The Canadian Updateable Model Output Statistics (UMOS) System: Design and Development Tests

LAURENCE J. WILSON AND MARCEL VALLÉE

Meteorological Service of Canada, Dorval, Quebec, Canada

(Manuscript received 19 April 2001, in final form 12 October 2001)

ABSTRACT

The use of model output statistics (MOS) in operational weather element prediction has been hindered since the mid-1980s by frequent changes in the operational numerical weather prediction models that supply the predictors for the weather element forecasts. Once the model changes, a new archive of model output must be collected for a long enough period that statistically stable equations can be developed. This paper describes a new statistical interpretation system that addresses this problem and permits the rapid adaptation of the statistical forecast to changes in the formulation of the driving model. In comparison with traditional MOS development, the new system incorporates two main features. First, the data are stored in the form of the cross-products matrices used in multivariate statistical techniques rather than as raw observations and forecasts. It is these matrices that are updated regularly with new output from the model. Second, equations are developed by a weighted blending of the new and old model data, with weights chosen to emphasize the new model data while including enough old model data in the development to ensure stable equations and a smooth transition to dependency on the new model. This paper describes the design of the new system and shows tests of the equation development method following a major change of the Canadian operational model. Tests were carried out for surface temperature, probability of precipitation, and wind direction and speed for about 200 Canadian stations that have a reliable observation record. For all three elements, the coefficients and predictors selected remained remarkably stable through the transition from dependence on old model data to new model data. Although some degradation of the goodness of fit was noticed during the period when new and old model forecasts were blended, especially for wind, these effects were minor, which means that useful MOS equations could be obtained relatively soon after a change of model. Results from a comparison of forecasts from the new system with operational perfect prog forecasts and direct model output forecasts are the subject of a second paper.

1. Introduction

Model output statistics (MOS; Glahn and Lowry 1972) has been the preferred formulation for statistical interpretation of model output, mainly because it accounts for persistent model biases by using model output variables as predictors. MOS equations may also incorporate model-derived variables that are not observed, for example, vertical velocity, stress, and fluxes. Furthermore, MOS equations can be used to evaluate the performance of the model variables, since they relate these variables to observations. MOS had been shown to provide weather element guidance forecasts that are competitive with local subjective forecasts as early as the late 1970s, using the models of that time (e.g., Zurndorfer et al. 1979). Thus, during the 1980s, many weather services developed and ran operational MOS systems (Lemcke and Kruizinga 1988; Carter et al. 1989; Wilson et al. 1986; Yacowar et al. 1985).

Corresponding author address: Laurence J. Wilson, Recherche en Prévision Numérique, Meteorological Service of Canada, 2121 Route Transcanadienne, Suite 500, Dorval, QC H9P 1J3, Canada.
E-mail: lawrence.wilson@ec.gc.ca

Two practical drawbacks of MOS led to a reduction in its use in operations in the 1990s. First, any significant change in the driving model implies a change in the statistical characteristics of the model output variables. Such changes not only affect the bias but also the variance of the model variables, their correlation structure, and, most important, the covariance structure with respect to observations. Inevitably, MOS equations must be rederived using data that are representative of the new model output. Experiments have shown that two seasons of homogeneous data (about 300 cases or so) are required for stable statistical relationships to be developed (Carter 1986), while bootstrap experiments with surface wind data showed that more than 200 cases would be required to control overfitting of the development sample (Wilson 1985). Using traditional MOS development methods, one would then have to wait at least two years after a model change to obtain statistically stable MOS equations using variables from the new model.

Second, in an operational setting where forecasts are required for many locations and many projection times, the development cost of a MOS system can be quite high because variation in the model bias characteristics with forecast projection requires that separate equations be developed for each different projection. This means that a single MOS system can consist of many thousands of equations, which increases the maintenance overhead of a MOS system in an environment of frequent model changes. By the late 1980s, in Canada and in other countries, significant model changes had become too frequent and the development cost too great to maintain a statistically stable MOS system.

To address this problem, one alternative is to revert to the older perfect prog method (PPM; Klein and Lewis 1970). Perfect prog equations, which use analyzed or observed data for equation development, do not require redevelopment after a model change, and considerably fewer equations are needed to provide a complete set of forecasts for each weather element. However, by substitution of model estimates of the predictor values to run the equations in forecast mode, the PPM makes a perfect model assumption; that is, the model forecast predictor values can be treated as if they were analyzed or observed values, as were used in equation development. Perfect prog equations do not take into account model biases, nor can they have access to model-derived variables that are not observed.

After the introduction of the Nested Grid Model (NGM), the National Oceanic and Atmospheric Administration's (NOAA) Techniques Development Laboratory reevaluated PPM forecasts, concluding that PPM NGM-based short-range temperature forecasts were inferior to MOS-based forecasts on the Limited Fine Mesh (LFM) model, but that the medium range (3–6 day) PPM forecasts were superior to MOS forecasts run on a different model from the one on which they were developed (Dallavalle 1988). Pending the accumulation of a stable NGM development dataset for MOS, a perfect prog system was implemented (Carter et al. 1989). In Canada parallel studies indicated roughly equivalent skill between MOS and perfect prog forecasts (Brunet et al. 1988), and all the operational MOS forecast products were replaced by perfect prog products, which have been running operationally throughout the 1990s.

A second option is to use PPM forecasts but to adapt or tune them to the characteristics of the new model in some way. Dallavalle (1988) used LFM initial fields to develop modified perfect prog equations for the NGM model, but biases in the NGM temperatures were not accounted for by this method. Sarrazin and Wilson (1989) used a 3-yr independent sample of model output to tune PPM wind forecasts. The resulting forecasts were of good quality, but tuning makes the forecasts model dependent as MOS equations, and the equations would be expected to need retuning after a major model change. The PPM wind forecasts also have run operationally throughout the 1990s, but without retuning, even though there have been several model changes.

A third option for statistical interpretation in an environment of frequent model changes is to use some sort of recursive updating method, the purpose of which is to ingest information about the characteristics of the new model variables into the forecast equations as early as possible, frequently updating the estimates of the coefficients in light of new data. Two methods of this type that have been applied to statistical weather element forecasting are the Kalman filter (Simonsen 1991) and Updateable MOS (UMOS; Ross 1987, 1989). After considerable evaluation of both approaches (Vallée et al. 1996), we opted to follow the UMOS route to develop a new statistical guidance package for use in Canada. The two methods are similar in that they both involve frequent updating of the coefficients of an empirical equation. In the Kalman filter method, the estimates of the coefficients are updated directly via a set of recursive equations, while in UMOS, it is the covariance (or correlation) matrix that is updated, then the statistical procedure, multiple linear regression (MLR) or multiple discriminant analysis (MDA), is rerun using the updated matrices. The Kalman filter is simple to implement and can be very responsive following a model change, especially for removing model bias. However, we found it unsuitable for use in an operational setting involving many equations, many projections, and many stations because the need to tune the parameters of the filter to optimize for each application leads to a large development and maintenance cost. The filter also becomes cumbersome to use with more than a few predictor variables.

As model resolution increased during the 1990s, as compared with surface and upper-air observation networks, we were encouraged to renew our efforts to mount a new MOS system using the higher-resolution model variables, even in an environment of frequent model changes. If accurate short-range forecasts (out to 2 days) could be made at a resolution much higher than the resolution of the upper-air observations or analyses used in perfect prog equation development, then this becomes an additional advantage of MOS as compared with PPM, provided there is useful predictive information in the smaller scales that the newer models are capable of simulating.

This paper describes the design and development of the new Canadian UMOS system. So far, we are processing five predictands: 3-h spot temperatures, 6-h probability of precipitation (POP), and 3-h wind direction and speed (three predictands—west and south vector components, and the scalar speed). These are treated using MLR, which is a separate component of the system. At present an MDA module is under development for use with multicategory predictands such as precipitation type and cloud amount.

The UMOS system is described in section 2, an evaluation of its performance in update mode is described in section 3, and in section 4 our plans for further development of the system are discussed.

2. System design

In a standard MOS development system, for example Glahn and Dallavalle (2000), a large archive of model output and observations is accumulated. Once there are enough data from a particular model, usually about two years, equations are developed. In such a system, the data management and statistical development portions of the system are kept separate. While data from several stations may be pooled to form the development sample for a single equation, normally each event of the development sample contributes equal weight to the statistical processing. The design of a UMOS system differs from this in two main aspects: First, the data management and statistical development portions of the system are not so clearly separable, in the sense that part of the data preparation for the statistical development is done in near–real time, and variables may be archived in the form needed for the statistical analysis. This makes it easier to update the equations, but it also means that some of the decisions about how the predictors are used must be made before starting the archiving. Second, an updating scheme is designed to allow control over the relative weights of the events in the development sample. For example, it must be possible to assign higher weight to data from a current model version as compared with an older model version, at least until the sample from the new model is large enough to produce stable equations. The design of our updating scheme is described in the next section, followed by a description of the overall structure of the system.

a. Updating concept

As described by Ross (1992), a UMOS system is intended to facilitate the rapid and frequent updating of a large number of MOS equations from a linear statistical model, either MLR or MDA. Both of these techniques use the sums-of-squares-and-cross-products matrix (SSCP), or components of it. The idea of the updating is to do part of the necessary recalculation of coefficients in near–real time by updating the SSCP matrix, and storing the data in that form rather than as raw observations.

Solving the multivariate regression model,

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon, \quad (1)$$

using least squares leads to the matrix form of solution for the coefficients,

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad (2)$$

where $\mathbf{b}$ is a vector of estimates of the coefficients, of length $P + 1$, where P is the number of predictors (X_j, $j = 1, \ldots, P$), and ($\mathbf{X}'\mathbf{X}$) is the SSCP matrix. The (j, k) element of the matrix is

$$\sum_{i=1}^{N} X_{ji} X_{ik} \qquad i = 1, \ldots, N$$

for a sample of size N. For compactness of computation, the SSCP matrix is augmented with the vector $\mathbf{X}'\mathbf{Y}$, the $P + 1$ dimension vector of cross products with the predictand Y. The first row and column of the SSCP matrix contain the current sample size on the diagonal and the sums of the P variables off the diagonal. Since the SSCP matrix is symmetric, a $P + 1$ by $P + 1$ matrix can be stored as $(P + 1)(P + 2)/2$ values, which greatly cuts down the storage requirements of UMOS as compared with to storing the original data.

We use a weekly update cycle; all the SSCP matrices are updated daily, and the equations are recomputed each week. To update the matrices, the elements of the matrix are incremented by the latest set of squares and cross products over all predictors and the predictand. Since we use separate equations for 0000 and 1200 UTC model runs, we accumulate one new realization per day for each station, for each model projection, for each predictand element. The advantage of a weekly equation update as compared with the monthly cycle of Ross (1992) is that the data are added in small enough increments that changes to the equation coefficients should be kept small, except perhaps near the beginning of a new cycle after a model change, when a week of data represents a more significant portion of the available knowledge about the model characteristics. When a model changes, we start accumulating another set of SSCP matrices for the new model. These are combined with those from the old model in a weighting system described below.

Another way in which our system differs from that of Ross (1992) is in the treatment of the annual climatological signal in the data. Ross (1992) uses a recursive least squares time series model to estimate the mean and variance of the observations for the next month. We treat the annual cycle by a combination of three methods. First, we stratify the data into two seasons, warm (23 April–6 November) and cold (7 November–22 April); second, we offer the sun angle as a variable; and third, we use a weighted blending of cold- and warm-season equations during spring and autumn transition periods, 1 April–14 May and 16 October–28 November.

The weighted blending of SSCP matrices was designed to ensure smooth transition both between seasons and following a major model change. The choice of parameters of the weighting scheme is based on several principles:

1) It is not feasible to begin development of new model equations until at least 30 days of data are available.
2) The effective sample size needed for stable equations is at least 300 (Carter 1986; Wilson 1985).
3) The effect of the old model on the equations should be phased out as the sample size from the new model approaches 300.
4) Seasonal weights should adjust for differences in sample sizes for the two seasons.

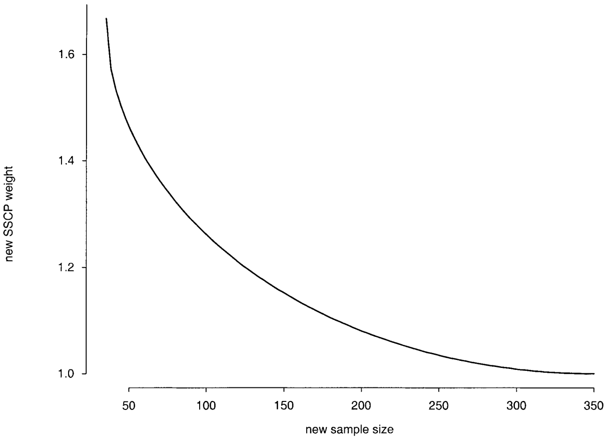


FIG. 1. The weights applied to dependent samples from a new model, as a function of the sample size, for POP equation development, following Eq. (3).

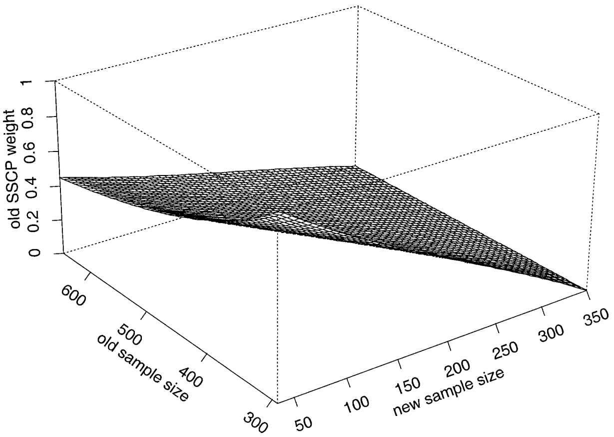


FIG. 2. The weights applied to development samples available from an old model, as a function of the sample size from the new and old models, for POP, following Eq. (4).

5) Data from a new model should be emphasized so that the equations respond to new model characteristics as quickly as possible.

The parameters of the weighting scheme are somewhat arbitrary and were adjusted experimentally following these general principles. The weights for the new (ω_1) and old models (ω_2) take the form

$$\omega_1 = \omega_{\max} + (1 - \omega_{\max})\sqrt{1 - \frac{(N_1 - N_{\max})^2}{(N_{\min} - N_{\max})^2}}, \quad (3)$$

$$\omega_2 = \frac{N_{\max} - \omega_1 N_1}{N_2}, \quad \text{for } N_1 \leq N_{\max}; \quad (4)$$

$$\omega_1 = 1, \quad \omega_2 = 0, \quad \text{for } N_1 > N_{\max}; \quad \text{and}$$

$$\text{SSCP} = \omega_1 \text{SSCP}_1 + \omega_2 \text{SSCP}_2,$$

where N_1 and N_2 are the sample sizes from the new and old models, respectively; $N_{\min}$ is the minimum permitted sample size from the new model for updating; $N_{\max}$ is the largest sample from the new model for which blending with old model data is considered necessary; and $\omega_{\max}$ is the largest weight applied to new model data. The parameters $N_{\min}$, $N_{\max}$, and $\omega_{\max}$ have been set according to the principles listed above and were adjusted for different elements based on development experience. For temperature and wind, $N_{\min} = 30$, while for POP, $N_{\min} = 35$. The value of $N_{\max}$ is presently set to 300 for temperature, 350 for POP, and 325 for wind. Experience shows POP is a more difficult predictand to fit statistically, and it takes longer for the equations to stabilize. The parameter $\omega_{\max}$ functions a little like the response parameter in the Kalman filter; it must be set high enough to cause a noticeable response to the new model data, but not so high that the equations are "misled" by the possible unrepresentativeness of a small sample from the new model. At present, $\omega_{\max}$ is set to 1.667, which makes the minimum sample of 30 look to the system as a sample of size 50. Figures 1 and 2 show the relationships between sample size and the weights for the new and old model, respectively. The figures use the threshold values for POP; corresponding figures for temperature and wind would be similar.

Seasonal weights ω_s are applied during spring and autumn transition periods of about 6 weeks in length, as mentioned above. The weights change stepwise through the period,

$$\text{SSCP} = \sum_{s=a}^{b} \omega_s \frac{\min(N_a, N_b)}{N_s}(\omega_{1s}\text{SSCP}_{1s} + \omega_{2s}\text{SSCP}_{2s}), \quad (5)$$

where subscripts a and b refer to the two seasons; ω_{1s} and ω_{2s} are the weights for the new and old model data, respectively, for season s; and ω_s is the seasonal weight for season s, which is currently a three-step function with values of 1/3 and 2/3 for seasons a and b for the first 2 weeks, ½ and ½ for the second 2 weeks, and 2/3 and 1/3 for the last 2 weeks of the 6-week changeover period. The role of the minimum function is to ensure that the resulting blended SSCP matrix is not dominated by the season with the higher sample size.

The weighting scheme described up to this point is intended for single station equation development. However, because of incomplete observation data and irregular observation schedules, we have sometimes found it necessary to blend data from one station with data from nearby stations that are climatologically similar. While this supports the equation development for less reliable observation locations, it adds to the variability in the predictand. To account for this, we adjusted $N_{\min}$ and $N_{\max}$ as a function of the number of stations that are grouped and the total available sample size. The grouped station value N_{gmin} is simply the number of stations times the single station value, while the adjustment for $N_{\max}$ recognizes that more that 300 events

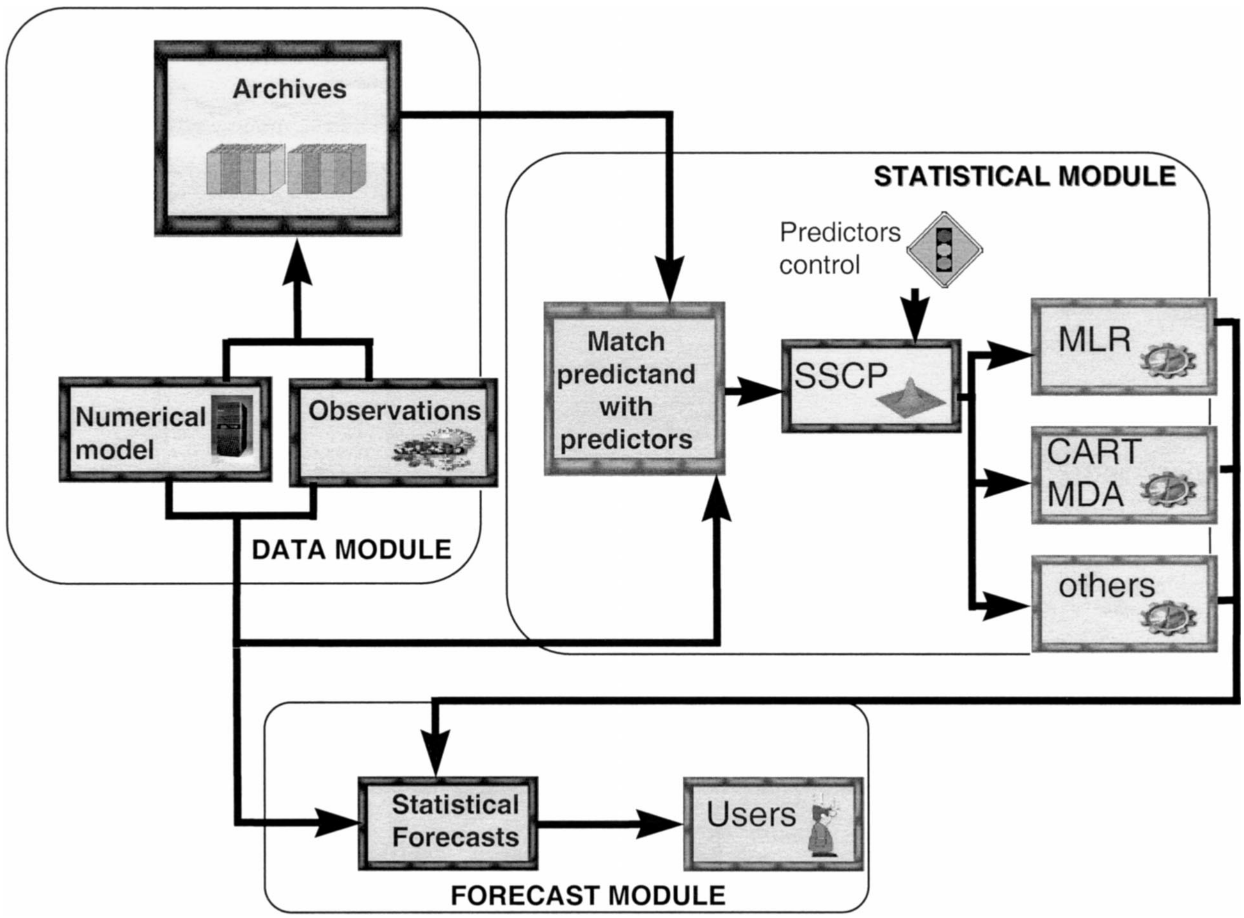


FIG. 3. A schematic diagram of the components of the UMOS development system and their relationships.

are needed to stabilize grouped station equations, but fewer than 300 events from every station. The formula is intended for groups of 2–20 stations:

$$N_{g\,\max} = N_{\max} + \sqrt{\left(\frac{19N_{\max}}{3}\right)^2\left[1 - \left(\frac{\mathrm{nstn} - 20}{19}\right)^2\right]}, \quad (6)$$

where nstn is the number of stations in the group. The constants in Eq. (6) were selected assuming that the maximum sample size needed for stable group equations based on 20 stations would be about 6 times the maximum needed for each station individually. For intermediate size groups, the response is a quadratic function of the number of stations in the group.

b. System components

Figure 3 shows the general structure of our UMOS system. The system consists of three main components: a data module, a statistical module, and a forecast module. In the data module, the main archive of forecasts and observations is created and updated. We archive a total of 177 variables from the model, all interpolated to station locations, in addition to the complete observation from each station. The full set of predictors, designed to be used for all surface weather element predictands, includes the following:

1) model diagnostic predictors—sea level and surface pressure, sea and soil temperature, total clouds, convective and total precipitation rates, convective and total precipitation over the last 6 h, orographic vertical motion, and Laplacian of pressure tendency;
2) predictors at standard levels (surface, 1000, 925, 850, 700, 500 hPa)—geopotential height, temperature, dewpoint depression, east and south wind components, vorticity advection, Laplacian of temperature advection, dewpoint temperature, wind speed, and vertical velocity (except surface); and
3) derived predictors—horizontal gradients of height, temperature, dewpoint temperature, and wind speed components; geostrophic wind speed and components; divergence components; thicknesses and temperature advection components; dewpoint temperature advection components; low-level wind shear; and lapse rates, relative humidity at six levels, and George *K* index.

Predictors are not smoothed before interpolation. However, the grid length for horizontal differencing is held constant at about 50 km for different model versions.

There are at present about 500 Canadian stations in the archive; model variables are also interpolated to an additional 200 locations that are not station locations, to provide a more complete spatial coverage. Storage of data in station-specific form reduces the size of the archive, but it means that all spatially derived predictors such as gradients, advection, and so on must be computed at the time the archive is updated. We cannot add a spatially derived predictor at a later time. The archive is automatically updated each day with the latest data. Observations and model output are kept separate in the archive. When a new station is added to the archive, SSCP matrices are started for that station from the time of its addition. While awaiting collection of sufficient data for equation development for a new station, interim equations are produced from the SSCP matrices of surrounding stations as a group.

Data for the statistical module may come from the archive if there is a need to catch up an update cycle, or it may come directly from the operational model output files and the observation files for the normal weekly update cycle. The statistical module runs approximately once a week and includes the matching of predictor and observation (predictand) data, the creation and updating of the SSCP matrices described above, and the statistical equation updating. To keep the storage requirements for the SSCP matrices to a manageable level, we prescreened the original set of 177 predictors to determine which were most often selected, using a subset of about 200 stations and a sufficiently large sample of old model data. The prescreening has and will be done only once for each predictand. So far, we have prescreened for temperature, POP, and wind, keeping 18, 35, and 33 of the original 177 predictors, respectively. To these were added persistence predictors, one for temperature and each wind component, and two for precipitation. The retained predictors for temperature are

predicted total cloud cover;
temperature at the surface, 1000, 925, and 850 hPa;
dewpoint depression at 1000 hPa;
geopotential height at 850 and 700 hPa;
850-hPa wind speed;
1000-hPa temperature advection;
relative humidity at 925, 950, 700, and 500 hPa;
thickness (1000–850) and (1000–925) hPa;
George *K* index;
sun elevation function (incorporates both diurnal and seasonal insolation); and
latest observation (at analysis time).

The retained predictors for POP are

cloud cover;
convective and total precipitation rates;
convective and total 6-h quantitative precipitation forecast;
Laplacian of pressure tendency;
850-hPa E–W temperature gradient;
700-hPa dewpoint depression;
N–S shear of zonal wind at 925, 850, and 700 hPa;
E–W shear of meridional wind at 925 and 700 hPa;
N–S shear of meridional wind at the surface and 700 hPa;
Laplacian of temperature advection at 925 and 500 hPa;
vertical velocity at 1000, 925, 850, 700, and 500 hPa;
geopotential height at 850 and 700 hPa;
N–S height gradient at 1000 hPa;
geostrophic vorticity at 1000 and 500 hPa;
surface wind speed;
divergence at 500 hPa;
relative humidity at 925, 850, 700, and 500 hPa;
vertical shear of meridional wind, (850–1000) hPa;
George *K* index;
observed precipitation at analysis time (binary); and
observed precipitation in last 6 h (binary).

The same predictors are offered for the three wind components, except that only the persistence predictor that corresponds to the predictand is offered. The retained predictors for wind are

Laplacian of pressure tendency;
E–W temperature gradient at 1000 hPa;
E–W dewpoint depression gradient at 925 hPa;
N–S dewpoint depression gradient at 1000 hPa;
west wind component at the surface, 1000, 925, 850, and 700 hPa;
south wind component at 1000, 925, 850, and 700 hPa;
N–S shear of west wind at 700 hPa;
E–W shear of south wind at 925 hPa;
vorticity advection at the surface and 925 hPa;
Laplacian of temperature advection at 925 hPa;
geopotential height at 850 hPa;
geostrophic vorticity at 1000 and 925 hPa;
orographic wind speed;
wind speed at the surface, 1000, 925, 850, and 700 hPa;
temperature advection at 1000 hPa;
geostrophic west wind component at 850 and 700 hPa;
geostrophic wind speed at 1000, 925, and 850 hPa; and
observed wind speed, west and south components (maximum over 3 h centered on analysis time).

The SSCP matrices are stored for each station, for each forecast projection (3 h for temperature and wind, 6 h for POP), for each model run, and for each predictand. When the equations are rerun each week, predictors are screened using the forward stepwise procedure (Draper and Smith 1998). The same predictor set is

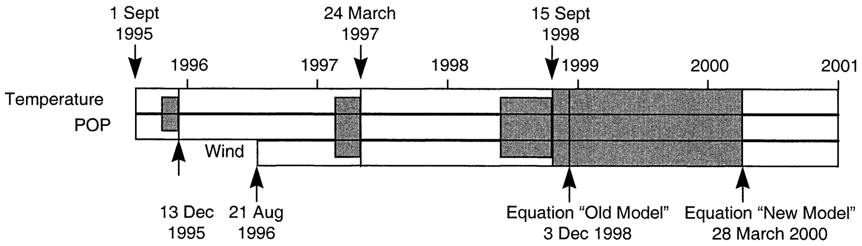


FIG. 4. Structure of the archives available for the UMOS development tests described in this paper. Vertical lines with dates indicate model changes; the new model data period is in black and the old model data period is shaded gray. Half-height shaded boxes indicate parallel run periods prior to each new model implementation.

offered for all projections, except that the persistence predictors are offered only for the first 24 h. Equations without the persistence predictors are also developed as backup for use when the persistence observation is missing. Predictors can also be forced into the equations or excluded from the equations if desired.

The statistical procedures (MLR and MDA so far) are kept separate from the rest of the system so that we can ''plug in'' different processing algorithms as required for different predictands. We can consider using only those statistical techniques for which partial data preparation and reduction can be carried out in real time. Any technique that must start from the original data matrix cannot be used in the system at present because the data handling requirements are too large for frequent updates.

The forecast module is used to produce the operational forecasts. Forecasts are produced for probability of precipitation greater than or equal to 0.2 mm, valid for 6-h periods, 0–48 h (8 projections); 3-h spot temperature forecasts, 0–48 h (17 projections); and 3-h wind speed and direction (0–48 h, 17 projections). Wind direction is derived from forecast equations for the west (U) and south (V) components. Following the development procedures of the U.S. MOS system (Glahn 1970), separate equations are needed for wind speed because computing speed from the equations for U and V does not lead to a least squares estimate of the speed. Forecasts are run twice per day using output from our regional model. All forecasts are archived for verification purposes.

3. Development results

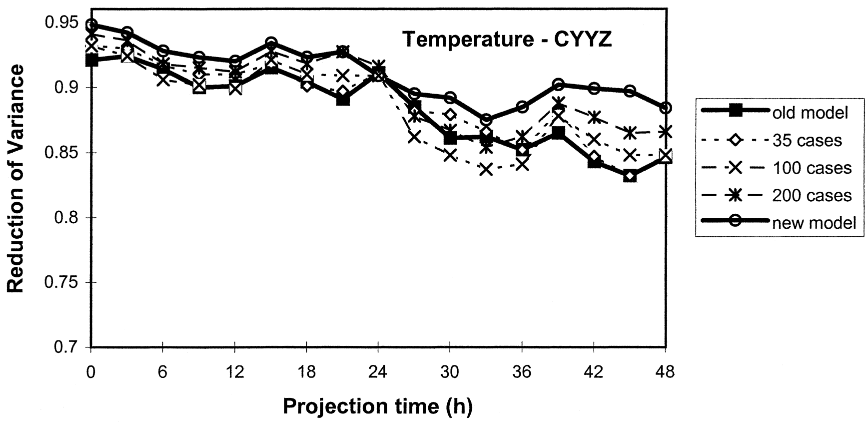
The Canadian regional operational Global Environmental Multiscale (GEM) model underwent a major change in 1998. The changes and their impact on precipitation forecasts from the model are discussed in detail in Bélair et al. (2000). The most significant changes were

1) an increase in the horizontal resolution from 0.33° latitude (about 37 km) to 0.22° latitude (about 24 km),
2) replacement of the Kuo convective precipitation scheme with the Fritsch–Chappell scheme,
3) an increase in the frequency of the radiation balance calculations from every 135 min to 60 min, and
4) an increase in the resolution of the geophysical fields (topography and roughness) to the high-resolution (1 km) U.S. Geological Survey database.

Results of tests indicated that the change in convection scheme had a positive impact in summer on the bias and threat scores, but a neutral impact in winter. The evaluation was limited to precipitation; the impact of the other changes on the precipitation forecasts was considered to be small.

This change of model, on 15 September 1998, gave us a good opportunity to test the ability of the UMOS system to cope with a major model change. In terms of the UMOS cycle, the model change came near the end of the summer season just before the blended-season period (16 October start). Thirty days after the implementation of the new model in this case coincides with the start of the autumn blending period. Thus, beginning 30 days after the implementation, the first equations to use any data from the new model would be made up of a blend of old model winter season data, a small sample of new model summer data, and old model summer season data. As the transition season progressed, the impact of the new summer season data would be reduced by the seasonal weighting, and we would have to wait until 30 days after the winter season began (about 6 December) before we could expect the new model to have a seasonally representative impact on the forecasts.

This problem can be alleviated under some conditions by archiving and using data from the parallel run period, that period of time before a new model implementation when new and old models are both run each day. Data from the parallel run can only be used if the new model remains unchanged during the parallel run. For the 15 September 1998 model change, we were able to use data from the 4-month period of the parallel run, but this had an impact only on the summer equations in this case, which were phased out during the autumn season transition. This is a consequence of our design: If the model changes near the end of one of the seasons, we



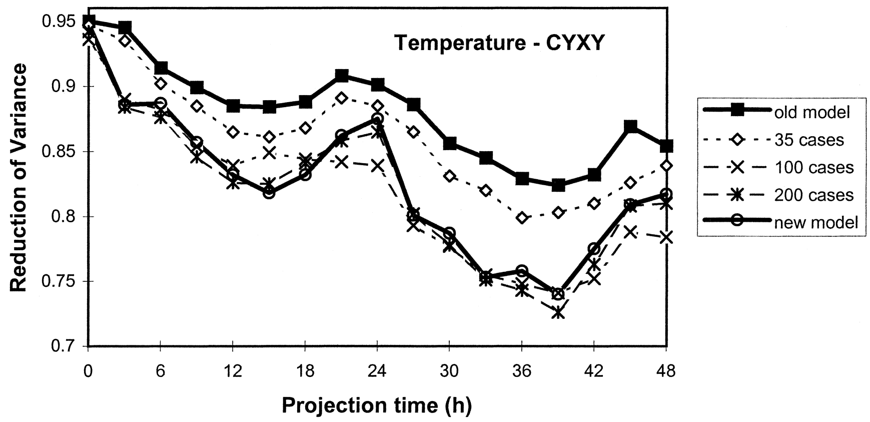


FIG. 5. Dependent sample variance explained as a function of projection time for temperature forecasts for (top) CYYZ and (bottom) CYXY, for different blends of new and old model data.

cannot expect the new model data to have a significant impact on the forecasts for perhaps 3 months after the implementation. By contrast, model changes near the beginning of either season mean that the new model data can noticeably affect the forecasts 30–35 days after the model implementation. In fact, if a model implementation occurs toward the middle of either season, and it is possible to use data from the parallel run period, it would be possible to implement MOS equations that use new model data at the same time as the model itself is implemented. In summary, even an unfavorable implementation date with respect to the season change leads to at most a 3-month wait for new model data to be used in UMOS equations, which represents a far superior response when compared with the 2 yr or more delay before new model output has an impact on standard MOS equations.

Figure 4 shows the archive period available for UMOS development, with respect to model changes that have taken place. Although there have been several model changes since 1995 when the archive started, only the latest is considered by the weighting scheme; once the model changes, all previous model versions are considered together as ''old model data.'' Since the SSCP matrices are stored separately for ''new'' and each version of the ''old'' model, it is always possible to simply drop the data from the oldest version of the model following the next model change. The decision to do this depends on the frequency of model changes; we must retain enough old model data to keep the equations stable during the transition period.

To illustrate the performance of the UMOS system, we conducted an update test on the winter equations, following the last major model change. Since two full seasons have passed since the model change, we have been able to test the system through a full transition from complete dependence on old model data to (nearly) complete dependence on new model data. To limit the test to pure winter data and equations, we waited until after the end of the autumn 1998 season blending period to begin the test. The first run was carried out on 3 December 1998, before enough winter data had accumulated from the new model. This run produced the last set of winter equations based solely on old model data.

TABLE 1. Results of UMOS POP equation development for update cycle during the model transition period from Dec 1998 to Mar 2000, for CYYZ, 6-h forecasts. Here RV is the reduction of variance on the development sample. After step 6, each second step is shown, except for step 17, the first step after seasonal transition. The predictand is expressed in percent (0%–100%).

Step	Run date	Constant	Coefficient	Predictor	Coefficient	Predictor	Coefficient	Predictor	Sample new	Sample old	Weight new	Weight old	RV
1	3 Dec 1998	2.271	0.301	Model pcpn	0.336	Obs pcpn	11.377	Rain rate	0	546	0.000	1.000	0.492
2	13 Dec 1998	1.669	0.380	Model pcpn	0.371	Obs pcpn			35	546	1.666	0.534	0.467
3	17 Dec 1998	1.627	0.379	Model pcpn	0.372	Obs pcpn			39	546	1.560	0.530	0.469
4	28 Dec 1998	1.570	0.380	Model pcpn	0.372	Obs pcpn			50	546	1.463	0.507	0.461
5	3 Jan 1999	1.584	0.393	Model pcpn	0.366	Obs pcpn			56	546	1.427	0.495	0.469
6	13 Jan 1999	1.461	0.345	Model pcpn	0.334	Obs pcpn	9.315	Rain rate	66	546	1.378	0.474	0.515
8	28 Jan 1999	1.430	0.345	Model pcpn	0.329	Obs pcpn	9.780	Rain rate	81	546	1.319	0.445	0.524
10	13 Feb 1999	1.396	0.361	Model pcpn	0.306	Obs pcpn	9.950	Rain rate	97	546	1.269	0.416	0.523
12	28 Feb 1999	1.664	0.358	Model pcpn	0.303	Obs pcpn	10.089	Rain rate	112	546	1.230	0.389	0.516
14	13 Mar 1999	1.372	0.345	Model pcpn	0.324	Obs pcpn	10.142	Rain rate	125	546	1.200	0.366	0.522
16	28 Mar 1999	1.248	0.347	Model pcpn	0.324	Obs pcpn	10.111	Rain rate	140	546	1.170	0.341	0.529
17	3 Dec 1999	−10.308	0.781	Model pcpn	0.354	Obs pcpn	0.395	RH 700	192	546	1.090	0.258	0.518
18	13 Dec 1999	−9.766	0.794	Model pcpn	0.368	Obs pcpn	0.375	RH 700	202	546	1.078	0.242	0.521
20	28 Dec 1999	−9.382	0.786	Model pcpn	0.385	Obs pcpn	0.362	RH 700	217	546	1.062	0.219	0.537
22	13 Jan 2000	0.895	0.375	Model pcpn	0.388	Obs pcpn			234	546	1.047	0.192	0.491
24	28 Jan 2000	0.804	0.382	Model pcpn	0.393	Obs pcpn			249	546	1.035	0.169	0.505
26	13 Feb 2000	1.174	0.384	Model pcpn	0.400	Obs pcpn			264	546	1.025	0.145	0.501
28	28 Feb 2000	0.838	0.393	Model pcpn	0.395	Obs pcpn			279	546	1.107	0.121	0.515
30	13 Mar 2000	0.655	0.377	Model pcpn	0.413	Obs pcpn			293	546	1.011	0.098	0.518
32	28 Mar 2000	0.913	0.370	Model pcpn	0.424	Obs pcpn			308	546	1.006	0.074	0.519

Subsequent runs were conducted approximately each week of the 1998/99 and 1999/2000 winter seasons, for about 200 Canadian stations (the exact number of stations varies slightly with the predictand), separately for the 0000 and 1200 UTC runs of GEM, and for temperature, wind, and POP. Again to limit the experiment to ''pure'' winter data, equations were not developed during the spring and autumn 1999 season blending periods. Winter season data from the blending period were added to the sample nevertheless, resulting in a jump in the new model sample size at the start of the 1999/2000 winter season. The reduced set of predictors was screened using forward stepwise regression at each step, with 0.5% additional reduction of variance (RV) as a criterion for stopping selection for temperature, and 3% additional reduction of variance as a stopping criterion for POP and wind. We felt the stopping criterion should be quite strict to avoid overfitting the data since relatively small samples are being used in development. Most of the equations contain fewer than six predictors, and we have encountered many equations where only one predictor was selected.

The results of equation development during the model transition period are described in this section, first for individual stations, then summarized over all 200 stations used in the experiment. Some examples of equation development through the season transition period are also shown to illustrate this aspect of the system.

a. Model transition development results

1) SINGLE-STATION ANALYSIS

Table 1 shows the weekly equation development for 0–6-h POP for Toronto (Pearson International Airport—CYYZ), through the two-season model transition period. At step one, only old model data were used, while at step 32, the development sample consisted nearly entirely of new model data. The predictors are listed in the order selected, along with the new and old model sample sizes, and the weights computed from Eqs. (3) and (4). The development sample variance explained by each equation is shown in the column at the right-hand side. For precipitation, 350 cases are needed before the old model data are eliminated from the development, but it can be seen from the table that the weight on the old model data is very small by the last step, with 308 new model cases accumulated. It should be noted that, in this table and all that follow, interpretation of the equations in absolute terms is difficult because the predictors are linearly coded with multipliers and offsets to economize on storage requirements. Since the regression is unaffected by a linear transformation, we did not need to transform the encoded predictors to actual values.

Changes in the equations for this short-range forecast period are remarkably small; even the coefficients change only slowly. Much of the time, the equation is essentially a nearly equally weighted balance between the persistence predictor and the model 6-h precipitation forecast, both binary predictors. During part of the period a third predictor is added, either the model's precipitation rate or the 700-hPa relative humidity. The reduction of variance increases slightly between the first (old model) and last step.

Table 2 shows the equations through the model transition period for the 36–42-h projection for POP, also for CYYZ. There are more changes to the predictor set in this case, especially near the end of the period, though

TABLE 2. Results of UMOS POP equation development for update cycle during the model transition period from Dec 1998 to Mar 2000, for CYYZ, 42-h forecasts. Here RV is the reduction of variance on the development sample. The predictand is expressed in percent (0%–100%).

Step	Run date	Constant	Coefficient	Predictor	Coefficient	Predictor	Coefficient	Predictor	Sample new	Sample old	Weight New	Weight old	RV
1	3 Dec 1998	−109.224	0.470	RH 700	0.022	*V* comp 850–1000 shear			0	548	0.000	1.000	0.213
2	13 Dec 1998	−3.444	0.494	RH 700	10.140	Rain rate			35	548	1.666	0.534	0.216
3	17 Dec 1998	−3.629	0.495	RH 700	10.206	Rain rate			39	548	1.560	0.528	0.218
4	28 Dec 1998	−2.300	0.473	RH 700	10.614	Rain rate			50	548	1.463	0.505	0.208
5	3 Jan 1999	−2.554	0.471	RH 700	10.725	Rain rate			56	548	1.423	0.493	0.210
6	13 Jan 1999	−2.936	0.474	RH 700	11.294	Rain rate			66	548	1.378	0.473	0.218
8	28 Jan 1999	−3.866	0.491	RH 700	10.455	Rain rate			81	548	1.319	0.444	0.233
10	13 Feb 1999	−4.707	0.514	RH 700	10.318	Rain rate			97	548	1.269	0.414	0.238
12	28 Feb 1999	−4.593	0.509	RH 700	10.394	Rain rate			112	548	1.230	0.387	0.234
14	13 Mar 1999	−4.956	0.520	RH 700	10.743	Rain rate			125	548	1.200	0.365	0.255
16	28 Mar 1999	−5.205	0.506	RH 700	11.124	Rain rate			140	548	1.170	0.340	0.252
17	3 Dec 1999	−7.802	0.748	Model pcpn	0.390	RH 850			192	548	1.090	0.257	0.257
18	13 Dec 1999	−8.095	0.759	Model pcpn	0.386	RH 850			202	548	1.078	0.241	0.263
20	28 Dec 1999	−8.672	0.765	Model pcpn	0.410	RH 850			217	548	1.062	0.218	0.267
22	13 Jan 2000	−10.092	0.461	RH 850	12.230	Rain rate			234	548	1.047	0.192	0.248
24	28 Jan 2000	−110.072	1.428	RH 700	0.379	RH 850	0.342	700 Dew Pt Dep	248	548	1.036	0.170	0.283
26	13 Feb 2000	−109.172	1.442	RH 700	0.361	RH 850	0.338	700 Dew Pt Dep	263	548	1.026	0.146	0.290
28	28 Feb 2000	−502.128	1.610	RH 700	0.101	E–W dT 850	0.361	700 Dew Pt Dep	278	548	1.018	0.122	0.299
30	13 Mar 2000	−509.435	1.616	RH 700	0.103	E–W dT 850	0.367	700 Dew Pt Dep	292	548	1.011	0.100	0.302
32	28 Mar 2000	−366.796	0.400	RH 700	11.174	Rain rate	0.091	E–W dT 850	307	548	1.006	0.075	0.307

many of the changes are limited to permutations among the same set of predictors, usually not involving the first-selected predictor. It is noteworthy that the predictors change immediately after the first batch of new model data is added. Although encouraged by the weighting scheme, we were surprised at the rapidity of the response to the new model data in some cases. The addition of the winter portion of the transition season data at step 17 led to a temporary shift in the predictors. These data, taken from the end and the beginning of winter, may be less seasonally representative than the rest of the data in the sample. For this projection and station, the 307 cases from the new model appear to provide a better fit to the predictand than the 548 cases from the old model. Such differences could be attributable to differences between the two samples, however. It would not be possible to separate the effects of sample differences from model-related differences without a long parallel run of the two models.

Tables 3 and 4 show the model transition results for 42-h temperature forecasts, for CYYZ and Whitehorse, Yukon (CYXY), respectively. CYXY was chosen for comparison because observations from that station are dominated by local effects, and we wanted to see whether a more ‘‘difficult’’ station would exhibit a greater variation in the coefficients and/or selected predictors through the model transition.

For CYYZ, once again, the predictors change completely as soon as new model data are added and remain remarkably constant afterward. This suggests that even a small sample can give a good indication of the predictor–predictand relationship. As the sample becomes dominated by new model data, the number of predictors selected reduces to two from four with the old model and the fit improves by a 5% additional reduction of variance. It is also notable that the equations with new model data prefer the model's forecast of surface temperature as the first predictor, whereas the model's estimate of surface temperature does not even appear in the equation based entirely on old model data.

For CYXY too, the selected predictors do not change much over the transition period. However, the number of predictors selected rises from one to five and the overall variance explained drops from 83% to 77%. Also, the model's surface temperature forecast is shunned as a predictor in all the equations. Clearly, the new model does not fit the temperature data from CYXY as well as the old model did, which is contrary to expectation. The higher resolution of the new model would have been expected to result in better resolution of the local effects at CYXY. While allowing for the effects of possible differences in the development samples, it also seems likely that the additional resolution in the new model has added more noise than signal to the predictor data, at least for this area.

Figure 5 shows the reduction of variance on the development sample as a function of projection time for CYYZ and CYXY surface temperature forecasts. The

TABLE 3. Results of UMOS temperature equation development during the model transition period from Dec 1998 to Mar 2000, for CYYZ, 42-h forecasts. Here RV is the reduction of variance on the development sample. The lower part of the table gives the results of continued equation development through the winter–summer season transition of 2000. The predictand is in tenths of kelvin.

Step	Run date	Constant	Coefficient	Predictor	Coefficient	Predictor	Coefficient	Predictor	Coefficient	Predictor	Sample new	Sample old	Weight new	Weight old	RV
1	3 Dec 1998	133.489	3.5860	Th 1000-925	−0.2030	Clouds	0.0979	Sun elev	0.0856	Hgt 700	25	503	0	1	0.843
2	13 Dec 1998	−90.777	0.7838	*T* surface	0.2283	*T* 850	0.068	Sun elev	−0.1676	Clouds	35	503	1.538	0.489	0.847
3	17 Dec 1998	−93.585	0.7913	*T* surface	0.2236	*T* 850	−0.1697	Clouds	0.0643	Sun elev	39	503	1.496	0.481	0.847
4	28 Dec 1998	−98.149	0.7816	*T* surface	0.2345	*T* 850	−0.1608	Clouds	0.065	Sun elev	50	503	1.414	0.456	0.847
5	3 Jan 1999	−118.753	0.7813	*T* surface	0.2424	*T* 850	0.0651	Sun elev	−0.1535	Clouds	56	503	1.381	0.443	0.854
6	13 Jan 1999	−98.808	0.7757	*T* surface	0.2396	*T* 850	0.0669	Sun elev	−0.1462	Clouds	66	503	1.334	0.421	0.862
8	28 Jan 1999	−101.473	0.7945	*T* surface	0.2199	*T* 850	−0.1651	Clouds	0.07027	Sun elev	81	503	1.276	0.391	0.860
10	13 Feb 1999	−98.815	0.8040	*T* surface	0.2093	*T* 850	−0.1706	Clouds	0.07005	Sun elev	97	503	1.227	0.360	0.860
12	28 Feb 1999	−88.000	0.8016	*T* surface	0.2063	*T* 850	−0.1656	Clouds	0.07312	Sun elev	112	503	1.188	0.332	0.860
14	13 Mar 1999	−82.38	0.7909	*T* surface	0.2156	*T* 850	−0.1678	Clouds	0.07196	Sun elev	125	503	1.159	0.308	0.860
16	28 Mar 1999	−83.249	0.7733	*T* surface	0.2283	*T* 850	0.0837	Sun elev	−0.16	Clouds	140	503	1.129	0.282	0.856
17	3 Dec 1999	−31.835	0.7909	*T* surface	0.1929	*T* 850	0.0729	Sun elev			192	503	1.056	0.194	0.876
18	13 Dec 1999	−41.194	0.7921	*T* surface	0.1960	*T* 850	0.0711	Sun elev			202	503	1.045	0.177	0.877
20	8 Dec 1999	−30.454	0.7919	*T* surface	0.1928	*T* 850	0.07	Sun elev			217	503	1.032	0.151	0.879
22	13 Jan 2000	−24.221	0.7877	*T* surface	0.1940	*T* 850	0.0807	Sun elev			233	503	1.021	0.124	0.880
24	28 Jan 2000	−52.176	0.7905	*T* surface	0.2009	*T* 850	0.0728	Sun elev			247	503	1.013	0.099	0.894
26	13 Feb 2000	−51.781	0.7848	*T* surface	0.2070	*T* 850	0.072	Sun elev			262	503	1.007	0.072	0.898
28	28 Feb 2000	−33.883	0.7844	*T* surface	0.2009	*T* 850	0.0717	Sun elev			277	503	1.002	0.044	0.898
30	13 Mar 2000	−18.803	0.8171	*T* surface	0.1964	*T* 850					291	503	1.0004	0.018	0.895
32	28 Mar 2000	−42.024	0.8241	*T* surface	0.1980	*T* 850					306	503	1	0	0.899
											Sample winter	Sample summer	Weight winter	Weight summer	RV
33	28 Mar 2000	−47.498	0.8278	*T* surface	0.1964	*T* 850					312	306	0.6571	0.33	0.9
34	13 Apr 2000	−42.200	0.8175	*T* surface	0.2048	*T* 850					322	306	0.4752	0.50	0.898
35	17 Apr 2000	−45.648	0.8200	*T* surface	0.2035	*T* 850					326	306	0.4693	0.50	0.898
36	28 Apr 2000	−37.809	0.8202	*T* surface	0.2004	*T* 850					331	312	0.4713	0.50	0.898
37	3 May 2000	−42.697	0.8244	*T* surface	0.1980	*T* 850					331	317	0.3160	0.67	0.902
38	13 May 2000	−40.924	0.8275	*T* surface	0.1941	*T* 850					331	327	0.3260	0.67	0.909
39	17 May 2000	−47.116	0.8350	*T* surface	0.1889	*T* 850					331	331	0	1.00	0.915

TABLE 4. Results of UMOS temperature equation development during the model transition period from Dec 1998 to Mar 2000, for CYXY, 42-h forecasts. Here RV is the reduction of variance on the development sample. The predictand is in tenths of kelvin.

Step	Run date	Constant	Coefficient	Predictor	Coefficient	Predictor	Coefficient	Predictor	Coefficient	Predictor	Coefficient	Predictor	Sample new	Sample old	Weight new	Weight old	RV
1	3 Dec 1998	−453.27	2.443	Thick 1000–850									25	501	0	1	0.832
2	17 Dec 1998	−369.33	2.379	Thick 1000–850									35	501	1.538	0.491	0.810
3	17 Dec 1998	−365.94	2.376	Thick 1000–850									39	501	1.500	0.482	0.809
4	28 Dec 1998	−216.67	4.181	Thick 1000–925									50	501	1.414	0.458	0.802
5	3 Jan 1999	−237.06	4.712	Thick 1000–925									56	501	1.381	0.445	0.798
6	13 Jan 1999	−217.01	4.680	Thick 1000–925									66	501	1.334	0.423	0.802
8	28 Jan 1999	−229.56	4.700	Thick 1000–925									81	501	1.276	0.392	0.803
10	13 Feb 1999	−606.62	4.173	Thick 1000–925	0.4610	Hgt 700	−0.404	Hgt 850					97	501	1.227	0.361	0.768
12	28 Feb 1999	−288.87	2.309	Thick 1000–850									112	501	1.188	0.333	0.743
14	13 Mar 1999	−429.64	4.619	Thick 1000–925	0.0811	Hgt 700							125	501	1.159	0.310	0.752
16	28 Mar 1999	−451.20	4.662	Thick 1000–925	0.0847	Hgt 700							140	501	1.130	0.283	0.758
17	3 Dec 1999	−530.73	4.722	Thick 1000–925	0.0954	Hgt 700							192	501	1.056	0.194	0.770
18	13 Dec 1999	−535.16	4.705	Thick 1000–925	0.1000	Hgt 700							202	510	1.045	0.177	0.763
20	28 Dec 1999	−550.03	4.708	Thick 1000–925	0.1040	Hgt 700							217	501	1.032	0.152	0.764
22	13 Jan 2000	−560.91	4.707	Thick 1000–925	0.1080	Hgt 700							233	501	1.021	0.124	0.761
24	28 Jan 2000	5587.83	4.789	Thick 1000–925	0.1100	Hgt 700	−15.512	*T*adv 1000					247	501	1.013	0.099	0.769
26	13 Feb 2000	5883.49	4.770	Thick 1000–925	0.1180	Hgt 700	−16.277	*T*adv 1000					262	501	1.007	0.072	0.764
28	28 Feb 2000	5905.18	4.898	Thick 1000–925	0.0986	Hgt 700	−16.317	*T*adv 1000	−0.378	RH 700			277	501	1.002	0.045	0.764
30	13 Mar 2000	6318.80	4.914	Thick 1000–925	0.1030	Hgt 700	−17.401	*T*adv 1000	−0.400	RH 700			291	501	1.0005	0.0197	0.761
32	28 Mar 2000	6772.13	5.101	Thick 1000–925	0.1260	Hgt 700	−16.423	*T*adv 1000	−0.246	Speed 850	−0.383	RH 700	307	501	1	0	0.775

heavy lines are for equations based on pure old and new model data, while the broken lines are for intermediate stages of mixed data. Generally, the RV at intermediate stages lies in between the RV achieved on data from either model alone, as expected. At CYXY, the RV levels are close to what is achieved on pure new model data by the time the new model sample size reaches 100 cases, but the fit to new model data is poorer than to the old model at all projections. There is also a diurnal variation in the RV at CYXY, with the highest values in the afternoon near the diurnal temperature maximum. The diurnal variation is stronger for the new model, again possibly reflecting the effects of the higher horizontal resolution. At CYYZ, there is more variation in the rate of transition from the old to the new model, but the new model explains more of the predictand variance at all projections except one. Thus, the results of Tables 3 and 4 are reflected at all other projection times.

Figure 6 gives RVs as a function of projection time for POP for CYYZ and CYXY. These results are generally consistent with Fig. 5: the fit is improved at CYYZ and is degraded slightly at CYXY. However, differences between old and new models are small, which is consistent with Bélair et al. (2000) who found the impact of the new precipitation scheme to be neutral in wintertime situations.

2) SUMMARY DEVELOPMENT RESULTS

Summary development results over 200 stations are shown in Figs. 7–10. Summary statistics on the frequency of selection of predictors are given in Figs. 7–9 for each of the three elements, for the 42-h projection time. In each of the figures, all predictors that appear in more than 10% of the equations based on new model data are listed. For each of these predictors, the five histogram bars show the frequency of first selection and the frequency of selection for the old model data, 40 new model cases, 100 new model cases, 200 new model cases, and 300 new model cases, from top to bottom. The predictors are listed from most frequently used to least frequently used from bottom to top. The first characteristic to note is that for all of the predictands, the shift in predictors over the model transition period is not large. Changes in frequency of selection of more than 20% are uncommon. This supports the results shown for single stations, which showed that the same few predictors would often be selected, but might change order. Second, the first predictor selected tends toward the corresponding model variable from the new model for all three elements. This supports the claim that the new version of the model provides an improved fit to station data in comparison with the old model. Third, for all three predictands, the preference for the model estimate of the corresponding element seems to increase more rapidly after 200 cases of the new model are included.

For temperature (Fig. 7), the model's 1000-hPa tem-

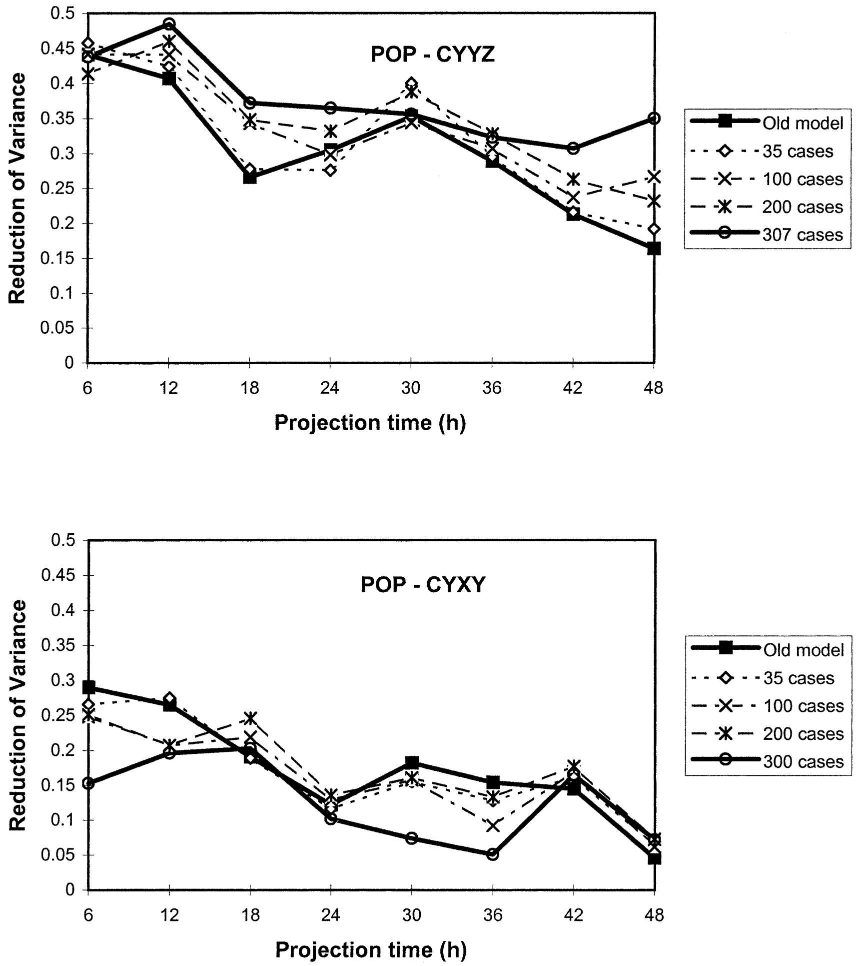


FIG. 6. Same as in Fig. 5 but for POP forecasts.

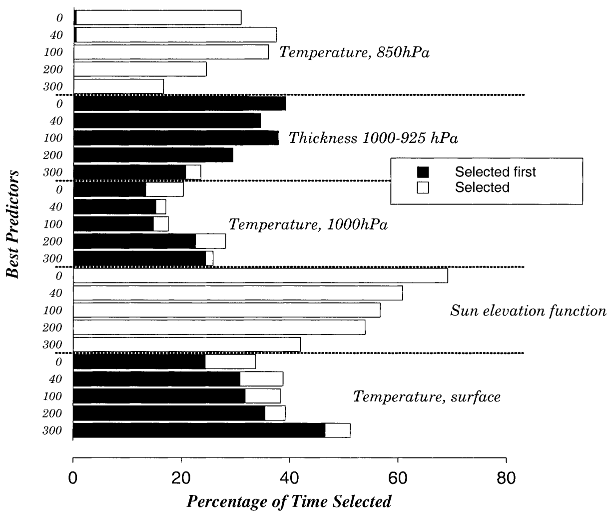


FIG. 7. Frequency of selection of predictors for 42-h temperature forecasts over 200 Canadian stations, for different blends of new and old model data. Predictors are listed in decreasing frequency of selection using new model data, from bottom to top.

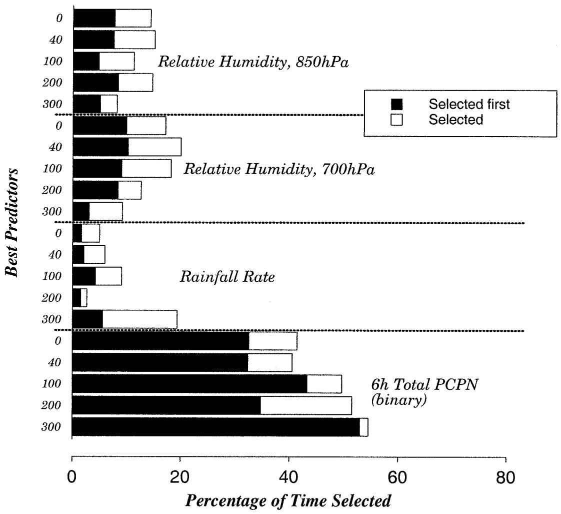


FIG. 8. Same as in Fig. 7 but for POP forecasts.

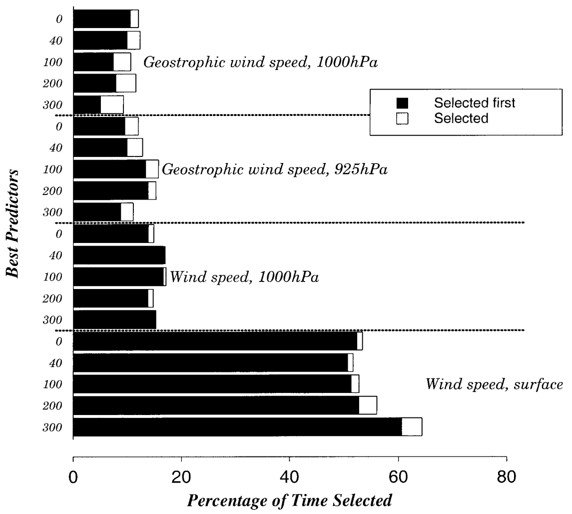


FIG. 9. Same as in Fig. 7 but for wind speed forecasts.

perature is also favored in the new model more than in the old, and these preferences appear to come at the expense of the sun elevation function and the low-level thickness. It is interesting that the sun elevation function, while almost as popular as the model's surface temperature, is never picked first.

For precipitation (Fig. 8), the model's 6-h precipitation forecast (a binary predictor) is picked as the first predictor in more than 50% of the equations using new model data. There seems also to be a slight increase in preference for the model's rainfall rate and a slight decrease in the preference for the relative humidity predictors as the new model sample increases.

For wind speed (Fig. 9), there is a slight tendency for the equations from data from a single model to prefer the model's wind speed more often than the equations based on mixed model data. Since wind speed is related to gradients mostly, and since we do not smooth the predictors, it is possible that the change in model resolution with a corresponding sharpening of gradients might have introduced some extra noise into the mixed samples.

Figure 10 shows the RV achieved on the development sample averaged over all 200 stations, through the model transition period. The steps indicated on the abscissa correspond to the steps represented by the different new model sample sizes shown in Figs. 7–9. In general, Fig. 10 indicates that no matter how mixed the development sample, the reduction of variance achieved by the regression does not change very much. This suggests that, statistically, the samples from the two models are not radically different and is consistent with the fact that the selected predictor sets also do not change significantly. The more sudden changes in RV at step 17 correspond to the addition of the winter data from the season transition periods of 1999, data that could be expected to change the sample characteristics somewhat. While the RV for temperature and POP remains nearly

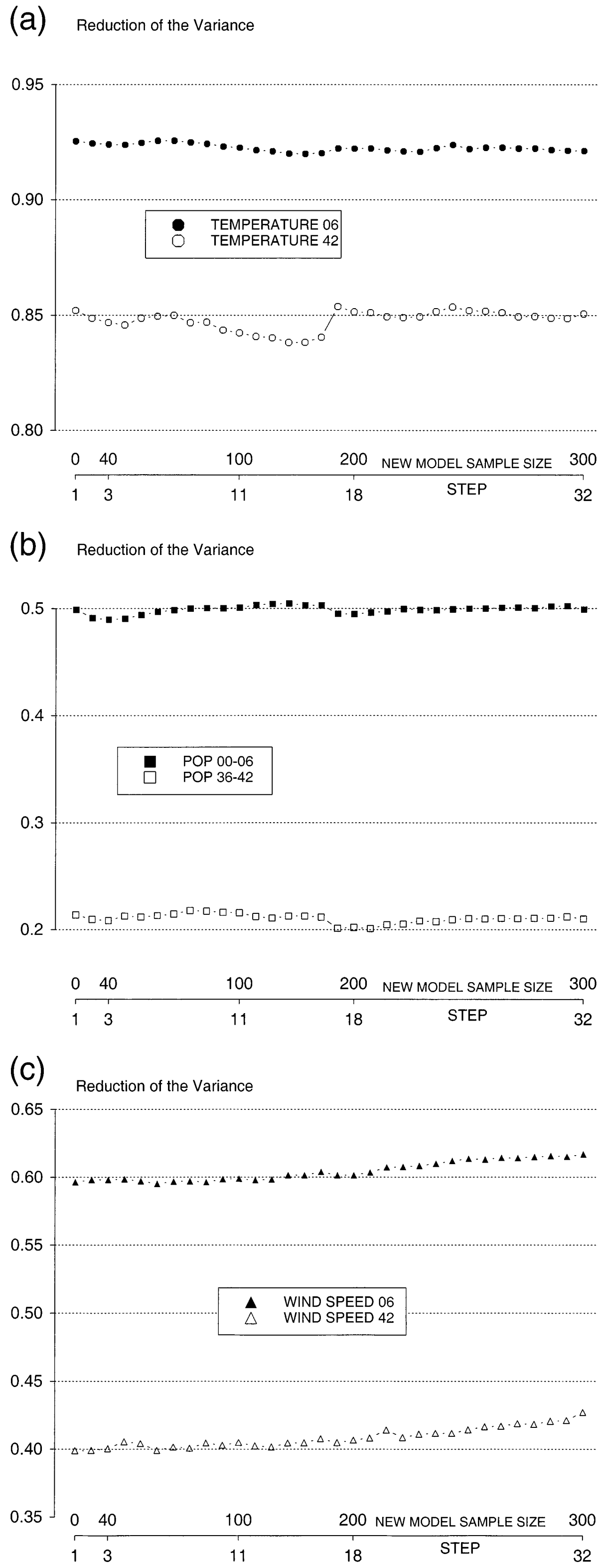


FIG. 10. Average dependent sample variance explained over 200 Canadian stations as a function of development step for the winter model transition period Dec 1998–Mar 2000. Black symbols are for 6-h forecasts and open symbols are for 42-h forecasts: (a) temperature, (b) POP, and (c) wind speed.

constant overall, the RV for wind speed increases steadily through the transition period. We attribute this to the increase in resolution of the new model, which permitted a better fit to station wind speed data.

b. Seasonal transition development results

The lower section of Table 3 shows the weekly updated equations, sample sizes, the seasonal weights applied, and RVs for CYYZ 42-h temperature forecasts. These equations were generated by continuing the update cycle through the spring 2000 seasonal transition. Since both summer and winter equations are based on new model data only by the beginning of the seasonal transition, it is necessary only to blend the two seasonal SSCP matrices. The nominal seasonal weights vary stepwise through the seasonal transition, in three steps: 1/3 to 2/3, ½ to ½, and 2/3 to 1/3, approximately 2 weeks at each step. The weights are modified to adjust for differing sample sizes according to Eq. (5). For CYYZ, again there is little change in the equations, coefficients, and the reduction of variance through this period.

The more complicated situation where both seasonal and model transitions occur concurrently is displayed in Table 5, showing the continuation of the POP update cycle for CYYZ for spring 2000. The summer equations are based on new model data alone, while the winter equations still use a blend of new and old model data. At 6 h, the equations continue to be a weighted combination of the 6-h model precipitation forecast and the latest observation, but the weight on the latest observation increases while the weight on the model precipitation decreases, until the last step, when the model precipitation predictor is no longer significant. At the 42-h projection, predictors change with the season change. In this case, the model precipitation forecast is used as the first predictor for the summer season equation, which is consistent with Bélair et al. (2000), who showed that the new convection scheme improved the simulation of summertime precipitation. The decrease in the explained variance from winter to summer is expected given the predominantly small-scale showery nature of summertime precipitation.

4. Discussion and future work

We have designed, developed, and tested a MOS forecast system that is responsive and adaptive to model changes. Regression equations for temperature, POP, and wind direction and speed are updated every week, after the addition of the latest model output data to the development sample. After a significant model change, data from the new and old models are blended according to a weighting scheme that emphasizes the new model data and gradually phases out dependence on old model data. The system also ensures a smooth transition between summer and winter equations using a second

TABLE 5. Results of UMOS POP equation development for the update cycle during the cool to warm season transition in 2000, for CYYZ, (top) 6- and (bottom) 42-h forecasts. Here *N* is the sample size and Wgt is the weight that is applied. The predictand is expressed in percent (0%–100%).

		Cool				Warm													
Step	Date	*N* new	*N* old	Wgt new	Wgt old	*N* new	*N* old	Wgt new	Wgt old	Wgt cool	Wgt warm	Constant	Coefficient	Predictor	Coeficient	Predictor	Coefficient	Predictor	RV
6 h																			
1	31 Mar 2000	311	546	1.005	0.068	356	597	1	0	0.67	0.320	2.211	0.484	Obs pcpn	0.331	Model pcpn			0.508
2	3 Apr 2000	314	546	1.004	0.063	356	597	1	0	0.67	0.324	2.198	0.486	Obs pcpn	0.331	Model pcpn			0.509
3	13 Apr 2000	324	546	1.002	0.046	356	597	1	0	0.5	0.491	2.836	0.545	Obs pcpn	0.298	Model pcpn			0.512
4	17 Apr 2000	328	546	1.002	0.039	356	597	1	0	0.5	0.492	2.821	0.547	Obs pcpn	0.297	Model pcpn			0.513
5	28 Apr 2000	333	546	1.001	0.031	362	597	1	0	0.5	0.483	2.746	0.556	Obs pcpn	0.291	Model pcpn			0.520
6	3 May 2000	333	546	1.001	0.031	367	597	1	0	0.33	0.639	3.303	0.602	Obs pcpn	0.266	Model pcpn			0.515
7	13 May 2000	333	546	1.001	0.031	377	597	1	0	0.33	0.622	3.433	0.602	Obs pcpn	0.267	Model pcpn			0.518
8	17 May 2000	333	546	1.001	0.031	381	597	1	0	0	1	6.000	0.875	Obs pcpn					0.503
42 h																			
1	31 Mar 2000	310	548	1.005	0.070	360	603	1	0	0.67	0.321	−8.313	0.386	RH 700	0.255	Model pcpn			0.242
2	3 Apr 2000	313	548	1.005	0.065	360	603	1	0	0.67	0.321	−8.285	0.385	RH 700	0.218	Model pcpn			0.241
3	13 Apr 2000	323	548	1.002	0.048	360	603	1	0	0.50	0.486	−346.8	0.395	RH 700	0.086	E–W dT 850	8.06	Rain rate	0.275
4	17 Apr 2000	327	548	1.002	0.041	360	603	1	0	0.50	0.486	−350.7	0.389	RH 700	0.087	E–W dT 850	8.12	Rain rate	0.275
5	28 Apr 2000	332	548	1.001	0.032	366	603	1	0	0.50	0.478	−7.157	0.359	RH 700	0.223	Model pcpn			0.235
6	3 May 2000	332	548	1.001	0.032	371	603	1	0	0.33	0.632	−6.774	0.347	RH 700	0.217	Model pcpn			0.230
7	13 May 2000	332	548	1.001	0.032	381	603	1	0	0.33	0.616	865.09	0.253	Model pcpn	−0.086	700 vert vel			0.228
8	17 May 2000	332	548	1.001	0.032	385	603	1	0	0	1	834.76	0.222	Model pcpn	−0.083	700 vert vel			0.209

weighting scheme during 6-week transition periods in spring and autumn.

The results of tests following a significant model change indicated that the forecast equations remained remarkably stable through the transition period. The reduction of variance changed little, and the predictor sets that were selected also changed only slightly. The model changes seem in this case to be quite significant: the convection scheme was changed completely and the horizontal resolution was increased significantly. However, the results suggest that the statistical characteristics of the model variables did not change greatly after the change of model. One could conceive of more radical changes, for example a complete change of model dynamics or physics package, that might significantly increase the unexplainable variance when samples from the new and old models are blended. There was little evidence in these results of increased difficulty in fitting the blended samples to the observations for the three elements tested. As a result, it was possible to obtain useful equations even within a few months after the model change. The results shown in this paper focus on the goodness of fit of predictors and predictand when development samples from different models are blended. In order to determine whether the forecasts from UMOS produce acceptable forecast guidance during the model or seasonal transition periods it is necessary to carry out a comparative verification with respect to the operational perfect prog forecasts using independent data. This will be the subject of a second paper.

The UMOS system became operational in late 2000 and has run in parallel with the perfect prog system since that time, for temperature, POP, and wind. Forecasters have notified us of several significant successes of the UMOS forecasts in comparison with the perfect prog forecasts, and they are used as input to CMC's automated forecast system called ''SCRIBE.'' Complete replacement of the perfect prog forecasts awaits completion of the system for all predictands that are required to drive SCRIBE.

To manage the UMOS system in operations, it is necessary to decide what constitutes a ''significant'' enough model change to warrant the start of a new set of SSCP matrices. From a statistical point of view, any model change that is expected to lead to a change in the bias characteristics of any variable used in the equations should be considered significant. Fortunately, the implementation process is complicated enough that our modelers tend to package several small changes together for each new implementation, rather than implement them one at a time. That being the case, perhaps all new implementations will be significant for UMOS.

All of the equations developed so far have used multivariate linear regression. We have been developing an updateable form of MDA for use with multicategory variables such as precipitation type, cloud amount, ceiling, and visibility. Since MDA also uses components of the SSCP matrix, the modifications required are not extensive. However, the handling of categorized predictands poses some practical problems for an update system: The system must be capable of changing the number of categories for which probabilities are estimated, depending on changes in the distribution of the predictand observations, to ensure that there are always enough cases in all categories to obtain a stable probability estimate. This is a particularly significant problem for predictands with highly skewed or exponential distributions such as precipitation amount. We have begun designing extensions to the system that will take account of this problem.

It also turns out that the data management portions of the system are flexible enough that we could use the UMOS framework to develop statistical interpretation equations on fixed datasets as well—for example, to use a reanalysis dataset to develop new perfect prog equations. Once these are available, real-time perfect prog forecasts could be produced and added as new predictors to a UMOS update cycle.

Acknowledgments. The authors thank Dr. William Burrows and two anonymous reviewers for their helpful comments on earlier versions of this paper.

REFERENCES

Bélair, S., A. Méthot, J. Mailhot, B. Bilodeau, A. Patoine, G. Pellerin, and J. Coté, 2000: Operational implementation of the Fritsch–Chappell convective scheme in the 24-km Canadian regional model. *Wea. Forecasting,* **15,** 257–274.

Brunet, N., R. Verret, and N. Yacowar, 1988: An objective comparison of model output statistics and perfect prog systems in producing numerical weather element forecasts. *Wea. Forecasting,* **3,** 273–283.

Carter, G. M., 1986: Towards a more responsive statistical guidance system. Preprints, *11th Conf. on Weather Forecasting and Analysis,* Kansas City, MO, Amer. Meteor. Soc., 39–45.

——, J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting,* **4,** 401–412.

Dallavalle, J. P., 1988: An evaluation of techniques used by the National Weather Service to produce objective maximum/minimum temperature forecasts. Preprints, *Eighth Conf. on Numerical Weather Prediction,* Baltimore, MD, Amer. Meteor. Soc., 572–579.

Draper, N. R., and H. Smith, 1998: *Applied Regression Analysis.* Wiley, 706 pp.

Glahn, H. R., 1970: A method for predicting surface winds. ESSA Tech. Memo. WBTM TDL 29, U.S. Department of Commerce, Washington, DC, 18 pp.

——, and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

——, and J. P. Dallavalle, 2000: MOS-2000. TDL Office Note 00-1, NOAA/Techniques Development Laboratory, 169 pp.

Klein, W. H., and F. Lewis, 1970: Computer forecasts of maximum and minimum temperatures. *J. Appl. Meteor.,* **9,** 350–359.

Lemcke, C., and S. Kruizinga, 1988: Model output statistics: Three years of operational experience in the Netherlands. *Mon. Wea. Rev.,* **116,** 1077–1090.

Ross, G. H., 1987: An updateable model output statistics scheme. Programme on Short- and Medium Range Weather Prediction, PSMP Report Series, No. 25, World Meteorological Organization, 25–28.

——, 1989: Model output statistics—An updateable scheme. Preprints, *11th Conf. on Probability and Statistics in Atmospheric Sciences*, Monterey, CA, Amer. Meteor. Soc., 93–97.

——, 1992: Probability of precipitation from an updateable MOS model. Preprints, *12th Conf. on Probability and Statistics in Atmospheric Sciences,* Toronto, ON, Canada, Amer. Meteor. Soc., 190–194.

Sarrazin, R., and L. J. Wilson, 1989: A tuned perfect prog wind forecast system for Canada. Preprints, *11th Conf. on Probability and Statistics in Atmospheric Sciences,* Monterey, CA, Amer. Meteor. Soc., 50–55.

Simonsen, C., 1991: Self adaptive output statistics based on Kalman filtering. Programme on Short- and Medium Range Weather Prediction Research, PSMP Report Series, No. 34, WMO TD No. 421, XX33–XX37.

Vallée, M., L. J. Wilson, and P. Bourgouin, 1996: New statistical methods for the interpretation of NWP model output at the Canadian Meteorological Center. Preprints, *13th Conf. on Probability and Statistics in Atmospheric Sciences,* San Francisco, CA, Amer. Meteor. Soc., 37–44.

Wilson, L. J., 1985: Application of statistical methods to short range operational weather forecasting. Preprints, *Ninth Conf. on Probability and Statistics in Atmospheric Sciences,* Virginia Beach, VA, Amer. Meteor. Soc., 1–10.

——, W. R. Burrows, and K. Macdonald, 1986: An assessment of operational MOS surface wind forecasts. Preprints, *11th Conf. on Weather Forecasting and Analysis,* Kansas City, MO, Amer. Meteor. Soc., 11–16.

Yacowar, N., G. Richard, N. Brunet, and H. Yang, 1985: Development of a MOS system to forecast the probability of precipitation. Canadian Meteorological Centre Information, Vol. 2, No. 4, 24–46.

Zurndorfer, E. A., J. R. Bocchieri, G. M. Carter, J. P. Dallavalle, D. B. Gilhousen, K. F. Hebenstreit, and D. J. Vercelli, 1979: Trends in comparative verification scores for guidance and local aviation/public weather forecasts. *Mon. Wea. Rev.,* **107,** 799–811.