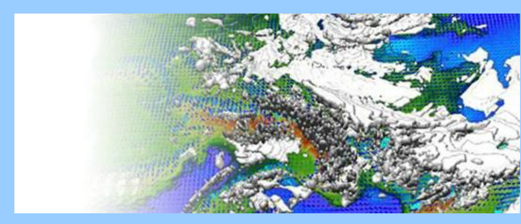
Variographic Models of Distribution of Meteorological Parameters on the Territory of Ukraine for Kriging-Interpolation



COSMO / CLM / ART - User Seminar 2015

Vitalii Shpyg and Lesia Katsalova

Ukrainian Hydrometeorological Institute, Kyiv, Ukraine (Vilal@rambler.ru)

Introduction.

Today in the world there is a small number of the atmospheric models of regional scale, which are enabling us to obtain the weather forecast for any territory. This list consists from ALADIN, HARMONIE, ETA, HIRLAM, WRF, COSMO etc. Practice has shown that the runs of the ETA, WRF and COSMO models for territory of Ukraine not enough takes into account the features of landscape, especially the underlying surface, which in turn leads to inaccuracy of weather forecast. Regional atmospheric models also give the value meteorological parameters on certain grid, which does not coincide with a grid of weather stations (observation network), which is irregular. The latter circumstance is somewhat difficult to work on the verification of the model and requires the use of interpolation methods. In 2013/2014 it was suggested using Kriging interpolation in order to bring the COSMO model data to grid of meteorological stations of Ukraine (see, Vitalii Shpyg and Lesia Katsalova, 2014).

Use of kriging involves building variographic models (Kanevskiy M.F. at al., 1999). The use of one or the other model affects on the accuracy of krging-interpolation and requires additional research. The purpose of the work is to build variographic models that are best suited to describe the spatial distribution of various meteorological parameters to the territory of Ukraine. To achieve this, the data of COSMO forecast of temperature, ground-level pressure and precipitation is researched and variographic models (variograms) were defined.

Variography.

One of the important properties of all natural phenomena is the spatial continuity: mutual dependence of the values at nearby points stronger than in remote points. Most often, the spatial continuity of the data described using semi-variograms (variograms). Determination of character dependence between data referred Variography. Variography essence is to determine the existence of the correlation structure of the data and its features. The ultimate goal of Variography is to build mathematical function that describes the spatial structure of the correlation data, which can then be used in geostatistical methods, including kriging. The quality of work of the method and the error value depends on the quality of the model.

The correlation analysis begins with the construction of the experimental variogram. Experimental variogram is calculated according to the formula for all pairs of points, the distance between which is equal to h:

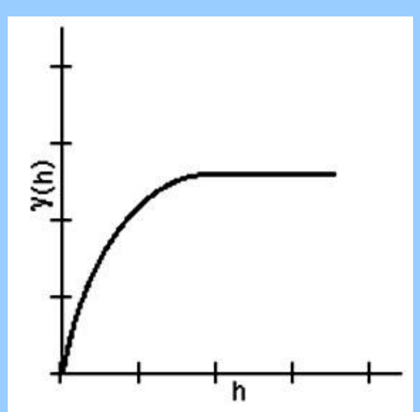
$$y(h) = 0.5 \cdot average\left(\left(V(x_i) - V(x_j)\right)^2\right) \qquad (1)$$

for all $i, j \in \{1,2,..n\}$, $|x_i - x_j| = h$.

The constructed spatial variogram, reflecting the spatial structure of the data are modeled using theoretical models. Variogram is a set of values for certain spatial steps, while the theoretical model allows to get value for each h.

The main types of variogram models are called according to the functions that describe them. There are models of the following types (McBratney A.B. and Webster R., 1986):

•Spherical model (1)

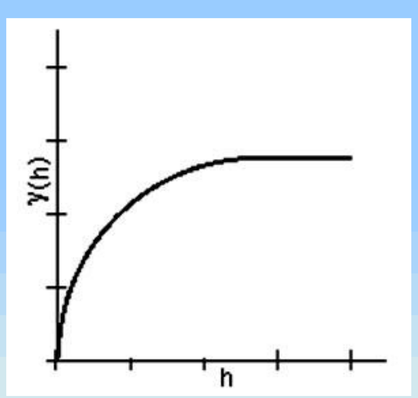


$$\gamma(h) = c_0 + c\left(\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right), \quad 0 < h \le a$$

$$\gamma(h) = c_0 + c, \quad h > a$$

$$\gamma(0) = 0$$

•Circular model (2)



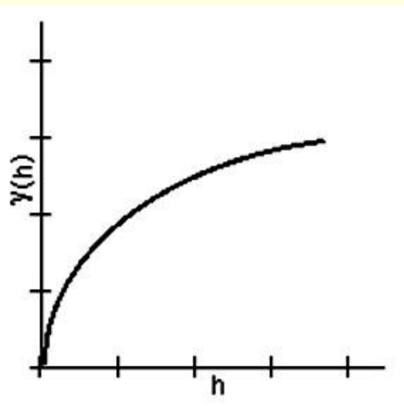
$$\gamma(h) = c_0 + c\frac{2}{\pi}\left(\frac{h}{a}\sqrt{1 - \frac{h^2}{a^2}} + \arcsin\left(\frac{h}{a}\right)\right),$$

$$0 < h \le a,$$

$$\gamma(h) = c_0 + c, \quad h > a,$$

$$\gamma(0) = 0$$

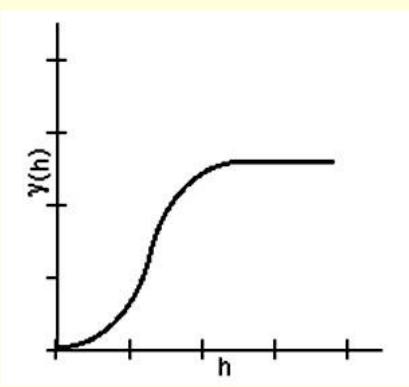
•Exponential model (3)



$$\gamma(h) = c_0 + c\left(1 - \exp\left(\frac{h}{r}\right)\right), \quad h > 0,$$

$$\gamma(0) = 0$$

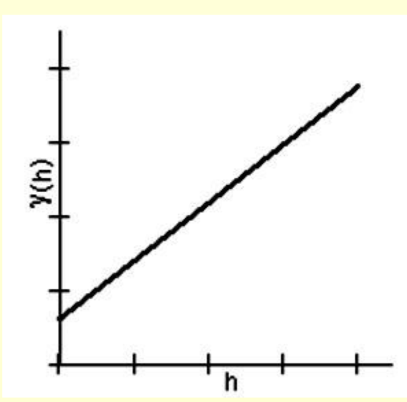
•Gauss model (4)



$$\gamma(h) = c_0 + c\left(1 - \exp\left(\frac{h^2}{r^2}\right)\right), \quad h > 0,$$

$$\gamma(0) = 0$$

•Linear model (5)

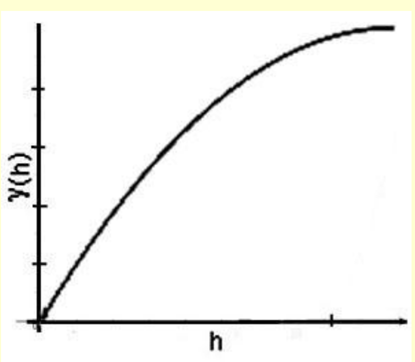


$$\gamma(h) = c_0 + c\left(\frac{h}{a}\right), \quad 0 < h \le a,$$

$$\gamma(h) = c_0 + c, \quad h > a,$$

$$\gamma(0) = 0$$

•Quadratic model (6)

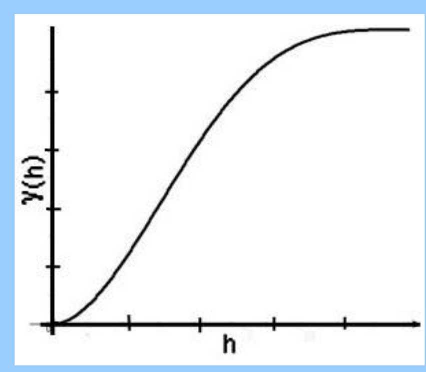


$$\gamma(h) = c_0 + c\frac{h}{a}\left(1 - \frac{h}{a}\right), \quad h > 0,$$

$$\gamma(h) = c_0 + c, \quad h > a,$$

$$\gamma(0) = 0$$

•Cubic model (7)



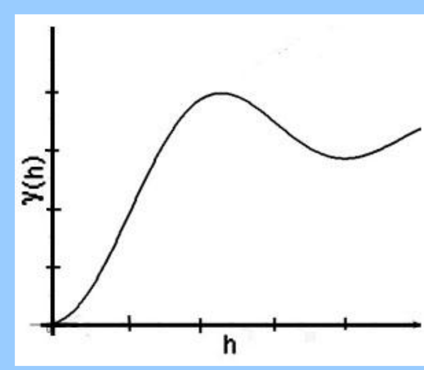
$$\gamma(h) = c_0 + c\left(\frac{7h^2}{a^2} - \frac{35h^3}{4a^3} + \frac{7h^5}{2a^5} - \frac{3h^7}{4a^7}\right),$$

$$h > 0,$$

$$\gamma(h) = c_0 + c, \quad h > a,$$

$$\gamma(0) = 0$$

•Nugget-effect model (8)

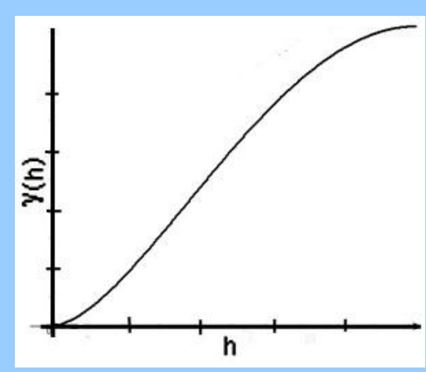


$$\gamma(h) = c_0 + c\left(1 - \frac{1}{h}\sin(h)\right),$$

$$h > 0,$$

$$\gamma(0) = 0$$

•Logarithmic model (9)



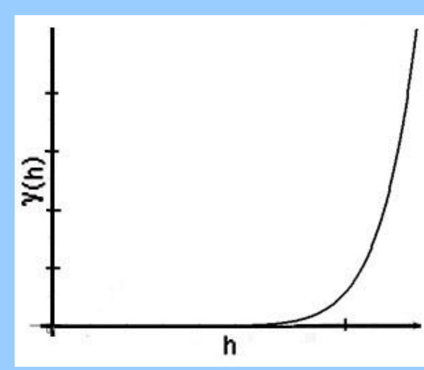
$$\gamma(h) = c_0 + c\left(\frac{h}{a}\right)^2\left(1 - \ln\left(\frac{h^2}{a^2}\right)\right),$$

$$h > 0,$$

$$\gamma(h) = c_0 + c, \quad h > a,$$

$$\gamma(0) = 0$$

•Power-law model (10)



$$\gamma(h) = c_0 + c(h)^a,$$

$$h > 0,$$

$$\gamma(0) = 0$$

a is valid correlation radius: the distance at which there is data dependency. For spherical and circular models $\gamma(a) = c_0 + c$, that is at a distance variogram value does not exceed a certain value (variography plateau). This means that the dependence of data on distance that more valid radius available. *r* is effective correlation radius: at this distance the variogram reaches 95% of its plateau. In Gauss and exponential models plateau is reached asymptotically. This means that the data dependence exists everywhere, but decreases with increasing distance between points. Linear and power-law model fit to the data dependency between them remains strong on long distances between points. For data dependence between them poor fit model "nugget-effect".

Determination of the model.

The distributions of values of various meteorological variables on the territory of Ukraine are widely divergent. The interdependence of data for each meteorological parameter for different distances differs. Therefore, using Kriging interpolation data for weather forecast needs a additional research that aims to determine the optimal variohraphic models for each parameter. To do this, the authors proposed the following approach: determine the optimal model for each data set of temperature, pressure and precipitation as the most important components of weather forecast, obtained using COSMO, and, after analysis of the results to determine the optimal model for each of the selected parameters as a whole.

To determine the optimal model uses a method of sorting theoretical models which are based on certain experimental variogram. To determine the optimal model using maximum $\varepsilon = \max_{i=1,N}\left|y(h_i) - \gamma(h_i)\right|$ and mean square $\sigma = \frac{\sum_{i=1}^{N}\left(y(h_i) - \gamma(h_i)\right)^2}{N}$ errors between the theoretical variogram values from the experimental variogram.

Although the selection of models for Kriging interpolation is used mainly mean square error, the authors believe that for forecast data interpolation more telling is the maximum error. Its value at some point can be compensated by greater accuracy in other points. In such cases, σ is small at large ε which can lead to large interpolation error at some point, so to inaccurate weather forecasts for specific areas.

The model is considered optimal for a given data set, if it is the smallest error.

Research conducted on a sample of data prediction of temperature, pressure and precipitation for April, July, October 2013 and January 2014. This sampling allows to determine the dependence of data not only in relation to individual meteorological parameters and the change it variography depending on the season.

Thus, the research conducted by the following algorithm:
•for each set of COSMO forecast data of temperature, pressure, precipitation is constructed experimental variogram;
•on the points of experimental variogram by the method of least squares build models (1) - (10) presented in the previous section (see Fig. 1);
•for each of the constructed models define the maximum and mean square error;
•define the model for which the smallest error value;
•perform the above procedure for each set of sampling.

<u>*Temperature.*</u> The results determine variographic model for kriging interpolation data of COSMO model forecast of temperature given in Table 1.

From Table 1 we can see that kriging interpolation data forecast of temperature is no one model that would give the best approximation for all data sets. This means that the dependence between data differs for different points in time. Physically, this is justified because the temperature field of Ukraine formed under different atmospheric processes with different scales, different length and other characteristics. In fact, the results indicate the dynamics of the temperature field and the constant change not only the temperature but also the degree of dependence between these values.

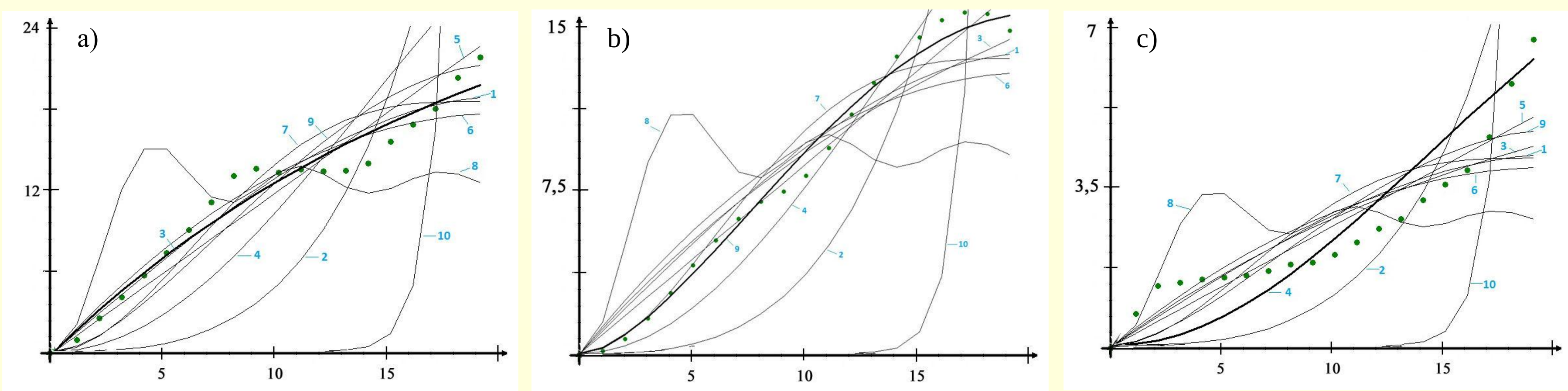


**Fig. 1. Charts models (1) - (10) constructed for experimental variogram:
a) temperature, b) pressure, c) precipitation, 9:00, 8.04.2012; thick lines show the optimal model.**

Table 1. Presentation of optimality models (1) - (10) for data of COSMO forecast of temperature as a percentage of the number of processed data sets (100%)

Model		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
April 2013 p.	ε	-	-	12.76	23.4	12.77	51.07	-	-	-	-
	σ	-	-	14.9	8.51	17.02	42.55	-	17.02	-	-
July 2013 p.	ε	-	-	18.18	18.18	20.45	43.19	-	-	-	-
	σ	-	-	43.19	-	13.64	34.09	-	9.08	-	-
October 2013 p.	ε	-	-	-	48.58	20	20	-	-	-	11.42
	σ	-	-	11.43	34.29	20	34.28	-	-	-	-
January 2014 p.	ε	-	-	12.77	27.66	34.04	8.51	-	-	-	17.02
	σ	-	-	23.4	29.79	31.92	14.89	-	-	-	-
All data	ε	-	-	10.93	29.45	21.82	30.7	-	-	-	7.11
	σ	-	-	23.23	18.15	20.65	31.45	-	6.52	-	-

Table 2. Presentation of optimality models (1) - (10) for data of COSMO forecast of pressure as a percentage of the number of processed data sets (100%)

Model		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
April 2013 p.	ε	-	-	-	45.1	27.45	27.45	-	-	-	-
	σ	-	-	-	17.65	11.76	37.35	-	9.7	23.54	-
July 2013 p.	ε	-	-	-	27.27	13.64	18.18	13.64	-	27.27	-
	σ	-	-	-	44.19	-	-	23.26	-	32.55	-
October 2013 p.	ε	-	-	-	28.58	-	37.14	-	17.14	17.14	-
	σ	-	-	-	20.0	-	28.57	11.43	22.58	17.14	-
January 2014 p.	ε	-	-	7.84	49.02	19.61	15.69	7.84	-	-	-
	σ	-	-	17.6	27.45	9.8	25.5	9.8	9.8	-	-
All data	ε	-	-	1.96	38	15.18	24.62	5.37	4.35	11.17	-
	σ	-	-	4.41	27.33	5.40	22.85	11.12	10.52	18.37	-

However, Table. 1 shows that the proposed ten variographic models not optimal for kriging interpolation of temperature field in Ukraine are quadratic, exponential, Gauss and linear models. The optimality of these models indicates that there is dependence between temperature values at the points that can asymptotically decrease with increasing distance between points. Differ only asymptotic of decrease of dependence.

The fact that the model (3) - (5) are optimal on the maximum error and on the mean square error also indicates the existence variographic dependence between data of forecast of temperature.

Analyzing the data by month, we see that there are some differences: for the autumn-winter period is characterized by a strong variography than for spring and summer. This is evidenced decline in the share of quadratic model and the presence of a power-law model in the optimal models for October and January. For spring and summer more suitable model, that point to reduced dependence data with increasing distance between points. However, the difference is negligible, therefore the authors recommended at kriging interpolation check models (3) - (5) for determine optimal.

Table 3. Presentation of optimality models (1) - (10) for data of COSMO forecast of precipitation as a percentage of the number of processed data sets (100%)

Model		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
April 2013 p.	ε	-	-	-	-	-	-	-	61.70	-	38.3
	σ	-	-	-	8.51	-	-	-	61.70	-	29.79
July 2013 p.	ε	-	-	-	9.09	-	-	-	77.27	-	13.64
	σ	-	13.64	-	-	13.64	-	-	72.72	-	
October 2013 p.	ε	-	-	-	-	-	-	-	60	-	40
	σ	-	31.42	-	22.87	-	-	-	45.71	-	
January 2014 p.	ε	-	27.45	-	13.72	-	7.84	-	50.99	-	-
	σ	8.50	27.66	-	17.03	-	-	-	46.81	-	-
All data	ε	-	6.87	-	5.7	-	1.96	-	62.49	-	22.98
	σ	2.13	18.18	-	12.1	3.41	-	-	56.74	-	7.44

<u>*Pressure.*</u> From Table 2 we see that kriging interpolation data forecast of pressure, as for temperature, is no one model that would give the best approximation for all data sets. From the experimental results follow that the dynamics of atmospheric circulation leads to more active changes in the correlation structure of the pressure field than the temperature field. To a large extent, this structure is described by a Gauss, quadratic, linear and logarithmic model. However, for some part of data, optimal models are also exponential, cubic models and model nugget-effect. For kriging interpolation of pressure field on a grid of weather stations of Ukraine recommended the selection of the optimal model among models (3) - (9), which will allow maximize precision interpolation.

<u>*Precipitation.*</u> From Table 3 we see that variographic analysis for data of forecast of precipitation is more definite than for data of forecast of temperature and pressure. In particular, there is a clear preference of "nugget-effect" model, which means no dependence between the data on the territory of Ukraine.

However optimality of other models for a number of observational indicates global processes responsible for precipitation on the settlement area, and which provide dependence between the data. With that in spring and summer dependence between these stronger, as evidenced by optimality of power-law model. For data for October and January there is optimality of spherical and circular models, that indicate that exist the dependence at a certain distance.

Conclusions. It is shown that none of meteorological parameters in question are not subject to the one model. However, there is a sampling of models that achieve optimality on data forecasting of certain weather parameters.