WORK ON THE STATISTICAL ADAPTATION FOR LOCAL FORECASTS IN FRANCE

D. Rousseau

Direction de la Météorologie, EERM/GMD

Paris, France

1. INTRODUCTION

This paper is limited to statistical weather prediction studies done in France and deals with local statistical adaptation of numerical weather predictions. All the methods discussed use the output of a numerical forecast model as the main predictors. Most of the operational experience has been obtained for short range forecasts. However the methods may be used easily for medium range forecasts and some limited experiments has been gained in this field using the ECMWF model forecasts.

The methods may be divided roughly in two classes:

- Perfect prog methods. The statistical relation is derived between the local predictand and observed meteorological predictors; output of a forecast numerical model is used in place of the unknown observed predictor in order to make a forecast. The quality of the statistical relation between the predictand and the predictors is kept only for a perfect numerical model: the decrease of the quality depends upon the quality of the forecast model.

- Model Output Statistics (M.O.S.) In this case the statistical relation is derived by using model output as predictors.

In each method the available predictors are different and the data bases available have different characteristics. Long time series may be used with perfect prog methods; M.O.S. methods generally use small data bases, but the available predictors may be more informative.

After a short description of the data bases available in France (hereafter referred to as files), we discuss the methods and results obtained for the forecast of the following local weather elements: the maximum, and minimum temperature, the occurrence of precipitation during a 12 h period, the amount of precipitation during a 24 h - period and the cloud amount.

2. FILES

The predictand files are derived from the national archive: this consists of files of the French synoptic observations and files of the climatological observations network. This last file has been used to forecast the amount of precipitation in a region (based on the French Department); in this case the predictand has been computed as the average of the values observed at the climatological stations.

The length of the available file is generally larger than the length of the predictor file (on computer from 1949 for the synoptic observations and from 1961 for climatological precipitation observations).

The form of the predictor file depends upon whether the perfect prog or M.O.S. method is used.

The most commonly used file in perfect prog studies is the "PANAL" file which contains analyses at grid points (381 km mesh size at 60°N) of geopotential and temperature on 10 standard levels. This file contains data from 1963. For each study a "learning" file of about 15 years and a test file of 1 or a few years have been used.

With these basic data some other derived predictors have been computed depending upon the predictand. A 10 x 10 grid defined on Fig.1 has been extracted for forecasting purposes in France.

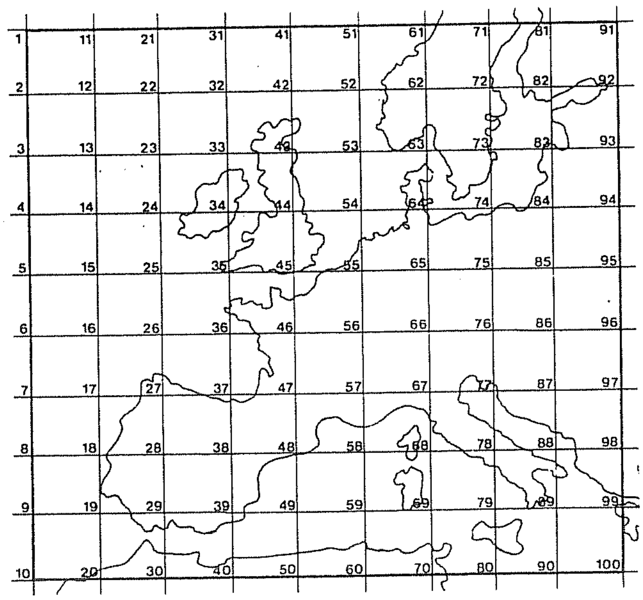

Fig.1 File PANAL grid

The first file available for M.O.S. studies was estblished in 1975 by archiving every 6 hours all historical variables of the primitive equation model of that time (5 levels, 190 km grid size) on a domain covering France. The "learning" files were 2 or 3 years long and test files were one year long. After changing the operational model (1979) new files have been established; no new operational methods have yet been applied to these files. Recently an experiment with cloud amount forecasts has been done by using two month of archives of the ECMWF model.

3. MAXIMUM AND MINIMUM TEMPERATURES

3.1 Perfect prog method

The predictors are the 10 x 10 fields included in file PANAL: geopotential at 10 standard levels (1000 mb to 100 mb), temperature at 9 standard levels (850 mb to 100 mb) and other derived fields - meridional and zonal geostrophic wind, absolute vorticity, advection of absolute vorticity and advection of temperature at 5 levels (850, 700, 500, 400 and 200 mb) and vertical velocity computed by a quasi geostrophic diagnostic equation at 4 levels (800, 600, 400 and 200 mb). The last observation of the predictand (persistence) has been added so that there are 4801 potential predictors.

A first contraction of the information is obtained by the Principal Component Analysis of the fields; the first 10 coefficients are retained. The vertical velocities were not treated in this way: only 9 grid point values covering France were kept.

Then a unique scalar variable called the "canonical variable" is computed for each field. This is the linear combination giving the best estimate of the predictand by linear regression.

The last step is to select the best amongst these last 49 predictors (48 canonical variables + the persistence) by a forward stepwise selection procedure. 10 predictors were chosen (Veysseire 1980, Javelle et al 1980 and 1981). Selection is done for each season and each station.

Table I gives the 10 predictors selected for Paris-Le Bourget. From an inspection of the selected predictors for each station, the following remarks are relevant.

The first selected predictor is the canonical variables of the 850 mb temperature field; the exceptions are in winter and sometimes in autumn for the maximum temperature when the persistence is then a better predictor. 850 mb temperature and persistence have been selected in all computed regressions. Then the more frequent predictors are the canonical variables of the following fields: 850 mb meridional wind, 1000 mb geopotential, 700 mb temperature, 200 mb temperature.

Fall	Winter	Spring	Summer
Minimum Temperature			
850 mb temperature	Persistence	850 mb temperature	850 mb temperature
Persistence	850 mb v_g wind	Persistence	Persistence
700 mb temp. advection	700 mb temp. advection	250 mb temperature	850 mb temp. advection
200 mb temperature	850 mb temperature	700 mb temp. advection	250 mb geopotential
1000 mb geopotential	400 mb vertical velocity	250 mb geopotential	700 mb vort. advection
250 mb temperature	400 mb geopotential	1000 mb geopotential	1000 mb geopotential
200 mb v_g wind	200 mb temperature	500 mb vort. advection	250 mb temperature
850 mb v_g wind	850 mb u_g wind	700 mb absolute vorticity	850 mb absolute vorticity
300 mb vort. advection	500 mb vort. advection	850 mb v_g wind	300 mb temp. advection
500 mb u_g wind	300 mb temperature	500 mb temperature	500 mb geopotential
Maximum Temperature			
Persistence	Persistence	850 mb temperature	850 mb temperature
850 mb temperature	700 mb absolute vorticity	Persistence	Persistence
200 mb temperature	700 mb temp. advection	100 mb temperature	850 mb geopotential
850 mb geopotential	850 mb temperature	700 mb temperature	300 mb geopotential
200 mb v_g wind	150 mb temperature	850 mb absolute vorticity	200 mb temperature
850 mb v_g wind	700 mb temperature	400 mb vertical velocity	500 mb vort. advection
500 mb temperature	400 mb vertical velocity	500 mb geopotential	700 mb v_g wind
400 mb geopotential	300 mb absolute vorticity	200 mb temperature	300 mb u_g wind
250 mb geopotential	100 mb geopotential	300 mb v_g wind	150 mb temperature
100 mb temperature	200 mb geopotential	500 mb v_g wind	500 mb u_g wind

Table I. Best predictors for Paris-Le Bourget (Perfect Prog Method)

It is striking that 200, 150, 100 mb temperature fields are often selected, when a lower level temperature has already been selected. Middle level fields are highly correlated with those at 850 mb, so they add less information than the high level fields. Fig.2 gives the frequency of the selection of the predictors in 72 regression equations (6 stations, 4 seasons, maximum and

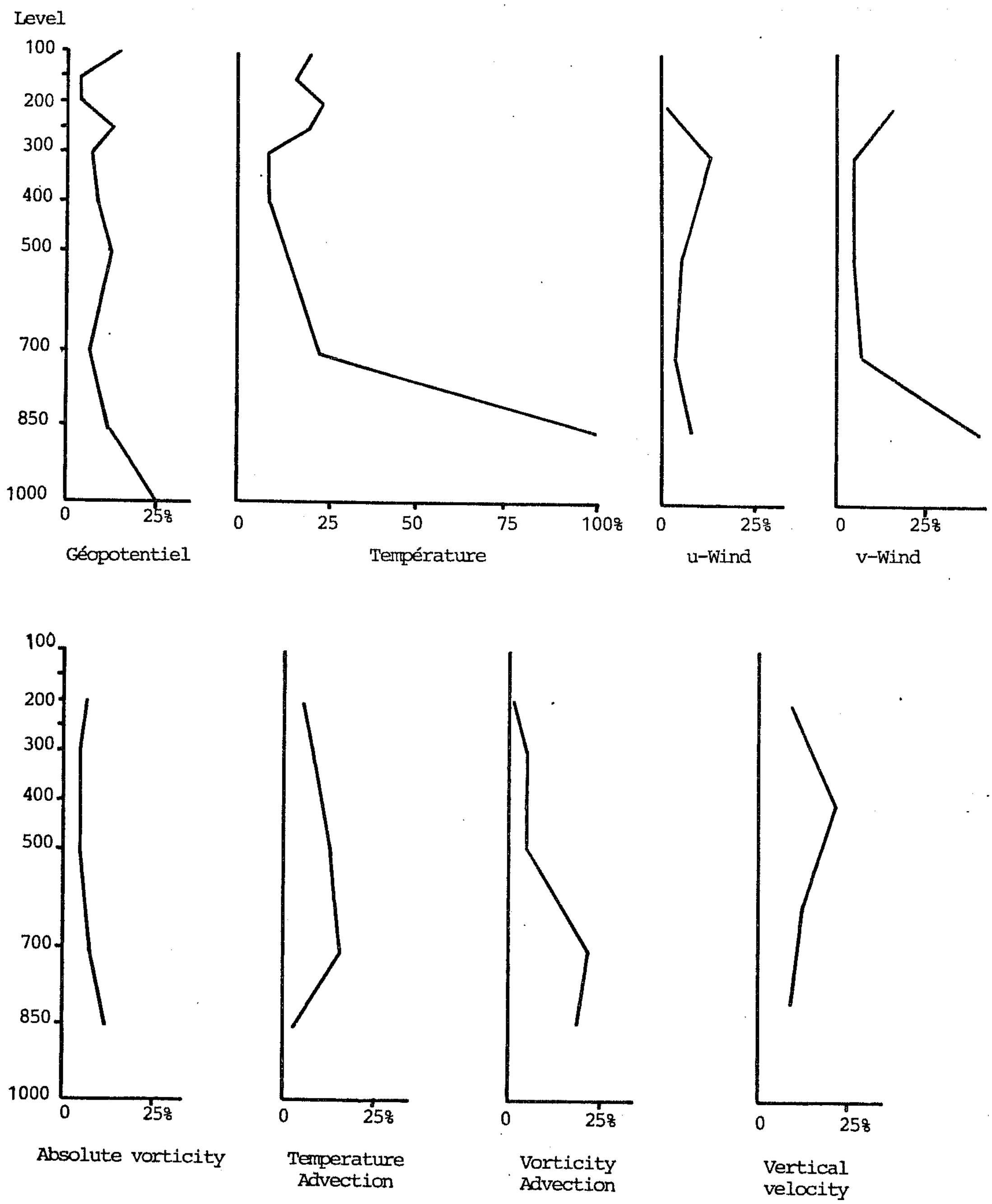

Figure 2. Frequency of selection in the ten best of the 48 canonical variables. (Perfect prog Method).

minimum temperature). The best predictors are obtained from the following levels: 1000 and 100 mb for height; 850, 700 and 200 for temperature, 850 and 200 for meridional wind; 300 mb for zonal wind; 850 mb for vorticity; 700 mb for temperature and vorticity advection; 400 mb for vertical velocity.

The results from the "learning" file give the limitations of this perfect prog method: an average of the mean absolute error of the order of 1.8°C for maximum and minimum temperature.

This perfect prog method has been applied since the end of 1979 for 6 French regional centres by using operational numerical model output. The annual mean of the abolute error for Paris-Le Bourget in 1980 and 1981 is given in Fig.3.

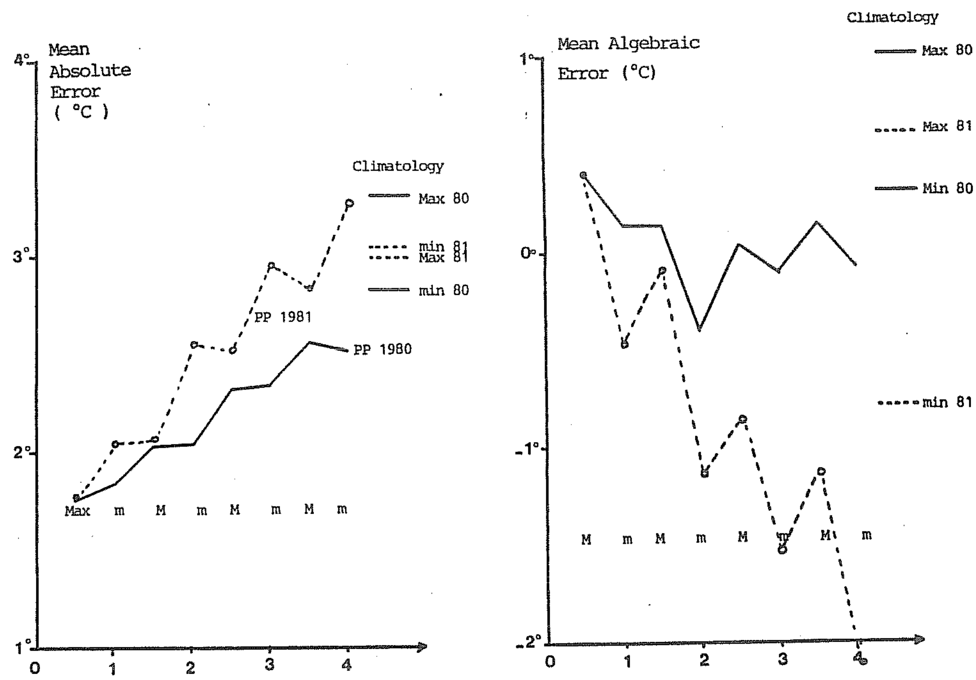

Fig.3 Perfect prog temperature forecast verification

There is a degradation of the results in 1981 due to an increase in a systematic error. The reason for this is partly due to a modification in the radiative scheme of the numerical forecast model, which had little effect on synoptic fields but gave a negative bias to the low level temperature. Since the 850 mb temperature is a very important predictor, this bias is reflected in the statistical forecast. This incident shows clearly how important it is to control systematic errors, which may be sensitive to a minor change in the forecast model.

To suppress this systematic error several procedures have been tested (Hazane and Taboulot 1982). The very simple exponential smoothing procedure has given good results. In this a correction C is applied to the forecast of day J using:

$$C\ (J) = a\ C\ (J-1) + (1-a)\ e\ (j-1)$$

where e is the error.

For operational use a value a = 0.95 has been chosen.

3.2 Medium range forecast

A first experiment in medium range local forecast has been performed using the ECMWF forecasts. The method used is the same as in 3.1; it has been applied to 93 stations in France with forecasts out to day 10. It has been run every day since 15 March 1982. As an example, Fig.4 gives the verification of the spring forecast for Paris-le Bourget. Comparison with the effective climatological forecast shows that the local predictability for temperature reaches 6 1/2 days, which is about what is found for predictability of large scale fields. Also it may be seen that the bias of the ECMWF adaptation is relatively moderate (about -1°C) with marked increase only at the end of the period (after 8 days). The verification at the 93 stations gives equivalent results. It may be concluded from this encouraging preliminary experiment that real skill is obtained for local temperature forecast at middle range by using ECMWF model output.

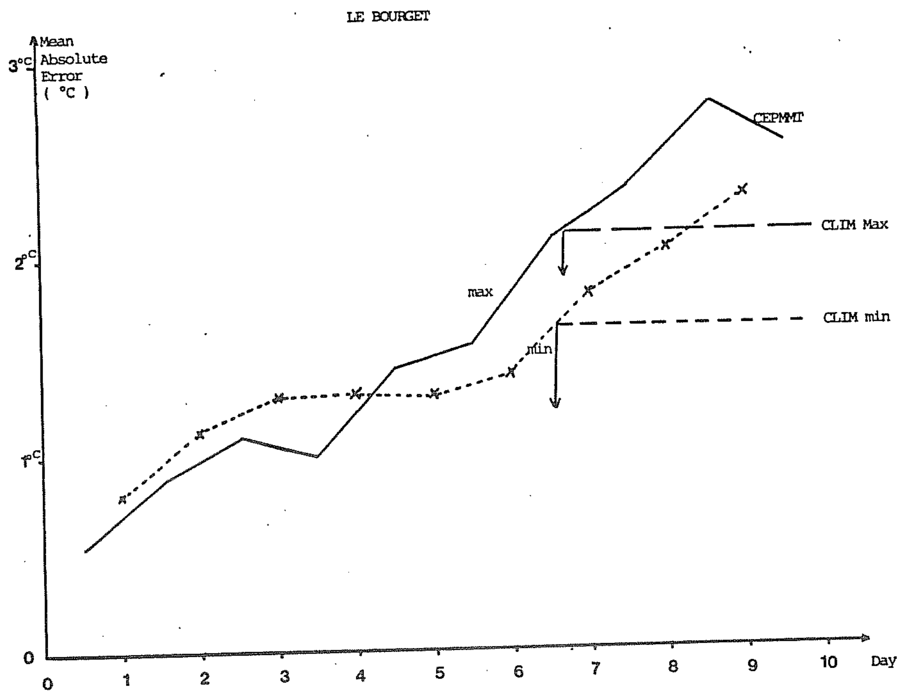

Fig.4a

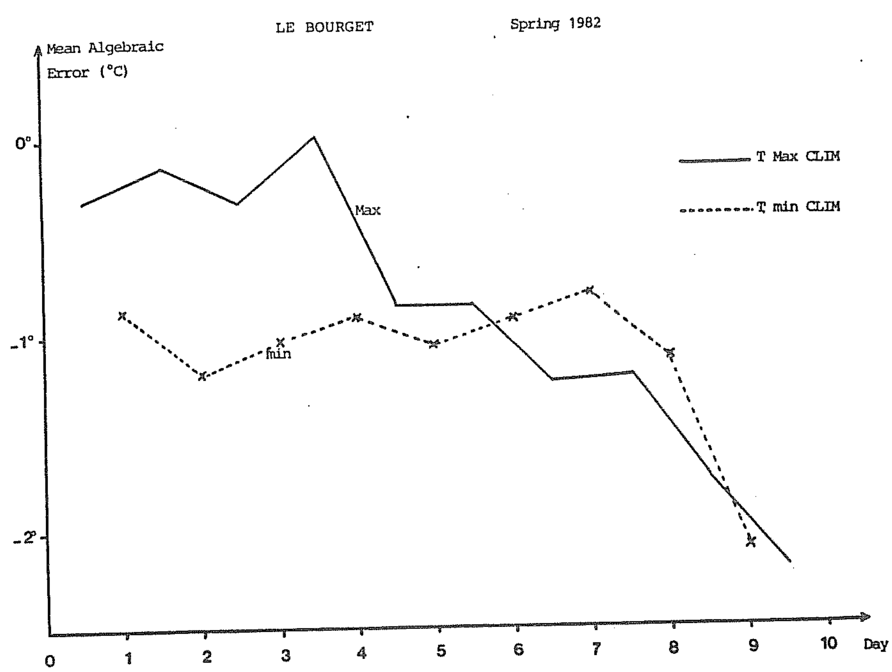

Fig.4b Perfect prog temperature forecast - (ECMWF output)

3.3 M.O.S. method

A method using local grid point predicted values as predictors has been devised. Predictors are meteorological parameters at the 4 grid points surrounding the station for 3 time levels (6 h, 12 h, and 18 h for instance to forecast the maximum temperature), that is 12 possible predictors for one parameter. The following parameters have been considered: vertically integrated relative humidity, 900, 700 and 500 mb relative humidity; 900, 700 and 500 mb temperature; 900-700 and 700-500 mb stability indeces: 900 mb wind components, Laplacian of surface pressure; surface pressure; 900 mb wind velocity. The last local 6 h observations have been added to these parameters: temperature, relative humidity, dew-point temperature, surface pressure, wind velocity and cloud amount. The climatological values are taken into account: cosine (day of year), sine (twice day of year) and climatological temperature. Finally persistence predictors are added: maximum temperature from yesterday and the day before, and yesterday's minimum temperature. A forward stepwise regression selected the best predictors among these 180 potential predictors. As an example Table II gives the predictors used for forecasting the maximum temperature of today and tomorrow and the minimum temperature of tomorrow, for Paris-Le Bourget.

Today's maximum	Tomorrow's minimum	Tomorrow's maximum
Yesterday's maximum	Observed temperature 6 H	Yesterday's maximum
900 mb temp. 18 H Prog. (1)	900 mb temp. 18 H Prog. (3)	900 mb temperature 36 H Prog. (3)
Cosine day of year	Mean rel. humidity 24 H Prog. (3)	Cosine day of year
Observed cloud cover 6 H	Climatological temperature	900 mb zonal wind 36 H Prog. (1)
Observed dew point 6 H	900 mb wind velocity 30 H Proj. (1)	Observed cloud cover 6 H
Yesterday's minimum	900 mb rel. humidity 18 H Prog. (1)	Laplacian of pressure 36 H Prog. (3)
900 mb U wind 12 H Prog. (2)	Cosine day of year	Sine of twice day
900 mb rel. humidity 12 H Prog. (2)	500 mb temperature 18 H Prog. (3)	900 mb temperature 30 H Prog. (1)
500 mb temp. 6 H Prog. (4)	Observed cloud cover 6 H	Laplacian of pressure 30 H Prog. (1)
900 mb temp. 12 H Prog. (1)	900 mb temperature 30 H Prog. (1)	Wind Velocity 30 H Prog. (4)

Table II. Best predictors for PARIS-LE BOURGET (Model Output Statistics). The numbers refer to one of the grid point surrounding the station :

1.	3.
	LE BOURGET
2.	4.

This regression has been computed with all data. Season by season regressions have also been computed: better results have been obtained with the 3 years long "learning" file which covers 3 years. However, the worst results were obtained with the test file (Giraud et Salengro, 1979). Fig.5 gives some information about the quality of the method. It shows the results from the learning file (1975-1976-1977), year 1978 when the operational model was the same as the model used to compute the regression equations and year 1981 when a new numerical model was in operation. For comparison the results of the perfect prog method for year 1981 are included. It can be seen that the change in the model degrades the M.O.S. method, but the method remains competitive at least for the first forecast time (Duvernet and Rousseau, 1980).

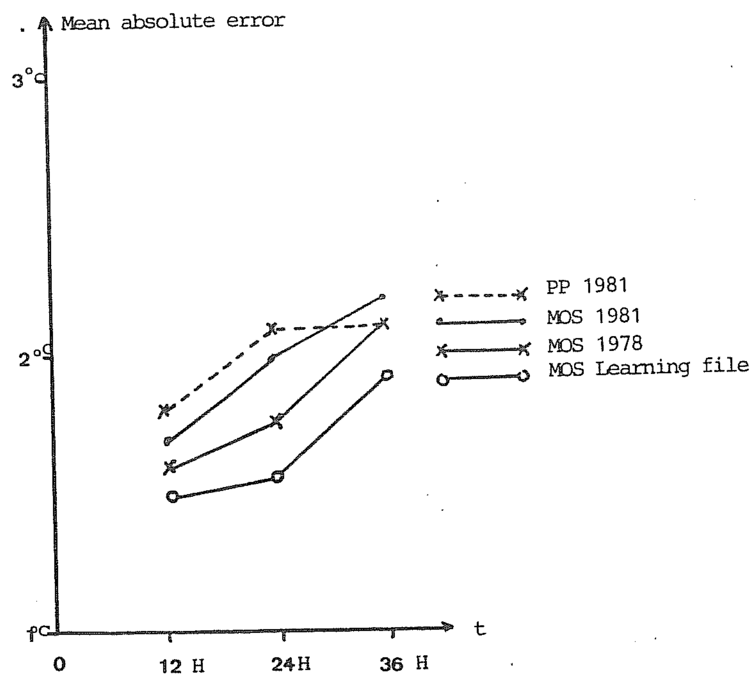

Fig.5 Temperature verification. (Average of 7 stations)

4. PRECIPITATION

4.1 Occurrence of precipitation

The forecast parameter is the occurrence of precipitation at a station during a 12 h period. As in the maximum and minimum temperature forecast, two methods have been studied: a perfect prog method used out to day 4 and a M.O.S. method to forecast for today, tonight and tomorrow.

In the perfect prog method, the initial predictor field and the procedures are identical to those of the temperature forecast except that a parametric linear discriminant analysis is used in place of multiple linear regression and persistence is not used as a predictor (Veysseire, 1980).

In the M.O.S. method the same parametric linear discriminant analysis is used to select the predictors. The predictors are again the values at the 4 grid point surrounding the station for the 3 projections in the forecast period. The potential predictors are 12 values of: vertically integrated humidity, 700-900 mb and 700-500 mb stability parameters 900 mb temperature, 900 mb u and v - wind, Laplacian of surface pressure; then 8 values of precipitation occurrence (0 or 1) at the 4 grid points surrounding the station for two successive 6 h periods and the square root of the amount of precipitation forecast by the model and interpolated to the station location. In all 93 potential predictors have been examined. After selection 8 predictors have been retained (Duvernet and Rousseau, 1980).

The model used to derive the M.O.S. relations is no longer in operation, so that a decrease in the quality occurred when the relations were applied to the new operational model. For the prediction of occurrence of precipitation the derived equations are very sensitive to changes in the model since important predictors such as model predicted precipitation occurrence, vertically integrated humidity or amount of precipitation, are very sensitive to the horizontal and vertical resolution of the model and the parameterisation of convection.

As a criterion for verifying the quality of precipitation occurrence prediction, the following score is used:

$$I = 100 \times \frac{4\ ad - (b+c)^2}{(2a+b+c)\ (2d+b+c)}$$

where a, b, c and d are the elements of the following contingency table.

		OBSERVED	
		Yes	No
Forecast	Yes	a	b
	No	c	d

This score has the property of being zero for random forecast following the global climatology and 100 for perfect forecast (Rousseau, 1980).

Fig. 6 gives the 1981 verification score (average of 7 stations) for perfect prog method and M.O.S. method.

For comparison, the persistence and the numerical model forecast (called AMH) scores are given. It may be seen that the M.O.S. method does not improve the raw model output, because of changes in the model; the M.O.S. equations need to be recomputed with the new model file. However, the perfect prog method gives a significant improvement upon the raw model output; the 4 day perfect prog forecast is better than the 2 day raw model output.

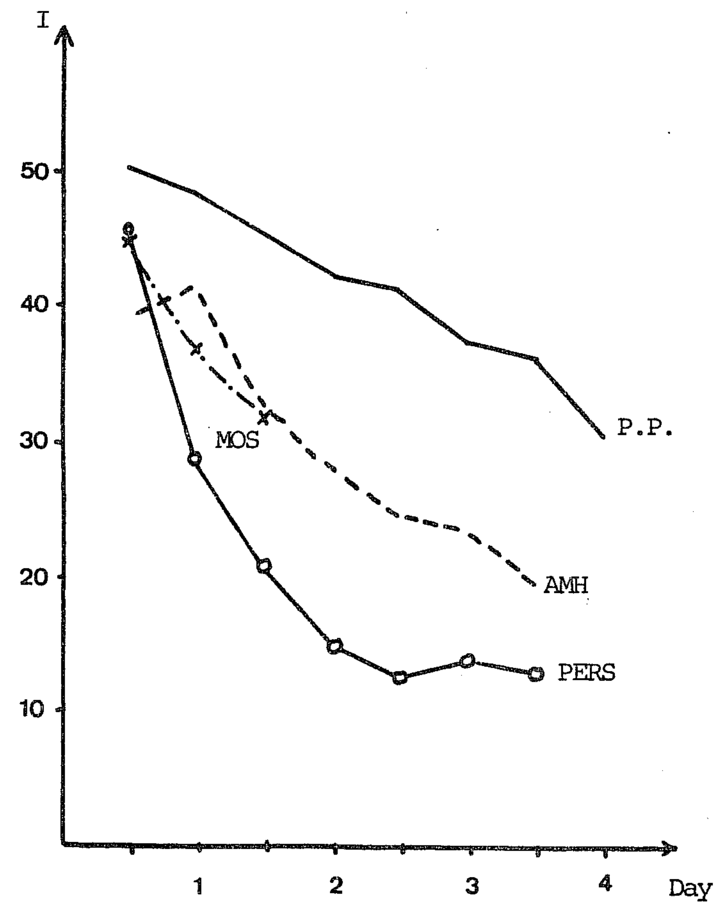

Fig.6 Occurrence of precipitation scores (1981)

4.2 <u>Precipitation amount - analogous situation method</u>

Such a method has been developed at E.D.F. (Electricité de France) and has been in operation for 10 years (Duband 1980, 1982). The predictands are the daily rainfall for 33 catchment areas. A principal component analysis has permitted the definition of 6 principal homogeneous rainfall regions (Fig.7).

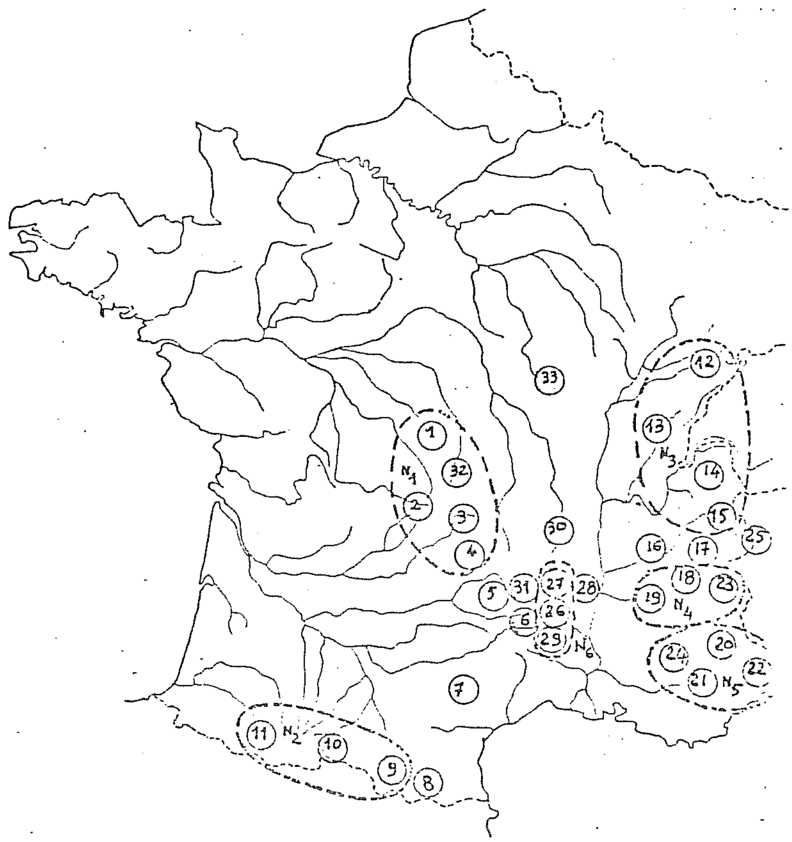

Fig. 7 The 33 catchment areas and the 6 principal homogeneous rainfall regions.

A regression equation calculates the precipitation in each of the 33 areas as a function of the index N for one or more of the regions. The basic file, divided into seasons, contains 700 and 1000 mb geopotential and 1000-700 mb thickness at 37 European radiosonde stations since 1953. Principal component analysis permits the contraction of these 37 data into 6 components. The first selection of analogous situations is done by using a proximity sphere in the 6-dimensional domain of the 700 mb field. To refine the selection a coefficient of correlation R is computed between the six first 700 mb geopotential components, the six first 1000 mb geopotential components, the first 1000-700 mb thickness components of the present day situation with every analogous situation extracted in the proximity sphere. If the coefficient is too small ($R^2 < 0.1$) or if $D^2/R^2 > 4$ (D distance of the situation in the 6-dimensional domain) the situation is rejected. If more than 16 analogous situations are found, a regression is established between $\sqrt{N}$ and the components of the 1000 mb geopotential 24 h variations. If the number of situations is less than 16, the mean and standard deviation of $\sqrt{N}$ is determined in the analogous situations sample. In both cases, the forecast precipitation may be given for the 10%, 50% and 90% probability levels. As an example Fig.8 gives the verification of precipitation forecasts for "today" in winter in region N_3.

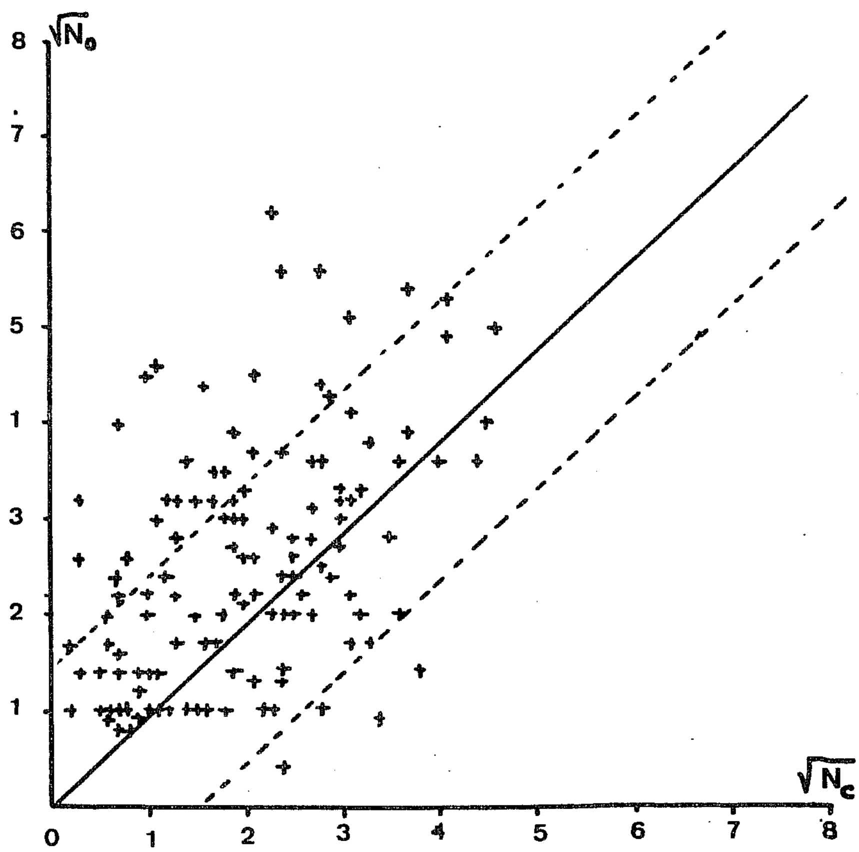

Fig.8 Verification of precipitation forecasts for "today" in winter in region N_3.

The forecast geopotential values necessary for the application of the method are given by the Meteorologie Nationale model. This method is applied operationally for today and tomorrow precipitation forecasts. As a perfect prog method it could be applied to longer forecasts.

Recently, a comparison has been made between this method and a simpler regression method. The same basic data have been used; the predictors are 1000 mb - 700 mb thickness, 1000 mb geopotential and 24 h variation principal components and yesterdays precipitation. The $\sqrt{N}$ value is obtained by a regression equation.

From a comparison for two winters, it has been found that the first method was better for regions N_5 and N_6 but the regression method was slightly better for regions N_1, N_2 and N_3. Also the first method was better for the zero precipitation forecast.

4.3 Precipitation amount. Canonical method

Another pefect prog method has been derived to predict precipitation amounts. The purpose is to forecast the daily value of precipitation amount averaged over each French Department. Therefore, the predictand field is a 90 dimensional vector. In this method the predictand field is reduced by a principal component analysis and the first 30 components are kept as the new predictands.

The basic predictor file is again derived from the "PANAL" file. After a study of the predictand characteristics the file has been divided into two periods: enlarged winter and summer. The potential predictors are the same as for the temperature, but now two consecutive observation times (12 h and 24 h are used) since the amount of precipitation is accumulated from 6 h to 30 h; the wind velocity has been added to the previous predictors and the vertical velocity is used at 0 h, 12 h, 24 h and 36 h. There are 108 potential predictor fields.

Again each field is replaced by a "canonical variable" which is obtained in the same way as for the extreme temperature forecast. Then a forward stepwise selection procedure selects the 8 best canonical variables among the 108 (Nuret and Stangret, 1980; Der Megreditchian et al., 1982); this is done for each of the 30 components.

Table III gives the frequency of selection of the various predictors.

It may be seen that the vertical velocity is indeed the most frequently chosen predictor for this quantitative precipitation forecast.

Correlation coefficients between forecast and observed values have been computed to check the quality of the method. There is a very large variation depending on the location of the Department. The coefficients vary from 0.51 to 0.76 in the 14 seasons long "learning" file and from 0.31 to 0.77 in the 1 season test file. This method will be implemented for operational use.

Predictor	Frequency of selection
Vertical velocity	21%
Temperature	18%
Absolute vorticity advection	18%
Temperature advection	13%
Wind velocity	11%
Geopotential	8%
Meridional wind	6%
Absolute vorticity	3%
Zonal wind	2%

Table 3

5. <u>CLOUD AMOUNT</u>

The purpose is to forecast cloud amount ⩾ 3/8 or cloud amount ⩽ 2/8. A parametric linear discriminant analysis has been used. The predictor file is constituted from the 24 h forecast output file of the ECMWF model. For a pilot study, a two month long predictand file of cloud amount over the United States has been used. There are 58 potential predictors: zonal, meridional and vertical wind, temperature, mixing ratio, geopotential and relative humidity at 8 standard levels (1000, 850, 700, 500, 400, 300, 250, 200 mb), vertically integrated humidity and persistence. The forward stepwise selection has been implemented by using all the stations together; the first 10 selected predictors are the following:

700 mb relative humidity

persistence

700 mb meridional wind

850 mb temperature

300 mb mixing ratio

850 mb humidity

1000 mb temperature

200 mb geopotential

200 mb zonal wind

1000 mb meridional wind

The first estimation of the quality on the "learning" file is the following: 78% successful forecast as compared to 66% for persistence and 66% for forecast using the first best predictor alone (Bernardet, 1982, Personal communication).

Other studies are in progress to apply the method to a homogeneous region selected by automatic classification.

6. CONCLUSION

Perfect prog methods have been tested for the following weather elements: the maximum and minimum temperature, the occurrence of precipitation, the amount of precipitation, M.O.S. methods have been used for the maximum and minimum temperature, the occurrence of precipitation and the cloud amount forecasts.

Some comparisons have been made between the two kinds of methods. Although the M.O.S. method may be better when applied under optimum conditions (no change in the numerical model), the operational constraints may deteriorate the quality of the forecasts. Perfect prog methods may also be sensitive to model modification (change in systematic errors). The first application of perfect prog method for medium range forecast has shown that useful information may be given up to 6 1/2 days for local temperature forecasts.

Developments of these methods at the national office and in regional offices (precipitation and avalanche forecasts in the Alps for instance) are now in progress.

Acknowledgement

This paper summarizes the main works done in France in the field of statistical adaptation for local forecasts. I would like to thank in particular G. der Megreditchian and J.P. Javelle from the Météorologie National and D. Duband from E.D.F. who gave me all elements for the preparation of this summary.

References

Der Megreditchian, G. et al, 1981: La prévision statistique des précipitations. La Houille Blanche 7/8 1981; 487-495.

Duband, D. 1980: Dynamic selection of analogue flow pattern to produce quantitative precipitation forecasts. Proceedings of the WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting, Nice 1980 487-492.

Duband, D. 1981: Prévision spatiale des hauteurs de précipitations journalières. La Houille Blanche 7/8 1981, 487-492.

Duvernet, F. and Rousseau, D. 1980: Statistical interpretation of numerical weather prediction in the French Meteorological Service. Proceedings of the WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting, Nice 1980, 397-400.

Giraud, G. and Salengro, R. 1979: Adaptation locale des prévisions par methode statistique. Application à la prévision de la température minimale et maximale. Rapport de stage - Météorologie Nationale. Paris.

Hazane, P. and Taboulot, S. 1982: Algorithms statistiques de détection et de correction automatique des erreurs dans un modéle de "prévision parfaite". Rapport de stage - Météorologie Nationale. Paris.

Javelle, J.P. et al, 1980: Opérational forecasting of local weather parameters through statistical interpretation of NWP using "Perfect prog" method. Proceedings of the WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting, Nice, 1980, 423-430.

Javelle, J.P. et al, 1981: Statistical and Dynamical Adaptation of the forecasts given by the French P.E. Model.

Nuret, M. and Stangret, R. 1980: Approche statistique de la prévision d'un champ de précipitations sur la France. Rapport de stage. Météorologie Nationale. Paris. 152 + pp 184.

Rousseau, D. 1980: A new skill score for the evaluation of yes/no forecasts. Proceedings of the WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting Nice 1980, 167-174.

Veysseire, J.M. 1980: PROJET AMETHYSTE Cahier N°6. Adaptation statistique des sorties du modèle AMETHYSTE pour le prévision locale de températures extrêmes et d'occurences de précipitation par la méthode de prévision parfaite. Note technique de l'EERM n°81. Direction de la Météorologie Nationale. Ministère des Transports. pp.68.