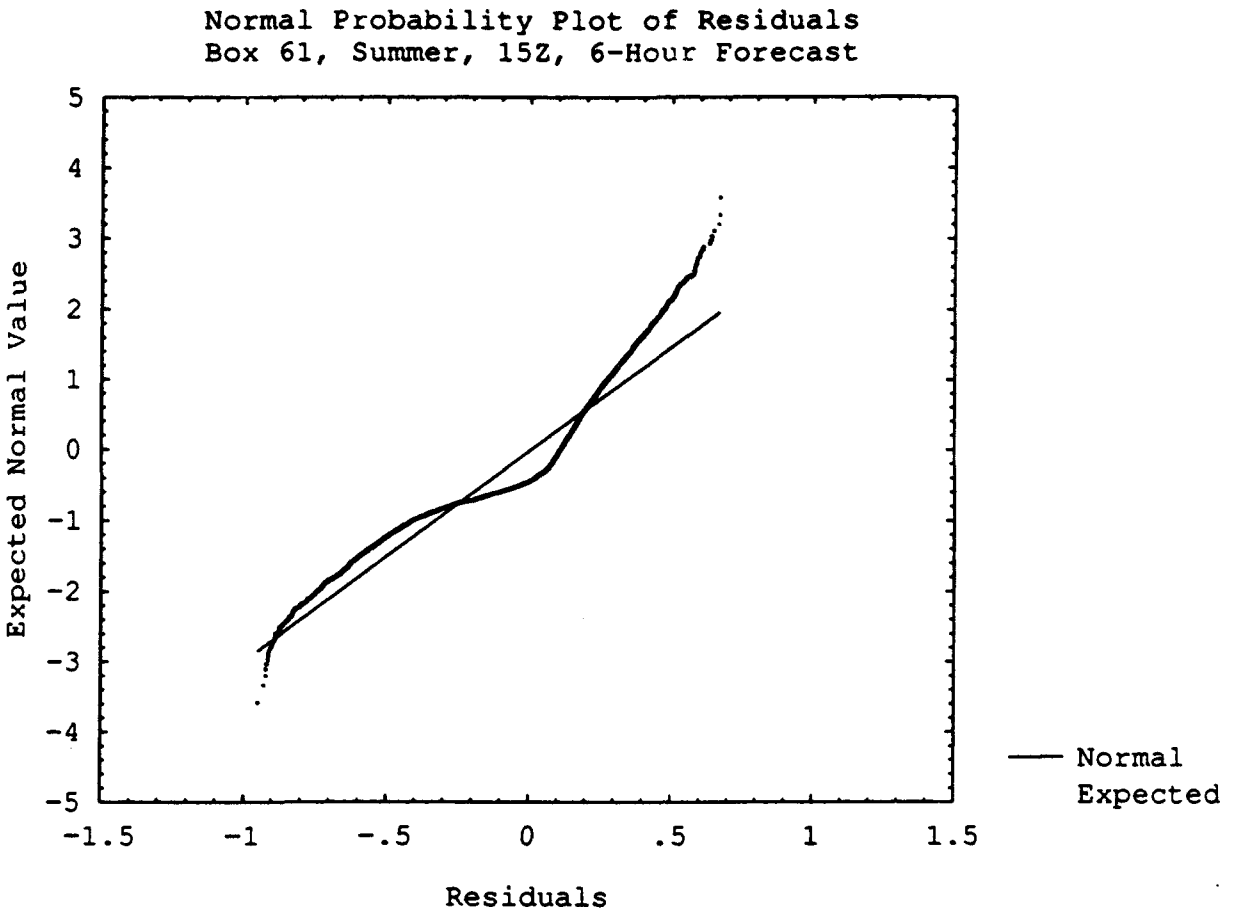


Figure 40 Normal Plot of the Residuals for Case #52 Showing a Poor Fit to the "Normal Line"

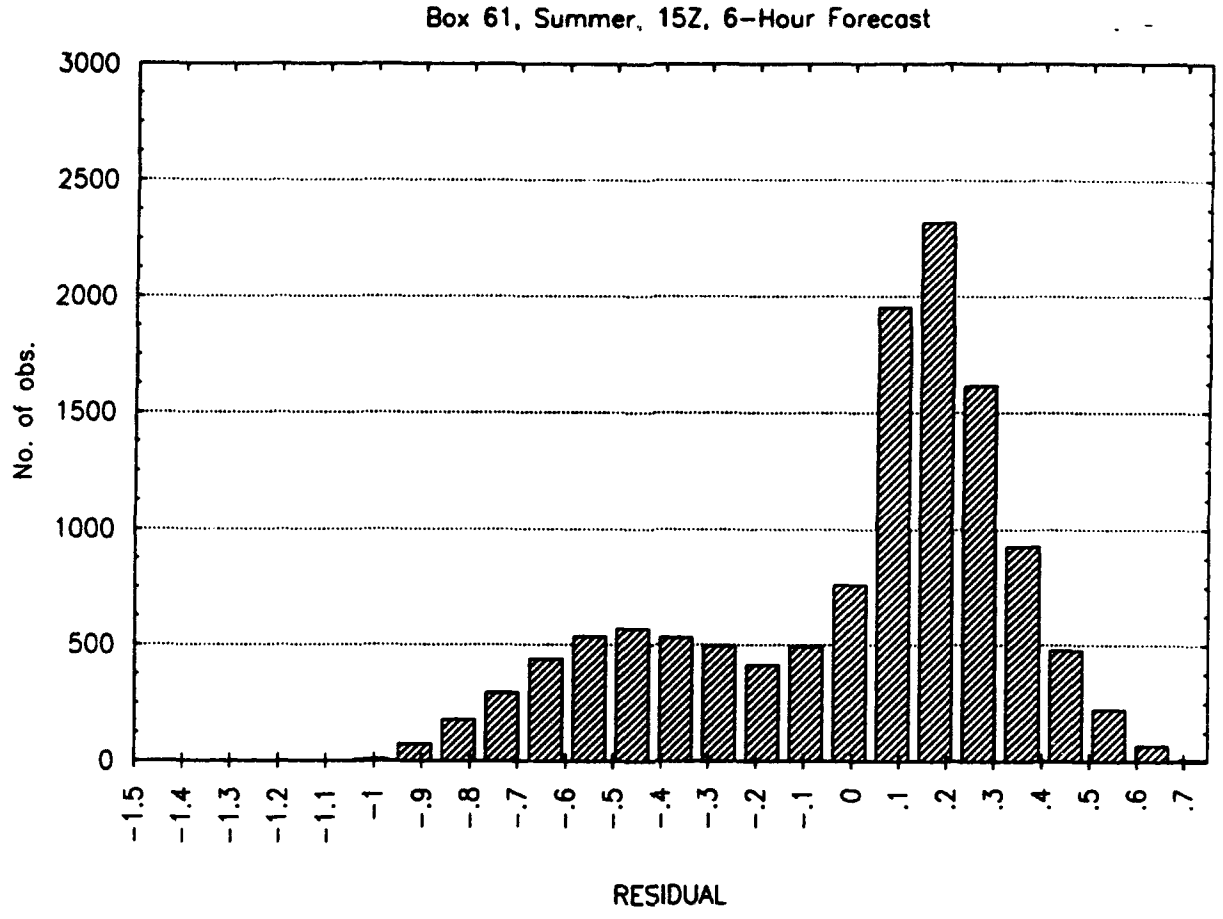


Figure 41 Histogram of Model Residuals for Case #52 Showing Distinct Non-Normalities

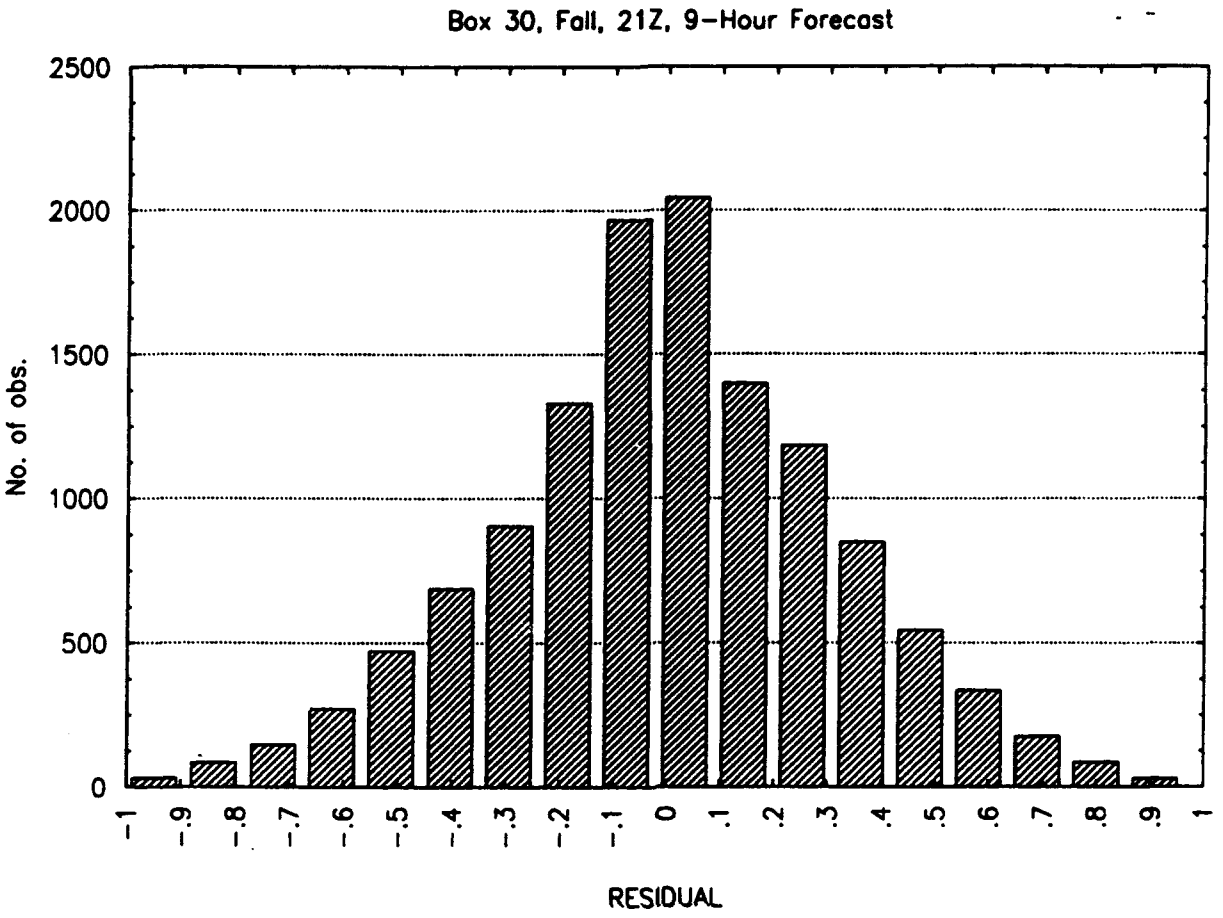


Figure 42 Histogram of Model Residuals for Case #27

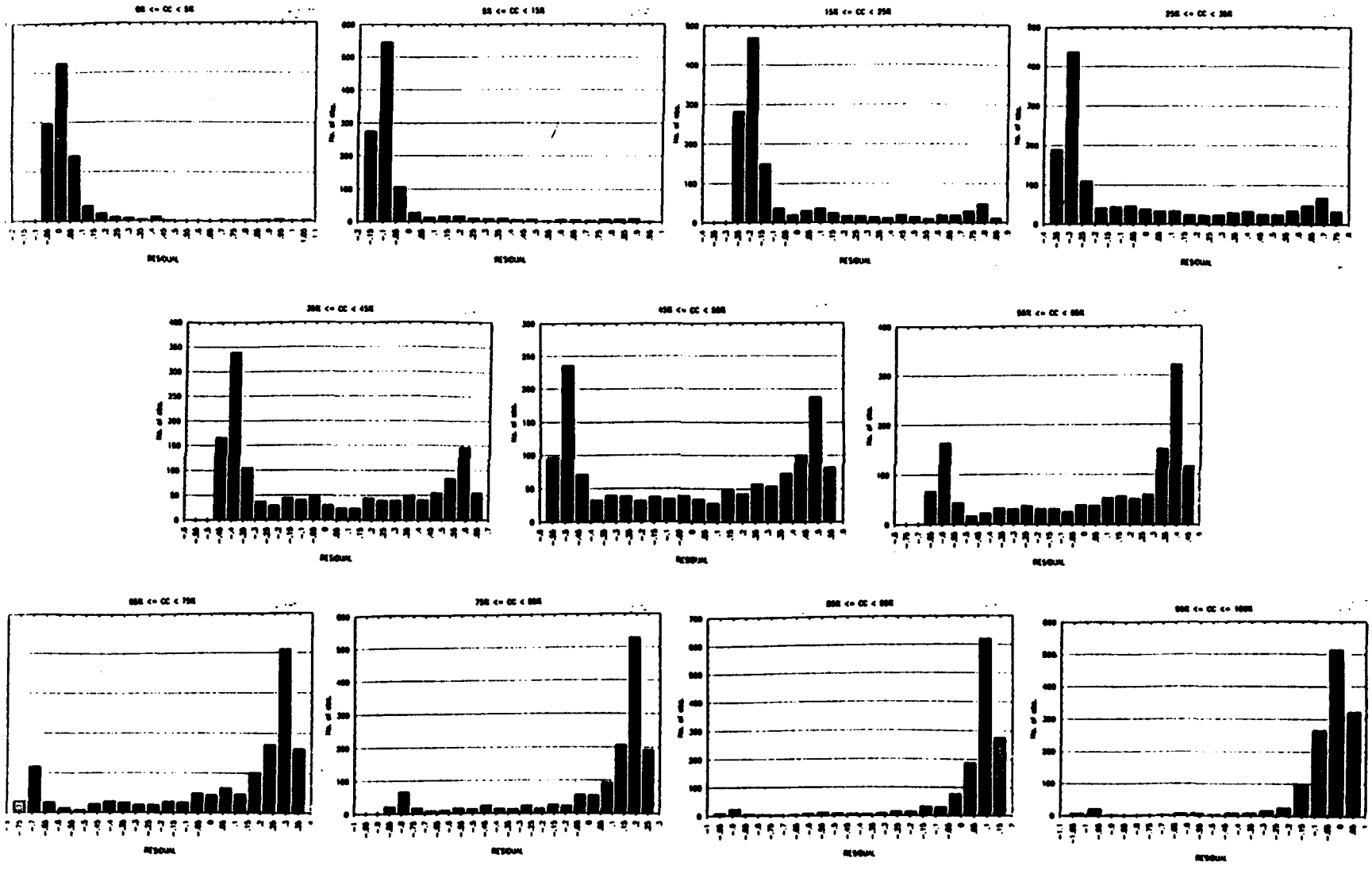


Figure 43 Sequence of Histograms Showing the Residuals for Box 30, Spring, 3Z, 6-hour Forecast Categorized by Predicted Cloud Cover (cc)

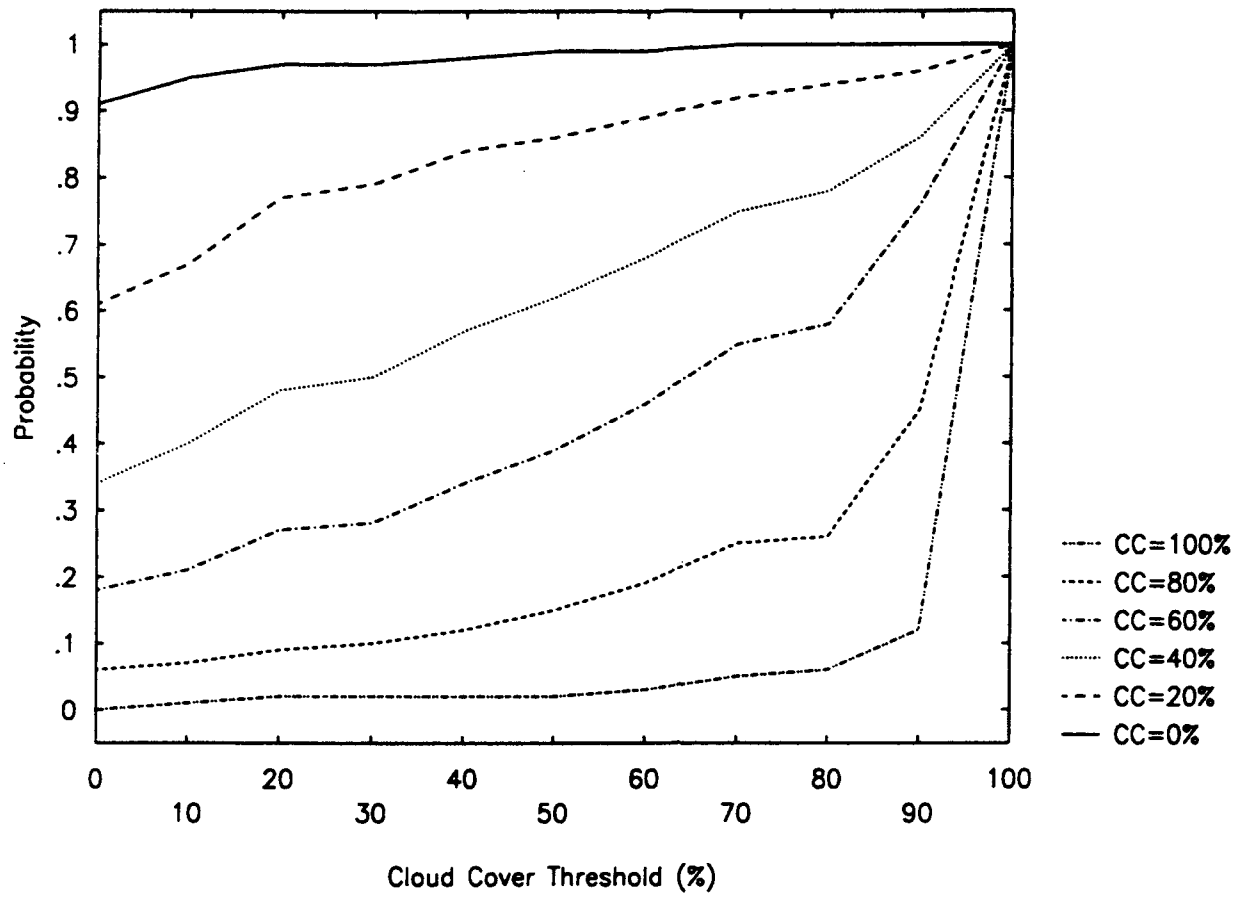


Figure 44 MOS-based Probability Forecast That the Observed Cloud Cover Will be Less Than or Equal to the Specified Cloud Cover Threshold Given the MOS-predicted Cloud Amount (cc) in Case #24

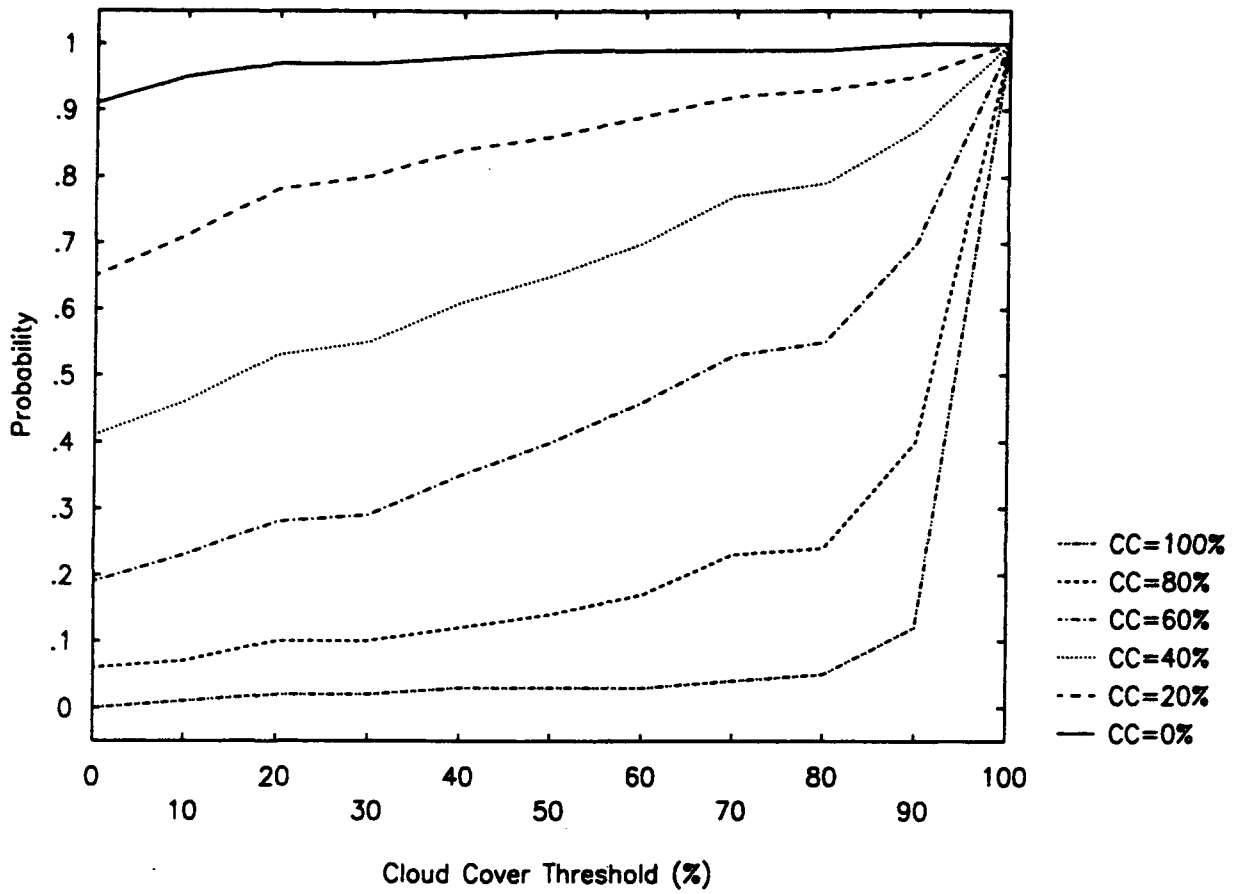


Figure 45 MOS-based Probability Forecast That the Observed Cloud Cover Will be Less Than or Equal to the Specified Cloud Cover Threshold Given the MOS-predicted Cloud Amount (cc) in Case #25

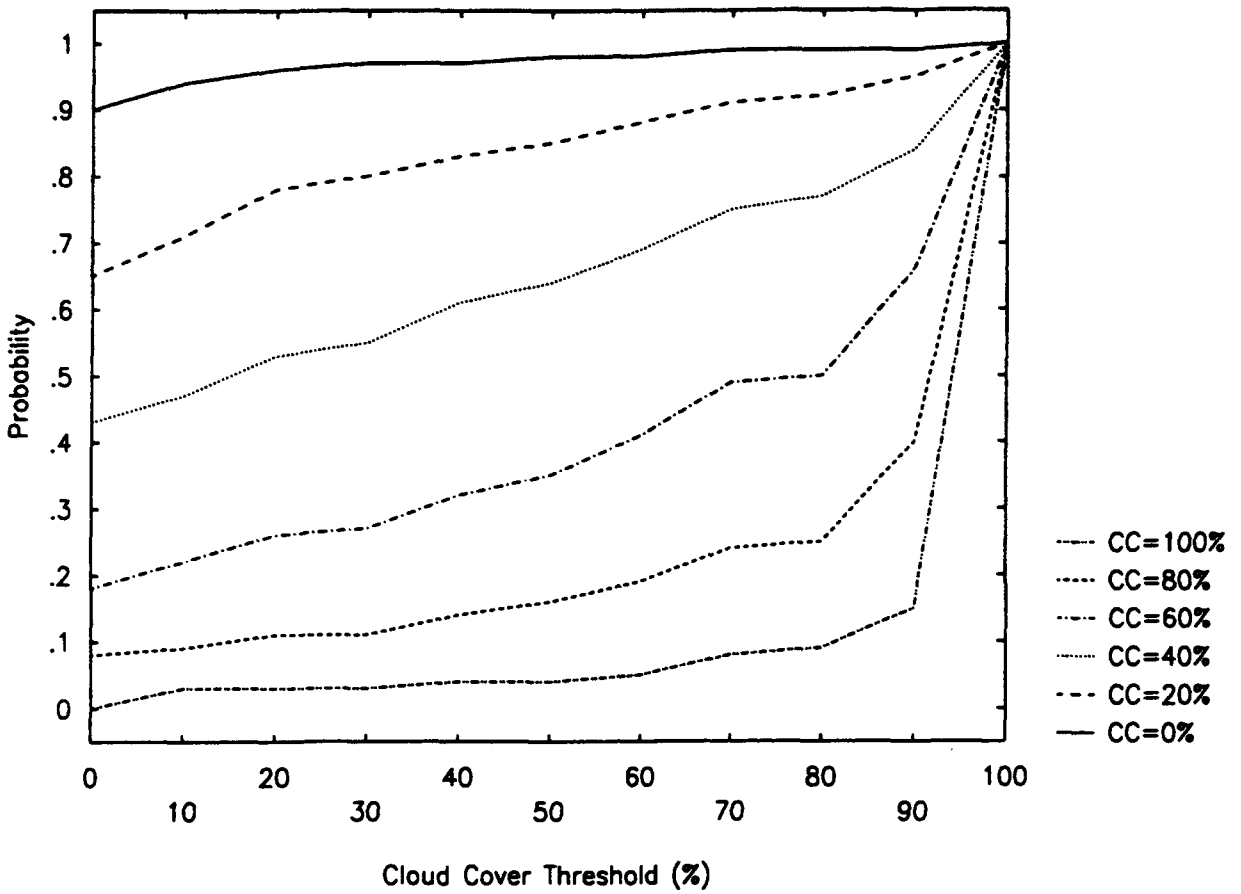


Figure 46 MOS-based Probability Forecast That the Observed Cloud Cover Will be Less Than or Equal to the Specified Cloud Cover Threshold Given the MOS-predicted Cloud Amount (cc) in Case #27

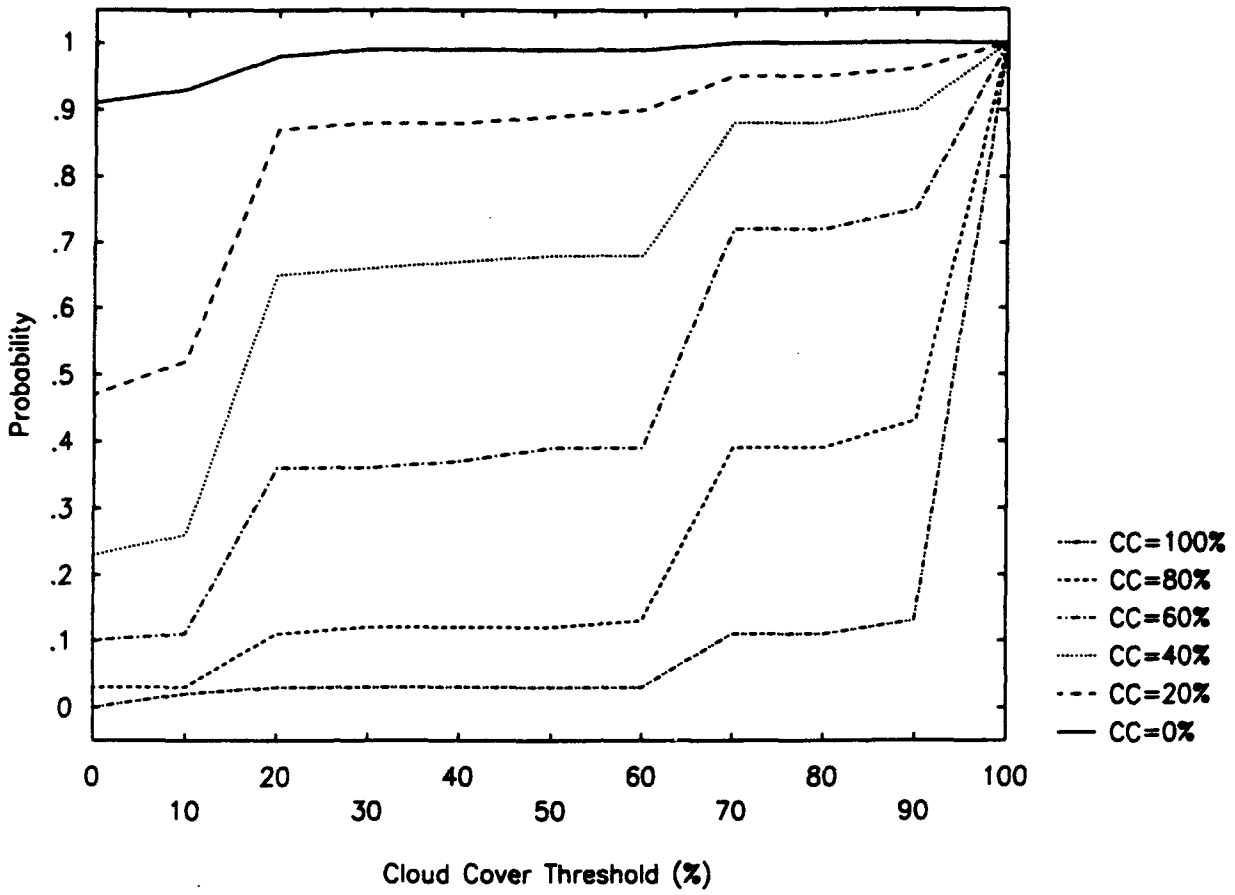


Figure 47 MOS-based Probability Forecast That the Observed Cloud Cover Will be Less Than or Equal to the Specified Cloud Cover Threshold Given the MOS-predicted Cloud Amount (cc) in Case #42

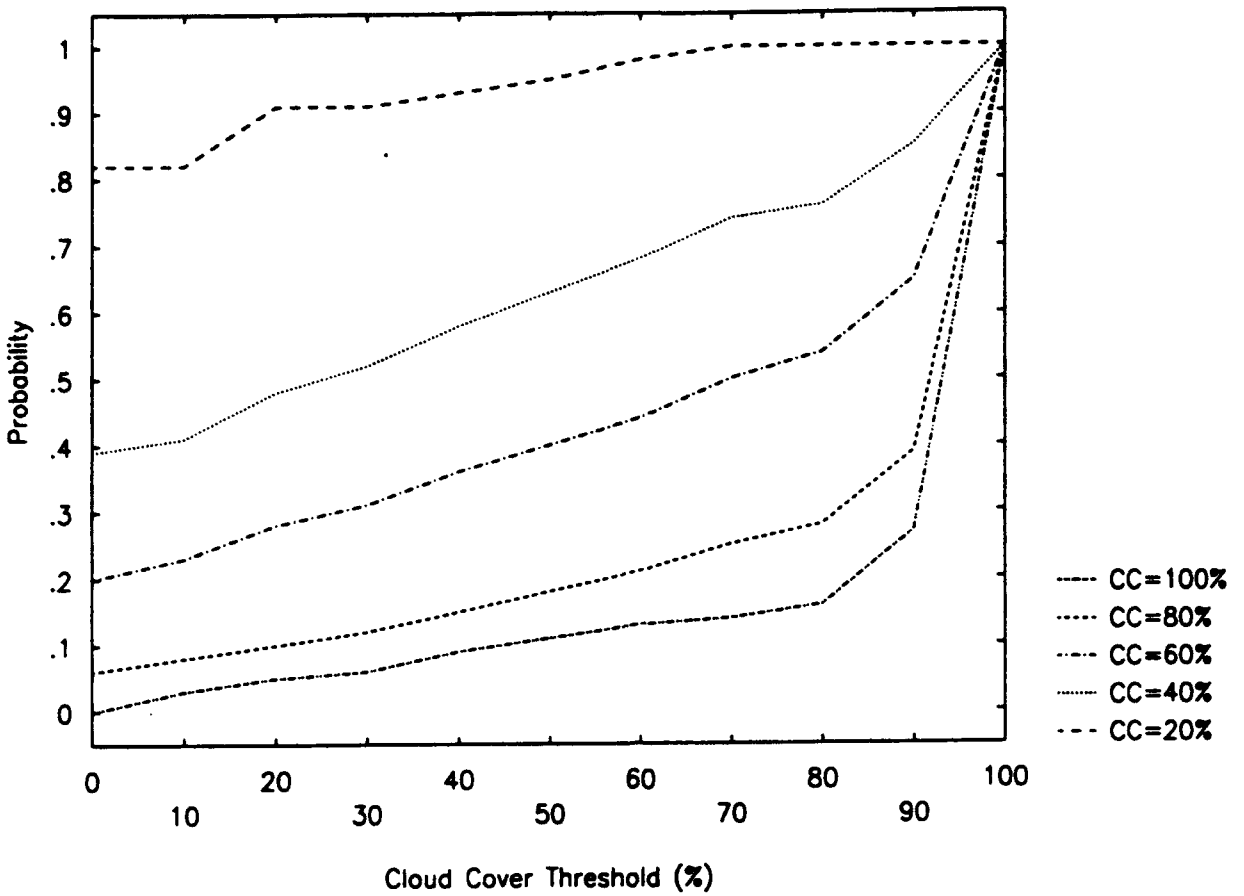


Figure 48 MOS-based Probability Forecast That the Observed Cloud Will be Less Than or Equal to the Specified Cloud Cover Threshold Given the MOS-predicted Cloud Amount (cc) in Case #52

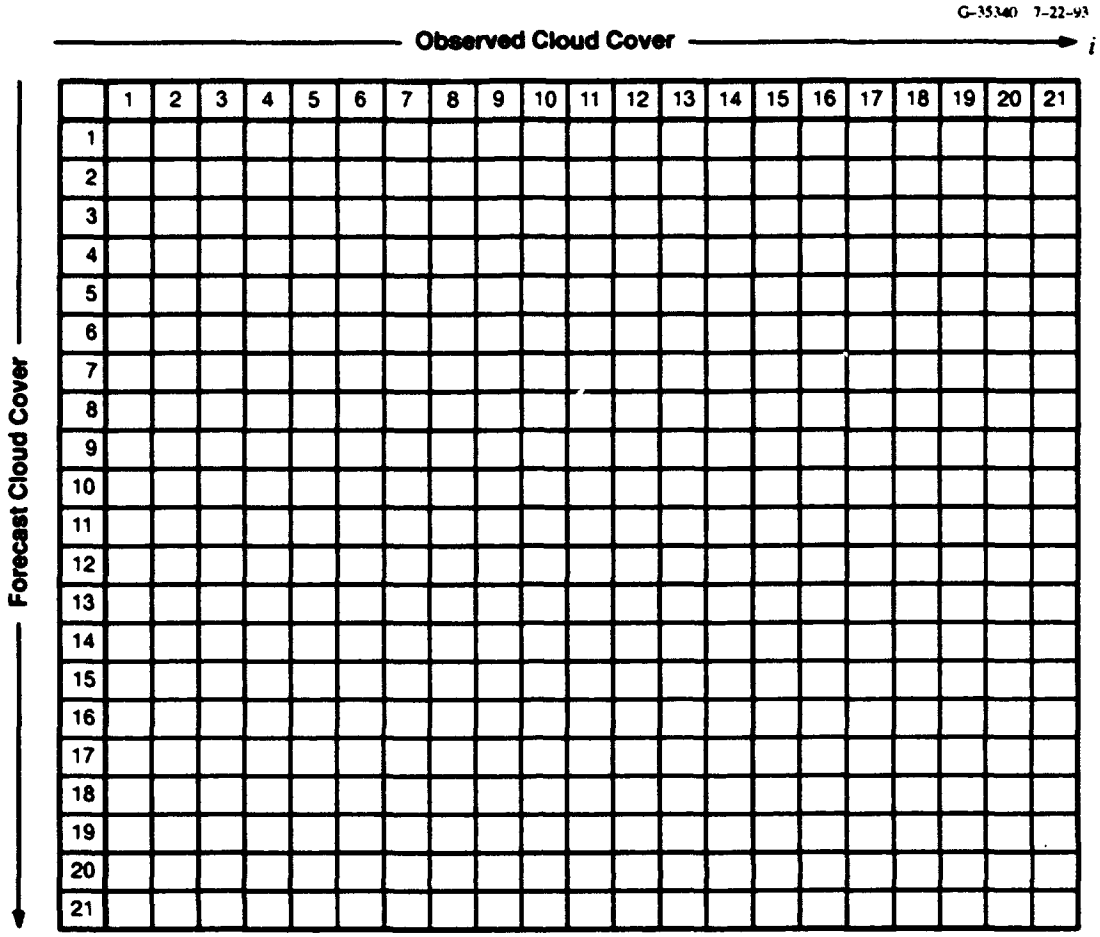


Figure 49 Schematic of a 21 × 21 Contingency Table Used to Compute Forecast Brier and 20/20 Scores, and Sharpness Measure

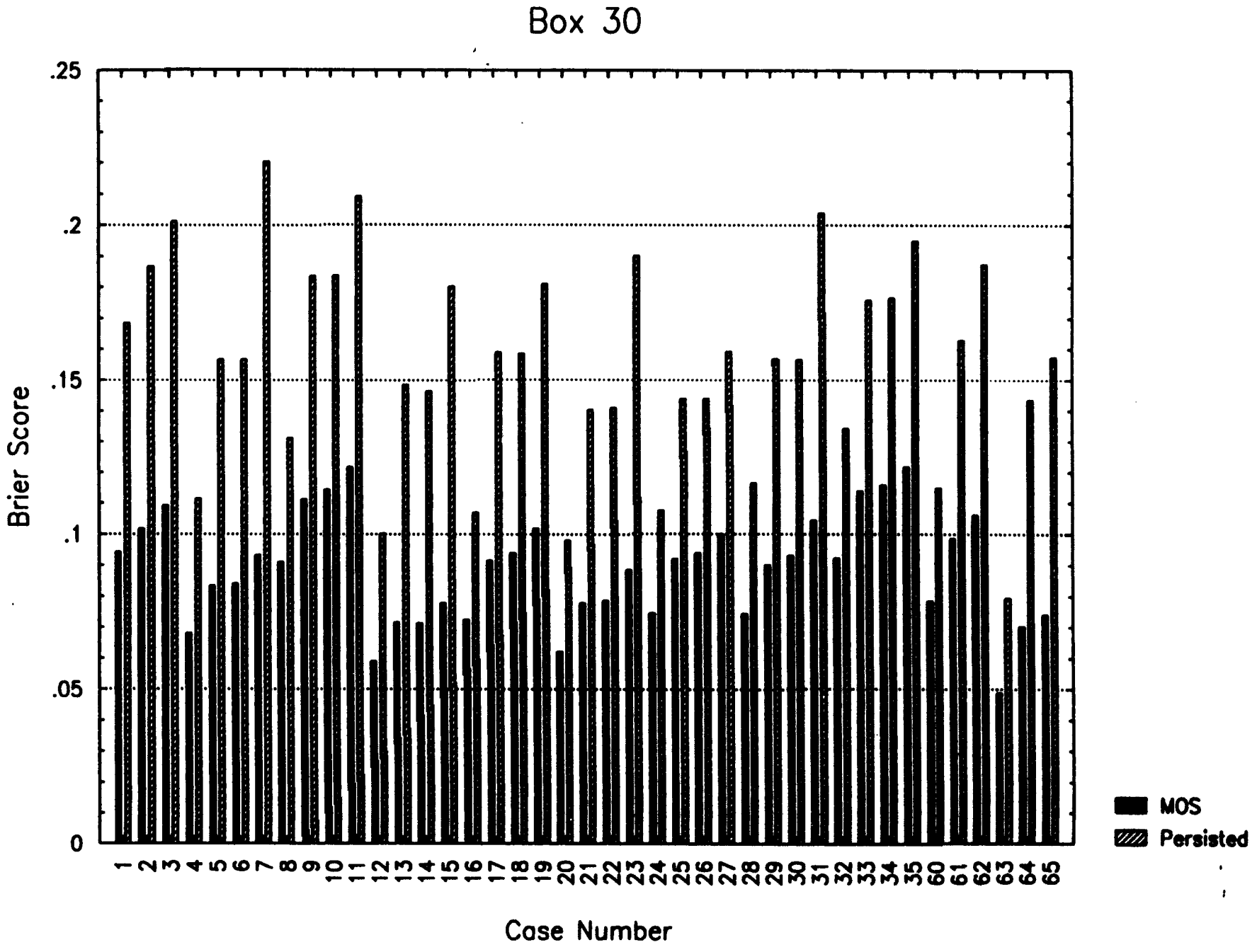


Figure 50 Comparison of Brier Scores for MOS-based and Persistence-based Forecasts for All Cases in Box 30

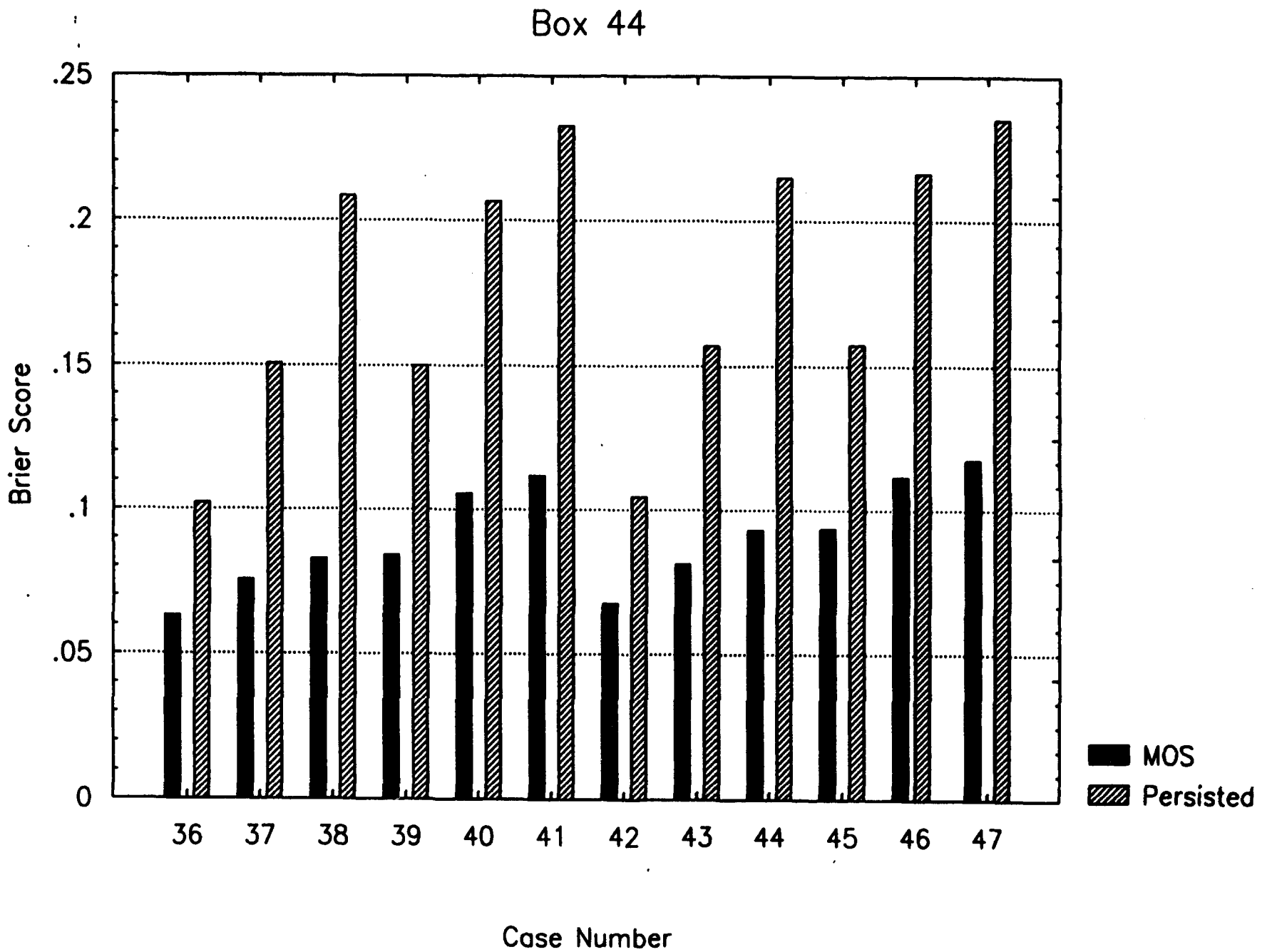


Figure 51 Comparison of Brier Scores for MOS-based and Persistence-based Forecasts for All Cases in Box 44

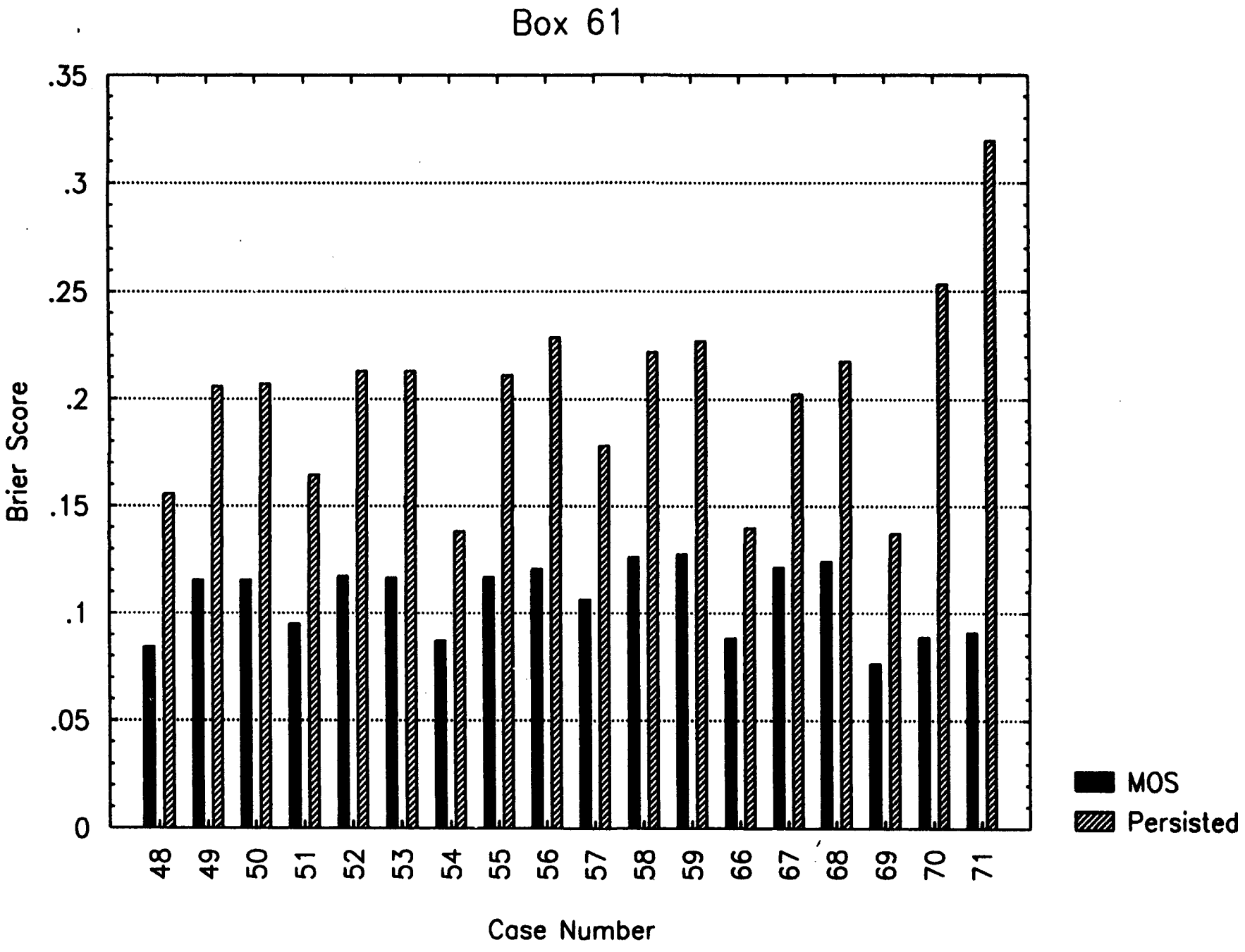


Figure 52 Comparison of Brier Scores for MOS-based and Persistence-based Forecasts for All Cases in Box 61

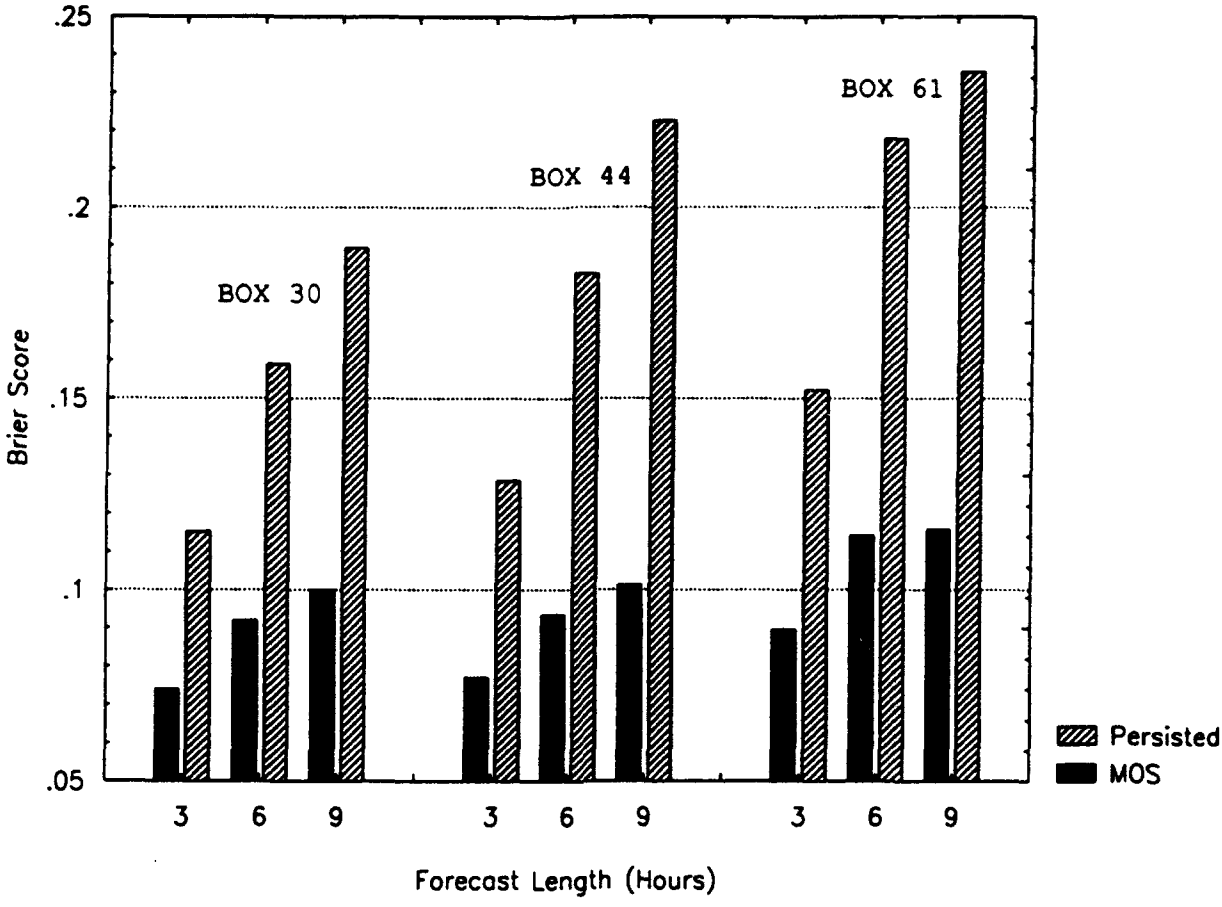


Figure 53 Brier Scores Averaged Over All Cases in a Given Box for 3-, 6-, and 9-Hour Forecasts

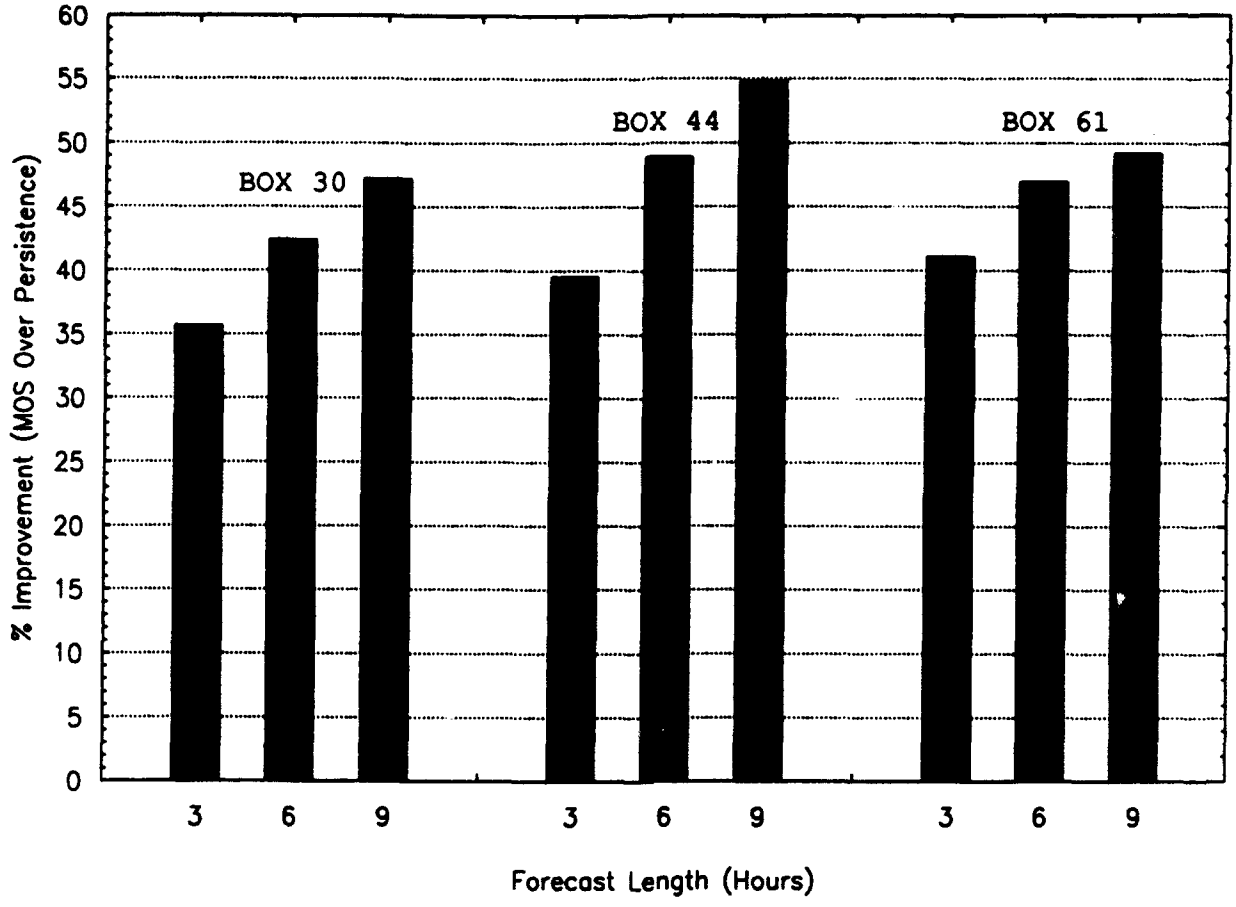


Figure 54 Averaged Percentage Improvement in Brier Scores of MOS Forecasts Over Persistence by Box and Forecast Length

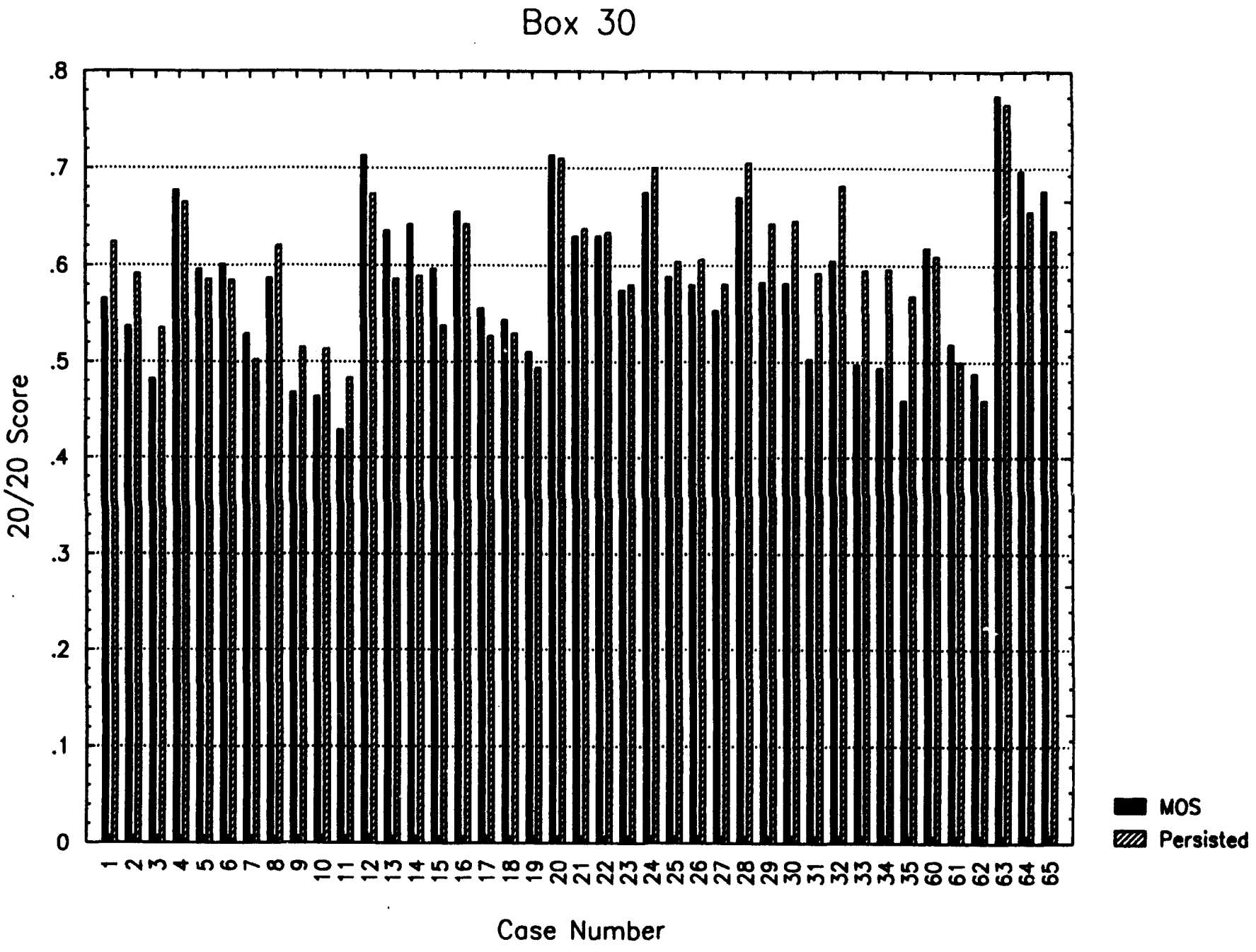


Figure 55 Comparison of 20/20 Scores for MOS-based and Persistence-based Forecasts for all Cases in Box 30

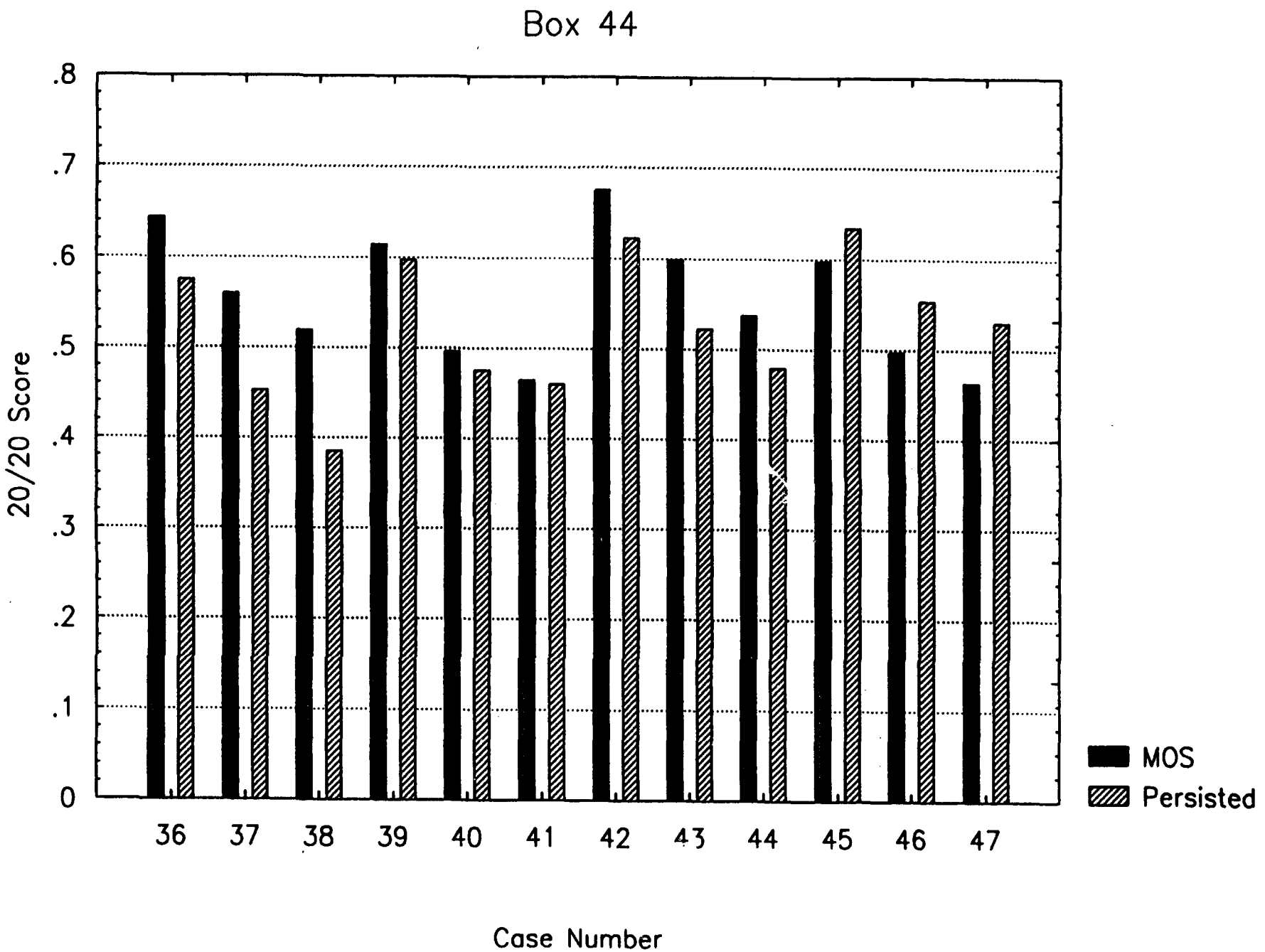


Figure 56 Comparison of 20/20 Scores for MOS-based and Persistence-based Forecasts for all Cases in Box 44

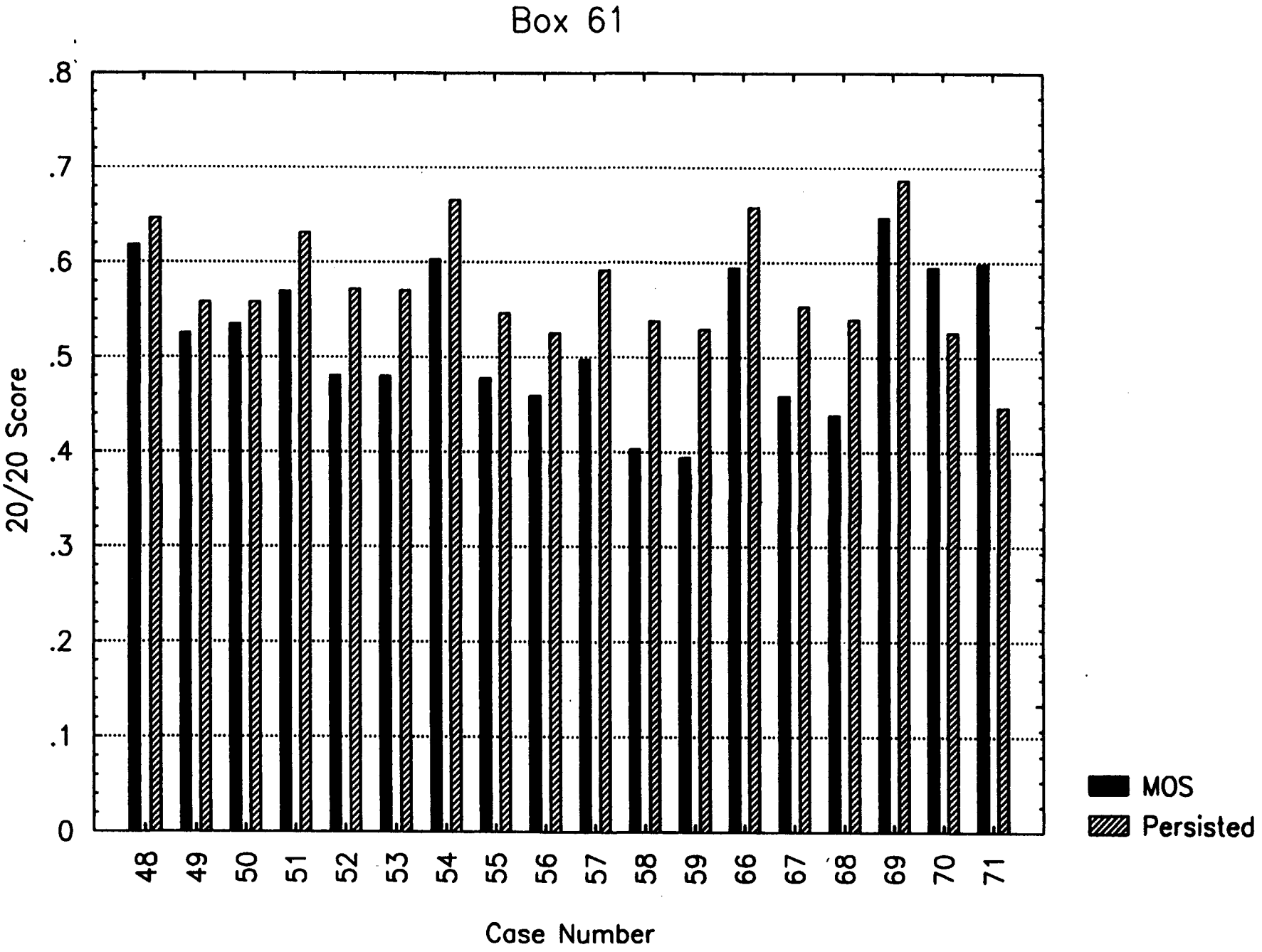


Figure 57 Comparison of 20/20 Scores for MOS-based and Persistence-based Forecasts for all Cases in Box 61

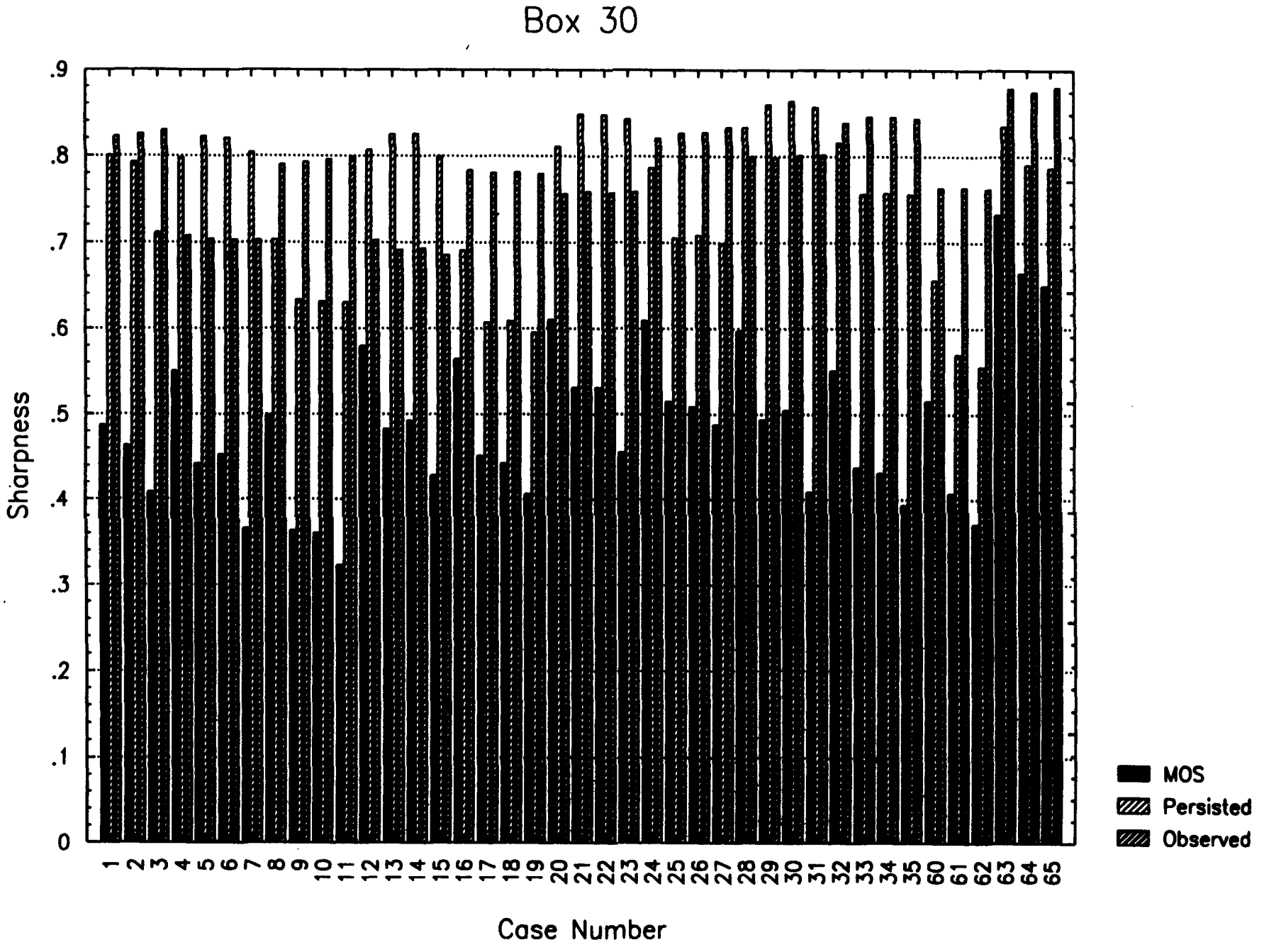


Figure 58 Comparison of Sharpness for MOS-based Forecasts, Persistence-based Forecasts and Observed Cloud Cover for all Cases in Box 30

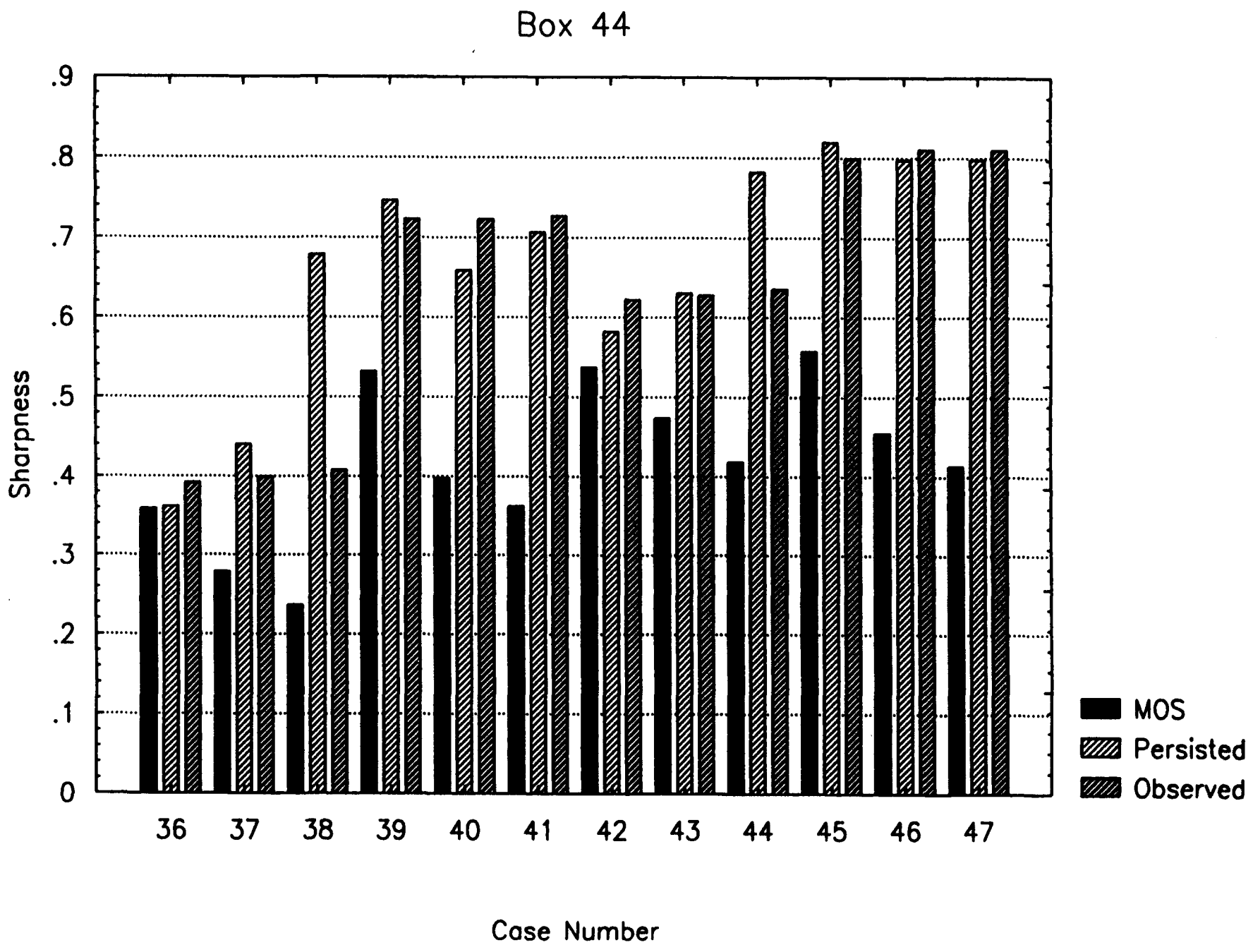


Figure 59 Comparison of Sharpness for MOS-based Forecasts, Persistence-based Forecasts and Observed Cloud Cover for all Cases in Box 44

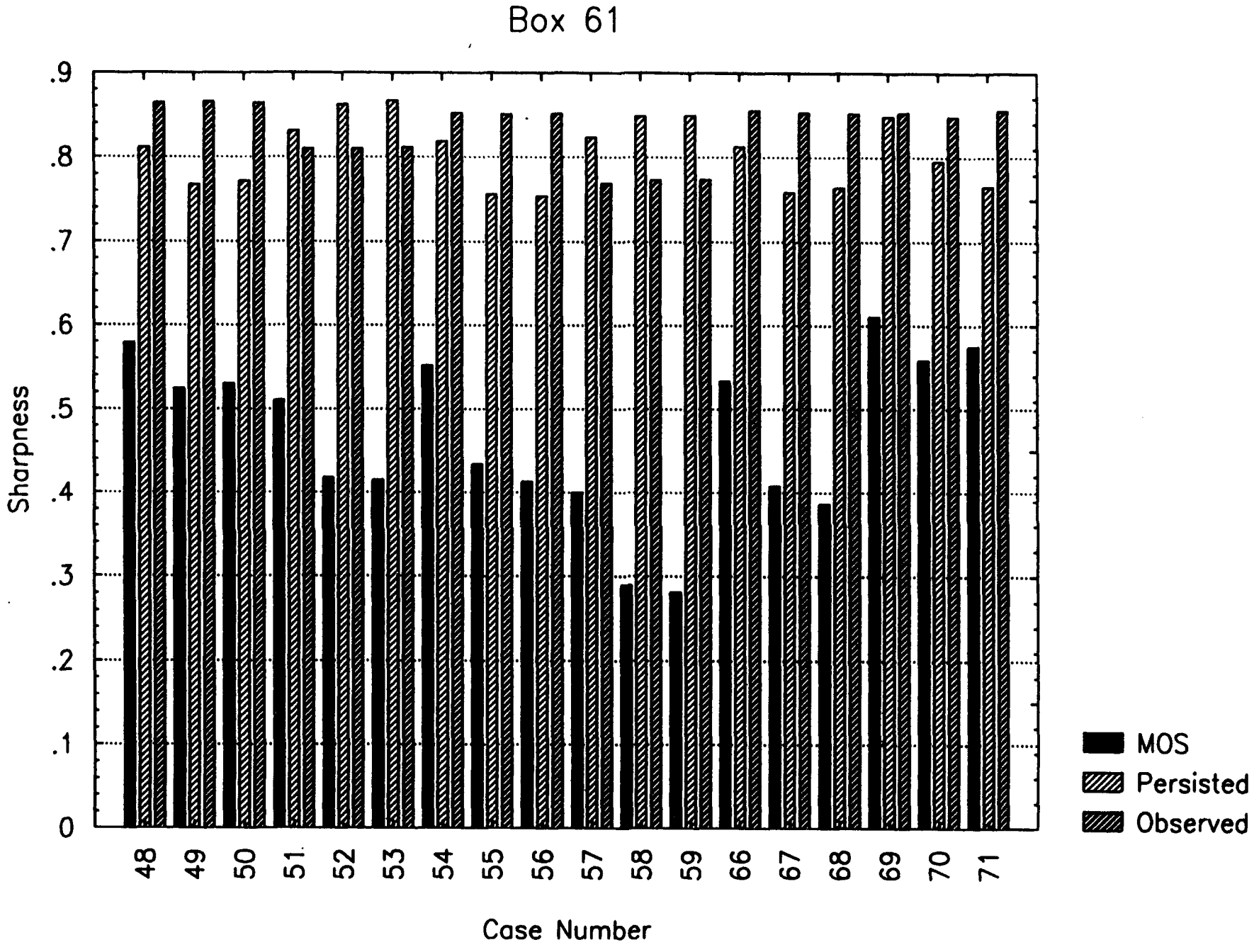


Figure 60 Comparison of Sharpness for MOS-based Forecasts, Persistence-based Forecasts and Observed Cloud Cover for all Cases in Box 61

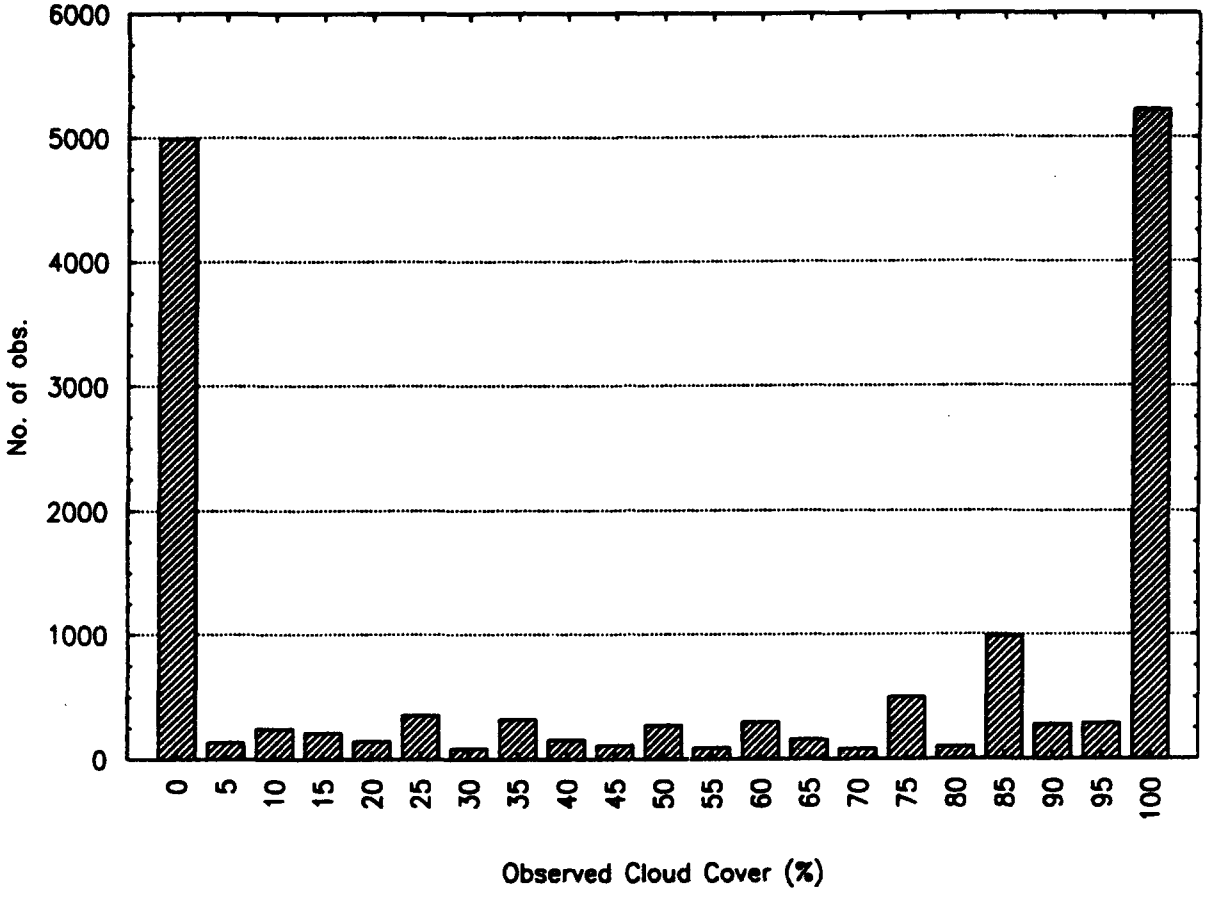


Figure 61 Histogram of Observed Cloud Cover Values for Case #32 (Box 30, Winter, 21Z) Showing a Very Sharp U-shaped Distribution

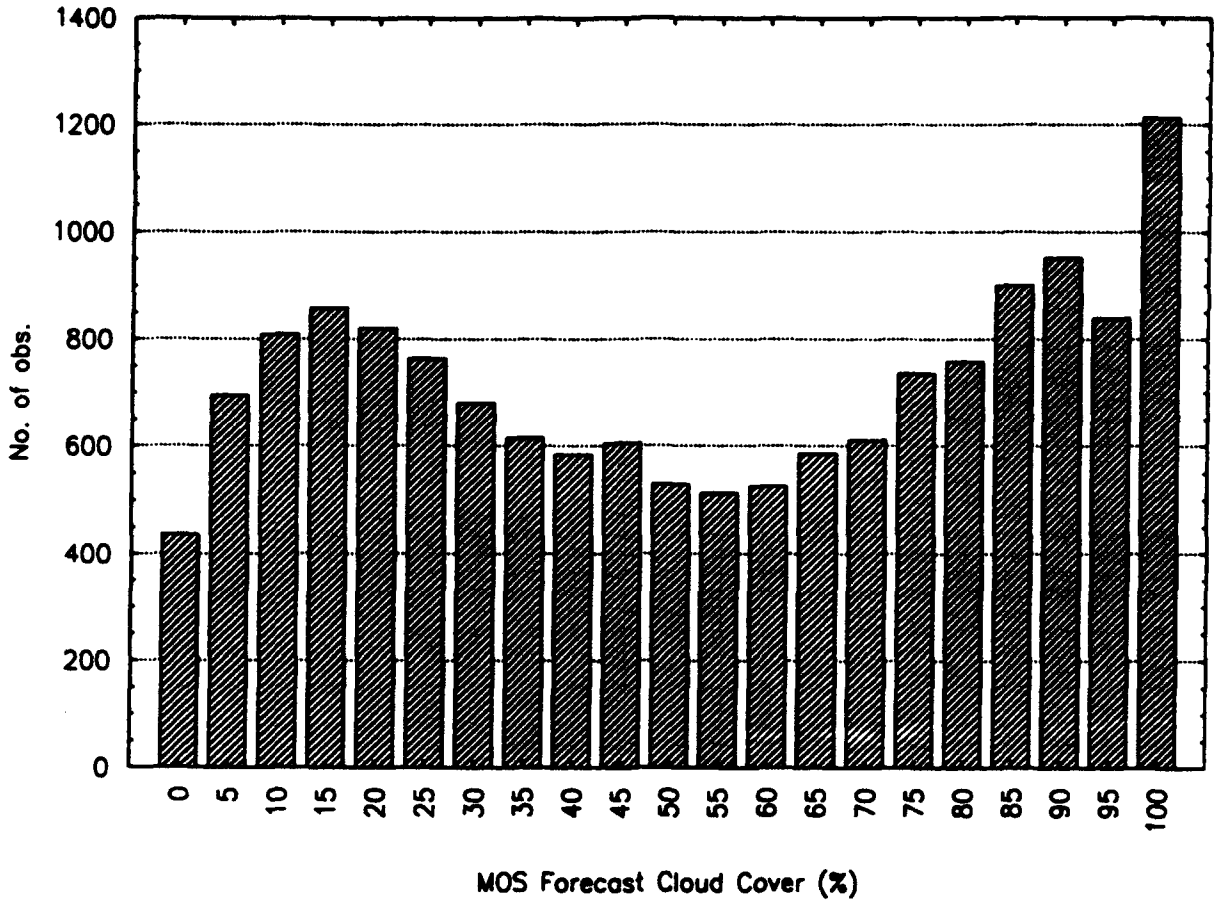


Figure 62 Histogram of 3-Hour Forecast Cloud Cover Values for Case #32 (Box 30, Winter, 21Z) Showing a Much Smoother Distribution Than the Observed Cloud Cover Shown in Figure 61

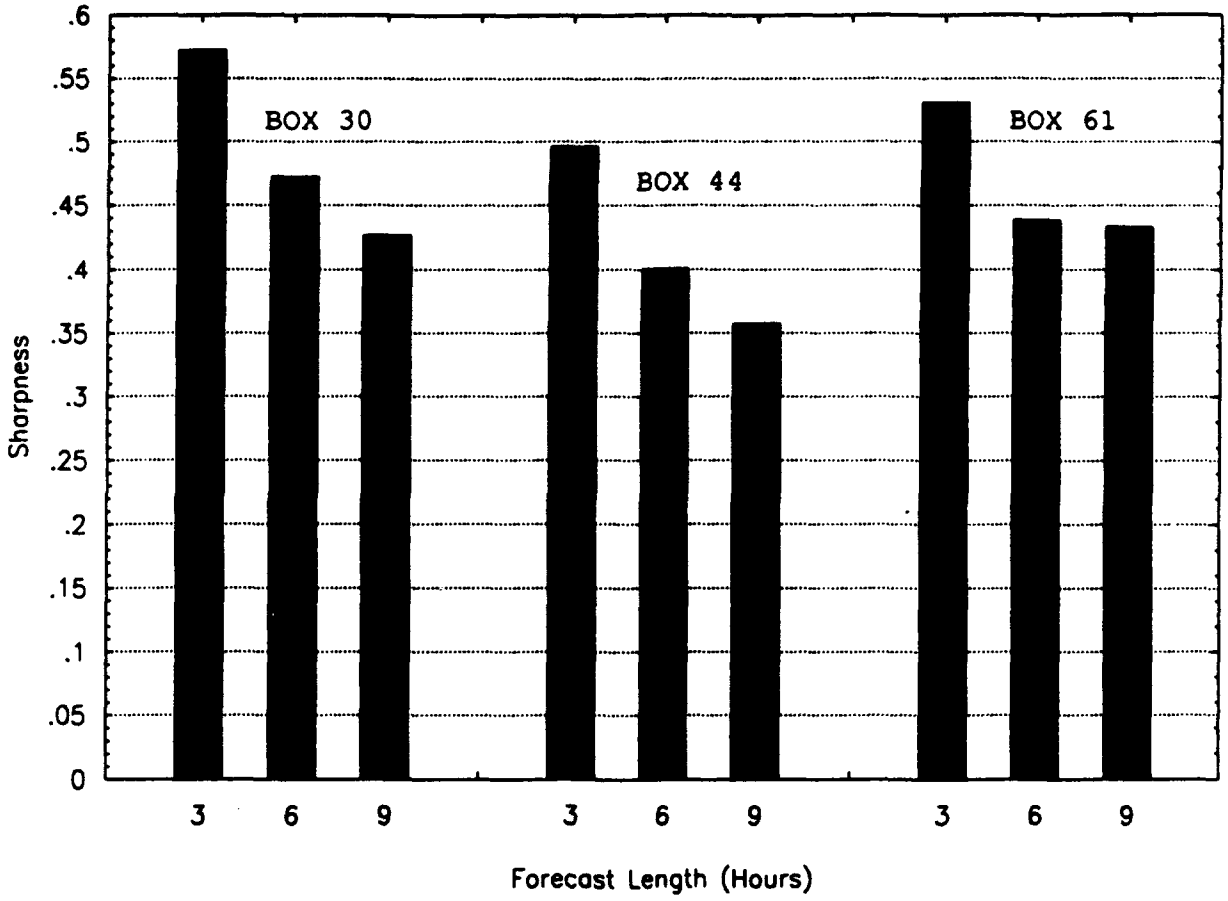


Figure 63 Averaged Sharpness for Each of the Box Regions Studied as a Function of Forecast Length

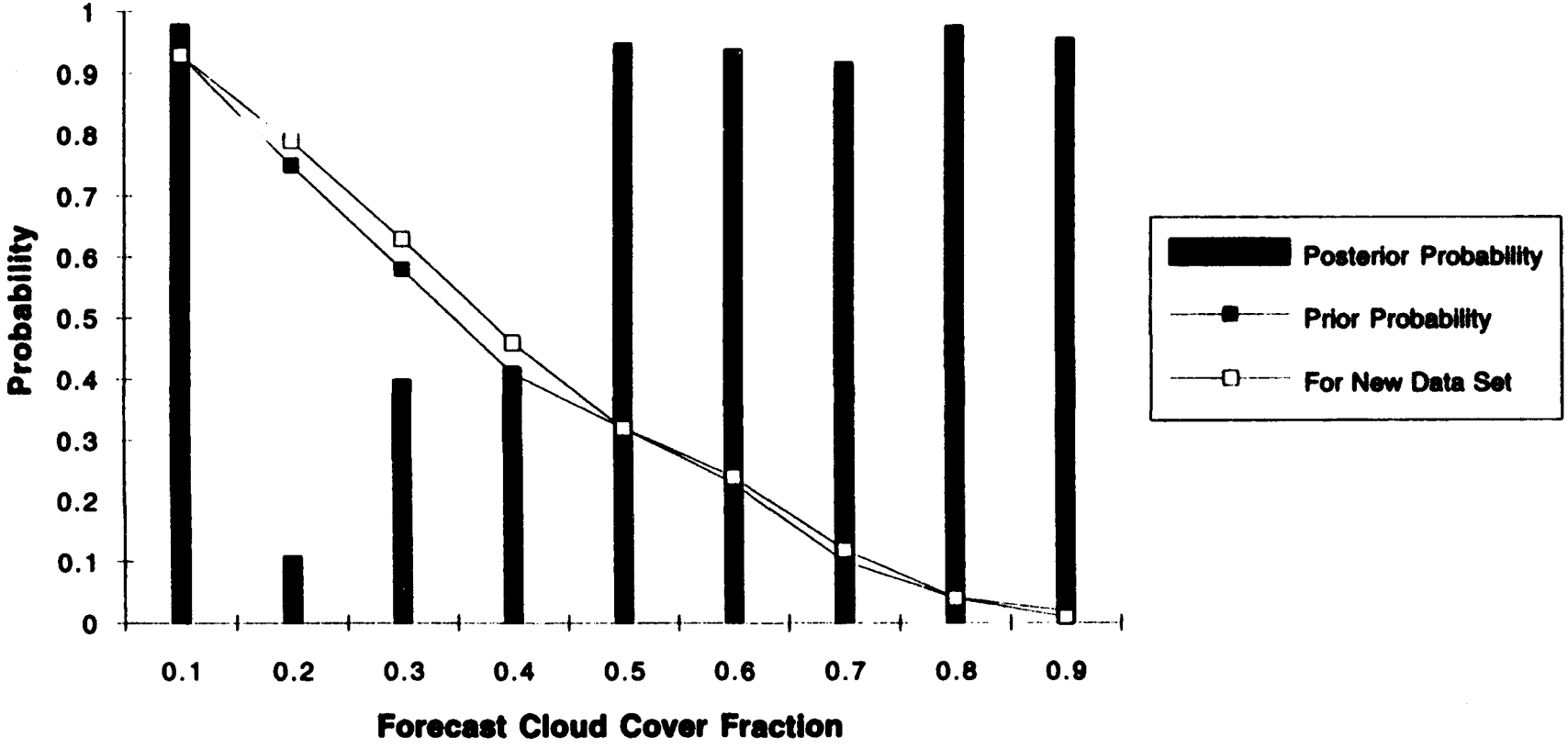


Figure 64 Verifying Forecast Cloud Cover Probability Case #12 Cloud Cover Threshold 30%

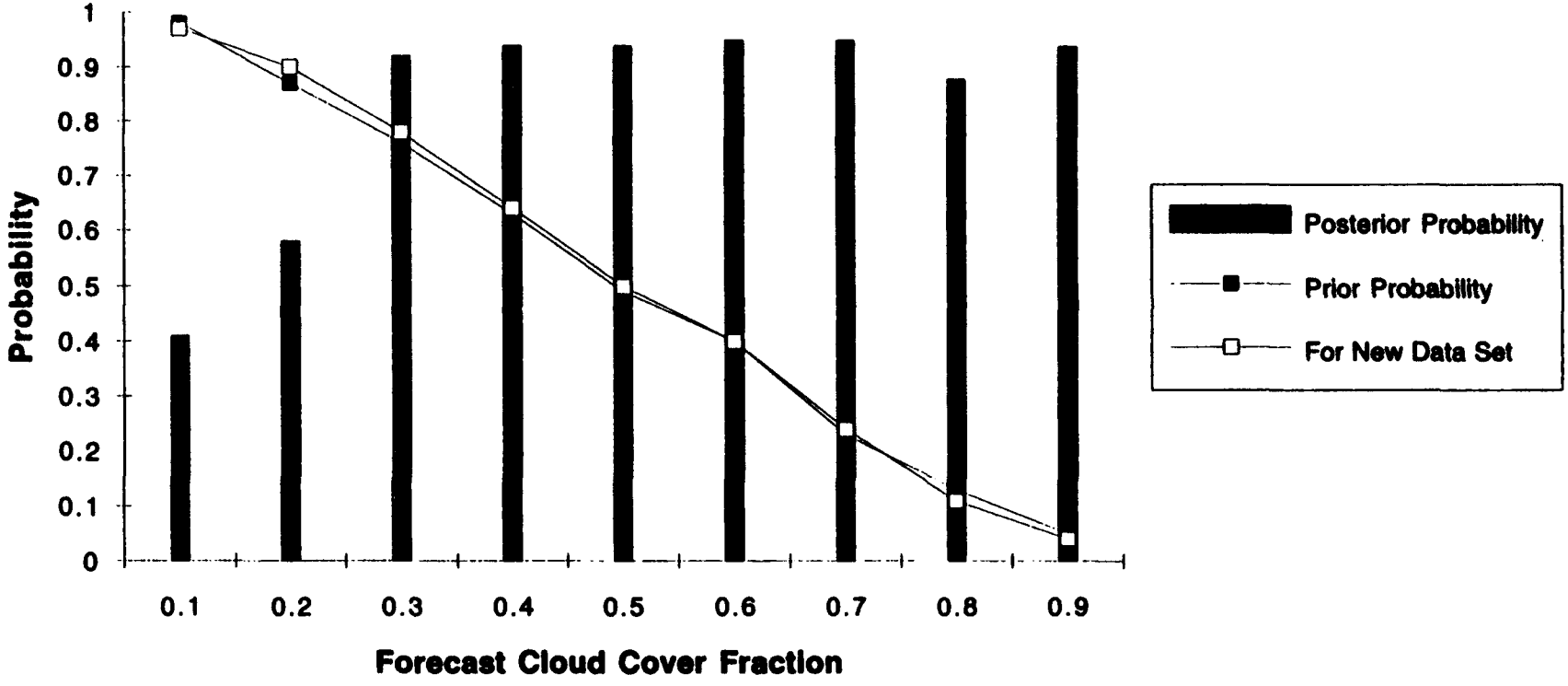


Figure 65 Verifying Forecast Cloud Cover Probability Case #12 Cloud Cover Threshold 50%

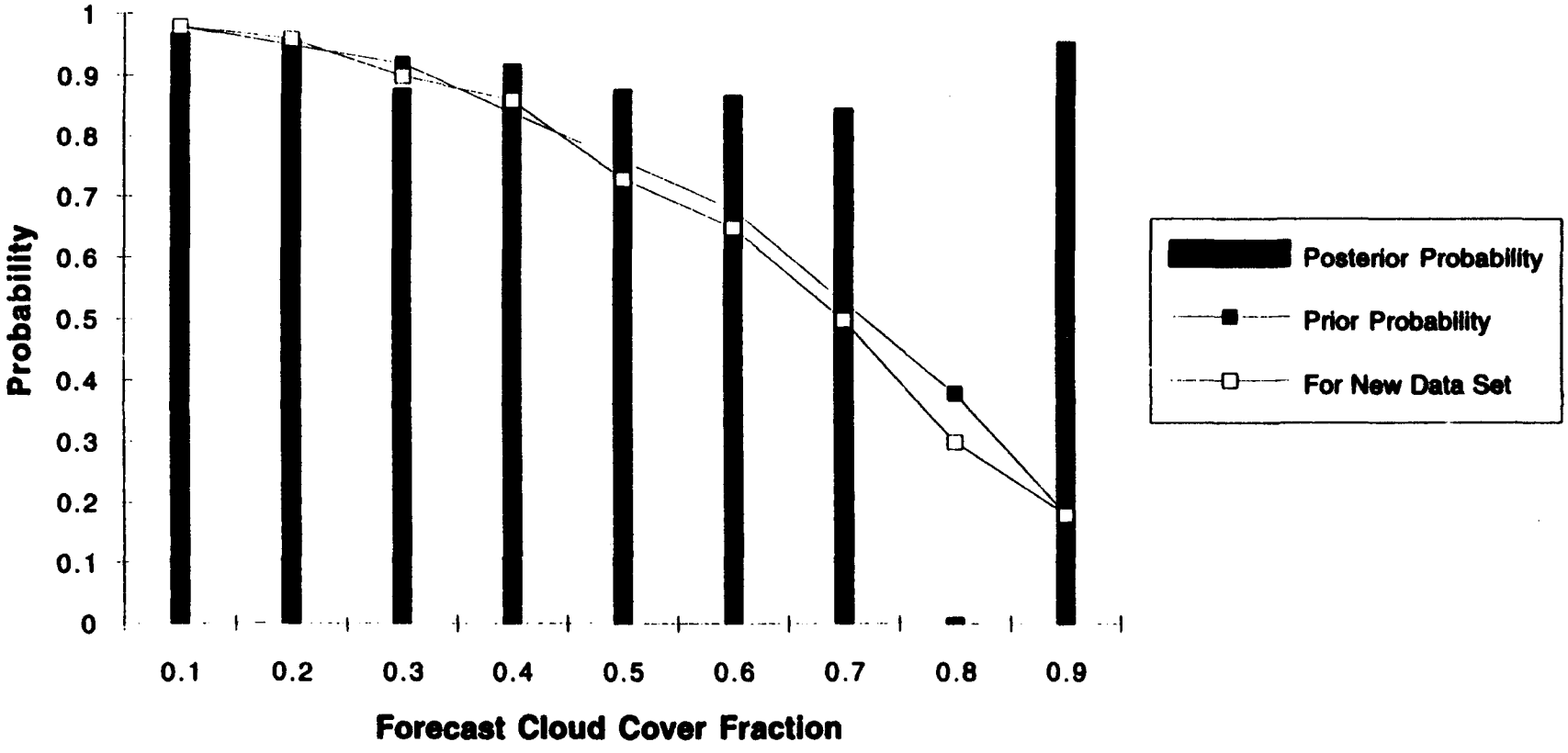


Figure 66 Verifying Forecast Cloud Cover Probability Case #12 Cloud Cover Threshold 70%

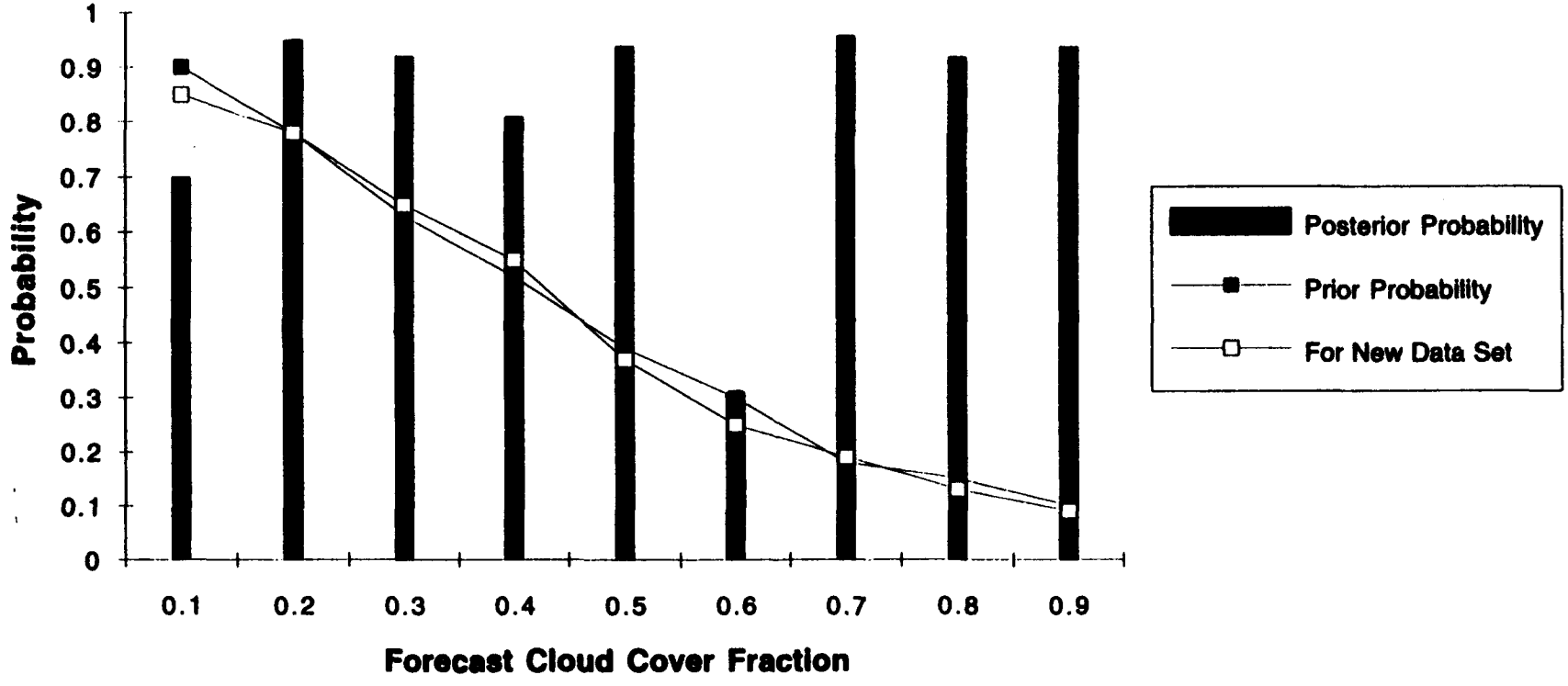


Figure 67 Verifying Forecast Cloud Cover Probability Case #59 Cloud Cover Threshold 30%

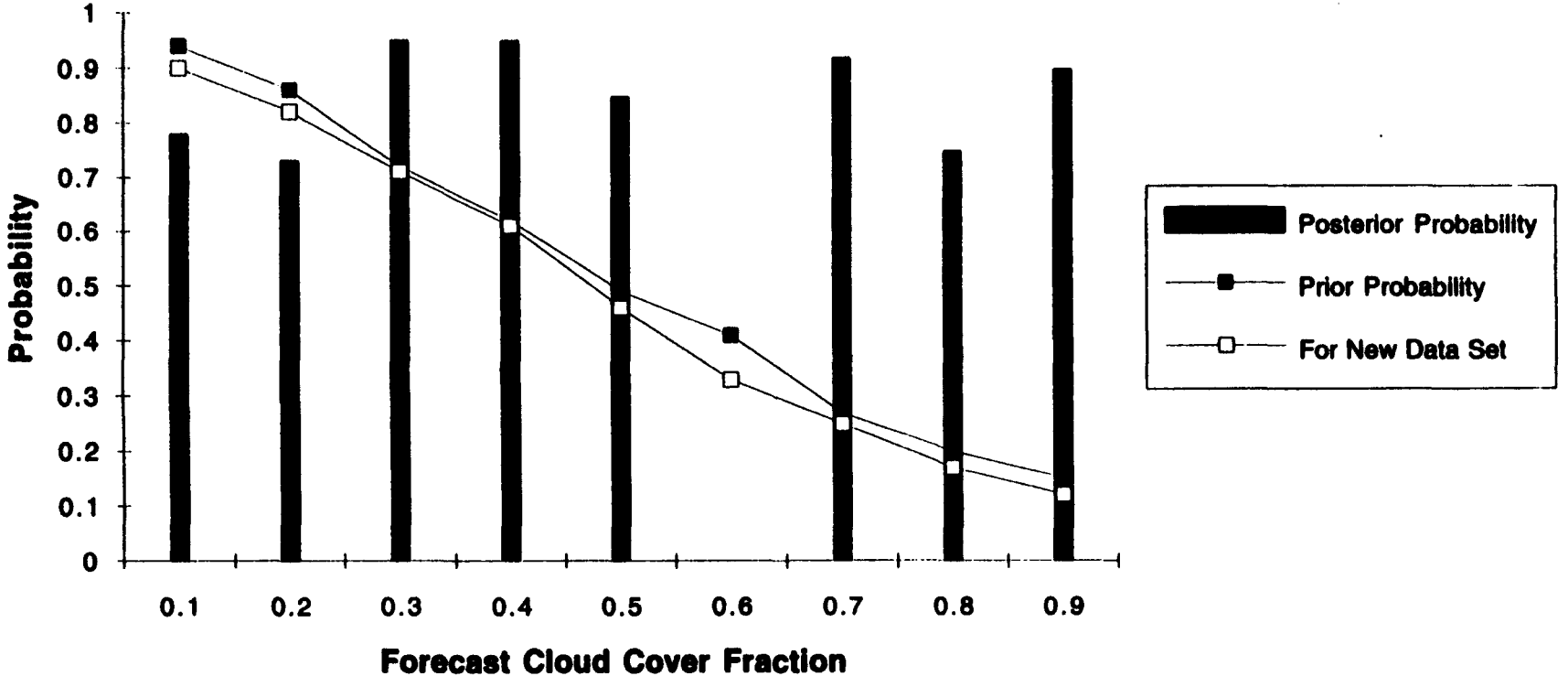


Figure 68 Verifying Forecast Cloud Cover Probability Case #59 Cloud Cover Threshold 50%

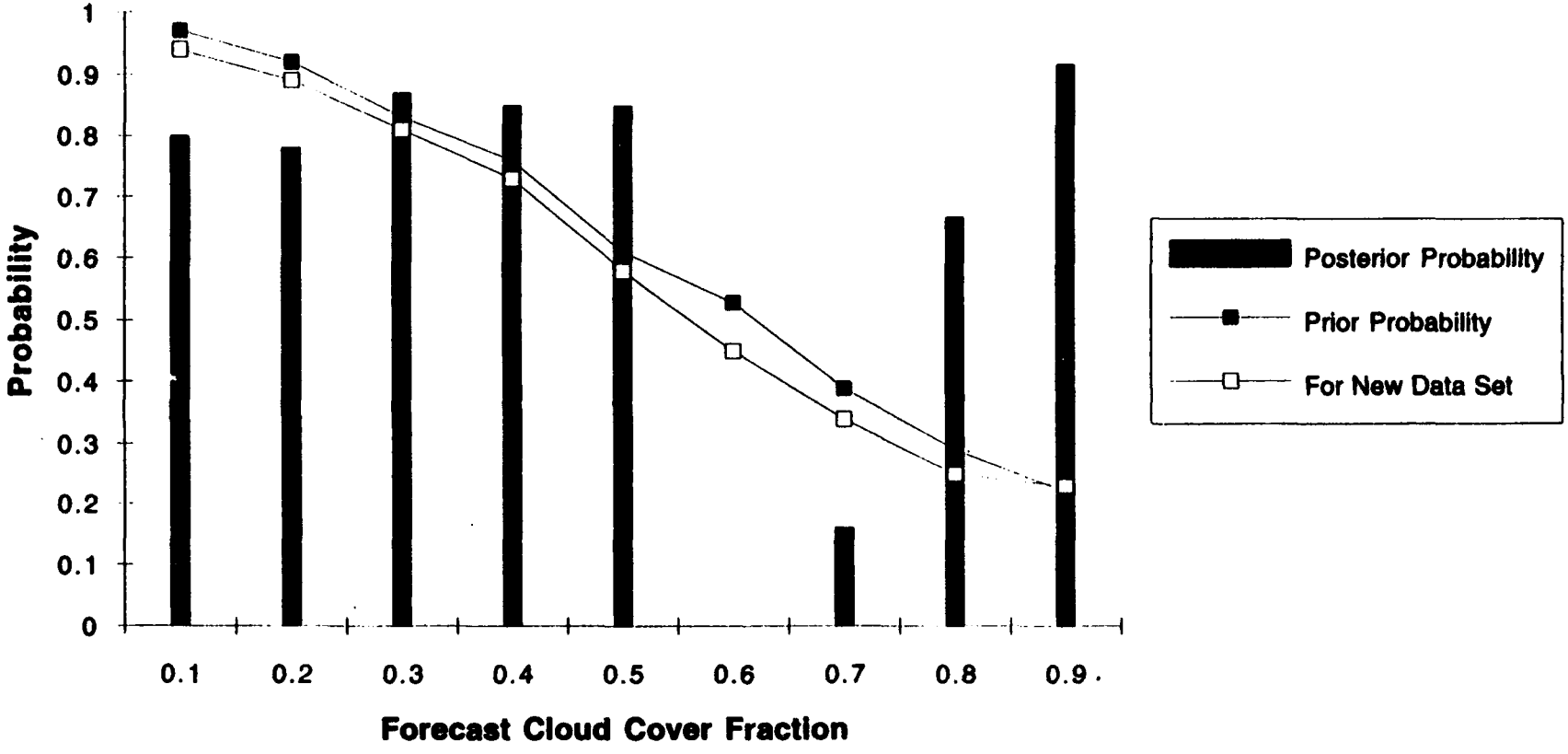


Figure 69 Verifying Forecast Cloud Cover Probability Case #59 Cloud Cover Threshold 70%

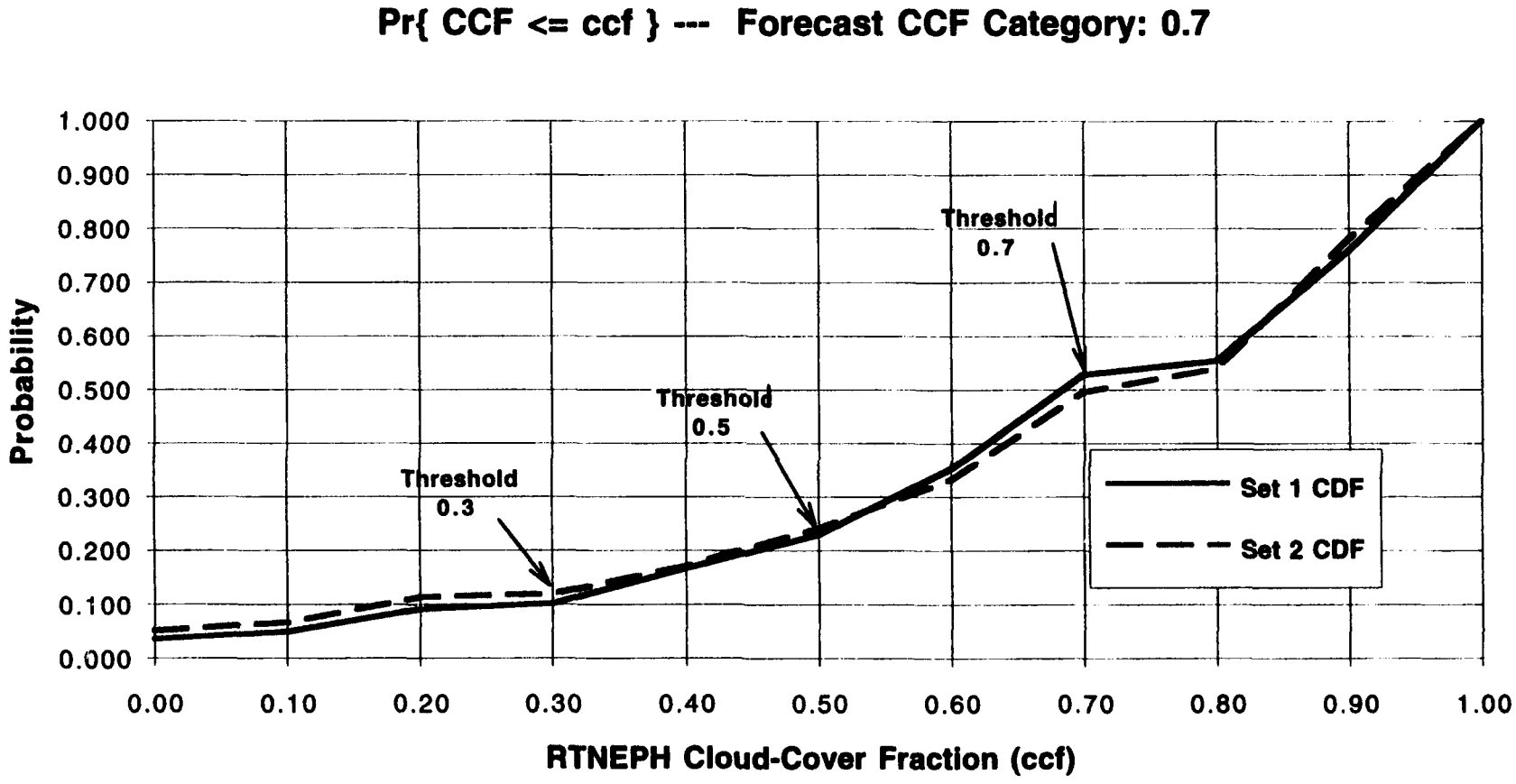


Figure 70 Experimental Distribution Functions of the True ccf for the MOS ccf Forecast Category of 0.7

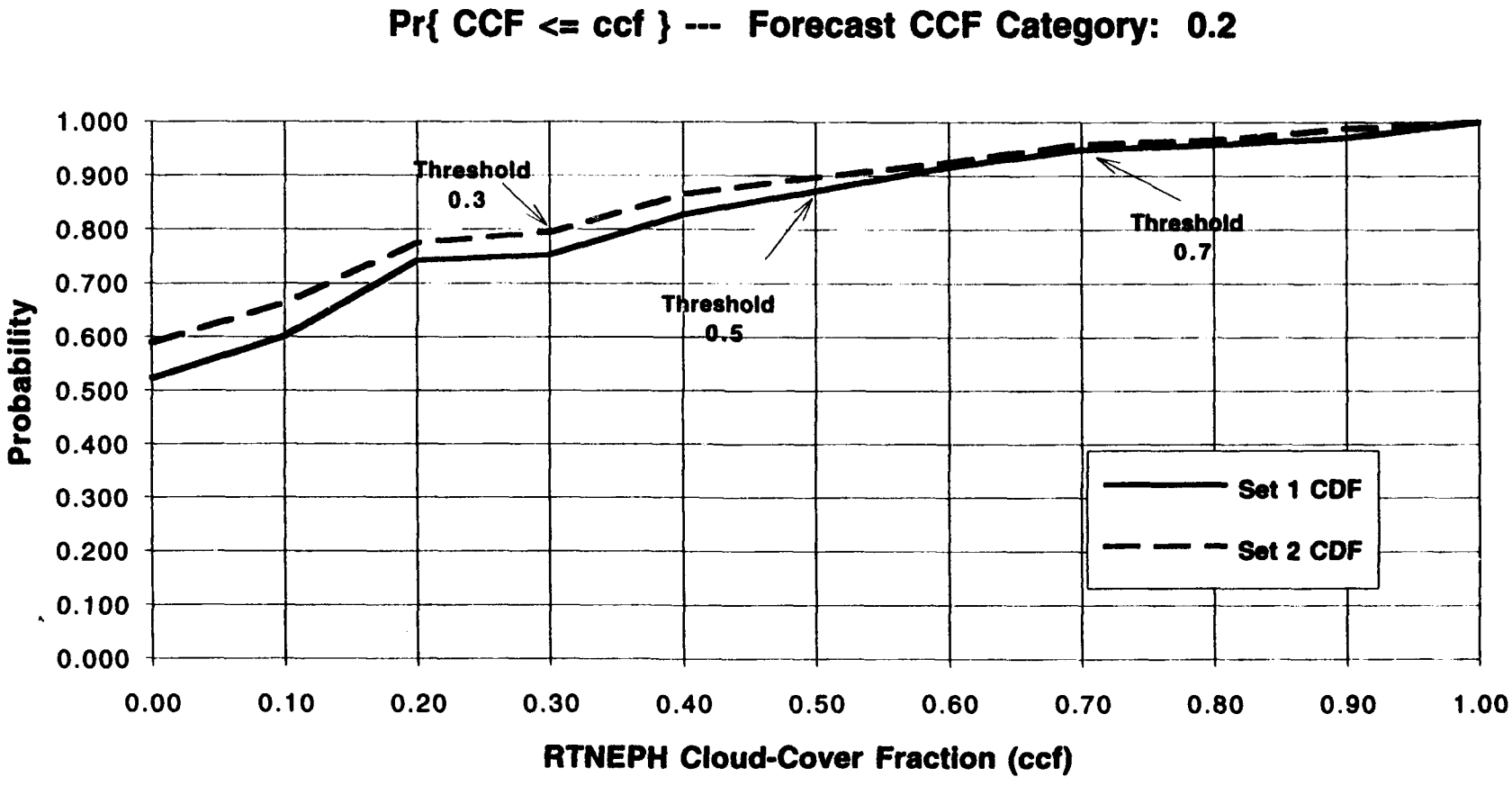


Figure 71 Experimental Distribution Functions of the True ccf for the MOS ccf Forecast Category of 0.2

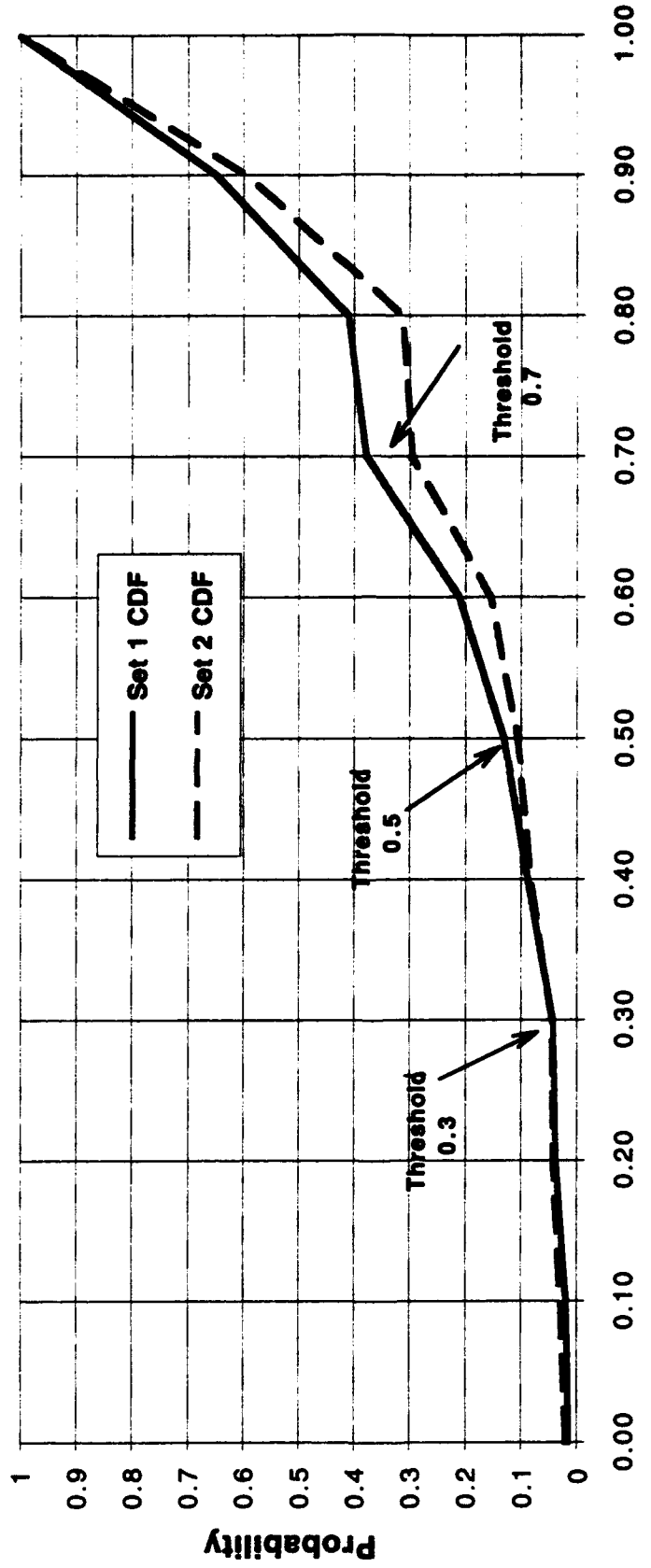


Figure 72 Experimental Distribution Functions of the True ccf for the MOS ccf Forecast Category of 0.8

4. RECOMMENDATIONS

As discussed in the previous section, the MOS-based forecast models developed under this effort performed well in terms of overall variance reduction and robustness. GSM variables added substantially to the skill of the cloud cover forecasts (we saw 35-55% improvement when compared with persistence forecasts alone). Our analysis showed that the MOS approach using Global Spectral Model variables in addition to persistence shows promise as a low cost cloud forecasting tool. (We use the term "low cost" because the MOS approach uses existing and previously verified forecast models along with existing cloud analysis databases. In addition, the MOS model as described here uses linear regression techniques that are computationally efficient.)

However, throughout this report we have also pointed out the limits of the models developed under this effort. In this section, we discuss those limitations and offer recommendations in three general areas: model development data, the modeling approach, and performance analysis and validation.

4.1 MODEL DEVELOPMENT DATA

Develop Models Using Higher-Resolution GSM Data

We generated MOS-based forecasts of total cloud cover at eighth-mesh grid resolution. Potential predictors at that resolution included terrain type, elevation, and persisted cloud amount. The remaining predictors used during development were derived from the 40-wave Global Spectral Model with a grid resolution *over four times more coarse than eighth-mesh*. Cloud processes important at eighth-mesh resolution (such as local terrain effects, sea breeze effects, and localized convection) are not modeled with the 40-wave GSM. Therefore the MOS equations developed with the 40-wave model data cannot support cloud formation and evolution at that resolution. In addition, the 2.5° grid used in the 40-wave spectral model effectively contains area averages of the model variables with respect to the much finer-resolution eighth-mesh grid, resulting in forecasts that are too smooth.

We recommend using higher-resolution GSM data in model development such as is currently available with the 80-wave (and higher) versions of the spectral model. The recently released Eta model from the National Weather Service is a good example of the

recent advances in global weather modeling; it has a horizontal grid resolution of approximately 80 km with 38 levels in the vertical. In the future, even higher-fidelity models will be available. Of course, higher model resolution alone is not sufficient to guarantee improved MOS forecasts. The models themselves must accurately predict the growth and decay of disturbances at those resolutions.

The increased spatial resolution would not necessarily increase data processing and storage costs if a similar number of grid point locations (such as the 15×15 grid points per RTNEPH box used in this analysis) were modeled. However, increases in the number of vertical levels can quickly increase the number of potential predictors. If necessary, the number of variables in the vertical could be reduced by using only their most significant principal components.

As with any significant change to the driving model, at least two years of stable development data would be required to develop operational MOS equations based on higher-resolution model output.

Include RTNEPH Layer Cloud Amounts and Types

Cloud *layer* fields reported in the RTNEPH analysis grids (or available from pilot reports, upper air analyses, etc.) could be used in conjunction with the humidity profiles generated in the GSM. Using cloud layer data would allow us to adjust humidity layers for the thickness of the cloud layer within each sigma level and thus eliminate any ambiguity concerning whether or not a cloud layer straddles two or more sigma levels. This would result in better estimates for the maximum layer humidity, a strong predictor for cloud amount. The quality of the cloud layer data, the additional value it might bring to the MOS equations, and the increased data processing and storage costs would all need to be weighed before deciding whether or not to include these data in future modeling efforts.

Use Independent Persistence and Response Data

In this effort we used the RTNEPH analysis grids both as persisted cloud cover and "ground truth." Thus, there is potentially a high correlation between the two. We attempted to remove the temporal correlation by removing all data points which had not been updated in the three hours prior to the forecast valid time. However, the time flag available to verify data age only indicates the age of the *newest* data source for the grid point; not necessarily the age of the total cloud cover reported for that grid point. Therefore, some old data may still be present in the development data sets.

In considering the results of this feasibility study, it is important to note the *relative* improvement of the MOS approach over and above persistence to account for potential artificial correlations in the development data sets due to the age of the data. In order to assess the *absolute* impact of persistence and MOS-based forecasts, independent persistence and response data sources should be used. Candidate data sources other than the RTNEPH analysis fields include: surface observations and GOES-based cloud analyses.

4.2 MODELING APPROACH

Include Additional Years of Data

It is well known that meteorological conditions vary significantly from year to year. We used one year of data to develop MOS models in this feasibility study. Although acceptable for a feasibility study such as this, it would be imperative to use additional years of data (at least two) for operational use. Once the modeling mechanism is defined, processing additional data is a mechanical process. By using additional years of data we could achieve more stable regression coefficients along with a predictor set that is more meaningful and less subject to any single year's meteorological peculiarities.

Minimize Absolute Error

Linear multiple regression algorithms, like the one used in this feasibility study, use least squares techniques to determine the MOS equations. This accentuates those cases with large residuals and can result in an overly conservative fit to the data (i.e., overly smooth forecasts). To make sharper forecasts, it may be desirable to minimize absolute error instead of squared error. The additional value of using absolute error would need to be evaluated against the possibly significant increase in computational time to perform the regression.

Use Logistic Regression

Logistic regression (Ref. 1) is a computationally intensive regression technique. In Section 3.1.1 we noted its advantages for modeling binary response functions such as categorical cloud amounts. The logistic function does not produce responses that lie outside the range [0, 1] as linear regression techniques do. Future efforts may evaluate the need for a non-linear regression technique such as logistic regression by weighing the significant additional computational costs (logistic regression is an iterative method which requires good initial estimates of the model parameters) against its advantages.

Use Neural Networks

Another non-linear modeling technique that is being used more and more, is neural networks. Neural networks provide a modeling capability that is different from conventional methods. Instead of using a particular functional dependence between the measured data and desired prediction, neural networks adjust themselves to give an optimal (with respect to some metric, such as the mean square error) non-linear transformation between predictors and the response variable. In addition, neural networks can be used to determine the minimum set of variables necessary to solve a problem.

The availability of COTS neural net software tools provides for a low risk/potentially high-payoff approach to the area of cloud cover prediction. One such software tool (ExploreNet from HNC) has already shown promise for this problem in a small test case performed at TASC.

4.3 PERFORMANCE ANALYSIS AND VALIDATION

Compute Other Performance Measures

In this effort we selected three performance statistics (Sharpness, Brier score, and 20/20 score) to measure the skill of the MOS-based forecasts. These three statistics were selected in consultation with PL personnel as meaningful in the context of potential applications for this MOS approach. Other applications may emphasize different aspects of the model forecasts. For example, certain satellite systems may place more emphasis on reliability measures. Future analyses could include more or different performance measures to accommodate the interests of the customer community.

Validate More Cases

We selected two cases for evaluation from the 71 development cases analyzed in this study. Those two cases were chosen because they had the highest and lowest performance of all 71 cases with respect to overall reduction in variance. It was our hope that by studying these two extreme cases we could bound our error estimates for both the total and categorical cloud amount forecasts. Additional cases should be analyzed in the future to get a better handle on the variation of the forecast strengths and weaknesses as a function of time of day, season, and region.

Validate Against More Data

In the two cases mentioned above we evaluated both total and categorical cloud amount forecasts by developing sample models using a portion of the data set and validating those against the remainder of the season of data.

It would be crucial if this approach was to be taken toward an operational model, that the models be tested against additional seasons of cloud data from independent years containing a greater variety of weather conditions. The estimates of model robustness determined in our analysis were most likely overly optimistic since the variability in one season is most likely less than across multiple seasons. We recommend developing models with at least two years of data and validating the models against at least two additional independent years.

APPENDIX A

A list of all of the variables (listed by their abbreviated names) used in the pool of potential predictors follows. A short description of each variable is provided along with the appropriate units and any specific numerical ranges if applicable.

VARIABLE NAME	VARIABLE DESCRIPTION	UNITS/ RANGE
DAY	day of the month (1–31)	[1,31]
TYPE	terrain type at eighth-mesh grid point at center (lower right) of half-mesh cell	0 = water, 1 = land, 3 = coast
ELEV	terrain elevation at eighth-mesh grid point at center (lower right) of half-mesh cell	meters
LAT	latitude at center of grid point	degrees
LON	longitude at center of grid point	degrees
ZENITH	sine of solar zenith	
UL	low level wind in zonal direction	meters/sec
UM	middle level wind in zonal direction	meters/sec
D3UL	3-hr change in low level zonal wind	meters/sec
D3UM	3-hr change in mid level zonal wind	meters/sec
D6UL	6-hr change in low level zonal wind	meters/sec
D6UM	6-hr change in mid level zonal wind	meters/sec
D9UL	9-hr change in low level zonal wind	meters/sec
D9UM	9-hr change in mid level zonal wind	meters/sec
VL	low level wind in meridional direction	meters/sec
VM	middle level wind in meridional direction	meters/sec
D3VL	3-hr change in low level meridional wind	meters/sec
D3VM	3-hr change in mid level meridional wind	meters/sec
D6VL	6-hr change in low level meridional wind	meters/sec
D6VM	6-hr change in mid level meridional wind	meters/sec
D9VL	9-hr change in low level meridional wind	meters/sec
D9VM	9-hr change in mid level meridional wind	meters/sec

VARIABLE NAME	VARIABLE DESCRIPTION	UNITS/ RANGE
SPEED	wind speed at the surface	meters/sec
SHEAR	wind shear between high and low layers	meters/sec
TL	low level average temperature	Kelvin
TM	middle level average temperature	Kelvin
TDIF1	temperature difference between sigma layers 1 and 0 (approx. $T_{850\ mb} - T_{surface}$)	Kelvin
TDIF2	temperature difference between sigma layers 4 and 1 (approx. $T_{500\ mb} - T_{850\ mb}$)	Kelvin
RHL	low level average relative humidity	[0.0,1.0]
LNRHL	natural logarithm of low level relative humidity	
RHL2	low level relative humidity squared	
RHL4	low level relative humidity raised to the power of 4	
RHM	middle level average relative humidity	[0.0,1.0]
LNRHM0	natural logarithm of middle level relative humidity	
RHM2	middle level relative humidity squared	
RHM4	middle level relative humidity raised to the power of 4	
RHMX	maximum relative humidity of six sigma layers	[0.0,1.0]
RHAB	relative humidity at layer above maximum	[0.0,1.0]
D3RHL	3-hr change in low level relative humidity	[-1.0,1.0]
D3RHM	3-hr change in middle level relative humidity	[-1.0,1.0]
D6RHL	6-hr change in low level relative humidity	[-1.0,1.0]
D6RHM	6-hr change in middle level relative humidity	[-1.0,1.0]
D9RHL	9-hr change in low level relative humidity	[-1.0,1.0]
D9RHM	9-hr change in middle level relative humidity	[-1.0,1.0]
THICK	850 mb thickness	meters
SFCP	surface pressure	Pascals
D3P	3-hr change in surface pressure	Pascals
D6P	6-hr change in surface pressure	Pascals
D9P	9-hr change in surface pressure	Pascals
SFCH	model surface height	meters
DVL	low level divergence	$second^{-1}$

VARIABLE NAME	VARIABLE DESCRIPTION	UNITS/ RANGE
DVM	middle level divergence	second^{-1}
DVH	high level divergence	second^{-1}
D3DVL	3-hr change in low level divergence	second^{-1}
D3DVM	3-hr change in middle level divergence	second^{-1}
D3DVH	3-hr change in high level divergence	second^{-1}
D6DVL	6-hr change in low level divergence	second^{-1}
D6DVM	6-hr change in middle level divergence	second^{-1}
D6DVH	6-hr change in high level divergence	second^{-1}
D9DVL	9-hr change in low level divergence	second^{-1}
D9DVM	9-hr change in middle level divergence	second^{-1}
D9DVH	9-hr change in high level divergence	second^{-1}
VTL	low level vorticity	second^{-1}
VTM	middle level vorticity	second^{-1}
VTH	high level vorticity	second^{-1}
D3VTL	3-hr change in low level vorticity	second^{-1}
D3VTM	3-hr change in middle level vorticity	second^{-1}
D3VTH	3-hr change in high level vorticity	second^{-1}
D6VTL	6-hr change in low level vorticity	second^{-1}
D6VTM	6-hr change in middle level vorticity	second^{-1}
D6VTH	6-hr change in high level vorticity	second^{-1}
D9VTL	9-hr change in low level vorticity	second^{-1}
D9VTM	9-hr change in middle level vorticity	second^{-1}
D9VTH	9-hr change in high level vorticity	second^{-1}
TAL	low level temperature advection	Kelvin/sec
TAM	middle level temperature advection	Kelvin/sec
TAH	high level temperature advection	Kelvin/sec
D3TAL	3-hr change in low level temperature advection	Kelvin/sec
D3TAM	3-hr change in middle level temperature advection	Kelvin/sec
D3TAH	3-hr change in high level temperature advection	Kelvin/sec
D6TAL	6-hr change in low level temperature advection	Kelvin/sec
D6TAM	6-hr change in middle level temperature advection	Kelvin/sec

VARIABLE NAME	VARIABLE DESCRIPTION	UNITS/ RANGE
D6TAH	6-hr change in high level temperature advection	Kelvin/sec
D9TAL	9-hr change in low level temperature advection	Kelvin/sec
D9TAM	9-hr change in middle level temperature advection	Kelvin/sec
D9TAH	9-hr change in high level temperature advection	Kelvin/sec
WL	low level vertical velocity	$second^{-1}$
WM	middle level vertical velocity	$second^{-1}$
WH	high level vertical velocity	$second^{-1}$
D3WL	3-hr change in low level vertical velocity	$second^{-1}$
D3WM	3-hr change in middle level vertical velocity	$second^{-1}$
D3WH	3-hr change in high level vertical velocity	$second^{-1}$
D6WL	6-hr change in low level vertical velocity	$second^{-1}$
D6WM	6-hr change in middle level vertical velocity	$second^{-1}$
D6WH	6-hr change in high level vertical velocity	$second^{-1}$
D9WL	9-hr change in low level vertical velocity	$second^{-1}$
D9WM	9-hr change in middle level vertical velocity	$second^{-1}$
D9WH	9-hr change in high level vertical velocity	$second^{-1}$
WRHL	product of low level vertical velocity and relative humidity	$second^{-1}$
WRHM	product of middle level vertical velocity and relative humidity	$second^{-1}$
DELEV	elevation gradient in local average wind direction (normalized to grid box size)	dimensionless
ACCL3	3-hr upwind NEPH cloud cover (averaged to half mesh) using average low level wind	[0.0,1.0]
ACCL6	6-hr upwind NEPH cloud cover (averaged to half mesh) using average low level wind	[0.0,1.0]
ACCL9	9-hr upwind NEPH cloud cover (averaged to half mesh) using average low level wind	[0.0,1.0]
ACCM3	3-hr upwind NEPH cloud cover (averaged to half mesh) using average middle level wind	[0.0,1.0]
ACCM6	6-hr upwind NEPH cloud cover (averaged to half mesh) using average middle level wind	[0.0,1.0]
ACCM9	9-hr upwind NEPH cloud cover (averaged to half mesh) using average middle level wind	[0.0,1.0]

VARIABLE NAME	VARIABLE DESCRIPTION	UNITS/ RANGE
ACCH3	3-hr upwind NEPH cloud cover (averaged to half mesh) using average high level wind	[0.0,1.0]
ACCH6	6-hr upwind NEPH cloud cover (averaged to half mesh) using average high level wind	[0.0,1.0]
ACCH9	9-hr upwind NEPH cloud cover (averaged to half mesh) using average high level wind	[0.0,1.0]
RH50	binary variable that is 1 if relative humidity is greater than 50%, else 0	0,1
RH70	binary variable that is 1 if relative humidity is greater than 70%, else 0	0,1
RH90	binary variable that is 1 if relative humidity is greater than 90%, else 0	0,1
NEPH3	RTNEPH cloud cover at the eighth-mesh grid point at the center (lower right) of the half-mesh cell 3 hours before valid time	[0.0,1.0]
NEPH6	RTNEPH cloud cover at the eighth-mesh grid point at the center (lower right) of the half-mesh cell 6 hours before valid time	[0.0,1.0]
NEPH9	RTNEPH cloud cover at the eighth-mesh grid point at the center (lower right) of the half-mesh cell 9 hours before valid time	[0.0,1.0]
NEPH12	RTNEPH cloud cover at the eighth-mesh grid point at the center (lower right) of the half-mesh cell 12 hours before valid time	[0.0,1.0]
NEPH15	RTNEPH cloud cover at the eighth-mesh grid point at the center (lower right) of the half-mesh cell 15 hours before valid time	[0.0,1.0]
NEPH18	RTNEPH cloud cover at the eighth-mesh grid point at the center (lower right) of the half-mesh cell 18 hours before valid time	[0.0,1.0]
NEPH21	RTNEPH cloud cover at the eighth-mesh grid point at the center (lower right) of the half-mesh cell 21 hours before valid time	[0.0,1.0]
NEPH24	RTNEPH cloud cover at the eighth-mesh grid point at the center (lower right) of the half-mesh cell 24 hours before valid time	[0.0,1.0]

APPENDIX B

In this appendix we present a derivation of the formula for posterior probability that produced the evaluation results on MOS forecast probabilities reported in Section 3.5.3. While elementary, the derivation requires careful use of delta functions to represent probability measures (equivalently, probability density functions) which are defined on a continuous domain (the interval [0, 1] in this case) yet have non-zero weight (or probability) at an isolated point.

In the derivation, we assume that the event of interest, E(i, j), has been fixed (i.e., a ccf bin, B_i, and a threshold value, T_j, have been selected) and suppress notation identifying these parameters. Define

p = the true probability that E(i,j) will occur
$\hat{p}$ = the MOS forecast probability of E(i,j) based on data set D_1

n = the number of observations in data set D_2 for which $\hat{p}$ is in bin B_i,
k = the number of observations in data set D_2 for which E(i,j) occurs.

Now, our Bayesian viewpoint picks, somewhat arbitrarily, the following prior probability measure on the unknown (true) probability, p:

$$f(p) = \begin{cases} 0.5\ \delta(p-\hat{p}) + 0.5, & 0 \le p \le 1.0 \\ 0 \quad , & \text{otherwise} \end{cases} \qquad \text{(B-1)}$$

where $\delta(x)$ is a delta function concentrated at $x = 0$. Once data set D_2 has been analyzed (i.e., the values of n and k have been determined for D_2), the likelihood of a particular value, y, is defined by

$$L(y/D_2) = y^k(1-y)^{n-k}. \qquad \text{(B-2)}$$

Note that the likelihood has the form of a binomial density function (lacking only a scaling constant) because data set D_2 is assumed to provide n independent trials. Each "trial" consists of comparing the actual cloud cover fraction at the forecast valid time with the selected threshold, T_j.

Using Eqs. B-1 and B-2, the *posterior probability* measure on p has the form

$$f_{p/D_2}(p) = \begin{cases} (0.5c)\cdot p^k(1-p)^{n-k}[\delta(p-\hat{p}) + 1.0], & 0 \le p \le 1.0 \\ 0 & , \text{ otherwise} \end{cases} \tag{B-3}$$

where the constant, c, must be chosen so that the integral over [0, 1] is equal to 1.0. Carrying out the integration results in

$$0.5c = \left\{ \hat{p}^k(1-\hat{p})^{n-k} + \int_0^1 x^k(1-x)^{n-k}dx. \right\}^{-1} \tag{B-4}$$

So, to calculate the posterior probability (p^*) that $p = \hat{p}$, we can evaluate the following

$$p^* = \lim_{\Delta \to 0} \int_{\hat{p}-\Delta}^{\hat{p}+\Delta} f_{p/D_2}(p)dp\,. \tag{B-5}$$

Substituting Eqs. B-3 and B-4 into Eq. B-5 and evaluating the limit yields the desired formula, Eq. 3-25.

REFERENCES

1. Aldrich, J.H., and F.D. Nelson, *Linear probability, logit, and probit models,* Sage Pubs, Beverly Hills, CA, 1984.
2. Brenner, S., Yang, C.H., and S.Y.K. Yee, The AFGL spectral model of the moist global atmosphere: documentation of the baseline version, Environmental Research Papers No. 815, AFGL-TR-82-0393, U.S. Air Force Geophysics Laboratory, Hanscom AFB, MA, 1982, ADA129283.
3. Brier, G.W., Verification of forecasts expressed in terms of probability, *Monthly Weather Review,* vol. 79, January 1950, pp 1–3.
4. Brunet, N., Verret, R., and N. Yacowar, An objective comparison of model output statistics and "perfect prog" systems in producing numerical weather element forecasts, *Weather and Forecasting*, vol. 3, December 1988, pp 273–283.
5. Burger, C.F., and I.I. Gringorten, Two-dimensional modeling for lineal and areal probabilities of weather conditions, Environmental Research Papers No. 875, AFGL-TR-84-0126, U.S. Air Force Geophysics Laboratory, Hanscom AFB, MA, 1984, ADA147970.
6. Canavos, G.C., *Applied probability and statistical methods*, Little, Brown, & Company, Boston, MA, 1984.
7. Carter, G.M., and H.R. Glahn, Objective prediction of cloud amount based on model output statistics, *Monthly Weather Review,* vol. 104, December 1976, pp 1565–1572.
8. Carter, G.M., Dallavalle, J.P., and H.R. Glahn, Statistical forecasts based on the National Meteorological Center's numerical prediction system, *Weather and Forecasting,* vol. 4, September 1989, pp 401–412.
9. Crum, T.D. (Editor), AFGWC cloud forecast models, Air Force Global Weather Central Technical Report AFGWC/TN-87/001, Offut AFB NE, April 1987.
10. Draper, N.R., and H. Smith, *Applied regression analysis,* Wiley, New York, 1966.
11. Erickson, M., private conversation with P. Janota.
12. Efron, B., *The jackknife, the bootstrap, and other resampling plans*, SIAM, Philadelphia, 1982.
13. Glahn, H.R., and D.A. Lowry, The use of model output statistics (MOS) in objective weather forecasting, *Journal of Applied Meteorology,* vol. 11, December 1972, pp 1203–1211.
14. Glahn, H.R., Problems in the use of probability forecasts, Proceedings of the fifth Conference on Weather Forecasting and Analysis, American Meteorological Society, Boston, March 1974, pp 32–35.

15. Glahn, H.R., An objective cloud forecasting system, Proceedings of the Fifth Conference on Weather Forecasting and Analysis, American Meteorological Society, Boston, March 1974, pp 79–80.

16. Glahn, H.R., and K.F. Hebenstreit, Design study for the development and use of model output statistics in automated aviation weather forecasting, Air Force Geophysics Technical Report AFGL-TR-79-0002, U.S. Air Force Geophysics Laboratory, Hanscom AFB, MA, December 1978, ADA072849.

17. Jacks, E., et al., New NGM-based MOS guidance for maximum/minimum temperature, probability of precipitation, cloud amount, and surface wind, *Weather and Forecasting*, vol. 5, March 1990, pp 128–138.

18. Lindley, D.V., and L.D. Phillips, Inference for a Bernoulli process (a Bayesian view), *The American Statistician*, August 1976, v. 30, no. 3, pp 112–119.

19. Mosteller, F., and J.W. Tukey, *Data analysis and regression, a second course in statistics*, Addison-Wesley, Reading, MA, 1977.

20. Muench, S., private conversation.

21. Perrone, T.J., and R.G. Miller, General exponential Markov and model output statistics: a comparative verification, *Monthly Weather Review*, vol. 113, September 1985, pp 1524–1541.

22. Pool, W., Regression estimation of event probabilities, Selected topics in Statistical Meteorology, Air Weather Service Technical Report AWS-TR-77-273, Scott AFB, IL, July 1977, ch. 5.

23. RTNEPH USAFETAC Climatic database user's handbook no. 1, Air Force Environmental Technical Applications Center Handbook USAFETAC/UH-86/001, Asheville, NC, September 1986.

24. Sela, J.G., The NMC spectral model, National Oceanic and Atmospheric Administration Technical Report NWS 30, Silver Springs, MD, May 1982.

25. Stipanuk, G.S., Algorithms for generating a skew-T, log P diagram and computing selected meteorological quantities, Army Electronics Commands Fort Monmouth, NJ, Report No. ECOM-5515, October 1973.

26. TASC Technical Proposal, A proposal for short-range cloud amount forecasting with model output statistics, TP-9826, June 1991.