A Vector Approach to the Analysis of (Patterns with) Spatial Dependence

Andrés Molina

Dpto. de Informática, Escuela Politécnica Superior,
Universidad de Jaén. Avda. de Madrid, 35,
23071 Jaén, Spain.
molina@ujaen.es

Abstract. . It is evident that the utility of an image or map will depend on the quantity of the information we can extract from it by the analysis of the spatial relationships of the phenomenon represented. For it, tools that describe aspects such as spatial dependence or autocorrelation in patterns are used. The statistic techniques that measure the spatial dependence are very varied, but all of them provide only scalar information about the variation of spatial properties in the pattern, without analyzing the possible directedness of the dependence mentioned. In this work, we make a vector approach to the analysis of spatial dependence, therefore, given a pattern, besides quantifying its autocorrelation level, we will determinate if statistics evidence of directedness exists, calculating the angle where the direction appears. For this we will use a parametric method when the normality of population can be assumed, and a non-parametric method for uniform distribution.

Keywords: Spatial Dependence, Anisotropy, Directional Trend, Circular Statistics.

1 Introduction and Previous Works

The interpretation of the spatial distribution of a phenomenon, can only be done by the evaluation of both the global (large scale trend or values in each point in the space) and local scale effects due to the interaction of each point with its neighboring points [1].

The absence of these local effects do that the values of the phenomenon vary depending on the place, in other words, the values observed in a window change systematically, hence, it does not exist spatial dependence among values, and the process is heterogeneous or non-stationary. On the contrary, if the existence of local effects is detected, the process is spatially homogeneous or stationary.

Spatial dependence is a particular case of homogeneity. In images whose elements show spatial correlation, it is verified that the existence of a concrete value of the phenomenon makes more credible this value to occur in near places. The existing statistics for determining the existence of spatial dependence among elements of an image are very varied [2][3], and they include non-spatial technics as ANOVA, error

terms in spatial autoregresive systems, Moran's I, Geary´s c, Getis´ G*, variograms, correlograms.... All of them determine the existence or absence of autocorrelation in an image, but they do not provide spatial information about the direction in which it is manifested, that is, the directional trend of the variation of parameters that define the thematic characteristics of the image.

We propose in this paper a method, based on first and second-order bivariate circular statistics, that will allow us to study the spatial variability of the phenomenon, establish the existence of spatial dependence in an image and calculate the direction in which this appears, using a parametric procedure based on the standard and confidence ellipses for normal samples, and a non-parametric test for directionality for samples in which the condition of normality cannot be assumed.

2 Difference Vectors Matrix

Let $W_{n,n}$ the matrix that keeps the values of a window obtained from an image. From $W_{n,n}$, the difference vectors matrix $V_{n,n}$ or matrix of first order vectors is obtained. For that, every element $w_{i,j}$ is compared with the element which is diametrically opposed to it; if $w_{i,j}$ is greater tham $w_{2k+2-i,2k+2-k}$ we will assign the difference between these two values ($w_{i,j}$- $w_{2k+2-i,2k+2-k}$) to $v_{i,j}$ and zero value to $v_{2k+2-i,2k+2-k}$. If $w_{i,j}$ is not greater than $w_{2k+2-i,2k-k}$, the assignment will be inversely.

$$W_{n,n} = \begin{bmatrix} w_{1,1} & .. & w_{1,c} & .. & w_{1,n} \\ : & & : & & : \\ w_{c,1} & .. & w_{c,c} & .. & w_{c,n} \\ : & & : & & : \\ w_{n,1} & .. & w_{n,c} & .. & w_{n,n} \end{bmatrix} = \begin{bmatrix} w_{c-d,c-d} & .. & w_{c-d,c} & .. & w_{c-d,c+d} \\ : & & : & & : \\ w_{c,c-d} & .. & w_{c,c} & .. & w_{c,c+d} \\ : & & : & & : \\ w_{c+d,c-d} & .. & w_{c+d,c} & .. & w_{c+d,c+d} \end{bmatrix}$$

The element $v_{i,j}$ of the matrix $V_{n,n}$ represents a vector with length $r = v_{i,j}$ and angle $\Phi = \text{ArcTan}\left[\frac{j-(k+1)}{-(i-(k+1))}\right]$, that can be interpreted as a variation of the characteristic with intensity $v_{i,j}$ in direction Φ . From the $V_{n,n}$, and taking values $d = 1, 2, 3, ..., (n-1)/2$, we obtain $V_{3,3}(1)$, $V_{5,5}(2)$, ... $V_{n,n}((n-1)/2)$, submatrix of $V_{n,n}$ whose central element (element that has d rows both up and down and d columns both right and left) is $v_{(2d+1)/2,(2d+1)/2}$. Each one of the matrix $V_{2d+1,2d+1}(d)$ will be formed by the elements $v_{i,j}$ whose distance to the central element is less or equal to d.

Given V(d), its mean variation vector $\overline{m}_d$ (of length r_d and mean angle $\overline{\Phi}_d$) is calculated as follows [4]:

$$N = \sum_{i,j}^{2d+1} b_{i,j}(d)v_{i,j} \tag{1}$$

$$\overline{x} = \frac{1}{N} \sum_{i,j}^{2d+1} b_{i,j}v_{i,j} \text{Cos}\left[\text{ArcTan}\left(\frac{j-(d+1)}{-(i-(d+1))}\right)\right] \qquad \overline{y} = \frac{1}{N} \sum_{i,j}^{2d+1} b_{i,j}v_{i,j} \text{Sin}\left[\text{ArcTan}\left(\frac{j-(d+1)}{-(i-(d+1))}\right)\right] \tag{2}$$

$$r = \sqrt{\overline{x}^2 + \overline{y}^2} \qquad \overline{\Phi}_d = \begin{cases} \text{ArcTan}\dfrac{\overline{y}}{\overline{x}} & \text{if } \overline{x} > 0 \\ 180 + \text{ArcTan}\dfrac{\overline{y}}{\overline{x}} & \text{if } \overline{x} < 0 \end{cases} \qquad (3)$$

$$b_{i,j}(d) = \begin{cases} 1 & \text{if } |i\text{-}d| = 0 \ \text{ or } \ |j\text{-}d| = 0 \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

If the mean variation vectors $\overline{m}_d$ [d = 1..8] are calculated from the image in Fig. 1 and 2, we will obtain respectively the values shown Tables 1 and 2.

Table 1. Mean variation vectors calculated from Fig. 1

d	1	2	3	4	5	6	7	8
r_d	0.79	0.78	0.77	0.66	0.61	0.66	0.63	0.44
$\overline{\Phi}_d$	292	290	281	270	253	237	225	222

Table 2. Mean variation vectors calculated from Fig. 2

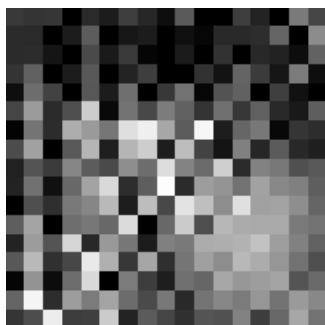| d	1	2	3	4	5	6	7	8
r_d	0.35	0.23	0.06	0.10	0.03	0.07	0.06	0.26
$\overline{\Phi}_d$	143	28	93	231	158	254	19	216

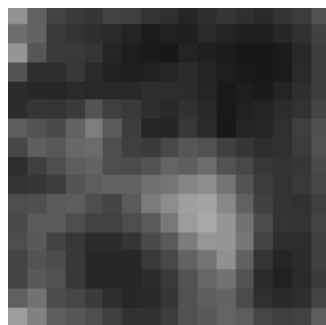**Fig. 1.** Pattern with Moran´s I = 0.34   **Fig. 2.** Pattern with Moran´s I = 0.82

If we calculate the first-lag autocorrelation of Fig. 1 and 2 using Moran's Index [5] examining the eight neighboring cells connected to each cell, we obtain the values shown in Table 3. We can observe that Fig. 1 presents a certain level of heterogeneity in the spatial distribution of the values. This fact is confirmed with a low index of autocorrelation (0.34). Fig. 2, on the contrary, shows an index near to one (0.82), what indicates a high spatial dependence.

Table 3. Autocorrelation parameters of Fig. 1 and 2

	Fig. 1	Fig. 2
Number of cells included	289	289
Mean of cells included	91.82	67.02
Standard Deviation (σ) of cell values	3.74	1.80
Moran´s I	0.34	0.82

3 Bivariate and Second-Order Analysis

It is evident that, from observations made for a concrete spatial lag, we cannot draw a conclusion about the behavior of values in the image for other lags no matter how many data are available and how sophisticated the analysis is. For this reason the statistical analysis must be performed in two steps called first and second-order analysis or first and second stage of analysis [6]:

- For each k-lag [$k = 1..d$] we reduce the variation vectors by calculating the mean variation vector $\overline{m}_k$.
- We combine the mean variation vectors $\overline{m}_1, \overline{m}_2, ..., \overline{m}_d$ of above step and test their significance. Only then can we make statistical inference about the directional behavior of data.

The vector m_d is described by an angle Φ_d and a module r_d; in other words, it has to be considered both the angles and the amplitude of his module. Under this condition, the pairs $(r_1, \Phi_1), (r_2, \Phi_2), ..., (r_d, \Phi_d)$ become bivariate and second-order sample, and his treatment is considered second-order analysis.

3.1 Standard Ellipse

Among the tools used for second-order analysis we find the standard ellipse [7]; it serves exclusively for descriptive purposes. The tips of the vectors of a second-order sample form a scatter diagram of data with standard deviations in the x and y directions and a certain trend upward or downward. The standard ellipse describes this behavior in a condensed form: assuming normality, roughly 40% of the data points fall inside the ellipse and 60% outside. The parent population need not to be normal, although it is desirable that it does not deviate too much from normality. For drawing the standard ellipse two means, two standard deviations and a correlation coefficient are required:

$$x_i = r_i \, \text{Cos} \, \Phi_i \qquad\qquad y_i = r_i \, \text{Sin} \, \Phi_i \qquad\qquad (5)$$

$$\overline{x} = \tfrac{1}{n}\sum x_i \qquad\qquad\qquad \overline{y} = \tfrac{1}{n}\sum y_i \qquad\qquad (6)$$

$$s_1^2 = \tfrac{1}{n-1}\sum(x - \bar{x}) \qquad\qquad s_2^2 = \tfrac{1}{n-1}\sum(y - \bar{y})$$ (7)

$$\text{Cov}(x, y) = \tfrac{1}{n-1}\sum(x - \bar{x})(y - \bar{y}) \qquad r = \text{Corr}(x, y) = \tfrac{\text{Cov}(x,y)}{s_1\,s_2}$$ (8)

The equation of the standard ellipse is

$$A(x - \bar{x})^2 + 2B((x - \bar{x})(y - \bar{y})) + C(x - \bar{x})^2 = D$$ (9)

where

$$A = s_2^2 \qquad B = -rs_1 s_2 \qquad C = -s_2^2 \qquad D = (1 - r^2)s_1^2 s_2^2$$ (10)

The midpoint of ellipse is at $(\bar{x}, \bar{y})$; the semi-axes a and b $(a < b)$ are

$$a = \left[\tfrac{2D}{A+C-R}\right]^{1/2} \qquad\qquad b = \left[\tfrac{2D}{A+C+R}\right]^{1/2}$$ (11)

where

$$R = \left[(A - C)^2 + 4B^2\right]^{1/2}$$ (12)

The sample shows the maximum variability in a direction of angle θ. This is the angle by which the major axis is inclined versus the X axis (-90° < θ < 90°).

$$\theta = \text{ArcTan}\left[\tfrac{2B}{A-C-R}\right]$$ (13)

The values of the parameters described, calculated from Table 1 and Table 2, are the ones shown respectively in Tables 4 and 5. Fig. 3 and 4 show the representation of the ellipses.

Table 4. Values of the standard ellipse calculated from vectors of Table 1

$\bar{x}$	$\bar{y}$	s_1	s_2	Cov	r	A	B	C	D	R	a	b	θ
0.073	−0.368	0.189	0.370	−0.004	−0.062	0.137	0.004	0.036	0.005	0.101	0.370	0.189	272

Table 5. Values of the standard ellipse calculated from vectors of Table 2

$\bar{x}$	$\bar{y}$	s_1	s_2	Cov	r	A	B	C	D	R	a	b	θ
−0.076	−0.596	0.295	0.164	−0.042	−0.860	0.027	0.042	0.087	0.0006	0.103	0.329	0.075	332

As we have commented in this paragraph, standard ellipse serves exclusively for descriptive purposes. Nevertheless, from it, we can estimate whether is reasonable the assumption of normality in the parent population. If the mean vector angles are uniformly spaced around the coordinates origin, we can consider that the population from which the sample is drawn do not differ from randomness or one-sideness, avoiding normality. When this occurs, the origin falls into the ellipse. For this reason we can consider, in a approximate way, that the parent population is normal if the standard ellipse contains the coordinates origin.
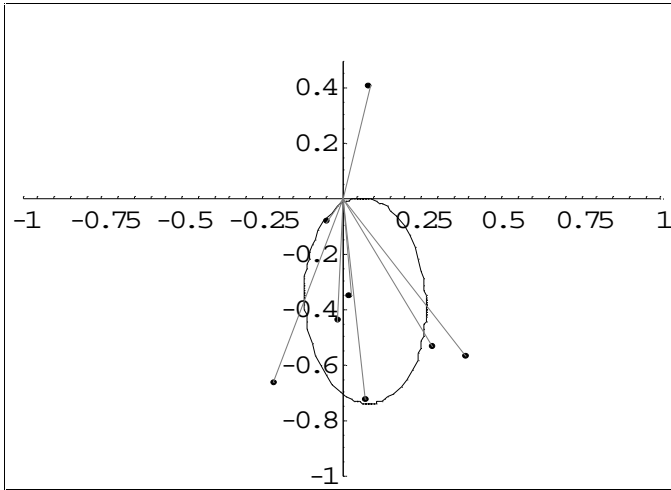
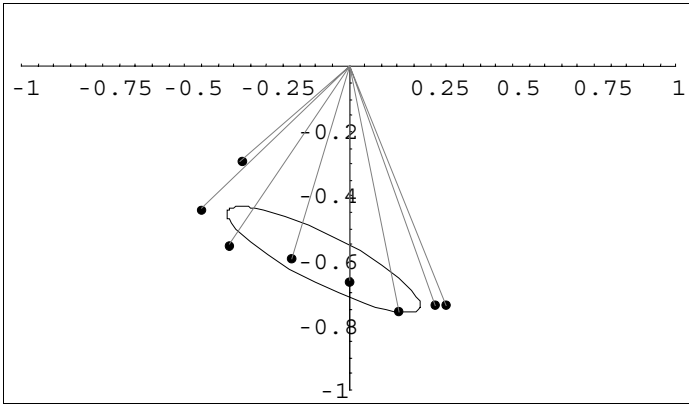**Fig. 3.** Standard ellipse depicted from parameters contained in Table 4

Fig. 4. Standard ellipse depicted from parameters contained in Table 5

3.2 Confidence Ellipse

By the standard ellipse, we have described the spatial behaviour of the mean variation vectors quantifying, by the calculation of θ, its directional trend. Nevertheless, we must determine if the directional trend is caused by random fluctuations of the vectors or whether, on the contrary, this is caused by the existence of directedness or anisotropic variability [8]. For testing directedness, the confidence ellipse is used.

Confidence ellipse includes a region in the xy-plane that covers the unknown population centre with a preassigned probability $Q = 1 - \alpha$ being therefore, a tool for

statistical inference. Assuming normality, this region has the shape of an ellipse which has the same centre as the standard ellipse and the same inclination of the major axis (θ). The problem of determining the existence of directedness is solve by generating the confidence ellipse and testing if the origin in its interior. If this fact does not occur, the population centre cannot coincide with the origin, and $(\bar{x}, \bar{y})$ is significantly different from it, concluding that the mean vectors as a group are oriented in the direction ψ

$$\psi = \text{ArcTan}\left[\frac{\bar{y}}{\bar{x}}\right] \tag{14}$$

For the confidence ellipse, the coefficients A, B and C in Eq. (9) have the same values for standard ellipse in Eqs. (10). Being the coefficient D

$$D = (1 - r^2)s_1^2 s_2^2 n^{-1} T^2 \tag{15}$$

where

$$T^2 = 2\frac{n-1}{n-2} F_{2, n-2} \tag{16}$$

Here, $F_{2, n-2}(\alpha)$ denotes the critical F value with 2 and n-2 degree of freedom and significance level α. Confidence ellipse has the same center and the same θ value as the standard, since it is independent on the variable D, as Eqs. (6) (13) reveal. In both ellipses, principal axes coincide. Only the semi-axes are variable. Let be the semi-axes with the special parameter $D = 1$. Then we obtain from Eqs. (11) for arbitrary values $D <> 0$

$$a = a_1 D^{\frac{1}{2}} \qquad\qquad b = b_1 D^{\frac{1}{2}} \tag{17}$$

Since a and b are proportional to $D^{\frac{1}{2}}$, we obtain

$$a = a_s n^{\frac{1}{2}} T \qquad\qquad b = b_s n^{\frac{1}{2}} T \tag{18}$$

where a_s, b_s are the semi-axes of the standard ellipse. From Eq. (9) is deduced that points which are inside the region limited by the confidence ellipse fulfil the inequality

$$A(x - \bar{x})^2 + 2B((x - \bar{x})(y - \bar{y})) + C(y - \bar{y})^2 < D \tag{19}$$

If the origin falls within the ellipse, the population centre could coincide with the origin and the sample is not directed. In this case, inequality (18) is fulfilled with the special values $x = 0$ and $y = 0$. From Eqs. (14)(18), the condition for the existence of directedness with a level of significance α is

$$T^2 < T^2(\alpha) \tag{20}$$

being

$$T^2(\alpha) = \frac{n}{1 - r^2}\left[\frac{\bar{x}^2}{s_1^2} - \frac{2r\bar{x}\bar{y}}{s_1 s_2} + \frac{\bar{y}^2}{s_2^2}\right] \tag{21}$$

Analysing the standard ellipse (Fig. 4) obtained from Fig. 2, we observe that the assumption of normality is reasonable, so that a confidence ellipse is applicable to it. We want to determine the confidence ellipse at a level of significance $\alpha = 5\%$ or a confidence coefficient of $Q = 95\%$ from Fig. 2. As a first step we calculate T^2 using Eq. (15). From a table of the F-distribution we read $F_{2,6} = 5.14$. Hence $T^2 = 11.99$. From Eqs. (14)(17) we obtain D, a and b parameters showed in Table 6. Our confidence ellipse (dashed curve depicted in fig. 5) is somewhat biggest than the standard ellipse. Since the confidence ellipse does not contain the origin in its interior, $(\overline{x}, \overline{y})$ differs significantly from the origin. Therefore, the mean vectors are oriented as a group, being the direction at $\psi = -97.25°$.

Table 6. Parameters of confidence ellipse calculated from vectors of Table 2

	Fig. 1 $\alpha = 0.5$	Fig. 1 $\alpha = 0.1$	Fig. 2 $\alpha = 0.5$
T^2	8.07	11.99	11.99
a	0.37	0.45	0.40
b	0.19	0.23	0.09

In Fig. 3 we observe that, although the standard ellipse does not contain the origin, this is very close to it. Consequently, the assumption of normality must be done with certain reservations. Fig. 6 shows the confidence ellipses with errors $\alpha = 0.05$ and $\alpha = 0.1$ drawn from the parameters of Tables 5 and 6. The inside ellipse does not contain the coordinates origin, while the exterior one does, assuming the existence of directedness ($\psi = -77.76°$) with $Q = 90\%$, but not with $Q = 95\%$.

3.3 Not Normal Bivariate Population

The procedure described in the paragraph bellow requires that the second order scatter diagram be a sample drawn from a normal bivariate population. Nevertheless, there are occasions in which, seeing the circular histogram of the sample, this condition of normality cannot be assumed. When this happens, it is possible by the Moore´s test [9], to know if the sample is directed or on the contrary, whether it presents a uniform distribution. This test is non parametric because only the ranks of r_i are used.

In this procedure, we must rank the r_i from smallest to largest, letting t_i denote the rank of the ith mean variation vector. The null hypothesis for this test states the $\overline{\Phi}_i$ are independently and uniformly distributed on the circle. Let

$$C = \sum t_i \, \mathrm{Cos}\,\overline{\Phi}_i \qquad S = \sum t_i \, \mathrm{Sin}\,\overline{\Phi}_i \qquad D = \sqrt{(C^2 + D^2)} \qquad D^* = D\Big/ n^{\frac{3}{2}} \qquad (22)$$

We reject the null hypothesis in favour of the hypothesis of directionality if

$$D^* \geq D^*(\alpha) \qquad (23)$$

As we have commented in the paragraph below, the proximity of standard ellipse to origin in Fig. 3, make us doubt about the suitability of considering the sample

drawn from a normal bivariate population. Table 7 shows the critical values of Moore´s Test and confidence ellipses for different errors α, being their statistic tests respectively $D^* = 1.249$ and $T^2 = 8.749$. With these values, according to the Moore´s Test, a directedness with an error $\alpha \geq 0.010$ exists. If, on the contrary, we use the confidence ellipse, we found directedness with $\alpha \geq 0.100$. The difference between both tests is due to the use of confidence ellipse supposes the assumption of normality, and this is a strong constraint that we cannot clearly establish visualising the standard ellipse.

Table 7. Critical values of D^* and T^2

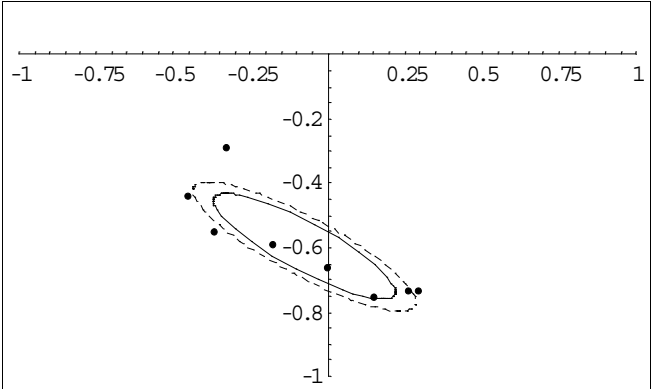| α	0.001	0.005	0.010	0.025	0.050	0.100
$D^*(\alpha)$	1.397	1.300	1.242	1.148	1.059	0.949
$T^2(\alpha)$	63.000	33.927	25.480	16.940	11.990	8.074

Fig. 5. Confidence (dashed curve) and standard ellipses depicted from values of Tables 4 and 5

4 Conclusion

Previous works about the evaluation of spatial dependence in patterns, based on statistics of spatial autocorrelation, provide an acknowledge exclusively quantitative about the topological structure and spatial relationships in the distribution of a spatial phenomenon, but they do not analyze neither the existence of spatial anisotropy nor the direction in which it manifest. In this paper we have described a method for:

- Visualizing the spatial variability of the pattern.
- Determining the normality of parent population.
- Testing the existence of spatial trend or directedness in patters that show spatial dependence.
- Calculating the direction ψ where that anisotropy is revealed .
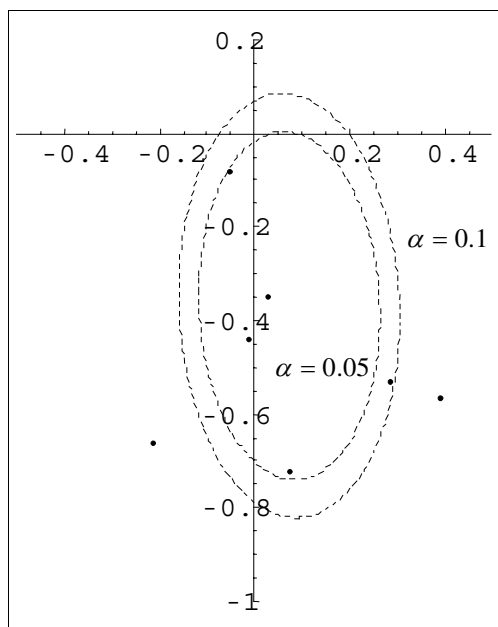
Fig. 6. Confidence ellipses with errors $\alpha = 0.05$ and $\alpha = 0.1$

For that, we have used a parametric method (confidence and standard ellipses) when the normality of population can be assumed, and a non-parametric method (Moore´s Test) for uniform distribution. Both are based in first and second-order bivariate circular statistics.

References

[1] Getis, A.: Homogeneity and Proximal Databases. In: Fotheringham, S., Rogerson, P. (eds.): Spatial Analysis and GIS. Taylor&Francis, London (1994) 105-120
[2] Bailey, T.C., Gattrel, A.C.: Ineractive Spatial Data Analysis. Longman, Edinburgh (1995)
[3] Wackernagel, H.: Multivariate geostatistics. Springer, Berlin Heidelberg New York (1995)
[4] Batschelet, E.: Recent statistical methods for orientation data. U.S. Government Printing Office, Washington (1965)
[5] Chow, Y,H.: Spatial Patern and Spatial Autocorrelation In: Frank, A.U., Kuhn, W. (eds.): Spatial Information Theory. Lectures Notes in Computer Science, Vol.988. Springer, Berlin Heidelberg New York (1995) 365-376
[6] Cox, D.R., Hinkley, D.V.: Theoretical Statistics. Chapman and Hall, London (1974)
[7] Batschelet, E.: Circular statistics in biology. Academic Press, London (1981)
[8] Molina, A., Feito, F.R.: Design of Anisotropic Functions Within Analysis of Data. In: Wolter, F.E., Patrikalakis, N.M. (eds.): Proc. IEEE Computer Graphics International (1998) 726-729
[9] Moore, F.R.: A modification of the Rayleigh test for vector data. Biometrika 67 (1980) 175-180