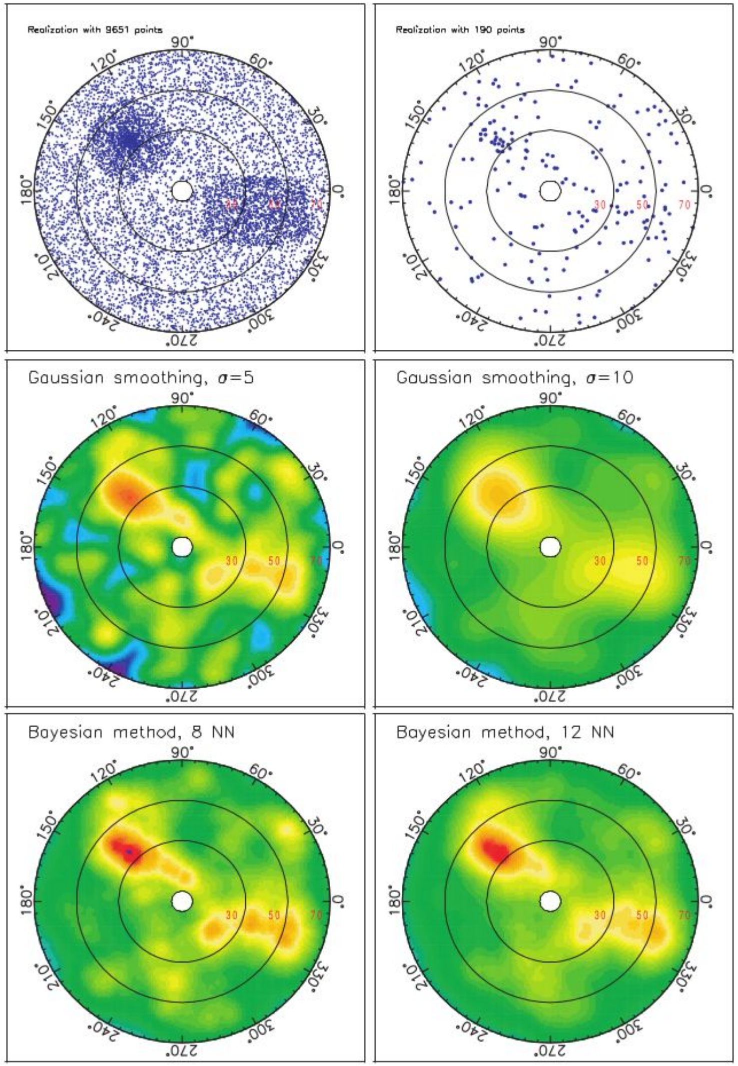
"On Estimating Non-Uniform Density Distributions Using N Nearest Neighbors" Woźniak & Kruszewski 2012

Radek Poleski

29 Oct 2025

What problem is discussed?

- Spatial studies of distributions of objects: RR Lyr, galaxies, GRBs, stars in globular clusters, finding star clusters…
- Simulations
- Color-magnitude diagrams of globular clusters

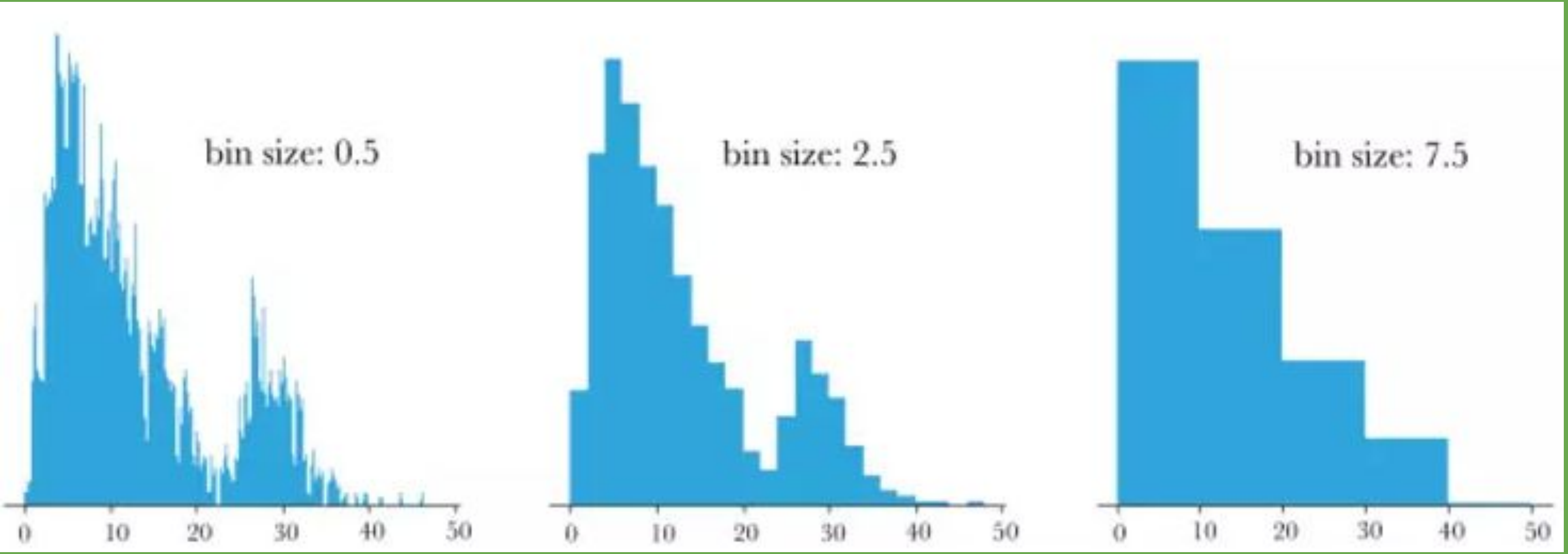


Ivezic et al. 2005

Solution 1 - binning

Obvious issues with bin size and phase.

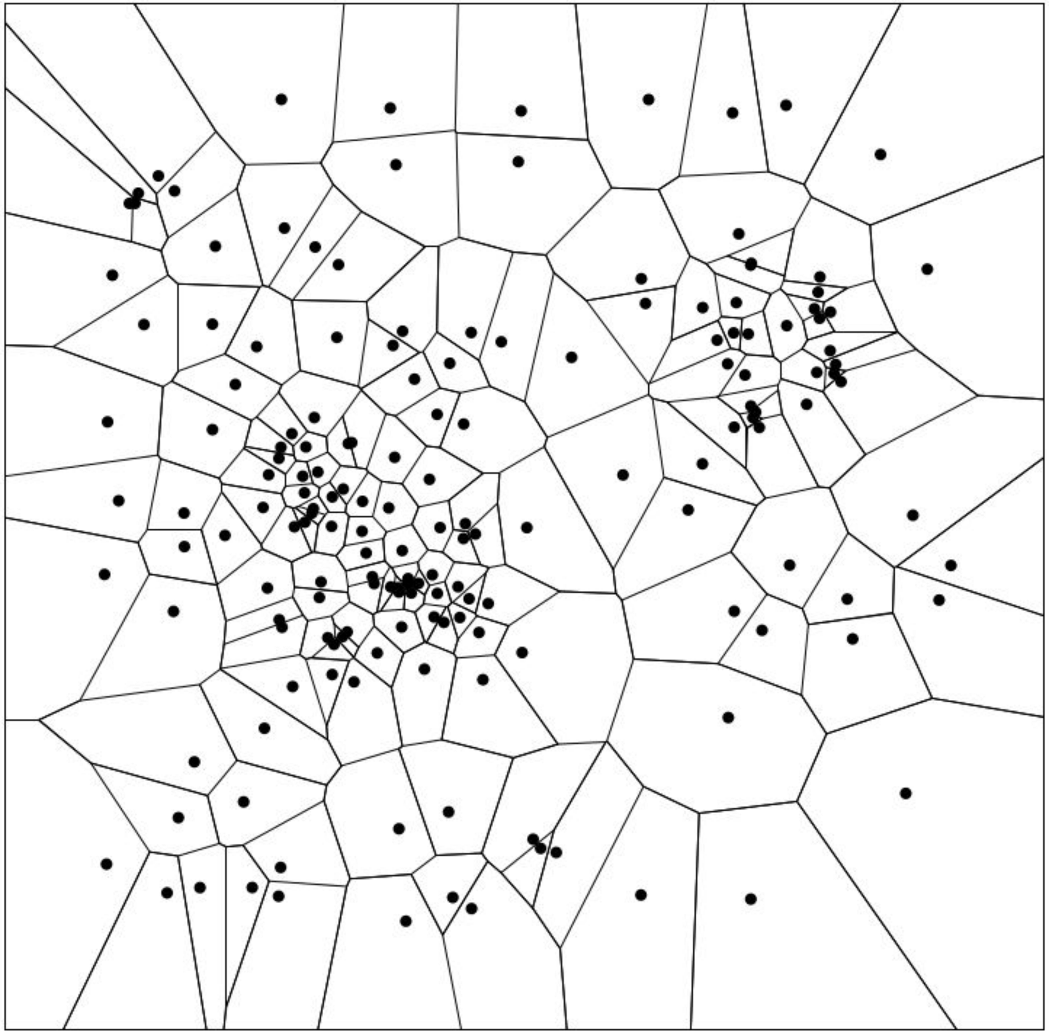
See SJC on 14 Dec. 2021 by P. Szewczyk



https://serc.carleton.edu/mathyouneed/geomajors/histograms/index.html

Solution 2 - Voronoi tessellation

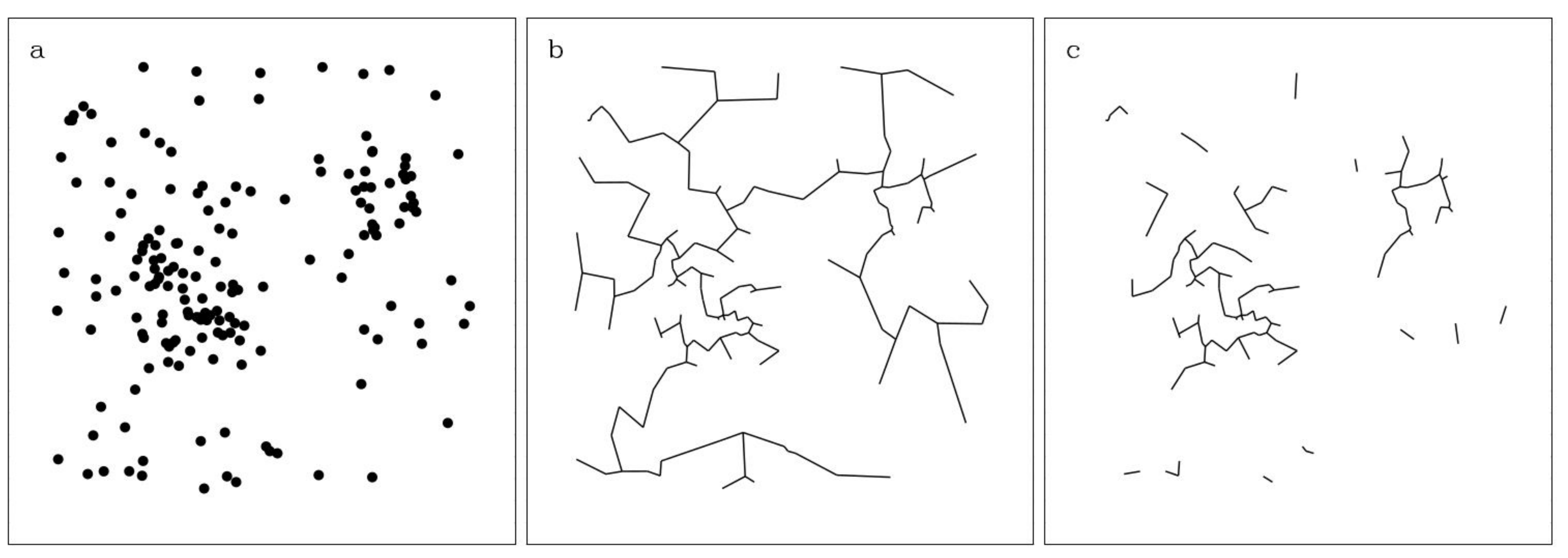
Very sensitive to small-scale fluctuations -> smoothing



Schmeja 2011

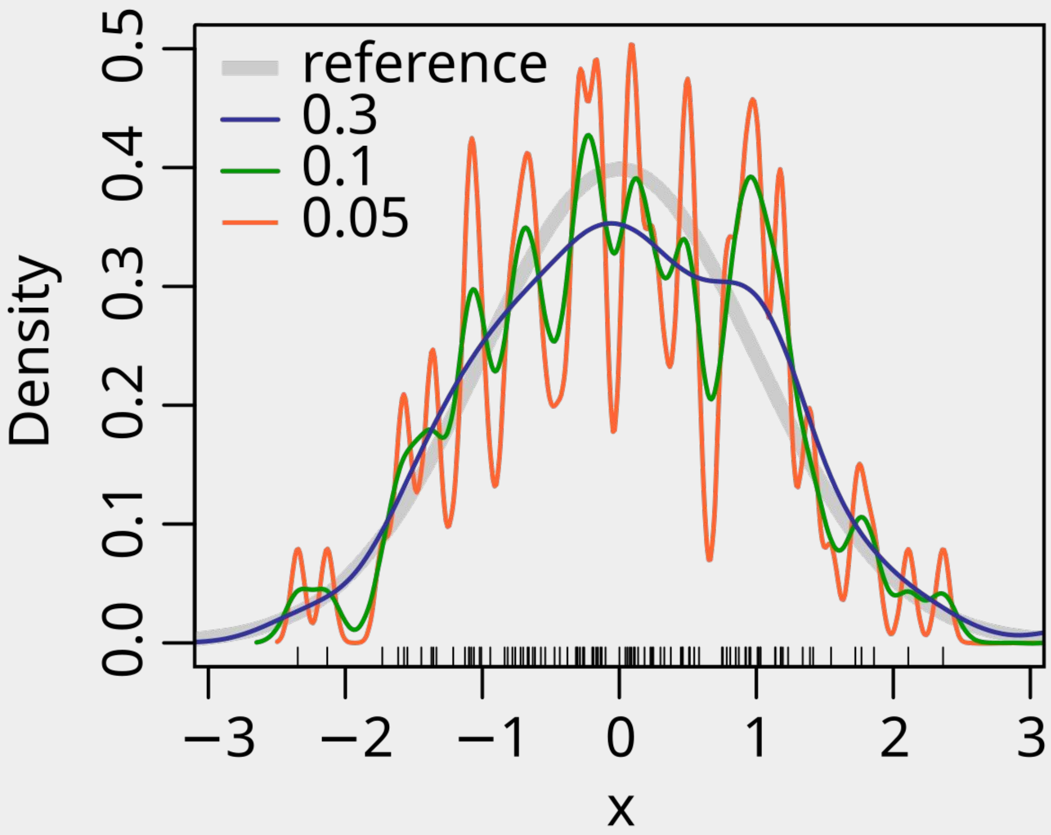
Solution 3 - minimum spanning tree separation

MST is the unique set of straight lines (“edges”) connecting a given set of points (“vortices”) without closed loops, such that the sum of the edge lengths is minimum.



Schmeja 2011

Solution 4 - kernel density estimation

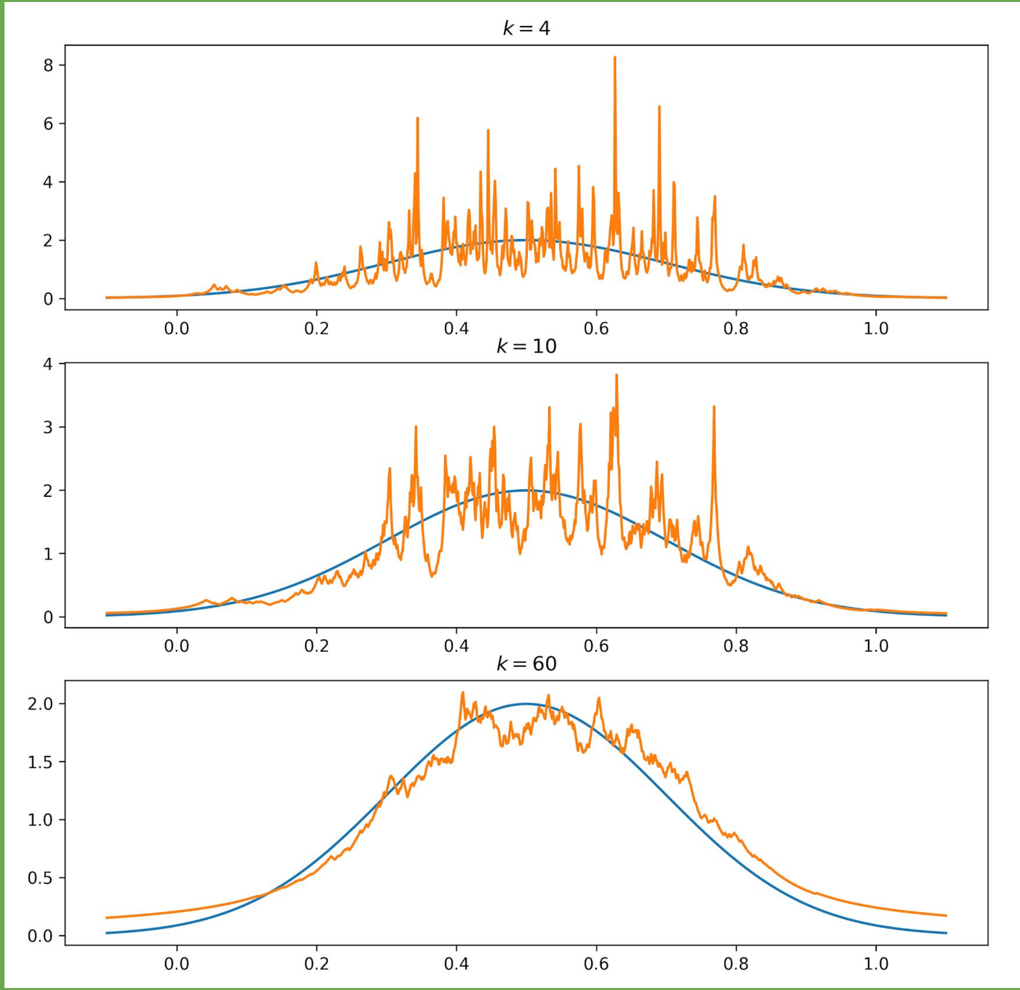


Toews/wiki

Solution 5 - nearest neighbour density

Depends only on a single parameter.

Slow, but faster than MST.



https://math.stackexchange.com/questions/4028633/which-is-the-algorithm-for-knn-density-estimator

5a - Bayesian approach to Nearest Neighbours

$$p(n_0|\{d_k;k=1,N\},I) \propto p(d_N|n_0;I)p(n_0|\{d_k;k=1,N-1\},I), \quad \text{(B2)}$$

$$p(n_0|\{d_k;k=1,N\},I) \propto p(n_0|I)\prod_{k=1}^{N} p(d_k|n_0,I). \quad \text{(B3)}$$

Ivezić et al. 2005

$$p(k|\mu) = \frac{\mu^k e^{-\mu}}{k!}. \quad \text{(B4)}$$

The number of expected points, $\mu = n_0 V_D(d)$, can be conveniently parametrized as $(d/d_0)^D$, where d_0 is the characteristic (mean) distance between two points [determined from $n_0 V_D(d_0) = 1$]. The probability that the distance to the kth neighbor is between d and $d+\delta d$ is the same as the probability that there are exactly $k-1$ neighbors enclosed by d and follows from equation (B4). (For a treatment of non-Poisson distributions, see White 1979.) With the change of variables from μ to d, the probability density distribution for the distance to the kth nearest neighbor becomes

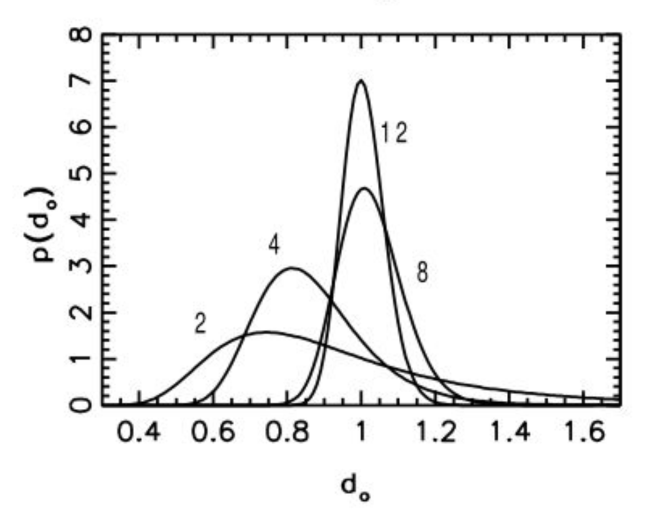
$$p(d_k|d_0) = \frac{De^{-(d_k/d_0)^D}}{d_0(k-1)!}\left(\frac{d_k}{d_0}\right)^{Dk-1}. \quad \text{(B5)}$$

5a - Bayesian approach to Nearest Neighbours

In the two-dimensional case adopted here to analyze the distribution of candidate RR Lyrae stars (because the SDSS Data Release 1 footprint consists of two elongated strips; see Fig. 5), the probability density distribution for d_0, which is related to the local density via $n_0 = \pi d_0^2$, is thus determined from[9]

$$p(d_0|\{d_k; k = 1, N\}) \propto \prod_{k=1}^{N} \frac{2e^{-(d_k/d_0)^2}}{d_0(k-1)!} \left(\frac{d_k}{d_0}\right)^{2k-1}. \qquad \text{(B6)}$$

A few steps in a typical realization of this computation are shown in the middle panel in Fig. 9, where $p(d_0)$ is evaluated for $N = 2, 4, 8$, and 12, for a random sample with true $d_0 = 1$.



[9] Eq. (B6) can be recast in a form that allows much simplified computation of the expectation value for the local density and its uncertainty (such that the dependences on d_k and d_0 are separated and the final expressions involve only a summation of d_k^D, without a need to evaluate the full posterior probability distribution [P. Wozniak 2004, private communication].

Ivezić et al. 2005

5a - Bayesian approach to Nearest Neighbours

Comments by Woźniak & Kruszewski 2012:

Another way of diminishing the smoothing bias was introduced by Ivezić *et al.* (2005) and Cowan and Ivezić (2008) who take a "Bayesian" approach to combine contributions from all N nearest neighbors. The net effect is, again, lower bias at the cost of increased variance. Here we propose a new method of dealing with the smoothing bias that captures the information on density variations contained in distances to all N nearest neighbors using the Legendre series expansion.

5b - use normalised volumes and Legendre polynomials

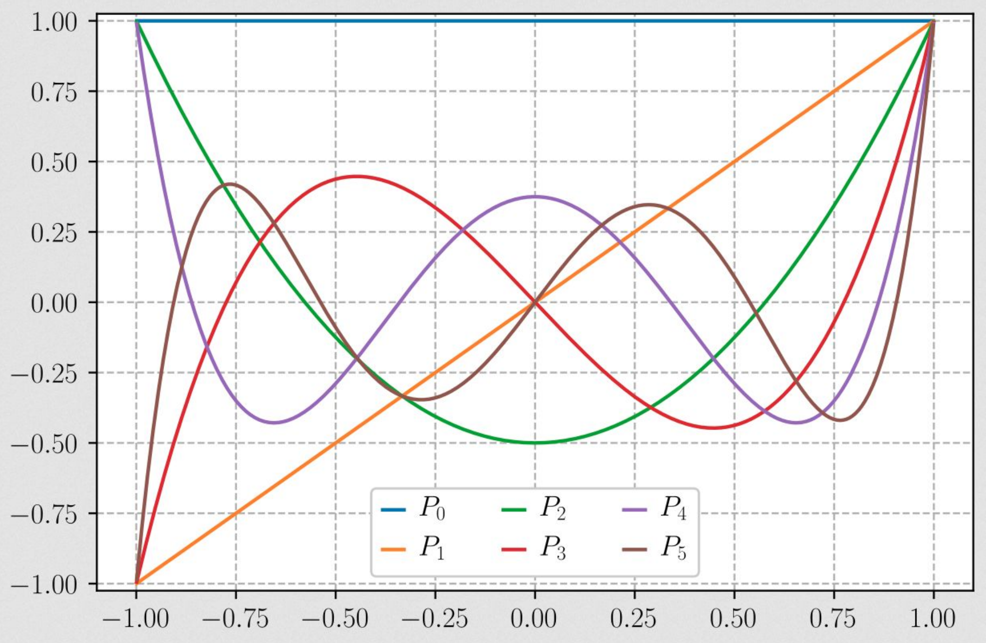
$$\hat{\rho}_{N,k} = \frac{1}{v_N} \sum_{i=1}^{N-1} \sum_{l=0}^{k} (-1)^l (2l+1) P_l(2y_i - 1)$$

Woźniak & Kruszewski 2012 (though the formula was known years earlier)

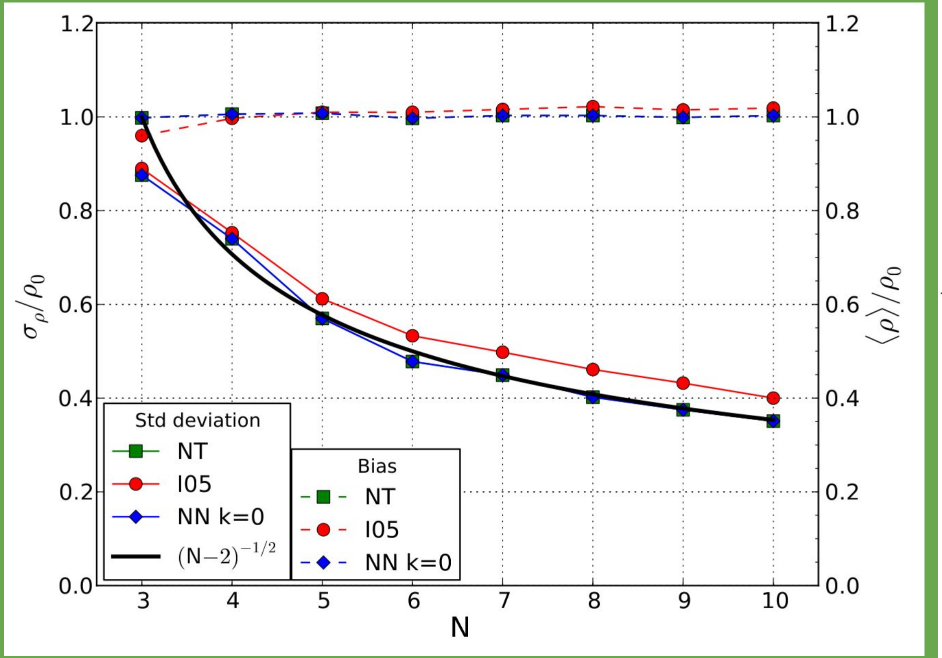
Watercolor caricature by Julien-Léopold Boilly (see § Mistaken portrait), the only Authenticated portrait of Legendre[2]

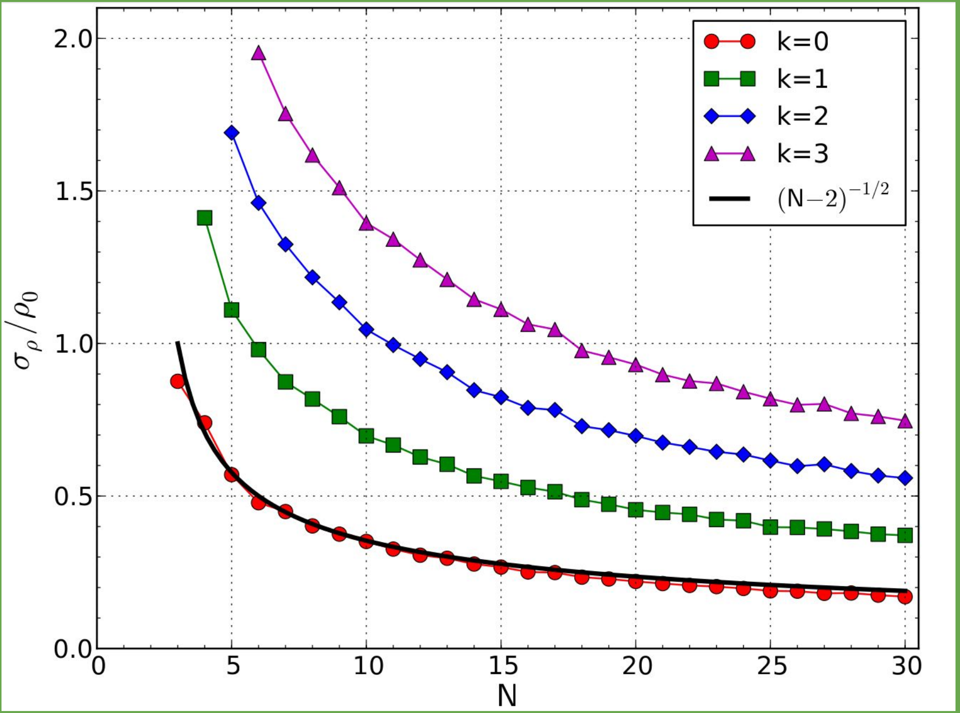
Portrait based on an anonymous and undated sketch, allegedly given by Legendre to François Arago in 1829, according to Arago's son.

Geek3/wiki



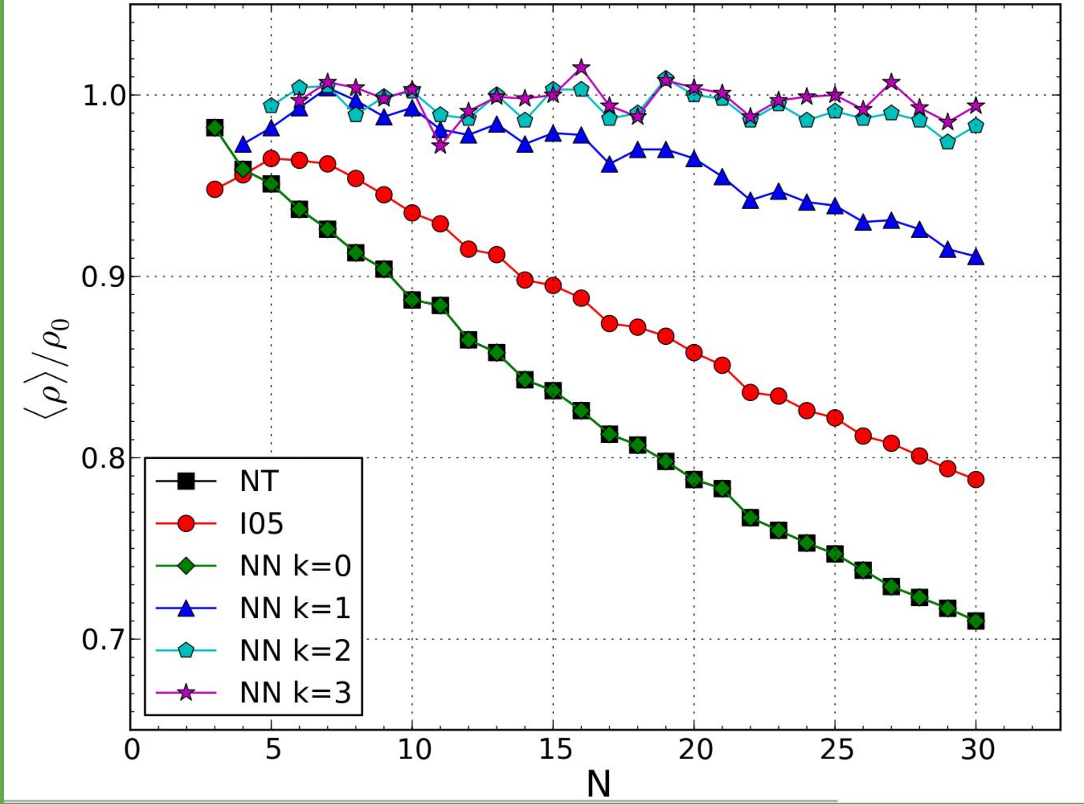
5b - use normalised volumes and Legendre polynomials



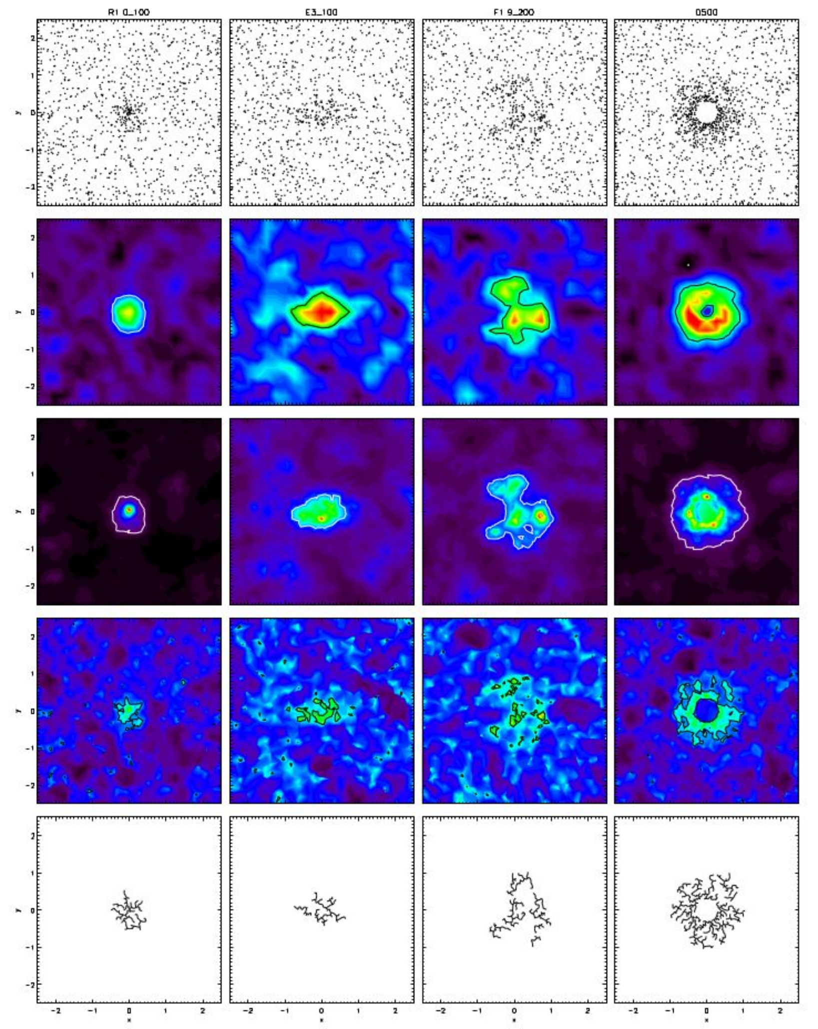


Woźniak & Kruszewski 2012

5b - use normalised volumes and Legandre polynomials



Woźniak & Kruszewski 2012



Model	SC	NN	VT	sVT	MST
R0.1_50	–	–	–	–	–
R0.1_100	○	○	–	–	–
R0.1_200	○	○	–	○	○
R0.1_500	●	●	○	●	●
R1.0_50	○	●	–	○	○
R1.0_100	●	●	○	○	●
R1.0_200	●	●	○	●	●
R1.5_50	●	●	●	●	●
R1.5_100	●	●	●	●	●
R1.5_200	●	●	●	●	●
E2_50	–	○	–	–	–
E2_100	○	●	–	–	○
E2_200	●	●	○	○	○
E3_50	○	○	–	–	○
E3_100	●	●	–	–	○
E3_200	●	●	○	○	●
F1.9_100	○	○	–	–	○
F1.9_200	○	●	–	○	○
F1.9_500	●	●	○	○	○
D100	○	○	–	–	–
D200	○	○	–	–	○
D500	●	●	○	○	○

Schmeja 2011