U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT OFFICE
TECHNIQUES DEVELOPMENT LABORATORY

TDL OFFICE NOTE 83-9

MOS SUPPORT FOR MILITARY LOCATIONS FROM THE TECHNIQUES DEVELOPMENT LABORATORY

Harry R. Glahn

June 1983

MOS SUPPORT FOR MILITARY LOCATIONS FROM THE TECHNIQUES DEVELOPMENT LABORATORY

Harry R. Glahn[1]

1. INTRODUCTION

The MOS system developed by the Techniques Development Laboratory (TDL) and implemented at the National Meteorological Center (NMC) on NOAA's large computer system furnishes guidance forecasts twice a day out through 48 or more hours for approximately 294 civilian and 190 military locations. All stations are located within the conterminous 48 states (CONUS) and Alaska. Different developmental procedures are used for different weather elements, mainly because of meteorological considerations. Also, the structure of the developmental and operational systems dictate that CONUS and Alaskan stations be separated in the developmental process. For some weather elements and projections, the current observation is used in the forecast algorithm. For these situations, we must have a backup procedure so that a forecast can be made whenever the observation is missing in real time. Finally, because of data availability, military stations have sometimes been treated differently than civilian stations in the developmental process. A relatively new data source--an hourly data archive--is now making it possible to develop for military and civilian stations at the same time for most weather elements; a new guidance package has recently been implemented for 27 civilian and 12 Department of Defense (DoD) locations in Alaska for which this was done.

This paper briefly describes the development and implementation processes of the MOS system, taking particular note of the differences between guidance availability and accuracy for civilian and military stations.

2. THE DEVELOPMENTAL PROCESS

MOS, being a statistical technique, requires great quantities of carefully quality controlled data for the development of a useful operational system. These data are generally of two types--those from numerical models (to be used as "predictors") and observations of weather elements (to be used as "predictands" and, in some cases, as predictors).

In order to develop the hundreds of thousands of predictor-predictand relationships necessary for providing quality guidance for many weather elements, projections, and locations, a well documented software system must be available. The data archives, software system, development of predictor-predictand relationships, and system documentation are explained below, particularly as they pertain to DoD interests.

A. Data Archives

<u>Models</u> - NMC's Limited Area Fine Mesh (LFM) model has been archived since October 1971 on a grid covering the CONUS. The grid was expanded to cover a much larger area including Alaska in September 1977. Initially, the LFM ran

[1] This report was prepared while the author was on duty with the Air Weather Service, U.S. Air Force.

only to 24 hours but has been run to 48 hours since February 1976. A total of 336 "fields" are being saved twice a day. These are forecasts of geopotential height, temperature, humidity, and wind at various levels at 6-h intervals. Although there have been numerous small changes in the LFM over the period of record, perhaps the most major changes were the decrease in grid mesh size in August 1977 and the simultaneous increase in grid mesh size and the use of 4th-order differencing in June 1981.

NMC's large-scale Primitive Equation (PE) model and its successor, the spectral model, have been archived twice daily since July 1969 on four different grids covering the CONUS, Alaska, Hawaii, and Puerto Rico. A total of 180 fields, similar to those saved from the LFM, cover projections out to 60 hours at 6- or 12-h intervals. In addition, selected fields for projections out to 240 hours are being saved at 0000 GMT only.

TDL's trajectory model has been archived since July 1969. This model was driven by PE winds until December 1980 and then was switched to the LFM. The trajectory model was developed specifically for forecasting thunderstorms and severe convective weather. A total of 22 forecast fields, such as the K Index, Total Totals Index, 850-mb 24-h net vertical displacement, and 700-mb to surface convective instability, are saved for the 24-h projection only.

Observations - The basic predictand data set is available since October 1972. It consists of observations of 18 weather elements at 3-h intervals for about 255 U.S. civilian stations, 14 of which are in Alaska, four in Hawaii, and one in Puerto Rico. Data also exist for these same stations at 6-h intervals for a 3-year period prior to October 1972. These data, dubbed "Asheville" data, are purchased from the National Climatic Data Center on a monthly basis and very carefully quality controlled by a machine-manual method. At the time the archive started, these 255 stations were operating 24 hours a day.

Another data collection was started in December 1976, and is now beginning to be used as predictand data. It consists of a real-time archive of essentially all hourly reports, including military, in North America. These hourly reports are augmented in real-time by synoptic reports. Besides location and identification, 27 weather elements are saved whenever available. On the order of 550 to 800 reports may be available at a given hour. This collection nearly triples the density of predictand data in both space and time, as well as provides observations of wind gusts and amount and height of individual cloud layers which were not previously archived. These data are also subjected to a careful quality control process, but because of the "on-line" source of the data, more care is needed in their use.

Other data being archived for specialized uses or of shorter periods of record include manually digitized radar (MDR) data, severe local storm reports, hourly precipitation amounts from climatological stations, radiosonde data, and GOES satellite data.

B. The Software System

The developmental and operational MOS software systems have evolved over a 15-year period. All development programs that have more than transient interest are documented according to rigid standards and entered into a TDL library. After entry, they are not modified without laboratory approval.

C. Development of Forecast Equations

Any statistical model can be used to develop predictor-predictand relationships. By far, the model most extensively used in TDL's MOS system is multiple regression. This model is very versatile and can be highly non-linear through special treatment of the input predictors and/or predictands. For a quasi-continuous predictand such as surface temperature and a quasi-continuous predictor such as 850-mb temperature, both being somewhat normally distributed, little is gained by transformations. However, when the predictand is binary, such as precipitation occurrence or ceiling height in a particular category, the predictors may need to be transformed non-linearly into new predictors before developing the regression equation.

Each weather element poses different problems in the developmental process. Largely for this reason, as well as to divide the development into manageable pieces, each weather element is dealt with somewhat independently of the others. There are exceptions, however. For instance, ceiling height and total cloud amount equations are developed at the same time, as well as visibility and obstruction to vision equations.

Whenever the sample of data is sufficient, a separate equation is developed for each weather element, projection, and station; these are called "single station" equations. When the data sample will not support single station equations, data for several stations in a "climatologically homogeneous" region are pooled to develop "regionalized" equations. If data from all stations available are pooled, the result is called a "generalized operator" equation.

Usually, a larger data sample is required when a predictand is binary than when it is quasi-continuous and somewhat normally distributed. Therefore, a sample of a few hundred cases may be sufficient to support single station equations for temperature, but data from the same period of record usually produces better equations for probability of precipitation when several stations are grouped together. Also, a single station equation cannot be developed to make a forecast valid at a particular time of the day if the station is closed at that hour, because no predictand data would be available.

For regionalized equations, the regions are determined subjectively after considering topography, the climatic frequency of predictand events, and certain important predictor-predictand relationships. What is needed is an area in which the relationship(s) between the predictand(s) and the predictors in the equation(s) is (are) constant. Suppose equations for PoP are being developed. We compute the relative frequency of precipitation at each station given an LFM relative humidity forecast at that station of, say, greater than or equal to 75%. An analysis of these relative frequencies plays a major role in the determination of regional boundaries.

Many weather variables are forecast in terms of probabilities of categories. This is done primarily to deal with highly nonlinear predictor-predictand relationships and because one portion of the range of the predictand may be much more important operationally than other portions. These probabilities are then usually used to make a categorical forecast by a thresholding technique. In this technique, if the probability of a category exceeds a certain threshold, the forecast will be of that category. The thresholds are selected according to criteria which vary by weather element. For instance, for ceiling height, we strive for a high Heidke skill score and at the same time to make about as many forecasts of each category as is observed.

A reasonable set of possible predictors is subjectively chosen for each predictand and then objectively screened to determine a good set for inclusion in the equations. For instance, prediction equations for wind components and speed would usually contain, together with other variables, wind components and speed forecast by a numerical model at one or more levels. For the shorter range projections, observed wind components at the predictand station might also be important.

Some aspects of station climatology are automatically built into single station equations and, to a lesser extent, into regional equations. Climatic relative frequencies or predictand means, computed by month or season, can be explicitly included as predictors if desired. These are helpful for regional and especially for generalized operator equations.

Tables 1 and 2 summarize some of the characteristics of the guidance and its development for the CONUS and Alaska, respectively. Development for Alaska has been done recently, and except for max/min temperature, civilian and military stations have been dealt with in the same manner. For the CONUS, no data for military stations have been used in the development. Therefore, no single station equations exist for military stations, and all regional equations are based only on data from civilian stations. The following sections elaborate and comment on the table entries by weather element. Except as noted, all forecasts are based on the LFM model; no attempt is made to specify which forecasts are based partly on initial observations.

Temperature and Dew Point - The max/min temperatures are for a calendar day and do not always meet the need for a daytime max and a nightime min. To enhance consistency of the forecasts, equations for several 3-hourly (or 6-hourly for Alaska) temperatures and dew points and the max and/or min that usually occurs within the time span covered by those 3-hourly projections are all developed together so that all have the same predictors. For instance, one would like the temperature forecast for a particular time to be not less than the dew point forecast for that same time. Forcing both equations to contain the same predictors does not ensure consistency on any particular day, but the number of inconsistencies is considerably reduced.

There have been no comprehensive comparisons of the accuracy of temperature forecasts at CONUS civilian stations produced by single station equations versus forecasts at military stations produced by a generalized operator equation or by an equation for a nearby civilian station, but previous experiments suggest a generalized operator equation can be considerably less accurate than a single station equation for temperature forecasting.

Probability of Precipitation (PoP) - The probabilities of measurable precipitation in a 12-h period and in the two included 6-h periods are predicted by three regionalized equations in which all predictors are forced to be the same to enhance consistency. The regionalized equations can be used effectively for both civilian and military stations.

Quantitative Precipitation (QPF) - For the CONUS only, the probabilities of about four categories of precipitation amount (liquid equivalent) in 6-, 12- and 24-h periods are forecast by regional equations. Thresholding is used to determine a categorical forecast. The 12-h and the two included 6-h equations contain the same predictors.

Probability of Precipitation Type (PoPT) - For the CONUS, the conditional probabilities of precipitation type (liquid, frozen, or freezing), given precipitation occurs, are produced by regional equations for the months September through April. Although these regional equations can be used effectively for military stations, some of the predictors are tailored to the climate of individual stations. Since data were not available for military stations, the climate for these locations was estimated. Thresholding is used to produce a categorical forecast.

Probability of Frozen Precipitation (PoF) - For Alaska, the probability of frozen precipitation is provided by regional equations. Like for PoPT, some predictors for PoF are keyed to station climatic factors; these factors were determined directly from data for military as well as for civilian stations.

Probability of Snow Amount (PoSA) - For the CONUS, the probabilities of snow amount in three categories, both unconditional and conditional on snow occurring, and a categorical forecast are produced for the months October through March. Forecasts are provided for only the 12 to 24 h projection, and no forecasts are made for stations along the Pacific, Gulf, and Southern Atlantic coasts.

Type of Liquid Precipitation - For the CONUS, the probabilities of three categories of liquid precipitation--drizzle, rain, showers--are provided. The equations were developed to provide input to the computer worded forecast.

Probability of Thunderstorms and of Severe Weather - For the CONUS, the probability of thunderstorms and of severe weather are provided by regional equations. The predictands were actually defined by MDR data and severe storm reports and refer to square areas about 80 km on a side. Quite recent data (1981) were used in development of equations for the western U.S.; equations for the eastern U.S. were developed on data no later than 1978. The trajectory model furnishes valuable input to these forecasts. Forecasts will apply equally well to military and civilian stations.

Ceiling Height and Cloud Amount - The probabilities of each of six discrete categories of ceiling height (seven cumulative categories for Alaska) and of each of four categories of cloud amount (three cumulative categories for Alaska) are provided. By thresholding, categorical forecasts are determined. All equations for ceiling height and cloud amount for a particular region and projection contain the same predictors for consistency of forecasts.

Visibility and Obstructions to Vision - The probabilities of each of six discrete categories of visibility (four cumulative categories for Alaska) and of each of four categories of obstructions to vision (none, haze or smoke, blowing phenomena, and fog) are provided. For both visibility and obstructions to vision, a categorical forecast is provided by thresholding.

Surface Wind - For civilian stations, the U- and V-wind components and wind speed forecasts are produced by three, single station equations, all containing the same terms to enhance consistency. For stations of particular interest to the AWS, the wind components and speed are forecast by three generalized operator equations, all three containing the same predictors. The components are used to determine direction. The speed is forecast directly because combining the regression estimates of the two vector components will underestimate the speed. The speed forecasts are inflated, a process which spreads the forecasts away from the mean such that the variabilities of the forecasts and observations are about equal; this actually increases the mean square error but produces about as many strong wind forecasts as occur. Indications are the accuracy of forecasts from the generalized operator equations is considerably less than the accuracy of the single station equations.

D. Documentation

The documentation of the MOS system is contained in TDL Office Note 74-14 which is kept up to date as changes are made. The developmental software is documented in TDL Office Note 75-2 and is also kept up to date. Development of specific forecast equations is documented in NWS Technical Procedures Bulletins, as indicated in Tables 1 and 2, and usually, in more detail, in TDL Office Notes and published articles.

3. OPERATIONAL PROCEDURES

Forecast equations are implemented on NOAA's IBM 360/195 system. The preparations necessary for implementation, the archival and dissemination of forecasts, and documentation are described in the following sections.

A. Implementation of Forecast Equations

Equation Packing - All MOS equations are "packed" onto a group of data sets on the 360/195 disks. All data sets are available on-line when forecasts are being made. The elements and projections to make forecasts for on a particular run are specified in a control data set. Because of the enormous number of equations and the structure of NMC's files, we must specify the elements and projections in approximately the order the corresponding equations exist on the data sets for computer efficiency. Also for efficiency, the packed equations are ordered from small to large projections. Equation packing is strictly controlled and is handled by one person within TDL.

Test Runs - After the equations are packed, test runs are made in which forecasts are produced for a test case. These forecasts, made with operational software by the Computer Systems Section of TDL, are compared with forecasts made from the same equations, but in their original unpacked form, with developmental software. This quality control step assures, insofar as possible, correct implementation. Software is available to assist in this comparison, which is made by the developer.

Implementation Software - All MOS forecasts are made by one program. This program may be run more than once per cycle as needed. For instance, one run is made to produce CONUS forecasts and another to produce Alaskan forecasts.

Input includes not only the equation files and control data sets, but the station list, climatic constants from a file dubbed CONMAT, surface weather observations, and the LFM, trajectory, and spectral model grid point fields. Interpolation into the grid point fields and computation of specialized predictors is done as needed. Identical subroutines are used for computing predictors in the development and implementation steps to help assure correct implementation. Forecasts are written to a data set with the generic name MOSMAT. Most forecasts are also written to a station-oriented data set called SOFTRIX. All equation data sets are on line when the operational program is run; when equations are seasonal (e.g., one set for summer and one set for winter), the correct set to use is determined automatically from the current date.

Adding Stations - If guidance forecasts are needed for an additional station or location, several steps are necessary to provide that guidance. It is assumed additional development will not be done, since the time and effort required would not usually be available.

Step 1:
The station identification and location has to be added to the directories of the MOSMAT and CONMAT files. This must be done by TDL's Computer Systems Section.

Step 2:
If climatic values are needed by the forecast equations, they must be determined for the new location and added to the CONMAT file. The values could be determined from data for the location, but data may not be readily available, and estimation from nearby stations would be necessary. After the values are determined, they are added to the file by the TDL personnel.

Step 3:
The new station number must be associated with each pertinent equation on the packed equation data set. If only one forecast were needed, say the total cloud amount for one projection for one cycle, and backup equations were not to be used, this would mean inserting the station number in one place only, although the data set would have to be rewritten since it is treated as a sequential file. However, when a full set of forecasts is wanted, the station number must be inserted in many places. This is a time consuming task and is done by the one person who has control over the equation packing operation.

Step 4:
The dissemination software must be modified. This is usually fairly trivial and amounts to adding the station number to one or two control data sets. The software that produces the SOFTRIX data set and transmits a portion of it is flexible, and the addition of a station is not difficult.

Step 5:
The new data sets--MOSMAT, CONMAT, and packed equations--and software must be submitted to NMC who enters them into the run stream. This requires considerable coordination between TDL and NMC, which is carried out by the Computer Systems Section.

B. Archival of Forecasts

All forecasts are permanently archived, first to on-line disk, then periodically to tape. The process is carefully quality controlled. The tape format is documented, and documented software is available for reading the archive tapes. Software is also available to collate these operational forecasts with verifying observations and with experimental forecasts for verification studies. Some verification programs are available but are not sufficient for all needs.

C. Dissemination

Although there are several ways users can obtain MOS forecasts, the primary distribution to civilian stations is in FOUS12 bulletins. These bulletins are station oriented and contain in one place most of the MOS forecasts made for that station. However, the bulletins and facsimile products containing MOS forecasts are not generally available at DoD locations.

Forecasts for 255 civilian stations and 146 locations of interest to the Air Weather Service (AWS) are transmitted to the Air Force Global Weather Central (AFGWC) twice daily in a very efficient format over the high speed line between there and NMC. One transmission is made for CONUS stations and one for Alaskan stations. The software that produces the SOFTRIX file and transmits the data to AFGWC is quite flexible, but considerable study is required to modify it or its input when changes are made to the forecast product. Changes can be made to add a new forecast or may be necessary when a change is made to existing forecasts. For instance, if the number of categories for which ceiling height probability forecasts are made were increased, changes to the dissemination software would be needed.

Forecasts for 46 locations of interest to the U.S. Navy are prepared in FOUS12 format and transmitted to Carswell AFB, where they are made available to Navy stations on by request-reply (ARQ). The software that produces these bulletins is nearly identical to that used for civilian stations and is not only easy to maintain, but several persons in TDL are familiar with its structure.

D. Documentation

Documentation for operational programs is maintained by the Computer Systems Section and generally is needed and used only by them. In some cases, this documentation has been prepared by other persons. For instance, the AWS liaison group has provided the documentation on the software that provides for dissemination to AFGWC.

Table 1. Characteristics of MOS forecasts and equation development for the CONUS. Information about civilian (CIV) and Air Force (AF) stations pertains to the SOFTRIX message sent to AFGWC. Information about Navy stations pertains to the message sent to Carswell AFB. Some forecasts may be missing for some elements and projections. SS stands for single station equations, R for regionalized, and GO for generalized operator; for most elements, regionalized or generalized operator equations are applied to all stations and are listed only once. The numbers of regions used, which may vary by season or projection, is indicated in parentheses after the R. The predictand data collections used are either Asheville (A) or MDR (R). No Technical Procedures Bulletin (TPB) exists for type of liquid precip; however, TDL Office Note No. 79-8 is pertinent. Altimeter setting forecasts are also provided but are direct transformations from the LFM model rather than being MOS forecasts.

Weather Element	No. Stations Guidance Provided For			Maximum Projection (h)			Projection Increment (h)			Type of Equation			Predictand Data Collection Used In Development (Last Date Used)	Seasonal Stratification	Relevant TPB No.
	CIV	AF	NAVY	CIV	AF	NAVY	CIV	AF	NAVY	CIV	AF	NAVY			
Surface Temperature	228	134	46	51	60	51	3	6	3	SS	GO	GO	A (2/80)	4	285
Surface Dew Point	228	134	46	51	60	51	3	6	3	SS	GO	GO	A (2/80)	4	285
Max/min Temperature	228	0	0	60	--	--	12	--	--	SS	--	--	A (2/80)	4	285
6-h PoP	228	134	46	48	48	48	6	6	6		R(26-27)		A (9/80)	2	289,299
12-h PoP	228	134	46	60	60	48	12	12	12		R(26-27)		A (9/80)	2	289,299
6-h QPF	228	134	46	42	42	36	6	6	6		R(9)		A (3/80)	2	283
12-h QPF	228	134	46	48	48	48	12	12	12		R(9)		A (3/80)	2	283
24-h QPF	228	134	0	60	60	--	24	24	--		R(9)		A (3/80)	2	227
PoPT	228	134	46	60	60	48	6	6	6		R(5)		A (3/81)	1	319
12-h PoSA	200	113	25	24	24	24	--	--	--		R(5)		A (3/81)	1	318
Type of liquid Precip.	228	134	0	54	54	--	6	6	--		GO		A (3/78)	2	--
Prob. of Thunderstorms	228	134	46	48	48	48	12	12	12		R(2)		R(9/78;9/81)	2	331
Prob. of Severe Weather	228	134	0	48	48	--	12	12	--		R(2)		R(9/78;9/81)	3	--
Ceiling Height/ Cloud Amt.	228	134	46	60	60	48	6	6	6		R(15-20)		A (3/80)	2	303
Visibility	228	134	46	60	60	48	6	6	6		R(15-20)		A (3/80)	2	303
Obstructions to Vision	228	134	46	48	48	48	6	6	6		R(15-20)		A (3/80)	2	303
Surface Wind	228	134	0	60	60	--	6	6	--	SS	GO	--	A (9/79)	2	316

Table 2. Generally the same information for Alaskan stations as Table 1 gives for CONUS stations. H stands for hourly data collection. TPB No. 329 does not discuss ceiling height.

Weather Element	No. Stations Guidance Provided For CIV	AF	Maximum Projection (h)	Projection Increment (h)	Type of Equation	Predictand Data Collection Used In Development (Last Year)	Seasonal Stratification	Relevant TPB NO.
Surface Temperature	26	12	54	6	SS	H (1/82)	4	329
Max/min Temperature	14	--	60	12	SS	A (1/82)	4	329
6-h PoP	27	12	54	6	R(5-6)	H (7/82)	4	329
12-h PoP	27	12	54	12	R(5-6)	H (7/82)	4	329
PoF	27	12	54	6	R(3)	H (7/82)	2	329
Ceiling Hgt/Cloud Amt	27	12	60	6	R(6-7)	H (7/82)	4	329
Visibility	27	12	60	6	R(6-7)	H (7/82)	4	--
Obstruction to Vision	27	12	48	6	R(6-7)	H (7/82)	4	--
Surface Wind	27	12	60	6	SS	H (1/82)	4	329