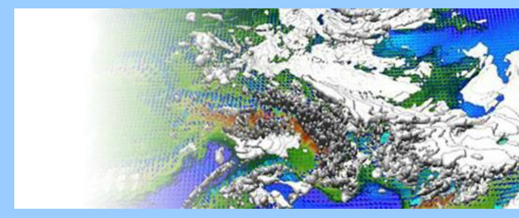
# The choice of the variographical model for Kriging interpolation of data of pressure, temperature and precipitation of model COSMO



COSMO / CLM / ART - User Seminar 2016


**Katsalova Lesia and Vitalii Shpyg**

*Ukrainian Hydrometeorological Institute, Kyiv*
*e-mail: Vilal@rambler.ru, vitold82@i.ua*


## Introduction

**Using mathematical models for receiving the weather forecast entered to the world practice long time ago. However there are not so many models in the world, that describe processes, occurring in the atmosphere and, properly speaking, determine the weather. Development, implementation and support of these models require considerable intellectual and material resources. Such models being developed in the USA, Japan, UK, Germany. Most world countries use such models' data for specification and improvement its forecasts. In particular, American model WRF works in UHMI, COSMO model adaptation is performed.**

**The practice has shown, that the work of these models does not include fully landscape of Ukraine, features of the underlying surface, that, in its turn leads for inaccuracies of the weather forecast. To correct these inaccuracies should be done post-processing of received data and its adaptation to the peculiarities of the territory of Ukraine. Also, meteorological models give values of meteorological parameters at some grid, that does not match with a grid of the weather stations of Ukraine. This complicates the work of the model verification for Ukraine and requires the use of interpolation methods.**

**Interpolation is a method of obtaining intermediate values of the parameters on a discrete set of known values. There are a lot of methods of interpolation, but its effectiveness largely depends on tasks and input data. One of the methods of interpolation is Kriging. This method optimizes interpolation procedure based on the statistical nature of the data (in this case meteorological parameter). Kriging similar to interpolation of weighted distances, it determines weight of surrounding measured nodes for determining the desired value in the unmeasured point. But definition of weight for surrounding units in the Kriging method is complicated compared to the method of weighted distances. In Kriging all depends on model of variogram and spatial distribution of measured units around the estimated point.**

## Value of the variographic model for Kriging-interpolation

**The components that determine the magnitude of the error of Kriging is described.**

- **Measurement error (input data error) can not affect the outcome of interpolation. This error is unrecoverable and it is clear that the error of results not will be less, than input data error.**

**For the case of data interpolation COSMO model: if input data have error concerning real data, then result of these data interpolation by Kriging also will have not less error.**

- **The error of machine's calculation is a rounding error. It is necessary to rotate the matrix during the implementation of Kriging, which requires the implementation of a significant number of transactions. Since the data are presented in the program are not accurate, but with some rounded, then the results of operations of the data will be obtained from some error. The larger the matrix, the more transactions to make, the greater the rounding error will be. Therefore it is advisable to use a block Kriging, which reduces this error.**
- **The error of the method. The method of Kriging based on assumptions of unbiasedness evaluation and minimization its dispersion that gives an opportunity to get optimally accurate interpolation under exact variogram (when the theoretical variogram describes accurately distribution of experimental data). Therefore Kriging error is essentially the error of variogram, that is used by Kriging. The more precisely variogram describes the distribution of input data, the less interpolation error. That is why the question of selection of the variographic model is very important while using Kriging.**

**In the paper results of application of Kriging-interpolation to results of COSMO model's forecast of pressure, temperature and precipitation are presented. Results,which are presented, were quite satisfactory, such, that do not inferior bilinear interpolation by accuracy, although Kriging method was used only the linear model.**

**Obviously it can be improvement of accuracy of interpolation while using variographic models, that describe more precisely distribution of input data. Determination of the optimal model is conducting by experimental means, based on COSMO model's forecast data. The work of determination of optimal models is extended on bigger sample data.**

## The optimality and the quasi-optimality of variographic models.

**As mentioned above, a great value in Kriging accuracy plays the variographic model, which reflect change of the parameter dependence in points from the distance between it. The variographic model is based on the experimental variogram, which in turn, is based on input data.**

**Forecast COSMO's data for each of meteorological parameters are submitted on the grid 101x209=N nodes with spatial step by latitude and longitude 0.125°. The grid covers area from 42.5° to 55° N and from 17° to 43° E, and thus the maximum distance between two points will be 28.85°. The experimental variogram is calculated by equation**

$$\gamma(\rho)=\frac{\sum\left(\left(\nu(x_i)-\nu(x_j)\right)^2\right)}{\mathrm{K}} \quad (1)$$

**for all** $i, j=\overline{1,N}:\left|x_i-x_j\right|=\rho$ **,** $\mathrm{K}$ **– the number of pairs of input points, the distance between it equal** $\rho$ **.**

**For COSMO's forecast data the number of pairs of the experimental variogram** $(\rho,\gamma(\rho))=10972$ **.**

**It is clear that building of the theoretical model with considering of such number of points is extremely difficult and resource intensive task. Also a large number of points may increase errors of variography. That is why for interpolation task of forecast data of meteorological parameters COSMO model, necessity of building of the smoothed experimental variogram is fully justified. We will choose lag for the smoothed experimental variogram, which will be equal to spatial step** $h=0.125$**.**

**On fig. 1 the experimental variogram and the smoothed experimental variogram for COSMO data forecast on 1 February 2014 00 UTC is presented.**

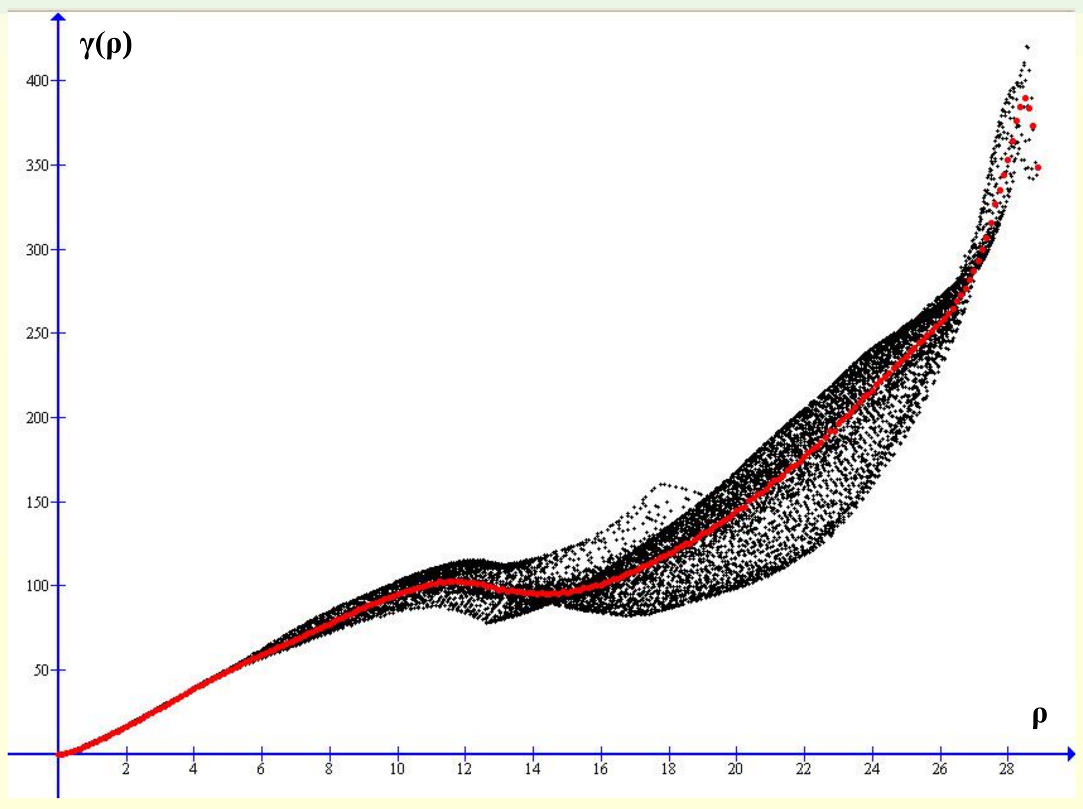



**Figure 1. The experimental (black) and the smoothed experimental (red) variograms for COSMO data forecast on 1 February 2014 00 UTC**

**Obviously, that building of the smoothed experimental variogram enables significantly simplify the procedure of coefficients determination of the theoretical model and reduce error, which arises because of the large number of nodes of the experimental variogram.**

**We build 10 theoretical models by the smoothed experimental variohgram: 0 - linear, 1 – exponential, 2 – Gaussian, 3 – spherical, 4 – quadratic, 5 – nagget-effect, 6 – cubic, 7 – logarithmic, 8 – power, 9 - circular. Among these models we need to choose that model, which is the best for displaying of the correlation data structure.**

**For model selection is necessary to establish criterion of optimality, i.e., to determine indicator, under which one model will be better than the other. It is natural in the choice of models to guided by certain deviations theoretical models from the experimental variogram. Determination absolute** $\varepsilon=\max_{i=\overline{1,N}}\left|y(h_i)-\gamma(h_i)\right|$ **and mean square errors** $\sigma=\sqrt{\frac{\sum_{i=1}^{N}\left(y(h_i)-\gamma(h_i)\right)^2}{N}}$ **is a widespread practice and appears quite justified in determining of the best model. However, numerical experiment, which is presented in showed, that models, which give least mean square error** $\sigma$ **and the smallest absolute error** $\varepsilon$ **on the same set of input data, as a rule, do not match. This is caused by presence of variogram's values, which strongly deviate from the general trend, but can not be rejected as false. Such extreme values give greater maximum error of variographic simulation, but their influence is reduced by smoothing in the interpolation process by Kriging method. The mean-square error gives better a better idea of the quality of experimental data modeling by theoretic function and its impact on the interpolation quality. Such conclusion confirms numerical experiment, which establishes dependence between minimum absolute and mean square errors of variography and Kriging-interpolation. Results of experiment are partly presented in table 1.**

**Necessary to pay attention, that mean-square errors of interpolation and variography are not comparable. If** $\sigma$ **interpolation determined on the values of meteorological variables with a range of changes [-a;a], then** $\sigma$ **of variography determined on values of experimental variogram, which has the range of variation** $\left[0,2a^2\right]$**. As a rule, for temperature and pressure the mean square error of variography will be bigger, than the mean square error of Kriging-interpolation, and for precipitation - conversely.**

**Table 1**

| Time | 03:00 | 06:00 | 09:00 | 12:00 | 15:00 | 18:00 | 21:00 |
|---|---|---|---|---|---|---|---|
| $\varepsilon_{var}$, | 3.0654 | 6.2571 | 10.6992 | 12.2096 | 8.3667 | 4.1142 | 5.7718 |
| № of model | 1 | 2 | 5 | 5 | 2 | 2 | 2 |
| $\varepsilon_{kr}$, | 7.4655 | 6.1451 | 9.1836 | 17.4249 | 13.0362 | 11.3223 | 10.8116 |
| № of model | 4 | 4 | 4 | 5 | 4 | 4 | 4 |
| $\sigma_{var}$, | 1.4348 | 2.9541 | 6.0682 | 6.7354 | 3.3970 | 2.8939 | 3.2002 |
| № of model | 1 | 0 | 5 | 0 | 0 | 2 | 2 |
| $\sigma_{kr}$, | 1.7875 | 1.7868 | 2.4779 | 2.9756 | 2.5154 | 1.9749 | 2.0331 |
| № of model | 1 | 0 | 5 | 0 | 0 | 2 | 2 |

**Minimum absolute** $\varepsilon_{var}$ **and mean square** $\sigma_{var}$ **deviations between experimental and theoretical variograms, minimum absolute** $\varepsilon_{kr}$ **and mean square** $\sigma_{kr}$ **deviations between results of Kriging-interpolation of forecast data COSMO and observation data of temperature for 8.04.2012. Models, on which the appropriate minimum deviations are received.**

**Experiment data show, that the least mean square error between Kriging-interpolation data and observation data achieved on variographic models, which have the lowest standard deviation from the numerical variogram. This statement is false, regarding the absolute error.**

**Experiment data show, that the least mean square error between Kriging-interpolation data and observation data achieved on variographic models, which have the lowest standard deviation from the numerical variogram. This statement is false, regarding the absolute error.**

**Guided by these considerations, we introduce the following definition.**

**Definition 1. Model, which has the smallest mean square deviation from the experimental variogram on dataset among models 0 – 9, will be considered the optimal model.**

**The minimum mean square deviation of theoretical variogram on dataset from the experimental will be denoted** $\sigma_{opt}$**, mean square error of Kriging-interpolation, conducted on the basis of the optimal model, will be denoted** $\sigma_{kr\_opt}$**.**

**Table 1 shows, that linear model is optimal on tree dataset. In table 2 are presented mean square errors, which linear model gives on another dataset for April 8, 2012, the difference between the least mean square error and errors of variogram and Kriging-interpolation, based on linear model.**

**Table 2**

| | 03:00 | 06:00 | 09:00 | 12:00 | 15:00 | 18:00 | 21:00 |
|---|---|---|---|---|---|---|---|
| $\sigma_{mod0}$ | 1.7028 | 2.9541 | 7.5292 | 6.7354 | 3.3970 | 3.5020 | 3.5887 |
| $\sigma_{kr0}$ | 1.7876 | 1.7868 | 2.4790 | 2.9756 | 2.5154 | 1.9751 | 2.0332 |
| $\lvert\sigma_{mod0}-\sigma_{opt}\rvert$ % from $\sigma_{opt}$ | 0.268 18.68% | 0 | 0.461 24,08% | 0 | 0 | 0.6081 21,01% | 0.3885 12,14% |
| $\lvert\sigma_{kr0}-\sigma_{kr\_opt}\rvert$ | 0.0001 | 0 | 0.0003 | 0 | 0 | 0.0002 | 0.0001 |

**The mean square deviation between experimental and linear variograms** $\sigma_{mod0}$ **the mean square deviation** $\sigma_{kr0}$ **between results of Kriging-interpolation, based on the linear model forecast data COSMO and observation data of temperature for 8.04.2012.**

**From table 2 we see, that mean square deviation of the linear variogram from the experimental is not so different from minimum on that dataset, where** $\sigma_{opt}$ **is received by another variographic model.** $\sigma_{kr0}$ **and** $\sigma_{kr\_opt}$ **differ among itselves still less. There is follow dependency: the smaller the difference between** $\sigma_{opt}$ **and** $\sigma_{mod0}$ **of linear variogram** $\lvert\sigma_{mod0}-\sigma_{opt}\rvert$**, the smaller the difference between the Kriging-interpolation, based on optimal and linear models.**

**Let** $f(\sigma_{opt})=a\sigma_{opt}+b$ **is linear function, which simulates the change of mean square error of Kriging-interpolation from the change of mean square error of variography per day,** $f$ **is based on a experimental data (fig. 2).** *fmax* **is maximum mean square error of Kriging-interpolation, based on optimal model for the indicated period. Let's determine, how much value** $\sigma_{opt}$ **can be changed, in order to value** $f\left(\sigma_{opt}+\Delta\sigma_{opt}\right)\le f_{\max}$ **. Let** $\Delta\sigma_{opt}=c\cdot\sigma_{opt}$**, where** $c>0$ **is some constant. Then** $f\left(\sigma_{opt}+\Delta\sigma_{opt}\right)=a\sigma_{opt}+a\cdot c\cdot\sigma_{opt}+b\le f_{\max}\Rightarrow c\le\frac{f_{\max}-b}{a\sigma_{opt}}-1$**.**

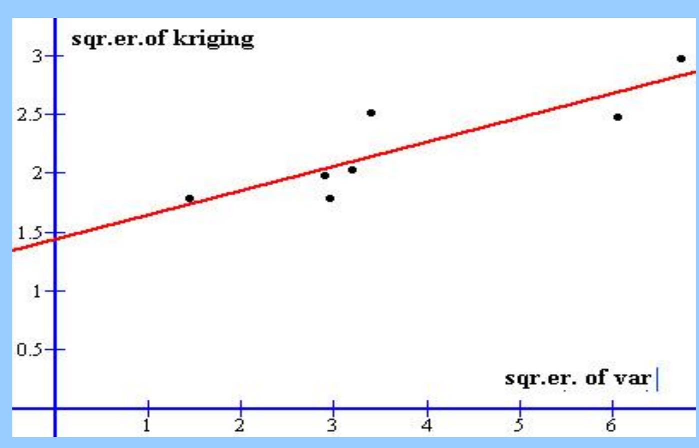



**Figure 2. Value of mean square errors of variography and Kriging-interpolation forecast data Cosmo for 8.04.2012 for temperature (°C), the function** $f(x)=0.21x+1.44$**, which simulates dependences between it**

**Constant** $c$ **is determined experimentally by formula**

$$c=\frac{f_{\max}-b}{a\cdot average\left(\sigma_{opt}\right)}-1, \quad (2)$$

**where** $average\left(\sigma_{opt}\right)$ **is average minimum mean square error of variography per day.**

**Definition 2. Model, mean square deviation of which from experimental variogram belong to segment** $\left[\sigma_{opt},\sigma_{opt}+c\cdot\sigma_{opt}\right)$**,** $c>0$ **will be called quasi-optimal variographic model.**

**It is clear that the value c for temperature, precipitation and surface pressure will differ. According to the forecast data observations for the 7-9-th April 2014 is identified, that for temperature** $c=0.3$**, for pressure** $c=0.9$**, for precipitation** $c=0.1$**.**

**From table 2 we can see, that quasi-optimal on COSMO model's forecast data of temperature for 8.04.2012 is linear model. The mean square error of Kriging-interpolation, based on this model is small, its value falls in the neighborhood errors of 30% , which are obtained on the basis of the optimal models in the respective datasets.**

## Results of the research

**For determination the optimal and quasi-optimal variographic models let's compute mean square errors of variography for models (1)-(10) on all datasets COSMO forecast for April 2013 - March 2014. Authors consider that this dataset is sufficient for determining and simulation the correlation structure of prognostic data for the year.**

**Temperature.**

**In table 3 is showed the percentage of data sets, in which each of the ten theoretical models is optimal and quasi-optimal (in brackets). In the case of, when model is not optimal on any dataset for the indicated month in table is specified sign "-".**

**Table 3**

**The results of determining optimal and quasi-optimal models, according to the COSMO model's forecast data of temperature for the period from April 2013 to March 2014**

| № of model | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| April 2013 | - | 6.7% (80%) | - | - | 30% (80%) | - | - | - | 10% (83.3%) | 53.3% (93.3%) |
| May 2013 | - | 16.7% (93.3%) | - | - | 13.4% (86.7%) | - | 13.3% (83.3%) | 13.4% (86.7%) | 16.7% (100%) | 26.5% (63.3%) |
| June 2013 | - | 14.3% (92.9%) | - | - | 10.7% (85.7%) | - | 32.1% (82.1%) | - | 14.3% (92.9%) | 28.6% (53.6%) |
| July 2013 | - | - | - | - | 25.9% (85.2) | - | 33.3% (85.2%) | - | 7.4% (100%) | 33.4% (66.7%) |
| August 2013 | - | - | - | - | 21.4% (50%) | - | 7.1% (50%) | - | - | 71.5% (100%) |
| September 2013 | - | - | - | - | 17.2% (100%) | - | - | - | 6.9% (100%) | 75.9% (96.5%) |
| October 2013 | - | 5% (65%) | - | - | 10% (55%) | - | - | - | - | 85% (100%) |
| November 2013 | - | 13.3% (96.7%) | - | - | 26.7% (90%) | - | - | - | - | 60% (100%) |
| December 2013 | - | - | - | - | - | - | - | - | - | 100% (100%) |
| January 2014 | - | 7% (83.3%) | - | - | 6.3% (73%) | - | - | - | 10% (86.7%) | 76.7% (100%) |
| February 2014 | - | - | - | - | - | - | - | - | - | 100% (100%) |
| March 2014 | - | - | - | - | - | - | - | - | - | 100% (100%) |

**Table 3 shows, that circular model is optimal in the cold half of the year (October - March) for most forecast datasets of temperature. This model is quasi-optimal on all data sets for the indicated period. Regarding the warm half of the year (April - September) - all is not so uniquely. Just as in the cold season, in April - September 2013 circular model is optimal at the majority datasets, however, it is already not quasi-optimal on all datasets for the warm period. In return, power model is quasi-optimal on all datasets in May, July and August. In April and June, none of the ten models are not quasi-optimal on all sets. To determine the model based on which is best for conducting Kriging-interpolation in these months are built the following diagrams: value the average per month σ for all models (fig. 3a, 4a), graphics of quasi-optimality for models, which are optimal in April and June at every dataset (fig. 3b, 4b). Information about optimality for models for MSLP and precipitation is shown on fig. 5.**

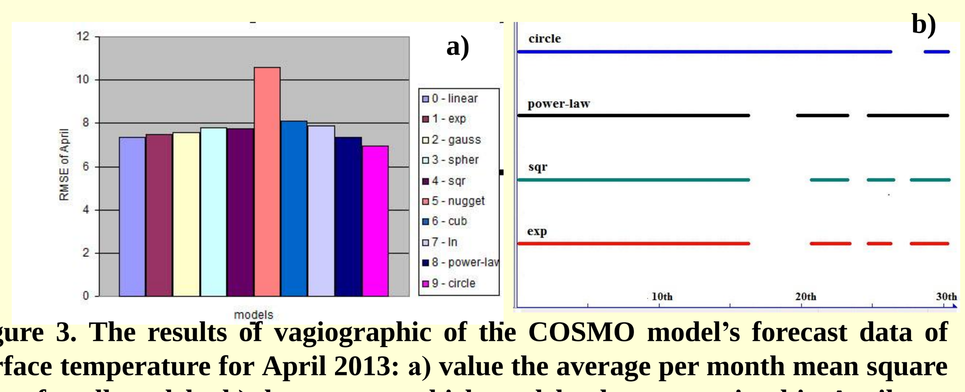



**Figure 3. The results of vagiographic of the COSMO model's forecast data of surface temperature for April 2013: a) value the average per month mean square error for all models; b) datasets, on which models, that are optimal in April, are quasi-optimal**

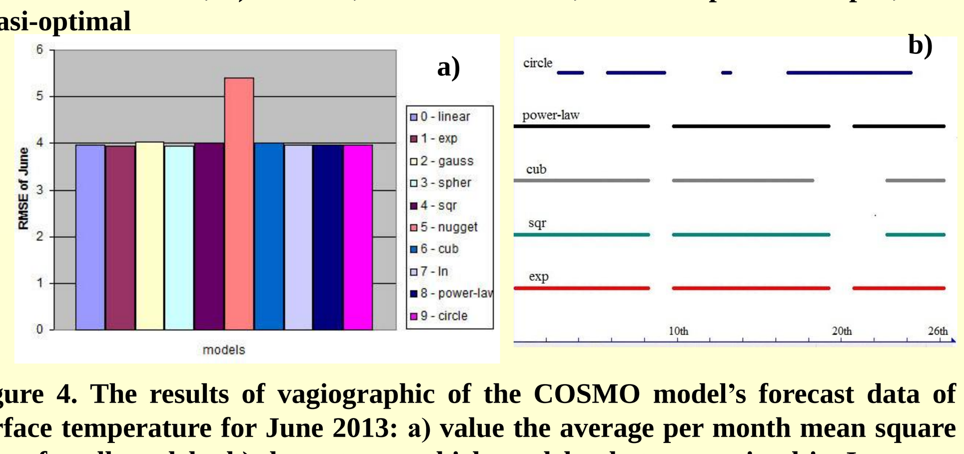



**Figure 4. The results of vagiographic of the COSMO model's forecast data of surface temperature for June 2013: a) value the average per month mean square error for all models; b) datasets, on which models, that are optimal in June, are quasi-optimal**

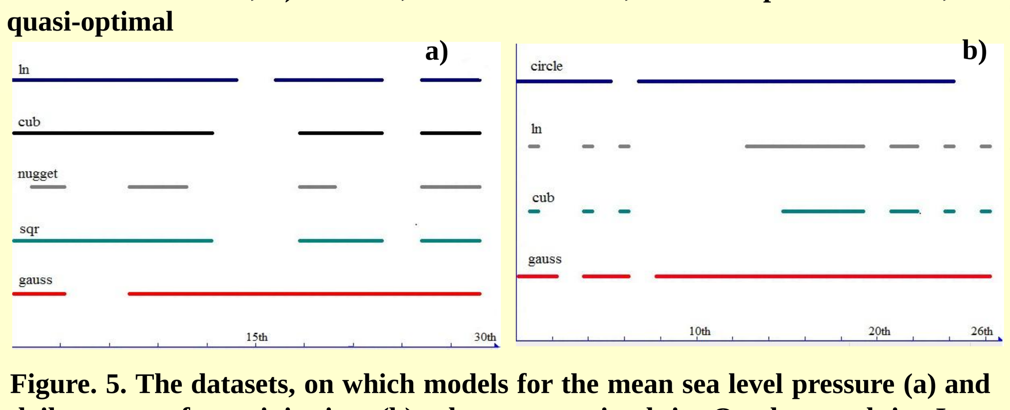



**Figure. 5. The datasets, on which models for the mean sea level pressure (a) and daily sums of precipitation (b), that are optimal in October and in June respectively, are quasi-optimal**

## Conclusions

**On the basis of theoretical considerations and numerical experiment, the authors set criteria of selection of the variographic models for Kriging-interpolation of meteorological parameters of COSMO model forecast data. According to the established criteria were defined concept of optimal and quasi-optimal models of variography.**

**The results of experimental determination of optimal and quasi-optimal variographic models for forecast data of MSLP, T2m and daily sums of precipitation by COSMO model were obtained for the period from April 2013 to March 2014. In result of the conducted analysis the following conclusions were done:**

- **For Kriging-interpolation forecast data of temperature is the best circular model, that is quasi-optimal in the most datasets per year. For magnification of precision in the process of interpolation by Kriging method data for temperature it should be conducted a choice among circular and power models, guided by the model's mean square deviation from experimental variogram as a selection criterion.**
- **For Kriging-interpolation forecast data of pressure the best is the Gaussian model. However, during Kriging-interpolation it is necessary to conduct a choice between the Gaussian and logarithmic model, that can to improve precision of interpolation.**
- **For Kriging-interpolation of forecast data of the amount of precipitation the best are circular and the Gaussian models.**