Looking closely at the dendrogram of similarity of the kernel estimators and making cuts at the level $d = 20\%$, we obtain the following division of estimators into taxa, i.e. into groups of objects with similar properties. Figures 4–7 show the obtained groups (taxa) of kernel estimators with similar behavior in the ranges of variability of the examined feature (groundwater level) defined by the experimenter. This analysis allows us to further reduce the set consisting of a large number of different kernel estimators obtained on the basis of different analytical and statistical concepts.

Note. All numerical calculations were performed using the authors' own original procedures on the R platform and using packages from the literature: [13–15].

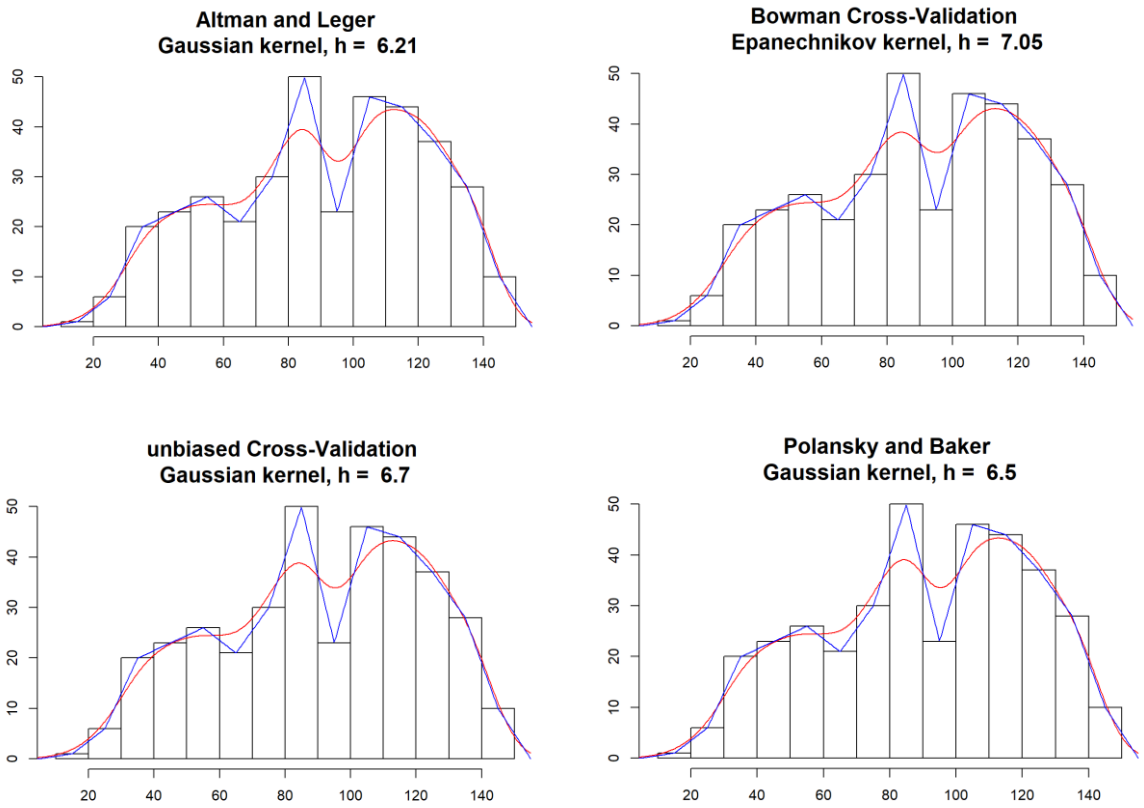

Fig. 4. Taxon I with similarity measure $s = 80\%$ (or equivalently distance $d = 20\%$).

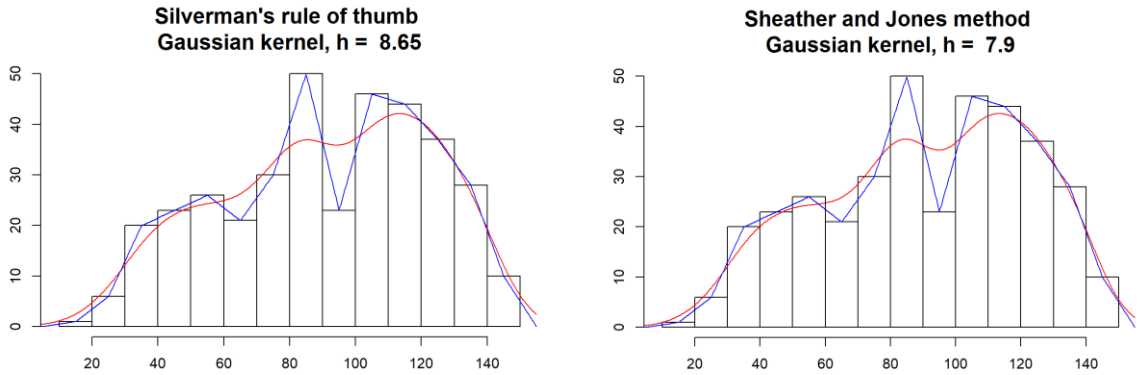

Fig. 5. Taxon II with similarity measure $s = 80\%$ (or equivalently distance $d = 20\%$).

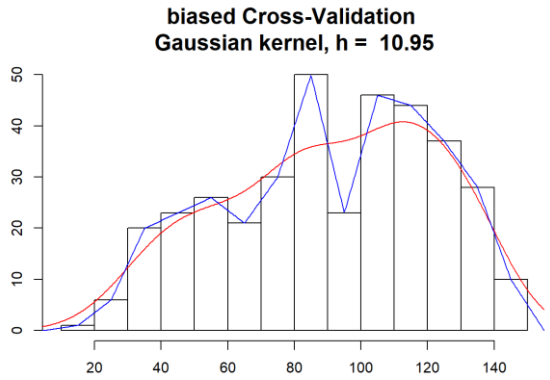

Fig. 6. Taxon III with similarity measure s = 20% (d = 80%).

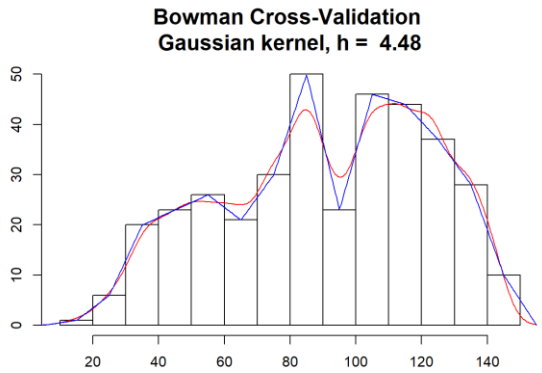

Fig. 7. Taxon IV with similarity measure s = 5% (d = 95%).

3 Conclusions

In the statistical literature there can be found a wealth of different approaches to obtaining the best estimate of the unknown density function of a continuous type random variable by nonparametric kernel estimation methods. Knowledge of the unknown density function that best reflects the probabilistic data structure is invaluable in the process of predicting values of the studied phenomenon. Using the example of hydrological data, the authors have suggested a way of classifying kernel estimators, chosen from those most often used in practice. Based on our calculations, we have shown that none of the considered estimators have optimal properties on the entire region. This is a known phenomenon in mathematical statistics related to the study of the admissibility of statistical decision rules. Each of the assessed estimators has good local properties, but their behavior is strictly dependent on empirical data. For example, Bowman et al. (1998) performed a simulation study comparing this method with the plug-in method of Altman and Leger. Better results are obtained, in general, with cross-validation (cf. [11]). Plug-in methods apply a pilot bandwidth to estimate one or more important features of the density function f. The bandwidth for estimating f itself is then estimated at a second stage using a criterion that depends on the estimated features. The best plug-in methods have proven to be very effective in diverse applications and are more popular than cross-validation approaches (see [4, 16]). However, other authors offer arguments against the uncritical rejection of cross-validation approaches.
Our results and the considerations of many authors give us the incentive to look for a solution that will allow us to use the best behavior of the tested estimators in a given area of variability, by suggesting, for example, an estimator that would be a convex linear combination of selected estimators. In 1989 Devroye introduced and developed the very interesting concept of the double kernel method for density estimation [17], and its usefulness has been demonstrated in extensive simulation studies [16]. In the double kernel method, we take two different kernels K and L whose characteristic functions do not coincide on any open neighborhood of the origin.

Only comprehensive knowledge of the efficiency and properties of the kernel density estimators for the one-dimensional case will allow us to consider cases of two-dimensional or three-dimensional random variables with greater awareness. The problem of stochastic modeling of hydrological or meteorological data using methods of multivariate density function estimation is more difficult and complex [18].

References

1. E. Gąsiorek, A. Michalski, and A. Pływaczyk, Zeszyty Naukowe AR **192** (1990)
2. A. Michalski, MHWM **Vol. 4**, 1, 41-46 (2016)
3. W. Feluch, J. Koronacki, Comput. Stat. Data Anal **13**, 143-151 (1992)
4. G.H. Givens, J.A. Hoeting, *Computational Statistics* (New York, Wiley & Sons, 2005)
5. B.W. Silverman, *Density Estimation for Statistics and Data Analysis* (London, Chapman & Hall, 1986)
6. M.P. Wand, M.C. Jones, *Kernel Smoothing* (London, Chapman & Hall, 1995)
7. E. Marczewski, H. Steinhaus, Colloq Math **6**, 319-327 (1958)
8. N. Altman, C. Léger, J. Stat. Plan, Inference **46**, 195-214 (1995)
9. D.W. Scott, G.R. Terrell, J. Amer. Statist. Assoc. **82**, 1131-46 (1987)
10. A.W. Bowman, Biometrika **71** (1984)
11. A.W. Bowman, P. Hall, T. Prvan, Biometrika **85** (1998)
12. A. Polansky, E.R. Baker, J. Stat. Comp. Sim **65** (2000)
13. R Core Team, *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2018)
    URL https://www.R-project.org/ (2018)

14. A. Quintela-del-Rio, G. Estevez-Perez, J. Stat. Softw. **50(8)**, 1-21 (2012)
15. A.C. Guidoum. *kedd: Kernel estimator and bandwidth selection for density and its derivatives*, R package version 1.0.3.
    http://CRAN.R-project.org/package=kedd (2015)
16. A. Berlinet, L. Devroye, Publications de l'Institut de Statistique de l'Université de Paris, vol. XXXVIII – Fascicule **3**, 3- 59 (1994)
17. L. Devroye, Annales de l'I H. P. **25**, 533-580 (1989)
18. D.W. Scott, *Multivariate density estimation. Theory, Practice and Visualization* (J. Wiley & Sons, Inc., 1992)