THE EXPERIMENTAL FORECAST FOR AIRPORT TOTAL CLOUD COVER USING THE LINEAR AND NONLINEAR MODULES

H. F. Wang*, B. Huang* **, J. W. Liu*, Y.B.Hu*, J.An*, X.T.Wang*
*** Beijing Aviation Meteorological Institute, China, ** Physical Institute Peiking University**

Keywords: total cloud cover, explanation forecast, BP neural network,support vector machines

Abstract

It is necessary to develop a forecasting method which can resolve nonlinear problems because there are limitations in statistic forecasting methods mostly based on linear correlation when dealing with complicated nonlinear problems. Airport total cloud cover categorical forecast modules were developed applying the nonlinear methods including the back propagation neural network (BPNN), support vector machines (SVM) etc. and the multiple linear regression method, respectively, with the T213 medium-range numerical forecast products between 2003 and 2005. The experimental forecast was applied with the datasets in 2006 using the linear and nonlinear forecast modules, respectively and the forecast accuracy of them was compared. The results show that the forecast accuracy of two nonlinear forecast modules applied BPNN and SVM is better than that of linear forecast module applied multiple linear regression when using the same factor screening method. While the grading of total cloud cover adds to more than four classes, the forecast accuracy of SVM forecast module is better than the BPNN module.

1 Introduction

It is necessary to develop a forecasting method which can resolve nonlinear problems because there are limitations in statistic forecasting methods mostly based on linear correlation when dealing with complicated nonlinear problems. As a representative of the linear forecast systems, MOS is widely used, well developed and skillful. It is running in many weather service organizations in different countries. Based on the raw forecast model output, MOS is able to correct for model bias and take into account some educed variables which are unobservable, such as vertical velocity, stress, and flux. However, in the operational forecast, it is expensive for developing a multi-station and multi-period of validity MOS system since the numerical models which provides the predictors change frequently. Therefore, the construction of statistic forecast model needs collection of the new forecast model products for long period.

Many new technologies are explored and developed for explanation forecast: (1) Recursive updating method. The characteristics of new model are obtained as early as possible and the parameters are updated based on the new datasets. These kinds of method used in the statistic forecast for weather elements is Kalman filters and continuously updated MOS. (2) PP method. The module which is developed using PP method with analysis and observing data needn't be developed a new model when the numerical model changed and don't need as many equations as MOS do. NOAA evaluates the PP method and the conclusion is that PP method is more poor than MOS method for short-range forecast but is better than the MOS method for the medium-range (three to six days) forecast. In all 1990s, operational MOS forecast products are mostly replaced by PP forecast products. But in the PP forecast method the model deviation and derived variable from numerical model is not considered. (3) Revised method for dealing with predictors. The nonlinear correlated model variable, for example, the quartic of relative humidity, is introduced as a new predictor. The typical correlation analysis reflecting all the information of field is introduced. (4) Nonlinear module. The linear equation can not reflect the nonlinear relation between the predictand and predictors. So the direction is that nonlinear relation is introduced and the BPNN and SVM methods are used to make module.

The BPNN method is superior to traditional linear regression method for the nonlinear problems. It is helpful in the atmospheric research, especially for nonlinear problems in the weather forecast. Caren Marzban[1] developed neural networks forecasting modules for cloud ceiling and invisibility prediction for 39 airports by BPNN method. Data from hourly surface observations and output from the model MM5 over the period 2001—2005 are combined. It is found that the performance of the neural networks is generally superior to logistic regression and MOS, especially at specific station with sufficient observations. Dustin Fabbian[2] applied BPNN to fog forecast at Caberra international airport in Australia based on 44 years' observation data. The BPNN was found to have good performance at various forecast periods.

The statistic learning theory which is firstly put forward by Vapnid et al. is a small sample theory. It is accepted as the optimal theory for the classification and regression for small samples, since it avoids some intrinsic limitations within the BPNN method, such as the uncertainty of the network construction, over-learned, under-learned and local minimum problems. In recent years, the SVM method which is based on the statistic learning theory provides a new solution for the nonlinear problem. By a nonlinear mapping function, SVM is able to map the sample space to a high dimensional feature space or infinite dimensional feature space. Therefore, nonlinear classifications and regressions in the feature space can be solved by linear learning algorithm.

Airport total cloud cover categorical forecast modules were developed applying the nonlinear methods including the back propagation neural network (BPNN), support vector machines (SVM) etc. and the multiple linear regression method, respectively, with the T213 medium-range numerical forecast products between 2003 and 2005. The experimental forecast was applied with the datasets in 2006 using the linear and nonlinear forecast modules, respectively and the forecast accuracy of them was compared. The results show that the forecast accuracy of two nonlinear forecast modules applied BPNN and SVM is better than that of linear forecast module which applied multiple linear regression when using the same factor screening method. While the grading of total cloud cover adds to more than four classes the forecast accuracy of SVM forecast module is better than the BPNN module.

2 Datasets Description and Preprocessing

2.1 Datasets Description

There are two types of datasets involved in this study. One is the hourly surface observation data in 2006 of Beijing airport, the other is T213 Medium-range forecast model products over the period 2003-2006 from China Meteorological Administration. The T213 model products have a horizontal resolution of 1.125° by 1.125°.The vertical coordinate is divided into 13 layers: 1000hPa, 925hPa, 850hPa, 700hPa, 600hPa, 500hPa, 400hPa, 300hPa, 250hPa, 200hPa, 150hPa, 100hPa and 50hPa.

2.2 Classification of Predictors

The predictand is the total cloud cover at 3-hour intervals within 24 hours at single station during summertime (from April to September). According to the airport forecasting criteria, the total cloud cover can be classified as 5 conditions. Forecast model is established for each classification.

Table 2-1 Classification of total cloud cover

Classification	Standard	Expression
2(fine day)	0-3, 4-10	1, 0
2(cloudy day)	0-7, 8-10	0, 1
3	0-3, 4-7, 8-10	1, 2, 3
4	0-3, 4-6, 7-9, 10	1, 2, 3, 4
11	0,1,2,3,4,5,6,7,8,9, 10	0,1,2,3,4,5,6 ,7,8,9,10

2.3 Selection of Predictors

Predictors include the total cloud cover at the initial time and 16 variables of T213 model output at 3-hour intervals, such as temperature, geopotential height, wind components u and v, vapor mixing ratio, vertical velocity, relative humidity, sea level pressure, vorticity, convergence, surface air temperature, surface pressure, 10 m wind components u and v, 2 m temperature, and 2 m humidity. Based on these variables, 13 diagnostic variables closely related to cloud formation are calculated, including vertical vapor transportation term (SQ1 and SQ2), total energy term (ZNL), atmospheric stratification parameter（such as K index, total layer vapor saturation term ZC, mid-low level vapor term (ZDC1 and ZDC2）, ascending motion parameter (upward motion term SSYD and spiral motion term LX), relative humidity between 700hPa and 500hPa, mean relative humidity between 1000hPa and 500hPa, temperature difference between 1000hPa and 850hPa, temperature difference between 850hPa and 700hPa[3]· Among these parameters, 36 variables are selected as initial predictor(Table 2-2).

Table 2-2 36 Initial predictors

Number	Code	Physics	Number	Code	Physics
1	h850	geopotential height at 850hPa	20	ζ700	relative vorticity at 700hPa（×105）
2	h700	geopotential height at 700hPa	21	ζ500	relative vorticity at 500hPa（×105）
3	h500	geopotential height at 500hPa	22	Pr	sea level pressure
4	ω850	vertical velocity at 850hPa	23	K	K index
5	ω700	vertical velocity at 700hPa	24	LX	Spiral motion term
6	ω500	vertical velocity at 500hPa	25	SQ1	vertical vapor transportation at 500hPa
7	V850	V component wind at 850hPa	26	SQ2	vertical vapor transportation at 700hPa and 850hPa
8	V700	V component wind at 700hPa	27	SSYD	ascending motion term
9	V500	V component wind at 500hPa	28	ZC	summary of difference between air temperature and dew point temperature at 500hPa, 700hPa and 850hPa
10	Rh850	relative humidity at 850hPa	29	ZDC1	mid-low level vapor term
11	Rh700	relative humidity at 700hPa	30	ZDC2	mid-low level vapor term
12	Rh500	relative humidity at 500hPa	31	ZNL	total energy term
13	T850	temperature at 850hPa	32	Cloud00	total cloud cover at 20h
14	T700	temperature at 700hPa	33	Rh700-500	mean relative humidity between 700hPa and 500hPa
15	T500	temperature at 500hPa	34	T850-1000	temperature difference between 850hPa and 1000hPa
16	U850	U component wind at 850hPa	35	Rh1000-500	mean relative humidity between 1000hPa and 500hPa
17	U700	U component wind at 700hPa	36	T700-850	temperature difference between 850hPa and 700hPa
18	U500	U component wind at 500hPa	educed variables are underlined		
19	ζ850	relative vorticity at 850hPa（×105）			

Canonical correlation analysis (CCA) has been developed as a new statistic approach recent years which can be used to find the pattern of highly correlation between two fields. In this section, CCA is performed between the selected predictors field and the total cloud cover filed within a coverage about 20°×20°around the specific station. Then, we derived 35 canonical parameters. Fig.2-1 shows the comparison of the correlation coefficients between the canonical parameters, interpolation parameters (except for the predictor Cloud00), and the total cloud cover at 03:00. The results show that canonical parameters which are combined all the information of the field have good performance to present the variation of the total cloud cover. Each correlation coefficient between the canonical parameters and the predictand is above

0.4. The maximum coefficient is 0.7, and mean coefficient 0.55. Each canonical parameter has better correlations with predictand than the interpolation parameter.

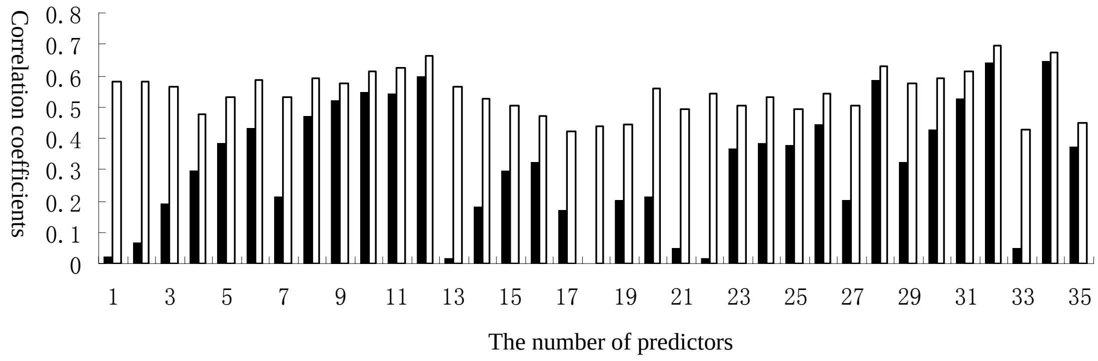

Fig. 2-1 Contrast of correlation coefficients between interpolation parameters, canonical parameters and the daily total cloud cover at 23:00 at summertime at Beijing airport station

Considering different period of validity and the rationality and stability of the algorithm, stepwise regression analysis is performed to the selected predictors, till the optimal predictors are found.

Table 2-3 The final selected canonical factors for the total cloud cover forecast at summertime.

Classification of the total cloud cover	The number of final selected canonical parameter
2(fine day)	1 7 10 12 16 21 24 36
2(cloudy day)	1 2 5 7 12 16 21 29 30 33
3	1 2 7 12 16 21 24 29
4	1 7 12 14 16 21 28 33
11	1 2 7 12 16 21 29 30 33 36

3 Construction of the Total Cloud Cover Forecast Module

Output of T213 medium-range forecast over the period 2003-2005 is utilized to construct the total cloud cover forecast model for 3-, 6-, 12-, and 18-h head times from 2000 local standard time for each classification case.

3.1 Multiple Linear Regression Forecast Model(MLR)

Least square method is used to determine the regression coefficients and significance test is performed by F distribution at 5% level of significance. Based on the final choice of the parameters, the total cloud cover is classified as 2, 3, 4, and 11 grades. For each graded, the linear regression forecast model for the total cloud cover is constructed.

3.2 Artificial Neural Network Forecast Model

The 3-layer back propagation neural network (BPNN) is utilized[4], and the number of hidden nodes is taken over the range from 3 to 10. The hidden level uses Tansig transport function, while linear transport function is used for output level. LM algorithm is adopted as learning algorithm by virtue of its rapid convergence, high precision and smaller mean square error. The process of learning, training and emulation is performed via MATLAB tools.

3.3 Support Vector Machines (SVM)

The radial radical kernel function and cross check are utilized to grid searching for two parameters—C and γ. In order to determine the optimal C and γ, grid searching is firstly performed on the coarse grid, then on a finer grid. Ultimately, the forecast model of the cloud fraction classification can result from the employment of the training process.

4 Comparison between Linear and Nonlinear Forecast Module

Comparison is performed based on data in the summertime in 2006 between the linear and nonlinear forecast module.

Forecast accuracy is taken as follows:

$$TT = \frac{N_{r_k}}{N_{f_k}} \times 100\% \qquad (1)$$

Here, N_{r_k} is the times of correct prediction for certain period of validity (times when the prediction of total cloud cover classification is same to observations), N_{f_k} is the total times for certain period of validity.

4.1 Comparison between Linear and Nonlinear Forecast Model of Different Classification Support Vector Machines

It can be seen from fig.4-1 to fig. 4-4 that dealing all five kinds of classification the nonlinear model is more skillful than linear model for the short-term forecast with period of validity within 24 hours. For those less than three-class, BP shows the best performance, with forecast accuracy above 70% for all periods of validity. While for those more than four-class, SVM is better. However, the linear forecast model shows the forecast accuracy below 69% when classification is above three-class.

For the two-class, all the three models display high (above 70%), the accuracy is above 80% when the period of validity is within 15 hours. In this case, SVM is not better than linear regression model.

When the predictand is taken as 3 classes, the two nonlinear models are superior to linear regression model (MLR). Forecast accuracy of the MLR is below 60% after 15 hours, however, BP and SVM represent good performance with the forecast accuracy above 65% within 24 hours. For the period of validity within 9 hours, nonlinear model has improved the forecast accuracy upon the linear model to a extent of 5%-9%. Besides, when the period of validity is more than 9 hours, the forecast accuracy of the nonlinear models has increased to 10%-18% against the linear model.

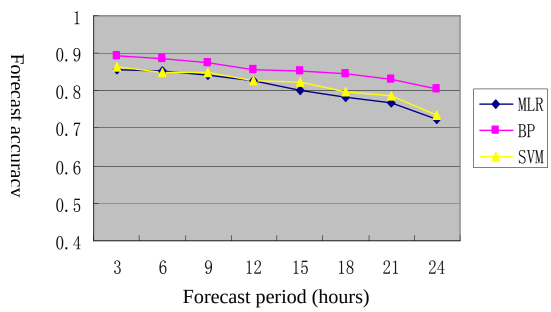

Fig. 4-1 Comparison among the three forecast models for two-class.

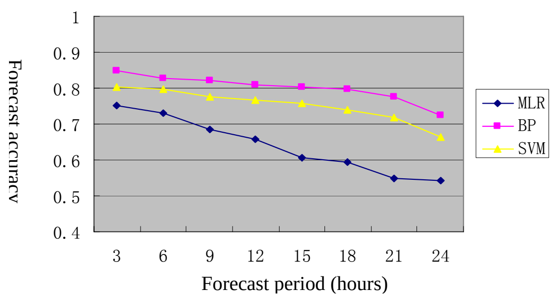

Fig. 4-2 As Fig. 4-1, but for three-class

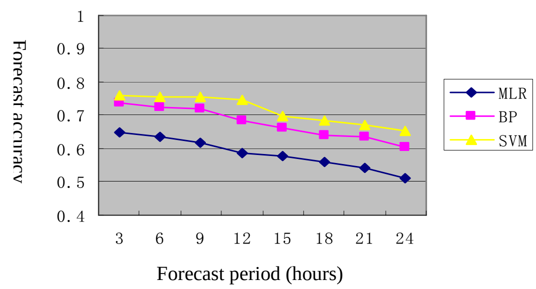

Fig. 4-3 As Fig. 4-1, but for four-class

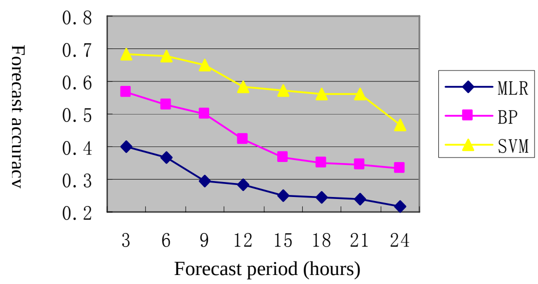

Fig. 4-4 As Fig. 4-1, but for eleven-class

Under the condition of four-class, the forecast accuracy of the three models drops obviously. For MLR, the forecast accuracy is less than 60% after 9 hours. On the other hand, BP and SVM seem to be more skillful with the forecast accuracy above 70% within 9 hours and above 60% within 24 hours. In general, forecast accuracy of the nonlinear models is 8%-16% higher than the linear model.

For the eleven-class, poor performance has been found from all three forecast model. The forecast accuracy of MLR and BP is less than 30% and 50%, respectively. At all periods of validity, the forecast accuracy of BP is 10% more than the linear forecast model. Nevertheless, SVM has the forecast accuracy 10%-20% higher than BP. The SVM displays good potential ability in the cloud cover forecast, whose forecast accuracy above 60% in 9 hours and above 55% within 21 hours.

4.2 Comparison between Linear and Nonlinear Forecast Model of Different Periods of Validity

On a whole, the forecast accuracy both of the nonlinear and linear forecast models drops along with the periods of validity.

If 70% is taken as a standard level of the forecast accuracy, the linear model can make three-class forecast within 9 hours, but two-class forecast beyond 9 hours.

There is little difference between the forecast accuracy of BP and SVM. Both of the two models can give four-class forecast within 9 hours, while three-class forecast beyond 9 hours.

Especially, SVM is able to give four-class forecast within 15 hours and two-class forecast beyond 15 hours, with the forecast accuracy above 70%. Moreover, it also can give four-class forecast with the forecast accuracy above 65%.

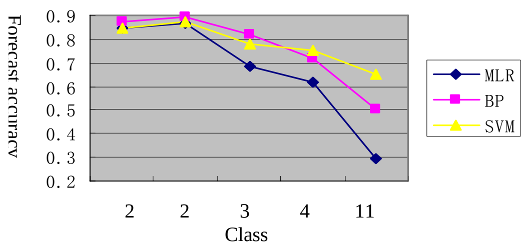

Fig. 4-5 Comparisons among the three models for period of validity within 9 hours.

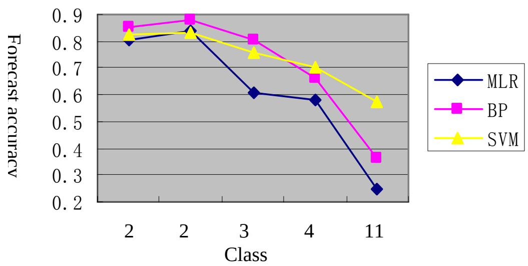

Fig. 4-6 As Fig. 4-5, but for 15 hours

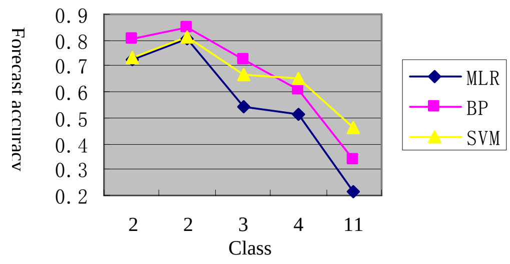

Fig. 4-7 As Fig. 4-5, but for 24 hours.

5 Conclusion

We can get the following conclusion:

(1) The forecast accuracy of two nonlinear forecast modules applied BPNN and SVM is better than that of linear forecast module applied multivariate linear regression when making 24 hours short-range forecast according to five kinds of grading of total cloud cover using the same factor screening method. According to the grading of total cloud cover more than four classes the forecast accuracy of SVM forecast module is better than the BPNN module. The forecast accuracy of linear and nonlinear forecast modules decrease when the period of validity increases.

(2) The forecast accuracy of three forecast modules in period of validity less than 24 hours is well good (more than 70%) and is more than 80% in period of validity less than 15 hours according to two classes of total cloud cover. The forecast accuracy of SVM is comparative to that of one linear regression module (MLR). Only in the period of validity less than 6 hours the forecast accuracy of MLR is more than 70% according to three classes of total cloud cover.

(3) According to two and three classes of total cloud cover the forecast accuracy of BPNN is better than that of other two modules and is more than 70% in all period of validity less than 24 hours. The forecast accuracy of SVM is better than that of other two modules when the classes of total cloud cover are more than 3.

(4) The method of combining the correlation analysis and the nonlinear modeling is a new and effective technology for explanation forecast of station total cloud cover.

References

[1] Caren M, Stepthen L and Brad C. Ceiling and visibility forecast via neural networks.*Weather and Forecasting*, Vol. 22,pp 466-479,2007.

[2] Dustin F, Richard D.D and Stepthere L. Application of artificial neural network forecasts to predict fog at Canberra international airport. *Weather and Forecasting*, Vol. 22,pp 372-381,2007.

[3] Barrs, J. A., and Mass C. F., 2005，Performance of national weather service forecasts compared to operational, consensus, and weighted model output statistics. *Weather and Forecasting*, Vol. 20, pp 1034-1047, 2005.

[4] Zhou K.L., Kang Y.H. *Neural network model and MATLAB simulation program design*. Qinghua University Press, 2006.

Copyright Statement