Results of statistical correction of air temperature forecasts for Siberian based on the COSMO-Ru_Sib model

Zdereva M., Tokarev V., Khluchina N., Voronina L.

Siberian Research Hydrometeorological Institute. Russia, Novosibirsk.

Technological complexes COSMO are installed in the SGI Altix 4700 Computer of the West Siberian Computing Center (Novosibirsk) with well-run initial and boundary data acquisition system, launch of calculation units on models, and resulting output in map, text, and GRIB formats. The spatial step in the modification of COSMO being run in the West Siberian Computing Center is 13.2 km.

There only deterministic forecasts are made in Novosibirsk. The temperature forecasts contain sometimes essential errors. The quality of model forecasts varies depending on specific terrain, season, and a lead time of forecast. The purpose of that presentation is to assess a possibility of improving model temperature forecasts using statistical technique (MOS) for Siberia region.

Data for 205 synoptic stations were involved in testing. For every station, the modeled temperature was presented by mean values from the nearest grid-box.

Figure 1 shows more or less robust estimates are observed during the summer months, though during other months the quality of forecasts decreases with lead time, especially for night hours (24h-48h-72h).

The result is improved after the application of the correction factors of the linear regression between the forecast and the actual temperature: Tfrcst=a*Tactl+b.

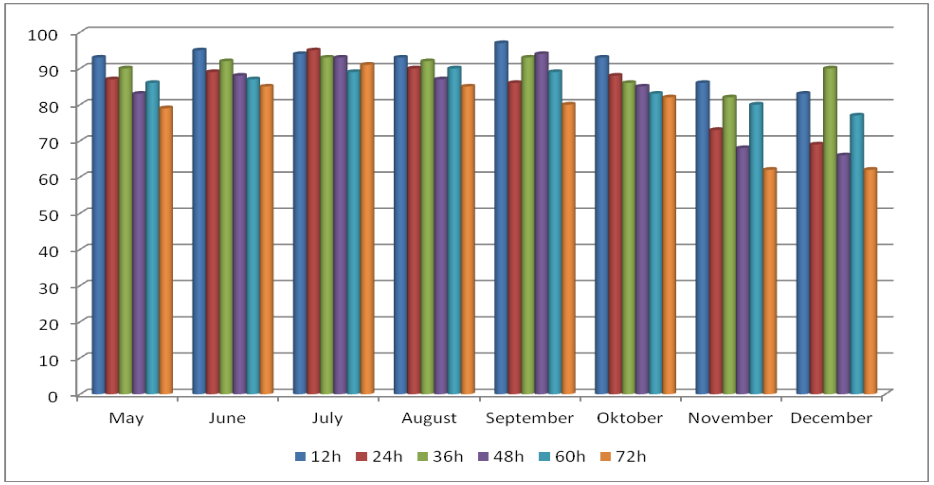


Fig.1 The percentage mean absolute error ≲3.5C for south-east of West Siberia (137 synoptic stations)

The example of summer forecasts for different regions before and after correction is presented in Figure 2. Forecasting is complicated for mountain stations. It is evident that the model overestimates the temperature during nighttime and underestimates during daytime. This shift is typical in the summer.

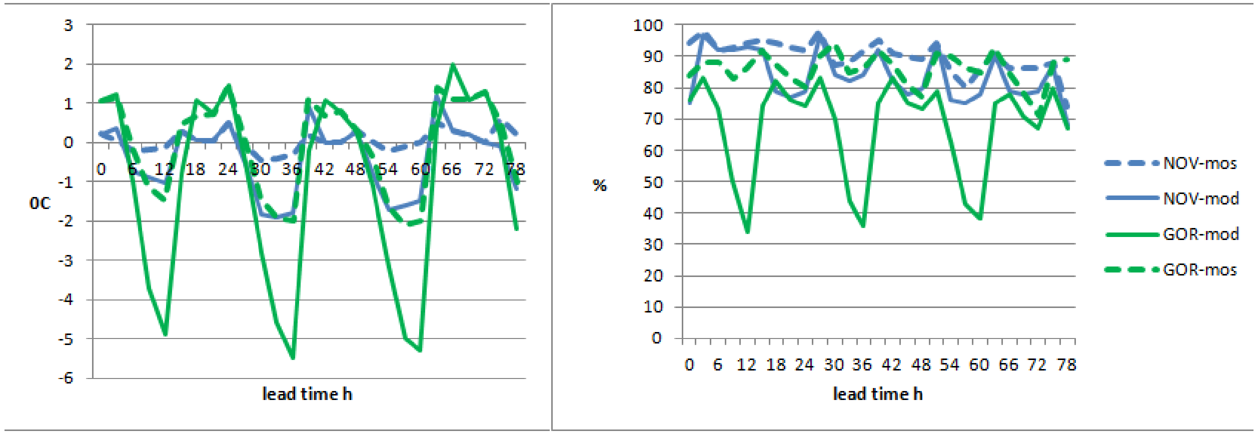


Fig.2 There are difference mean arithmetic errors (left) and percentage of mean absolute error ≲3.5C (right) for Novosibirsk region (NOV) and Republic Altai (GOR). The solid line is for COSMO model forecast, the dashed line is for corrective forecast.

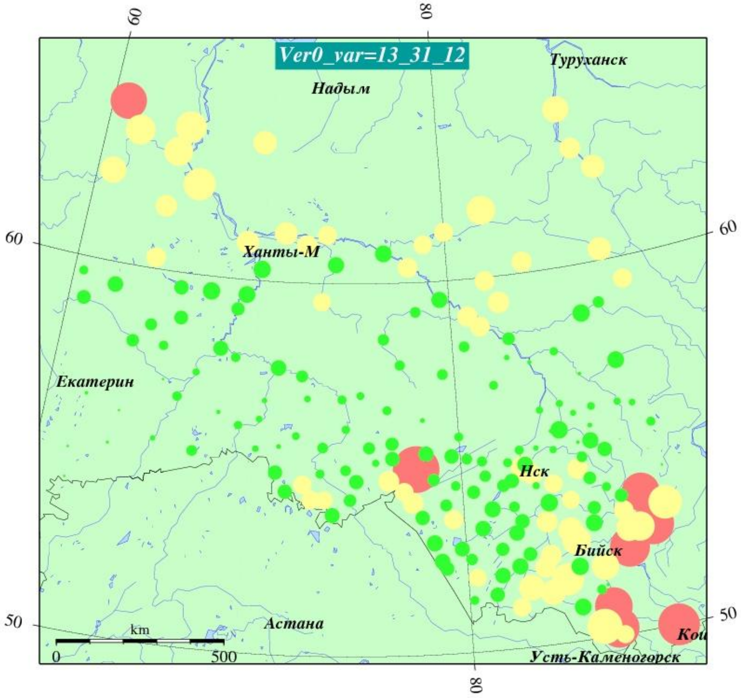
In Novosibirsk, the operational technology of daily calculations of factors is set up based on a relatively small moving learning sample. It contains 40 pairs of actual and forecasted values preceding the calculation. The factors are calculated for each station for each lead time. Three-year testing of this approach confirms the summertime forecasts improving.

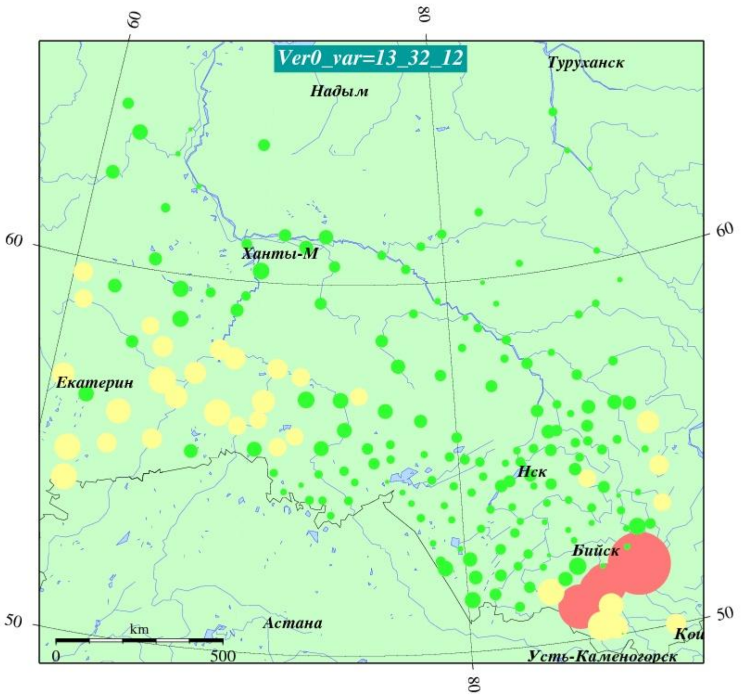
The task is complicated with transition to wintertime forecasting. The amplitude of temperature changes greatly increases over the region territory, both in space and in time. Stable anticyclone weather may cause severe frosts with temperatures below -40 degrees Celsius. Model calculations in these cases significantly overestimate the temperature. And conversely, upon the approaching of warm front, the model captures the warming with great accuracy. The shift toward underestimation is more frequent in forecasts with statistical correction, but it is associated with a standard approach to rare but large negative errors in the learning sample and only confirms the strong statistical heterogeneity of the error array.

To improve the homogeneity of the learning sample, we solved the problem of classification based on unknown beforehand values of the model forecast error. Apriori hypothesis assumes that large model errors are somehow connected with the peculiarities of surface and upper-air thermodynamic fields and sub-grid scale topography. To find a connection with meteorological fields, we screened a large complex of model output parameters and their complexes at grid points nearest to each station, and subgrid topography was taken into account indirectly by binding errors to a particular station. Initial parameters (from model COSMO): pressure on the sea level and its derivatives – Laplacians, gradients, changing for 12 hours; temperature at 850hPa, and vertical gradient; wind speed modulus at 500hPa; gradients and laplacians of geopotential at 500 hPa. In all there are 13 parameters.

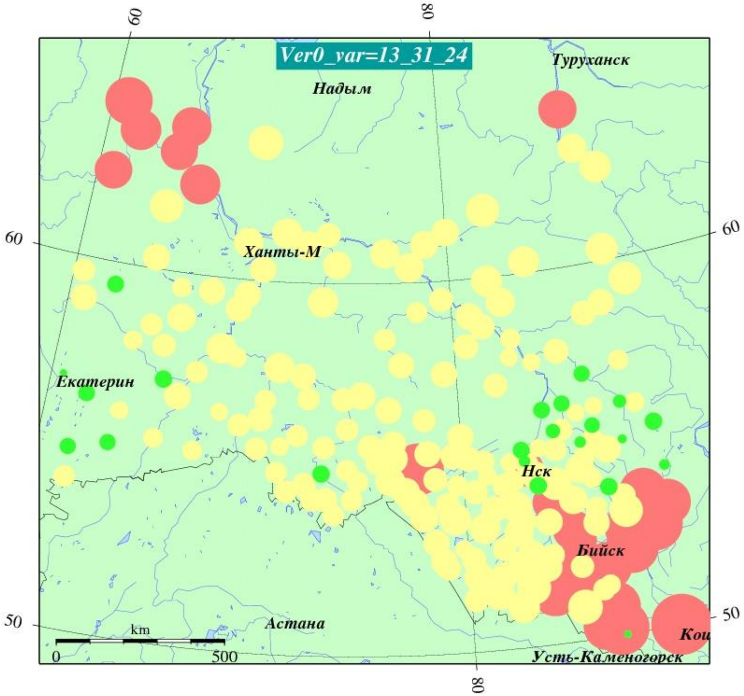
Further classification was performed by construction of a binary decision tree with probabilities of each of the two classes on terminal branches. Thus, after numerous experiments on fitting the lists of appropriate attributes, we obtained logical decision trees which allow, based on prognostic output data at grid points, to attribute a given station to one of potential classes specified for the value of expected model errors: "< -5", "< -3", "< |3|", "> +3", "> +5".

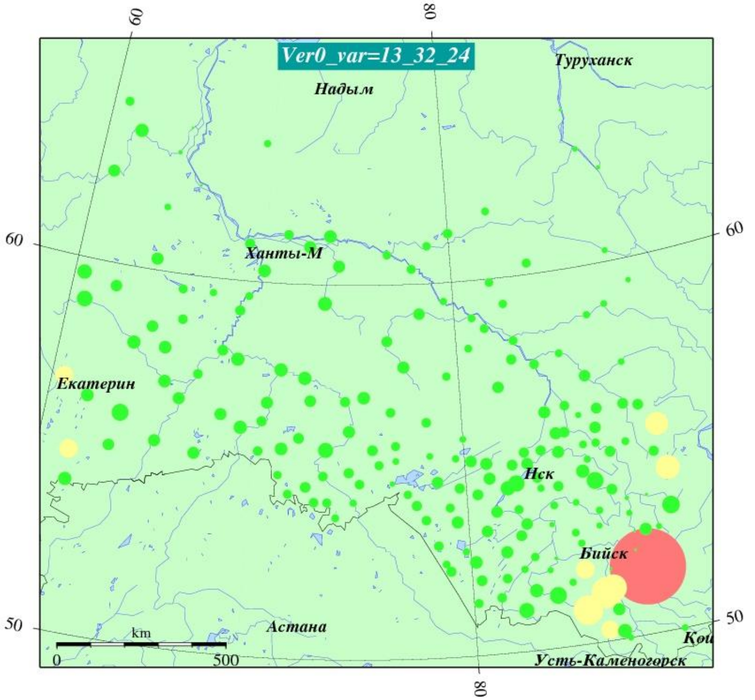
Initial probabilities of each class relative to the whole sample which is about 300 cases for the period 2012 -2014, or for 10 winter months, were obtained for stations in the region. The Figure 3 shows distribution of these probabilities over the territory. Further, prognostic linear regressions were built for each class (lead time, station) for interpretation and correction of the COSMO model forecasts at 220 stations in Western Siberia (Fig.4).


Ver0_var=13_31_12
Надым
Туруханск
Ханты-М
Екатерин
Нск
Бийск
Астана
Усть-Каменогорск
km
0
500
60
80
50

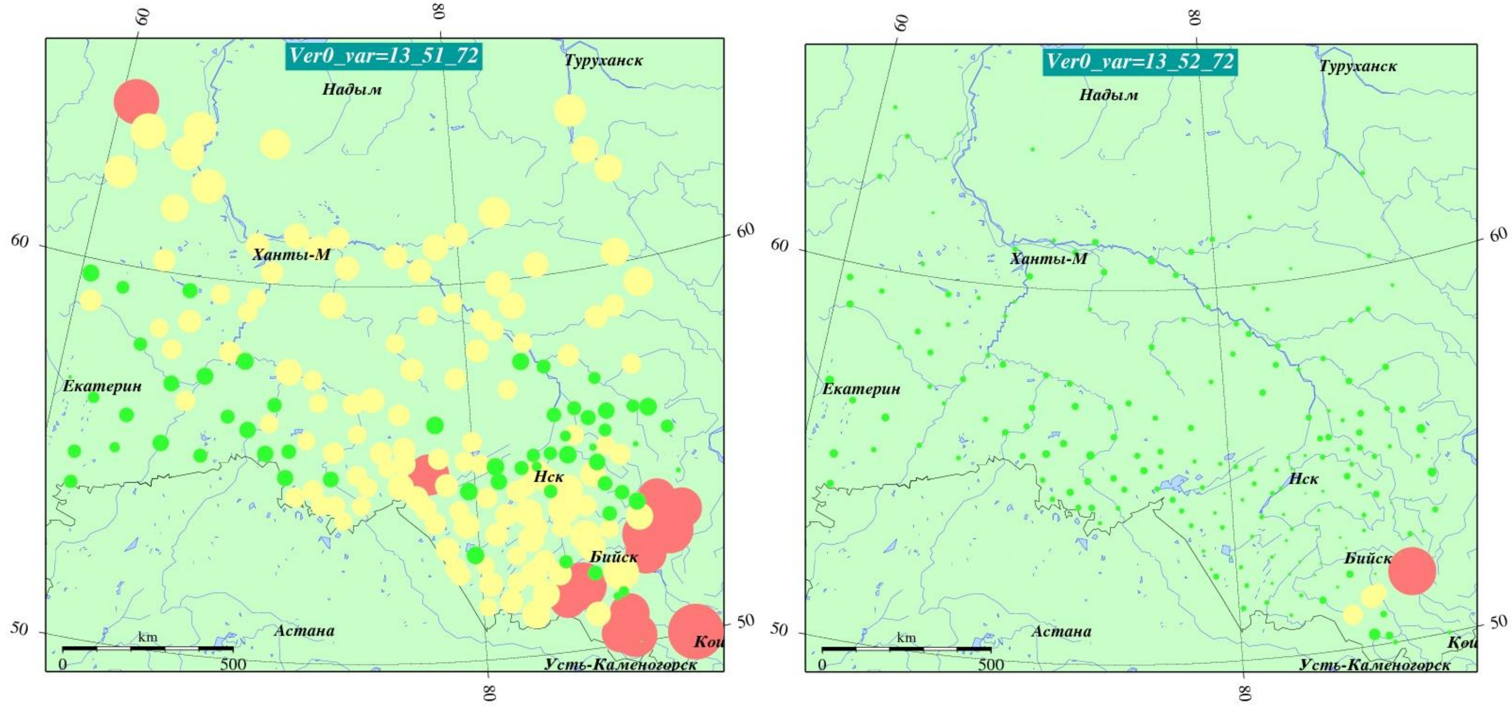

Ver0_var=13_32_12
Надым
Туруханск
Ханты-М
Екатерин
Нск
Бийск
Астана
Усть-Каменогорск
km
0
500
60
80
50

00+12h


Ver0_var=13_31_24
Надым
Туруханск
Ханты-М
Екатерин
Нск
Бийск
Астана
Усть-Каменогорск
km
0
500
60
80
50


Ver0_var=13_32_24
Надым
Туруханск
Ханты-М
Екатерин
Нск
Бийск
Астана
Усть-Каменогорск
km
0
500
60
80
50

00+24h



00+72h

Fig.3- Median error (%) for model temperature forecasts >+3 (left), <-3 (right) and >+5,<-5 (at the bottom). Data for 10 winter months 2012-2015. Lead time is in title: the last number after stressing.

- **reliability <20%**

- **20%≤ reliability >40%**

reliability ≥40% **Diameter is proportion by value.**

The COSMO-RU-Sib model in the first-day evening forecast (12h) underestimates the temperature at two-thirds of Western Siberia's stations, though by the third day the shift is almost leveled. As for next morning forecasts (24h), the COSMO-RU-Sib model overestimates the wintertime temperature almost everywhere in Siberia, at 80 % of stations on the first day and at 90 % - on the third day

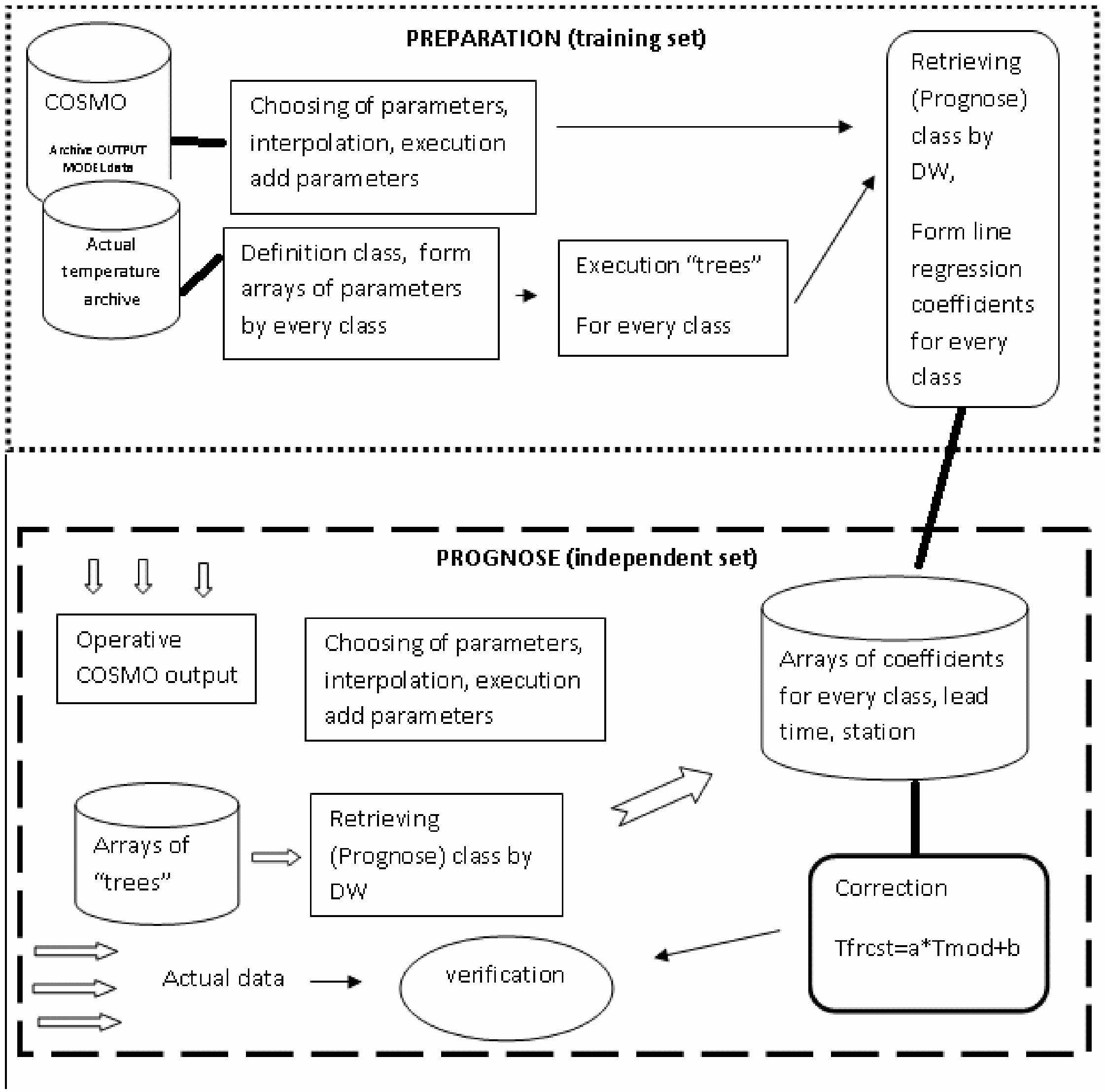
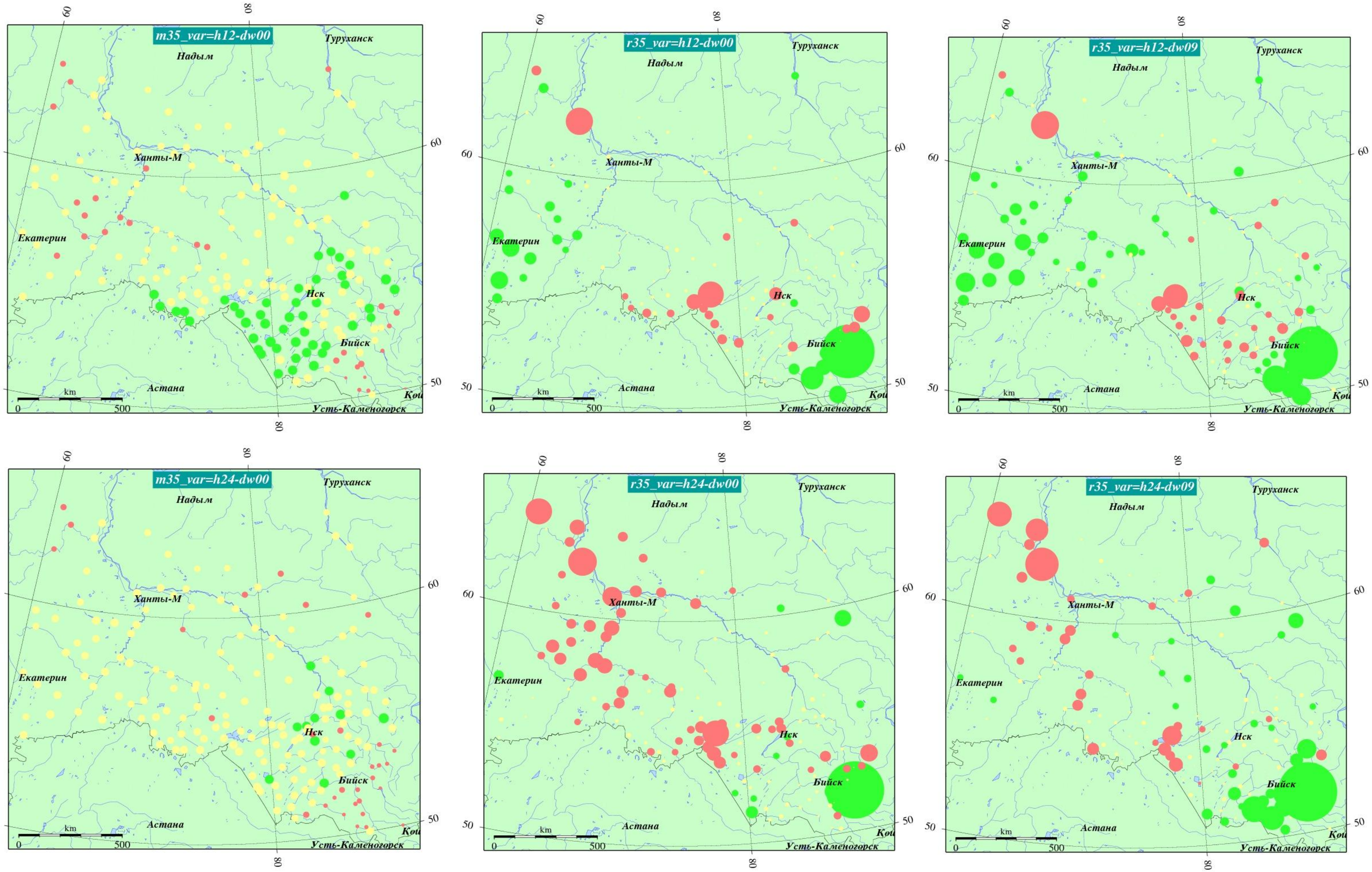


Fig. 4 Technological scheme of executing statistical correction in SibNIGMI

The examination of developing algorithm was conducted on winter data 2015-2016 (150 situations). The results of testing are presented below (Fig.5).



00+12h

00+24h

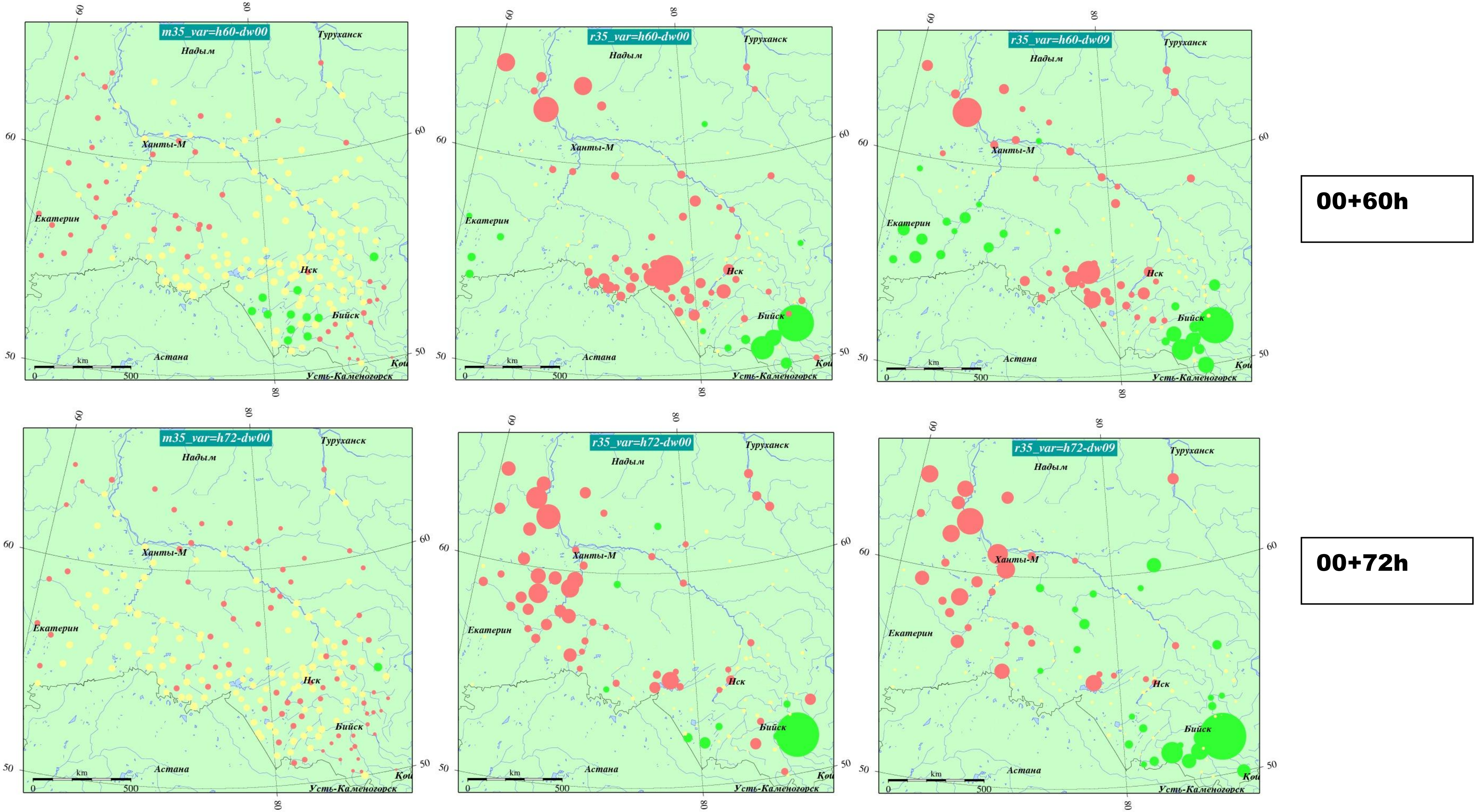


Fig.5 **Left column: forecast success rate (%error <|3.5|) of the COSMO model** –>80%, -60%...80%, -<60%

The deviation success rate of statistical correction by linear regression without clustering (middle column) and after clustering at first (right column) from COSMO model forecasts : >5%, -5%...+5%, -<-5%

So, the green circles show results, when correction is better, the red – opposite results.

Conclusions:

1. Mapping of prognostic model errors allows to analyze possible impact of topography and features of the location of weather stations on the quality of forecasts. Preliminary results suggest that hydrodynamic model (HDM) is very sensitive to the mountain topography, channels of great rivers, and lake shores of Siberia.
2. Statistical correction using a simple linear regression on a relatively small number of stations can improve the HDM results due to substantial heterogeneity of the learning sample on the array of stations.
3. Preliminary classification of sampling within situations with potentially large/small sign errors can significantly improve the uniformity and quality of correction using simple linear regression. The DW-algorithm is very suitable for such classification.
4. Detailed physical and statistical (geographical) analysis of HDM errors before and after correction is capable of providing useful information to HDM developers to improve model results in difficult physico-geographical conditions of Siberia.