Probabilistic Precipitation Forecasting via Post processing of an Ensemble Output over Iran

Saeed vashani, Majid Azadi

***Abstract*— An attempt is made to obtain calibrated probabilistic numerical forecasts of 24-hour accumulated precipitation over north of Iran, using artificial neural network (ANN) and rank-histogram calibration methods. The forecasts were obtained from an eight-member ensemble using three limited area models of WRF and MM5 used five and two times respectively with different configurations. Initial and boundary conditions are obtained from the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS). In order to remove the systematic error in the deterministic output of each member in the raw ensemble, each member's output was first postprocessed using the ANN technique (E1). Results show that the ANN technique is successful in removing the systematic error in the precipitation forecasts of each member in the raw ensemble such that mean absolute error in the precipitation forecasts are decreased from 1.8 mm to 1.4 mm, from 4 mm to 2 mm and from 4.2 mm to 2.2 mm for the first, second and third day of forecasts. Then rank-histogram calibration method was then applied on the output of E1 to obtain the calibrated probabilistic forecast (E2). Statistical scores including Brier score calculated for the raw ensemble, E1 and E2 show significant improvement is in the reliability of the probabilistic forecasts, for example, the amount of BS for raw ensemble 0.42 decreased to 0.29 for using both E1 and E2 for the second forecast day in precipitation less than 0.1 mm.**

***Index Terms*— Artificial neural network, calibrated probabilistic forecast, rank-histogram.**

I. INTRODUCTION

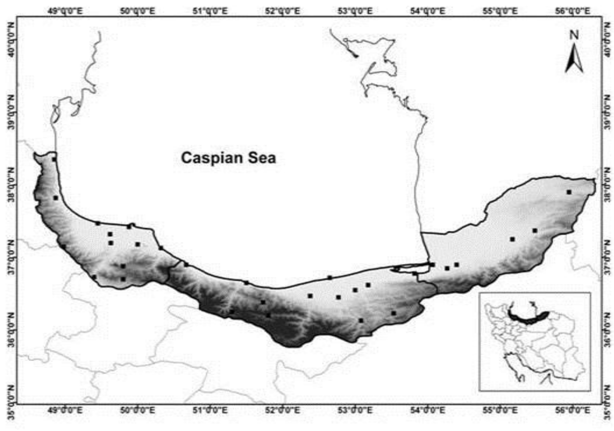
Accurate quantitative precipitation forecasts (QPFs) have been always a demanding and challenging job in numerical weather prediction (NWP). The outputs of ensemble prediction systems (EPSs) in the form of probability forecasts provide a valuable tool for probabilistic quantitative precipitation forecasts (PQPFs). But the ensemble biases in the form of under dispersion or over dispersion, aroused mainly from deficiencies in models physics and less than optimum ensemble initial perturbations, limits their more effective use. In the last couple of years various statistical methods such as artificial neural network (ANN)[1], logistic regression[2]-[3], Bayesian model averaging[4], non-homogeneous Gaussian regression[5] and Gaussian ensemble dressing [6]-[7], Rank histogram calibration[8], among others, have been developed for postprocessing the raw EPSs outputs. Reference 9 shows successfully applied ANN to correct temperature forecast and found that ANN technique outperforms the Kalman filter in removing the temperature systematic error.

In this paper, the ANN technique as used in [1], and ran-histogram calibration method proposed by Hamill and Cloucci in [8], hereafter HC98, have been used to calibrate the output of a multi model EPS to produce calibrated PQPFs, over north of Iran. Fig. 1 shows the area of study in the northern part of Iran. This region is almost a uniform region and most of the precipitation over Iran occurs in this region. Maximum amount of annual precipitation is this region is exceeding 1900 mm year.

II. DATA

The data used in this study consists of 24-hour accumulated precipitation measured at 33 irregularly spaced synoptic meteorological stations scattered in the northern part of the country from first November 2008 to 30 April 2009 and corresponding 72-hour numerical prediction of precipitation from eight members of the ensemble system, bilinearly interpolated to the observation sites.For producing PQPFs over north of Iran, 72-hour ahead forecasts of the Weather Research and Forecasting (WRF) [10] model with five different configurations and the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5) [11]-[12], with three different configurations, have been used to build an eight member ensemble. The model settings are presented in Table 1 and Table 2. As seen in the table the main differences between different model setups pertain to convective and boundary layer parameterization schemes The initial and boundary conditions come from the operational 1200 UTC runs of global forecasting system (GFS) of NCEP (National Center for Environmental Prediction). The integration period goes from first November 2008 to the 30 April 2009 (182 days).

Fig. 1. location stations



Geographical location of the 33 stations in north of Iran

Manuscript received November 21, 2014.

Saeed vashani Assit. Professor, Department of agriculture , college of agriculture ,Islamic Azad University, Varamin-Pishva Branch, Varamin, Tehran, Iran.Tel: 00989122844563

Majid Azadi Associ. Professor, Atmospheric Science and Meteorological Research Center (ASMERC), Tehran, Iran Tel: 00989127054765

Table 1. configuration of model WRF

Different configuration derived from model WRF (members 1-5).

No.	Micro physics	LW_RA	SW_RA	Surface	PBL	cum ulus
1	Ferrier	RRTM	CAM	RUC	YSU	KF
2	WSM6	RRTM	Dudhia	Thermal	MYJ	KF
3	WSM5	RRTM	Dudhia	Noah	YSU	KF
4	WSM5	RRTM	Dudhia	Noah	MYJ	KF
5	Lin	RRTM	Goddard	Noah	MYJ	KF

Table 2. Configuration of model MM5

Different configuration derived from model MM5 (members 6-8)

No. member	Microphysics	cumulus	PBL	LW_RA
6	Dudhia	KF	ETA	RRTM
7	Dudhia	Grell	MRF	RRTM
8	Dudhia	KF	MRF	RRTM

Both WRF and MM5 are used with non-hydrostatic option and were run with two nested domains, with the larger domain covering the south-west middle east from 10°N to 51°N and from 20°E to 80°E and the smaller domain covers Iran from 23°N to 41°N and from 42°E to 65°E. The spatial resolutions are 45- and 15-Km for the coarser and finer domains respectively. Forecasts out to +72 hour ahead from the inner domains have been used to form the raw ensemble forecasts.

III. Methods

A. Artificial Neural Network

Reference [1] shows a feed-forward ANN with one input layer of neurons, one hidden layer, and one output layer was used to post process the QPF of each individual member. In the first layer a sigmoid activation function and in the second and third layers linear transfer functions were used. For each of the 33 station locations used in this study, model forecasts consisting of quantitative precipitation; 1000-, 850-, and 500-hPa air temperature (K); 1000-, 850-, and 500-hPa vertical velocity(m s^{-1}); 1000-, 850-, and 500-hPa relative humidity(%); 1000-, 850-, and 500-hPa specific humidity (kg kg^{-1}); 1000-, 850-, 700-, and 500-hPa geo potential height (m) were bilinearly interpolated to the station locations to provide 17 inputs (predictors) to the NN. The ANN method was applied for each station and member separately. The output of the NN for each station location is bias-corrected 24-h accumulated QPF up to 72 hours ahead.

B. Rank histogram calibration

The bias correction method using ANN described above, effectively removed the systematic biases but the corrected probability forecasts might still not be reliable. In order to get calibrated probabilistic precipitation forecasts, the rank histogram calibration technique proposed by Hamill and Colucci in [8] was implemented both on the raw and postprocessed (using ANN) ensemble forecasts. This method uses the information in the raw ensemble rank histogram in the training period to establish higher reliability in probabilistic precipitation forecasts. Fig. 2 shows an example of the rank histogram for the raw ensemble. As seen, the shape of the rank histogram is highly non uniform and under dispersive, such that, due to systematic errors in the forecasts, around 50% of the times the verifying observation falls outside of the range of forecast values. Using this fact the subsequent forecasts of the ensemble can be better interpreted and calibrated. Since distribution of one rank histogram calculated from the past forecasts in the training period might not be not representative of all the subsequent forecasts and verifying observation, generally more than one rank histograms are used depending on the ensemble variability. Based on the value of the standard deviation of the ensemble about its mean, s, two different rank histograms were constructed (in the training period) and used (in the test period) for low ($s < 0.45$) and high ($s > 0.45$) variability in the ensemble. Suppose vector X represents the N sorted ensemble precipitation forecasts, V the verifying observation and vector **R** the N+1 ranks in the representative verification rank histogram distribution (representing the relative frequency of verifying observation in the bins), the probability of precipitation for a quintile q is then estimated as:

$$\Pr(v \le q) = \sum_{j=1}^{i} R_j + R_{j+1}\left(\frac{q - x_i}{x_{i+1} - x_i}\right) \qquad x_i \le q \le x_{i+1} \qquad (1)$$

Following in [8] the following assumptions are used to calculate probabilities for q falling outside the range of all ensembles forecast values:

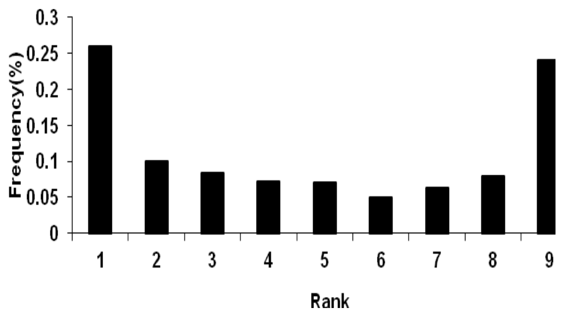


Fig. 2. Rank Histogram

Rank histogram distribution for precipitation forecast with 8 members

1. For q smaller than the lowest N ensemble forecasts, a uniform distribution between zero and the lowest ensemble member is assumed and the probability is thus estimated as

$$\Pr(0 \le v \le q) = \left(\frac{q}{x_1}\right) R_1 \qquad 0 \le q \le x_1 \qquad (2)$$

2. For q falling in the upper tail, i. e. larger than the highest N ensemble forecasts, the rank histogram is assumed to follow a Gumble distribution and the probability is thus estimated as

$$\Pr(x_N \le v \le q) = \frac{F(q) - F(x_N)}{1.0 - F(x_N)} R_{N+1} \qquad q \succ x_N \tag{3}$$

Where F denotes the fitted Gumble distribution. For more explanation on the above mentioned method, referred to [8].

1. **Verification procedure**

For comparing the deterministic forecasts associated with individual member forecasts, the mean absolute error (MAE) is calculated over all the test period and over all the 33 observation locations. It is calculated as:

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|o_k - f_k\right| \tag{4}$$

A commonly used verification measure for probabilistic forecasts is the Brier score [13], which is essentially the mean squared error of the probabilistic forecasts and is defined as the average of the differences between the forecast probability and the corresponding binary observation:

$$BS = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2 \tag{5}$$

where, y_k is the forecast probability and o_k is the corresponding binary observation, assuming that o_k =1 if the observed precipitation exceeds an established threshold, and o_k =0 if it does not and k is the index number of the forecast/event observation pair.BS ranges between 0 and 1 and is a negatively oriented score with values close to 0 indicating better forecasts. BS was evaluated for raw ensemble, post processed ensemble using ANN method, post processed ensemble using HC method and post processed ensemble using HC&ANN methods for mentioned thresholds from forecast days1-3.

IV. Results

A. Estimation of the training period for ANN

To establish an optimum length of the training period for ANN, several experiments with varying training periods from 20 to 80 days were performed. Fig. 3 shows the MAE for the member-1 forecasts after postprocessed using ANN. As seen from the Fig. 3, for training periods less than around 40 days the MAE decreases with increasing the number of days used as training period. Beyond 40 days the MAE remains about constant.

Similar results (not shown here) are obtained for other ensemble members. Therefore we chose a window of 40-days for as training the ANN.

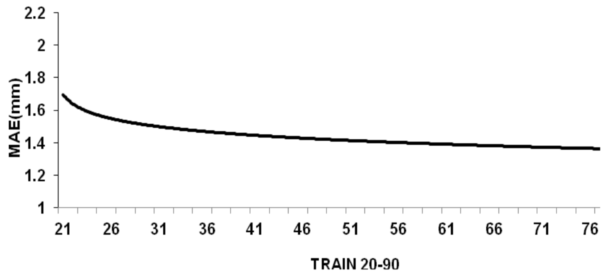


Fig. 3. Training period
MAE of post processed QPF using ANN for member-1 with training period 20-80 days for 33 stations.

B. Deterministic forecasts

Fig. 4 presents the MAE calculated for each member forecasting the raw ensemble for 1-3 days forecasts of precipitation. As seen, the MAE was generally lower for first forecast day than for second and third forecast days. All members were nearly equal in keeping MAE under 2 mm for the first forecast day, and performed nearly equally well in keeping MAE between 3.9 and 4.2 mm for second and third forecast days. The difference between the lowest MAE for member-8 (1.58 mm) and the highest MAE for member-4 (1.79 mm) is about 0.2 mm for the first day. It is thus clear that there is no much deference between the raw ensembles members forecast.

Output of each member in the raw ensemble for precipitation forecasts was postprocessed using ANN with 17 predictors as described in section 2. Fig. 5 shows the calculated MAE for each of the post processed forecasts for1-3 forecast days. Examining the Fig. 5 reveals that the value of MAE for the postprocessed forecasts ranged between 1.4 to 1.5 mm, 2.3 to 2.6 mm and 2.6 to 2.7 mm for first to third day of forecasts respectively. Again, there is no significant difference between MAE of different postprocessed forecasts members after implementing ANN, but member 5 shows better by a slight margin for the first day of forecast.

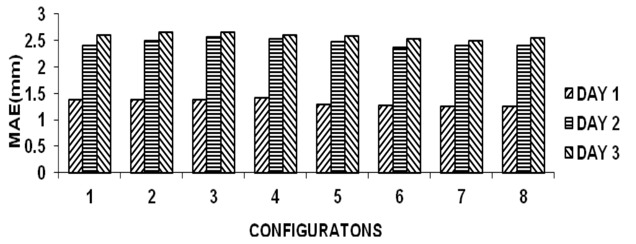


Fig. 4. MAE calculated for the raw ensemble
MAE of QPF for eight members for all forecast periods of days 1-3.

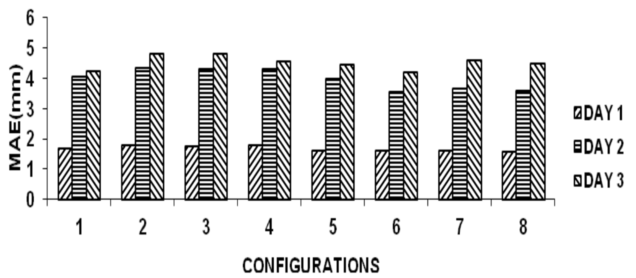


Fig. 5. Calculated MAE after postprocessed QPF
MAE of post processed QPF using ANN for all forecast periods of days1-3.

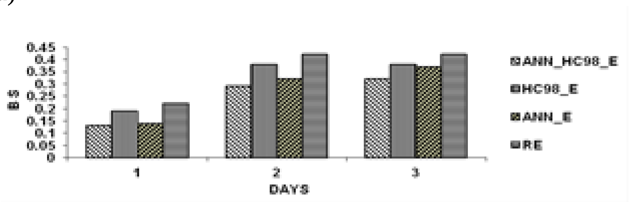
C. Probabilistic forecasts

1) Brier Score (BS)

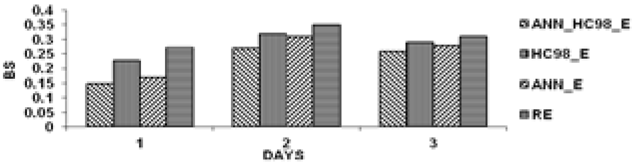
Fig. 6 presents the BS for probabilistic forecasts from RE, ANN_E, HC98_E and ANN_HC98_E. As seen in the Fig. 6 there is a significant increase in BS value from first forecast day to third forecast day in all the ensemble system used. The calculated value of BS for RW is always higher compared to the other ensemble systems for all precipitation thresholds and forecast days. After implementing the ANN, Fig. 6 shows significant improvement in quality of the forecasts almost for all thresholds and forecast days. For example, The BS calculated for precipitation less than 0.1 mm for the RW and

ANN_E are 0.21 and 0.13 respectively for the first forecast day. The BS calculated for HC98_E shows a small but consistent increase when compared to ANN_E. In other words, using ANN was more effective than using HC on the RW in our case. The best BS score is obtained when the each raw ensemble member forecast is first postprocessed using ANN and then the HC98 method is implemented to get probabilistic forecasts. Results of BS for ANN_HC98_E are most effective compared to all the ensemble configurations and for all forecast days for all thresholds considered here.

a)



b)



c)

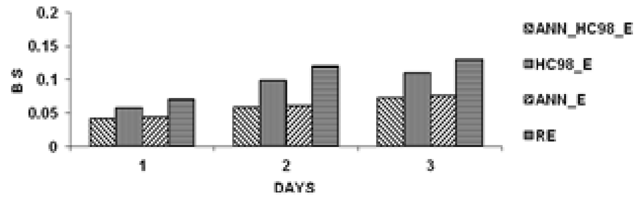


Fig. 6. Brier Score

calculated BS the for raw ensemble (RE), post processed ensemble using ANN method (ANN_E), post processed ensemble using HC method(HC98_E) and post processed ensemble using HC& ANN methods(ANN_HC98_E) for precipitation less than 0.1 mm(a),between 0.1 mm to 10 mm(b) and more than 10 mm(c) from forecast days1-3.

2) Relative value (RV)

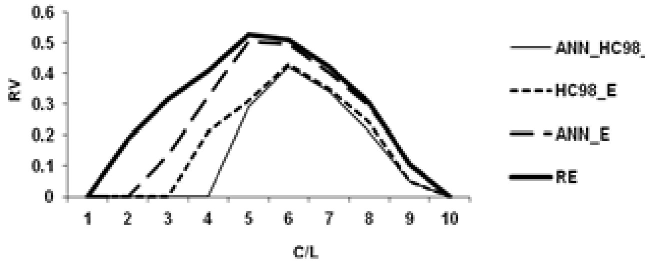
Another way of evaluating both deterministic and probabilistic forecasts is through the use of economic value analysis of the forecasts. References [14] - [17] show that for most forecast events and for most users the probabilistic forecasts offer higher economic impact on potential users than the deterministic forecasts from a higher resolution model. In this section the verification results of economic value analysis for the probabilistic forecast of four different ensemble forecasts considered in this study are presented. The relative value, RV, of a forecast system can be defined as the reduction in mean expense relative to the reduction that would be obtained by having access to perfect forecasts [16] .Fig. 6 presents the calculated RV versus cost-loss ratio (CL^{-1}) for 24-h probabilistic forecasts and three different precipitation thresholds. Larger area under the RV curve means higher economic value for potential users. It is to be mentioned that a negative RV is considered zero on the graphs. It is seen that for values of CL^{-1} close to both 0 and 1 the RV is zero. This means that potential users with small values of CL^{-1} should always take protective action and on the contrary those with large values of CL^{-1} should never take protective action, or the forecasts have no value for users with very small and large cost of protective actions. Only, where RV is positive the user can make a decision based on the forecast.

Examining Fig. 7 shows that value of RV calculated for ANN_HC98_E and RW are consistently highest and lowest respectively for all potential users and all three precipitation thresholds. It is seen also, that ANN technique is more effective compared to HC98 in getting forecasts with higher economic values. Similar results (not shown here) are found for other forecast ranges.

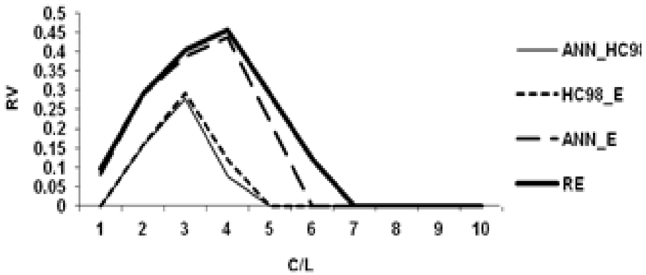
V. Conclusion and Discussion

In this paper we used the artificial neural network ANN and rank histogram calibration method for of output of deterministic and post processing of ensemble forecasting system to get probabilistic precipitation forecast in north of IRAN from period a November 2008 to 30 April 2009.

a)



b)



c)

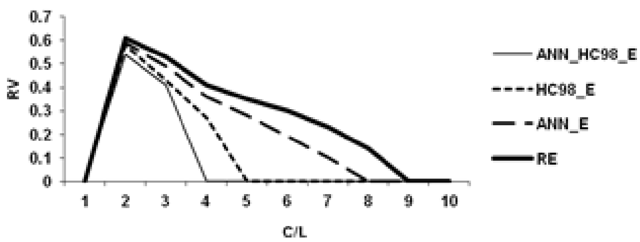


Fig. 7. Relative Value

The RV calculated raw ensemble (RE), post processed ensemble using ANN method (ANN_E), post processed ensemble using HC method (HC98_E) and post processed ensemble using HC& ANN methods (ANN_HC98_E) for precipitation less than 0.1 mm (a), between 0.1 mm to 10 mm (b) and more than 10 mm(c) from forecast days1-3.

Totally the conclusions of this research show that ANN could decrease the error of raw ensemble so that the MAE for the first day of forecast achieve under 1.5 mm and for second and third forecast days is about 2.5 mm. Clearly the results was obtained in first forecast day is better than the next.

In term of MAE , all members errors are similarly for all forecast days, but it seems that the members related to the MM5 model(members 6,7,8) produce the better forecasts, while after using the post processing methods the result of MAE are nearly similar. BS was calculated for RE, ANN_E, HC98_E and ANN_HC98_E for mentioned thresholds from forecast days1-3. Having performed ANN method, the forecast quality increased significantly. for example, the amount of BS for raw ensemble 0.42 decreased to 0.32 for post processed ensemble using ANN method for the first forecast day in precipitation less than 0.1 mm. also the BS calculated before and after using rank histogram method proposed by HC98 shows the increasing of probabilistic forecast quality such that the amount of BS for raw ensemble 0.42 decreased to 0.29 for post processed ensemble using both ANN and HC98 for the second forecast day in precipitation less than 0.1 mm.

The RV was evaluated as a measure of forecast value. Forecast value is related to the cost that user will pay if he uses forecast is making decisions. The results show that value of calibrated probabilistic forecast is more than uncalibrated one. The increasing of forecast value post processed using ANN&HC98 is very well (Fig. 7).

Briefly the selection of different configurations does not have much effect on decreasing error and difference between observation and DMO increases from the first to the third forecast days in all members. The ANN and HC98 as two post processing methods can significantly decrease the systematic error of DMO, but the ANN method can remove systematic error better than the HC98 method. We can produce more accurate probabilistic forecast using ANN for raw ensemble output and the calibrating the post processed output using HC98 method.

VI. Acknowledgments

The authors express their appreciation to I. R. Iran Meteorological Organization (IRIMO) for their financial support.

References

[1] W. Yuval and H. William,"An adaptive nonlinear MOS scheme for precipitation forecasts using neural networks," Wea. Forecasting., 2003, vol. 18, pp. 303-310.
[2] W. M. Brelsford, and R. H. Jones, "Estimating probabilities," Mon. Wea. Rev., 1967, vol. 95, pp. 570-576.
[3] D. S. Wilks, and T. M. Hamill, "Comparison of ensemble-MOS methods using GFS reforecasts," Mon. Wea. Rev. 2007, vol. 135, pp. 2379-2390.
[4] A. E. Rftery, T. Gneiting, and F. Balabdoui, "Using Bayesian model averaging to calibrate forecast ensembles," Mon. Wea. Rev., 2005, vol. 133, pp. 1155-1174.
[5] T. Gneiting, A. E. Rftery, A. H. Westved, and T. Goldman,"Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation," Mon. Wea. Rev., 2005, vol. 133, pp. 1098-1118
[6] M. S. Roulston, and L. A. Smith. "Evaluating probabilistic forecasts using information theory," Mon. Wea. Rev., 2002, vol. 130, pp. 1653-1660.
[7] X. Wang, and C. H. Bishop, "Improvement of ensemble reliability with a new dressing kernel," Quart. J. Roy. Meteor. Soc. 2005, vol. 131, pp. 965-986.
[8] T. M. Hamill, and S. J. Colucci, "Evaluation of Eta-RSM Ensemble Probabilistic Precipitation Forecast," Mon. Wea. Rev., 1998, vol. 126, pp. 711-724.
[9] S. Vashani, M. Azadi, and S. Hajjam, "Comparative Evaluation of Different Post Processing Methods for Numerical Prediction of Temperature Forecasts over Iran,"Res. J. Environ. Sci. 2010, vol. 4 pp. 305-316.
[10] W. C. Skamarock et al., "A description of the Advanced Research WRF," Version 3. NCAR Tech. Note NCAR/TN-4751STR, 2008, 113 pp
[11] J. Dudhia, "A nonhydrostatic version of the Penn State/NCAR mesoscale model: Validation tests and the simulation of an Atlantic cyclone and cold front," Mon. Wea. Rev. 1993, vol. 121, pp. 1493-1513.
[12] G. A. Grell, J. Dudhia, D. R. Stauffer, "A Description of the Fifth-generation Penn State/NCAR Meso Scale Model (MM5)," NCAR Tech. Note NCAR/TN-3981STR, 1994, 117 pp.
[13] G. W. Brier, "Verification of forecasts expressed in terms of probability," Mon. Wea. Rev., 1950, vol. 78, pp. 1-3.
[14] K. R. Mylne, "The use of forecast value calculations for optimal decision making using probability forecasts," Seventieth Conf. on Weather Analysis and Forecasting, Denver, CO, Amer. Meteor. Soc., 1999, pp. 235-239.
[15] Y. Zhu, Z. Toth, R. Wobus, D. Richardson, K. Mylne, "The economic value of ensemble based weather forecasts" Bull. Amer. Meteor. Soc., 2002, vol. 83, pp.73-83.
[16] D. S. Richardson, "Skill and relative economic value of the ECMWF Ensemble Prediction System," Quart. J. Roy. Meteor. Soc., 2000, vol. 126, pp. 649-668.
[17] K. K. Lee, and J. W. Lee, "The economic value of weather forecasts for decision-making problems in the profit/loss situation," Meteor. Appl., 2007, vol. 14 pp. 455-463.

Saeed vashani received Ph.D. in meteorology from azad Islamic university branch of science &research in 2010. He is teaching as assistance professor in department of agricultural in azad Islamic university branch of varamin-pishva. He has presented few papers in journals. His area interest of is numerical prediction, artificial neural network.

Majid azadi received Ph.D. in meteorology from jawaharlal nehro university in 2001. He is working as associate professor in Atmospheric Science and Meteorological Research Center (ASMERC). He has presented many papers in national, international conferences and journals. His area interest of is numerical prediction, probabilistic forecasting.