MULTI-CRITERIA SELECTION OF HYDRODYNAMIC PREDICTORS IN A STATISTICAL LONG-RANGE FORECAST SCHEME

A.Muraviev, I.Kulikova, E.Kruglova
Hydrometeorological Research Centre of the Russian Federation
Bol. Predtechensky per., 11-13, 123242 Moscow, Russia
muravev@mecom.ru, kulikova@mecom.ru, kruglova@mecom.ru

By a ***hydrodynamic predictor*** a GCM output product is meant, which is used in the statistical long-range monthly mean forecast scheme based on a Perfect Prognosis (PP) approach and operationally run in the Hydrometeorological Center of the Russian Federation (*Muraviev, 2001*).

The diversity and abundance of hydrodynamic predictors necessitates their ranking, or ***ordering***, on the basis of skill scores both for verification and optimal forecast scheme construction. One of the approaches to tackle the problems is the theory of multi-criteria decision making (*Brussilovsky, 1986; Noghin, 1997*). Here ***solutions*** $\mathbf{X}=(x_1,\ldots,x_n)$, $n>1$, are the final forecasts under evaluation, and ***attributes*** $f_1,\ldots,f_m$, $m>1$, comprise the quality criteria vector $\mathbf{F}=(f_1,\ldots,f_m)$.

In the multi-criteria technique every k-th attribute has its own ***preference relation*** R as a subset of the Descartes product $\mathbf{X}\times\mathbf{X}$, built under the condition of ***linear order*** in the set of all solution pairs: $R_k=\{(x_i,x_j) \in\mathbf{X}\times\mathbf{X} : f_k(x_i) \gtrless f_k(x_i)\}$. The inequality sign between the criteria values corresponds to the *preference* defined. Every relation R_k may be rewritten in the form of a $n\times n$-***matrix of preference*** $M_k=\{\mu_k(x_i,x_j)\}$, composed of units and zeros in correspondence to belonging $(x_i,x_j)\in R_k$ or $(x_i,x_j)\notin R_k$, respectively. Let us denote the corresponding linear order via $\boldsymbol{r_k}$.

If the ***distance*** between two orderings $\mathbf{r_1}$ and $\mathbf{r_2}$ is defined by the formula

$$d(r_1,r_2) = 0.5\cdot\sum_{i\neq j}|\,\mu_1(x_i,x_j)-\mu_2(x_i,x_j)\,|,$$

we may obtain the final ordering of solutions $\mathbf{r_0}$ (***Kemeny median***) through the equation

$$\sum_{i=1}^{m} d(r_0,r_i) = \min_{r}\sum_{i=1}^{m} d(r,r_i).$$

The solution set $\mathbf{X}$ is composed of statistical monthly surface air temperature forecasts at 120 stations of the former USSR. The multiple regression coefficients are estimated with the help of the temperature series (*VNIIGMI archive*, *2007*) and reanalysis data (*Kanamitsu et al,2002*), for the period 1974-2005.

The 500 hPa heights and air temperatures at 850 hPa obtained from the GCMs are used as initial hydrodynamic fields in the PP-procedure. Two spectral global models of the T41L15 and T85L31 types, and the semi-Lagrangean model SLAV were used for generating the hydrodynamic predictors. The resulting station temperature values from the three model outputs were also averaged and evaluated as a separate scheme (ENSEM).

The diversity of the predictors was provided by two ***regression bases*** (5 and 10 days averages) and by different boundary conditions in the SST fields in the T41L15 integrations (statistical forecast – *frc*, persistence of the previous month anomaly – *per*, and climatic values – *cli*).

The vector criterion $\mathbf{F}=(\rho$, Q, MSSS$)$ is composed of three scores: the anomaly sign correlation coefficient ρ, the relative anomaly forecast error Q standardized by station temperature variances and the mean squared skill score MSSS with respect to the climate forecast.

Three main problems were aimed in using the multi-criteria approach: (1) optimal choice of the SST field and the regression base for the T41L15 integration, (2) forecast verification for the models and their post-processed average over the test period, and (3) construction of an adaptive forecast scheme using the Kemeny median with an evaluation of the approach.

The results for the first two problems are shown in the Table. The selection was performed among three models and their ensemble as well as only among the models. The multiple selected SSTs, predictors and regression bases for one initial date may be explained by the non-strict linear order in the relation R.

As it is seen in the Table the preferred SST in the T41L15 integrations in most cases is the persistence of the previous month anomaly with no distinct regression base.

In selection of the models and the post-processed average the SLAV model may be preferred, whereas the inclusion of the model ensemble shifts the regression base definitely to the preferred 10 days averaging period.

Table

The multi-criteria selection of hydrodynamic predictors (the first table line) using the vector quality evaluation of monthly surface air temperature forecasts for the stations of the former USSR.

	T41L15		T41L15, T85L31, SLAV, ENSEM		T41L15, T85L31, SLAV	
initial date	***SST***	***regr base***	***PREDICTOR preferred***	***regr base***	***PREDICTOR preferred***	***regr base***
20070927	*frc, per*	10	SLAV	10	SLAV	10
20071030	*cli*	5	T85L31, SLAV	5	T85L31, SLAV	5
20071129	*per*	5	SLAV	5	SLAV	5
20071226	*per*	5	T85L31	10	T85L31	5
20080130	*cli, frc, per*	10	SLAV, ENSEM	5, 10	SLAV	5
20080227	*frc, per*	10	ENSEM	10	SLAV	5, 10
20080330	*per*	10	SLAV	10	SLAV	10
20080428	*per*	5	T85L31	10	T85L31	5
20080529	*cli, per*	5	ENSEM	10	T41L15	10
20080629	*frc*	10	T41L15	10	T41L15	10
20080730	*per*	10	T41L15	10	T41L15	10
20080830	*per*	10	T41L15, SLAV	5, 10	T41L15, SLAV	5, 10
20080929	*cli, frc, per*	5	SLAV	10	SLAV	5
20081030	*frc, per*	10	T85L31	10	T41L15, SLAV	10
20081129	*cli, frc, per*	10, 5	SLAV	10	SLAV	10
20081228	*per*	10	T41L15, SLAV	10	T41L15, SLAV	10

The most simple adaptive forecast technique was tested based on the multi-criteria selection for the next month. The averaged monthly mean air temperatures, obtained with the Kemeny median over the test period, yielded a poor skill: ρ = 0.31, Q =1.30, MSSS =0.03. But the study of some score curves gives the impression that the adaptive approach decreases risks in possible forecast failures, as shown in Figure.

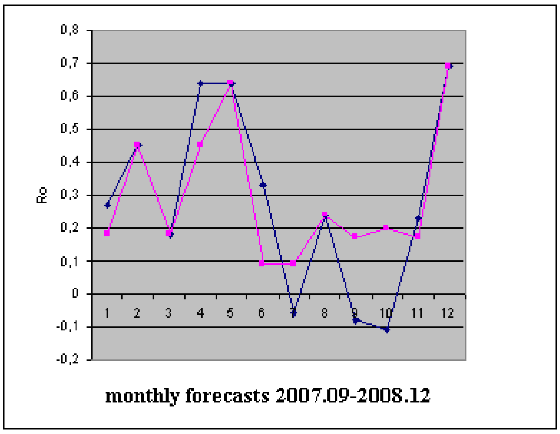

Figure. Anomaly sign correlation coefficient ρ for the SLAV predictors (blue) and the Kemeny-median persisted forecasts (pink).

1. Brussilovsky P.M. Multi-criteria forecast verification and the aggregation problem for alternative forecasts // The Russian J. of Meteorology and Hydrology, 1986, 10, p.18-24 (in Russian).
2. Kanamitsu M., Ebisuzaki W.I, Woollen J., Yang Shi-Keng, Hnilo J.J., Fiorino M., and Potter G.L. NCEP–DOE AMIP-II Reanalysis (R-2) // Bull. Amer. Met. Soc., 2002, vol. 83, 11, p.1631–1643.
3. Muraviev A. Hydrodynamic-statistical long-range prediction scheme: practical stability and predictability study. The first APEC Climate Network Working Group Meeting, 17-18 May 2001, Seoul, Republic of Korea, 91-105.
4. Noghin V.D. Relative importance of criteria: a quantitative approach // J. Multi-Criteria Decision Analysis, 1997, vol. 6, p.355-363.
5. VNIIGMI-archive, Russian Institute for Hydrometeorological Information – World Data Center, (Obninsk), 2007, http://meteo.ru/english/.