{MERAA} Mathematics in Education, Research and Applications ISSN 2453-6881
Math Educ Res Appl, 2019(5),1, 1-8

Received: 2019-01-30
Accepted: 2019-06-18
Online published: 2019-10-15
DOI: https://doi.org/10.15414/meraa.2019.05.01.1-8

Original Paper

Nonparametric distribution of the daylight factor

Dušan Páleš[1*], Milada Balková[2]

[1] Slovak University of Agriculture in Nitra, Faculty of Engineering, Department of Machine Design,
[2] Slovak University of Agriculture in Nitra, Faculty of Engineering, Department of Building Equipment and Technology Safety

ABSTRACT

Kernel density estimation (KDE) approximates the distribution of statistical data similar to the histogram. The histogram of data is a special kind of the Kernel density. In the reconstructed building of stall in Oponice (Slovakia), we measured the values of daylight factor. The obtained data proved a bimodal distribution, so it was not appropriate to use some of the usual parametric distributions. This paper describes how Kernel density can be applied to measured results. We find out the values of the cumulative distribution function of such density, by probability procedures, that serves us comparison with the prescribed values of the daylight factor in the standard, on the one hand for animals (1.0%) and on the other hand for the people (1.5%) who care for animals. The results obtained from the measurements and the same ones approximated by KDE are in good agreement.

KEYWORDS: Kernel density estimation, bandwidth, daylight factor

JEL CLASSIFICATION: C13, C16

INTRODUCTION

Nonparametric probability density estimation is an important tool in statistical data analysis [2]. It allows capture multimodality, skewness or other irregular structure of obtained measurements. Compared to the classical parametric estimation of distributions, it has greater flexibility and efficiency. With the parametric approach [6], it is necessary to predict some distribution model, whose small number of parameters is estimated by the likelihood principle. A nonparametric approach does not require an a priori assumption of probability density distribution [4]. The probability density distribution is created directly from the examined data. The most popular nonparametric approach to estimate probability density is Kernel density estimation (KDE). The data, that are possible to model by KDE, can come from very complicated distributions that mathematically do not even have to be described exactly. In this paper, we fit KDE to the distribution of the daylight factor for cattle.

* Corresponding author: Dušan Páleš, Slovak University of Agriculture, Faculty of Engineering, Department of Machine Design, Trieda Andreja Hlinku 2, 949 76 Nitra, Slovakia, e-mail: dusan.pales@uniag.sk

ISSN 2453-6881
Math Educ Res Appl, 2019(5),1, 1-8

MATERIAL AND METHODS

Kernel density estimation

The histogram is the simplest nonparametric estimate of the probability density and concurrently it belongs to the most used. To create a histogram, we need two parameters: the starting position of the first bin and the bin width. However, the histogram has several drawbacks. Density estimate depends on the starting point of histogram. The discontinuities of the estimate are not created by underlying density, but only by the bin width of the histogram. The histogram is only suitable for one- or two-dimensional problems, because with a larger number of dimensions it becomes unclear.

A more convenient tool for nonparametric modelling of the distribution density is Kernel density estimation. Its advantage over the histogram lies in that it is smooth and continuous [8]. KDE assembles the measurement of values and creates so called kernels of the bandwidth in these values, which in the sum form the Kernel density estimation. If we have a set of data x_i, i = 1, ..., n, the KDE is expressed according to [12] by the relation

$$KDE(x) = \frac{1}{n.bw}\sum_{i=1}^{n} K\left(\frac{x-x_i}{bw}\right), \quad (1)$$

where $K\left(\frac{x-x_i}{bw}\right)$ represents the kernel and symbol bw means bandwidth. Figure 1 illustrates a method of producing KDE from individual kernels; on the x axis there are selected these points {14, 16, 20, 30, 33, 35}.

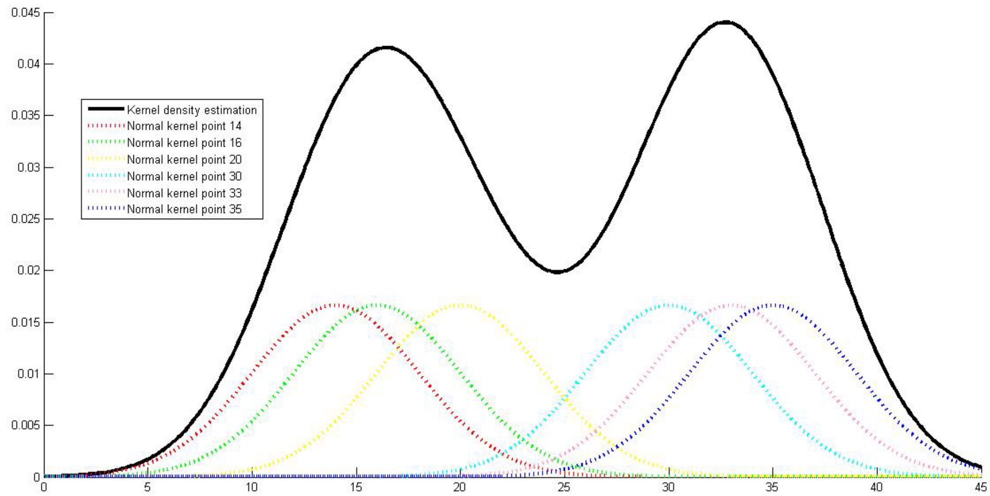

Figure 1 Example of the sum of individual kernels to the resulting KDE, probability density functions

The most used four types of kernels are (in parametric form)

$$\text{Normal:} \qquad K(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} \quad (2)$$

Box: $K(t) = 0.5$ (3)

Triangular: $K(t) = 1 - |t|$ (4)

Epanechnikov: $K(t) = \frac{3}{4}(1 - t^2)$ (5)

As Breheny stated in [3], each kernel must satisfy three conditions:

$$\text{symmetric to } 0; \quad \int K(u)\,du = 1; \quad \int u^2 K(u)\,du > 0 \tag{6}$$

In the Figure 2 the various types of kernels (2) - (5) are drawn and the property (6) can be demonstrated.

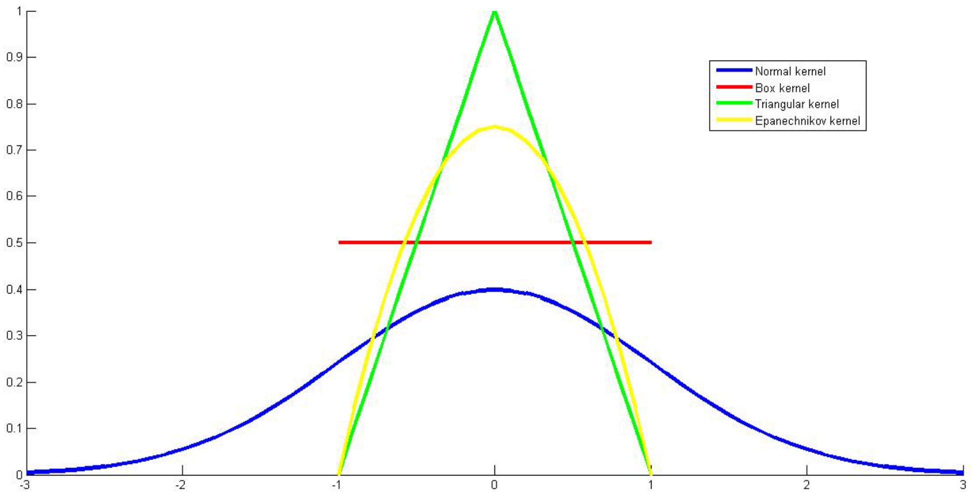

Figure 2 Types of kernels (2) - (5), probability density functions

It turns out that selection of the kernel is not as substantial as choice of the bandwidth bw. By choosing an incorrect value of bw, we can oversmooth the distribution by overvaluing it and undersmooth by undervaluing. The bandwidth can be optimized, for example, for the normal kernel [8].

Another important parameter for representation of KDE is the number of evenly spaced points on the x axis, in which the values of KDE are computed. Obviously, with an increasing number of points, the performance of the density will be more accurate.

Realization of the experiment

Daylighting was measured in the stall for cattle in Oponice (Slovakia). The stall has undergone extensive reconstruction to improve the conditions for animals. Satisfactory lighting is the part of a successful breeding. A detailed description of the building can be found in [1] and [5]. The Figure 3 shows ground plan and cross-section of the studied building also with the position of the measured points.

We measured the internal illumination E and at the same time the external comparison of light E_h in the depicted points. As it is determined in [11], the basic condition for performance of the measurements is the evenly cloudy sky with a brightness distribution for dark landscape. The measurements were made in the precision class 3, because the cattle conditions are degraded in comparison with residential homes [9]. We measured with two identical luxmeters Testo 545 [7]. We calculated the daylight factor by the formula

$$D = \frac{E}{E_h}.100, \tag{7}$$

where D is daylight factor [%], E is internal illuminance [lx], E_h is external comparison illuminance [lx].

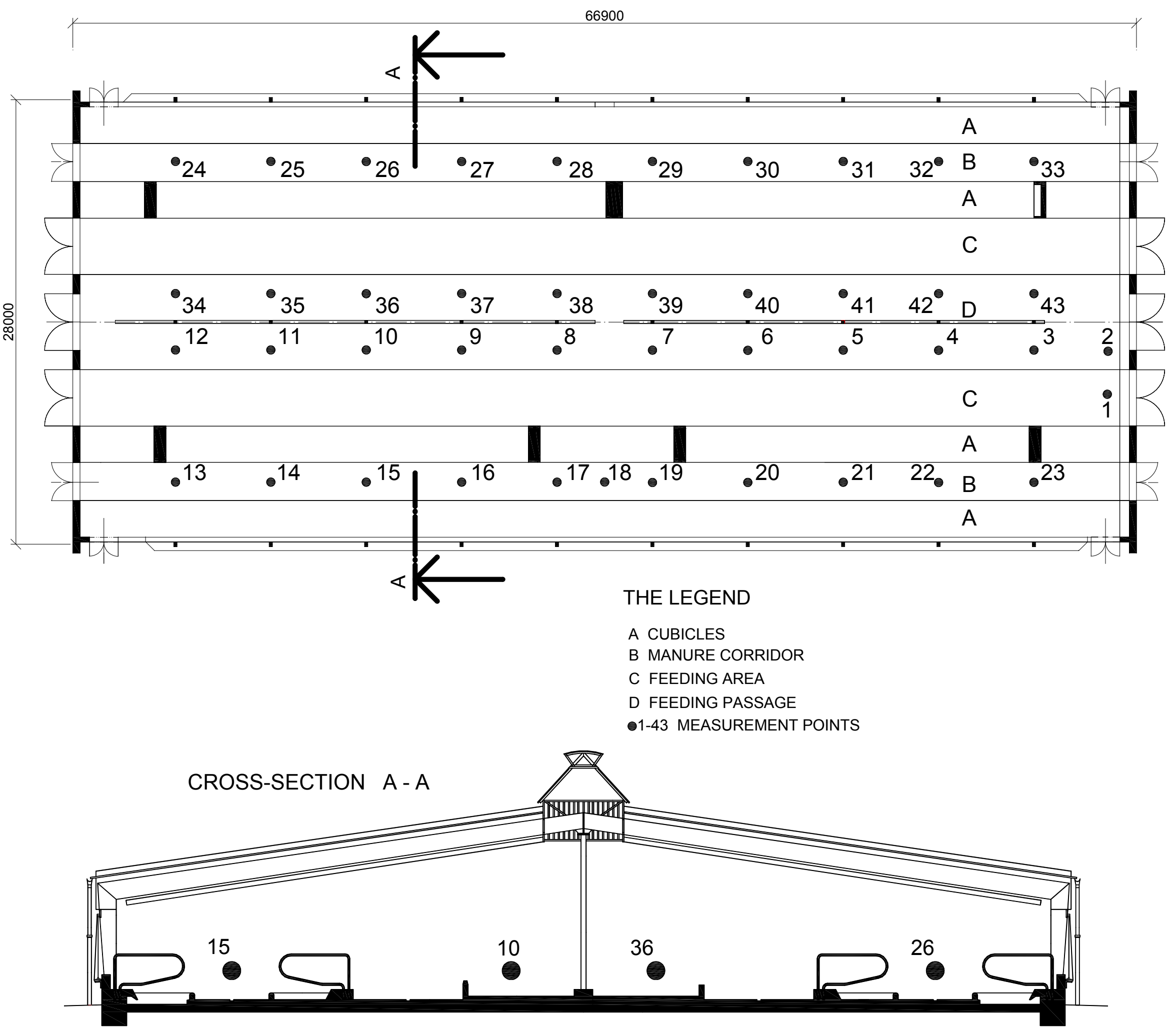

Figure 3 Ground plans and the cross-section of the measured building

The standard values of the daylight factor are listed in [10]. For a stall with dairy cows in free housing, the minimum D value should be 1.0 %. The value of D = 1.5% is needed for cattle workers.

RESULTS AND DISCUSSION

The resulting data in the measurement points transformed by the relation (7) are presented in the Table 1. At each point in the Figure 3 illuminance was measured five times, giving a total number of 215 measurements. As can be seen from the histograms of these data (Figures 4–6) the values have a bimodal probability density distribution. Such distribution shows the heterogeneity of data, probability feature is composed of two phenomena and it is worth considering to divide the data into two more homogeneous sets [13]. To fit on these values the classical parametric distribution is practically impossible. Kernel density estimation, however, without difficulty creates a probability density distribution over such data.

Table 1 Values of the daylight factor organized in ascending order, 215 values, obtained from measurements at 43 points in Figure 3

0.2195	0.6925	1.9011	2.3853	2.8451	3.8000	6.3425	6.8849	7.0987
0.2493	0.6938	1.9162	2.4482	2.8966	3.8211	6.3441	6.8872	7.0991
0.2524	0.6968	1.9336	2.4553	2.9219	3.8349	6.3457	6.8902	7.1041
0.2527	0.7005	1.9645	2.4593	2.9433	3.8356	6.4251	6.9383	7.1443
0.2534	0.7036	2.0274	2.4814	2.9801	3.8410	6.4270	6.9814	7.1684
0.2609	0.7167	2.0410	2.4861	2.9814	4.2793	6.4671	6.9870	7.1768
0.2624	0.7767	2.0644	2.5052	2.9952	4.4198	6.4889	6.9946	7.1811
0.2667	0.8051	2.1371	2.5095	3.0132	4.4201	6.4989	6.9977	7.1885
0.2696	0.8124	2.1492	2.5223	3.0212	4.4337	6.4996	7.0027	7.2394
0.2779	0.8266	2.1753	2.5279	3.0267	4.4873	6.5003	7.0059	7.2926
0.4505	0.8368	2.1765	2.5434	3.0332	5.9975	6.5095	7.0177	7.3122
0.4613	1.1022	2.2032	2.5502	3.0752	6.1256	6.5562	7.0191	7.3755
0.4679	1.1198	2.2202	2.5510	3.1553	6.1597	6.6477	7.0287	7.3833
0.4810	1.1486	2.2453	2.5770	3.2882	6.1828	6.7038	7.0414	7.4780
0.4941	1.3133	2.2494	2.5777	3.3212	6.1876	6.7145	7.0531	7.4851
0.6006	1.3657	2.2527	2.5878	3.3369	6.2071	6.7248	7.0543	7.5118
0.6197	1.6054	2.2544	2.6134	3.3614	6.2113	6.7422	7.0566	7.6676
0.6312	1.6172	2.2661	2.6240	3.3724	6.2427	6.7524	7.0585	7.7039
0.6323	1.6989	2.2688	2.6246	3.3827	6.2604	6.7616	7.0585	8.1081
0.6346	1.7200	2.2828	2.6415	3.6708	6.2733	6.7879	7.0689	8.1120
0.6718	1.7480	2.3068	2.6581	3.7432	6.2939	6.8052	7.0853	8.1128
0.6752	1.7847	2.3118	2.7064	3.7685	6.2959	6.8380	7.0856	8.2334
0.6792	1.8516	2.3205	2.7462	3.7798	6.3017	6.8768	7.0863	8.2486
0.6808	1.8562	2.3423	2.8032	3.7955	6.3370	6.8775	7.0873	

On the Figure 4 KDE is created using data with the optimized bandwidth bw = 1.2853 [8]. Obviously, by reducing the bw to 0.5, the distribution gets closer to the columns of the histogram and the number of its peaks increases, the distribution is undersmoothed. On the contrary, by increasing the value of bw to 2, the distribution is more aligned, the number of peaks decreases, the distribution is oversmoothed [12].

The Figure 5 presents the effect of choosing a different kernel on the form of the probability density. It turns out that the box kernel is quite jump, and also the triangular and

{MERAA} Mathematics in Education, Research and Applications ISSN 2453-6881
Math Educ Res Appl, 2019(5),1, 1-8

Epanechnikov kernels are not quite ideal. The smoothest shape has the normal kernel. But it is confirmed that the shape of KDE is not too sensitive to the kernel type, because the functions are not different.

The Figure 6 illustrates the effect of the number of evenly distributed points in which kernels are counted. Obviously, the highest value of 100 points gives smoother shape than the lower values 10 and 20 points. With a further increase in points, over 100, there is no more smoother function, so we do not show the higher number of points in the figure.

For KDE, normal kernel and optimized bandwidth bw = 1.2853 (Figures 4 - 6 depicted in blue) we calculated the cumulative distribution function values for daylight factor 1.0 % and 1.5 % as mentioned above [10]. In the Table 2, we compare them with the cumulative frequency calculated directly from the data. The value 1.0 %, acceptable for the animals, is approximated by KDE almost exactly and the value 1.5 %, ideal for the people, who care of animals, differs slightly.

Table 2 Comparison of the cumulative distribution function calculation from data and from Kernel density estimation

	1.0 %	1.5 %
Fractile from data	16.51 %	18.84 %
Fractile from KDE	16.28 %	22.30 %

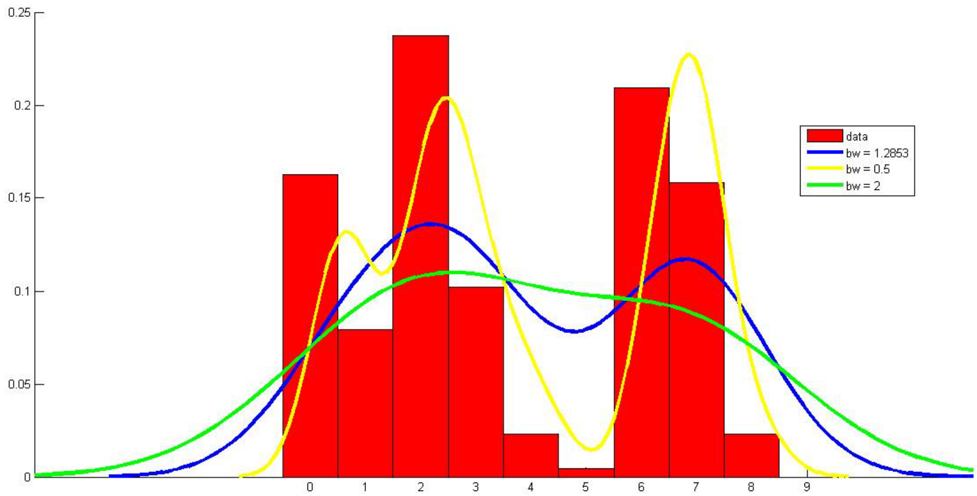

Figure 4 Histogram of the daylight factor together with Kernel distribution estimation for different bandwidth values

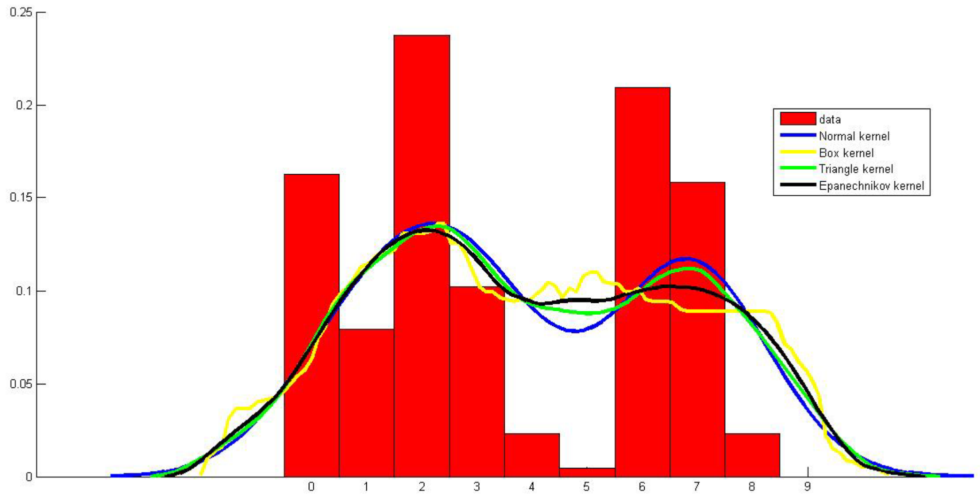

Figure 5 Histogram of the daylight factor together with Kernel distribution estimation for different types of kernels

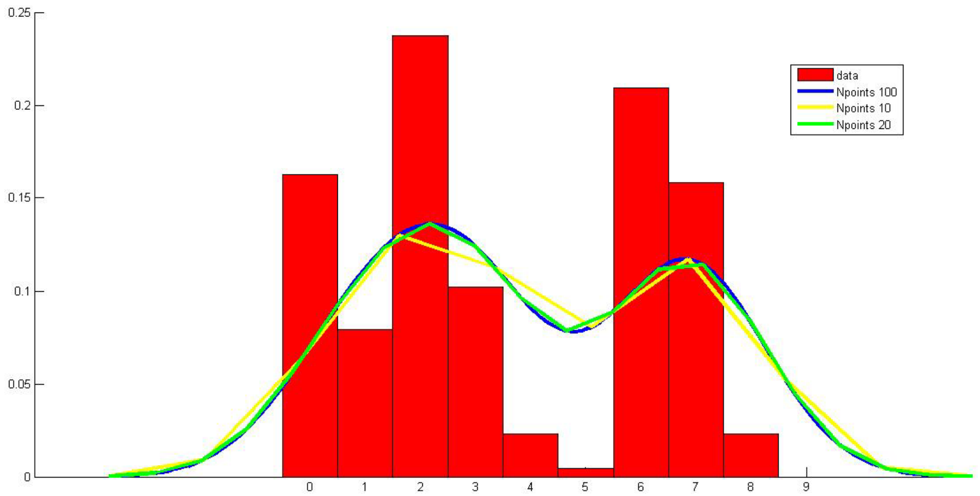

Figure 6 Histogram of the daylight factor together with Kernel distribution estimation for different numbers of approximate points

CONCLUSIONS

We applied the Kernel density estimation for the data obtained by measuring the daylight factor. The effect of bandwidth, kernel and number of points was examined on the shape of the density. It was particularly important to adjust the bandwidth, Figure 4, which the most influences the resulting shape of the distribution. By computing fractiles of the density, we found that roughly 16 % of the values measured in the examined building did not meet the light requirements for the animals and about 20 % of the values did not meet the requirements of the staff. The fractiles calculated from the Kernel density estimation are only slightly different from the fractiles calculated directly from the measured data and can be considered as a suitable approximation.

REFERENCES

[1] Balková, M. & Páleš, D. (2015). Influence of material solution of cover shells on stable daylight. *Research Agricultural Engineering*, Vol. 61, Special Issue, pp. S67-S73. Retrieved 2018-01-15 from https://doi.org/10.17221/47/2015-RAE

[2] Botev, Z., Grotowski, J. F. & Kroese, D. P. (2010). Kernel density estimation via diffusion. *Annals of Statistics*, 38, 2916-2957.

[3] Breheny, P. (2018). *Kernel density estimation.* Department of Biostatistics, University of Iowa. Retrieved 2018-10-12 from http://web.as.uky.edu/statistics/users/pbreheny/621/f10/notes/10-28.pdf

[4] Chen, Y. Ch. (2017). *A Tutorial on Kernel Density Estimation and Recent Advances*. Department of statistics, University of Washington. Retrieved 2018-01-15 from https://arxiv.org/pdf/1704.03924.pdf

[5] Karandušovská, I., Mihina, Š. & Boďo, Š. (2012). Effect of differently designed systems of housing for dairy cows on ammonia and carbon dioxide concentration. *Acta Technologica Agriculturae,* Vol. 15, pp. 106 – 109.

[6] Rédl, J. (2012). Stochastic processes in vehicle dynamics. In *Technika v technológiach agrosektora,* 2012, Slovak University of Agriculture in Nitra, pp. 182-187.

[7] Rybár, P. (1997). *Štandartná metodika. Meranie a hodnotenie osvetlenia.* Ministerstvo zdravotníctva SR, 35 pp.

[8] Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[9] STN EN 12464-1 (2004). *Svetlo a osvetlenie. Osvetlenie pracovných miest. Časť 1: Vnútorné pracovné miesta.*

[10] STN 36 0088 (1973). *Osvětlování v zemědělských závodech.*

[11] STN 73 0580 (1986, 1992). *Denné osvetlenie budov. Časť 1 – základné požiadavky, 1986. Časť 2 – Denné osvetlenie obytných budov, 1992.*

[12] Vermeesch, P. (2012). On the visualisation of detrital age distributions. *Chemical Geology*, Vol. 312-313, pp. 190-194. doi: 10.1016/j.chemgeo.2012.04.021.

[13] Zlacká, A. *Základy kvantitatívnych metód*. Part 2: Charakteristika jednorozmerných štatistických súborov. Katedra geografie a geoekológie, Prešovská univerzita. Retrieved 2018-01-15 from http://www.fhpv.unipo.sk/cvt/statistika/zlacka/skripta.htm