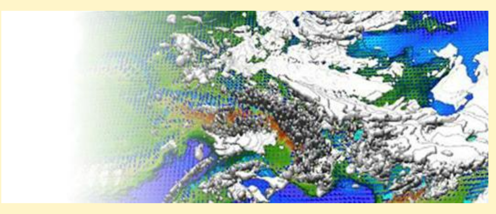

Application of Kriging-Interpolation for COSMO Weather Forecast

Vitalii Shpyg and Lesia Katsalova

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

COSMO / CLM / ART - User Seminar 2014

Introduction.

Modern international practice of weather forecast based on the use of mathematical models whose work is going in three stages: model pre-processing model implementation, model post-processing [1]. Weather forecast models are based on multi-dimensional hydrodynamic equations. In an unpredictable changes of meteorological parameters this equations do not have analytical solutions [2]. Numerical methods are used for solving such equations require a certain kind of input and output data. Discretization of space (grid partition), at the points where the initial data presented and forecast got, carried out. But as a rule, the measurement data are not evenly distributed. And the forecast is often neededs in the points of space that do not coincide with the grid points.

That is why the important problem of numerical weather prediction is the problem of analysis of spatially distributed information on meteorological data. This problem is represented by a lot of specific tasks:

- estimate the value in the point at which the measurements were not (model pre-processing);
- determine the value in the point which does not belong to model grid (model post-processing);
- evaluate the interpolation error.

Data analysis and results depend both on the quality and quantity of input data and methods of their processing.

Today there is a wide range of methods and tools of analysis, processing and presentation of spatially distributed information. Interpolators, methods, based on the statistical interpretation of data, artificial intelligence algorithms among them.

For better spatial estimation (in a statistical sense) using kriging. An important property of kriging is an exact recreation of measurement values at the points where they are known. Kriging estimation accompanied by interpolation error estimate at each point unlike many interpolation methods.

In this article essence of kriging and results of its application for interpolation of meteorological parameters (pressure, temperature and precipitation) are displayed. Data obtained by the model COSMO and interpolated on non-uniform grid of weather stations in Ukraine. The objective of this work is to determine the efficiency of the method for interpolation of meteorological parameters and the appropriateness of its application to adapt the world's best mathematical weather forecast models for the territory of Ukraine.

Kriging.

Consider the problem of assessing the value of a meteorological parameter V at some point in the computational grid x. Basis for this will serve as a set of n measurements of meteorological parameters that create one at the points $x_1, x_2, \ldots, x_n$ in the computational domain. It is necessary to build an interpolation function V^* to be a good estimate of the unknown function V. The basic method of estimation in geostatistics is kriging. The term kriging was taken to refer to a family of algorithms of spatial regression [3]. The word "kriging" comes from the name of the scientist D. Krige, in honor of which G. Materon called the basic geostatistic method.

Like other classic interpolation methods kriging based on the calculation of weights for each point x. Estimation of the value of meteorological parameter at the point x is determined as a linear combination of known values with these weights

$$V^*(x) = \sum_{i=1}^{n} w_i(x) V(x_i) \qquad (1)$$

The method relies on the assumption that the function $V(x)$ is a random process. Accordingly $V(x_1), V(x_2), \ldots, V(x_n)$ is random values. Then linear combination $V^*(x)$ is random value too. Coefficients $w_i(x)$ are calculated so that the expectation value of $V^*(x)$ was equal to the expectation value of $V(x)$ at this point

$$M(V^*(x)) = M(V(x)) \qquad (2a)$$

Condition (2a) provides unbiasedness property of method. Optimal estimates provided by the following condition: the variance of the difference $V^*(x) - V(x)$ should be minimal

$$D(V^*(x) - V(x)) \to \min \qquad (2b)$$

Thus kriging provides the best unbiased estimate.

To fulfill the conditions (2) kriging should have some idea about the nature of random process $V(x)$, to know the model of random process. As the model uses function $\gamma(a,b)$ that characterizes the dependence of the difference of values at the points a and b of the distances between these points.

$$\gamma(a,b) = D(V(a) - V(b)) \qquad (3)$$

Then, according to [4], requiring the fulfillment of conditions (2), we find a system of kriging equations. We get optimal values of weights $w_1(x), w_2(x), \ldots, w_n(x)$ from this system:

$$\begin{pmatrix} w_1(x) \\ \ldots \\ w_n(x) \\ \mu(x) \end{pmatrix} = A^{-1} \begin{pmatrix} \gamma(x, x_1) \\ \ldots \\ \gamma(x, x_n) \\ 1 \end{pmatrix}$$

$$A = \begin{pmatrix} \gamma(x_1, x_1) & \ldots & \gamma(x_1, x_n) & 1 \\ \ldots & \ldots & \ldots & \ldots \\ \gamma(x_n, x_1) & \ldots & \gamma(x_n, x_n) & 1 \\ 1 & \ldots & 1 & 0 \end{pmatrix} \qquad (4)$$

$\mu(x)$ is Lagrange multiplier that allows to determine kriging error. The resulting weights $w_1(x), w_2(x), \ldots, w_n(x)$ from system (4) substitute into the formula (1) and obtain required $V^*(x)$.

Thus for calculating of $w_1(x), w_2(x), \ldots, w_n(x)$ are not used values $V(x_1), V(x_2), \ldots, V(x_n)$. Use only stationing of points $x_1, x_2, \ldots, x_n$ and model of random process – variogram. As to construct the variogram use values $V(x_1), V(x_2), \ldots, V(x_n)$, it is clear that the result of interpolation depends nonlinearly on these quantities.

In [4,5] it is shown that the variogram is constructed on the following algorithm:

- first, according (3), construct experimental variogram using $V(x_1), V(x_2), \ldots, V(x_n)$;
- experimental variogram approximated model function values which are used in the system (4).

Experimental variogram is calculated by the following formula for all pairs of points, the distance between which is equals to h:

$$\gamma(h) = 0.5 \cdot average\left(\left(V(x_i) - V(x_j)\right)^2\right) \qquad (5)$$

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

In the case where each pair has a unique distance, placing all pairs depending of the distance between points becomes unmanageable. This is typical for the spatial placing of meteorological stations. In this case distances are grouped in lag bins [5]. For example, find average variogram for all pairs, the distance between which is more than 10 km and less than 20 km. Thus experimental variogram is chart: distance along the axis and average values (5) along the axis y.

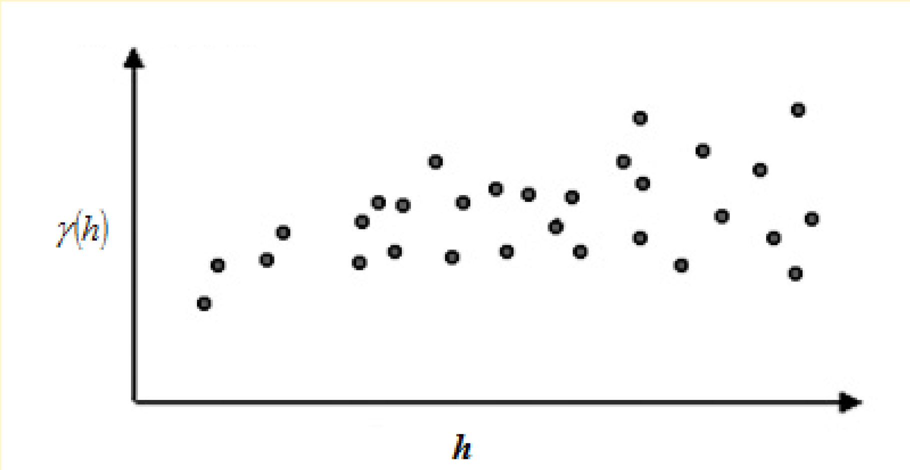

Fig. 1. Example diagram of experimental variogram

The next step is establishing a model (continuous function) for points of experimental variogram. To do this, choose the function that will serve as a model. There is a large selection of variogram models, but most have used following models:

- circular
- spherical
- exponential
- Gause
- linear

Model selection affects the prognosis of unknown values. The steeper the curve at the beginning of the variogram, the greater the impact on the forecast will have immediate environment. As a result, the resulting surface will be less smooth. Each model is developed to more closely correspond to different types of phenomena.

Also forecast meteorological parameter values kriging allows to estimate the error itself. Error is defined as the dispersion of deviation of the calculated values from the true unknown

$$D(V^*(x) - V(x)) = \begin{pmatrix} w_1(x) \\ \ldots \\ w_n(x) \\ \mu(x) \end{pmatrix} \begin{pmatrix} \gamma(x, x_1) \\ \ldots \\ \gamma(x, x_n) \\ 1 \end{pmatrix} \qquad (5)$$

But, according to [3], the value of the error given by the formula (6) may be understated. Therefore it is necessary conduct additional research of kriging accuracy.

Model post-processing data interpolation.

Four mesoscale atmospheric numerical weather prediction model (WRF ARW v.2.2.1, WRF ARW v.3.3.1, WRF NMM v.3.3.1 and COSMO [6]) and forecast wind and wave mode Black and Azov Seas model operated in UHMI. In the process of adaptation and verification of the model for the territory of Ukraine output data interpolated in the points, in which the meteorological station of our country are located. This is necessary to make a comparison of the forecast model results with actual measurements of meteorological parameters. It characterizes the model quality for the territory of Ukraine. Today, for interpolation of all atmospheric models used a local bilinear interpolation. The essence of this interpolation consists in carrying out local interpolation horizontal axis first, and then the vertical axis. Considering the distance between the model grid (more 10 km), it is clear that such an approach can give results with large error. Application of kriging in this situation seems more excuses.

Kriging was used to interpolate of the results of numerical weather forecast model COSMO [7]. This model is created within the eponymous consortium (Consortium for Small-scale Modeling), which was founded in October 1998. The main purpose of the consortium is to create, maintain and improvement non-hydrostatic atmospheric models for a limited territory for the needs of operational practices and research. The main members of the consortium are the meteorological services of Germany, Switzerland, Poland, Romania, Russia and four research centers in Germany and Italy. Model COSMO for the first time in the operational practice was introduced in 1999 in Germany. Since then, the model has proven itself as a reliable method for weather forecast, and on the characteristics and quality of predicting surpassed the model, which was developed by the German Weather Service, and which operated in 29 countries (mainly on commercial basis). In UHMI COSMO was implemented in computer architecture x86_64 in OC CentOS v.6.2 environment using the system compilers that are installed on the platform gcc and gfortran. Support of multiprocessor calculations produced using the library OpenMPI. The computational domain - 209 x 101 nodes, steps in horizontal – 0,125° (~14 km), 50 levels in the vertical. Coordinates of extreme south-west nod: 42,5° north latitude and 17° west longitude. Forecast is given for 75 hours.

At realization of kriging linear model was used. The linear model is used in cases when the model of the spatial distribution of the interpolated values not found.

Test results.

For testing were chosen COSMO forecast results for 08.04.2012. On this day, the dynamics and the distribution of values of meteorological parameters on the territory of Ukraine distinguished by particular instability and sudden changes. According to the authors, it is this synoptic situation will allow fully evaluate the work of method and check its efficiency.

Testing was conducted in two stages:

- comparison of kriging interpolation results with measurements at meteorological stations in Ukraine;
- solution of inverse problem: interpolation values, obtained by kriging on the grid weather stations of Ukraine, in grid nodes of model COSMO. Comparison of inverse interpolation results with model forecast results.

In the first stage were compared kriging interpolation results with real data and with bilinear local interpolation results which use for model post-processing in UHMI. The table 1 presents maximum δ_M and average error σ, which are computed on non-uniform grid of weather stations for temperature and precipitation for a certain time:

$$\delta_M = \max_{0 \le i \le N} \left| \Re(x_i, T) - y_{i,T} \right|, \quad \sigma = \frac{\sum_{i=1}^{N} \left(\Re(x_i, T) - y_i^K \right)}{N}$$

$\Re(x_i, T)$ is value of some meteorological parameter in the point x_i at time T obtained by interpolation, y_i^T is value of the same parameter at time T measured at i-th meteorological station (measured at the meteorological station), $T \in \{03, 06, 09, 12, 15, 18, 21\}$ hrs, N is number of weather stations in Ukraine, M is grid node number, which was obtained by the maximum deviation y from $\Re$.

Table 1. Average and maximum deviation bilinear and kriging interpolation results from values measurements of meteorological parameters (temperature and precipitation) on grid of weather stations in Ukraine

time	temperature				precipitation			
	$\delta_{bl,\max}$	$\delta_{kr,\max}$	$\sigma_{bl,av}$	$\sigma_{kr,av}$	$\delta_{bl,\max}$	$\delta_{kr,\max}$	$\sigma_{bl,av}$	$\sigma_{kr,av}$
03	7.50	7.46	1.27	0.76	-	-	-	-
06	6.00	6.14	1.32	0.71	10.90	10.80	1.16	1.14
09	9.60	9.18	1.85	1.24	-	-	-	-
12	17.30	17.42	2.14	1.67	4.40	3.90	0.43	0.46
15	13.10	13.03	1.91	1.29	-	-	-	-
18	11.70	11.32	1.47	0.69	35.20	35.20	2.07	2.02
21	11.20	10.81	1.30	0.63	-	-	-	-

Note: T is period of observation (UTC); $\delta_{bl,\max}$ is maximum error of meteorological values obtained using bilinear interpolation; $\delta_{kr,\max}$ is maximum error of meteorological values obtained kriging; $\sigma_{bl,av}$ is average error of meteorological values obtained using bilinear interpolation; $\sigma_{kr,av}$ is average error of meteorological values obtained kriging.

Data in the table shows that kriging interpolates data somewhat better than bilinear interpolation. It should be noted that the large maximum errors are caused by model COSMO error for mountain and coastal meteorological stations. It is assumed that the use of kriging interpolation with consideration of height will improve the forecast for the respective stations. The problem goes beyond the subject article and will be examined by the authors in the future.

In the second stage were compared inverse problem results with COSMO forecast results. As before, maximum δ_M and average error σ, which are computed on non-uniform grid of weather stations for meteorological parameters for a certain time are found. The table 2 and figures 2, 3 presents its errors for comparison.

Table 2. Average and maximum deviation inverse problem results from COSMO forecast result

time	pressure		temperature		precipitation	
	δ_M	σ	δ_M	σ	δ_M	σ
03	2.99	0.37	6.74	0.69	1.59	0.04
06	3.97	0.44	6.18	0.83	1.31	0.03
09	4.89	0.54	9.15	1.3	1.03	0.04
12	5.53	0.60	8.92	1.33	1.23	0.06
15	5.87	0.49	8.7	1.23	1.75	0.07
18	5.02	0.46	7.08	0.98	1.82	0.05
21	5.26	0.44	6.46	0.83	0.98	0.04

Note: T is period of observation (UTC); δ_M is maximum error of meteorological values; σ is average error of meteorological values.

a) b)

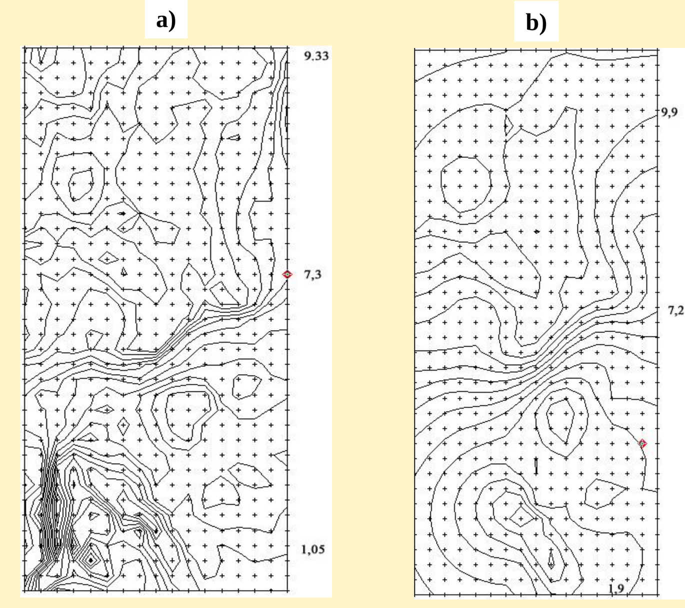

Fig. 2. Fragment of the spatial distribution of surface air temperature [°C] 08.04.2012 (21:00 UTC)
a) map, based on COSMO forecast data;
b) map, based on inverse problem results.

a) b)

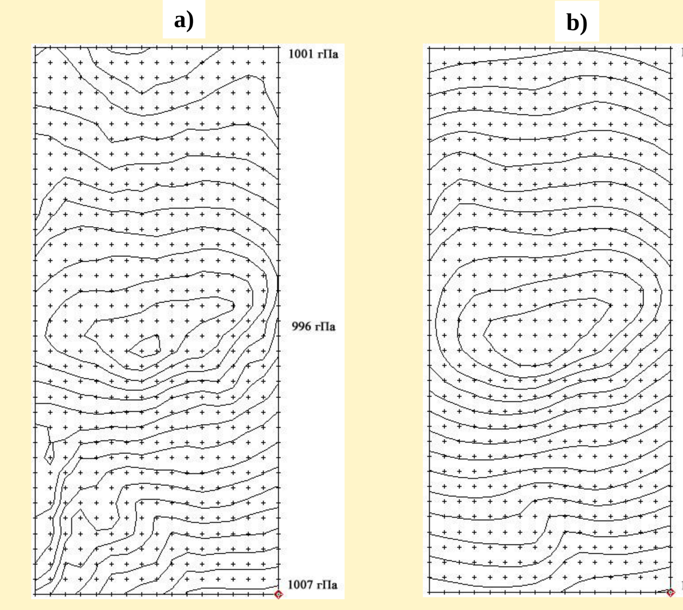

Fig. 3. Fragment of the spatial distribution of reduced atmospheric pressure [hPa] 08.04.2012 (21:00 UTC)
a) map, based on COSMO forecast data;
b) map, based on inverse problem results.

Conclusions.

Errors obtained during the solution direct and inverse problems are acceptable and allow to assert that kriging can be used for forecast data interpolation. Kriging little blurs field values. This is natural, since the method was based on a linear model. At the same time, in the course of experimentation dependence of interpolation efficiency by model selection for different meteorological parameters was found by the authors.

The authors were invited to use the Kriging method for model post-processing for the following reasons:

- kriging is based on the conditions and provides the best unbiased estimate of value of parameter at some point;
- error of the method does not depend on distances between the input data and the point where estimated value of parameter;
- kriging allows interpolation in 3-dimensional space (in two horizontal and vertical directions).

It should be noted that, depending on the problem, which is solved, the use of kriging interpolation is more expensive in terms of computer time than most other interpolation methods (triangulation, the method of inverse squares, splines, etc.).

Analysis of test results showed that the method gives a good estimate. Makes sense to use kriging-interpolation for weather forecast model post-processing.

References.

1. *Prusov V.A., Doroshenko A.Y.* Simulation of natural and anthropogenic processes in the atmosphere. – Kyiv: Naukova Dumka, 2006. – 542 p.
2. *Marchuk G.I.* Numerical methods in weather forecasting. – Leningrad: Gidrometeoizdat, 1967. – 353 p.
3. *Oliver M.A.* Kriging: A Method of Interpolation for Geographical Information Systems // International Journal of Geographic Information Systems. – 1990. –Vol. 4. – P. 313-332.
4. *Kanevskiy M.F., Demyanov V.V., Savelyev E.A., Chernov S.Y., Timonin V.A.* An elementary introduction to geo statics // VINITI. Series "Problems of the Environment and Natural Resources", Moscow. – 1999. – №11. – 135 p.
5. Help of ArcGIS 10.1: http://resources.arcgis.com
6. Official website UHMI: http://uhmi.org.ua/
7. *Doms G.* A Description of the Nonhydrostatic Regional COSMO-Model: www.cosmo-model.org