# Extension of mathematical background for Nearest Neighbour Analysis in three-dimensional space

**Eva Stopková**

PhD. student of Geodesy and Cartography
Department of Theoretical Geodesy
Faculty of Civil Engineering
Slovak University of Technology in Bratislava

## Abstract

*Proceeding deals with development and testing of the module for GRASS GIS [1], based on Nearest Neighbour Analysis. This method can be useful for assessing whether points located in area of interest are distributed randomly, in clusters or separately. The main principle of the method consists of comparing observed average distance between the nearest neighbours $\bar{r}_A$ to average distance between the nearest neighbours $\bar{r}_E$ that is expected in case of randomly distributed points. The result should be statistically tested. The method for two- or three-dimensional space differs in way how to compute $\bar{r}_E$. Proceeding also describes extension of mathematical background deriving standard deviation of $\bar{r}_E$, needed in statistical test of analysis result. As disposition of phenomena (e.g. distribution of birds' nests or plant species) and test results suggest, anisotropic function would represent relationships between points in three-dimensional space better than isotropic function that was used in this work.*

**Keywords:** 3D GIS, spatial analysis, Nearest Neighbour Analysis

## 1. Introduction

The purpose of this work is to outline the way how to implement Nearest Neighbour Analysis (NNA) in three-dimensional space into the geographical information systems (GIS) environment. At first, the article summarizes derivation of mathematical background in case of isotropic phenomenon. In the next part, there is described the module of open source software *GRASS GIS* [1] that was developed on the base of these relationships. Finally, the results of moduleś tests are analysed.

NNA helps to assess whether points located in tested area are distributed randomly, in clusters or separately. It can be useful in biology to monitor behaviour of plant or animal populations [2], in stellar statistics, in chemistry to analyse atomic structures, etc. In GIS, NNA may be helpful in answering questions in biology (mentioned above) or in solving social problems (e.g. crime analysis). In case of analysing vertically divided phenomena (as artifacts in archaeological trench, birds' nests, behaviour of plant or animal populations in 3D space), three-dimensional distance should be considered in NNA. This distance depends also on the difference of elevation between objects.

If we assume the phenomenon to be isotropic, it is possible to express NNA as function of distance $r$. It is more probable that natural phenomena are anisotropic - they behave

differently in horizontal and vertical direction. NNA then could be a function of horizontal distance and vertical difference of elevation (or zenith angle). In another case, the phenomenon may be anisotropic in three dimensions.

## 2. Keynote

NNA in three-dimensional space was outlined in proceedings [3] and [4]. The main principle of this work is based also on [2] that describes NNA for two-dimensional space. The goal of this article on theoretical level is to complete the idea of statistical testing in three-dimensional space.

There are $N$ points located in the area of interest. Their average isotropic distance between the nearest neighbours $\bar{r}_A$ is expressed using arithmetic mean. If these points are randomly distributed, average isotropic distance of the nearest neighbours in three-dimensional space is expected to be (according to [4], formulas (674), (675)):

$$\bar{r}_E = \int\limits_0^\infty 4\pi n \cdot r^3 \cdot e^{-\frac{4\pi n}{3} \cdot r^3} dr = \frac{1}{\sqrt[3]{\frac{4\pi n}{3}}} \int\limits_0^\infty e^{-x} \cdot \sqrt[3]{x} dx = \frac{1}{\sqrt[3]{\frac{4\pi n}{3}}} \cdot \Gamma(\frac{4}{3})$$

where $n$ is density of points in volume unit and $\Gamma(\frac{4}{3})$ is the gamma function[1]. This formula is based on the probability that the nearest neighbour of a point is located on the boundary of its spherical surrounding with radius $r$. More detailed explanation of the topic can be found in [4] or in [3].

The measure of degree to which observed points are randomly distributed $R$ is expressed as the ratio of observed and expected distance [2]. It may acquire these values:

$R = 1 \to$ points are distributed randomly, because $\bar{r}_A = \bar{r}_E$

$R = 0 \to$ points are identical, because $\bar{r}_A = 0$

$R = max \to$ it is necessary to derive maximal value of ratio $R$. According to [2], if points are located on a hexagonal pattern, $R$ in two-dimensional space reaches maximal value

$$R_{2Dmax} = \frac{2.1491}{\sqrt{k}}$$

where $k$ means the number of segments in circle of infinite radius with a center in the observed point. The results of module testing (Chap. 6), as well as the results obtained using analytical tool *Average Nearest Neighbor* [6], do not correspond with this value. Determining of $R$ maximum will be purpose of further work (Chap. 7).

## 3. Test of statistical significance

Derivation of standard deviation of the average distance expected in case of randomly distributed points $\sigma_{\bar{r}_E}$ in three-dimensional space belonged in the purposes of this article. $\sigma_{r_E}$ is needed to enumerate test statistic of Student's t-test of significance of the mean [7] that helps to assess the statistical significance of the deviation from a random distribution of points (clustering or, on the other hand, separation). Method of derivation (Appendix A) is analogical to method for two-dimensional space [2].

---

[1] *Gamma function* $\Gamma(n)$ means extension of factorial for complex and real numbers [5].

Null hypothesis $H_0$: $\overline{r}_A = \overline{r}_E$ means that points are randomly distributed in the space (according to [8]). Alternative hypothesis $H_a$: $\overline{r}_A < \overline{r}_E \vee \overline{r}_A > \overline{r}_E$ and statistical tables (e.g. [9]) of the cumulative standard normal distribution are base for determination of critical values $W$:

$$W = (-\infty, u_{\alpha/2}) \cup (u_{(1-\alpha/2)}, \infty)$$

If null hypothesis about randomness of distribution in point dataset is rejected with positive values of the test statistic $c$, the points are assumed to be separated (patterned). If the values of the test statistic $c$ are negative and the null hypothesis is rejected, the points are assumed to be clustered.

## 4. Implementation of 3D Nearest Neighbour Analysis

The theoretical background described in the previous chapters was implemented in the module *v.nn_spatial_stat* developed in the environment of open-source software *GRASS GIS* [1]. The module contains functionality of NNA in two-dimensional space (implemented also in the analytical tool [6] of *ArcGIS* [10]) and solution for objects in three-dimensional space. The most significant complications during development were connected with determining of Minimum Bounding Box (MBB)[2].

In two-dimensional space, the density of points depends on the area of surface where the points are located. According to [11], the area could be set up by user or it is possible to use area of Minimum Bounding Rectangle (MBR). Analogically, the density of points in three-dimensional space could be determined using volume of box set up by user or volume of MBB.

Methods how area (volume) of MBR (MBB) can be determined are principally quite similar:

- Coordinates of convex hull[3] that covers the point set must be obtain. In 3D case it is necessary to know also reference of each vertex to the faces. Partially modified functions of the module *v.hull* [13] that enables to build convex hull in new vector layer were used. Output of modified functions is represented by the list of coordinates of vertices (or faces), not by new vector body.

- The coordinates of vertices must be <u>transformed</u> to coordinate systems

    - which $x$ axes are parallel to lines between neighbouring vertices in 2D case

    $$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = R_x(\sigma) \cdot \begin{pmatrix} x \\ y \end{pmatrix}$$

    where $\sigma$ is bearing of the line,

    - which $xy$ planes are parallel to planes of faces of the convex hull in 3D case

    $$\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = R_{xy}(\sigma_x, \sigma_y) \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

---

[2]Minimum Bounding Box encloses all features in vector dataset. Its faces are paralel to planes of 3D coordinate system in special case only.

[3]Convex hull is a cycle graph (body) that encloses all features in dataset. Its vertices are located on the features in such way that all interior angles are less than 180° (according to [12]).

where the angles $\sigma_x$, $\sigma_y$ may be expressed:

$$tan(\sigma_x) = \frac{z_1 - z_0}{y_1 - y_0}$$

$$tan(\sigma_y) = \frac{z_2 - z_0}{x_2 - x_0}$$

and $x$, $y$, $z$ are coordinates of vertices belonging to triangular face of convex hull. Transformation matrix is based on rotation of $x$ and $y$ axes:

$$R_{xy}(\sigma_x, \sigma_y) = R_x(\sigma_x) \cdot R_y(\sigma_y) =$$

$$= \begin{pmatrix} cos(\sigma_y) & 0 & sin(\sigma_y) \\ sin(\sigma_x) \cdot sin(\sigma_y) & cos(\sigma_x) & sin(\sigma_x) \cdot cos(\sigma_y) \\ cos(\sigma_x) \cdot sin(\sigma_y) & -sin(\sigma_x) & cos(\sigma_x) \cdot cos(\sigma_y) \end{pmatrix}$$

– <u>Extent of transformed coordinates</u> should be determined,

– <u>Area / volume of the extent</u> must be counted and the values should be compared to obtain the smallest one. This value becomes input in determining the density of points.

## 5. Testing of the module *v.nn_spatial_stat*

For points located in two-dimensional space, the module was tested comparing numerical results with outputs of the analytical tool *Average Nearest Neighbor* [6] that is part of the *Spatial Statistics* toolbox in the software *ArcGIS 10.1* [10]. 3D variant of NNA is not implemented in any of accessible softwares. That was the reason to verify numerical results by comparing them with values computed by scripts in the software *Mathematica* [14] and *Matlab* [15].

### 5.1. Testing in two-dimensional space

The module was tested on various sets of synthetic data. Configuration of observed points in the area of interest 20 km x 20 km was designed to represent all possible cases: randomly distributed points, separated points and clustered points.

The samples of randomly distributed points were generated in the software *GRASS GIS* [1] using the module *v.random* [16]. Except numerical accuracy, process speed was tested too. The results are summarized in tables 1a and 1b.

| | n = 2000 | | | |
|---|---|---|---|---|
| | *v.nn_spatial_stat* | *Average Nearest Neighbor* [6] | difference | |
| $\overline{r}_A[m]$ | 225.858627 | 225.858627 | 0.000 | mm |
| $\overline{r}_E[m]$ | 223.227944 | 223.227944 | 0.000 | mm |
| $R$ | 1.011785 | 1.011785 | 0.000000 | |
| $c$ | 1.008239 | 1.008245 | 0.000006 | |
| $A[m^2]$ | 398645718.650660 | 398645718.650315 | -345 | mm$^2$ |
| $t[s]$ | 0.107 | 3 | | |

Table 1a: The results of testing of the module *v.nn_spatial_stat* in two-dimensional space using 2000 randomly distributed points

| | n = 5000 | | | |
|---|---|---|---|---|
| | *v.nn_spatial_stat* | *Average Nearest Neighbor* [6] | difference | |
| $\overline{r}_A[m]$ | 140.981861 | 140.981861 | 0.000 | mm |
| $\overline{r}_E[m]$ | 141.307761 | 141.307761 | 0.000 | mm |
| $R$ | 0.997694 | 0.997694 | 0.000000 | |
| $c$ | -0.311984 | -0.311986 | -0.000002 | |
| $A[m^2]$ | 399357668.904297 | 399357668.904439 | 142 | $mm^2$ |
| $t[s]$ | 0.324 | 5 | | |

Table 1b: The results of testing of the module *v.nn_spatial_stat* in two-dimensional space using 5000 randomly distributed points

According to the test statistic $c \in (-1.96; 1.96)$, it can be assumed that in both cases the points were randomly distributed.

Differences in the values of MBR area may be caused by the different way of storing data in computer memory, as the results of next experiments show. Significantly shorter processing time (compared with the analytical tool *Average Nearest Neighbor* [6]) may be reached, because the module does not generate report with graphical outputs.

The condition of maximized separation in two-dimensional space is accomplished by points arranged in the pattern of equilateral triangles, i.e. the nearest neighbouring points are located around the observed point in the shape of regular hexagon (proofed in [2]). Two datasets were generated, seven points arranged in the hexagon with centre and many points arranged in the hexagonal pattern. Coordinates of the points were computed in the Matlab [15] environment. Tables 2a and 2b summarize the results of the tests.

| | r = 250 m | | | |
|---|---|---|---|---|
| | n = 7 | | | |
| | *v.nn_spatial_stat* | *Average Nearest Neighbor* [6] | difference | |
| $\overline{r}_A[m]$ | 250.000000 | 250.000000 | 0.000 | mm |
| $\overline{r}_E[m]$ | 87.933894 | 87.933894 | 0.000 | mm |
| $R$ | 2.843045 | 2.843045 | 0.000000 | |
| $c$ | 9.328528 | 9.328585 | 0.000057 | |
| $A[m^2]$ | 216506.350945 | 216506.350945 | 0.000 | $mm^2$ |

Table 2a: The results of testing of the module *v.nn_spatial_stat* in two-dimensional space using maximally separated points (hexagon with centre)

| | r = 250 m | | | |
|---|---|---|---|---|
| | n = 4732 | | | |
| | *v.nn_spatial_stat* | *Average Nearest Neighbor* [6] | difference | |
| $\overline{r}_A[m]$ | 250.000000 | 250.000000 | 0.000 | mm |
| $\overline{r}_E[m]$ | 140.773069 | 140.773069 | 0.000 | mm |
| $R$ | 1.775908 | 1.775908 | 0.000000 | |
| $c$ | 102.108229 | 102.108855 | 0.000626 | |
| $A[m^2]$ | 375097253.014138 | 375097253.014137 | -0.954 | $mm^2$ |

Table 2b: The results of testing of the module *v.nn_spatial_stat* in two-dimensional space using maximally separated points (hexagonal pattern)

Because of the test statistic $c \in < 2.58; \infty)$, null hypothesis about random distribution of the points is rejected on the confidence level $\alpha = 0.01$. The points are separated.

Clusters of points were created locating points around each of sample of $n_0$ randomly generated points. Coordinates of new points were computed in *Matlab* [15] using bearings with step 36° and random distances (table 3a, table 3b).

| | $n_0 = 100$ | | |
| --- | --- | --- | --- |
| | $\sigma = [0°; 36°; \ldots 324°]$ | | |
| | $r = random('Normal', 30, 10)$ | | |
| | *v.nn_spatial_stat* | *Average Nearest Neighbor* [6] | difference |
| $\overline{r}_A[m]$ | 20.121996 | 20.121996 | 0.000 mm |
| $\overline{r}_E[m]$ | 311.387518 | 311.387518 | 0.000 mm |
| $R$ | 0.06462 | 0.06462 | 0.000000 |
| $c$ | -56.586926 | -56.587273 | -0.000347 |
| $A[m^2]$ | 387848745.955230 | 387848745.955141 | -89 $mm^2$ |

Table 3a: The results of testing of the module *v.nn_spatial_stat* in two-dimensional space using clusters of points with shorter distances from local centres

| | $n_0 = 100$ | | |
| --- | --- | --- | --- |
| | $\sigma = [0°; 36°; \ldots 324°]$ | | |
| | $r = random('Normal', 300, 1000)$ | | |
| | *v.nn_spatial_stat* | *Average Nearest Neighbor* [6] | difference |
| $\overline{r}_A[m]$ | 301.259436 | 301.259436 | 0.000 mm |
| $\overline{r}_E[m]$ | 351.460991 | 351.460991 | 0.000 mm |
| $R$ | 0.857163 | 0.857163 | 0.000000 |
| $c$ | -8.641085 | -8.641138 | -0.000053 |
| $A[m^2]$ | 494099313.011801 | 494099313.011630 | -171 $mm^2$ |

Table 3b: The results of testing of the module *v.nn_spatial_stat* in two-dimensional space using clusters of points with longer distances from local centres

Null hypothesis about randomly distributed points may be rejected on confidence interval $\alpha = 0.01$ because $c \in (\infty; -2.58 >$. Negative values of the test statistic $c$ indicate that the samples are clustered. The sample in table 3a in which clusters were generated using random distances with normal distribution $N(30, 10)$ is characterized by significantly lower value of test statistic $c$ as the sample in table 3b in which distances with normal distribution $N(300, 1000)$ were used.

## 5.2. Testing in three-dimensional space

Nowadays, there is no accessible tool for process NNA in three-dimensional space, so it is not possible to compare the results of the module with outputs of any verified software. That is the reason to control the results using parts of code scripted in mathematical software *Mathematica* [14] or *Matlab* [15]. This method helped to verify numerical accuracy and to repair few bugs.

The most complicated task was verification of volume of Minimum Bounding Box (MBB). This value is based on coordinates of vertices belonging to faces of convex hull. Coordinates of

convex hull are transformed to coordinate systems with plane of $x$ and $y$ axes parallel to plane of each face. Output is volume of the smallest box extending the transformed coordinates.

Functions of module *v.hull* [13] have been used to obtain vertices belonging to faces of convex hull. These functions were modified to output not new vector layer with convex hull but matrices containing coordinates of vertices and faces. Numerical accuracy of transformations and determining of volume of MBB were verified comparing results of module to outputs computed by *Matlab* [15]. This method enables also to verify coordinates exported of convex hull that we suppose to be created correctly.

Other functions of the module were tested while debugging or they were tested as part of 2D NNA functionality. For example, the function for computing average distance of the nearest neighbour $\bar{r}_A$ that is identical for 2D and 3D case (only input $z$ coordinates differ, they are zeros for two-dimensional space). Numerical accuracy of formulas for determining expected average distance between the nearest neighbours in set of randomly distributed points $\bar{r}_E$, ratio $R$ and test statistics $c$ were tested comparing to results obtained by *Mathematica* [14] while deriving of formulas and debugging.

The results of testing randomly distributed points are summarized in tables 4a and 4b. The same datasets as while testing in 2D space were used but $z$ coordinates were considered.

|  | n = 2000 | | |
|---|---|---|---|
|  | *v.nn_spatial_stat* | *Matlab* [15] script | difference |
| $\bar{r}_A[m]$ | 346.071782 | | |
| $\bar{r}_E[m]$ | 323.531486 | | |
| $R$ | 1.06967 | | |
| $c$ | 0.191691 | | |
| $V[m^3]$ | 398423031180.489 | 398423031174.718 | -5.771    $m^3$ |
| $t[s]$ | 0.093 | | |

Table 4a: The results of testing of the module *v.nn_spatial_stat* in three-dimensional space using 2000 randomly distributed points

|  | n = 5000 | | |
|---|---|---|---|
|  | *v.nn_spatial_stat* | *Matlab* [15] script | difference |
| $\bar{r}_A[m]$ | 251.243012 | | |
| $\bar{r}_E[m]$ | 238.607135 | | |
| $R$ | 1.052957 | | |
| $c$ | 0.145707 | | |
| $V[m^3]$ | 399562811897.870 | 399562811870.293 | -27.577    $m^3$ |
| $t[s]$ | 0.496 | | |

Table 4b: The results of testing of the module *v.nn_spatial_stat* in three-dimensional space using 5000 randomly distributed points

Considering value of test statistics $c \in (-1.96; 1.96)$, it is possible to assume that points were distributed randomly in both cases. Differences in volume of MBB are inappreciable comparing them to the values.

Coordinates of point clusters in three-dimensional space were computed in surrounding of randomly generated points using random distances, bearings with step 36° and zenith angles

with step $10°$. The results of tests are shown in tables 5a and 5b.

| | $n_0 = 100$ $\sigma = [0°; 36°; \dots 324°]$ $z = [-90°; -80°; \dots 90°]$ $r = random('Normal', 30, 10)$ | | |
|---|---|---|---|
| | $v.nn\_spatial\_stat$ | $Matlab$ [15] script | difference |
| $\overline{r}_A[m]$ | 7.03274 | | |
| $\overline{r}_E[m]$ | 155.180422 | | |
| $R$ | 0.04532 | | |
| $c$ | -2.626741 | | |
| $V[m^3]$ | 430279744723.451 | 430279744656.574 | -67 $m^3$ |
| $t[s]$ | 6.773 | | |

Table 5a: The results of testing of the module $v.nn\_spatial\_stat$ in three-dimensional space using clusters of points with shorter distances from local centres

| | $n_0 = 100$ $\sigma = [0°; 36°; \dots 324°]$ $z = [-90°; -80°; \dots 90°]$ $r = random('Normal', 300, 1000)$ | | |
|---|---|---|---|
| | $v.nn\_spatial\_stat$ | $Matlab$ [15] script | difference |
| $\overline{r}_A[m]$ | 182.221953 | | |
| $\overline{r}_E[m]$ | 349.816030 | | |
| $R$ | 0.520908 | | |
| $c$ | -1.318191 | | |
| $V[m^3]$ | 4784510371770.96 | 4784510371960.23 | 189 $m^3$ |
| $t[s]$ | 6.407 | | |

Table 5b: The results of testing of the module $v.nn\_spatial\_stat$ in three-dimensional space using clusters of points with longer distances from local centres

## 6. Outline of future work

It will be appropriate to enlarge testing of the module adding sample of points with maximal separation in three-dimensional space. Analogically to two-dimensional space where neighbouring points are arranged in hexagon around observed point, it will be necessary to find convex body with vertices located on equilateral triangles and plane cutting its vertices and center should also be composed of equilateral triangles. Body fulfilling condition that distances between vertices and centre definitely cannot be composed only of regular hexagons (proofed e.g. [17]: vertex of body is intersection of three polygons (faces) and sum of interior angles must be less than $360°$). Except that, in case of putting few pieces of these bodies together, there should be no empty spaces between them. Analysis of properties of truncated regular Plato's bodies described in Timaeus [18] or semi-regular Archimedes' bodies [19] will be purpose of future work.

The next item is to develop the mathematical background to better model fact that most of the phenomena may behave differently in horizontal and vertical direction. This analysis will be based on derivation of average distance of the nearest neighbours expected in case of

randomly distributed points $\overline{r}_E$ as function of two variables, horizontal distance h and vertical difference of elevation $z$:

$$\overline{r}_E = f(h, z)$$

Isotropic model of NNA that had been derived in the past and in this article was tested and completed for 3D GIS purposes may be useful for the applications mentioned in introduction, but it may be inappropriate for many phenomena on the Earth's surface. This assumption may be supported by e.g. differences found in tests' results. In test of randomly distributed points, there were both the samples detected as randomly distributed points in two- and also in three-dimensional space (see tables 1a, 1b and tables 4a, 4b). But in NNA of clustered points, results indicated randomness of distribution (tables 5a, 5b). This hypothesis may be verified testing the module on various samples including points located on three-dimensional pattern to compare behaviour of maximally separated points in 2D and 3D space.

**Appendix A** Derivation of $\sigma_{\overline{r}_E}$, the standard deviation of average distance between the nearest neighbours expected in case of randomly distributed points. The formulas have been derived analogically to the method [2] for two-dimensional space. Variance of average distance between the nearest neighbours in dataset of randomly distributed points:

$$var(r) = E(r^2) - (\overline{r}_E)^2$$

(A.1)

where $E(r^2)$ means dispersion. Standard deviation is then:

$$\sigma_{\overline{r}_E} = \sqrt{var(r)}$$

(A.2)

Derivation of dispersion $E(r^2)$

$$E(r^2) = \int\limits_0^\infty \left( \frac{4\pi n}{3} \cdot 3r^2 \cdot e^{-\frac{4\pi n}{3} \cdot r^3} \right) \cdot r^2 dr$$

(A.3)

It is necessary to express integral (A.3) as gamma function according to [5]:

$$\Gamma(z) = \int\limits_0^\infty e^{-x^2} \cdot x^{2z-1} dx$$

(A.4)

If $z = 2/3$,

$$\int\limits_0^\infty 4\pi n \cdot r^4 \cdot e^{-\frac{4\pi n}{3} \cdot r^3} dr = 2 \cdot \int\limits_0^\infty e^{-x^2} \cdot x^{1/3} dx$$

(A.5)

Then

$$x^2 = \frac{4\pi n}{3} \cdot r^3$$

$$(A.61)$$

$$r = \left( \sqrt{\frac{3}{4\pi n}} \cdot x \right)^{2/3} = \left( \frac{3x^2}{4\pi n} \right)^{1/3}$$

$$(A.62)$$

and also ($a$ means substitution constant):

$$4\pi n \cdot r^4 = 2a \cdot x^{1/3}$$

$$(A.7)$$

*Derivation of constant a*
To simplify derivation, (A.7) may be expressed as $4\pi n \cdot r^4 = 2a \cdot x^{1/3} \cdot x^{1/3} = 2a \cdot \left(x^2\right)^{1/3}$ (A.8) where $a_{A.7} \neq a_{A.8}$.

$$a = \frac{4\pi n \cdot r^4}{2 \cdot \left(x^2\right)^{1/3}} = 2\pi n \cdot \frac{\left(\frac{3x^2}{4\pi n}\right)^{4/3}}{\left(x^2\right)^{1/3}} = \sqrt[3]{\frac{3^4}{2^5 \cdot \pi n}} \cdot x^2 = \sqrt[3]{\frac{3^4}{2^5 \cdot \pi n}} \cdot \frac{4\pi n}{3} \cdot r^3 = \sqrt[3]{6\pi^2 n^2} \cdot r^3$$

$$(A.9)$$

After substituting (A.5) to (A.9):

$$\int\limits_0^\infty 4\pi n \cdot r^4 \cdot e^{-\frac{4\pi n}{3 \cdot r^3}} \, dr = \frac{2 \cdot \int\limits_0^\infty e^{-x^2} \cdot x^{1/3} dx}{\sqrt[3]{6\pi^2 n^2}} = \frac{\Gamma(2/3)}{\sqrt[3]{6\pi^2 n^2}} = \frac{1.35411793942640}{\sqrt[3]{6\pi^2 n^2}}$$

$$(A.10)$$

Derivation of variance $var(r)$

$$var(r) = E(r^2) - (r_E)^2 = \frac{\Gamma(2/3)}{\sqrt[3]{6\pi^2 n^2}} - \left( \frac{\Gamma(4/3)}{\sqrt[3]{4\pi n/3)}} \right)^2 =$$

$$= \frac{\frac{1}{\sqrt[3]{2\pi^2}} \cdot \left( \frac{\Gamma(2/3)}{\sqrt[3]{3}} - \frac{\sqrt[3]{9} \cdot (\Gamma(4/3))^2}{2} \right)}{\sqrt[3]{n^2}} = \frac{0.040536}{\sqrt[3]{n^2}}$$

$$(A.11)$$

Standard deviation of average distance between the nearest neighbour in randomly distributed set of points will be:

$$\sigma_{\overline{r}_E} = \sqrt{var(r)}$$

$$(A.12)$$

## References

[1] GRASS Development Team (2013): *Geographic Resources Analysis Support System (GRASS) Software* [computer software]. Open Source Geospatial Foundation Project. Available at: http://grass.osgeo.org.

[2] CLARK, P. J., EVANS, F. C. (2003): Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. In *Ecology* [online]. Vol. 35, Is. 4. October 1954 [cit. 2013-03-21], pp. 445-453. Available at: https://courses.washington.edu/bio480/Week1-PAPER-Clark_and_Evans1954.pdf. ISSN 0012-9658.

[3] HERTZ, P. (1909): Über den geigenseitigen durchschnittlichen Abstand von Punkten, die mit bekannter mittlerer Dichte im Raume angeordnet sind. In *Mathematische Annalen*, 67: 387-398. According to: CLARK, P. J., EVANS, F. C. (2003). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. In Ecology [online]. Vol. 35, Is. 4. October 1954 [cit. 2013-03-21], pp. 445-453. Available at: https://courses.washington.edu/bio480/Week1-PAPER-Clark_and_Evans1954.pdf. ISSN 0012-9658.

[4] CHANDRASEKHAR, S. (1943): The Law of Distribution of the Nearest Neighbor in a Random Distribution of Particles. In *Reviews of Modern Physics*. Stochastic Problems in Physics and Astronomy. Vol. 15, 1-89, 1943 [cit. 2013-03-21], pp. 86-87. Available at: http://rmp.aps.org/abstract/RMP/v15/i1/p1_1. ISSN 1539-0756.

[5] WEISSTEIN, E. W. (2013): Gamma Function. In *MathWorld*. A Wolfram Web Resource [online]. 2013 [cit. 2013-03-21]. Available at: http://mathworld.wolfram.com/GammaFunction.html.

[6] ESRI (2013): Average Nearest Neighbor [computer software]. In *ArcGIS Desktop: Release 10.1*, Spatial Statistics Toolbox [cit. 2013-05-14]. Redlands, CA: Environmental Systems Research Institute.

[7] HUGHES, I. G., HASE, T. P. A. (2010): *Measurements and their Uncertainities* : A practical Guide to Modern Analysis. 1. edition. New York : Oxford University Press Inc., New York, 2010. 136 p. ISBN 978-0-19-956633-4.

[8] ESRI (2012): What is a z-score? What is a p-value? In *ArcGIS Help 10.1* [online]. 2012 [cit. 2013-03-21]. Available at: http://resources.arcgis.com/en/help/main/10.1/index.html#/What_is_a_z_score_What_is_a_p_value/005p00000006000000/.

[9] KARPÍŠEK, Z. (2004): *Statistické tabulky* [Statistical tables]. Institute of Mathematics FSI VUT in Brno [online]. 2004 [cit. 2013-06-03]. Available at: http://mathonline.fme.vutbr.cz/default.aspx?section=2&article=121&highlighttext=tabulky

[10] ESRI (2013). *ArcGIS Desktop: Release 10.1*. Redlands, CA: Environmental Systems Research Institute.

[11] ESRI (2013): Average Nearest Neighbor (Spatial Statistics). In *ArcGIS Help 10.1* [online]. 2013 [cit. 2013-05-14]. Available at: http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000008000000

[12] NELSON, M. (2007): The Convex Hull. In *Computer Science*, Web Articles [online]. 2007 [cit. 2013-05-30]. Available at: http://marknelson.us/2007/08/22/convex/

[13] AIME, A., NETELER, M., DUCKE, B., LANDA, M. (2010): v.hull [computer software]. In *Geographic Resources Analysis Support System (GRASS) Software*. Available at: http://trac.osgeo.org/grass/wiki/DownloadSource#GRASS7

[14] Wolfram Research, Inc. (2008). *Mathematica* [computer software]. Version 7.0. Champaign : Wolfram Research, Inc.

[15] MathWorks, Inc. (2010). *MATLAB* [computer software]. Version 7.11.0. Natick, Massachusetts: The MathWorks, Inc.

[16] McCAULEY, J. D., LANDA, M. (2010): v.random [computer software]. In *Geographic Resources Analysis Support System (GRASS) Software*. Available at: http://trac.osgeo.org/grass/wiki/DownloadSource#GRASS7

[17] VALLO, D. et al. (2012): *Geometria telies...všeobecne a pútavo* [Geometry of bodies...in general and grippingly]. Faculty of Natural Sciences, Constantine the Philosopher University in Nitra [online]. 2012 [cit. 2013-05-30]. Available at: http://www.km.fpv.ukf.sk/admin/upload_pdf/20121108_145712___0.pdf. ISBN 978-80-558-0106-3.

[18] ZEYL, D. (2013) Plato's Timaeus. In *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition) [online], Edward N. Zalta (ed.). [cit. 2013-05-30].Available at: http://plato.stanford.edu/archives/spr2013/entries/plato-timaeus/.

[19] WEISSTEIN, E. W. (2013): Archimedean Solid. In *MathWorld. A Wolfram Web Resource* [online]. 2013 [cit. 2013-05-30]. Available at: http://mathworld.wolfram.com/ArchimedeanSolid.html.