ANISOTROPIC KERNEL ESTIMATES OF 2D RANDOM FIELDS

Alexey A. Lyubushin

Russian Academy of Sciences, Institute of the Physics of the Earth,
123995, Moscow, Russia, Bolshaya Gruzinskaya, 10;
e-mail: lubushin@mtu-net.ru

A new method for solution of the classic geo-statistical problem – interpolating data from random non-regular set of observation points into the nodes of regular grid (so called gridding) is proposed. The method is based on using kernel estimates which does not need traditional empirical variogram analysis which is used in kriging approach [Deutsch, Journel, 1997]. It is essential to underline that an anisotropic kernel functions are used and that parameters of anisotropy are estimated in parallel to constructing the main estimate of the random field. For comparison – in traditional kriging approach anisotropy parameters must be introduced into the geo-statistical model at the stage of variogram analysis, before estimating (for instance, using some expert knowledge) and this anisotropy must be uniform within the estimated domain. In the proposed method the anisotropy parameters are defined automatically and they can vary from point to point within the estimated domain, i.e. the anisotropy could be non-uniform.

The method is illustrated by application it to a set of 2D-observations: rainfall data in Switzerland, the depth of soils at the bottom of Aral sea and the radionuclide pollution at the Bryansk region of Russia after Chernobyl catastrophe.

Let $(\mathbf{x}_\alpha, \mathbf{y}_\alpha, \mathbf{Z}_\alpha), \alpha = 1,...,\mathbf{n}$ be the results of measurements of some random field $\mathbf{Z}(\mathbf{x},\mathbf{y})$ in the points $(\mathbf{x}_\alpha, \mathbf{y}_\alpha)$ of the irregular network. Let $(\mathbf{x}_c, \mathbf{y}_c)$ be some point for which we are interesting in obtaining some estimate $\hat{\mathbf{Z}}(\mathbf{x}_c, \mathbf{y}_c)$ of field. We shall use a method, based on kernel estimate, which use anisotropic Gaussian kernel functions.

For this purpose let us consider the function of 2D Gaussian probability density:

$$\begin{aligned}
&\Phi(\mathbf{x},\mathbf{y}\,|\,\vartheta) = \frac{1}{2\pi\sigma_1\sigma_2}\cdot \exp(-\mathbf{Q}(\mathbf{x},\mathbf{y}\,|\,\vartheta)/2), \quad \mathbf{Q}(\mathbf{x},\mathbf{y}\,|\,\vartheta) = \\
&= \mathbf{x}^2\left(\frac{\sin^2\varphi}{\sigma_2^2} + \frac{\cos^2\varphi}{\sigma_1^2}\right) + \mathbf{y}^2\left(\frac{\sin^2\varphi}{\sigma_1^2} + \frac{\cos^2\varphi}{\sigma_2^2}\right) + 2\mathbf{xy}\cdot\sin\varphi\cdot\cos\varphi\cdot\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) \\
&\vartheta = (\sigma_1, \sigma_2, \varphi); \quad 0 < \sigma_2 \le \sigma_1; \quad 0 \le \varphi \le \pi
\end{aligned} \tag{1}$$

Here $(\sigma_1, \sigma_2), 0 < \sigma_2 \le \sigma_1$ are standard deviations of 2 orthogonal principal components of Gaussian 2D distribution. By other words (σ_1, σ_2) are the half-lengths of axes of 2D Gaussian scattering ellipse. The parameter $0 \le \varphi \le \pi$ is the angle between the large ellipse axis and the X-axis.

Let $\mathbf{m}, \mathbf{m} \le \mathbf{n}$ be the number of nearest neighbors of the considered point $(\mathbf{x}_c, \mathbf{y}_c)$ (parameter of the method). Let us denote by $\mathbf{B}_\mathbf{m}(\mathbf{x}_c, \mathbf{y}_c)$ the set of those $\mathbf{m}$ points of observations $(\mathbf{x}_\alpha, \mathbf{y}_\alpha)$ which are more close to the point $(\mathbf{x}_c, \mathbf{y}_c)$. The kernel estimate is defined by the formula:

$$\hat{\mathbf{Z}}(\mathbf{x_c}, \mathbf{y_c} \mid \vartheta) = \frac{\sum\limits_{\alpha \in \mathbf{B_m}(\mathbf{x_c}, \mathbf{y_c})} \mathbf{Z}_\alpha \cdot \Phi(\mathbf{x}_\alpha - \mathbf{x_c}, \mathbf{y}_\alpha - \mathbf{y_c} \mid \vartheta)}{\sum\limits_{\alpha \in \mathbf{B_m}(\mathbf{x_c}, \mathbf{y_c})} \Phi(\mathbf{x}_\alpha - \mathbf{x_c}, \mathbf{y}_\alpha - \mathbf{y_c} \mid \vartheta)} \tag{2}$$

Thus, we must previously define the vector of parameters ϑ. For this purpose we shall use the jack-knife method. Let us consider the kernel estimate (2) for some point $(\mathbf{x}_\beta, \mathbf{y}_\beta)$, which belongs to the set $\mathbf{B_m}(\mathbf{x_c}, \mathbf{y_c})$, using the information from all other points of this set:

$$\hat{\mathbf{Z}}^{(\mathbf{m})}_{(\mathbf{x_c}, \mathbf{y_c})}(\mathbf{x}_\beta, \mathbf{y}_\beta \mid \vartheta) = \frac{\sum\limits_{\alpha \in \mathbf{B_m}(\mathbf{x_c}, \mathbf{y_c}) \setminus \{\beta\}} \mathbf{Z}_\alpha \cdot \Phi(\mathbf{x}_\alpha - \mathbf{x}_\beta, \mathbf{y}_\alpha - \mathbf{y}_\beta \mid \vartheta)}{\sum\limits_{\alpha \in \mathbf{B_m}(\mathbf{x_c}, \mathbf{y_c}) \setminus \{\beta\}} \Phi(\mathbf{x}_\alpha - \mathbf{x}_\beta, \mathbf{y}_\alpha - \mathbf{y}_\beta \mid \vartheta)} \tag{3}$$

Applying the estimate (3) to all points of the set $\mathbf{B_m}(\mathbf{x_c}, \mathbf{y_c})$ we can calculate a jack-knife function of the vector ϑ and estimate its components from minimizing problem:

$$\sum_{\beta \in \mathbf{B_m}(\mathbf{x_c}, \mathbf{y_c})} \left| \hat{\mathbf{Z}}^{(\mathbf{m})}_{(\mathbf{x_c}, \mathbf{y_c})}(\mathbf{x}_\beta, \mathbf{y}_\beta \mid \vartheta) - \mathbf{Z}_\beta \right|^2 \to \min_{\vartheta} \tag{4}$$

Solution of the problem (4) provides the estimate:

$$\hat{\vartheta}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) = (\,\hat{\sigma}_1^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}), \hat{\sigma}_2^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}), \hat{\varphi}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c})\,) \tag{5}$$

After substituting the estimate (5) into formula (2) we will obtain the estimate $\hat{\mathbf{Z}}(\mathbf{x_c}, \mathbf{y_c} \mid \hat{\vartheta}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}))$ which we are seeking for.

An important by-product of this estimate is the value of the scale of kernel estimate which could be defined as the square root of the scattering ellipse square:

$$\hat{\mathbf{s}}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) = \sqrt{\pi \cdot \hat{\sigma}_1^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) \cdot \hat{\sigma}_2^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c})} \tag{6}$$

Another interesting statistics are the components $(\mathbf{V_x}, \mathbf{V_y})$ of the vector of anisotropy:

$$\begin{aligned} &\hat{\mathbf{V}}_{\mathbf{x}}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) = \hat{\varepsilon}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) \cdot \mathbf{cos}(\hat{\varphi}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c})) \\ &\hat{\mathbf{V}}_{\mathbf{y}}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) = \hat{\varepsilon}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) \cdot \mathbf{sin}(\hat{\varphi}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c})), \quad \hat{\mathbf{V}}_{\mathbf{y}}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) \geq \mathbf{0} \end{aligned} \tag{7}$$

where

$$\hat{\varepsilon}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) = \sqrt{\mathbf{1} - \frac{(\hat{\sigma}_2^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}))^2}{(\hat{\sigma}_1^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}))^2}}, \quad \mathbf{0} \leq \hat{\varepsilon}^{(\mathbf{m})}(\mathbf{x_c}, \mathbf{y_c}) \leq \mathbf{1} \tag{8}$$

is the eccentricity of the scattering ellipse.

REFERENCES.

Deutsch, C., A.G.Journel. (1997) GSLIB, Geostatistical Software Library and User's Guide. Oxford University Press.
Hardle W. (1989) Applied nonparametric regression. Cambridge University Press, Cambridge, New York, New Rochell, Melbourne, Sydney
Nadaraya E.A. (1964) On Estimating Regression. Theory Probab. Appl. Vol.9, pp.141-142
Watson G.S. (1964) Smooth Regression Analysis. Sankhya Ser.A Vol. 26, 359-372.

FIGURES.

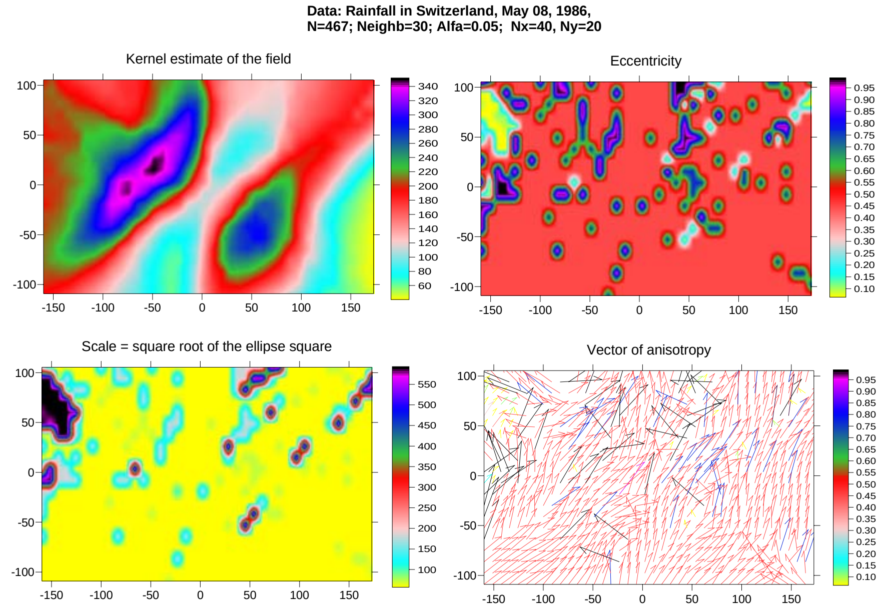


Fig.1

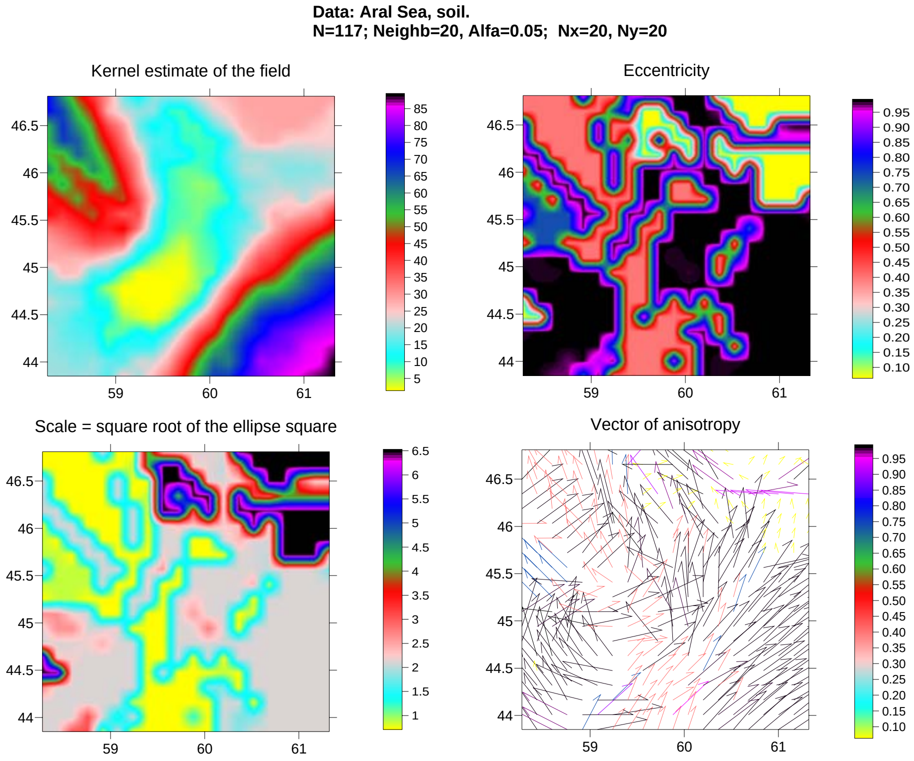
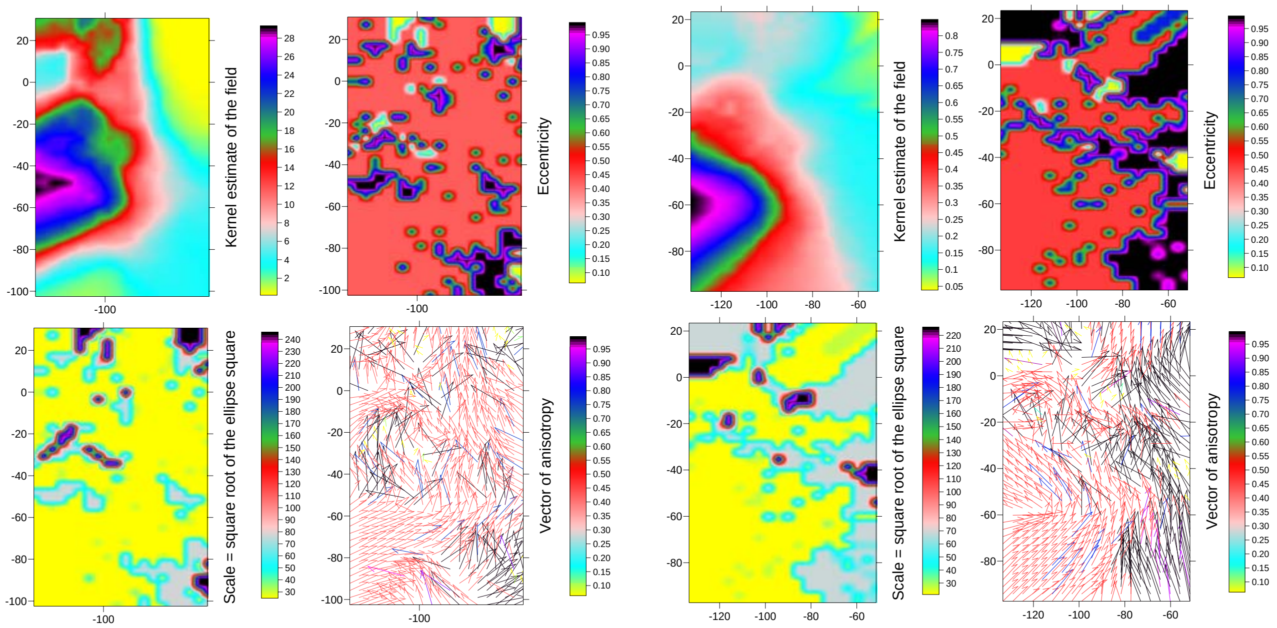


Fig.2



Data: Cs radionuclides pollutions (after Chernobyl) in Bryansk region of Russia; N=680; Neighb=30; Alfa=0.05; Nx=20, Ny=40

Data: Sr radionuclides pollutions (after Chernobyl) in Bryansk region of Russia; N=286; Neighb=30; Alfa=0.05; Nx=20, Ny=40

Fig.3.