An area-weighted particle-size peak from image analysis predicts total dissolved solids across grinder types

Author: Naoya Oec **Affiliation:** Dogrun Inc., Suruga-ku, Shizuoka, Japan **ORCID:** 0000-0002-7491-4994

Keywords: coffee extraction; particle size distribution; image analysis; total dissolved solids; kernel density estimation; grind size

---

Abstract

Coffee extraction is rate-limited by solute diffusion from particle surfaces, yet the particle size distribution (PSD) is commonly summarized by a single central value such as D50, which discards information about the shape of the distribution. Image analysis now enables inexpensive acquisition of the full PSD, but it remains unclear which of the many computable features best predicts total dissolved solids (TDS). In this study, 12 samples from two grinders (conical burr and flat burr, six settings each) were imaged and processed through a pipeline based on Fiji/ImageJ and Python. Spearman rank correlations between PSD features and TDS were compared. The area-weighted KDE peak (area_weighted_peak), defined as the mode of a kernel density estimate weighted by each particle's measured projected area, yielded the strongest and most consistent correlation (pooled $\rho = -0.949$; $|\rho| > 0.94$ for both grinders). Other features, including volume-weighted D50 (Dv50), De Brouckere mean diameter (D43), and count-weighted KDE peak (kde_peak), showed weaker or grinder-dependent correlations, whereas the Sauter mean diameter (D32) and area-weighted D50 (Da50) — the area-weighted mean and median of the same distribution — performed nearly as well as the area-weighted KDE peak (its mode). Among the evaluated descriptors, the area-weighted KDE peak showed the strongest and most consistent performance. It is independent of grinder geometry and can be computed from low-cost imaging equipment.

---

1. Introduction

Coffee extraction is rate-limited by solute diffusion from the surfaces of ground particles. Finer grinding increases the total surface area, leading to higher extraction strength as measured by TDS.

D50, the volume-weighted median diameter, is widely used to summarize particle size distributions. However, D50 does not reflect the shape of the distribution, such as the proportion of fines or the coarse tail. Two samples with identical D50 values can yield different extraction outcomes if their distributions differ in shape.

Image analysis enables inexpensive acquisition of the full particle size distribution, but which of the many computable features correlates most strongly with TDS has not been systematically compared.

This study systematically compares multiple PSD features computed from a single imaging pipeline and identifies the descriptor that best predicts TDS. Two grinder types (conical and flat burr) are used to test whether the prediction is consistent across grinder geometries.

---

2. Materials and Methods

2.1 Coffee samples and brewing

A single-origin coffee from Buenos Aires Farm, Nicaragua (Fully Washed, medium roast) was used throughout. Extraction followed a cupping-based immersion protocol. Ground coffee (8 g) was placed in a cupping bowl and 150 g of water at 96 °C was poured over it (brew ratio 1:18.75). After 4 min, the surface was stirred four times with a spoon and the floating crust was removed. At 5 min, approximately 2 mL of brew was drawn with a syringe, passed through a 0.45 µm syringe filter, and collected in a microcentrifuge tube. After cooling to room temperature (25.2–25.4 °C), TDS was measured three times with a DiFluid R2 Extract refractometer and the median value was recorded.

2.2 Grinders and grind settings

Two grinders were used:

- LAGOM Casa (65 mm conical burr): settings 8, 13, 15, 17, 19, 21
- Timemore Sculptor 078 (SSP flat burr): settings 2, 4, 7, 11, 14, 17

One grind-and-brew cycle was performed per setting, yielding 12 samples in total.

2.3 Imaging and particle measurement

An iPhone 14 Pro fitted with a 10× optical zoom lens (equivalent focal length 75 mm) via an adapter was used to photograph ground coffee from directly above under ring LED illumination.

Ground coffee was spread with a spatula onto a paper stage bearing a circle of 50 mm diameter. Only particles within the circle were analyzed. The known diameter of the circle (50 mm) served as the pixel-to-millimeter calibration reference, providing self-calibration on every shot.

Segmentation was performed with headless Fiji (ImageJ). Images were converted to 8-bit grayscale and scaled using the stage circle diameter. Binarization used ImageJ's built-in automatic threshold (Default / IsoData method), which computes a threshold dynamically for each image. Pixels outside the circle were excluded with Clear Outside. A downstream Python step filtered particles by Feret diameter (0.1 mm ≤ Feret ≤ 3.0 mm) and projected area (≥ 0.01 mm²) to remove noise and obvious non-particles.

Particle size was quantified as the Feret diameter (maximum caliper diameter). For each sample, ground coffee was spread three times and photographed separately; the detected particles from all three images were pooled at the particle level (replicate_combined). The number of particles per sample after filtering ranged from 138–662 particles.

2.4 Particle size features

The following features were computed from the filtered Feret diameter set of each sample.

KDE peak features (three variants)

The mode of the particle size distribution was estimated as the peak of a kernel density estimate (KDE). A Gaussian kernel was used with bandwidth selected by Scott's rule (scipy.stats.gaussian_kde). The KDE was evaluated on a grid of 1024 equally spaced points spanning the observed diameter range, and the peak was taken as the argmax on this grid.

The three variants differ only in their weighting:

- kde_peak: unweighted (all particles contribute equally; count-weighted)
- sqrt_area_weighted_peak: weighted by the square root of measured projected area ($\sqrt{\text{Area}}$)
- area_weighted_peak: weighted by measured projected area (Area in mm², as reported by Fiji)

Here, Area is the projected area measured by Fiji for each individual particle and reflects the actual irregular shape, not a circular approximation ($\pi d^2/4$). Stronger area weighting increases the contribution of larger particles. kde_peak is dominated by the numerous fine particles, whereas area_weighted_peak captures the diameter class that accounts for the largest share of total particle area.

Summary statistics (computed from Feret diameters)

- mean: arithmetic mean
- median: count-weighted median (equivalent to Dn50)
- std: sample standard deviation
- cv: coefficient of variation (std / mean)
- skew: skewness
- kurtosis: excess kurtosis

Reference diameters (computed from equivalent circle diameter, for comparison)

Using a separate script (recompute_d50.py), the equivalent circle diameter ($ECD = 2\sqrt{(\text{Area}/\pi)}$) was computed for each particle, and the following statistics were derived. Prior to this calculation, one oversized non-particle artifact per replicate (area ≈ 2550–2950 mm², attributable to ROI-boundary detection; genuine particles did not exceed 5.5 mm²) was excluded using an area threshold (filt_max_area = 50). Units are millimeters, using the same stage-circle calibration described in Section 2.3.

- D32 (Sauter mean diameter): $\Sigma d^3/\Sigma d^2$
- D43 (De Brouckere mean diameter): $\Sigma d^4/\Sigma d^3$
- Dv50 (volume-weighted D50): weighted median with weights d^3
- Da50 (area-weighted D50): weighted median with weights d^2
- Dn50 (count-weighted D50): unweighted median

2.5 Statistical analysis

Spearman rank correlation coefficients between each feature and TDS were computed using scipy.stats.spearmanr.

- Pooled: all 12 samples
- Per-grinder: LAGOM Casa (n = 6) and Timemore Sculptor (n = 6) separately

Because the per-grinder analyses have only n = 6, p-values have low statistical power and are reported for reference only. Results were interpreted primarily on the basis of the magnitude and sign consistency of ρ.

---

3. Results

3.1 Particle size distributions

Figure 1 shows the area-weighted KDE of Feret diameters for each grinder setting. For both grinders, finer settings shifted the KDE peak toward smaller diameters.

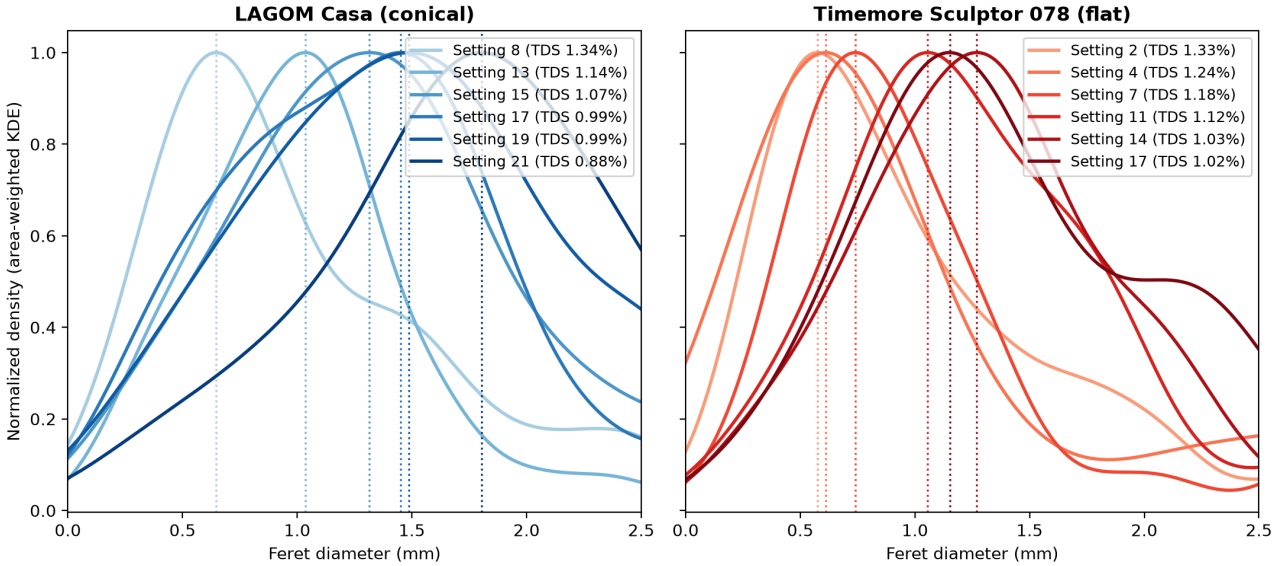

Figure 1. Area-weighted KDE of particle-size distributions

Figure 1. Area-weighted KDE of Feret diameters for LAGOM Casa and Timemore Sculptor 078, color-coded by grind setting with TDS values in the legend.

3.2 Correlation between particle size features and TDS

Table 1 lists the Spearman rank correlation coefficients between each feature and TDS.

Feature	Pooled ρ	p	Casa ρ	p	Sculptor ρ	p
area_weighted_peak	−0.949	<0.001	−0.986	<0.001	−0.943	0.005
std	−0.855	<0.001	−0.899	0.015	−0.886	0.019
mean	−0.802	0.002	−0.986	<0.001	−0.886	0.019
sqrt_area_weighted_peak	−0.473	0.121	−0.609	0.200	−0.943	0.005
kde_peak	−0.112	0.729	−0.812	0.050	+0.143	0.787
D32	−0.893	<0.001	−0.899	0.015	−0.943	0.005
D43	−0.434	0.158	−0.841	0.036	−0.314	0.544
Dv50	−0.448	0.144	−0.754	0.084	−0.314	0.544
Da50	−0.890	<0.001	−0.899	0.015	−0.943	0.005
Dn50	−0.060	0.854	−0.382	0.454	+0.314	0.544
median	−0.021	0.948	−0.232	0.658	−0.771	0.072

Per-grinder p-values are reported for reference (n = 6).

area_weighted_peak showed the highest absolute correlation in the pooled analysis (ρ = −0.949) and maintained |ρ| > 0.94 for both grinders. D32 (Sauter mean diameter) and Da50 (area-weighted D50) followed closely (pooled ρ = −0.893 and −0.890) and remained comparably strong for both grinders (|ρ| ≥ 0.899).

kde_peak exhibited a sign reversal between grinders (LAGOM Casa −0.812, Sculptor +0.143). Dv50 and D43 showed moderate pooled correlations but weakened substantially for Sculptor (ρ = −0.314 for both).

3.3 area_weighted_peak versus TDS

Figure 2 shows the relationship between area_weighted_peak and TDS.

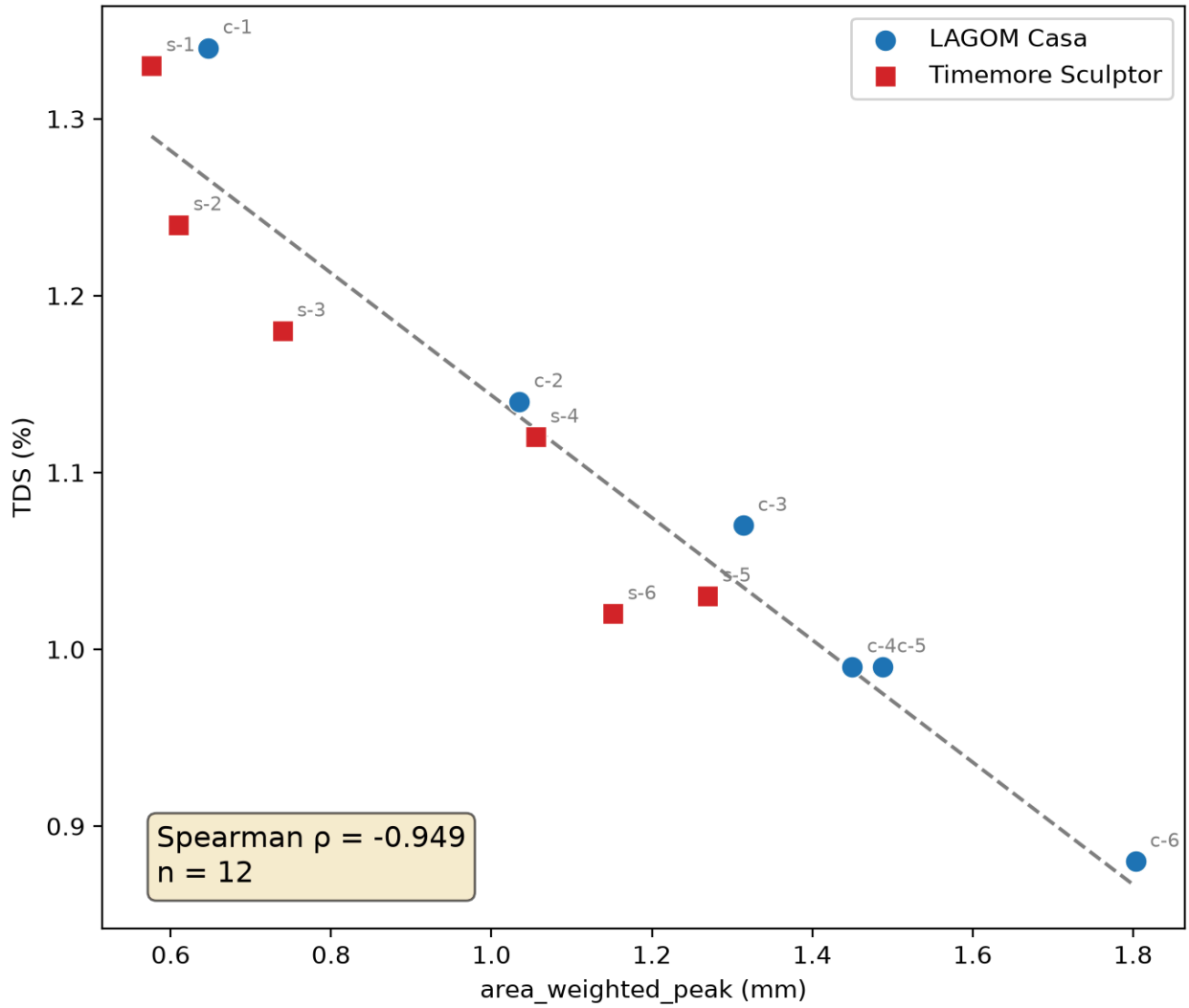

Figure 2. Relationship between area_weighted_peak and TDS

Figure 2. Scatter plot of area_weighted_peak (mm) versus TDS (%). Blue circles = LAGOM Casa; red squares = Timemore Sculptor. Dashed line: linear fit to all 12 samples.)

Data points from both grinders fell along the same negative trend, indicating that area_weighted_peak functions as a TDS predictor regardless of grinder type (conical or flat burr).

---

4. Discussion

4.1 Why the area-weighted KDE peak best predicts TDS

Coffee extraction is rate-limited by solute diffusion from particle surfaces. The diameter range where the extractable surface area is concentrated should therefore govern TDS.

The count-weighted kde_peak is pulled toward the fine-particle mode, where particle counts are highest. The sign reversal observed between grinders (−0.812 for LAGOM Casa vs. +0.143 for Sculptor) indicates that the peak position is dominated by grinder-specific fine-generation patterns, breaking its correspondence with TDS.

Volume- and mass-weighted representative diameters differed markedly depending on the weighting exponent. D32 (Sauter mean diameter, weighted by d^3/d^2) and Da50 (area-weighted median, weighted by d^2) achieved correlations nearly as strong and as consistent across grinders as area_weighted_peak (pooled $\rho = -0.893$ and -0.890; $|\rho| \geq 0.899$ for both grinders), consistent with the interpretation that weighting by particle area captures the diameter class governing extractable surface area, whether expressed as a KDE mode or a weighted quantile.

In contrast, Dv50 (volume-weighted median, weighted by d^3) and D43 (De Brouckere mean diameter, weighted by d^4/d^3) place even greater emphasis on a small number of coarse particles. These features showed moderate pooled correlations but weakened substantially for Sculptor ($\rho = -0.314$ for both), suggesting that part of their pooled correlation reflected systematic between-grinder differences in the coarse tail rather than a within-grinder relationship.

area_weighted_peak, D32, and Da50 therefore form a family of area-proportional descriptors that are effective, grinder-independent TDS predictors, in contrast to count-weighted (kde_peak, Dn50) or over-weighted (Dv50, D43) alternatives. Within this family, area_weighted_peak achieved the single highest correlation in every comparison (pooled and per-grinder) and, as a continuous KDE mode, does not depend on the choice of quantile or the interpolation scheme required by weighted-quantile estimators such as D32 and Da50. We therefore highlight it as the primary predictor, while noting that D32 and Da50 are close alternatives that may be preferable when a simpler, KDE-free calculation is desired.

Notably, these three descriptors correspond to the classical mean, median, and mode of the same underlying distribution: because Fiji's measured Area is proportional to d^2 by the ECD definition, D32 ($= \Sigma d^3/\Sigma d^2$) is exactly the area-weighted arithmetic mean of diameter, Da50 is its area-weighted median, and area_weighted_peak is its area-weighted mode. Their convergent performance therefore supports area weighting itself — rather than the specific choice of location statistic — as the operative principle linking particle size to TDS.

4.2 Relation to prior work

Battefeld-Montgomery et al. (2024) reported that the volume share of particles below 100 µm (share of fines, Q100µm) predicted espresso extraction more strongly than the median particle size. This finding demonstrated that the contribution of fines — which D50 does not capture — is important for predicting extraction, consistent with our results.

area_weighted_peak differs from the share-of-fines metric in that it is determined continuously from the entire distribution via KDE, without requiring an arbitrary size threshold (100 µm). It can be viewed as a continuous generalization of the same underlying insight: surface-relevant quantities, rather than central values, predict extraction strength.

4.3 Limitations and future work

This study has several limitations. The sample size is N = 12 (two grinders × six settings), and grinder diversity is limited. Validation with four or more grinders is underway.

This study aimed to identify predictive features and did not compare the imaging pipeline's diameter measurements against an external reference method such as laser diffraction.

The experiments used a cupping-based immersion protocol. Applicability to other brewing methods remains to be tested.

A single origin and roast level were used. The effect of origin and roast degree on the relationship between features and TDS is a topic for future work.

The robustness of area_weighted_peak to variation in lighting conditions and particle spreading methods is also under investigation.

---

5. Conclusions

Among the particle size features computed from the imaging pipeline, area_weighted_peak (area-weighted KDE peak) showed the strongest and most consistent Spearman rank correlation with TDS (pooled ρ = −0.949). This correlation was maintained across both conical and flat burr grinders.

Volume-weighted representative diameters such as Dv50 and D43 showed grinder-dependent correlations and were insufficient as TDS predictors, whereas D32 and Da50 — the area-weighted mean and median of the same distribution as area_weighted_peak (its mode) — performed nearly as well as area_weighted_peak itself.

area_weighted_peak can be computed from low-cost imaging and is a practically promising extraction-strength predictor that does not depend on grinder type; D32 and Da50 offer comparable performance without requiring KDE estimation.

Acknowledgements

The author thanks Roaster Muramatsu at hug coffee, Shizuoka, Japan, for providing the coffee beans used in this study.

Declaration of competing interest

The author is the developer of GrindGuide, an image-based particle size measurement service. The feature investigated in this study is intended for implementation in that service.

AI use disclosure

Generative AI tools, including ChatGPT, Claude, and Gemini, were used to support language editing, translation, manuscript organization, and formatting. The author critically reviewed and revised all AI-assisted text and verified the data, analyses, references, figures, and final interpretation. The experimental design, data collection, statistical analysis, and conclusions are the responsibility of the author. No generative AI tool is listed as an author.

Data availability

The datasets generated and analyzed during this study are available from figshare: https://doi.org/10.6084/m9.figshare.32897561. The deposited data include particle-level measurements (labeled_csvs/), TDS measurement records (tds_log.jsonl), derived particle-size statistics (particle_size_stats.csv), and the combined per-sample feature table

(feature_table_for_publication.csv), which contains all per-sample variables used in the statistical analyses reported in this study.

The analysis scripts used to reproduce the published analyses and figures are available from the GitHub repository: https://github.com/dogrun-inc/coffee-tds-particle-features.

References

Coffee extraction physics and modeling

[1] K.M. Moroney, W.T. Lee, S.B.G. O'Brien, F. Suijver, J. Marra, Modelling of coffee extraction during brewing using multiscale methods: An experimentally validated model, *Chemical Engineering Science* **137** (2015) 216–234. https://doi.org/10.1016/j.ces.2015.06.003

[2] K.M. Moroney, W.T. Lee, S.B.G. O'Brien, F. Suijver, J. Marra, Asymptotic analysis of the dominant mechanisms in the coffee extraction process, *SIAM Journal on Applied Mathematics* **76** (6) (2016) 2196–2217. https://doi.org/10.1137/15M1036658

[3] J. Melrose, B. Roman-Corrochano, M. Montoya-Guerra, S. Bakalis, Toward a new brewing control chart for the 21st century, *Journal of Agricultural and Food Chemistry* **66** (21) (2018) 5301–5309. https://doi.org/10.1021/acs.jafc.7b04848

[4] M.I. Cameron, D. Morisco, D. Hofstetter, E. Uman, J. Wilkinson, Z.C. Kennedy, S.A. Fontenot, W.T. Lee, C.H. Hendon, J.M. Foster, Systematically improving espresso: Insights from mathematical modeling and experiment, *Matter* **2** (3) (2020) 631–648. https://doi.org/10.1016/j.matt.2019.12.019

Particle size and fines in coffee extraction

[5] N. Battefeld-Montgomery, N. Cocconcelli, J. Keim, M. Perren, C. Yeretzian, S. Smrke, The role of fines in espresso extraction dynamics, *Scientific Reports* **14** (2024) 5612. https://doi.org/10.1038/s41598-024-55831-x

[6] S. Smrke, N. Battefeld-Montgomery, N. Cocconcelli, M. Perren, C. Yeretzian, Extraction of single serve coffee capsules: linking properties of ground coffee to extraction dynamics and cup quality, *Scientific Reports* **10** (2020) 17220. https://doi.org/10.1038/s41598-020-74138-1

Image-based particle size analysis

[7] ISO 13322-1:2014, Particle size analysis — Image analysis methods — Part 1: Static image analysis methods, International Organization for Standardization, Geneva, 2014.

Kernel density estimation

[8] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, London, 1986.

[9] D.W. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization, second ed., Wiley, New York, 2015.

Immersion brewing and TDS

[10] W.J. Lingle, The Coffee Brewing Handbook, Specialty Coffee Association of America, Long Beach, CA, 1996.

[11] L. Navarini, D. Rivetti, Water quality for Espresso coffee, *Food Chemistry* **122** (2) (2010) 424–428. https://doi.org/10.1016/j.foodchem.2009.04.019

Software

[12] C.T. Rueden, J. Schindelin, M.C. Hiner, B.E. DeZonia, A.E. Walter, E.T. Arena, K.W. Eliceiri, ImageJ2: ImageJ for the next generation of scientific image data, *BMC Bioinformatics* **18** (2017) 529. https://doi.org/10.1186/s12859-017-1934-z

[13] P. Virtanen, R. Gommers, T.E. Oliphant, et al., SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nature Methods* **17** (2020) 261–272. https://doi.org/10.1038/s41592-019-0686-2