Development of the new ensemble weather prediction system at the Hydrometcentre of Russia

Vasily Mizyak¹, Vladimir Rogutov², Ksenia Alipova³
Hydrometeorological Research Center of Russia, 11-13, Bolshoy Predtechensky per., Moscow, 123242, Russia
E-mail: ¹vmizyak@mecom.ru, ²rogutovv@mail.ru, ³ksusha_ast@mail.ru

Abstract. The new ensemble weather prediction system developed at Hydrometcentre of Russia is presented. It is based on the global atmospheric model with the resolution of 0.9×0.72 degrees in longitude and latitude, respectively, 96 vertical levels, and the Local Ensemble Transform Kalman Filter (LETKF) algorithm implemented by the authors. The key feature of this system is the centering of the initial data ensemble onto the operational analysis of the Hydrometcentre of Russia, which allows to take into account some satellite observations not assimilated in LETKF. Some improvements in assimilating radiosounding data has increased the accuracy of short range weather forecasts. The first results of ensemble medium-range forecasts are presented. Now this ensemble prediction system is ready for a quasi-operational use.

1. Introduction
Increasing the numerical weather prediction accuracy is a complex problem. Due to inaccurate representation of the atmosphere initial state and nonlinear dynamics even very small perturbations in the initial state lead to significant error growth after several days of forecast. The 5-10 days forecasts are usually probabilistic ones, implemented with the ensemble prediction system. The ensemble approach implies the calculation of a set (so-called ensemble) of slightly different predictions, which can be produced by different models, or using different parameters of the same model, or by integrating an atmosphere model that starts with a set of slightly different initial data. Stochastic perturbations of some parameters are introduced into the atmosphere model in order to take into account the uncertainties in the description of physical processes in this model. In this paper, we present a new ensemble prediction system of the Hydrometcentre of Russia.

Section 2 provides a general description of this system, its structure and features. The data assimilation system and aspects of some assimilated types of observations are described in Section 3. We present some results and conclusions in Section 4.

2. Ensemble prediction system
The ensemble prediction system described here is based on the global operational atmosphere model called SL-AV (semi-Lagrangian based on absolute vorticity equation) described further in Subsection 2.3 [1], which is used to obtain an ensemble of forecasts. The ensemble of initial data required to start the model is generated by the data assimilation system [2] which is our
implementation of the well-known Local Ensemble Transform Kalman Filter (LETKF) algorithm [3].

2.1. Structure

6-hour analysis-forecast cycle step consists of the following sub-steps.

- Analysis ensemble calculation (temperature $T$ and humidity $RH$ both in the atmosphere and at the surface, zonal wind $U$ and meridional wind $V$ in the atmosphere, surface pressure $Ps$) using LETKF data assimilation.
- Ensemble centering to the Hydrometcenter of Russia operational analysis.
- Soil variables analysis.
- SL-AV atmosphere model forecast.

Cycle run schematic is shown in fig. 1.

2.2. Features

2.2.1. Ensemble centering

The key feature of our system is a centering of initial data ensemble onto the operational analysis of the Hydrometcenter of Russia (HMC) [4]. To do this, the deviations of the operational HMC analysis fields from the mean LETKF ensemble analysis fields are added to respective ensemble members. This allows to correct the ensemble mean state which becomes the same as the operational analysis while the ensemble spread is taken from LETKF assimilation. This increases the forecast accuracy due to information from those satellite observations assimilated in the operational analysis that cannot yet be used in the LETKF assimilation system.

The mean analysis coincides with the operational one and the ensemble spread is generated in the ensemble assimilation system. This leads to the overestimated ensemble spreads in the regions where the number of assimilated observations in HMC operational data assimilation is considerably higher compared to LETKF data assimilation.

2.2.2. Soil variables initialization

The presented ensemble prediction system uses the soil assimilation module [5]. 2-meter temperature and humidity analysis increments at every grid point are used as input information to initialize the surface and deep soil temperature, surface and deep soil moisture content following [6].
2.3. The atmosphere model

Our system applies the Russian global atmosphere model SL-AV [1] developed at Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences and Hydrometcenter of Russia. This model is applied for operational medium-range and long-range weather forecasts at Hydrometcenter of Russia. It consists of the original dynamical core [7] and parameterizations of subgrid-scale processes mostly developed by ALADIN/LACE consortium [8]. The version used in this study has the horizontal resolution of 0.9×0.72 degrees in longitude and in latitude respectively, and 96 vertical levels.

3. Ensemble data assimilation system

3.1. Brief description

Our implementation [2] of the local ensemble transform Kalman filter [3] is a core of ensemble data assimilation system. This system, in turn, is used to generate an initial data ensemble. The propagation step of this system uses SL-AV model to produce the ensemble of backgrounds, which are 6-hour model forecasts. The ensemble size is 60 members.

Multiplicative [9] and additive [10] background covariance inflations are used to treat the ensemble spread underestimation and prevent from filter algorithm divergence [2]. The multiplicative inflation helps to account for possible undersampling, while the additive inflation accounts for the model error.

The observation localization is applied in our data assimilation system to reduce the spurious correlations problem and also to increase the ‘effective size’ of the ensemble. Namely, the analysis at every grid point is produced independently from other grid points, using observations from a local region only with localization radius depending from the type and height of observation. This guarantees the high level of parallelism but limits the system capability: non-local observations (like satellite radiances) can not be assimilated easily.

Thus the observation types used in this assimilation scheme are those that can be localized in the physical (and hence in the model) space:

- radiosoundings (about 2000 observations);
- land stations and ship stations (about 10000 observations);
- aircraft observations (about 10000 observations);
- satellite derived atmospheric motion vector observations (800000 observations);
- satellite derived advanced scaterrometer sea surface winds speed observations (700000 observations).

We further describe the particular features of our data assimilation system.

3.2. Radiosoundings data assimilation

The values of many variables, such as wind speed and direction, may differ significantly above and below the tropopause, and assimilation of such data may increase the analysis error. To reduce the influence of those observations that are on the other side of the tropopause from the grid point where the data assimilation is performed, our implementation of LETKF artificially increases the error covariances of these observations.

In addition, radiosounding data of geopotential, temperature and wind as compared with our previous implementation [2] is incorporated. Such observations using special representation are now available at 299 vertical levels, while conventional radiosounding observations are available at only 16 standard levels. As the observations with frequent vertical disposition are correlated, all the available observations are thinned by rejecting most of levels and then the rest of observations are assimilated at 32 pressure levels.
A series of numerical experiments are carried out with the ensemble weather forecast system described above for two periods: from February 1 to March 1, 2020 and from April 11 to May 11, 2020 applying 6-hour analysis-forecast cycle. The root mean squared errors for 48-hour forecasts with respect to operational analyses are calculated for Northern hemisphere extratropics (20-90N) for air temperature and wind speed. The results are presented at fig. 2. One can see that the new way to use radiosoundings observations somewhat increases the accuracy of the short-range forecasts. The similar conclusions can be drawn for Tropics and Southern extratropics.

3.3. Assimilation of the Atmospheric Motion Vector observations

Atmospheric Motion Vector observations (AMV) are obtained by tracking the motion of identical clouds or water vapor structures on two or more consecutive images derived from satellite. They measure the speed and wind direction in the upper-air. The big advantage of this observation type is the global Earth coverage. Often, AMVs are the only source of data on troposphere winds over some parts of the Ocean and in the upper latitudes including Arctic. In addition, this type of observation has high spatial and temporal resolution, and the measured characteristics can be easily transformed into model variables.

One of the main challenges while assimilating AMVs is that it is difficult to obtain proper height of observation from 2D image and therefore the errors of height assignment arise. We use the AMV height reassignment scheme from [11]. This allows to significantly improve the accuracy of produced initial data [12].

Another difficulty is the observation errors which are correlated in time and space (the correlation radius is about 700 km). In our data assimilation system, we simulate the observation error correlations using Second Order Autoregressive function (SOAR) [13], which was previously used for other types of satellite observations. This approach helps to slightly increase the accuracy of produced initial data and so the model forecast starting from that data [14].

3.4. Assimilation of the Advanced Scatterometer observations

Advanced Scatterometer (ASCAT) observations are the other satellite type of observation, which we have implemented in our data assimilation system. The spectrometer emits the electromagnetic wave and measures the signal scattered in the backward direction from the water surface. This signal magnitude depends on the surface shape which depends on the surface wind speed. This magnitude allows to find the wind speed and direction over the sea.
ASCAT data are grouped into superobservations (each superobservation is the mean ASCAT observation in the cell with the cell size comparable to the model grid cell). Observations from inner lakes, with low surface pressure and with big difference between observation height and model orography are not used in assimilation.

4. Numerical experiments and results
The code of all ensemble prediction system components is written using hybrid MPI and OpenMP parallel programming technologies. 10-days forecasts are computed once per day, at 12 hours UTC (universal coordinated time, coincides with Greenwich mean time). 6-hour forecasts are computed for all other standard synoptic times (00, 06, and 18 hours UTC) to obtain background fields and to advance data assimilation system. Runtime of whole program complex is about 1.5 hour for 12 UTC run including 10-days forecasts at 992 processor cores of Cray XC40 supercomputer.

Forecasts of August 2020 have been computed. The plots of the root-mean-squared error of the mean forecast for geopotential at 500 hPa, temperature and meridional wind at 850 hPa, mean sea level pressure averaged over the Northern extratropics and forecast ensemble spread averaged over region as functions of forecast lead time are shown in fig. 3. The plots for other regions and surfaces demonstrate a similar behavior. Similar plots are published at the WMO Lead Center on Verification of Ensemble Prediction System site [15]. It can be noted that our results are still worse in comparison with the results of most other forecast centers, but we are working to decrease this gap.

Since the ensemble forecasts are probabilistic, probabilistic scores of forecasted event probability should be used for its verification. Such scores presented here are Reliability diagrams and Relative operating characteristic (ROC) for the events when the forecasted value is less or greater than some target value usually measured in terms of standard deviation units (see [15] for their definition and details). This is especially useful when predicting some weather anomaly event. Binary classification for score calculation is performed using bins obtained by dividing the probabilistic interval from 0 to 1 by the ensemble size. Plots of these characteristics for temperature anomaly at 850 hPa exceeding one standard deviation predicted by our system are presented at fig. 4. Data for plot are aggregated over the period of August 2020.
Figure 4. Reliability diagrams and Relative operating characteristic (ROC) for temperature at 850 hPa surface. The lines at the left plot should be as close as possible to diagonal. The lines at the right plot should be as close as possible to the upper left corner.

5. Conclusions
The current version of the presented ensemble prediction system is already capable to provide a satisfactory ensemble forecast, but further improvements are required. We plan to improve accounting for model uncertainty to reduce the difference between the curves at plots in fig. 3. Currently, the ensemble prediction system is deployed in a quasioperational mode and is ready for trials at the Hydrometcenter of Russia.

This study is carried out with the support of Russian Fund for Basic Researches grant 17-05-01227.

References
[1] Tolstykh M A, Fadeev R Y, Shashkin V V, Goyman G S, Zaripov R B, Kiktev D B, Makhnorylova S V, Mizyak V G and Rogutov V S 2018 Russ. Meteor. Hydrol. 43 773–79
[2] Shlyaeva A, Tolstykh M, Mizyak V and Rogutov V 2013 Russ. J. Num. Analysis & Math. Mod. 28 419–42
[3] Hunt B R, Kostelich E J and Szunyogh I 2007 Physica D: Nonlinear Phenomena 230 112–26 ISSN 01672789
[4] Tsyrul’nikov M D, Svirenko P I, Gorin V E, Gorbunov M E, Ordin A L and Bagrov A N 2016 Results of Testing New and Improved Technologies, Methods for Hydrometeorological Forecasts [in Russian] 3–12
[5] Bogoslovskii N N, Shlyaeva A V and Tolstykh M A 2008 Computational Technologies 13 111–16
[6] Giard D and Bazile E 2000 Mon. Wea. Rev. 128 997–1015
[7] Tolstykh M, Shashkin V, Fadeev R and Goyman G 2017 Geosci. Mod. Dev. 10 1961–83
[8] Termonia P, Fischer C, Bazile E, Bouyssel F, Brožková R, Bénard P, Bochenek B, Degrauwe D, Derková M, El Khatib R et al. 2018 Geosci. Mod. Dev. 11 257–81
[9] Anderson J L and Anderson S L 1999 Mon. Wea. Rev. 127 2741–58
[10] Houtekamer P, Mitchell H L and Deng X 2009 Mon. Wea. Rev. 137 2126–43
[11] Yang C, Lu Q and Zhang P 2012 Acta Meteorologica Sinica 26 614–28
[12] Mizyak V G, Shlyaeva A V and Tolstykh M A 2016 Russ. Meteor. Hydrol. 41 439–46
[13] Stewart L M, Dance S and Nichols N 2008 Int J Numer Methods Fluids 56 1521–27
[14] Mizyak V G, Shlyaeva A V and Tolstykh M A 2017 The use of correlated observation errors in satellite AMV data for the ensemble data assimilation system based on LETKF algorithm CITES 2017 pp 156–60
[15] WMO lead center on verification of ensemble prediction system URL http://epsv.kishou.go.jp/EPSv/
[16] Palmer T, Buizza R, Hagedorn R, Lawrence A, Leutbecher M and Smith L 2006 ECMWF newsletter 106 10–17