Section 6

Developments in global forecast models, case studies, predictability investigations, global ensemble, monthly and seasonal forecasting

Medium-Range Ensemble Prediction at the Hydrometcenter of Russia

E.D. Astakhova and A.Yu. Bundel

Hydrometcenter of Russia, Moscow, Russia, helen@mecom.ru

An ensemble system for medium-range global weather forecasting has been developed at the Hydrometcenter of Russia. The system is based on the spectral T85L31 model, presently used for operational deterministic medium-range forecasting (Frolov et al., 2004). The ensemble is formed of 13 members — 12 perturbed forecasts and a control (unperturbed) forecast. Only uncertainties inherent in initial conditions are taken into account. Initial state perturbations are calculated using the breeding method (Toth and Kalnay, 1997). The 12-h breeding cycle is applied. Temperature, wind, and surface pressure are perturbed at each level and at each model gridpoint. The starting perturbations are determined as a difference between an analysis and a 12-h forecast valid for the same date. The perturbations are symmetrically added to/subtracted from the analysis data forming initial conditions for a pair of perturbed forecasts. Next perturbations are determined as a half-difference between the results of the pair of 12-h perturbed forecasts. The perturbations are scaled down at each step of the breeding cycle using the global total energy norm.

The ensemble system is implemented on one node of a four-node computer based on Quad-Core Intel Xeon5345 (2.33 GHz) processors (2 per node) in Linux. Although the model used in the runs is parallelized using MPI, it is still running in a one-processor mode because only 2 PCs are attributed to the job.

The system runs in quasi-operational mode from April 2008. Because of the limited computer resources, 10-day ensemble forecasts are issued only once a day for 12 UTC. The forecasts starting from 00 UTC are run for 12 h. The ensemble forecasts of 850-hPa temperature, mean sea level pressure, 500-hPa geopotential height, and convective, large-scale, and total precipitation are accumulated in a 40-day database.

The analysis of the EPS results shows that the ensemble mean is obviously better than the control run for lead times exceeding 96 h (Fig. 1).

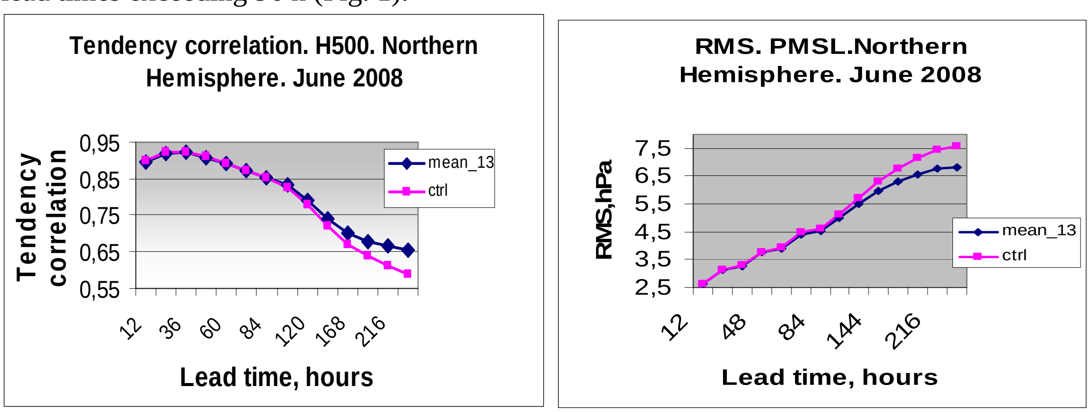


Figure 1: A comparison of verification scores for the ensemble mean and control forecasts

Varying the ensemble size from 9 to 13 members only slightly affected the skill of the ensemble mean. However, spaghetti maps clearly show that the ensemble range is insufficient and often does not embrace the analysis thus indicating the necessity of a larger ensemble.

A special study was made for 6-h precipitation totals. The verification period was from April 2 to September 29, 2008, 171 days in total. The station observations averaged over the model grid cells were used for comparison. The deterministic scores included a set of measures based on the contingency table (in particular, the Peirce skill score), the RMSE, and some others. Figure 2 shows that the ensemble mean outperforms individual ensemble members and that the precipitation quality is better in the day time.

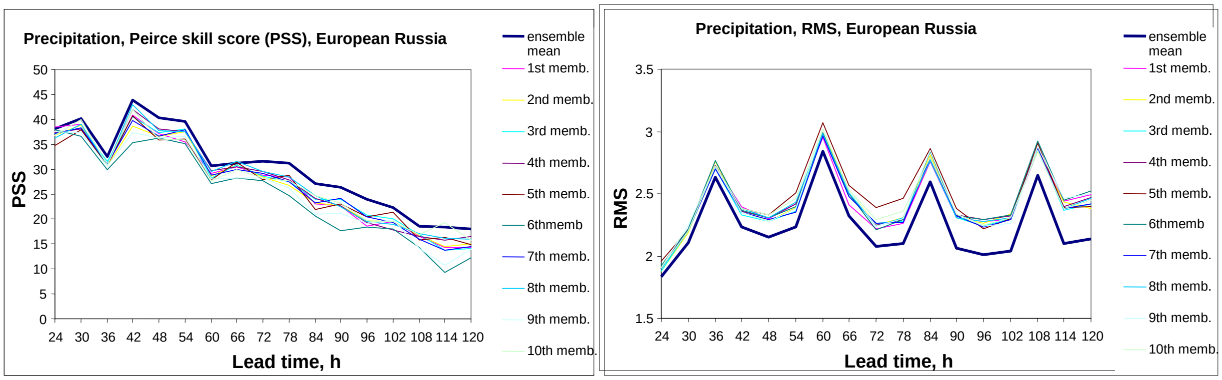


Figure 2: Deterministic scores for 6-h precipitation forecasts for ensemble mean and different ensemble members.

The probabilistic measures included the ROC curves and the area under the ROC curve, the Brier score, reliability diagrams, and frequency histograms. The binary events were precipitation exceeding 0.1; 1; 2.5; and 5 mm/6h. Figure 3 shows that the precipitation forecast in European Russia is better than in Western Siberia and that the intense precipitation is more difficult to predict. Overall, the probabilistic forecast is good up to three days. The reliability diagrams showed that the system is overconfident in predicting precipitation events or their absence.

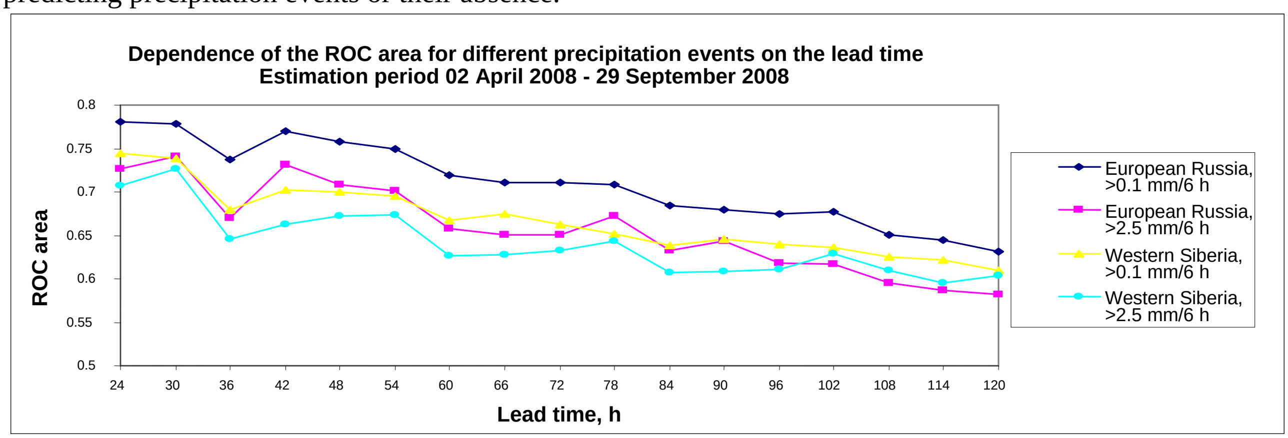


Figure 3: ROC areas for different precipitation thresholds and regions

An increase in the ensemble size and an improvement of the model resolution will be possible this year after the ensemble system is implemented on a new powerful computer recently installed at the Hydrometcenter of Russia. Introduction of another model, namely, the semi-Lagrangian model (Tolstykh, 2001), to the ensemble is planned as well.

Acknowledgements

The study was supported by the Russian Foundation for Basic Research (project no. 07-05-13607-ofi_ts).

References

1. Frolov, A.V., Astakhova, E.D., Rozinkina, I.A., et al., *Russian Meteorology and Hydrology*, 2004, 5, 5-20.
2. Toth, Z. and Kalnay, *Mon. Wea. Rev.*, 1997, 125, 3297—3319.
3. Tolstykh, M.A. Semi-Lagrangian high-resolution atmospheric model for numerical weather prediction. *Russian Meteorology and Hydrology*, 2001, 4, 1- 9.

Seasonal forecasting and daily probability

Michel Déqué and Jean-Philippe Piedelievre
Météo-France/CNRM, CNRS/GAME, Toulouse France (deque@meteo.fr)

Seasonal prediction is not a deterministic prediction. One can also consider that a 24h forecast contains a probabilistic part, in particular as far as local extreme events are concerned. But seasonal forecast is a climatology forecast. One attempts to predict some statistical properties of one or several seasons in advance, but not the chronology. Predicting the date (or rather the week) of the monsoon onset, even with a probability aspect, is a meteorological forecast. It should be considered as monthly or extended-range forecast, not as seasonal forecast. The validity of seasonal forecasting has first be proved in deterministic approach. Indeed, if the deterministic approach fails completely, the probabilistic approach cannot work. The probability theory does not create any information, it expresses our partial knowledge which leads to imperfect deterministic prediction. In order to validate the deterministic prediction in a robust way, scientists have used root mean square error or anomaly correlation of the seasonal mean. When going to probabilistic prediction (Doblas Reyes et al., 2000), they attempted to predict the seasonal mean with a probability density function (pdf).

From a scientific point of view, this approach is partial, because the model (or the nature) offers a succession of daily situations which form the weather (if chronology is taken into account) or the climate (if only the statistical properties are considered). From a user point of view, the seasonal average does not represent a useful information, in particular if it is given by a pdf. For heating or agriculture, degree days above or below a threshold, number of consecutive dry days ... are more useful. One case in which seasonal average is the needed product is management of water in big dams: the seasonally accumulated precipitation is the right input variable in the decision making process.

EUROSIP is a consortium born after the DEMETER European project. Three partners (ECMWF, Met Office and Météo-France) agree to produce each month at the same date 41 members of a 7-month forecast. A hindcast period (to evaluate model climatology and prediction scores) is based on 1987-2007 with 11 member ensembles. Here we examine 2 m temperature forecasts for December January February obtained from the early November issue.

Verifying a probability forecast is less straightforward than a deterministic one. A good choice is to use square differences because they are additive with respect to time and space. The square difference between observed and predicted pdf could have been a simple and good idea, but in the case of temperature and precipitation, it does not take into account the fact that they are ranked variables. For example if the predicted and observed pdfs are quasi-deterministic (Dirac distributions), an error for temperature of 1°C or an error of 10°C yield the same distance. For this reason, the distance between the cumulative density functions (cdf) is preferred. Such a distance is named ranked probability score (rps). It is often used for a sall number of categories, traditionnally quantiles:

$$\mathrm{rps} = \frac{1}{n}\sum_{i=1}^{n}\left[\mathrm{prob}(T_{pre} < Q_i) - \mathrm{prob}(T_{obs} < Q_i)\right]^2 \tag{1}$$

where Q_i are the quantiles of the climatological distribution. Remember that here we deal with daily data of a season, so observation is also probabilistic. The continuous form (corresponding to $n \to \infty$) is:

$$\mathrm{rps} = \int_{-\infty}^{+\infty} \mathrm{pdf}_{cli}(t)\left[\mathrm{cdf}_{pre}(t) - \mathrm{cdf}_{obs}(t)\right]^2\ dt \tag{2}$$

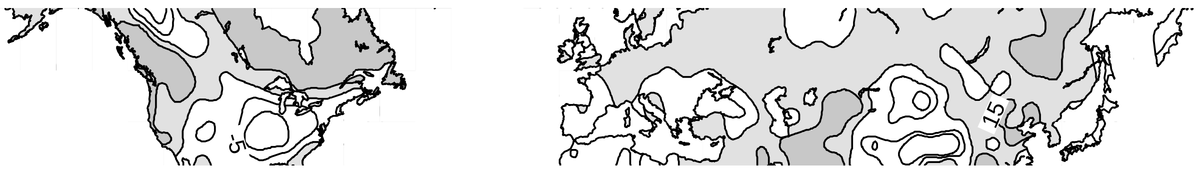


Figure 1: Ranked probability skill score for winter daily 2m temperature (land points only); contours ±5 and ±15%, shading above 5%.

In (2) the integral is calculated as a sum for bins of 0.5°C wide between -60°C and 30°C. The rps is a dimensionless quantity, with the size of a square difference between probabilities. To measure forecast skill, we need a reference. This reference is the minimum expected rps when prediction and observation are independent (i.e. no skill at all). It is obtained by replacing cdf_{pre} by cdf_{cli} in equation (2), which yields rps_{cli}. One can define a simple skill score as

$$\mathrm{ss} = (\mathrm{rps}_{cli} - \mathrm{rps})/\mathrm{rps}_{cli} \tag{3}$$

When this score is negative, the prediction is worse than the climatology prediction, which does not require costly numerical models. In this study we want to evaluate the skill for the mid-latitude (30°N-60°N) land points in EUROSIP. The first stage consists of calculating the climatological pdf on 21 winters (1987-2007) for ERA40 (extended by ERA interim beyond 2001), and the 3 models. Then a quantile-quantile correction (Déqué, 2007) is applied to each model, in order to correct the bias in pdf. For each temperature bin described above, the pdf and cdf are calculated by simple counting. Since the resulting pdfs are jaggy, a gaussian kernel smoothing is applied with 1 K standard deviation. Then the rps and ss are computed according to equations (2) and (3). The area mean rps is 0.0113 for climatology and 0.0106 for multi-model, leading to a skill score of 6.3%.

This method is not fully fair because the climatological pdf used in the correction introduces data to be predicted, that are not available in a real prediction. In fact, we do not correct the predictive behavior of each model, but only its climate properties, irrespective of the year-to-year chronology. In addition, this "cheat" favors also the climatological prediction which is the competitor. We have recalculated the scores, using the traditional "leave but one" method. The climatology rps becomes 0.0126 and the multi-model rps is 0.0118. The skill score remains unchanged. Thus the mid-latitude score is significantly positive and higher than traditional scores obtained with seasonal means, in particular over Europe. The geographical distribution (figure 1) indicates that the score is mainly positive (75% of the domain). This distribution depends not only on the models, but also on the period 1987-2007. Because of the reduced sample size, we did not attempt to improve the scores by *a posteriori* correction of the reliability of probability forecasts.

Conclusion

The winter daily temperatures provided by EUROSIP have a non-negligible skill versus climatology over most parts (except eastern US and western China) of the midlatitudes.

References

Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. *Global and Planetary Change*, **57**, 16–26.

Doblas Reyes, F.J., Déqué, M. and Piedelievre, J.P., 2000: Model and multimodel spread in the PROVOST seasonal forecasts : application to probabilistic forecasts. *Quart. J. Roy. Meteor. Soc.*, **126**, 2069–2088.

Ensemble prediction at Météo-France: a progress report

L. Descamps, C. Labadie, A. Joly, E. Bazile, J. Nicolau
Groupe d'étude de l'Atmosphère Météorologique (GAME), Météo-France et CNRS,
42 avenue G. Coriolis, F-31057 Toulouse cedex 1, France
Corresponding author: alain.joly@meteo.fr

1. Changes to the operational system

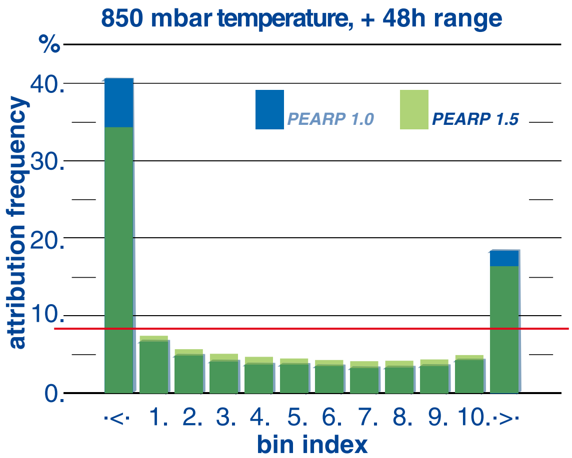


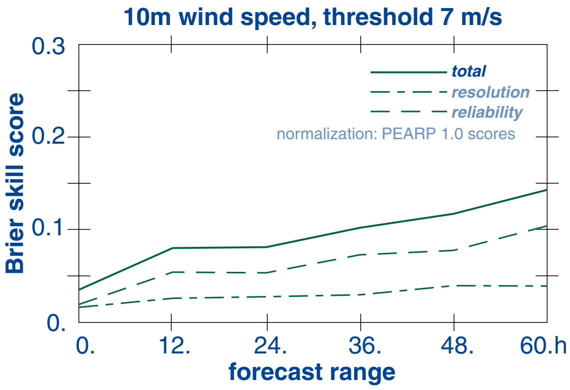
Figure 1: *Rank histogram contrasting the operational ensemble forecast of 2007 (labelled PEARP 1.0) with the version that went into operations at the end of january 2008 (labelled PEARP 1.5). The parameter is 850 mbar temperature at 48 h range. The histograms result from a one month validation period.*

Météo-France runs operationally since 2004 an ensemble forecast based on its global model Arpege, developed within the forecast department by J. Nicolau. It is called *PEARP*, french acronym for *Arpege ensemble forecast*. The ensemble has 10 members, it runs once per day following the 18UTC production assimilation. It is initialized with singular vector perturbations computed over the Northern Atlantic and Europe with an optimization time of 12 h. The model was unperturbed and exactly identical to its deterministic version. This ensemble was primarily meant to provide possible alternative scenarios in situations favourable to rapid storm development.

As the interest for producing actual probabilistic forecasts raises, it has been decided to hand the future evolution of the ensemble forecasting tool to the research department, while the forecast department would concentrate on developing methodologies and probabilistic products from both PEARP at short range and the ECMWF EPS at longer range. On 28 january 2008, a first step of evolution within the new organization has been completed. A new version, termed *PEARP 1.5* has been declared operational.

It includes the following changes. *(a):* although ensemble size is still 10, the horizontal resolution of the base PEARP model has been uncoupled from that of the deterministic version and fixed at T358 with the geometric factor C of the Schmidt-Courtier-Geleyn transform at $C = 2.4$. However, the vertical resolution continues to follow that of the other Arpege-Aladin models (changing from 41 to 60 levels), except that the mesosphere is removed. As a result, a control member has been added. It was planned to expand ensemble size later in 2008. Although indeed 10 further forecasts are performed in our experimental framework, the operations departments have other priorities and were no able to include these extra members in the operational suite. *(b):* further singular vectors are computed in all parts of the globe, although at low resolution (T44). The most significant change to the initialization, however, is the inclusion of some form of breeding. Indeed, initialization now includes the 24 h evolved perturbations from the previous run, combined to the singular vectors. *(c):* furthermore, the final anomalies added to the analysis are scaled to an amplitude sized using error variances derived from the 4D-Var assimilation cycle. *(d):* the complete

Figure 2: *Brier skill score of the event* 10 m wind speed faster than 7 ms^{-1} *as a function of the forecast range. The reference used for normalizing the score is the operational ensemble (PEARP 1.0). Also shown is the decomposition of the Brier score into its reliability and resolution components.*



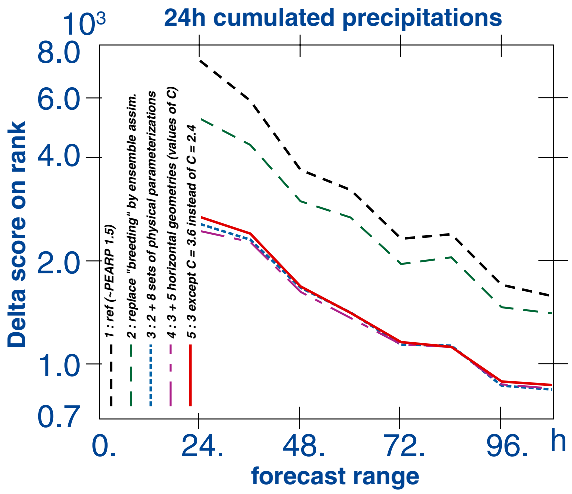


Figure 3: *A result from the PEARP 2009 preparatory experiment: the so-called "Δ" score of 24 h accumulated precipitations as a function of the forecast range (see Candille and Talagrand, 2005). This score measures the departure of the rank histogram from flatness. It has been derived from a series of experiment covering one month. Experiment definitions are recalled in the figure itself. Both the use of ensemble assimilation realizations and multiple parameterization sets bring key improvement to the probabilistic forecast of the parameter. All experiments are performed with 21 members.*

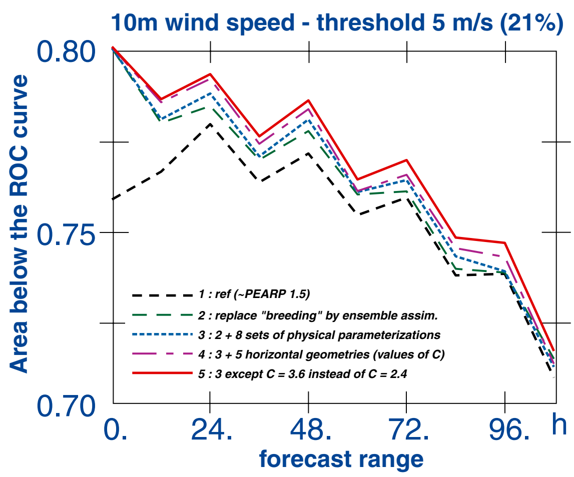
TIGGE requirements for output have been implemented, coupling files for the Hungarian Met Service limited area Aladin ensemble forecast are produced.

Figures 1 and 2 summarize the one month (05/01/2007 to 05/02/2007) evaluation period of this new version. The main goal of these changes was to turn *PEARP* into a true global ensemble forecast system as part of Météo-France contribution to the *TIGGE* project. However, the introduction of the semi-breeding especially, improved dispersion and resolution over the area of main interest.

2. Preparing the 2009 version

Preparations for turning *PEARP* into a more state-of-the-art ensemble prediction system that is both global and mesoscale over North-Atlantic and Europe are under way. One important innovation that has been introduced in our operational system in 2008 is a 6-members 3D-Var FGAT ensemble assimilation running parallel to the main 4D-Var Arpege assimilation cycle (Berre *et al.*, 2007). One goal of this assimilation ensemble with perturbed observations is to feed the 4D-Var with error statistics of the day. But another, equally important goal, is to provide better initial conditions for ensemble forecasting. Replacing the semi-breeding with perturbations from the assimilation ensemble is one primary feature of the 2009 changes. Another one is to include some kind of model error. One source will be to use several sets of physical parameterizations. According to another one month test period (march 2008, figures 3 and 4), these two aspects are able to bring very significant improvements, in particular with the probabilistic forecast of "actual weather" parameters (rainfall, near-surface state). Playing with the horizontal resolution (by changing the geometric factor C) is also considered: this can also improve the mesoscale aspects of the forecast. Finally, the ensemble size should be significantly enlarged up to between 30 and 40 members.

Figure 4: *Same as fig. 3 except that the score is now the area below the ROC curve, the parameter is the 10 m wind speed and the result is tied to the specific event that this speed is larger than 5 ms^{-1}. The frequency of that event is 0.21.*



Reference

Berre, L., O. Pannekoucke, G. Desroziers, S.E. Ştefănescu, B. Chapnik and L. Raynaud, 2007: A variational assimilation ensemble and the spatial filtering of its error covariances: increase of sample size by local spatial averaging. *Proceedings of the ECMWF workshop on flow-dependent aspects of data assimilation*, 11-13 June 2007, 151-168.

Candille, G. and O. Talagrand, 2005: Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. Roy. Meteor. Soc.*, **131**, 2131–2150.

Recent Improvements to the JMA Global NWP Model

Kengo Miyamoto[*†]
[*]Advanced Earth Science and Technology Organization
[†]Numerical Prediction Division, Forecast Department, Japan Meteorological Agency

1. Introduction

The JMA global NWP suite consists of a four-dimensional variational data assimilation system, a deterministic prediction system and two ensemble prediction systems. One of the two ensemble prediction systems is for one-week prediction, and the other is for predicting possible typhoon tracks over a short time span, where typhoons are defined as severe tropical cyclones in the western North Pacific. All these systems are based on a global spectral atmospheric model referred to as JMA-GSM, which has been in operation since 1 March, 1988. The T959 linear Gaussian grid model is currently used in the deterministic prediction system, and also serves as the outer model in the data assimilation system. As the inner model, the T159 quadratic Gaussian grid model is employed, while the ensemble prediction systems use the T319 linear Gaussian grid model. The number of vertical layers in all systems in the global NWP suite is 60. JMA-GSM is also used in the JMA seasonal prediction suite, and is further adopted for research on climate change.

The model has been improved many times over the last 20 years (e.g., Iwamura and Kitagawa, 2008). However, the basic structure of its source program remained old-fashioned in many ways. Additionally, repeated efforts to improve the model's accuracy and efficiency had left the source program disorganized. Accordingly, we decided to renovate the implementation of the model, thus making it faster and more accurate without changing its fundamental principles.

The renovated model is currently employed in the deterministic prediction system and the one-week ensemble prediction system. It is also used as the outer model in the data assimilation system. The inner model is scheduled for renovation in the near future (Kadowaki, 2009), and preparations are also under way to replace the typhoon ensemble prediction system.

2. Changes from the previous model

Major changes to JMA-GSM under the renovation are as follows: the reduced spectral transformation (Juang, 2004; Miyamoto, 2006) was introduced; a two-dimensional decomposition method was employed to decompose the calculation domain of the model for MPI parallelization; the procedures of inter-node communication were refined; and OpenMP directives were adopted for shared memory parallelization. Additionally, a number of deficiencies were corrected, and the source program was reorganized.

The introduction of the reduced spectral transformation lowered the model's number of grid points and the number of wave number components. Accordingly, it now has a kind of reduced Gaussian grid system. The number of grid points is 28.8% (22.9%) smaller than that of the conventional model in the case of the T959 (T319) linear Gaussian grid. As a result, the model's execution time has been shortened.

The employment of a two-dimensional decomposition method, the refinement of inter-node communications and the adoption of OpenMP directives improved computational efficiency, especially on scalar-type computer systems with a large number of computational nodes. The strategy for decomposing the calculation domain of the model and the procedures of inter-node communications were designed to minimize the occurrence of such communications. Load balancing among computational nodes and minimization of the amount of data communicated among such nodes were also considered. OpenMP directives allow us to parallelize outer loops in the manner expected, regardless of the kind of compiler used.

The renovation corrected the following deficiencies: the highest degree of integrands in spherical harmonic transformation exceeded the limit of Gauss-Legendre quadrature; the tables of Gaussian latitude, Gaussian weight and associated Legendre functions were evaluated in double precision arithmetic in the same way as the other variables; negative hydrometeor values were occasionally caused by unnatural local minima occurring in the semi-Lagrangian advection process; the coefficient

Corresponding author address: Dr. Kengo Miyamoto, Numerical Prediction Division, Forecast Department, Japan Meteorological Agency, 1-3-4 Otemachi, Chiyoda-ku, Tokyo 100-8122, Japan
E-mail: miyamoto@naps.kishou.go.jp

of horizontal diffusion for divergence was double that for vorticity as a relic of the implicit advection scheme previously employed. In the renovated model, the tables of Gaussian latitude, Gaussian weight and associated Legendre functions are evaluated in quadruple precision arithmetic. Additionally, a monotonic semi-Lagrangian advection scheme is employed.

The source program was also reorganized. A spectral model includes grid space and wave number space; the primary focus was changed to the grid space from the wave number space. This makes the source program clearer and allows numerical modelers to design physical parameterizations more flexibly. The reorganization reduced the minimum memory space needed to operate the model. The renovated version can be run on 12 nodes of HITACHI SR11000K1 at JMA, whereas the older model needed at least 30 nodes to implement.

3. Improved execution time

The renovation shortened the execution time for an 84-hour deterministic prediction by an average of 8 minutes and 19 seconds. The previous model took 38 minutes and 47 seconds to perform 84-hour deterministic prediction based on the average of the 36 days before the day of the replacement. The renovated model took 30 minutes and 28 seconds based on the average of the 36 days after it. At the time of the replacement, every deterministic prediction produced a special dataset for a scientific project. Accordingly, the extra time taken was included in the execution time. After the project, the renovated model took 28 minutes and 43 seconds based on the 36-day average.

4. Improved accuracy

The renovation improved RMSE in almost all the variables we investigated by an average of a few percentage points in the resolutions of both the T959 and T319 linear Gaussian grids. Tables 1 and 2 show the values of the improvement rate achieved by the renovation. The improvement rate shows the extent to which modification of the model diminishes the RMSE of each variable on average over a 216-hour prediction. The values were evaluated though two one-month assimilation / prediction experiments. One of the two experiments used the original model, and the other used the modified version. In the experiment, the four-dimensional variational data assimilation system was used with the model to produce assimilated fields at intervals of six hours (00, 06, 12, 18 UTC) over a one-month period. A 216-hour prediction was then started from each of the assimilated fields at 12 UTC with the same model.

Table 1: Improvement rate [%] resulting from the renovation on T959

	Psea	T850	Z500	Wspd850	Wspd250
Aug. 2006	1.49	1.46	1.57	1.06	0.95
Jan. 2007	1.95	1.78	2.1	1.39	0.92

Table 2: Improvement rate [%] resulting from the renovation on T319

	Psea	T850	Z500	Wspd850	Wspd250
Aug. 2004	1.26	1.27	1.39	0.78	0.69
Jan. 2006	0.62	0.63	0.52	0.52	0.39

References

Iwamura, K. and H. Kitagawa, 2008: An upgrade of the JMA Operational Global NWP Model. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, **38**, 06-03 – 06-04.

Juang, H.-M.H., 2004: A Reduced Spectral Transform for the NCEP Seasonal Forecast Global Spectral Atmospheric Model. *Mon. Wea. Rev.*, **132**, 1019-1035.

Kadowaki, T., 2009: Development of a Semi-Lagrangian Inner Model for Improving the Inner Resolution of the JMA Global Analysis System. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, **39**, in printing.

Miyamoto, K., 2006: Introduction of the Reduced Gaussian Grid into the Operational Global NWP Model at JMA. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, **36**, 06-09 - 06-10.

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*** $\mathbf{X}=(x_1,\dots,x_n)$, $n>1$, are the final forecasts under evaluation, and ***attributes*** $f_1,\dots,f_m$, $m>1$, comprise the quality criteria vector $\mathbf{F}=(f_1,\dots,f_m)$.

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 $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ 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 $\boldsymbol{r_0}$ (***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 **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 **F**=(ρ, 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.

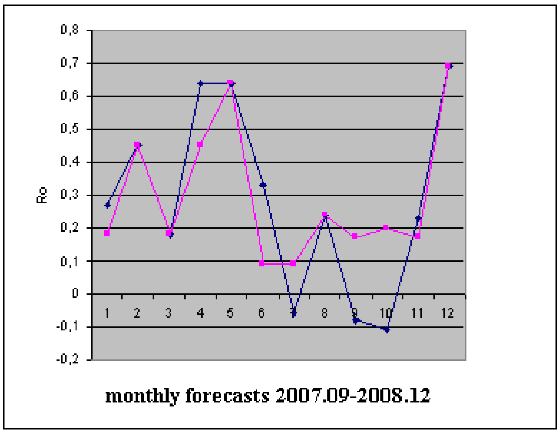


Figure. Anomaly sign correlation coefficient ρ for the SLAV predictors (blue) and the Kemeny-median persisted forecasts (pink).

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/.

Verification of Quantitative Precipitation Forecasts over Japan from Operational Numerical Weather Prediction Models

Takashi Nakamura
Numerical Prediction Division, Japan Meteorological Agency
Email: tks-nakamura@met.kishou.go.jp

1. Introduction

In 1995, the Working Group on Numerical Experimentation (WGNE) initiated a project for the verification and intercomparison of quantitative precipitation forecasts (QPFs) from operational NWP global models over different areas of the globe.

The Japan Meteorological Agency (JMA) has verified QPFs over Japan provided by operational NWP centers since 2002 (Hirai and Sakashita, 2004).

This paper briefly describes the verification results for 2007 and 2008.

2. QPF Data and Verification Method

(1) QPF Data

Table 1 shows the specifications of QPF data provided by operational NWP centers as of February 2009 and the methods of transforming QPF data.

The reference is taken from observational precipitation data derived from the surface rain gauge network over Japan that has been in operation since 1974. The density of the stations corresponds to a resolution of approximately 17×17 km.

(2) Verification Method

As the horizontal resolutions of QPF data differ among models, they must be transformed to a common verification grid system. In this activity, an 80-km mesh on a polar stereo projection is used as the verification grid system, and the two transformation methods outlined below are adopted.

1) Interpolation

This method is used to transform low-resolution QPF data from the original grid systems to the verification grid system. Each QPF value on a verification grid point is the interpolation of the raw QPF values on the four original grid points surrounding the verification grid point.

2) Averaging

This method is used to transform high-resolution QPF data from the original grid systems to the verification grid system. Each QPF value on a verification grid point is the average of the raw QPF values on the original grid points included in the verification grid point.

The methods used to transform QPF data are shown in Table 1. The observational data were transformed to the same verification grid system using the averaging method by regarding the stations as grid points.

Table 1 Specifications of QPF data provided by operational NWP centers as of February 2009, and the methods of transforming QPF data to the verification grid system.

NWP Center	Horizontal resolution (degrees)	Forecast time (hours)	Transformation method
BoM[1]	1.25×1.25	12, 24, 36, …, 120	Interpolation
DWD[2]	0.36×0.36	6, 12, 18, …, 174	Averaging
ECMWF[3]	0.50×0.50	6, 12, 18, …, 72	Averaging
NCEP[4]	1.00×1.00	6, 12, 18, …, 72	Interpolation
UKMO[5]	0.56×0.38	6, 12, 18, …, 96	Averaging
JMA[6]	0.25×0.25	6, 12, 18, …, 84	Averaging

3. Verification Results

(1) Time Series of Verification Results

Figure 1 shows a time series of the monthly bias score (BIAS) and equitable threat score (ETS) for precipitation exceeding 1 mm/24h for the Japan area (the forecast time is from 24 h to 48 h).

All models show seasonal variation in both BIAS and ETS. Concerning BIAS, there are peaks for summer in all models and for winter in some models. All models show poor ETS in summer.

(2) Six-hour Verification Results

Figures 2(1) and 2(2) show the frequency and ETS, respectively, of precipitation exceeding 1 mm/6h with respect to forecast time in summer 2007. Figure 3 shows the same data as Figure 2 for summer 2008.

From Figure 2(1), a clear tendency is seen for summer 2007: most models predicted precipitation the most frequently during the daytime and the least frequently at night. JMA predicted precipitation the most frequently in the evening though, and the frequency during the daytime is almost as high as that in the evening.

From Figure 3(1), a different tendency is found for summer 2008: JMA predicted precipitation the least frequently during the daytime, even though the other models predicted precipitation the most frequently during the daytime as with summer 2007.

[1] Australian Bureau of Meteorology
[2] Deutscher Wetterdienst (Germany)
[3] European Centre for Medium-Range Weather Forecasts
[4] National Centers for Environmental Prediction (United States)
[5] United Kingdom Meteorological Office
[6] Japan Meteorological Agency

All models overestimated the frequency of precipitation exceeding 1 mm/6h during the day for both the summers of 2007 and 2008.

From Figures 2(2) and 3(2), a dependency on local time can be found for the ETS. Some models predict precipitation better in the morning than with other local times. The dependency can be seen more clearly in summer 2008 than that of 2007.

4. Discussion

ETSs in summer are smaller than in other seasons for all models, and the precipitation they predicted have different dependencies on local time from observational precipitation. These facts indicate that it is still difficult for models to accurately estimate areas or frequencies of precipitation in summer.

A unique feature was found in JMA's model for 2008 whereby the frequency of precipitation showed minimum values during the daytime in summer. This feature was not found in the other operational models or for summer 2007. In November 2007 and January 2008, the cumulus parameterization scheme in the JMA model (GSM) was revised (Iwamura and Kitagawa, 2008; Nakagawa, 2008). Since no other changes related to precipitation processes were implemented in the JMA model, this is presumed to be part of the cause of this feature.

References

Hirai, M. and T. Sakashita, 2004: Verification of Quantitative Precipitation Forecast from Operational Numerical Weather Prediction Models over Japan (WGNE precipitation intercomparison project). *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, **34**, 6.01-6.02.

Iwamura, K., and H. Kitagawa, 2008: An upgrade of the JMA Operational Global NWP Model. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, **38**, 6.03-6.04.

Nakagawa, M., 2008: Improvement of the Cumulus Parameterization Scheme of the Operational Global NWP Model at JMA. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, **38**, 4.09-4.10.

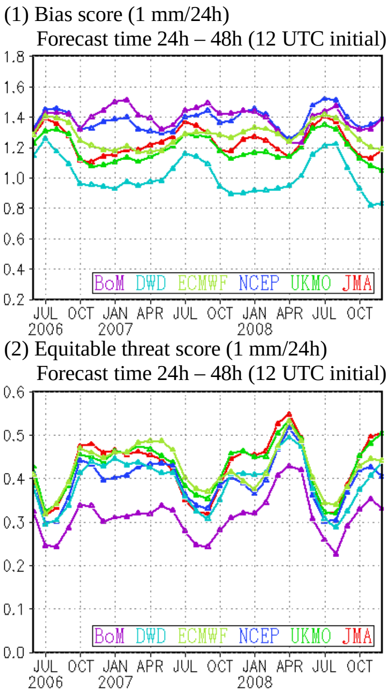


Figure 1 Time series of monthly bias scores (1) and equitable threat scores (2) for precipitation exceeding 1mm/24h for the Japan area from June 2006 to December 2008. The forecast time is from 24h to 48h (12 UTC initial). Scores are calculated for three consecutive months from the previous month to the next month.

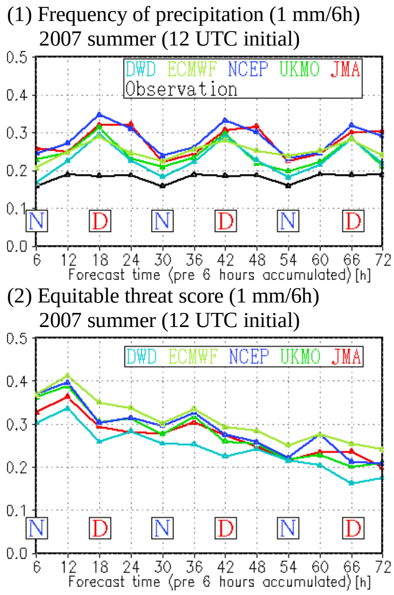


Figure 2 Frequency of predicted and observed precipitation (1) and equitable threat scores (2) for precipitation exceeding 1mm/6h (accumulated over the previous 6 hours) for the Japan area with respect to forecast time in summer 2007 (from June 2007 to August 2007). D and N correspond to local times of 09–15 (daytime) and 21–03 (night), respectively.

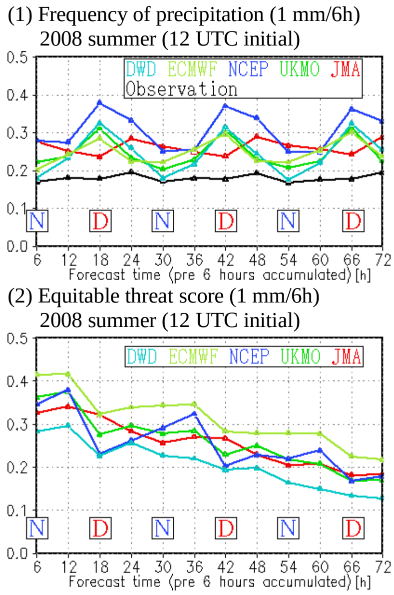


Figure 3 Frequency of predicted and observed precipitation (1) and equitable threat scores (2) for precipitation exceeding 1mm/6h (accumulated over the previous 6 hours) for the Japan area with respect to forecast time in summer 2008 (from June 2008 to August 2008). D and N are the same as in Figure 2.

A New Version of the Operational Global Spectral NWP Model of the Hydrometcenter of Russia

Rozinkina I., Astakhova E., Tsvetkov V., Ponomareva T., and Mikhailov A.
Hydrometeorological Research Center of the Russian Federation, Moscow, Russia

e-mail: inna@mecom.ru

1. Introduction

The spatial resolution of the global spectral model of the Hydrometcenter of Russia, used for operational medium-range forecasts, was improved from T85L31 to T169L31 in 2009. The step of the new Gaussian grid is about 0.7°. A more detailed shoreline description in the land-sea mask was introduced and the relief was adjusted to the new spectral resolution T169. The procedures of horizontal and vertical interpolation in preprocessing were modified. A new algorithm of defining initial meteorological values on lower model layers was developed.

The effects of these changes are analyzed in the present paper based on the results of numerical experiments with T85L31 and T169L31 models.

2. Improvement of horizontal resolution

The T169L31 model outperforms T85L31 in predicting dynamical fields in the free atmosphere. The advantage of T169L31 is seen in the behavior of most verification scores (see Fig. 1). The improvement is less pronounced for upper-tropospheric and stratospheric levels.

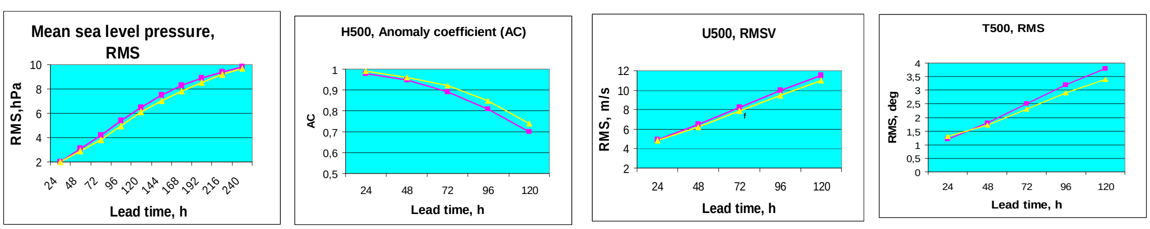


Figure 1: Verification scores for T169L31 (yellow) and T85L31 (magenta). Northern hemisphere. July-December 2008.

The increased horizontal resolution also resulted in better and more detailed prediction of near-surface weather elements. The precipitation fields now better correspond to the regions of atmospheric fronts, lines of instabilities, including those of orographic origin. Figures 2a and 2b demonstrate “the appearance” of really observed precipitation over the Caucasus in T169L31 forecasts.

(a) (b) (c) (d)

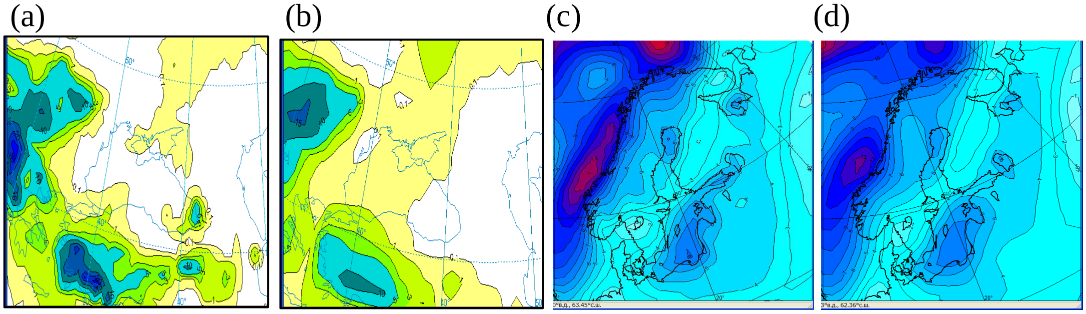


Figure 2: 6-h precipitation totals predicted by T169L31 (a) and T85L31(b). The 24-h forecast of the surface wind over sea by T169L31 (c) and T85L31 (d). The storm-wind regions are colored red.

The surface wind over sea is another resolution-sensitive element. The storm-wind regions are predicted much more realistically in the new model (Figs. 2c and 2d). Figure 3 demonstrates a successful T169L31 forecast of the storm wind zone with wind velocities of up to 40 m/s in the cyclone that approached Kamchatka in December 2008.

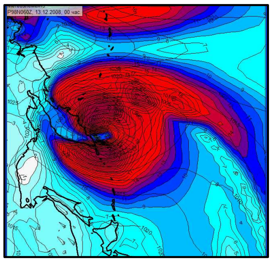


Figure 3. A T169L31 successful forecast of strong winds (red area) in the cyclone near Kamchatka in December 2008.

3. Modifications in model preprocessing

As the objective analysis data used for constructing initial datasets for the model is presented on a grid and levels that are different from those applied by the model, an accurate horizontal and vertical interpolation becomes of great importance. A new procedure has been developed for the spectral interpolation of two-dimensional fields on a sphere. Instead of traditionally applied associated Legendre polynomials, the Chebyshev polynomials of the first and second kinds are used in this procedure. The vertical interpolation scheme was also modified, and now vertical interpolation is performed using the Chebyshev-Laguerre polynomials combined with spline methods for the upper model levels. With this new scheme, the model spin-up was reduced. Higher-resolution model was more sensitive to the changes in the interpolation procedures.

4. Further research

Numerical experiments showed that the model is highly sensitive to initial fields of surface characteristics (sea surface temperature, albedo, roughness, soil moisture). Therefore, a new high-resolution archive of land-surface properties has been prepared using some results of the Institute of Geography, Russian Academy of Sciences. An example of information from this archive is given in Fig. 4. The new archive will help to prepare better initial surface data for the model.

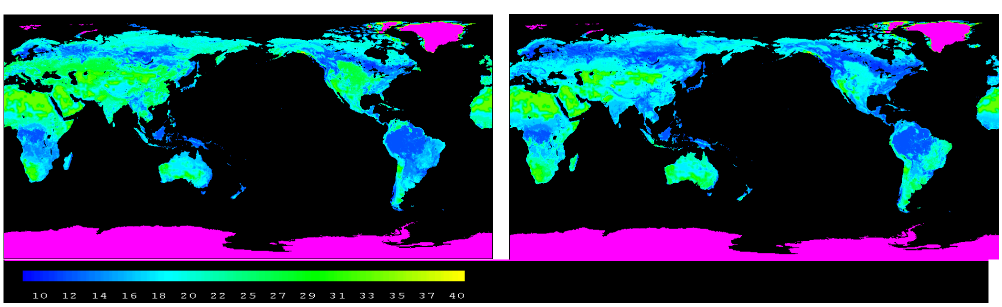


Figure 4: An example of information from the detailed archive of land-surface properties. Surface albedo for June (left) and December (right). Glaciers are shown in magenta.

Acknowledgements

The study was supported by the Russian Foundation for Basic Research (project 07-05-13607-ofi_ts).

New version of the seasonal forecast model at Hydrometcentre of Russia

M.Tolstykh, D.Kiktev, R.Zaripov, M. Zaichenko, V. Shashkin
Hydrometcentre of Russia 11-13 B. Predtecenskii per. 123242 Moscow Russia
Email: tolstykh@inm.ras.ru

The seasonal forecast model at Hydrometcentre of Russia is a version of the SL-AV model [1] with the resolution 1.125x1.40625 degrees lat-lon and 28 levels. It uses parameterizations of subgrid-scale processes developed in Météo-France and LACE consortium for ARPEGE/ALADIN NWP model. The seasonal forecast SL-AV model participated in SMIP2 project; also experiments were carried out using SMIP2/HFP protocol. The model produced reasonable fields in midlatitudes, however, the following drawbacks were noticed:

- Unrealistic high precipitation in tropics, wrong geographical distribution (lack of precipitation in continental tropics),
- T850 is too warm over Antarctida, too cold (by 2 degrees) over tropics,
- H500 is 30-40 m lower over tropics.

All this was attributed mostly to the absence of soil-vegetation-snow parameterization in the model version which participated in these experiments. In 2007, the ISBA parameterization [2] was implemented in the SL-AV model, including also soil freezing/melting according to [3] and the snow albedo parameterization [4]. First experiments have shown significant improvements for all the fields in the tropics, especially for precipitation. The seasonal prediction version of the SL-AV model was then further upgraded with the recent version of the shortwave and longwave radiation [5] developed by LACE consortium. This upgrade further improved tropical scores of H500 and T850.

The experiments according to SMIP2/HFP protocol were repeated with the new version of the model using 25 years of NCEP/NCAR reanalysis-2 data for all seasons. Each 4-month hindcast consists of 10-member ensemble. The model ensemble-mean data for the months 2 to 4 are averaged over 25 years and 4 seasons and compared with the corresponding data from reanalysis. The RMS error and bias for the old and new version of the model are given in Figs. 1 and 2 respectively. One can see that both error measures are improved, especially in tropics. Still there is a room for further improvement. Our work will include the increase of the vertical resolution and better account for ozone.

This work was supported by Russian RFBR grant 07-05-00893.

References

1. Tolstykh M.A. Semi-Lagrangian high-resolution atmospheric model for numerical weather prediction - Russian Meteorology and Hydrology. 2001, N 4. P. 1 - 9
2. Noilhan J. and Planton S. A simple parameterization of land surface processes for meteorological models - Mon. Wea. Rev. 1989, Vol 117, P. 536 – 549.
3. Boone A. , Masson V. , Meyers T., Noilhan J. The Influence of the Inclusion of Soil Freezing on Simulations by a Soil–Vegetation–Atmosphere Transfer Scheme - J. of Applied Meteorology. 2000. V.39. P. 1544–1569
4. Bazile E., Haiti M. El, Bogatchev A. and Spiridonov V. Improvement of the snow parameterization in ARPEGE/ALADIN. Proceedings of SRNWP / HIRLAM Workshop on Surface Processes, Turbulence and Mountain Effects; Madrid, 22-24 October 2001. January 2002.
5. Ritter B. and Geleyn J.-F. A comprehensive radiation scheme of numerical weather prediction with potential application to climate simulations - Mon. Weather Rev. 1992. V. 120. P. 303-325.

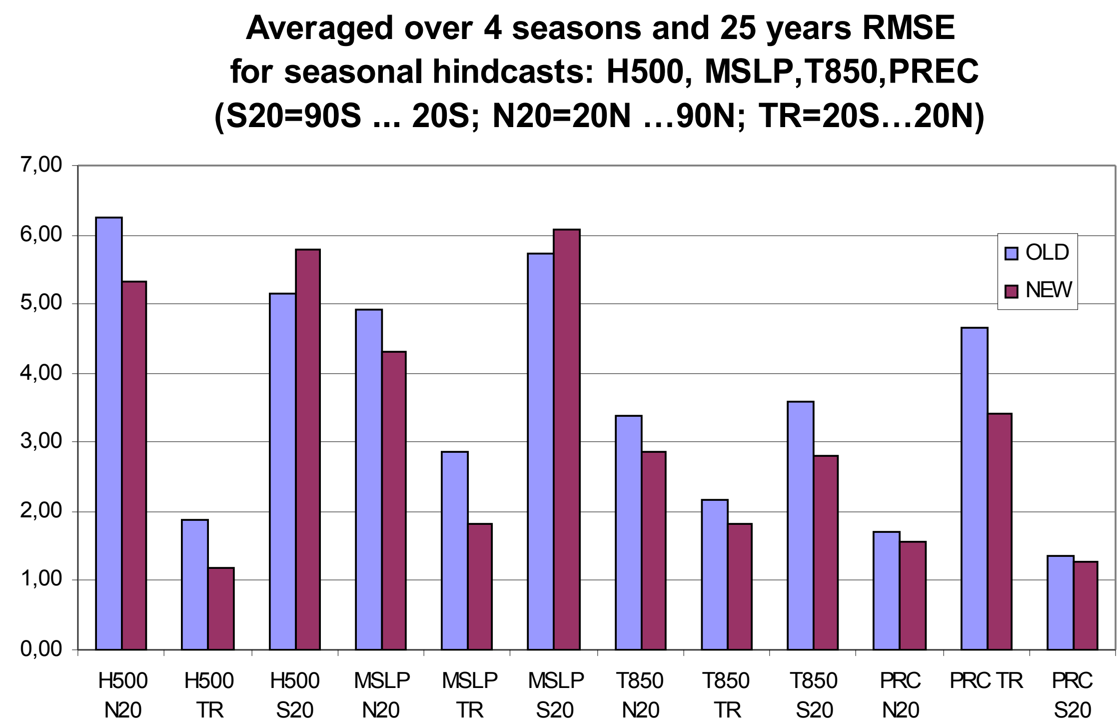


Fig. 1 RMS errors. Units: H500 – dam, MSLP – mb, T850 – K, PREC – mm/day.

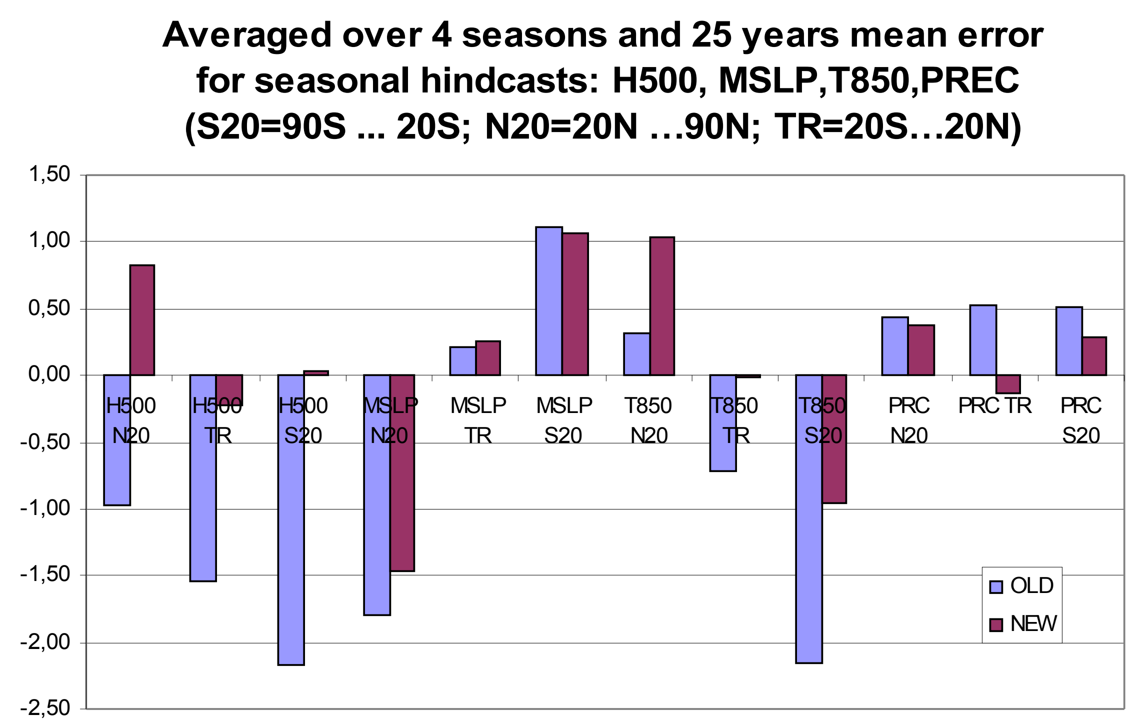


Fig. 2 Mean errors. Units: H500 – dam, MSLP – mb, T850 – K, PREC – mm/day.