*Article*

# Automated Procedure for Centre Localization, Noise Removal, and Background Suppression in Two-Dimensional X-ray Diffraction Patterns

**Massimo Ladisa**

Istituto per le Applicazioni del Calcolo Mauro Picone, Consiglio Nazionale delle Ricerche, Via Giovanni Amendola 122/D, 70126 Bari, Italy

**Abstract**

We present a comprehensive and automated methodology for processing two-dimensional X-ray diffraction (2D-XRD) patterns. The proposed workflow involves three sequential stages: (i) precise localization of the diffraction center, (ii) removal of high-frequency noise, and (iii) suppression of non-physical background signals. This method enables improved data quality for subsequent quantitative analysis such as radial integration, phase identification, and structural refinement. Application to experimental datasets from both the Synchrotron Radiation Facility and a table-top X-ray diffractometer demonstrates the method's robustness, accuracy, and computational efficiency.

**Keywords:** X-ray diffraction; image processing; center localization; background suppression; denoising

## 1. Introduction

The use of two-dimensional (2D) area detectors in synchrotron and laboratory-based X-ray diffraction (XRD) experiments has become increasingly widespread, driven by the availability of fast and high-efficiency detectors. Consequently, the volume and rate of data acquisition have increased substantially, particularly in time-resolved studies of dynamic processes such as phase transitions, melting and chemical reactions. Real-time data analysis is therefore essential for interpreting sample behaviour and for adjusting experimental parameters during acquisition.

Despite the availability of established software such as *Fit2D* [1], existing tools often exhibit limitations in processing speed, calibration flexibility and user interaction. *Fit2D*, though long regarded as a standard in the field, suffers from an outdated interface, slow image integration and challenges in calibrating complex detector geometries, particularly when the beam centre lies outside the image or the detector is strongly tilted. These issues can restrict the design and flexibility of XRD experiments, especially under demanding conditions such as high pressure or temperature.

To overcome these limitations, modern software platforms have been developed for rapid, interactive and versatile 2D XRD data processing—such as `Dioptas` [2]—often implemented in Python and based on a modular Model–View–Controller (MVC) architecture. These tools achieve high computational performance while maintaining extensibility and ease of use. Core tasks, including integration, calibration and visualisation, are accelerated through established scientific libraries (e.g. NumPy, SciPy, pyFAI) [3,4], while graphical interfaces enable fast rendering and intuitive data exploration.

While computational speed and automation represent major advances, the development of analytically rigorous and statistically robust routines is equally critical for quantitative accuracy and reproducibility. Rather than focusing solely on optimisation of existing algorithms, the present work introduces new mathematical and statistical methods designed to enhance the reliability of key analytical steps, including centre identification, background estimation and calibration refinement. Such approaches enable automated yet interpretable data evaluation, even in the presence of noise, overlapping reflections or incomplete detector coverage.

Combining high computational efficiency with mathematically grounded analysis establishes a comprehensive framework for accurate and reproducible 2D XRD data processing. The integration of speed, interactivity and statistical rigour enhances both the practical efficiency and scientific reliability of modern diffraction analysis. The present paper focuses primarily on the mathematical aspects rather than on computational or visualisation tasks, and is not intended as a review or promotion of existing software, although comparison with at least one available package is included for validation.

## 2. Centre Localization

### *2.1. User-Supervised Aspects of Centre Localization in the Calibration Procedure*

The process of centre localization constitutes one of the most user-sensitive stages of the calibration procedure. Although the implemented workflow is generally comprehensive and adaptable, it still depends strongly on user supervision at several points. In particular, the initialization of the calibration relies on user-supplied parameters such as approximate beam centre coordinates, detector distance, and orientation. Any inaccuracy in these preliminary entries can propagate through the subsequent steps, thereby compromising the overall calibration quality.

In situations where automatic algorithms fail to converge or misidentify the diffraction centre—often due to low image contrast, incomplete ring patterns, or high noise levels—manual intervention becomes essential. Users are typically required to manually adjust the estimated centre or select specific diffraction rings/peaks for guidance. This introduces a degree of subjectivity and demands a certain level of expertise, since the visual assessment of whether the rings appear concentric or whether *cake lines are straight* serves as the primary qualitative verification method. Such inspection-based validation, while practical, is inherently prone to interpretation errors and lacks quantitative robustness.

Automation, although available, is not fully reliable. The automatic centre-finding routines may require user supervision, such as toggling between single-ring and multi-ring/peak detection or refining the selection of regions of interest. The sensitivity of the localization process to data quality further limits its autonomy: when dealing with weak or incomplete diffraction patterns, users often need to adjust thresholds or apply spatial masks to constrain the fit. The effectiveness of these manual corrections, however, depends on the user's experience and understanding of the image features.

The absence of diagnostic feedback mechanisms compounds these challenges. When centre localization fails to yield consistent results, the current approach generally resorts to a trial-and-error exploration of parameter combinations, which can be time-consuming and inefficient. Moreover, the system does not provide clear indications of the cause of failure—whether stemming from poor data quality, misestimated input parameters, or algorithmic limitations—thus placing the diagnostic burden entirely on the user.

Finally, reproducibility remains an issue. Since manual centre adjustments and ring selections vary from one operator to another, the resulting calibrations may differ slightly across users. Additionally, the lack of automated logging means that user actions and parameter settings during the centre localization step are not systematically recorded, hindering the ability to reproduce or audit previous calibration sessions. Consequently, while the procedure enables flexible and accurate centre determination, it still relies heavily on informed user participation to ensure both reliability and reproducibility.

For instance, the results of applying the same `Dioptas` unsupervised calibration routine to identical diffraction datasets are shown in Figure 1. These include patterns collected at the Synchrotron Radiation Facility (*S04387*, top row), which are affected by experimental setup artefacts and exhibit additional noise, shadows, and generally poor data quality, as well as patterns acquired using a table-top X-ray diffractometer (*152762*, bottom row), characterised by low counting statistics. The four panels illustrate how the routine may yield different ring selections (red markers), leading to variability in centre localisation and profile extraction, whereas user-supervised intervention (single-peak search) ensures consistent ring detection.

*2.2. Unsupervised Centre Localization*

Unlike several existing approaches—including radial symmetry analysis and ring fitting [5]—that optimise the beam centre with subpixel accuracy, the `sniper2D` function developed in this study provides a novel, unsupervised algorithm for centre localisation in two-dimensional images, such as X-ray diffraction (XRD) patterns, based on correlation analysis. The algorithm identifies the horizontal and vertical center of an image by correlating it with its shifted versions and selecting the shifts that result in the highest correlation. The result is a set of coordinates [xc, yc] that represent the center of the image. Here is a breakdown of what the algorithm does (see the supplementary material for the full algorithm):

- 1. Image Preparation: imgLR (resp. imgUD) is the horizontally (resp. vertically) flipped version of the image.
- 2. Variable Initialization: ny, nx store the height and width of the input image. cx, cy are initialized to zero and will store the maximum correlation values for the x and y directions. sx, sy are initialized to zero and will be used to iterate through shifts in the x and y directions.
- 3. Horizontal Search (X-axis): the algorithm performs a search for the center of the image along the x-axis using sx, which is incremented starting from 0. It creates two shifted versions of the horizontally flipped image (auxp and auxm) using positive and negative shifts by sx. It computes the correlation between the original image and these shifted versions (auxp, auxm), and updates the center xc and the maximum correlation value cx if the correlation exceeds previous values. This process continues until sx exceeds nx/8.
- 4. Vertical Search (Y-axis): the same process is repeated for the y-axis, where sy iterates over vertical shifts. It computes auxp and auxm for the vertically flipped image (imgUD) and performs the same correlation comparison to find the vertical center yc and the maximum correlation value cy.
- 5. Output: the final output is the coordinates [xc, yc], representing the center of the image or the location of the peak feature based on correlation.

In the analysis of X-ray diffraction (XRD) two-dimensional (2D) patterns, enhancing the information content of the measured intensity distribution can significantly improve the accuracy of centre localisation. To this end, we apply a theorem demonstrating that a scale-free, uncorrelated modulation of a normalized positive function can reduce its entropy. In practice, this principle is employed to preprocess an XRD 2D pattern prior to applying the centre localization algorithm (`sniper2D`). By selectively concentrating probability mass, the modulation increases structural contrast and lowers the pattern's entropy, thereby enabling more accurate and robust localisation of the diffraction centre.

The modulation is realised through a scale-free function defined as

$$\varphi(x,y) = \mathrm{Arg}\big(\partial_x I + i\,\partial_y I\big),$$

which depends solely on the local angular variations of the intensity field $I(x,y)$ and is independent of its absolute scaling. Although $\varphi$ is ideally[1] uncorrelated with $I(x,y)$, their interaction through the modulated field

$$f_s(x,y) = \frac{I(x,y)}{s}\,\varphi(x,y),$$

where $s > 0$ denotes a scaling factor, redistributes the probability density in a manner that selectively emphasises spatial features associated with meaningful structural variations. This process effectively compresses redundant information while amplifying directional gradients that encode the intrinsic geometry of the diffraction pattern.

[1] In a radially symmetric two-dimensional pattern, $\varphi(x,y) \times I(x,y) = \mathrm{Arg}[(x-xc)+i(y-yc)] \times I\Big(\sqrt{(x-xc)^2+(y-yc)^2}\Big) = \vartheta \times I(r)$, and therefore it factors out during integration over the domain $dx\,dy = r\,dr\,d\vartheta$, indicating statistical independence.

The entropy reduction ensured by the theorem is therefore not merely a mathematical artefact of normalization but reflects a genuine improvement in the organisation of information. For diffraction analysis, this results in a clearer and more interpretable 2D pattern from which the `sniper2D` algorithm can extract the diffraction centre with greater precision and stability. The scale-free nature of $\varphi$ ensures that this enhancement is robust to intensity scaling, making the approach particularly suitable for experimental data characterised by broad dynamic ranges or nonuniform exposure.
For completeness, a formal proof of the theorem is outlined in the Appendix .1.
The results are shown in Figure 2, where the centres estimated by the `sniper2D` algorithm using intensity or intensity–phase data are juxtaposed with the corresponding 2D diffraction patterns. These include patterns collected at the Synchrotron Radiation Facility (*S04388*, affected by experimental setup artefacts, and *S04387*, exhibiting additional noise, shadows, and poor data quality) and at a table-top X-ray diffractometer (*152762*, with low counting statistics, and *resummed*, with higher counting statistics). Quantitative metrics for the diffraction ring analysis and ring centres compared with `Dioptas`, are reported in Table 1 for a scaling factor $s = 1$.

## 3. Denoising

A recent study [6] introduced the Lattice Boltzmann Method (LBM) [7–9] as an innovative approach to enhance the quality of two-dimensional X-ray diffraction (XRD) patterns. XRD is a key technique in materials science and crystallography for determining atomic structures, yet its effectiveness is often limited by the presence of noise that obscures subtle but critical features. Traditional denoising techniques typically struggle to preserve the fine details of diffraction data, motivating the use of alternative numerical models. In this context, LBM, originally developed for simulating fluid dynamics on discrete lattices, is adapted to model diffusion processes in XRD images by treating the signal and noise as distinct interacting components. Through the simulation of diffusion dynamics, the method efficiently separates noise from the true diffraction signal, producing cleaner and more accurate data. Tests reported in the original work demonstrated substantial improvements in signal-to-noise ratio and robustness against Poissonian noise, while maintaining the fidelity of structural features. Owing to its numerical stability, efficiency, and feature-preserving properties, LBM represents a valuable tool for improving XRD data interpretation. In the present work, this method is employed with the same parameter settings described in [6] to process comparable two-dimensional diffraction patterns.
The algorithm implements a two-dimensional nine-velocity (D2Q9) Lattice Boltzmann scheme for denoising two-dimensional patterns, following the Bhatnagar–Gross–Krook (BGK) approximation [10]. The parameters $\alpha = 0.85$, $\beta = 1.25$, and $\Omega = \alpha$ define the relaxation rate ($\Omega = 1/\tau_c$), while the discrete velocity set $(c_x, c_y)$ and corresponding weights $W = [16, 4, 1, 4, 1, 4, 1, 4, 1]/36$ define the D2Q9 lattice.

The equilibrium distribution $N_{eq}$ is initialized from $\log(I)$ (being $I$ the original image[2]) and weighted by $W$, while $N$ is randomly initialized. At each time step, the *streaming* step propagates $N_i$ along $(c_x(i), c_y(i))$ using periodic boundary conditions. The macroscopic fields are updated as $\rho = \sum_i N_i$ and $(u_x, u_y)$ from the equilibrium moments. A discretized Maxwell–Boltzmann equilibrium function, denoted $I_{eq}$, is constructed as

$$I_{eq} = W\left[1 + 3(\mathbf{u}\cdot\mathbf{c}) + 4.5(\mathbf{u}\cdot\mathbf{c})^2 - 1.5|\mathbf{u}|^2\right]\log(I).$$

In the *collision* step, the algorithm computes local gradients $(N_x, N_y)$ and a scale parameter $\Lambda_i = 0.9\,\langle\sqrt{N_x^2 + N_y^2}\rangle$, then derives adaptive weights $w_i = [1 + (N_x^2 + N_y^2)/\Lambda_i^2]^{-1}$. The weighted divergence term $\nabla\cdot(w\nabla N)$ is used to relax the system:

$$N \leftarrow N + \Omega\left[\frac{\nabla\cdot(w\nabla N)}{\beta} + (I_{eq} - N)\right].$$

[2] The addition of one count to each bin does not introduce any statistical artefact; rather, it remains fully compatible with maximum-likelihood estimation. A detailed discussion of this procedure and its justification can be found in [11].

After $t_f$ iterations, the reconstructed (denoised) density field is computed as

$$I^{den} = \exp(\rho).$$

This process iteratively separates signal from noise while preserving structural details in the input pattern.
A schematic outline of the implemented algorithm is provided in the Appendix .2 for reference.
For the denoising procedure, a direct comparison with an alternative method has been omitted, as the original manuscript [6] provides a detailed analysis of this aspect. Interested readers are referred to this work for further details.

## 4. Background Suppression

### *4.1. User-Dependent Aspects in Background Subtraction within the Integration Procedure*

The background subtraction stage of the integration procedure remains highly user-dependent. Sternberg's 1983 work [12] introduced the rolling-ball algorithm, modelling images as three-dimensional intensity surfaces and estimating the background via a simulated rolling sphere. This elegant and robust method became foundational in microscopy and biomedical imaging, and is widely adopted in software such as ImageJ [13,14].

Although these methods offer flexibility, they require substantial manual tuning of parameters such as smoothing width, iteration number, and polynomial order, guided primarily by visual inspection and prior experience. This introduces variability and limits quantitative assessment, as objective measures of fit quality are generally absent.

Selection of the region of interest for fitting adds further user dependence. Inaccurate choices can include noise or omit relevant intensity features, distorting the integrated profile. Consequently, visual comparison between original and background-subtracted patterns remains the principal, though subjective, validation approach, especially in noisy or complex diffraction data.

Algorithm parameters, such as the interplay between smoothing width and polynomial order, are often opaque, risking over- or underfitting. The absence of warnings or safeguards allows distortions or unrealistic background shapes to go unnoticed. Moreover, parameter values are not automatically logged, hindering reproducibility; different users may obtain divergent results from the same dataset.

The integration stage, including tools such as supersampling, also demands caution. Supersampling can enhance resolution and smoothness but may generate artefacts if misapplied, and no quantitative diagnostics or automated checks exist. The lack of preview or undo functionality further limits reversibility.

In summary, background subtraction within the integration workflow remains a user-sensitive process requiring careful parameter tuning, informed judgment, and systematic record-keeping. Enhanced guidance, quantitative feedback, and automated safeguards would improve robustness and reproducibility.

### *4.2. Data-driven Background Estimate*

In the current work, we begin by considering an intensity measurement $I(\vartheta, r)$ that is composed of two unknown or *blind* contributions: a slowly varying *background* $b(\vartheta, r)$ and a localized *signal* $s(\vartheta, r)$. In other words, the observed intensity at each point is the sum of these two components, whose individual shapes are not known a priori.

Now consider the perturbed intensity product

$$\mathrm{P}(\vartheta, r) = \prod_{i=1}^{n} \left[ b(\vartheta, r + \epsilon_i) + s(\vartheta, r + \epsilon_i) \right],$$

where the $\epsilon_i$ represent small random shifts along the radial coordinate. Physically, this models multiple measurements in which the underlying background is nearly constant, while the localized signal features are narrow and randomly displaced.

Because the background changes only slowly with $r$, it can be treated as essentially constant over each perturbation. The localized signals are sparse and do not significantly overlap, so their contributions remain limited. As a result, when taking the product over many measurements, the relative influence of the signals diminishes compared to the background.

Mathematically, factoring out the background from the product and bounding the residual signal terms yields

$$b(\vartheta, r) \leq \sqrt[n]{\mathrm{P}(\vartheta, r)} \leq b(\vartheta, r) \cdot n^{1/n} \cdot \left[\frac{1}{n} + \frac{\max(s(\vartheta, \cdot))}{b(\vartheta, r)}\right]^{1/n}.$$

As the number of measurements $n$ increases, both $n^{1/n}$ and the term involving the maximum signal contribution approach 1, leading to the convergence

$$\lim_{n\to\infty} \sqrt[n]{\mathrm{P}(\vartheta, r)} = b(\vartheta, r).$$

In practical terms, this shows that taking the geometric mean of multiple perturbed intensity measurements effectively isolates the background, *washing out* the localized signals and revealing the slowly varying component as the dominant contribution.
For completeness, a formal proof of the theorem is outlined in the Appendix .3.
The results are shown in Figures 3 and 4, where the 2D backgrounds estimated by the present algorithm are compared with the corresponding 2D patterns computed by `ImageJ` for three choices of the rolling-ball radius (25/50/100 pixels). These include a pattern collected at the Synchrotron Radiation Facility (*S04387*, affected by experimental setup artefacts and exhibiting additional noise, shadows, and poor data quality), and another one collected at a table-top X-ray diffractometer (*sam4*). A comparison among the corresponding 1D semilogarithmic profiles is shown in Figure 5.
Quantitative metrics for the comparison of the same samples with `ImageJ` are reported in Table 2. The apparent discrepancy between 2D Pearson's correlation coefficient and the 2D Structural Similarity Index (SSIM) [15] for the *S04387* sample backgrounds is explained as follows. A high correlation reflects a strong linear relationship but does not guarantee a high SSIM, which also takes into account perceptual factors such as luminance, contrast, and texture. Thus, the patterns may share similar global intensity variations but differ in local structure, leading to high correlation and lower SSIM. This highlights the need to consider both metrics in image analysis.

The metrics reported in Table 2 indicate a generally strong agreement between the background estimations produced in the present work and those obtained using *ImageJ* across a range of rolling–ball radii. For sample *S04387*, the background constitutes only a small fraction of the overall signal (3%), and the correlation with the *ImageJ* backgrounds remains consistently high (approximately 0.9) regardless of radius. However, the SSIM values are more modest (around 0.55), suggesting that although the global trends are well captured, finer structural details differ to a greater extent. The residuals are correspondingly small, remaining close to 1% of the total intensity.
In contrast, sample *sam4* exhibits a substantially larger background contribution (59%), and here both the correlation and SSIM values indicate an excellent match to the *ImageJ* estimates, particularly for radii of 50–100 pixels (Corr $\approx$ 0.96; SSIM $\approx$ 0.91). The residuals decrease markedly with increasing radius, falling from 0.25 at 25 pixels to around 0.10 at 100 pixels, implying that larger radii provide a closer approximation to the background retrieved by the present method. Collectively, these results demonstrate that the proposed background–recovery approach performs robustly and is broadly consistent with established methods, with the level of agreement improving for samples with more substantial background structure.

## 5. Results

To evaluate the robustness of the proposed denoising approach under realistic experimental conditions, the model was applied to X-ray diffraction (XRD) patterns obtained from tendon-derived collagen molecules (*rat tail* in this paper). Under varying biochemical environments, these molecules self-assemble into a hierarchical superstructure of triple helices, exhibiting a pronounced preferred orientation. This orientational order is evident in the $\pi$-symmetric partial arcs observed in Figures 6, 7-left, 8-top-left, which replace the fully $2\pi$-symmetric diffraction rings typically associated with isotropic samples. Such superstructural organization, together with the high degree of crystallinity within the fibrillar domains, contributes significantly to the mechanical stiffness of the tissue (for details of sample preparation and experimental procedures, see [16]). Figure 6 compares the corresponding XRD patterns: the original dataset (top left panel) was acquired over an integration time of 2400 s, yielding a maximum count below 1000.

For the sake of completeness, two additional figures are presented in Figure 9: 1D semilogarithmic profiles obtained using the present approach (right) and 2D original maps (left) for the *rat tail* sample, acquired at varying exposure times: 171568 s (centre) and 327350 s (bottom), yielding a maximum count around 60 000, along with the analogous profiles obtained using `Dioptas` for reference.

Small-angle and wide-angle X-ray scattering (SAXS/WAXS) measurements were performed at the X-ray Micro Imaging Laboratory (XMI-LAB [17]) on both raw collagen flakes and processed collagen films. The experimental setup comprised a Fr-E+ SuperBright rotating Cu anode microsource ($\lambda = 0.154$ nm) equipped with Confocal Max-Flux optics and a SAXS/WAXS three-pinhole camera. WAXS data were recorded on a 250 mm × 160 mm image plate detector (pixel size 100 µm) and read using an off-line RAXIA system. Each sample was mounted on a fixed holder, and data were collected from three distinct positions. The incident beam spot size was approximately 200 µm, with a sample-to-detector distance of 10 cm, providing a scattering vector ($q$) range of 3 $nm^{-1}$ to 35 $nm^{-1}$, corresponding to $d$-spacings of 0.18 nm to 2.5 nm.

The X-ray diffraction images *152762* and *resummed* in Figure 2 (bottom panels) correspond to vesicle-in-gel samples also measured on the diffractometer at the XMI-LAB [18]: the diffraction pattern *152762*, with an acquisition time of 45 seconds, exhibits a ring of very low intensity in Figure 2 (bottom left), which is nonetheless discernible in the 1D radial integration plot, where a peak appears at approximately $q = 0.1$ $nm^{-1}$. The pattern obtained by summing all acquisitions of the sample clearly highlights the presence of the aforementioned ring (*resummed*) in Figure 2 (bottom right).

Another example of wide-angle X-ray diffraction (WAXD) fibre-pattern denoising is provided by the linen samples characterized in [19]. The specimens (*sam4* in this paper) consisted of threads approximately 1 cm in length and 0.2 mm to 0.6 mm in width. WAXS measurements were performed on all linen samples at the X-ray Micro Imaging Laboratory (XMI-LAB) using an experimental setup similar to that described above, equipped additionally with an X-ray scanning microscope. Data were collected on an image plate (IP) detector and read off-line using a Rigaku RAXIA-Di system. Each linen thread, mounted directly on the sample holder, was exposed for 1800 s. The incident beam spot size at the sample position was approximately 200 µm, and the image plate detector was positioned at a distance of 10 cm. This geometry provided access to scattering vector magnitudes defined as

$$q = \frac{4\pi \sin\theta}{\lambda},$$

where $\theta$ is half the scattering angle, *i.e.*, the angle between the incoming beam and the detector direction and $\lambda$ is the wavelength of the incident radiation. It covers a range from approximately 1.5 $nm^{-1}$ to 35 $nm^{-1}$. Following geometric calibration, the two-dimensional (2D) diffraction images were azimuthally integrated to yield one-dimensional (1D) intensity profiles. Further details of the experimental configuration and data analysis procedures can be found in the original publication. The outcome of the image processing algorithm is presented in Figures 10, 7-right, 8-top-right.

The algorithm scheme can also be fruitfully applied to the XRD patterns collected at the Synchrotron Radiation Facilities (namely *S04387* and *S04388* patterns in Figure 2-top). The Pilatus II is a single photon counting pixel detector, *e.g.* used at the PSI Facility [20], renowned for its high spatial resolution and efficiency in detecting low-intensity X-ray beams. It features an advanced hybrid pixel array that counts individual photons, offering exceptional precision and dynamic range. This detector is pivotal for experiments requiring detailed imaging and analysis, such as those in structural biology and materials science. Its performance is enhanced by its low noise and high count rate capabilities, making it a versatile tool for cutting-edge research in various scientific fields. However, low statistics two-dimensional images, sometimes affected by artefacts (*e.g.* experimental setup shadows), could benefit from a pre-processing denoising: in Figure 2-top the application of the `sniper2D` algorithm is illustrated in such a case, while the result of the entire image processing is shown in Figures 11, 12 and 8 (bottom panels).

These issues have collectively met the need for better automation, real-time validation, and user guidance to reduce the burden and risk of incorrect user supervision. The application of the `sniper2D` algorithm to all aforementioned 2D XRD patterns is summarised in Table 1, which also provides a numerical comparison with the results obtained using the `Dioptas` software. The corresponding one-dimensional semilogarithmic profiles are presented in Figure 8.

The metrics summarised in Table 1 provide an overview of the information content and angular–intensity relationships present in the diffraction images, together with a comparison of ring centre positions determined in the present work against those obtained using *Dioptas*. The intensity entropy values span a broad range, reflecting differing levels of structural complexity across the samples: *sam4* and the long-exposure *rat-tail* datasets exhibit relatively high entropy (around 0.30–0.35), whereas sample *152762* shows very low entropy (0.01), consistent with a markedly simpler intensity field.
A key observation is that the entropy computed on the joint distribution $I \times \varphi$ is systematically lower than the entropy of the intensity pattern alone for every sample. This reduction is expected, as conditioning on the angular variable restricts the available variability within the data, thereby lowering the overall information content of the resulting distribution. The consistent decrease across all entries supports the interpretation that the angular dependence introduces additional structure rather than additional randomness.
The correlations between the intensity gradients and the azimuthal angle $\varphi$ are generally weak and hover close to zero, indicating that directional variations in intensity are not strongly aligned with the angular coordinate. This behaviour is compatible with the moderate joint entropy values, which reflect increased structural organisation in more complex patterns.
Finally, the comparison of ring centres reveals good agreement between the two methods: for most samples, the centre coordinates determined here differ from the *Dioptas* values by only a few pixels. The *rat-tail* datasets show the largest offsets (up to roughly 10 pixels), whereas samples such as *S04387*, *sam4*, and *152762* exhibit close numerical correspondence. Overall, these results demonstrate that the present analysis method provides centre estimates consistent with established software while offering additional quantitative insight through entropy-based characterisation of diffraction-pattern structure.

## 6. Conclusions

The calibration system presented in this work minimises reliance on user expertise, judgement, and trial-and-error, reducing inconsistency and inefficiency. By automating common failure recovery, providing real-time quality metrics, and offering clearer guidance, the workflow requires less direct supervision while maintaining high accuracy and computational efficiency. Application to both synchrotron and laboratory XRD datasets demonstrates its robustness and versatility, facilitating reliable quantitative analysis. This automated methodology has the potential to streamline routine 2D XRD processing and improve reproducibility in structural characterisation studies.

## 7. Figures, Tables

**Figure 1.** Results of applying the same `Dioptas` unsupervised calibration routine to identical diffraction datasets (*S04387*, top row, and *152762*, bottom row). The four panels illustrate how the routine can produce different ring selections (red markers), leading to variability in center localization and profile extraction, whereas user-supervised intervention (single peak search) ensures consistent ring detection.

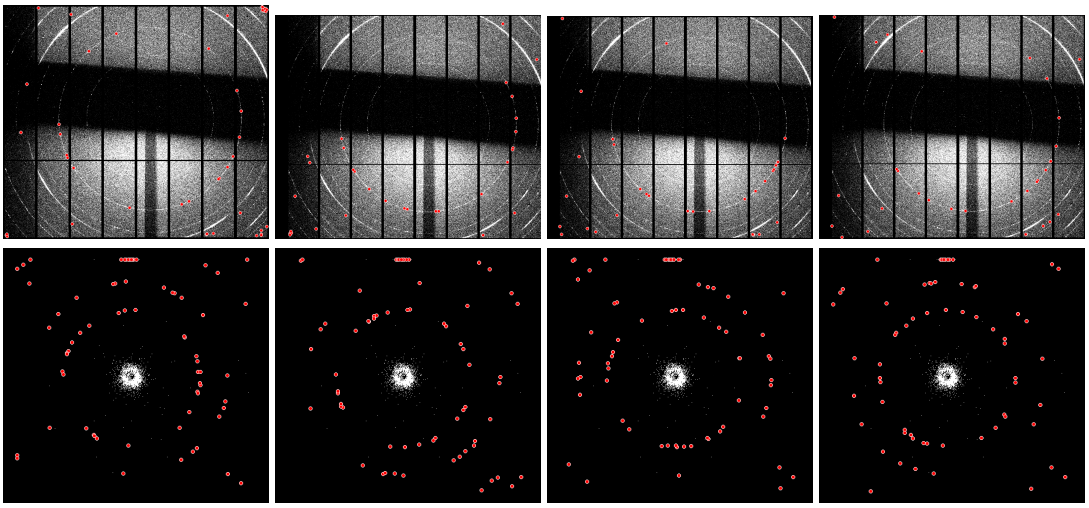

**Figure 2.** 2D pattern obtained for the *S04388* (top left), *S04387* (top right) samples, both collected at the Synchrotron Radiation Facility, and *152762* (bottom left), *resummed* (bottom right) samples (together with centre closeup), collected at a table-top X-ray diffractometer. The red "×" symbol indicates the centre estimated by the (unsupervised) `sniper2D` algorithm when only the 2D intensity pattern is used as input, whereas the blue "+" symbol shows the centre obtained by applying the same algorithm to the combined 2D intensity–phase pattern.

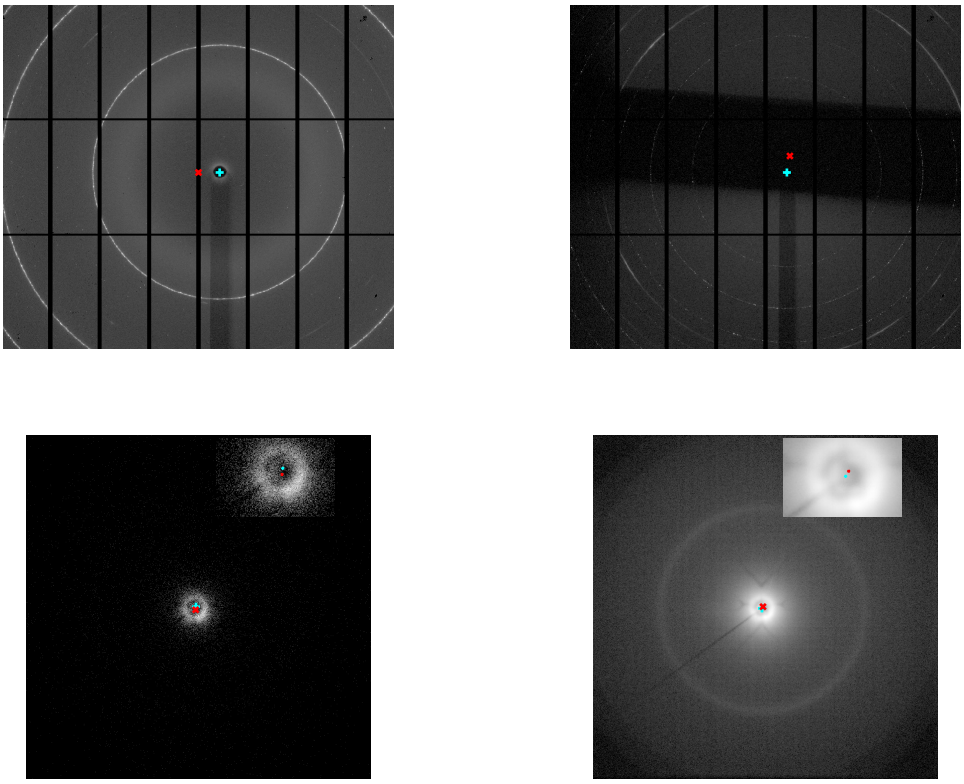

**Figure 3.** 1D profiles and 2D conformal maps for the *S04387* sample background: (top left) this work, (top right) `ImageJ` with a 25-pixel ball radius, (bottom left) `ImageJ` with a 50-pixel ball radius, and (bottom right) `ImageJ` with a 100-pixel ball radius.

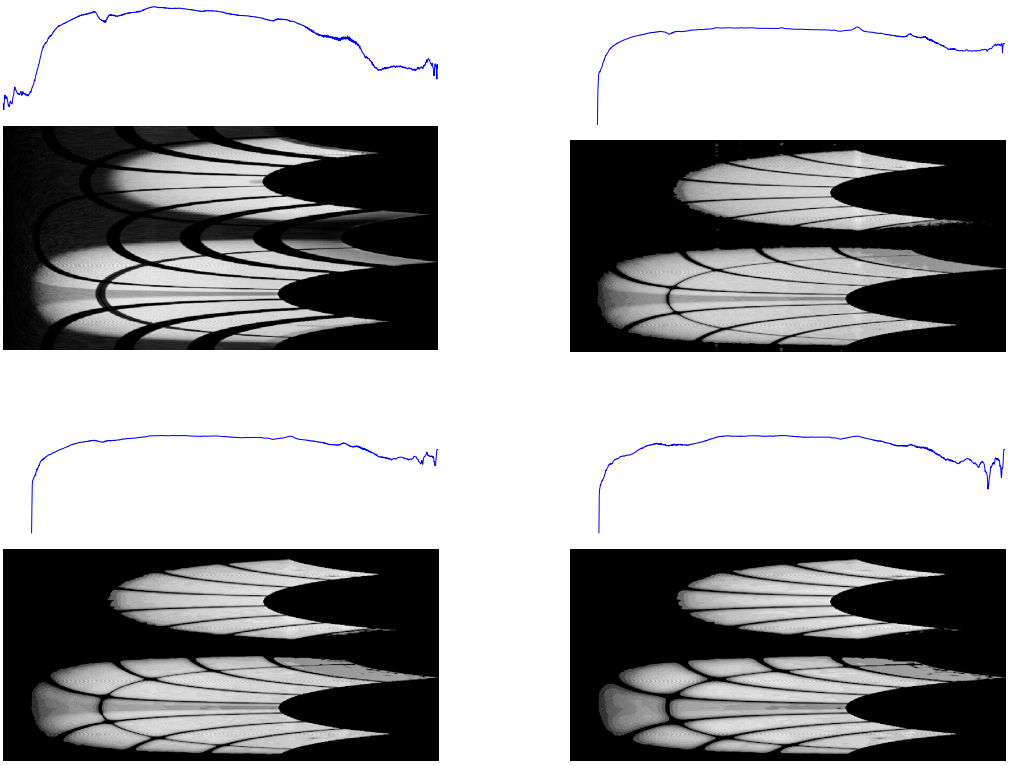

**Figure 4.** 1D profiles and 2D conformal maps for the *sam4* sample background: (top left) this work, (top right) `ImageJ` with a 25-pixel ball radius, (bottom left) `ImageJ` with a 50-pixel ball radius, and (bottom right) `ImageJ` with a 100-pixel ball radius.

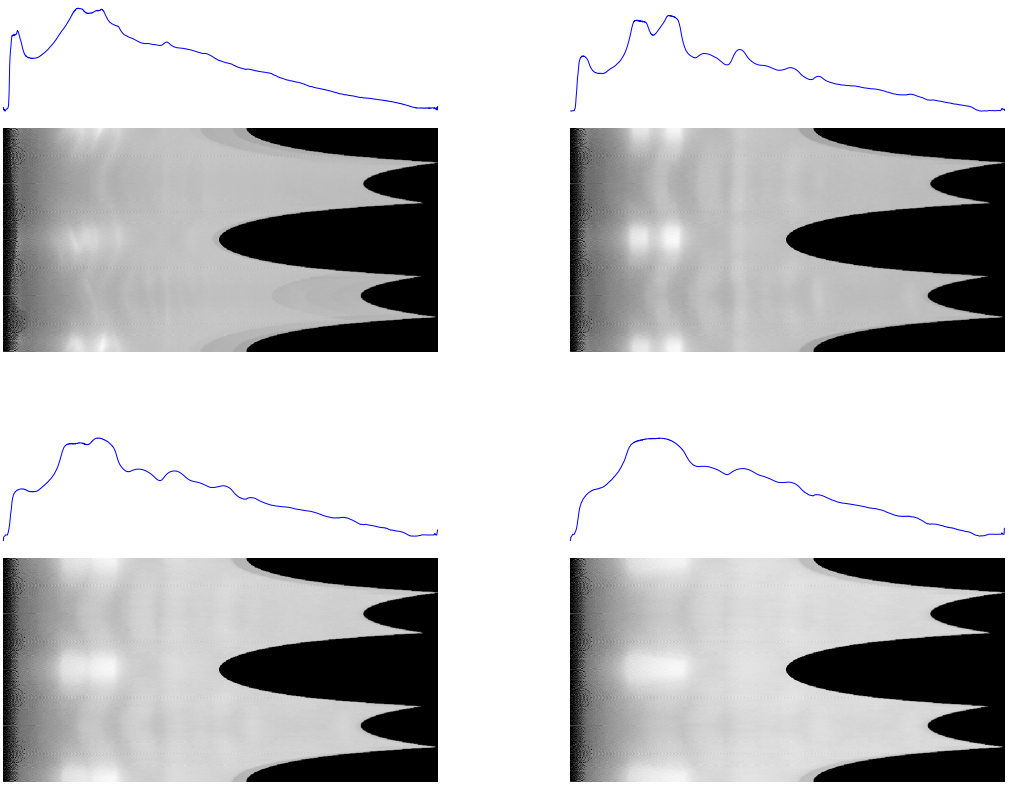

**Figure 5.** Semilogarithmic profiles obtained for the *sam4* (left) and *S04387* (right) sample backgrounds.

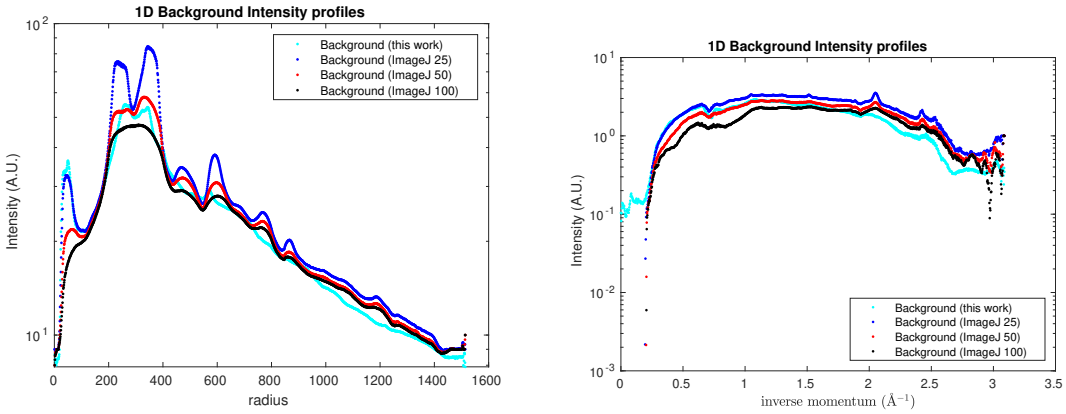

**Figure 6.** 1D profiles and 2D conformal maps for the *rat tail* sample: unprocessed (top left), denoised (top right), background (bottom left), denoised and background subtracted (bottom right).

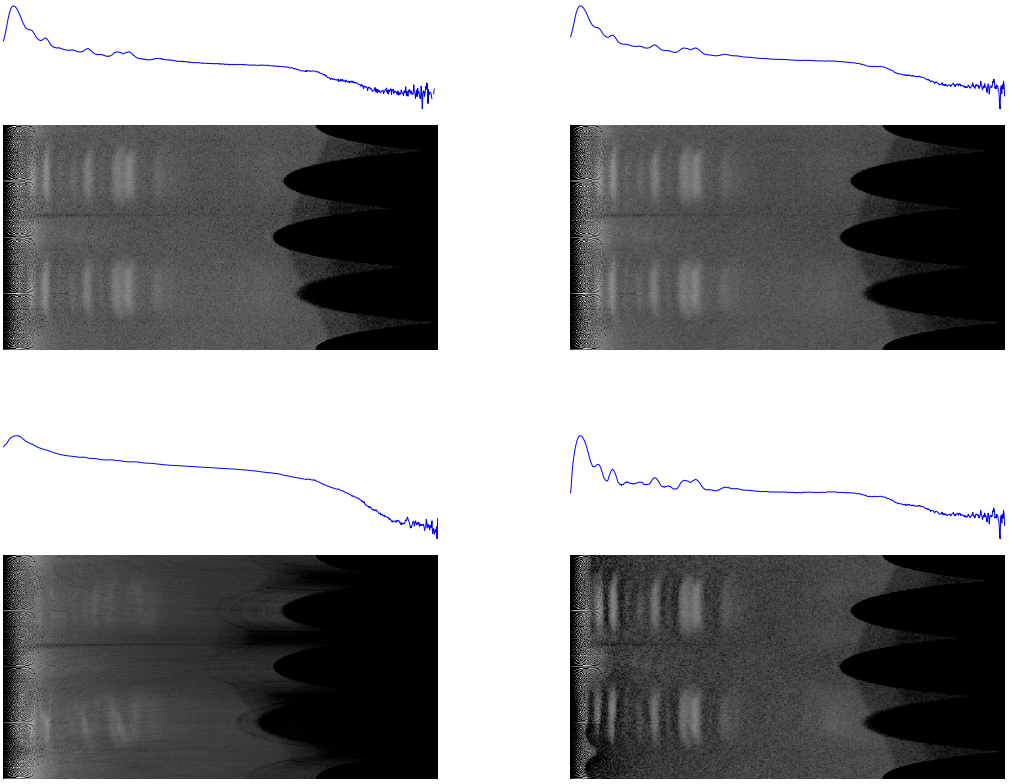

**Figure 7.** 1D semilogarithmic profiles and 2D original maps obtained for the *rat tail* (left) and *sam4* (right) samples.

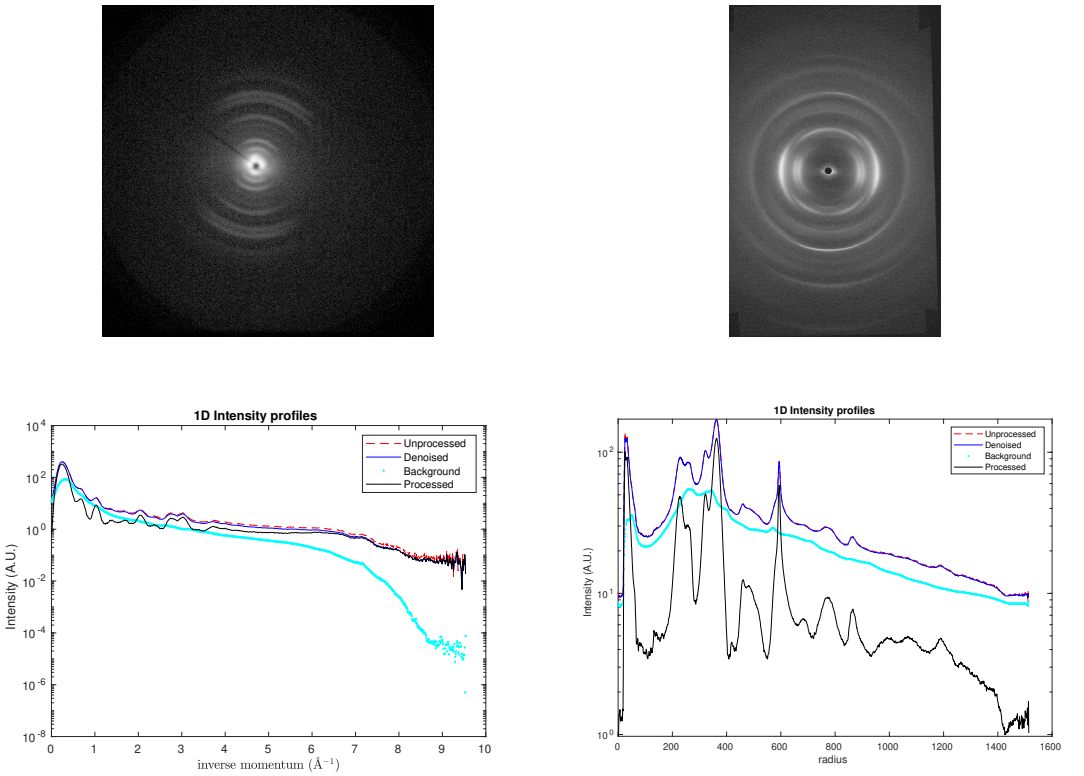


**Figure 8.** Semilogarithmic intensity profiles for the *rat tail* (top left), *sam4* (top right), *S04387* (bottom left), and *S04388* (bottom right) samples. For reference, analogous profiles obtained with `Dioptas` are also included.

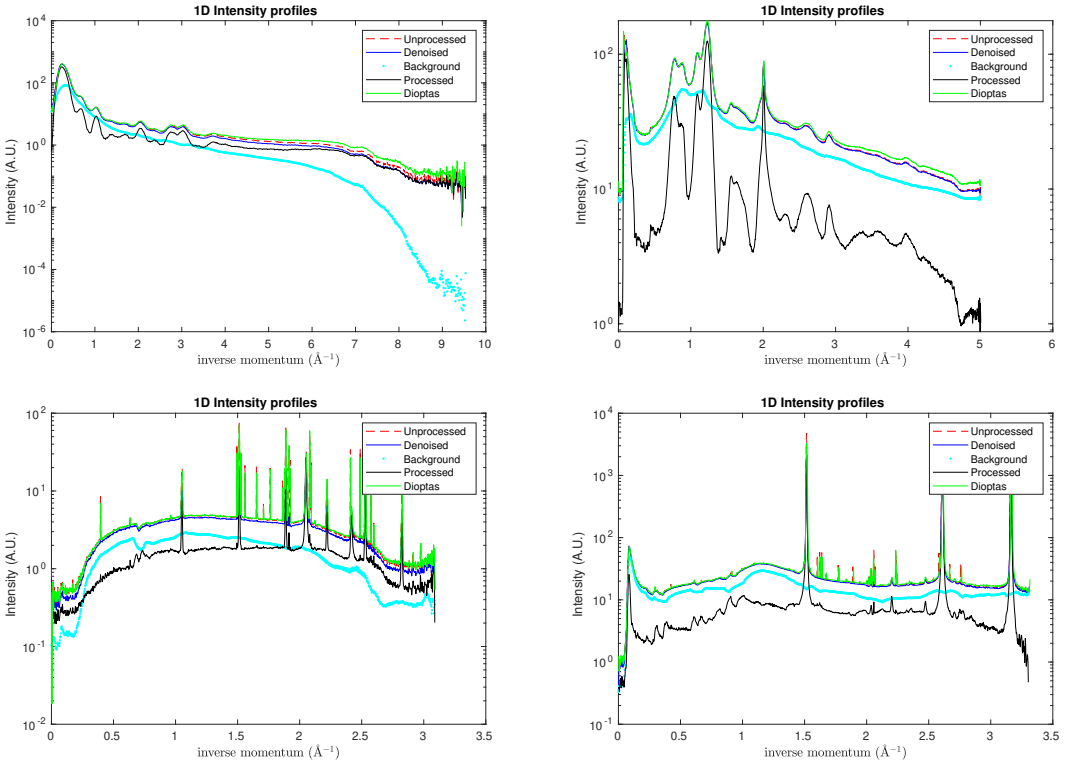

**Figure 9.** 1D semilogarithmic profiles (right) and 2D original maps (left) obtained for the *rat tail* sample at different exposure times: 2400 s (top), 171568 s (centre), and 327350 s (bottom). For reference, analogous profiles obtained with `Dioptas` are also included.

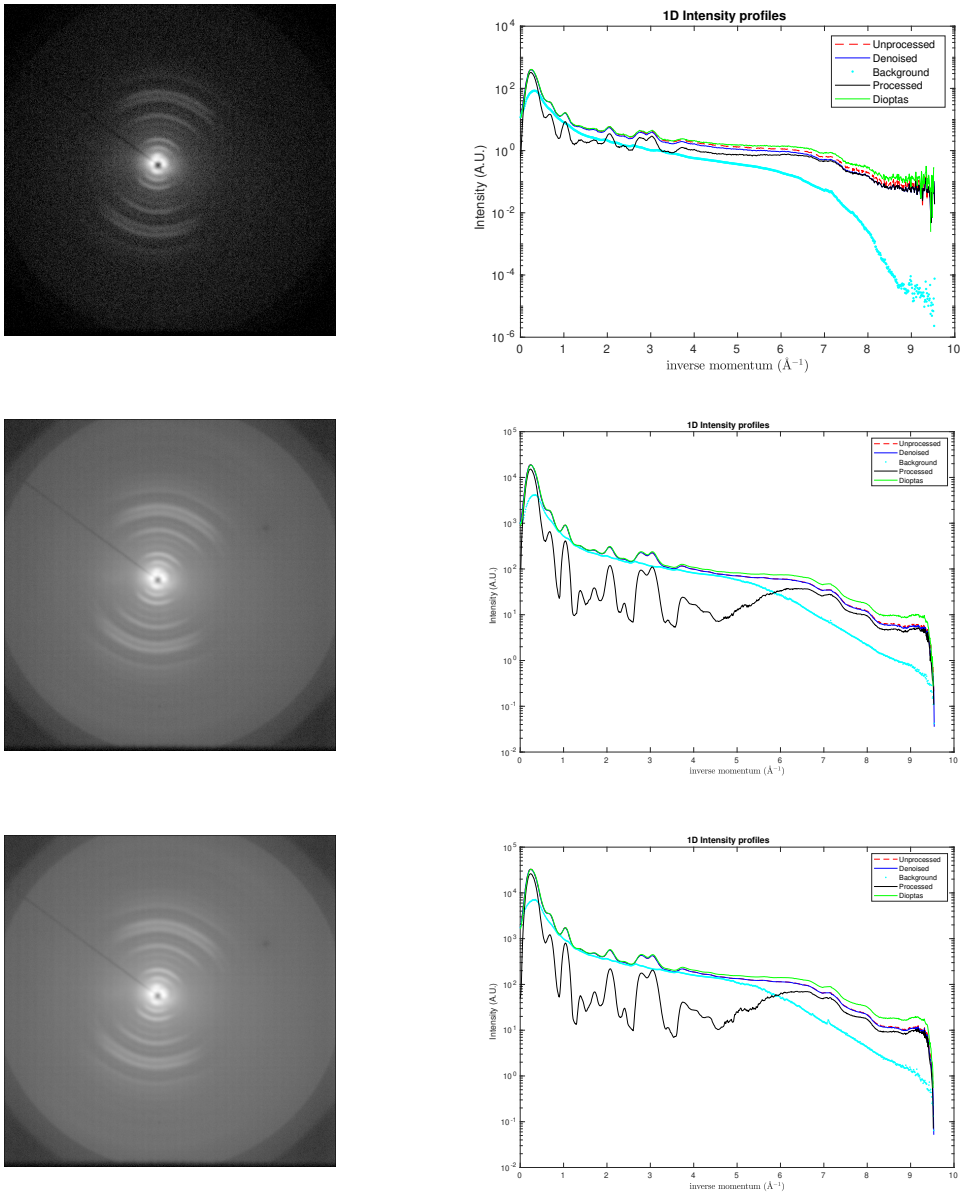

**Figure 10.** 1D profiles and 2D conformal maps for the *sam4* sample: unprocessed (top left), denoised (top right), background (bottom left), denoised and background subtracted (bottom right).

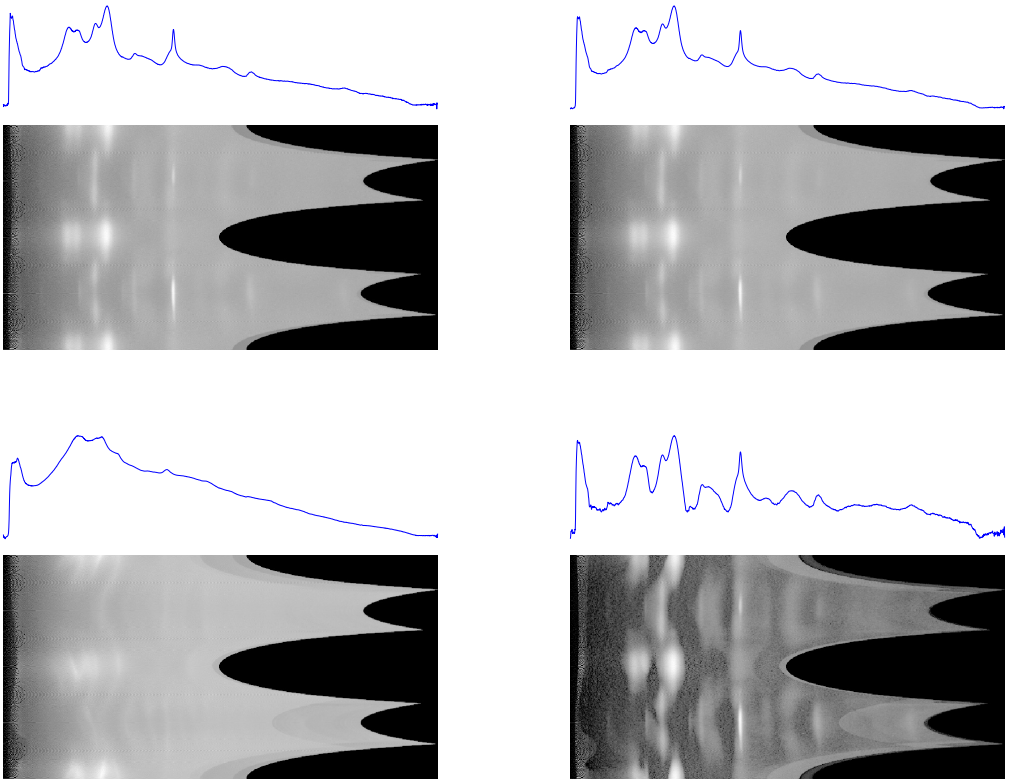

**Figure 11.** 1D profiles and 2D conformal maps for the *S04387* sample: unprocessed (top left), denoised (top right), background (bottom left), denoised and background subtracted (bottom right).

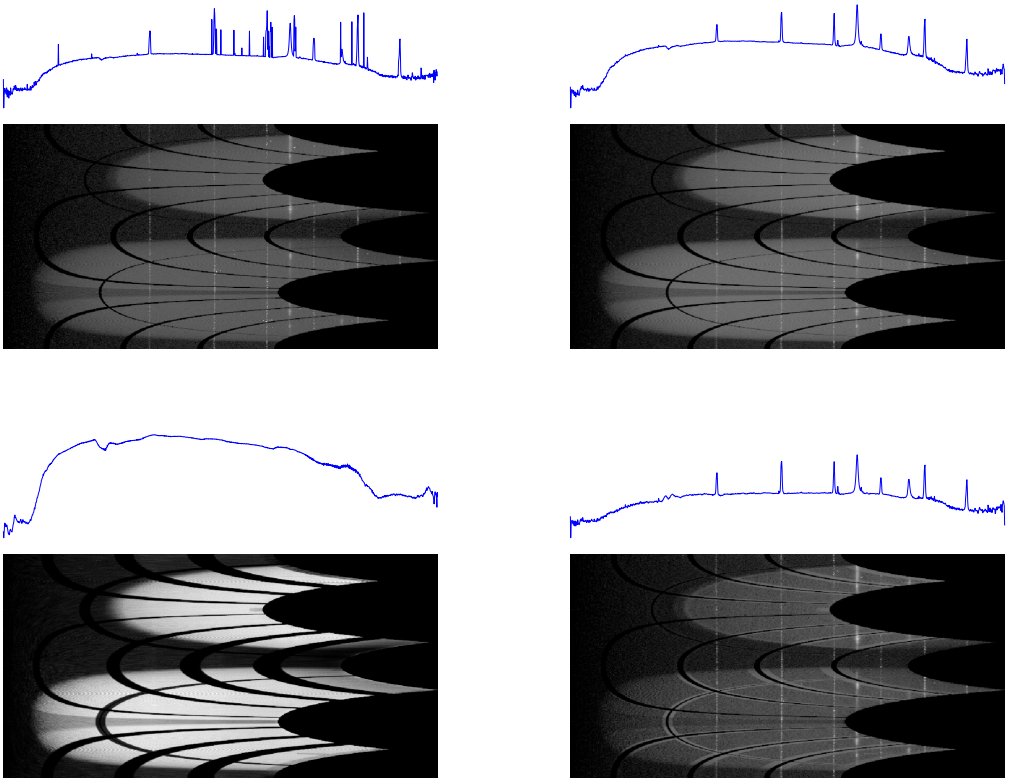

**Figure 12.** 1D profiles and 2D conformal maps for the *S04388* sample: unprocessed (top left), denoised (top right), background (bottom left), denoised and background subtracted (bottom right).

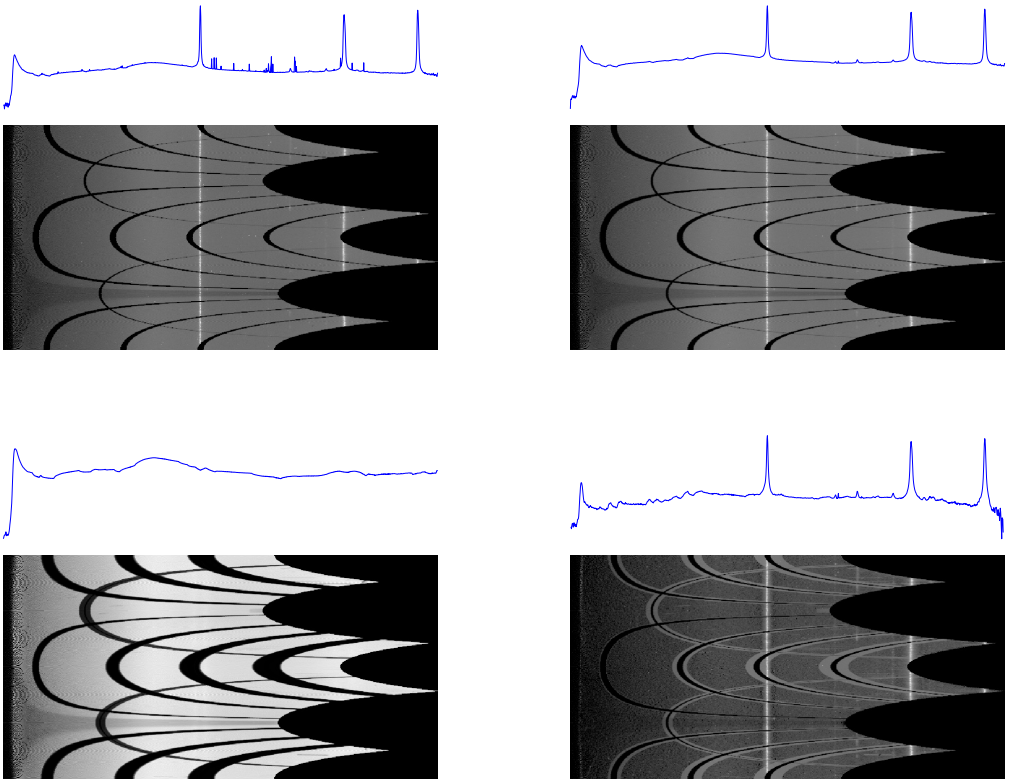

**Table 1.** Metrics for diffraction ring analysis: intensity entropy Ent(I), correlation with $\varphi$ ($= \mathrm{Arg}(\partial_x I + i\,\partial_y I)$), joint entropy Ent($I \times \varphi$) and ring centers determined by this work versus `Dioptas`.

| Sample | exp. time (s) | Ent($I$) | $I$–$\varphi$ corr. | Ent($I \times \varphi$) | Centre (this work) | Centre (`Dioptas`) |
|---|---|---|---|---|---|---|
| *S04387* | - | 0.20 | 0.113 | 0.14 | (932,717) | (931,717) |
| *S04388* | - | 0.33 | 0.017 | 0.24 | (931,717) | (935,720) |
| *rat-tail* | 2400 | 0.19 | -0.006 | 0.14 | (475,494) | (486,497) |
| *rat-tail* | 171568 | 0.35 | -0.005 | 0.28 | (474,493) | (486,497) |
| *rat-tail* | 327350 | 0.35 | 0.004 | 0.29 | (476,493) | (486,497) |
| *sam4* | 1800 | 0.35 | -0.010 | 0.30 | (753,1255) | (757,1254) |
| *152762* | 45 | 0.01 | 0.021 | 0.008 | (504,510) | (504,515) |

**Table 2.** Metrics for the *S04387* and *sam4* sample 2D backgrounds: Pearson's correlation (*Corr*), Structural Similarity Index (*SSIM*) and residuals (*Res*) are shown for a comparison between the results of the present work (Bckg) and those obtained using the `ImageJ` software (Bckg ImJ$_x$) with three different choices of rolling-ball radius (x=25/50/100 pixels).

| Sample | $\frac{\|\text{Bckg}\|}{\|\text{Intensity}\|}$ | *Corr*(Bckg,Bckg ImJ$_x$) x=25/50/100 | *SSIM*(Bckg vs Bckg ImJ$_x$) x=25/50/100 | *Res* = $\frac{\|\text{Bckg}-\text{Bckg ImJ}_x\|}{\|\text{Intensity}\|}$ x=25/50/100 |
|---|---|---|---|---|
| *S04387* | 0.03 | 0.91/0.92/0.90 | 0.55/0.55/0.53 | 0.015/0.012/0.012 |
| *sam4* | 0.59 | 0.90/0.96/0.96 | 0.90/0.91/0.91 | 0.25/0.11/0.10 |

**Funding:** This research received no external funding.

**Data Availability Statement:** The MatLab code and XRD patterns are available.

**Acknowledgments:** The author wishes to acknowledge D. Altamura, T. Sibillano, and C. Giannini for providing him with XRD patterns.

**Conflicts of Interest:** The author declares no conflict of interest.

## Supplementary Methods

*Appendix .1. Center Localization*

**Theorem** (Entropy Reduction via Uncorrelated Scale-Free Modulation)**.** *Let $\rho(x)$ be a positive definite function over a domain $D \subseteq \mathbb{R}^n$, normalized such that $\max_{x\in D} \rho(x) = 1$. Let $h(x)$ be a scale-free, positive definite function over the same domain, also satisfying $\max_{x\in D} h(x) = 1$. Assume:*

1. *$h(x)$ is computed based on $\rho(x)$ but is scale-free, i.e., independent of the magnitude scaling of $\rho$;*
2. *$\rho$ and $h$ are uncorrelated in the sense that:*

$$\int_D \rho(x)h(x)\,dx = \int_D \rho(x)\,dx \cdot \int_D h(x)\,dx;$$

3. *Let $\rho_s(x) := \rho(x)/s$ for some $s > 0$.*

*Then there exists a scale factor $s > 0$ such that the entropy of the modulated function*

$$f_s(x) := \frac{\rho(x)}{s} \cdot h(x)$$

*is strictly smaller than the entropy of the unmodulated, scaled function:*

$$H[f_s] < H[\rho_s],$$

*where entropy is defined (up to a constant) as:*

$$H[f] = -\int_D f(x)\log f(x)\,dx.$$

**Proof.**

1. **Normalization:** Since both $\rho$ and $h$ are normalized to have unit maximum, their product $f_s(x) = \rho(x)h(x)/s$ remains bounded by $1/s$. Hence, entropy is well-defined for finite domains and positive definite functions.
2. **Entropy decomposition:** Consider the entropy difference:

$$H[f_s] - H[\rho_s] = -\int_D \left[\frac{\rho(x)}{s}h(x)\log\left(\frac{\rho(x)}{s}h(x)\right) - \frac{\rho(x)}{s}\log\left(\frac{\rho(x)}{s}\right)\right]dx.$$

   This simplifies to:

$$H[f_s] - H[\rho_s] = \frac{H[f_1] - H[\rho]}{s} - \frac{\log(s)}{s}\int_D [\rho(x) - \rho(x)h(x)]dx.$$

3. **Effect of $h(x)$:** Because $h(x) \leq 1$ and is uncorrelated with $\rho(x)$, the last integral is positive.
4. **Existence of optimal $s$:** For large enough $s$, both terms are dominated by the scaling factors. However, since $1/s$ is subleading with respect to $\log(s)/s$, there exists a threshold $s^*$ beyond which the entropy of $f_s$ remains strictly below that of $\rho_s$.
5. **Conclusion:** Therefore, such a scale $s$ exists.

□

**Remark.** *The key insight is that even though $h(x)$ is scale-free and uncorrelated with $\rho(x)$, its interaction with the scaled $\rho$ selectively reduces entropy by concentrating probability mass, a form of information compression.*

*Appendix .2. Denoising*

**Algorithm** D2Q9 Lattice Boltzmann Simulation with Adaptive Relaxation

Initialize parameters: $\alpha = 0.85$, $\beta = 1.25$, $\Omega = \alpha$, $f_0$ original image.
Define D2Q9 weights $W = [16, 4, 1, 4, 1, 4, 1, 4, 1]/36$, and lattice velocities $(c_x, c_y)$.
Set equilibrium distribution $N_{eq} = W \log(f_0)$, and initialize $N = W \cdot \text{rand}$.
**for** $t = 1$ to $t_f$ **do**
    **Streaming:** shift $N_i$ by $(c_x(i), c_y(i))$ (periodic Boundary Condition).
    Compute macroscopic fields: $\rho = \sum_i N_i$, $u_x = \frac{\sum_i N_{eq} c_x}{\sum_i N_{eq}}$, $u_y = \frac{\sum_i N_{eq} c_y}{\sum_i N_{eq}}$.
    Construct local equilibrium (BGK) distribution:
    $f_{eq} = W\big[1 + 3(u \cdot c) + 4.5(u \cdot c)^2 - 1.5|u|^2\big] \log(f_0)$.
    **Collision:** for each $i = 1..Q$,
        compute $N_x, N_y$ as weighted moments, local scale $\Lambda_i = 0.9\langle\sqrt{N_x^2 + N_y^2}\rangle$,
        weight $w_i = 1/[1 + (N_x^2 + N_y^2)/\Lambda_i^2]$, divergence $\nabla \cdot (w \nabla N)_i$.
    Relaxation step: $N \leftarrow N + \Omega\big[(\nabla \cdot (w \nabla N))/\beta + (f_{eq} - N)\big]$.
**end for**
Output: $f_0^{den} = \exp(\rho)$.

*Appendix .3. Background suppression*

**Theorem** (Asymptotic Dominance of Background in Perturbed Intensity Product)**.** *Let the perturbed intensity product in polar coordinates be given by*

$$\mathrm{P}(\vartheta, r) = \prod_{i=1}^{n} [b(\vartheta, r + \epsilon_i) + s(\vartheta, r + \epsilon_i)], \tag{A1}$$

*where:*

- $b(\vartheta, r)$ *is the background function, assumed to be slowly varying in* $r$*,*
- $s(\vartheta, r)$ *is a localized signal function,*
- $\epsilon_i$ *are random perturbations along the r-coordinate, intermediate in scale between the widths of s and b.*

*Then under the assumptions below, the geometric mean of* $\mathrm{P}(\vartheta, r)$ *converges to the background* $b(\vartheta, r)$ *as* $n \to \infty$*:*

$$\lim_{n \to \infty} \sqrt[n]{\mathrm{P}(\vartheta, r)} = b(\vartheta, r).$$

**Proof.** By the slow variation of the background, we approximate:

$$b(\vartheta, r + \epsilon_i) \approx b(\vartheta, r) \quad \Rightarrow \quad \mathrm{P}(\vartheta, r) \approx \prod_{i=1}^{n} [b(\vartheta, r) + s(\vartheta, r + \epsilon_i)].$$

Factoring out the constant background term:

$$\mathrm{P}(\vartheta, r) \approx [b(\vartheta, r)]^n \prod_{i=1}^{n} \left[1 + \frac{s(\vartheta, r + \epsilon_i)}{b(\vartheta, r)}\right].$$

Assume that the signal contributions are localized and non-overlapping, i.e.,

$$i \neq j \Rightarrow s(\vartheta, r + \epsilon_i) \cdot s(\vartheta, r + \epsilon_j) \approx 0.$$

Then applying the first-order product approximation:

$$\prod_{i=1}^{n} (1 + \delta_i) \approx 1 + \sum_{i=1}^{n} \delta_i,$$

we obtain:

$$\mathrm{P}(\vartheta, r) \approx [b(\vartheta, r)]^n \left[1 + \sum_{i=1}^{n} \frac{s(\vartheta, r + \epsilon_i)}{b(\vartheta, r)}\right].$$

Bounding the sum using $s(\vartheta, r + \epsilon_i) \leq \max(s(\vartheta, \cdot))$, we get:

$$\mathrm{P}(\vartheta, r) \leq [b(\vartheta, r)]^n \left[1 + n \cdot \frac{\max(s(\vartheta, \cdot))}{b(\vartheta, r)}\right].$$

Thus, we establish the bounds:

$$[b(\vartheta, r)]^n \leq \mathrm{P}(\vartheta, r) \leq [b(\vartheta, r)]^n \cdot n \left[\frac{1}{n} + \frac{\max(s(\vartheta, \cdot))}{b(\vartheta, r)}\right].$$

Taking $n$-th roots and analyzing the limit:

$$\lim_{n\to\infty} b(\vartheta, r) \leq \lim_{n\to\infty} \sqrt[n]{\mathrm{P}(\vartheta, r)} \leq \lim_{n\to\infty} b(\vartheta, r) \cdot n^{1/n} \cdot \left[\frac{1}{n} + \frac{\max(s(\vartheta, \cdot))}{b(\vartheta, r)}\right]^{1/n}. \tag{A2}$$

Now use standard asymptotic results:

- $n^{1/n} \to 1$,
- $\left[\frac{1}{n} + \frac{\max(s(\vartheta,\cdot))}{b(\vartheta,r)}\right]^{1/n} \to 1$,

to conclude:

$$\lim_{n\to\infty} \sqrt[n]{\mathrm{P}(\vartheta, r)} = b(\vartheta, r).$$

□

**Assumption** (Summary of Key Conditions)**.**

- *The background is slowly varying:* $b(\vartheta, r + \epsilon_i) \approx b(\vartheta, r)$,
- *Signal overlaps are negligible:* $i \neq j \Rightarrow s(\vartheta, r + \epsilon_i) \cdot s(\vartheta, r + \epsilon_j) \approx 0$,
- *Signal values are bounded:* $s(\vartheta, r + \epsilon_i) \leq \max(s(\vartheta, \cdot))$,
- *Asymptotic properties:* $n^{1/n} \to 1$, $c^{1/n} \to 1$ *as* $n \to \infty$.

This shows that the background dominates the product asymptotically, and the geometric mean of the product converges to $b(\vartheta, r)$.

## References


1. Hammersley, A.P.; Svensson, S.O.; Hanfland, M.; Fitch, A.N.; Hausermann, D. Two-dimensional detector software: From real detector to idealised image or two-theta scan. *High Pressure Res.* **1996**, *14*, 235–248.
2. Prescher, C.; Prakapenka, V.B. DIOPTAS: a program for reduction of two-dimensional X-ray diffraction data and data exploration. *High Pressure Res.* **2015**, *35*, 223–230. https://doi.org/10.1080/08957959.2015.1059835.
3. Kieffer, J.; Karkoulis, D. PyFAI, a versatile library for azimuthal regrouping. *J. Phys.: Conf. Ser.* **2013**, *425*, 202012.
4. Kieffer, J. PyFAI Documentation, 2024. Available online: `https://pyfai.readthedocs.io/`.
5. Loy, G.; Zelinsky, A. Fast radial symmetry for detecting points of interest. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 959–973. https://doi.org/10.1109/TPAMI.2003.1217601.
6. Ladisa, M. Denoising X-Ray Diffraction Two-Dimensional Patterns with Lattice Boltzmann Method. *Crystals* **2025**, *15*, 51. https://doi.org/10.3390/cryst15010051.
7. Chen, H.; Chen, S.; Matthaeus, W.H. Recovery of the Navier–Stokes equations using a lattice-gas Boltzmann method. *Phys. Rev. A* **1992**, *45*, R5339.
8. He, X.; Luo, L.S. Lattice Boltzmann Model for the Incompressible Navier–Stokes Equation. *J. Stat. Phys.* **1997**, *88*, 927–944.
9. He, X.; Luo, L.S. A priori derivation of the lattice Boltzmann equation. *Phys. Rev. E* **1997**, *55*, R6333.

10. Bhatnagar, P.; Gross, E.; Krook, M. A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems. *Phys. Rev.* **1954**, *94*, 511–525.
11. Mighell, K. J., 1999, *Parameter Estimation in Astronomy with Poisson-Distributed Data. I. The* $\chi^2_\gamma$ *Statistic*, Astrophysical Journal, **518**, 380–393.
12. Sternberg, S.R. Biomedical image processing. *Computer* **1983**, *16*, 22–34.
13. Rasband, W.S. ImageJ; U.S. National Institutes of Health: Bethesda, MD, USA, 1997–2018. Available online: `https://imagej.net/ij/`.
14. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675.
15. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. https://doi.org/10.1109/TIP.2003.819861.
16. Terzi, A.; Storelli, E.; Bettini, S.; Sibillano, T.; Altamura, D.; Salvatore, L.; Madaghiele, M.; Romano, A.; Siliqi, D.; Ladisa, M.; De Caro, L.; Quattrini, V.; Valli, L.; Sannino, A.; Giannini, C. Effects of processing on structural, mechanical and biological properties of collagen-based substrates for regenerative medicine. *Sci. Rep.* **2018**, *8*, 2045–2322. https://doi.org/10.1038/s41598-018-19786-0.
17. Altamura, D.; Lassandro, R.; Vittoria, F.A.; De Caro, L.; Ladisa, M.; Giannini, C. X-ray microimaging laboratory (XMI-LAB). *J. Appl. Crystallogr.* **2012**, *45*, 869–873. https://doi.org/10.1107/S0021889812025733.
18. Scattarella, F.; Altamura, E.; Albanese, P.; Siliqi, D.; Ladisa, M.; Mavelli, F.; Giannini, C.; Altamura, D. Table-top combined scanning X-ray small angle scattering and transmission microscopies of lipid vesicles dispersed in free-standing gel. *RSC Adv.* **2020**, *11*, 484–492. https://doi.org/10.1039/d0ra08581b.
19. De Caro, L.; Giannini, C.; Lassandro, R.; Scattarella, F.; Sibillano, T.; Matricciani, E.; Fanti, G. X-Ray Dating of Ancient Linen Fabrics. *Heritage* **2019**, *2*, 2763–2783. https://doi.org/10.3390/heritage2040171.
20. Broennimann, C.; Eikenberry, E.; Henrich, B.; Horisberger, R.; Hülsen, G.; Pohl, E.; Schmitt, B.; Schulze-Briese, C.; Suzuki, M.; Tomizaki, T.; Toyokawa, H.; Wagner, A. The PILATUS 1M detector. *J. Synchrotron Radiat.* **2006**, *13*, 120–130. https://doi.org/10.1107/S0909049505038665.