

Article

A Framework for Global Background Subtraction in Fiber Diffraction Patterns from Striated Muscle

Irina Klein ¹, Gady Agam ¹ and Thomas Irving ^{2,*}

¹ Computer Science Department, Illinois Institute of Technology, Chicago, IL 60616, USA; iklein@illinoistech.edu (I.K.); agam@illinoistech.edu (G.A.)

² Biology Department, Illinois Institute of Technology, Chicago, IL 60616, USA

* Correspondence: irving@illinoistech.edu

Abstract

Accurate background subtraction is critical in the analysis of X-ray fiber diffraction patterns from striated muscle. Diffuse scattering and low signal-to-noise ratio complicate the separation of signal from background. Existing methods are applied heuristically, without guidance on parameter selection, leading to variability in following analyses. We propose a framework for background removal in muscle fiber diffraction images that evaluates methods and selects optimal parameters. It combines quantitative metrics and validation using synthetic features to guide the optimization. Minimizing the aggregate loss produced visually plausible results in the representative cases examined by the authors; independent blinded validation was not performed. The loss metric and its components allow currently used approaches to be compared within a single reproducible scheme. Individual metrics show trade-offs among approaches, such as preservation of detail or oversubtraction. Additionally, we evaluate parametric fitting of the equatorial streak and isotropic general background, which may be combined with the optimization framework. The framework improves the analysis of fiber diffraction patterns by providing a reproducible quantitative basis for the choice of background subtraction method and its parameters.

Keywords: background subtraction; synchrotron X-ray diffraction; fiber diffraction; striated muscle; quantitative evaluation; image analysis

1. Introduction

Scientific studies in materials science and structural biology that rely on the analysis of diffraction patterns from complex, composite systems, such as muscle tissue, face the challenge of extracting weak signals that overlap complex and non-uniform backgrounds. The background comes from the amorphous material within the sample, lattice disorder, and scattering from the experimental environment. It may obscure subtle diffraction features that are critical for quantitative analysis. X-ray fiber diffraction studies have provided unique insight into the sarcomere-level molecular structural dynamics of muscle and can contribute to understanding the molecular mechanisms underlying diseases ranging from nemaline myopathy to heart failure [1]. Interpretation of such diffraction patterns requires accurate measurement of feature intensity.

Muscle diffraction patterns have, in addition to equatorial peaks, meridional peaks, and continuous layer lines arising from the quasi-helical actin and myosin filaments, amorphous scattering from non-coherently diffracting material. In small-angle X-ray fiber diffraction experiments on muscle fibers, the signal of interest consists of diffraction peaks



Academic Editor: Constantinos Koutsojannis

Received: 5 August 2026

Revised: 22 September 2026

Accepted: 24 September 2026

Published: 29 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

of varying intensity superimposed on a strong, spatially non-uniform diffuse background that must be removed prior to analysis [2].

Analysis of two-dimensional fiber diffraction patterns from partially crystalline structures such as muscle fibers using standard crystallographic techniques is challenging. The patterns show continuous bands of intensity along layer lines and meridional reflections on top of diffuse background scattering. As a result, fiber diffraction data are usually analyzed using one-dimensional projections which are effective for sharp reflections along the meridional and equatorial axes, where a large fraction of the signal is concentrated. However, one-dimensional projections are less suitable for studying continuous features such as layer lines, where separating signal from background is difficult. A robust approach for two-dimensional background subtraction would allow us to extract structural information for features that are not analyzed systematically.

Background removal strategy in diffraction studies depends on the dimensionality of the data. Usually, when working with one-dimensional intensity profiles, diffraction peaks are analyzed individually. In this case, the background intensity is estimated by averaging values in regions adjacent to the peak, which is then subtracted from the integrated peak intensity [3]. This approach is simple and effective for individual reflections, but does not provide information on spatial variation in the background. Another commonly used technique is the convex hull algorithm [4] applied to one-dimensional profiles to separate diffraction peaks from a smoothly varying background [5]. The convex hull method allows us to subtract high-intensity background, such as the equatorial streak; however, it is affected by background noise, which may introduce errors in measurements. Some parametric background models, such as low-order polynomials or spline functions, may be used to remove slowly varying backgrounds, whereas steep exponential components may be used to model high-intensity variations [6]. One-dimensional approaches are effective for sharp reflections on the equator or the meridian; however, the two-dimensional spatial information is inherently discarded. As a result, such methods provide limited information on continuous diffraction features such as layer lines.

Two-dimensional background subtraction methods are applied directly to diffraction patterns and preserve the spatial information. Before processing the image, a sample-free pattern may be subtracted to remove air scattering and instrumental artifacts. Most two-dimensional approaches assume that the image is centered. Some methods are based on specific assumptions about the geometry or smoothness of the background. The radially symmetric method assumes that the background is symmetric about the beam and estimates it by averaging intensities in concentric rings [2]. The method is computationally efficient; however, this assumption does not hold for anisotropic background features, such as the equatorial streak characteristic of muscle fiber diffraction. Partition-based methods, such as the roving window method [2], which is attributed there to P. Langan, divide the pattern into regions, estimate the background locally, and interpolate between regions. Another example is the radial convex hull approach, which applies the convex hull algorithm to angular projections [5] and then interpolates the two-dimensional background. The performance of these methods depends on partition size; they may create artifacts in areas with rapidly changing backgrounds. The performance of radial convex hull decreases at high angles because angular sectors cover larger portions of the pattern and the interpolation is no longer smooth.

Some methods utilize image processing techniques, for example, the iterative low-pass filtering method estimates the background through repeated smoothing [7]. The white top-hat morphological filter [8] removes small details, using the result as background. For both methods, the performance depends on the size and the shape of the filter; inappropriate choices may oversubtract smooth low-intensity diffraction signals. Frequency-based

approaches, such as Fourier or wavelet transforms, isolate low-frequency components as background. Such techniques are effective for crystalline diffraction patterns with sharp peaks [9,10], but perform poorly on fiber diffraction patterns, where signal and background overlap. A related strategy is to re-express the pattern in polar coordinates and estimate the background along the radial direction [11]. This is efficient for powder-like samples, in which the crystallites have no preferred orientation and the pattern is, therefore, a set of isotropic, centrally symmetric rings. Oriented fiber specimens do not satisfy these conditions, as discussed in Section 4. Similarly, the rolling-ball method [12] performs well on patterns from crystalline samples, but is less suitable for fiber diffraction data.

More recently, data-driven approaches have been explored. A supervised deep learning model has been used to learn muscle fiber diffraction backgrounds based on non-parametric method estimates [13]. However, with fiber diffraction patterns, the lack of ground-truth background and the difficulty of generating representative training datasets limit the practical applicability of such approaches.

In current approaches to two-dimensional background removal, the operator visually assesses the result and adjusts algorithm parameters. Manual processing introduces several challenges. Firstly, weak signals may be lost due to the large intensity contrast between background and features. Secondly, noise may obscure low-intensity features. Next, visual assessment introduces subjectivity and may result in suboptimal solutions. Finally, reproducibility depends on thorough documentation and reuse of parameters. We ask whether we can find an optimal solution to the background subtraction problem without ground truth. A lack of a standardized evaluation framework prevents objective comparison of background subtraction methods and parameters.

This work presents a framework for improving the analysis of diffraction patterns that contain complex, non-uniform backgrounds. The main contributions are as follows:

1. Quantitative assessment of background subtraction quality given lack of ground truth: a set of quantitative metrics is defined based on the properties of the expected background-subtracted patterns. These measures give a reproducible way to assess background subtraction and allow for unsupervised comparison of algorithms.
2. Sampling framework for measuring fine detail preservation: injection of faint synthetic features to penalize loss of signal.
3. Optimization procedure for background subtraction on muscle fiber diffraction patterns: a unified loss and optimization pipeline for automated method-specific parameter selection and automated method selection. The procedure has been incorporated in MuscleX, a software package for muscle fiber diffraction analysis, and is available for use.

Background subtraction in diffraction X-ray images is typically an interactive and dataset-dependent process. The analyst must choose among alternative algorithms, adjust multiple parameters, visually inspect the resulting image, and repeat these steps until the estimated background is satisfactory. We formulate this task as a human-in-the-loop numerical optimization problem. The proposed framework searches across alternative algorithms and parameter configurations while allowing the analyst to fix or constrain selected parameters and specify a small set of readily interpretable weights in the objective, or loss, function. These weights express physically meaningful preferences concerning the desired result, thereby preserving expert control while avoiding exhaustive manual exploration of the full configuration space.

The principal efficiency gain provided by the framework is a reduction in operator effort, rather than an assertion that manual tuning cannot achieve an equivalent solution. Given sufficient time and experimentation, an expert may obtain a background estimate comparable to that identified by the proposed framework. In the conventional work-

flow, however, the expert must repeatedly select algorithms, vary parameters, and inspect candidate outputs. Our framework automates much of this search and presents an optimized configuration for expert verification. The improvement in operational efficiency, therefore, follows directly from replacing repeated manual search operations with automated optimization. Our claim concerns the efficiency, systematicity, and reproducibility of the workflow, rather than necessarily reducing computational cost or producing a better optimum than could ultimately be found through careful manual tuning.

The framework further exploits the coherence of diffraction image series. Images acquired under the same or similar experimental conditions frequently share background characteristics, making it unnecessary to perform a complete optimization independently for every image. After the analyst optimizes and verifies a configuration on a representative image, the selected algorithm and parameters can be propagated automatically to the remaining images in the series. The objective value is then monitored as a consistency criterion. If it becomes sufficiently large, or deviates substantially from the values observed during calibration, the framework alerts the analyst that the propagated configuration may no longer be appropriate and that inspection or reoptimization may be required. The initial calibration effort can, thus, be amortized across a series while retaining human oversight when the assumptions supporting propagation cease to hold.

At the same time, our experiments show that no single set of objective-function weights or background-subtraction parameters was suitable for all of the experimental datasets examined. This variability is expected because specimens, acquisition conditions, noise characteristics, and background structure differ among experiments. Operator intervention, therefore, remains necessary to define the appropriate tradeoffs and verify the resulting background estimate for each new dataset. The goal is not to produce a fully automatic algorithm that can be applied without supervision to any experimental diffraction pattern. Rather, the framework provides a robust and systematic tool that guides the analyst toward an optimized background-subtraction scheme for a particular dataset using physically motivated criteria and consistently applied quantitative metrics.

Although the framework employs an objective function, parameter optimization, and default initialization values, it is not a machine-learning method and does not train a predictive model intended to generalize across experimental systems. Directly observed ground-truth background-subtracted images are generally unavailable for the range of experimental systems considered, limiting both the development and objective validation of conventional supervised-learning formulations. In our method, the optimized quantities are processing parameters for a particular dataset, while the user-defined objective-function weights encode explicit, interpretable preferences rather than learned model parameters. Any default values serve only as starting points for dataset-specific optimization and are not presented as universally applicable settings. Adaptation to the characteristics of an individual dataset is, therefore, an intended feature of the method, rather than overfitting that must be prevented through training-validation separation or leave-one-dataset-out evaluation.

The present work also does not claim that use of the framework necessarily improves the accuracy of downstream measurements, such as diffraction intensities or spacings, relative to a background estimate carefully optimized by an expert. Instead, its contribution is to reduce manual search, apply physically motivated criteria consistently, support reproducible selection of a background-subtraction scheme, propagate a verified configuration across compatible images, and identify cases requiring renewed intervention. Accordingly, the relevant evaluation questions are whether the framework can efficiently identify satisfactory configurations, whether its components contribute to systematic and reproducible optimization, whether a verified configuration can be applied reliably within a compat-

ible image series, and whether the monitoring criterion can identify images for which reoptimization is warranted.

Relation to prior work

None of the background subtraction algorithms evaluated here are new. Circularly symmetric subtraction, the roving window method, the radial convex hull, iterative low-pass filtering and white top-hat filtering either originate in the CCP13 software [2,6] or developed for earlier versions of MuscleX [14], where they are applied with default or manually chosen parameters. Our earlier data-driven work [13] learned a background from the estimates produced by those same non-parametric methods and, therefore, inherited, rather than assessed, their quality. What is introduced here is the evaluation layer built on top of them: the synthetic-feature injection used as a surrogate ground truth (Section 2.2), the composite loss constructed from it (Section 2.3), the automated per-image parameter and method selection driven by that loss, and the iterative parametric fit of the equatorial streak together with the general background (Section 2.6). The subtraction algorithms themselves are treated as interchangeable components of the pipeline.

2. Materials and Methods

This part explains the approach to assess background subtraction for muscle fiber diffraction patterns. We describe qualitative objectives and define quantitative metrics that reflect successful background removal. We then combine these metrics into a single loss that balances background subtraction and signal preservation and is minimized for each method using a coordinate-descent search over the parameters. Figure 1 shows the overall workflow.

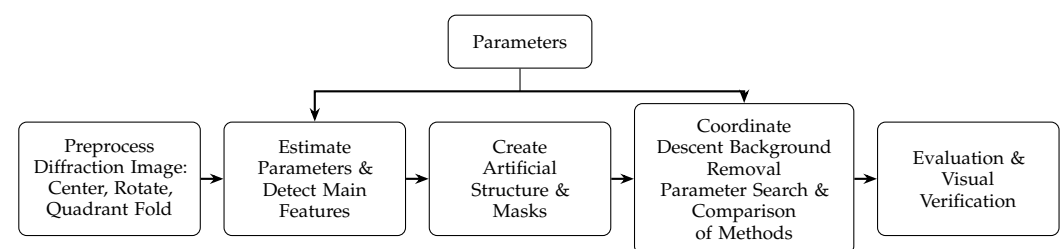


Figure 1. General flow of operations. For each image, following initial preprocessing using standard techniques, relevant parameters and features are estimated, synthetic signal components are generated, and a systematic parameter search is performed.

2.1. Evaluation Framework

2.1.1. Qualitative Objectives

A background-subtracted diffraction pattern should fulfill several qualitative criteria: preserve true diffraction features without distortion and avoid introduction of artifacts. In regions with no diffraction signal, the residual should be smooth and near-zero, and for patterns taken under similar conditions, background subtraction results and parameters should be consistent. These qualitative objectives motivate the following metrics in order to capture deviations from the described objectives in a reproducible way.

2.1.2. Quantitative Metrics

Throughout, I denotes the input pattern, B the estimated background, $D = I - B$ the background-subtracted image, and $\mathcal{M}$ the set of unmasked pixels over which metrics are evaluated (Section 2.2.3), with $|\mathcal{M}|$ its cardinality. All metrics are computed on the raw quantities defined below and are subsequently divided by fixed normalization constants (Section 2.3.1).

Oversubtraction, denoted $f_{\text{neg,con}}$, measures systematic negative excursions of the residual. The negative pixels $\{p \in \mathcal{M} : D(p) < 0\}$ are grouped into connected components using 4-connectivity (nearest-neighbour structuring element); components smaller than $n_{\text{min}} = 9$ pixels are discarded, and

$$f_{\text{neg,con}} = \frac{|\{p \in \mathcal{M} : D(p) < 0, p \in \mathcal{C}, |\mathcal{C}| \geq n_{\text{min}}\}|}{|\mathcal{M}|}, \quad (1)$$

where $\mathcal{C}$ denotes the connected component containing p . The size threshold excludes isolated negative pixels caused by random noise, so that the metric responds to spatially extended oversubtraction rather than to noise; its ideal value is zero.

Residual background level is measured by the fraction of unmasked pixels whose residual intensity exceeds a small baseline threshold τ ,

$$1 - f_B = \frac{|\{p \in \mathcal{M} : D(p) > \tau\}|}{|\mathcal{M}|}, \quad (2)$$

so that f_B is the complementary fraction of pixels at or below τ . A high f_B , and, therefore, a low value of $1 - f_B$, indicates minimal residual background. The threshold τ is estimated from the average intensity near the edge of the pattern, where the background is expected to be minimal, and is bounded below by a small positive constant so that it does not collapse to zero on very clean patterns. It is the quantity $1 - f_B$ that enters the loss function.

Background smoothness, denoted S_{BG} , is computed on the estimated background and normalized by the mean intensity of the input pattern so that it is invariant to exposure. It is defined as

$$S_{\text{BG}} = \frac{\sum_{(x,y) \in \mathcal{M}_v} |B(x,y+1) - B(x,y)|}{|\mathcal{M}| \bar{I}}, \quad \bar{I} = \frac{1}{|\mathcal{M}|} \sum_{p \in \mathcal{M}} I(p), \quad (3)$$

where $\mathcal{M}_v$ is the set of vertically adjacent pixel pairs that both lie in $\mathcal{M}$. A smooth background gives a low value for this metric. The derivative is taken along the vertical direction only. In these patterns the residual structure that we most want to penalize varies principally along this direction. That structure is the incomplete removal of background between layer lines, together with the vertical banding left by filters whose extent is mismatched to the layer-line spacing. The dominant horizontal variation, by contrast, is the intended fall-off of the equatorial streak. A comparison against a full two-dimensional gradient measure was not carried out and is noted in Section 4.

2.1.3. Validation Using Synthetic Structures

Because ground-truth background is unknown for diffraction patterns, synthetic diffraction features are added to the pattern before processing so that signal preservation can be quantified against a known reference. Each image is processed twice, once with and once without the injected features, and the recovered synthetic signal is obtained as the difference of the two background-subtracted results, $R = D^{\text{syn}} - D$. Let S be the injection template, that is, the image containing only the synthetic features, and let $\mathcal{M}_{\text{syn}}$ be the intersection of the synthetic-feature mask with the general evaluation mask $\mathcal{M}$. Both signals are clipped at zero, $a = \max(R, 0)$ and $b = \max(S, 0)$, so that the metric scores

recovery of positive structure and oversubtraction is left to $f_{\text{neg,syn}}$. Signal preservation is then measured by the offset-invariant normalized mean squared error

$$\text{NMSE}_{\text{syn}} = \frac{\sum_{p \in \mathcal{M}_{\text{syn}}} (a(p) - c - b(p))^2}{\sum_{p \in \mathcal{M}_{\text{syn}}} b(p)^2 + \varepsilon}, \quad c = \frac{1}{|\mathcal{M}_{\text{syn}}|} \sum_{p \in \mathcal{M}_{\text{syn}}} (a(p) - b(p)), \quad (4)$$

where c is the constant offset that minimizes the numerator in the least-squares sense and $\varepsilon = 10^{-12}$ guards the denominator. Subtracting c makes the metric insensitive to a uniform residual pedestal, so that it responds to distortion of the injected features rather than to a constant baseline error, which is instead captured by $1 - f_B$. Its value is expected to be zero, indicating perfect preservation of the synthetic signal. Because the normalization is by the energy of the reference features, NMSE_{syn} becomes large and poorly conditioned when the injected signal is weak relative to the residual; the lower bound imposed on the synthetic amplitude (Section 2.2.1) keeps the denominator bounded away from zero, and the large standard deviations of this column in the results reflect the remaining heavy-tailed behavior on low-intensity patterns. Although synthetic features are placed in regions without real features, if they happen to overlap real structure, NMSE_{syn} would increase.

We measure feature oversubtraction, $f_{\text{neg,syn}}$, as the fraction of synthetic-mask pixels at which the recovered signal is negative,

$$f_{\text{neg,syn}} = \frac{|\{p \in \mathcal{M}_{\text{syn}} : R(p) < 0\}|}{|\mathcal{M}_{\text{syn}}|}. \quad (5)$$

All metrics are summarized in Table 1. The framework supports visual assessment with these quantitative, reproducible criteria; it does not remove the need for operator verification of the outcome (Section 4).

Table 1. Quantitative metrics used for evaluating background subtraction. Each metric is divided by a fixed normalization constant, equal to its mean value over a calibration subset, before aggregation in the loss function (Section 2.3.1). All metric values reported in this paper are, therefore, dimensionless ratios and not raw pixel fractions; a value greater than one indicates a metric worse than the calibration mean.

Purpose	Description	Symbol
Ensure background smoothness	Mean absolute vertical derivative of the residual background.	S_{BG}
Avoid oversubtraction of connected regions	Fraction of pixels belonging to 4-connected clusters of negative pixels of at least nine pixels, indicating systematic oversubtraction artifacts.	$f_{\text{neg,con}}$
Ensure signal preservation	Normalized offset-invariant mean squared error between the synthetic data in the background-subtracted image and the reference image containing only synthetic features.	NMSE_{syn}
Avoid signal oversubtraction	Fraction of pixels within the synthetic-feature mask at which the recovered synthetic signal is negative, emphasizing oversubtraction of faint features.	$f_{\text{neg,syn}}$
Ensure low residual baseline	Penalty for structured residual background: the fraction of unmasked pixels whose intensity exceeds a small baseline threshold, equal to $1 - f_B$ with f_B the fraction of pixels at or below that threshold.	$1 - f_B$

2.2. Sampling Framework for Assessing Detail Preservation

This subsection describes the design of synthetic signal used for objective evaluation and region-specific masking used to isolate background regions for metric calculation.

2.2.1. Synthetic Signal Validation Design

To address the main challenge of missing ground-truth, we add synthetic diffraction features similar to real signal prior to background subtraction, as summarized in Algorithm 1. Synthetic features provide a controlled way to measure signal loss and over-subtraction. They are approximated using two-dimensional Gaussian functions elongated along the horizontal axis, given by

$$G(x, y) = A \exp\left(-\frac{(x - x_0)^2}{2\sigma_x^2} - \frac{(y - y_0)^2}{2\sigma_y^2}\right), \quad (6)$$

where A is the amplitude, (x_0, y_0) is the feature center, and σ_x and σ_y are the standard deviations. The shape parameters are selected automatically for each image to match its scale. The extents of the feature are set as full widths at half maximum, taken as a fifth of the radial position $S_{1,0}$ of the 1,0 equatorial reflection and a tenth of the M1 meridional spacing. Since Equation (6) is parameterized by standard deviations, these widths are converted using the Gaussian relation $\text{FWHM} = 2\sqrt{2\ln 2}\sigma \approx 2.355\sigma$:

$$\text{FWHM}_x = \frac{S_{1,0}}{5}, \quad \text{FWHM}_y = \frac{\text{M1}}{10}, \quad \sigma_{x,y} = \frac{\text{FWHM}_{x,y}}{2\sqrt{2\ln 2}}. \quad (7)$$

The amplitude of the injected features is tapered radially as $1/r$, where r is the distance from the pattern center, so that features far from the center are faint in the same way as real high-angle reflections. We adopt this taper empirically rather than as a physical correction. For equatorial reflections the fiber-diffraction Lorentz factor does behave approximately as $1/r$ at small angles, but on and near the meridian the appropriate form is anisotropic, scaling as $1/R_\perp$ with $R_\perp$ the reciprocal-space distance from the meridian, and it is not defined on the meridian itself [15]. The isotropic taper used here is, therefore, a convenience that imposes a plausible, monotonically decreasing amplitude profile on the surrogate features; it is not intended to reproduce the intensity of any real reflection, and the metrics computed from the injected features do not require it to be physically exact.

The amplitude of the synthetic features is set from the two strongest equatorial reflections, conventionally indexed 1,0 and 1,1 [1]. We set the maximum synthetic feature amplitude to $A_{\max} = 5\% I_{1,0} + 2\% I_{1,1}$, subject to a lower bound, which makes the synthetic features faint yet still detectable. In the implementation the two reflections are identified as the two most prominent maxima of the equatorial projection, so their assignment can exchange between the 1,0 and 1,1 positions as the relative peak intensities change with physiological state; since $A_{\max}$ is a weighted sum of both and is bounded below, this affects the injected amplitude only mildly. Synthetic features are placed on a regular grid; the sampling frequency (f_x, f_y) controls density and is chosen to avoid known diffraction features and layer lines. Such an approach allows sampling of the entire pattern and ensures reproducibility of results in contrast to random feature placement.

Algorithm 1 Procedure for synthetic-feature generation and mask construction used for validation in the absence of ground-truth background. The scale controls $(z_x, z_y) = (5, 10)$ and amplitude fractions $(\alpha, \beta) = (0.05, 0.02)$ reproduce Equation (7) and $A_{\max}$; $A_{\min}$ is the amplitude floor.

Require: Image I , prior peak positions $(S'_{1,0}, m'_1)$, controls $(f, \alpha, \beta, z_x, z_y)$

Ensure: Synthetic image I^{syn} , artificial-only image I^{art} , mask M

```

1: Detect equator peaks  $S_{1,0}, S_{1,1} \leftarrow \text{FINDEQUATORPEAKPOSITIONS}(I)$ 
2: if  $|S_{1,0} - i'_{1,0}| > \delta$  then
3:    $S_{1,0} \leftarrow S'_{1,0}$ 
4: end if
5: Detect meridional peak  $m_1 \leftarrow \text{FINDMERIDIONALPEAKPOSITIONS}(I)$ 
6: if  $|m_1 - m'_1| > \delta$  then
7:    $m_1 \leftarrow m'_1$ 
8: end if
9:  $\sigma_x \leftarrow S_{1,0} / (z_x \cdot 2\sqrt{2\ln 2})$ 
10:  $\sigma_y \leftarrow m_1 / (z_y \cdot 2\sqrt{2\ln 2})$ 
11:  $A \leftarrow \max(A_{\min}, \alpha \cdot \text{Equator}(S_{1,0}) + \beta \cdot \text{Equator}(S_{1,1}))$ 
12:  $K \leftarrow \text{GAUSSIANKERNEL}(A, \sigma_x, \sigma_y)$ 
13:  $G \leftarrow \text{SAMPLINGGRID}(f, S_{1,0}, m_1)$ 
14: Initialize  $I^{\text{art}} \leftarrow 0, M \leftarrow 0$ 
15: for all  $(x, y) \in G$  do
16:    $I^{\text{art}} \leftarrow I^{\text{art}} + \text{PASTEKERNEL}(K, x, y)$ 
17:   Mark support of  $K$  at  $(x, y)$  in  $M$ 
18: end for
19:  $I^{\text{art}} \leftarrow \text{INTENSITYDECAY}(I^{\text{art}})$ 
20:  $I^{\text{syn}} \leftarrow I + I^{\text{art}}$ 
21: return  $(I^{\text{syn}}, I^{\text{art}}, M)$ 

```

2.2.2. Parameter Sensitivity

The sensitivity to synthetic signal parameters was tested by varying amplitude, feature extent, and sampling frequency, as shown in Figure 2. Table 2 shows that these variations lead to similar loss values. Since the default settings resulted in low loss values, they were used for the subsequent analyses.

Table 2. Sensitivity of the aggregate loss to variations in synthetic signal parameters evaluated on a subset of the data. Average loss values are reported for ten images from three datasets. For each parameter variation, all other parameters were held at their medium setting. Results show that the default settings produce consistently low loss values and stable method rankings. The variations cover feature extent, amplitude and sampling density; randomized spatial positions and orientations were not tested (Section 4). ‘LL’ stands for layer lines.

Parameter	Variation	Setting	Loss
Horizontal extent σ_x	Short	$S_{1,0}/7$	0.527 ± 0.14
	Medium	$S_{1,0}/5$	0.523 ± 0.11
	Long	$S_{1,0}/3$	0.550 ± 0.14
	Extra Long	$S_{1,0}/1$	0.525 ± 0.09
Vertical extent σ_y	Thin	M1/20	0.520 ± 0.11
	Medium	M1/10	0.523 ± 0.11
	Thick	M1/5	0.530 ± 0.11
Sampling frequency (f_x, f_y)	Sparse	Every two LL.	0.523 ± 0.11
	Medium	Every LL.	0.523 ± 0.11
	Dense	Two between LL.	0.535 ± 0.12

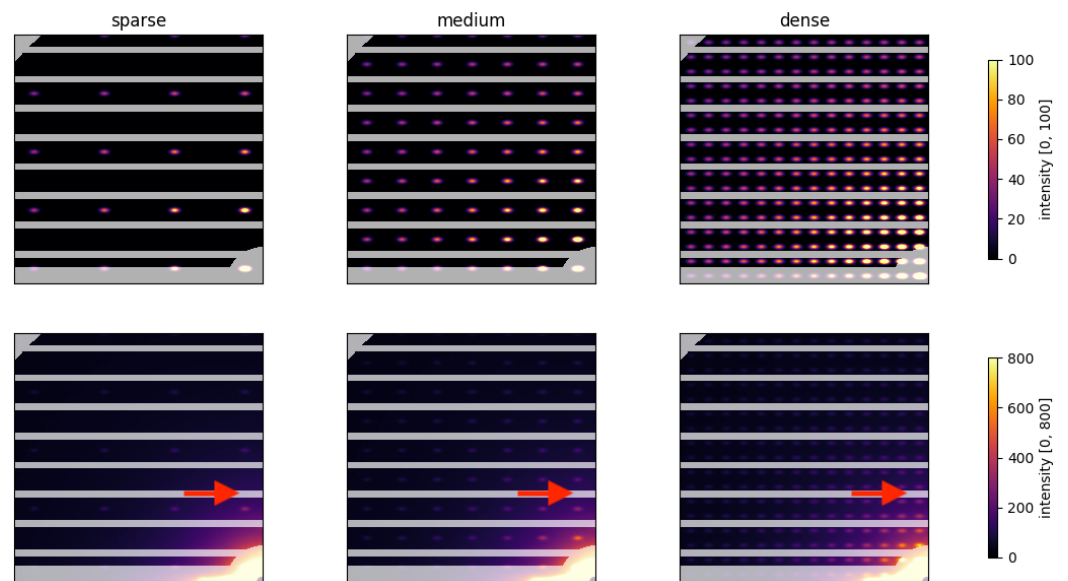


Figure 2. Representative placement of synthetic features, visible as small bright elliptical dots. The figure also shows masking of the central beam region, equatorial streak background, and layer-line regions (see Section 2.2.3). The synthetic features are distributed to sample background-dominated regions while avoiding overlap with real diffraction signal. (**Top**): synthetic features. (**Bottom**): real diffraction image with superimposed synthetic features. The three columns represent different synthetic feature placement frequencies: sparse, medium, and dense. The intensity of the synthetic features is increased for better visibility. Red arrow indicates the position of the third myosin layer line and meridional reflection at 14.3 nm.

2.2.3. Region-Specific Masking

Accurate calculation of evaluation metrics requires distinguishing regions with signal from background or artifacts. To achieve this, several complementary masks are applied. Main signal mask is an annulus that excludes non-representative signal: the beam center and detector edges. Equator mask is a horizontal band along the equatorial streak, which has a much higher intensity than the rest of the background. Synthetic mask consists of rectangles that extend 2σ around each injected synthetic feature. Real-feature mask consists of horizontal bands at layer-line positions; it is used to avoid contamination from real diffraction signal. The masks restrict metric evaluation specifically to background regions.

2.3. Optimization Procedure

This subsection describes how the metrics are combined into an objective function, which is minimized to tune each method's parameters.

2.3.1. Aggregate Loss Function

The metrics are combined into a unified loss function, defined as

$$L = \lambda_1 S_{BG} + \lambda_2 f_{neg,con} + \lambda_3 f_{neg,syn} + \lambda_4 NMSE_{syn} + \lambda_5 (1 - f_B), \quad (8)$$

where each raw metric is first divided by a fixed normalization constant equal to its mean value over a calibration subset of images, so that the five terms are dimensionless and of comparable magnitude. The constants used throughout this work, and stored as defaults in the implementation, are

$$\begin{aligned} \bar{S}_{BG} &= 1.4868 \times 10^{-3}, & \bar{f}_{neg,con} &= 0.1888, & \bar{f}_{neg,syn} &= 0.5133, \\ \overline{NMSE}_{syn} &= 1.4245 \times 10^{-2}, & \overline{(1 - f_B)} &= 0.34. \end{aligned} \quad (9)$$

Because the tabulated metrics are the normalized quantities, values greater than one indicate a metric worse than the calibration mean and must not be read as raw pixel fractions. The calibration subset consists of 74 patterns with 45 images coming from nine datasets used in the evaluation, each contributing 5 patterns, and 29 patterns from six other datasets. The calibration subset spans a range of intensities and covers various sample environments and detectors. It partially covers the optimization and evaluation sets.

The weights λ_i encode the relative importance of the five criteria and are exposed to the user. The values used here for all methods other than CS, $[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] = [0.1, 0.05, 0.05, 0.01, 0.8]$, place most of the weight on the residual-baseline term, which is the criterion that most directly reflects whether the diffuse background has been removed, while the two synthetic-feature terms carry small weights and act as guards against over-subtraction rather than as primary objectives. For the CS method, a slightly different weight vector is used, $[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] = [0.1, 0.05, 0.05, 0.001, 0.8]$, which places less weight on synthetic-feature preservation as the method affects the features more strongly by design and placing more weight on that results in less visually plausible results. The weights were selected by inspecting background-subtraction results on a small set of images and were then applied across datasets; the one deliberate departure is the equatorial region, where the smoothness term is inapplicable and its weight is reallocated to the over-subtraction terms (Section 3.5). The summary results Tables 3–5 use the same weight vector $[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] = [0.1, 0.05, 0.05, 0.01, 0.8]$ for the ordering of all methods.

Two consequences of this design should be stated plainly. First, because the weights define the objective that is subsequently minimized and reported, the ranking of closely-spaced methods is conditional on them: re-scoring the cross-dataset results of Section 3.3 with the equatorial weighting $[0, 0.1, 0.1, 0.01, 0.8]$ places RW first and G-ILPF second, ahead of B-ILPF and WTH. Second, because the weights were informed by visual inspection, agreement between the aggregate loss and visual assessment is not independent evidence that the loss captures expert judgment. Both points are taken up in Section 4.

Table 6 shows the systematic sensitivity analysis over the weight simplex. Since the loss is a weighted sum of five normalized metrics and the weights are a modeling choice, the sweep evaluates which method wins across all sampled weight vectors, and how often. The weight vector is sampled from a Dirichlet distribution since it is non-negative and sums up to 1. Three variants of the mean weight priors m are evaluated: (1) equal weights with $m_1 = [0.2, 0.2, 0.2, 0.2, 0.2]$, (2) $1 - f_B$ is up-weighted with $m_2 = [0.125, 0.125, 0.125, 0.125, 0.5]$ and (3) $1 - f_B$ is up-weighted and NMSE_{syn} is down-weighted with $m_3 = [0.15, 0.15, 0.15, 0.05, 0.5]$. Under equal weights, the average baseline ranks first most often due to its smoothness and no distortion of synthetic features; the win probability is spread across methods rather than concentrated on one. Setting prior mean to place most of the weight on subtraction $1 - f_B$ and less weight on synthetic feature distortion NMSE_{syn} , which matches closer to the selected weight vector, gives higher probability of WTH scoring first. The per-dataset per-method probabilities are shown in Figure A3.

Visual inspection remains necessary to confirm the absence of artifacts that the loss does not capture, in particular subtle layer-line distortion. Visualization of the results based on the optimization of individual metrics, shown in Figure A1, reflects that no single metric is sufficient to guide optimization.

Table 3. Comparison of background subtraction methods across all evaluated images. Values are reported as mean $\pm$ st.dev. across images to emphasize the variability across datasets rather than absolute rankings. All metric columns are normalized as described in Section 2.3.1 and are dimensionless ratios, not raw pixel fractions; values above one indicate a metric worse than the calibration mean. Lower values indicate better performance for each metric. The ordering is conditional on the weight vector of Equation (8) with the weights $[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] = [0.1, 0.05, 0.05, 0.01, 0.8]$; see Section 4. The AVG method is a constant-background baseline included as a reference and is not expected to perform well on spatially varying backgrounds; it might however perform well on patterns with low background intensity. Best values are highlighted in bold and second-best are underlined.

Method	Loss	S _{BG}	f _{neg,con}	f _{neg,syn}	NMSE _{syn}	1 – f _B
AVG	0.73 $\pm$ 0.12	0.00 $\pm$ 0.00	4.02 $\pm$ 0.11	1.34 $\pm$ 0.08	0.90 $\pm$ 9.67	0.56 $\pm$ 0.08
B-ILPF	0.62 $\pm$ 0.24	1.99 $\pm$ 0.70	1.00 $\pm$ 0.52	1.19 $\pm$ 0.14	1.69 $\pm$ 2.73	0.36 $\pm$ 0.30
CS	0.67 $\pm$ 0.20	2.24 $\pm$ 0.94	1.89 $\pm$ 1.32	0.97 $\pm$ 0.29	1.76 $\pm$ 2.61	0.36 $\pm$ 0.22
G-ILPF	<u>0.61 $\pm$ 0.24</u>	2.01 $\pm$ 0.71	<u>0.96 $\pm$ 0.59</u>	1.20 $\pm$ 0.14	1.65 $\pm$ 2.47	<u>0.36 $\pm$ 0.29</u>
RCH	1.74 $\pm$ 0.49	3.92 $\pm$ 1.70	0.10 $\pm$ 0.08	<u>0.80 $\pm$ 0.26</u>	12.45 $\pm$ 25.14	1.48 $\pm$ 0.74
RW	0.65 $\pm$ 0.25	3.47 $\pm$ 1.23	1.60 $\pm$ 0.78	0.59 $\pm$ 0.35	<u>1.00 $\pm$ 1.51</u>	0.23 $\pm$ 0.18
WTH	0.59 $\pm$ 0.23	<u>1.67 $\pm$ 0.50</u>	1.11 $\pm$ 0.67	0.93 $\pm$ 0.17	3.37 $\pm$ 4.31	0.36 $\pm$ 0.26

Table 4. Cross-dataset comparison of background subtraction methods over the general background region of the pattern (equator masked). RCH is the radial convex hull baseline from Table 8; Fit is the iterative fitting (Section 2.6) without further refinement; the remaining rows combine the fitting with each framework method applied to the residual background. Values are mean $\pm$ st.dev. across images, using the normalized metrics of Section 2.3.1. Best values are highlighted in bold and second-best are underlined. The top rows of this table are constrained to a background model that must also represent the equatorial streak and are, therefore, not directly comparable with the standalone methods of Table 3, which are tuned for the equator-masked region alone; the comparable baseline here is RCH. The standalone methods are included for reference in the bottom section but these rows are excluded from best and second-best rating.

Method	Loss	S _{BG}	f _{neg,con}	f _{neg,syn}	NMSE _{syn}	1 – f _B
Fit	1.29 $\pm$ 0.79	3.57 $\pm$ 1.29	0.03 $\pm$ 0.06	0.02 $\pm$ 0.02	0.09 $\pm$ 0.13	1.17 $\pm$ 0.92
Fit+B-ILPF	0.82 $\pm$ 0.38	<u>3.68 $\pm$ 1.28</u>	1.03 $\pm$ 0.44	0.36 $\pm$ 0.14	2.12 $\pm$ 1.74	0.45 $\pm$ 0.40
Fit+G-ILPF	<u>0.81 $\pm$ 0.38</u>	3.70 $\pm$ 1.30	0.99 $\pm$ 0.45	0.35 $\pm$ 0.14	2.08 $\pm$ 1.66	<u>0.44 $\pm$ 0.40</u>
Fit+RW	0.89 $\pm$ 0.25	3.73 $\pm$ 1.30	2.22 $\pm$ 0.56	0.75 $\pm$ 0.21	13.92 $\pm$ 10.72	0.29 $\pm$ 0.31
Fit+WTH	0.80 $\pm$ 0.38	3.69 $\pm$ 1.28	0.93 $\pm$ 0.52	<u>0.32 $\pm$ 0.15</u>	<u>1.66 $\pm$ 1.34</u>	0.44 $\pm$ 0.39
RCH	1.74 $\pm$ 0.49	3.92 $\pm$ 1.70	<u>0.10 $\pm$ 0.08</u>	0.80 $\pm$ 0.26	12.45 $\pm$ 25.14	1.48 $\pm$ 0.74
B-ILPF	0.62 $\pm$ 0.24	1.99 $\pm$ 0.70	1.00 $\pm$ 0.52	1.19 $\pm$ 0.14	1.69 $\pm$ 2.73	0.36 $\pm$ 0.30
G-ILPF	0.61 $\pm$ 0.24	2.01 $\pm$ 0.71	0.96 $\pm$ 0.59	1.20 $\pm$ 0.14	1.65 $\pm$ 2.47	0.36 $\pm$ 0.29
RW	0.65 $\pm$ 0.25	3.47 $\pm$ 1.23	1.60 $\pm$ 0.78	0.59 $\pm$ 0.35	1.00 $\pm$ 1.51	0.23 $\pm$ 0.18
WTH	0.59 $\pm$ 0.23	1.67 $\pm$ 0.50	1.11 $\pm$ 0.67	0.93 $\pm$ 0.17	3.37 $\pm$ 4.31	0.36 $\pm$ 0.26

Table 5. Comparison of background subtraction methods restricted to the equatorial region, using the normalized metrics of Section 2.3.1. Best values are highlighted in bold and second-best are underlined. The loss omits the smoothness term, with weights $[0, 0.1, 0.1, 0.01, 0.8]$.

Method	Loss	S _{BG}	f _{neg,con}	f _{neg,syn}	NMSE _{syn}	1 – f _B
Fit	1.33 $\pm$ 0.56	27.09 $\pm$ 27.25	0.04 $\pm$ 0.12	0.02 $\pm$ 0.05	0.27 $\pm$ 0.86	1.66 $\pm$ 0.69
Fit+B-ILPF	1.02 $\pm$ 0.46	27.22 $\pm$ 27.22	0.68 $\pm$ 0.34	0.26 $\pm$ 0.12	0.95 $\pm$ 1.53	1.15 $\pm$ 0.58
Fit+G-ILPF	0.99 $\pm$ 0.46	27.27 $\pm$ 27.20	0.69 $\pm$ 0.30	0.27 $\pm$ 0.11	1.16 $\pm$ 1.78	<u>1.10 $\pm$ 0.56</u>
Fit+RW	1.11 $\pm$ 0.39	27.41 $\pm$ 27.18	1.66 $\pm$ 0.61	0.61 $\pm$ 0.25	10.31 $\pm$ 15.69	0.97 $\pm$ 0.55
Fit+WTH	<u>1.00 $\pm$ 0.44</u>	27.10 $\pm$ 27.25	0.63 $\pm$ 0.31	<u>0.25 $\pm$ 0.12</u>	<u>0.82 $\pm$ 1.36</u>	1.13 $\pm$ 0.54
RCH	2.07 $\pm$ 4.85	18.45 $\pm$ 9.34	<u>0.04 $\pm$ 0.03</u>	0.72 $\pm$ 0.36	67.20 $\pm$ 487.71	1.66 $\pm$ 0.48

Table 6. Weight-simplex sweep results: probability that each background-removal method ranks first, averaged over all images used in the evaluations and 1000 Dirichlet weight vectors per prior. Columns differ in the prior mean assigned to the metric weights; unpinned metrics share the remaining mass equally. The highest probability for each prior setup is in bold.

Method	$P(\text{Rank} = 1)$		
	Equal (0.2 Each)	$1 - f_B$ 0.5	$1 - f_B$ 0.5, NMSE _{syn} 0.05
AVG	0.280	0.249	0.225
RCH	0.114	0.039	0.065
CS	0.091	0.113	0.074
RW	0.196	0.261	0.211
B-ILPF	0.115	0.103	0.096
G-ILPF	0.083	0.096	0.099
WTH	0.121	0.138	0.228

2.3.2. Parameter Optimization

The loss function is minimized to find optimal parameters for each subtraction method using a coordinate-descent strategy in which parameters are optimized sequentially, with all other parameters held fixed at their current best values. Since background subtraction methods typically involve non-differentiable operations and discrete parameter choices (e.g., window size or number of smoothing cycles), gradient-based optimization is not well suited. For this reason, we use direct search strategies that rely on function evaluations, despite higher computational cost.

The protocol is fully specified by the following elements, which are reported here so that the search can be reproduced. Initial values and search ranges are fixed per method and are listed in Table A2; they are the defaults distributed with the implementation and are not adapted per image or per dataset. Parameter order is also fixed per method and is given in the same table; parameters believed to have the largest effect on the loss are visited first, and several methods revisit their parameters in a second pass, so that a parameter tuned early can be corrected once the others have moved. Step sizes follow a declining schedule, by default $\{100, 50, 25, 10, 5, 3, 1, 0.1\}$; at each stage the three candidates $v - \Delta$, v , $v + \Delta$ are evaluated and clipped to the parameter bounds. Restart occurs whenever a candidate improves the loss: the schedule is reset to the coarsest step, so that the search can leave a shallow local minimum rather than immediately refining within it. If no candidate improves the loss, the step size is reduced to the next value in the schedule. Stopping occurs when the schedule is exhausted, when an improvement smaller than the early-stopping tolerance (5×10^{-3} in normalized loss units) is obtained, when a per-parameter iteration cap of 20 is reached, or when a wall-clock budget is exceeded. Discrete parameters, such as the number of smoothing cycles or a window size in pixels, are handled by rounding candidate values to integers and by restricting the step schedule to steps that fit within the parameter bounds; the convex-hull degree, which takes values on a short discrete list, is enumerated. Since coordinate descent depends on initialization and visit order, all methods are started from the same fixed defaults in every run, so that the reported comparisons are made under identical search conditions. Sensitivity to this starting point is method-dependent (Table A11): G-ILPF and WTH are insensitive to initialization, whereas B-ILPF is sensitive, with a single favorable start outweighing the others. The default parameters for B-ILPF were adjusted accordingly, though a user may still choose to run a multi-start search. For CS and RW, multiple starts are used: one start is the default parameter vector and nine candidate parameter vectors are drawn with a stratified, Latin-hypercube-style sample of the search space: each parameter's range is split into nine equal-width strata, one value is drawn per stratum, and the per-parameter draws are shuffled and paired across parameters, so the nine candidates jointly cover the space. The loss is evaluated for

each candidate. Coordinate descent is then run to convergence from the three best-scoring candidates, and the run with the lowest final loss is kept as the result. In the experiments, the same seed was used for the multi-start vectors, so that the reported comparisons are made under identical search conditions.

Images and methods are processed in parallel to lower runtime. Additionally, the background is estimated on a downsampled pattern; it is then upsampled and subtracted from the full-resolution image to preserve diffraction signal fidelity. Downsampling is applied when it reduces the loss, or when it leaves the loss essentially unchanged while lowering the computational cost: a factor of 2 is used for the background subtraction methods G-ILPF, B-ILPF, RW, CS and WTH, and no downsampling is applied for RCH or AVG (Table A9). For WTH the two settings are practically equivalent in loss, while downsampling reduces the runtime (Table 7); the factor of 2 was, therefore, adopted for WTH as well. For RCH downsampling clearly increases the loss and is not used.

Table 7. Average runtime per image for each background subtraction method as a function of the downsampling factor (1 = no downsampling), together with the average number of objective function evaluations of the coordinate-descent search at a downsampling factor of 2. A subset of 10 images from each dataset is used for the evaluation. Times are wall-clock seconds per image on the workstation described above. The AVG baseline is not parameter-optimized and, therefore, has no search evaluations.

Method	Time Per Image (s)		Evaluations
	×1	×2	
AVG	41	41	–
CH	65	43	–
CS	1257	570	54.4
RW	576	339	81.4
SB	1337	691	63.3
SG	1317	621	39.6
WTH	438	122	15.4

2.4. Computational Environment

All experiments were run on a single workstation with an AMD Ryzen 9 5950X (16 cores/32 threads, 2.2–5.08 GHz) and 128 GB of RAM, running Ubuntu 22.04.5 LTS (Linux kernel 6.8.0). The pipeline is implemented in Python 3.10.12 and driven by shell scripts that invoke the MuscleX 2.1.0 command-line interface. Per-dataset JSON settings files specify the background-subtraction method, the masking parameters, the optimization metric weights and the downsampling factor. All computation was performed on CPU; no GPU acceleration was used. The average processing time per image per method, average number of objective function evaluations and the effects of downsampling on runtime are shown in Table 7.

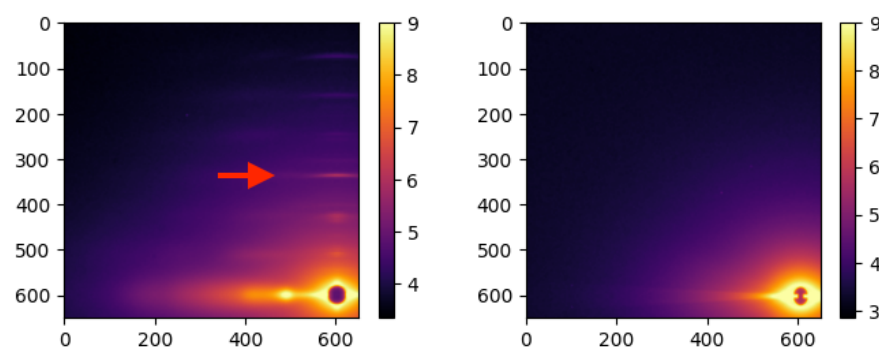
2.5. Experimental Design

We optimize a set of background subtraction algorithms used for fiber diffraction analysis, implemented in MuscleX Quadrant Folding [14], based on the CCP13 algorithms [6] or developed for previous versions of MuscleX (RCH and WTH). The tested algorithms are summarized in Table 8. The proposed framework allows evaluation of new background subtraction methods.

Table 8. Background subtraction algorithms for fiber diffraction analysis optimized and tested in this study. Mean value is used as a baseline reference without parameter optimization.

Abbreviation	Method	Notes
CS	Circularly Symmetric	Assumes radial symmetry of background
RW	Roving Window	Local averaging using sliding windows
RCH	Radial Convex Hull	Convex hull subtraction on angular projections
G-ILPF	Gaussian Iterative Low-Pass Filter	Iterative smoothing with Gaussian kernel
B-ILPF	Boxcar Iterative Low-Pass Filter	Iterative smoothing with boxcar kernel
WTH	White Top-Hat Filter	Morphological background removal
AVG	Mean Value of Non-Masked Pixels	Baseline reference; no parameter optimization

To assess robustness, we selected a set of muscle fiber diffraction datasets acquired with different detectors and under different sample environments. The data used in our study were taken on either the Mar 165 CCD detector (Rayonix Inc., Evanston, IL, USA) or the EIGER X 9M pixel array detector (Dectris AG, Baden-Dättwil, Switzerland). As shown in the example in Figure 3, images were manually filtered to exclude frames without clearly defined diffraction features. Preservation of diffraction peaks is required for meaningful evaluation of background subtraction performance. Patterns with more clearly defined features are used to show the performance of the framework on those patterns a user would want to analyze. No quantitative signal-strength proxy was used for the filtering. Table 9 lists the datasets used.

**Figure 3.** Example of included and excluded images based on manual assessment of the presence of well-defined diffraction features. Left image has clear equatorial and layer-line features, whereas right image lacks diffraction structure. Red arrow indicates the position of the third myosin layer line and meridional reflection at 14.3 nm.**Table 9.** Experimental datasets used in the analysis. Images lacking clearly defined diffraction features were excluded. A random subset was used for experiments.

Name	Muscle Type	Detector	Environment	Total	Filtered	Subselected
2025_0227	Skinned Pig Cardiac	Mar 165	Solution	685	345	50
P2_F6_639	Intact Pig Cardiac	EIGER X 9M	Solution	120	120	50
P3_F4_675	Intact Pig Cardiac	EIGER X 9M	Solution	120	120	50
2024_0508	Skinned Pig Cardiac	Mar 165	Solution	267	246	50
2024_1213	Skinned Pig Cardiac	Mar 165	Solution	242	188	50
2024_1213	Skinned Pig Cardiac	Mar 165	Air	287	237	50
2025_0402	Skinned Pig Cardiac	Mar 165	Solution	640	640	50
2026_0409	Skinned Mouse Cardiac	Mar 165	Air	304	304	50
2026_0409	Skinned Mouse Cardiac	Mar 165	Solution	325	325	50
2022_1028	Intact Mouse Skeletal	Mar 165	Solution	1282	1282	47
MPLL	Skinned Pig Cardiac	Mar 165	Solution	63	63	50
Total	–	–	–	4335	3870	547

This selection is a deliberate restriction of scope and should be read as such. The metrics that drive the optimization presuppose that a pattern with identifiable equatorial and meridional features is present: the geometry and amplitude of the synthetic features are derived from the position of the 1,0 reflection and the M1 spacing (Section 2.2.1), and the layer-line masks are placed relative to M1. On a frame with no usable diffraction structure these quantities are not defined, and the loss would not be meaningful. The results reported here, therefore, characterize behavior on patterns that a user would in practice submit for analysis; they are not evidence that the framework detects or recovers from failed exposures. Automatic rejection of unusable frames is a separate problem that the present framework does not address, the inclusion criterion was applied by visual assessment rather than by an objective threshold, and its inter-observer reproducibility was not measured (Section 4).

Because the criterion is subjective, the reported results could depend on where the boundary was drawn. We tested the framework on rejected low-quality patterns. Dataset 2025_0227 has the largest exclusion fraction and was used for this purpose. The rejected frames span a range from patterns with faint but discernible equatorial reflections to patterns with no diffraction structure at all. Only the former are meaningful inputs to the framework, since the 1,0 reflection sets the geometry and amplitude of the synthetic features (Section 2.2.1). We drew 20 frames at random from the rejected frames in which equatorial reflections could still be seen. All the user-set parameters for masking were reused from the 2025_0227 dataset; all methods were optimized with the same weight vector and search protocol used throughout. Table A14 reports the results alongside the included subset of the 2025_0227 dataset. The ordering of the methods is preserved (with only B-ILPF and CS exchanging places). The aggregate loss is not higher on the low-quality frames, but this should not be read as better subtraction: the residual-baseline term, which carries most of the weight, is easier to satisfy when there is less diffuse scattering.

Prior to background estimation, each folded pattern was smoothed with a self-guided filter, using the input as its own guide image, a window radius of 7 pixels and regularization $\varepsilon = 1$ in units of squared intensity. For some methods the pattern was then downsampled by a factor of two. The same guided filter was applied for every method, so smoothing itself is not a source of difference between them. Downsampling has a similar effect: it reduces computation and smooths the pattern. Both operate only on the input to the background estimator; the estimated background is upsampled where necessary and subtracted from the original, full-resolution, unsmoothed pattern, so no diffraction signal is lost to the preprocessing itself.

Downsampling by two was used for G-ILPF, B-ILPF, RW, CS and WTH, and was not used for RCH or AVG. This makes effective estimation resolution a method-specific setting rather than a controlled variable, and the setting was chosen using the same loss that is subsequently reported. Part of the difference between methods in Table 3 may, therefore, be attributable to preprocessing rather than to the background estimator itself; a factorial ablation separating the two was not performed (Section 4).

2.6. Iterative Two-Stage Background Subtraction

One of the challenges in background subtraction for muscle fiber diffraction is the presence of a pronounced equatorial streak, arising from aligned cylindrical structures within muscle fibers, namely the actin-containing thin filaments and myosin-containing thick filaments. The equatorial streak is characterized by high intensity and anisotropic shape. The radial convex hull (RCH) approach has been most effective at subtracting the equatorial streak and produces visually plausible results at low angles. If the goal is the analysis of low-angle data, RCH is often a good choice. However, RCH introduces artifacts at higher angles, where angular projection cones cover larger areas. Figure 4 shows an

example of the equatorial streak background subtraction using the RCH method. Combining RCH with other methods can produce plausible patterns, but may add artifacts in the transition regions.

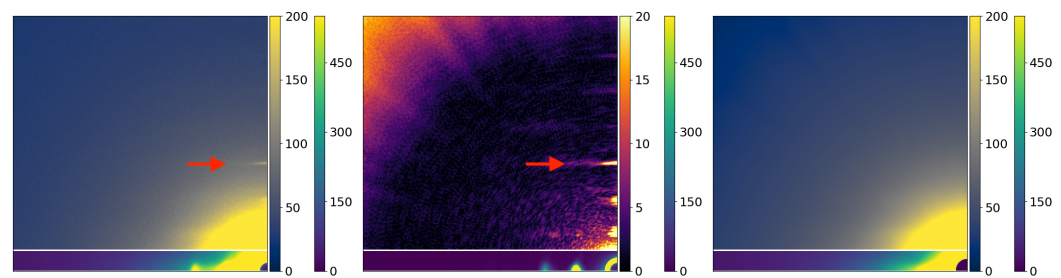


Figure 4. Example of the equatorial streak background subtraction using the RCH method. Left to right: original image, background-subtracted image using RCH, background. The equatorial streak is removed but radial streak artifacts are introduced at higher scattering angles. Two scales are used for visualization due to the large intensity differences between the equatorial streak and the rest of the pattern. Red arrow indicates the position of the third myosin layer line and meridional reflection at 14.3 nm.

In addition to the described framework, we propose an iterative procedure to jointly model the anisotropic equatorial streak and the surrounding isotropic general background. Fitting each component once is not sufficient as the components overlap and bias one another. For this reason, we iteratively refine the equatorial and general background fits, feeding each estimate into the other until they converge. This iterative fitting allows us to perform background subtraction on the whole pattern and may be combined with the optimization framework.

2.6.1. Parametric Background Modeling

Each iteration round requires an initial estimate of the general background to isolate the equatorial streak for fitting. In diffraction, scattering from randomly oriented components produces a circular pattern, which can be described in polar coordinates. In muscle fiber diffraction, the general background is nearly circularly symmetric, usually slightly elongated along the equatorial direction. The model that is used as the very initial fit combines a power-law decay with a low-order orthogonal polynomial expansion,

$$B(q) = \sum_{j=0}^3 a_j T_j(\tilde{q}) + A_p (q + q_0)^{-p}, \quad (10)$$

where T_j denotes Chebyshev polynomials and q is the radial coordinate. Parameters are estimated using least-squares fitting to radial intensity projections between angular ranges 0–5°, 5–10°, and 10–60° from the meridian. The two-dimensional background is then interpolated and subtracted from the original image. This step ensures that the subsequent modeling of the equatorial streak is not biased by the isotropic background. This projection-based estimate, B_0 , is used only in the first round of the iterative refinement described in the next section.

2.6.2. Equatorial Streak Modeling

Parametric modeling of equatorial streaks has been developed for polymer diffraction data using elliptical coordinates and Pearson VII functions [16]. Both polymer and biological samples with rod-like particles produce streak-like diffraction features in diffraction patterns, so the same model may be applied to muscle fiber diffraction. We model the equatorial streak with the elliptical-coordinate model [16] and fit it by least squares.

2.6.3. Iterative Refinement

Figure 5 summarizes the fitting procedure. In each iteration, we fit the equatorial streak E_k on $\text{img} - G_{k-1}$ and the general background G_k on $\text{img} - E_k$, using the result of the previous step. The general background is parameterized with three components: exponential decay, constant baseline, and an additional term chosen per image from Lorentzian, power-law, or stretched-exponential. The form that reduces oversubtraction is selected for each image, since no single form was found to be best across all datasets. An adaptive approach works best for an accurate background fit. The functional form can instead be kept fixed to study how the background shape changes with experimental conditions. The iterative refinement is done on smaller, downsampled images to speed up calculations. An example of the fitted backgrounds is shown in Figure 6.

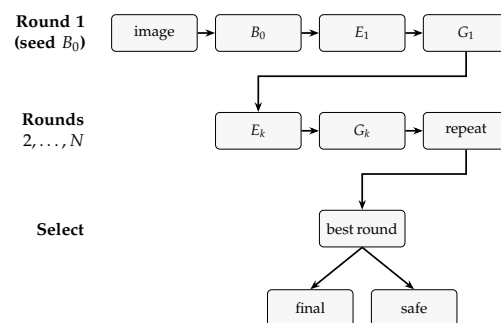


Figure 5. Iterative background fitting procedure. Round 1 is seeded by the projection-based general background B_0 , giving the first equatorial fit E_1 and general background fit G_1 . In rounds $2, \dots, N$, E_k is fitted on $\text{img} - G_{k-1}$ and G_k on $\text{img} - E_k$. The best round is selected using an oversubtraction-based criterion, giving the final residual $\text{img} - E - G$ (“final”) and a scaled version with the background scaled to reduce oversubtraction if present (“safe”).

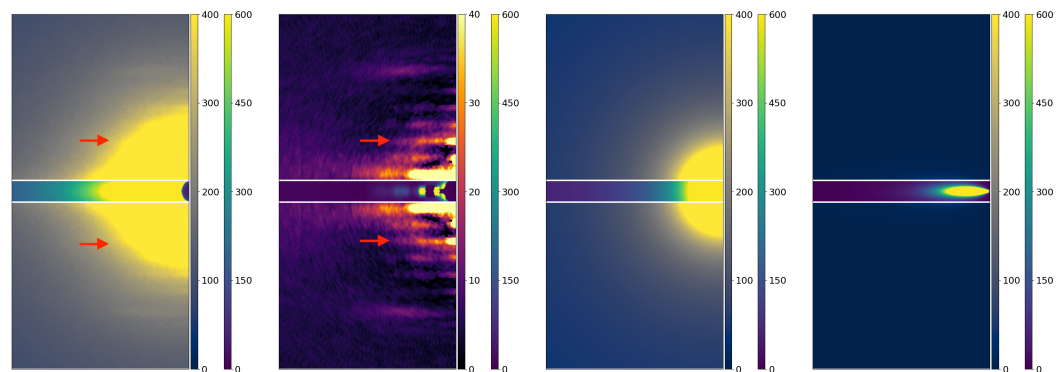


Figure 6. Examples of parametric background fitting results. Left to right: original image before subtraction, background-subtracted pattern, fitted general background, and fitted equatorial streak. Multiple scales are used for visualization due to high intensity differences between the equatorial streak and the rest of the pattern. Most of the background is subtracted with some high-angle residual background left over that can be removed by a final pass of a non-parametric method. Red arrow indicates the position of the third myosin layer line and meridional reflection at 14.3 nm.

3. Results

This section summarizes the performance of the framework in finding an optimal solution to the problem of background subtraction given the lack of ground-truth background. The goal is to demonstrate how the framework provides a consistent way to compare and analyze methods using the aggregate loss function and the individual metrics, with lower values indicating better performance.

We first evaluate representative visual examples. Then, we demonstrate consistency of the outputs, followed by a comparison of the methods across datasets and analysis of optimized parameters. Finally, we evaluate the iterative fitting combined with the optimization framework, both visually and using the defined metrics.

3.1. Visual Correspondence with Quantitative Metrics

Figure 7 illustrates representative examples corresponding to low-loss background subtraction outcomes for images from three different datasets. The diffuse background is removed whereas diffraction features are retained and the residual baseline is smooth. Figure A2 shows an example of background subtraction outcomes for different methods applied to the same pattern, allowing us to appreciate the differences in methods and performance. In the examples with higher loss values, there is either incomplete background removal with remaining residual intensity between layer lines or excessive background removal.

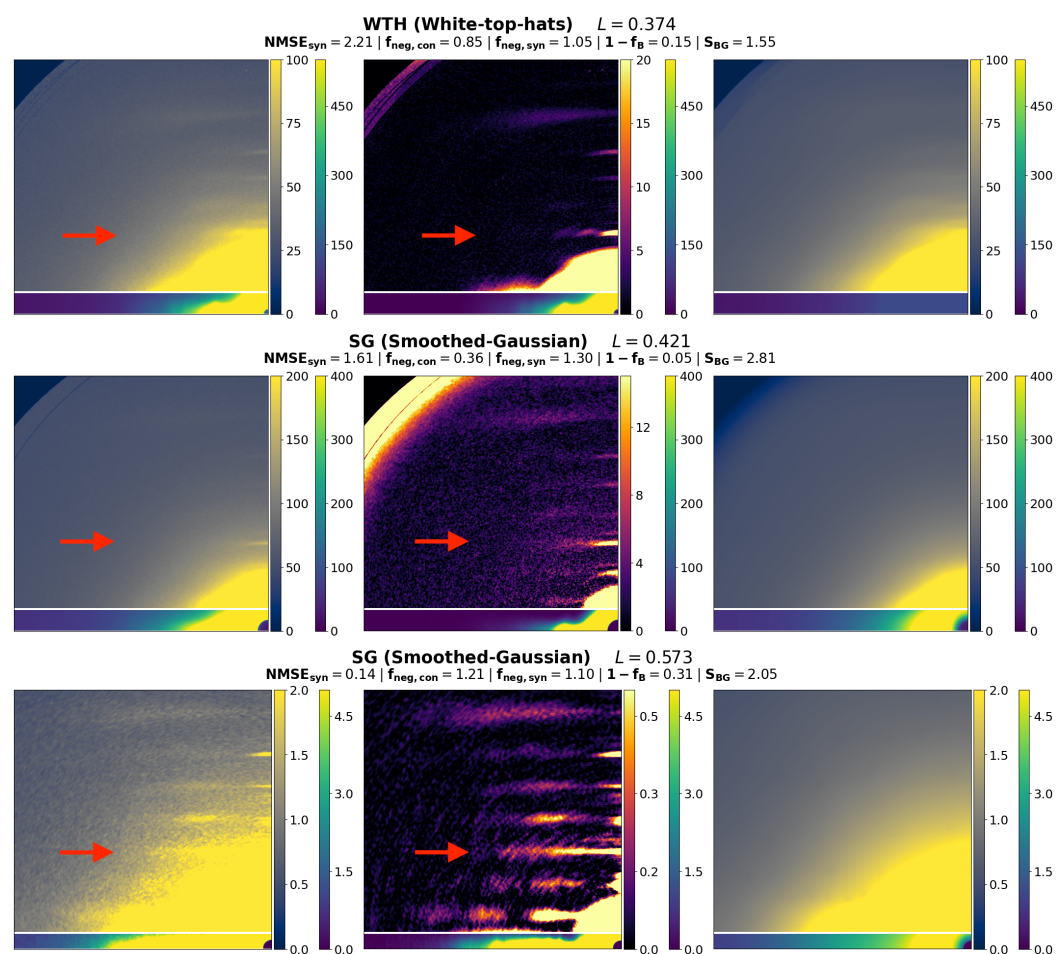


Figure 7. Examples of optimized background subtraction outcomes for different images. Each row is a different pattern from three datasets. Left to right: original image before subtraction, background-subtracted pattern, and background. Multiple scales are used for visualization due to high intensity differences between the background, the equatorial streak and the rest of the pattern. The selected method and parameters effectively remove the background while preserving faint features for both high-intensity patterns in the top two rows and the low-intensity pattern in the bottom row. Red arrow indicates the position of the third myosin layer line and meridional reflection at 14.3 nm.

3.2. Convergence and Consistency of Optimization

Figure 8 shows the low average dispersion of the integrated background intensity per dataset, which is expected to be consistent within datasets, since patterns are taken with

the same exposure time. The caption shows that although the best-performing method varies with dataset, WTH is ranked first in most datasets, with the remaining datasets split among G-ILPF, RW, and AVG. The search is stable in the sense that the methods converge to comparable loss values from their default starting points, and it terminates either when no candidate step improves the loss or when the improvement falls below the early-stopping threshold (Section 2.3.2). Table A10 confirms that this optimization reduces the loss relative to the default parameters in nearly every dataset and method. Although this reflects successful minimization of the loss, this should not be interpreted as an independent evidence of an optimal background subtraction in absolute terms.

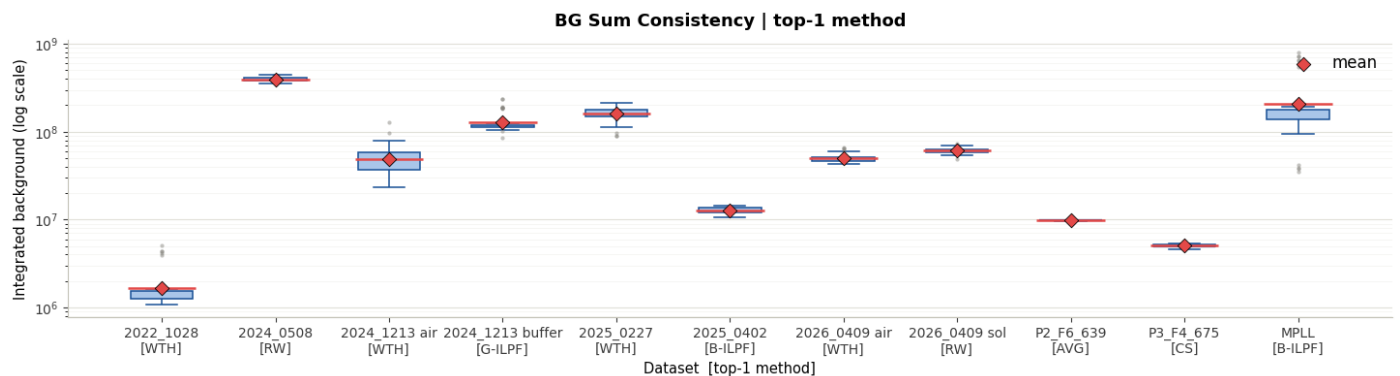


Figure 8. Consistency of optimization results within datasets. Distribution of average integrated background intensity per dataset. The background intensity is consistent within datasets. Best-performing method per dataset is shown in brackets.

3.3. Quantitative Comparison of Methods

Table 3 reports the aggregate losses and metrics across datasets. Lower values indicate better performance. Under the weight vector used here, the WTH approach resulted in the lowest average loss, followed by the iterative low-pass filtering approaches. The margins between the leading methods are small relative to their spread across images, and the images are grouped into datasets rather than sampled independently, so we present this ordering as indicative rather than as a demonstration that one estimator is superior (Section 4). CS and RCH methods do not perform well due to the synthetic data distortion and background roughness. Since the equatorial streak is masked out, the evaluation focuses on high-angle background, where the RCH method introduces artifacts. The average baseline confirms that spatially structured background modeling is essential for reliable subtraction. The AVG method is expected to perform poorly on patterns with high background intensity; however, it may do well on low-intensity patterns with little background. The individual metrics show trade-offs between methods. For example, G-ILPF and WTH approaches achieve low feature oversubtraction, while the CS method tends to oversubtract signal. Two of the columns are won by the AVG baseline for structural rather than substantive reasons: a constant background has zero vertical derivative, so S_{BG} vanishes by construction, and $NMSE_{syn}$ is invariant to a constant offset, so a uniform subtraction leaves the injected features undistorted. This illustrates directly why no single metric is enough and a compound criterion is needed (see Figure A1). Table A3 reports the statistical comparison at the dataset level, and Table A4 the corresponding per-dataset win rates, i.e., the share of images on which each method attains the lowest loss.

Table 10 shows the per-term loss decomposition. The ranking is affected by two terms much stronger than others: $\lambda_5(1 - f_B)$ and $\lambda_1 S_{BG}$ together account for ~60–90% of every row. The margins between the two leading methods, WTH and G-ILPF, are small and

depending on the user-selected weight vector, as is described in Section 2.3.1. The AVG baseline wins on $\lambda_1 S_{BG}$ by construction but loses on $\lambda_2 f_{neg,con}$ more than other methods.

Table 10. Per-term decomposition of the aggregate loss of Table 3. Each cell gives the weighted contribution $\lambda_i \bar{x}_i$ of one normalized metric to the mean loss, with its percentage share of the row total beneath it; the weights are those of Equation (8), $[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] = [0.1, 0.05, 0.05, 0.01, 0.8]$, applied to all methods. Rows are ordered by total loss.

Method	$\lambda_1 S_{BG}$	$\lambda_2 f_{neg,con}$	$\lambda_3 f_{neg,syn}$	$\lambda_4 NMSE_{syn}$	$\lambda_5 (1 - f_B)$	Total
AVG	0.000 (0.0%)	0.201 (27.7%)	0.067 (9.2%)	0.009 (1.2%)	0.450 (61.9%)	0.727
B-ILPF	0.199 (32.2%)	0.050 (8.1%)	0.059 (9.6%)	0.017 (2.7%)	0.291 (47.3%)	0.616
CS	0.224 (33.5%)	0.095 (14.1%)	0.048 (7.2%)	0.018 (2.6%)	0.284 (42.5%)	0.669
G-ILPF	0.201 (33.0%)	0.048 (7.9%)	0.060 (9.8%)	0.017 (2.7%)	0.284 (46.6%)	0.610
RCH	0.392 (22.5%)	0.005 (0.3%)	0.040 (2.3%)	0.124 (7.1%)	1.181 (67.8%)	1.743
RW	0.347 (53.4%)	0.080 (12.3%)	0.029 (4.5%)	0.010 (1.5%)	0.184 (28.3%)	0.651
WTH	0.167 (28.1%)	0.055 (9.3%)	0.047 (7.8%)	0.034 (5.7%)	0.291 (49.0%)	0.594

3.4. Optimized Parameter Variability

Variability of the optimized parameters for each method is summarized in Table 11. The moderate variability across images, with only three of the eighteen parameters exceeding a coefficient of variation of one, reflects differences in experimental conditions and sample preparation. Optimal parameters depend on dataset-specific characteristics: the mean coefficient of variation computed within datasets (0.46) is lower than that computed across all images (0.61), although the difference is modest and no formal test of it was carried out. This observation suggests that further improvements may be achieved by clustering diffraction patterns and selecting parameters per cluster. The variability may also be due to parameter degeneracy, since the loss and integrated background intensity do not vary as much. In general, the optimized values are not tightly determined, and the reduction in variability within datasets is consistent in direction across every method and parameter, which supports dataset- or image-subset-specific strategies.

Table 11. Coefficient variability (st.dev./mean) across images and datasets. Even though variability is lower within datasets, adaptive dataset-specific and image-subset-specific tuning should be considered. No inferential test of the difference between the two columns was performed.

Method	Parameter	Across-Images CV	Between-Datasets CV	Difference
RCH	degree	0.00	0.00	0.00
CS	smooth	1.25	1.18	−0.07
CS	radial bin	1.20	1.12	−0.08
CS	cirmin	0.89	0.57	−0.32
CS	cirmax	0.38	0.29	−0.08
RW	cirmin	0.95	0.65	−0.30
RW	cirmax	0.68	0.50	−0.18
RW	window sep _x	0.57	0.35	−0.22
RW	window sep _y	0.34	0.22	−0.11
RW	tension	0.70	0.52	−0.18
RW	window size _x	1.11	0.54	−0.57
RW	window size _y	0.49	0.35	−0.14
B-ILPF	cycles	0.47	0.47	0.00
B-ILPF	boxcar _x	0.35	0.24	−0.11
B-ILPF	boxcar _y	0.31	0.14	−0.17
G-ILPF	FWHM	0.57	0.52	−0.05
G-ILPF	cycles	0.44	0.39	−0.05
WTH	tophat	0.29	0.23	−0.06
Mean		0.61	0.46	−0.15

3.5. Evaluation of Iterative Background Fitting

The iterative fitting of the equatorial streak E and the general background G produces a residual, $\text{img} - E - G$. We evaluate this approach in both the high-angle region with the equator masked (same as in the previous section) and additionally in the equatorial region. For the equatorial region, the loss is computed with the smoothness term dropped ($\lambda_1 = 0$), since the background is expected to be steep. The weight removed from the smoothness term is reallocated to the two oversubtraction terms, $f_{\text{neg,con}}$ and $f_{\text{neg,syn}}$, resulting in $[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] = [0, 0.1, 0.1, 0.01, 0.8]$. The general background region is evaluated in the same way as in the previous section.

An important advantage of the iterative fitting is the ability to subtract the background on the whole pattern. Examples of backgrounds produced by the iterative fitting are shown in Figure 9.

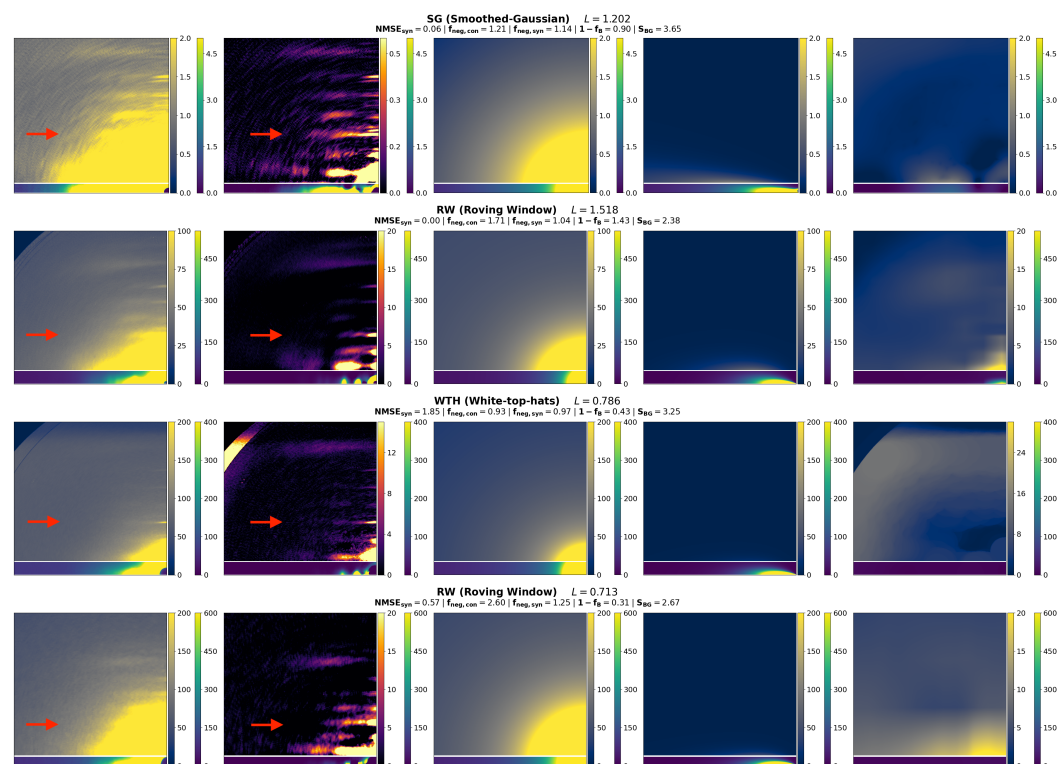


Figure 9. Examples of successful background subtraction outcomes for different patterns with fitted equatorial streaks. Each row is a different pattern from four datasets. Left to right: original image before subtraction, background-subtracted pattern, fitted general background, fitted equatorial streak, and the non-parametric background for the remaining residual. Multiple scales are used for visualization due to high intensity differences between the equatorial streak and the rest of the pattern. The equatorial streak and the isotropic general background are removed; the remaining low-intensity smooth background is removed using a non-parametric method. Red arrow indicates the position of the third myosin layer line and meridional reflection at 14.3 nm.

3.5.1. General Background Comparison

Table 4 reports the resulting metrics over the same region as in the previous analysis (beam center, detector edges and the equator region masked) for the radial convex hull baseline (RCH), for the iterative fitting alone (Fit), and for the iterative fitting combined with each framework method applied to the residual general background (Fit+G-ILPF, Fit+WTH, Fit+B-ILPF, Fit+RW). Table A7 reports the statistical comparison at the dataset level.

The iterative fitting results in low oversubtraction and produces the smoothest background, but leaves the highest residual baseline, which is expected by design. Applying

non-parametric methods using the optimization framework to the residual reduces the loss: all combinations outperform the RCH baseline.

The losses in Table 4 are higher than those reported in Table 3 for the corresponding standalone methods, and the two tables should not be read as a like-for-like ranking. Although both are evaluated over the equator-masked region, the rows of Table 4 are constrained to a background model that must simultaneously account for the equatorial streak, whereas the standalone methods of Table 3 are free to tune their parameters for the equator-masked region alone and are never required to represent the streak at all. The appropriate baseline within Table 4 is, therefore, RCH, the only other approach evaluated here that subtracts background across the whole pattern; all four combinations improve on it. The benefit of the two-stage approach is the additional coverage rather than a lower value of a loss that is computed only where the non-parametric methods already operate. Therefore, the iterative fit removes the high-intensity anisotropic equatorial streak and the isotropic background, and a slowly varying residual that remains can be removed with a non-parametric method, producing a cleaner background than RCH alone or the iterative fit alone.

3.5.2. Equatorial-Region Comparison

To evaluate the equatorial streak fit, Table 5 shows the same methods evaluated on the equatorial region. However, in the aggregate loss the smoothness term is dropped. Table A8 reports the statistical comparison at the dataset level.

In the equatorial region, the iterative fit achieves a comparable loss to the RCH baseline with far lower variation, suggesting that RCH's performance varies across images whereas the parametric fit is more stable. Similarly to the general background comparison, applying a non-parametric method to the residual after the fit reduces the loss further.

4. Discussion

The results show that the proposed optimization framework provides a consistent and reproducible basis for comparing background subtraction methods across datasets. Since the quality of background subtraction is evaluated with quantitative metrics, different methods can be placed on a common scale. In the cases we examined, patterns that a practitioner would judge to be well subtracted receive a low aggregate loss, and patterns with visible residual structure or oversubtraction receive a high one (Section 3.1). We did not carry out a blinded rater study, so this is a qualitative correspondence rather than a measured agreement with expert judgment, and the weights of the loss were themselves chosen with reference to visual inspection (Section 4). What the framework does establish is reproducibility: the same images, metrics and weights yield the same ranking and the same parameters, which is not the case when an operator tunes parameters by eye.

No single background subtraction method outperforms the others. Under the weights used here the white top-hat method has the lowest average loss; however, the iterative low-pass methods and the roving window method are competitive, and the separation between them is small compared with the spread across images. Individual metrics are not sufficient to guide optimization (Figure A1), which emphasizes the need for a combined metric.

The parametric and non-parametric approaches work together. The iterative two-stage fit removes the anisotropic equatorial streak and most of the general background, but intentionally leaves a residual baseline. A non-parametric method then removes the remaining slowly varying component. Among the approaches evaluated here that subtract background across the whole pattern, this combination gives the lowest loss in both the general-background region and the equatorial region, improving on the RCH baseline and on the parametric fit alone (Tables 4 and 5). It does not give a lower loss than the

best non-parametric method applied on its own to the equator-masked region (Table 3), and we do not claim that it does. The two are not alternatives for the same task: the non-parametric methods do not model the equatorial streak and are evaluated only where it has been masked out, so they are optimized against a strictly easier problem. The loss as defined here rewards a low, smooth residual over the region on which it is computed and gives no credit for the additional coverage that the fit provides, which is the principal benefit of the two-stage approach. Extending reliable subtraction to the whole pattern is an improvement over existing approaches, which usually work well only on either the high- or the low-angle region. A more accurate separation of background should allow more accurate analysis of two-dimensional diffraction features and could contribute to future studies of how background structure changes with experimental conditions. We have also tested whether it improves an established diffraction measurement: propagating three background treatments through the MuscleX Equator module (Table A12) leaves the fitted peak positions, and hence $d_{1,0}$, unchanged, but shifts the integrated intensities and the $I_{1,1}/I_{1,0}$ ratio, most strongly for the two-stage fit. A lower loss should, therefore, not be read as implying a more accurate downstream measurement, and a reported intensity ratio should state which background treatment produced it (Section 4).

The appropriate choice of background subtraction strategy still depends on the data analysis goal: for low angles only, the RCH method is often sufficient and robust; for quantification of layer lines and higher-angle meridionals, the equator may be masked and the optimization framework applied to the rest of the pattern; for background subtraction across the whole pattern, the iterative fitting approach combined with the framework may be used.

Limitations

The evidence presented here is narrower than the framework's eventual aim, and several qualifications should be stated plainly.

The loss weights were chosen by inspecting background-subtraction results, and visual inspection is still used to detect artifacts that the loss does not capture, so agreement between the aggregate loss and visual assessment does not represent independent validation. A properly independent test could determine the weights from a calibration set, apply them unchanged to unseen datasets scored by several blinded raters, and report rank correlation together with inter-rater agreement. No such study is presented here due to the unavailability of a pool of independent expert raters. The ranking is likewise conditional on the weights: re-scoring the cross-dataset comparison with the alternative equatorial weighting [0, 0.1, 0.1, 0.01, 0.8] places RW first and G-ILPF second, ahead of B-ILPF and WTH; a sensitivity analysis over the weight simplex is presented in Section 2.3.1, but results it reports remain conditional on the choice of weight mean priors. Normalization constants are chosen based on the data that overlaps evaluation data and parameters are selected by minimizing the loss and the optimized methods are then compared using that same loss on the same images. Due to this, the reported values are in-sample; a held-out image set, or leave-one-dataset-out parameter selection and normalization, would appear to be a stronger design; however, adaptation to the characteristics of specific datasets is an intended feature of the method, rather than being concerned about overfitting.

More fundamentally, the framework evaluates algorithms through surrogate properties expected of a satisfactory residual, so lower values after optimization demonstrate that the objective was minimized and not that the estimated background is physically correct. Table A13 demonstrates this: cosmetic post-processing—a constant offset or blurring the background—lowers the loss on roughly half the images, confirming that a lower score is not on its own a certificate of accuracy. Establishing absolute accuracy would require a

controlled dataset with a known background, built either from measured blank patterns combined with independently measured or simulated diffraction features, or from a forward physical simulation with separately known scattering components. These approaches were not employed since blank patterns lack the diffuse scattering backgrounds coming from the sample and representative forward physical simulations are not available. A lower loss also does not, by itself, imply a better measurement of spacings and intensities of diffraction features. Propagating three background treatments through the MuscleX Equator module on one of the datasets (Table A12) shows that fitted peak positions, and hence the $d_{1,0}$ lattice spacing, are insensitive to the treatment, whereas integrated intensities and the $I_{1,1}/I_{1,0}$ ratio are not. A reported intensity ratio must, therefore, be accompanied by a statement of the background treatment used.

The individual criteria are also partial. The injected features are elongated Gaussians on a regular grid, placed away from layer lines, so they probe preservation of weak isolated signal rather than of the continuous layer-line structure that motivates two-dimensional subtraction in the first place; randomized positions and orientations were not tested, and an interaction between the regular grid and the spatial filters being compared cannot be excluded. S_{BG} uses the vertical derivative alone, so a background with substantial horizontal structure but small vertical variation would escape penalty, and no comparison against a two-dimensional gradient measure was made. The reasons for such approach are discussed in the Methods.

The experimental comparison carries its own qualifications. It rests on means and standard deviations pooled across images that belong to eleven acquisitions differing in muscle type, detector and environment, and which are, therefore, not independent units; Table A5 reports dataset-stratified paired differences and Table A6 a hierarchical bootstrap treating datasets, not images, as the unit of resampling, and under both analyses small differences between the leading methods should not be read as evidence of superiority. Preprocessing is partly confounded with the estimator, since downsampling was applied to five of the seven methods and was selected using the same loss that is subsequently reported. Frames without clearly defined diffraction features were excluded by visual assessment before evaluation. There was no attempt to measure inter-observer reproducibility of this criterion. Performance near the inclusion boundary is characterized on 20 frames drawn at random from the rejected frames of the dataset with the largest exclusion fraction in which equatorial reflections remained faintly discernible (Table A14); the method ordering is preserved on them. This analysis bounds the dependence of the reported results on the inclusion criterion. Optimized parameters vary considerably across datasets and, to a lesser extent, within them (Table 11), and we cannot presently separate genuine dependence on experimental conditions from parameter degeneracy, in which many parameter combinations give a similar loss. Finally, two recent alternatives, polar-coordinate background removal [11] and learned background estimation [13], were not included in the benchmark, and the reason in each case is that their assumptions do not transfer to the patterns studied here. The polar-coordinate method is designed for two-dimensional scattering patterns from samples without preferred orientation. Such patterns are isotropic and centrally symmetric about the beam and contain sharp rings. Fiber diffraction patterns from oriented specimens violate both parts of this description. The learned approach is our own earlier work and was trained in a supervised manner on target backgrounds produced by the same non-parametric estimators that are assessed in this study (Section 1). Its output is, therefore, bounded by the quality of those estimates and reproduces their systematic errors, so it cannot serve as an independent baseline for evaluating them.

Given these limitations, we recommend a hybrid workflow rather than unattended automation: run the automated search, verify the results on a small subset of images,

adjust the weights if a systematic artifact is visible, and then process the full dataset. The framework automates the parameter search and makes the criterion explicit and reproducible; it does not remove the operator from the loop.

5. Conclusions

We present a quantitative framework to evaluate and optimize background subtraction methods in muscle fiber diffraction images. It reduces reliance on subjective visual assessment by providing reproducible metrics, validation using synthetic signal, and an automated parameter search; operator verification of the outcome remains part of the recommended workflow. Out of the tested non-parametric background subtraction methods, no single method outperformed the others, which is the central argument for a framework that selects and tunes methods per dataset.

Combining the iterative fitting procedure with a non-parametric method extends background subtraction across the whole pattern, including the equatorial streak region, and improves on the RCH baseline, which is the comparable whole-pattern approach. The parametric and non-parametric approaches address different failure modes and are effective in combination.

Overall, the proposed framework provides a quantitative and reproducible basis for background subtraction in muscle fiber diffraction. Although the process is not fully automated, it is supported by a quantitative metric for evaluating subtraction quality. The new approach makes any required manual adjustment more intuitive, since the parameters have clear physical meaning.

The framework and the iterative fitting have been integrated into MuscleX version 2.1.0 [17], where background subtraction may be applied with parameter optimization and quantitative evaluation, allowing users to review results and make adjustments (see Supplementary Materials).

What has been demonstrated here is specific to small-angle X-ray fiber diffraction from striated muscle: eleven acquisitions from two detector families, predominantly from pig cardiac muscle, in air and in solution. The metrics and the optimization procedure themselves rest on general properties of signal and background rather than on muscle-specific assumptions, so extension to other diffraction geometries is plausible. That extension has not been tested here, and the synthetic-feature construction in particular would have to be redesigned for patterns whose features are not organized around an equator and a meridian.

Supplementary Materials: Code for the optimization framework and the evaluation metrics is available at <https://github.com/biocatiit/muscleX> (accessed on 23 September 2026).

Author Contributions: Conceptualization and methodology, G.A., T.I. and I.K.; software, I.K. and G.A.; validation, T.I.; formal analysis, G.A., T.I. and I.K.; resources, T.I.; data curation, T.I.; writing—original draft preparation, I.K.; writing—review and editing, G.A. and T.I.; visualization, I.K. and T.I.; supervision, G.A. and T.I.; funding acquisition, T.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the US National Institutes of Health grant number R01GM144555 (TI).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The source code for the optimization framework and the evaluation metrics is part of MuscleX and is available at <https://github.com/biocatiit/muscleX>; the release used in this study is archived at <https://doi.org/10.5281/zenodo.21709989> [17]. Sample patterns from the data used in this study and the corresponding metrics are available at CC-BY license at <https://doi.org/10.5281/zenodo.21709060>. Any additional data are available on request from the corresponding author.

Acknowledgments: During the preparation of this manuscript, the author(s) used Claude Opus 5 (Anthropic) for language editing of the text. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AVG	Mean value of non-masked pixels (baseline)
B-ILPF	Boxcar iterative low-pass filter
CCD	Charge-coupled device
CS	Circularly symmetric
CV	Coefficient of variation
FWHM	Full width at half maximum
G-ILPF	Gaussian iterative low-pass filter
LL	Layer line
NMSE	Normalized mean squared error
RCH	Radial convex hull
RW	Roving window
WTH	White top-hat filter

Appendix A

Table A1 shows sensitivity to the synthetic signal parameters. The small variation in loss reflects that the framework is not sensitive to the choice of synthetic signal parameters.

Table A1. Sensitivity of the evaluation framework to variations in synthetic signal parameters. (σ_x, σ_y) are quoted as the divisors of Equation (7), so that $(S_{1,0}/5, M1/10)$ denotes the default setting. Values are reported as mean $\pm$ standard deviation across datasets, using the normalized metrics of Section 2.3.1. The purpose of this table is to demonstrate robustness of the aggregate loss to reasonable changes in synthetic feature configuration.

Samp. Freq.	(σ_x, σ_y)	NMSE _{syn}	f _{neg,con}	f _{neg,syn}	1 − f _B	S _{BG}
dense	$(S_{1,0}/5, M1/10)$	5.10 $\pm$ 4.57	0.904 $\pm$ 0.65	1.092 $\pm$ 0.39	0.245 $\pm$ 0.12	1.886 $\pm$ 0.33
medium	$(S_{1,0}/5, M1/5)$	4.17 $\pm$ 4.21	0.894 $\pm$ 0.66	1.070 $\pm$ 0.41	0.260 $\pm$ 0.14	1.819 $\pm$ 0.26
medium	$(S_{1,0}/1, M1/10)$	4.30 $\pm$ 3.67	0.942 $\pm$ 0.65	1.042 $\pm$ 0.20	0.240 $\pm$ 0.11	1.911 $\pm$ 0.28
medium	$(S_{1,0}/3, M1/10)$	5.36 $\pm$ 5.22	0.900 $\pm$ 0.67	1.094 $\pm$ 0.35	0.266 $\pm$ 0.17	1.837 $\pm$ 0.19
medium	$(S_{1,0}/7, M1/10)$	3.70 $\pm$ 4.47	0.906 $\pm$ 0.67	1.006 $\pm$ 0.40	0.260 $\pm$ 0.17	1.869 $\pm$ 0.20
medium	$(S_{1,0}/5, M1/10)$	3.61 $\pm$ 3.52	0.906 $\pm$ 0.67	1.054 $\pm$ 0.40	0.255 $\pm$ 0.15	1.850 $\pm$ 0.20
medium	$(S_{1,0}/5, M1/20)$	3.33 $\pm$ 3.41	0.908 $\pm$ 0.67	1.054 $\pm$ 0.40	0.253 $\pm$ 0.15	1.859 $\pm$ 0.20
sparse	$(S_{1,0}/5, M1/10)$	3.61 $\pm$ 3.52	0.906 $\pm$ 0.67	1.054 $\pm$ 0.40	0.255 $\pm$ 0.15	1.850 $\pm$ 0.20

Table A2 gives the search configuration used for the coordinate-descent optimization (Section 2.3.2). Initial values and ranges are the defaults distributed with the implementation and are identical for every image and dataset.

Table A2. Coordinate-descent search configuration. “Visit positions” gives the positions in the fixed sweep at which each parameter is optimized; parameters appearing twice are revisited in a second pass after the remaining parameters have moved. The step-size schedule ($\{100, 50, 25, 10, 5, 3, 1, 0.1\}$, restarted at the coarsest value on every improvement), the early-stopping tolerance (5×10^{-3}) and the per-parameter iteration cap (20) are shared by all methods. The CS tension parameter is held fixed and the RW smoothing parameter has a degenerate range; both are retained so that the parameter vectors match the implementation.

Method	Parameter	Initial	Search Range	Visit Positions
RCH	degree	1	[0.5, 2]	1
WTH	tophat size	50	[1, 200]	1
G-ILPF	FWHM	15	[2, 150]	1, 3
	cycles	150	[2, 500]	2, 4
B-ILPF	cycles	150	[5, 500]	1, 4
	boxcar _x	15	[2, 150]	2, 5
	boxcar _y	15	[2, 150]	3, 6
CS	smooth	0.1	[0.1, 50]	1, 5
	radial bin	10	[2, 200]	2, 7
	cirmax	25	[5, 50]	3, 8
	cirmin	0	[0, 5]	4
	tension	1	fixed	6
RW	window size _x	10	[5, 100]	1, 9
	window size _y	10	[5, 100]	2, 10
	window sep _x	40	[2, 50]	3, 11
	window sep _y	40	[2, 50]	4, 12
	cirmax	25	[5, 50]	5, 13
	smooth	0	[0, 0]	6, 14
	tension	1	[1, 5]	7
	cirmin	0	[0, 5]	8

Figure A1 shows an example of background subtraction outcomes for a single image from dataset 2025_0227 when guided using metrics individually or a combination of the two most significant. No metric is sufficient to guide optimization.

Figure A2 shows an example of background subtraction outcomes for different methods on a single image from dataset 2025_0227.

Figure A3 shows the per-dataset probabilities of different methods scoring first across 1000 Dirichlet weight vectors per prior. Table A3 breaks the aggregate loss down by dataset and Table A4 reports the matching win rates at the fixed weight vector. The two views agree on which method leads in each dataset. Table A5 treats the dataset, rather than the image, as the unit of analysis and tests paired differences against WTH, and Table A6 reports a hierarchical bootstrap that resamples datasets and, within each, their images, to give a rank-1 probability that accounts for the non-independence of images within a dataset.

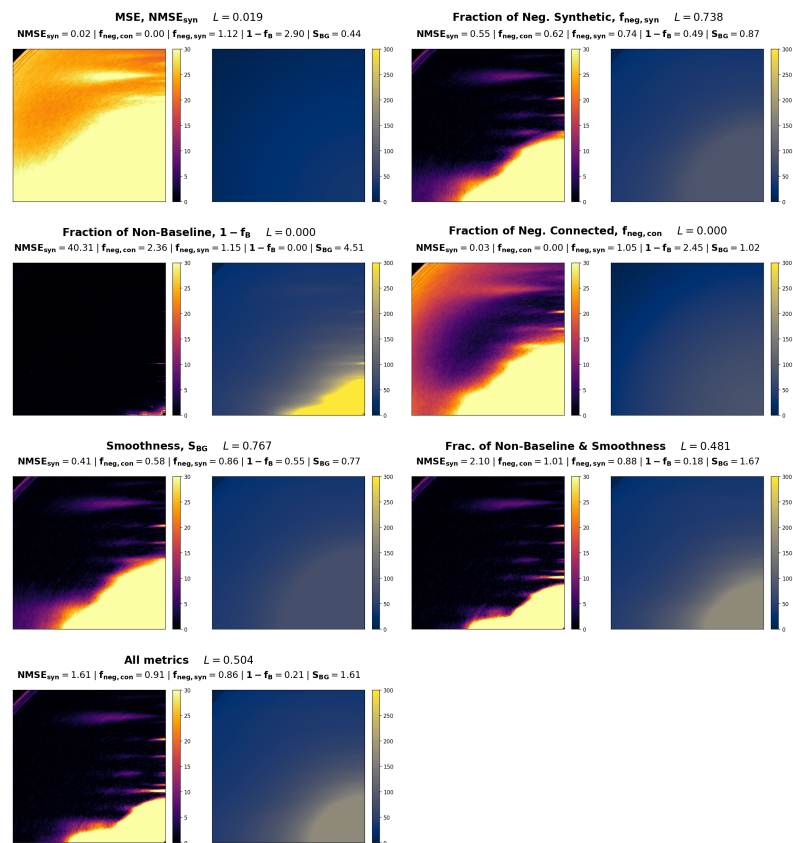


Figure A1. Image sample from dataset 2025_0227 when guided using metrics individually or a combination of the two most significant. Left of each pair: Background removed image. Right: Background.

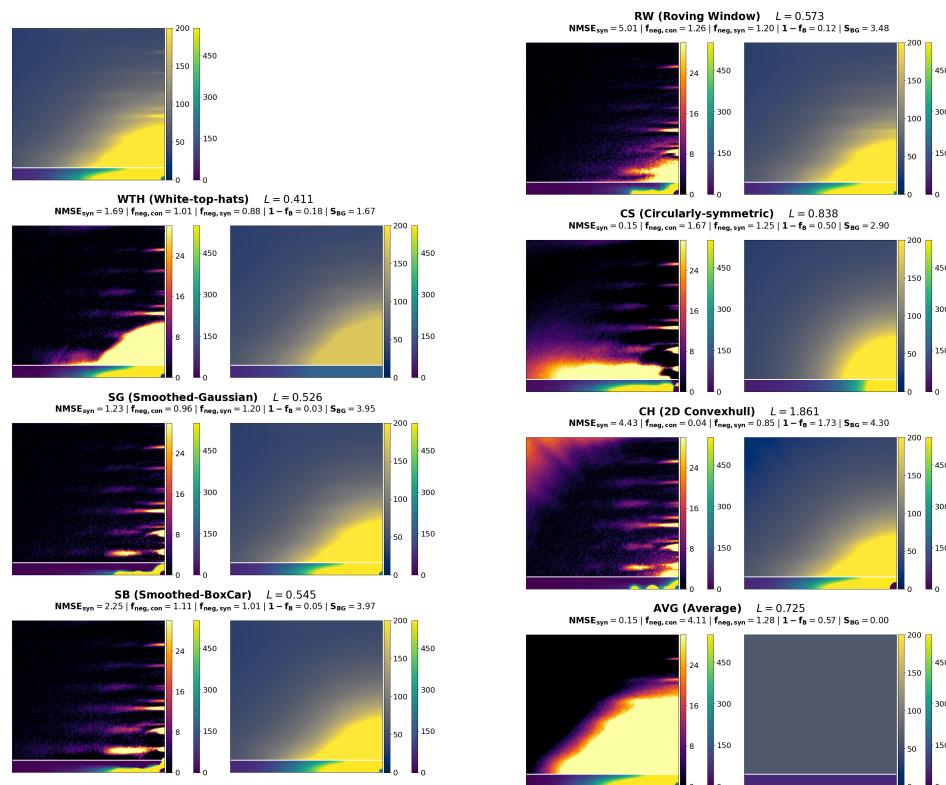


Figure A2. Visual comparison of background subtraction outcomes for image from dataset 2025_0402. Top left: original image. Rest: background-subtracted pattern (left), background (right). Two scales are used for visualization due to high intensity differences between the background, the equatorial streak and the rest of the pattern.

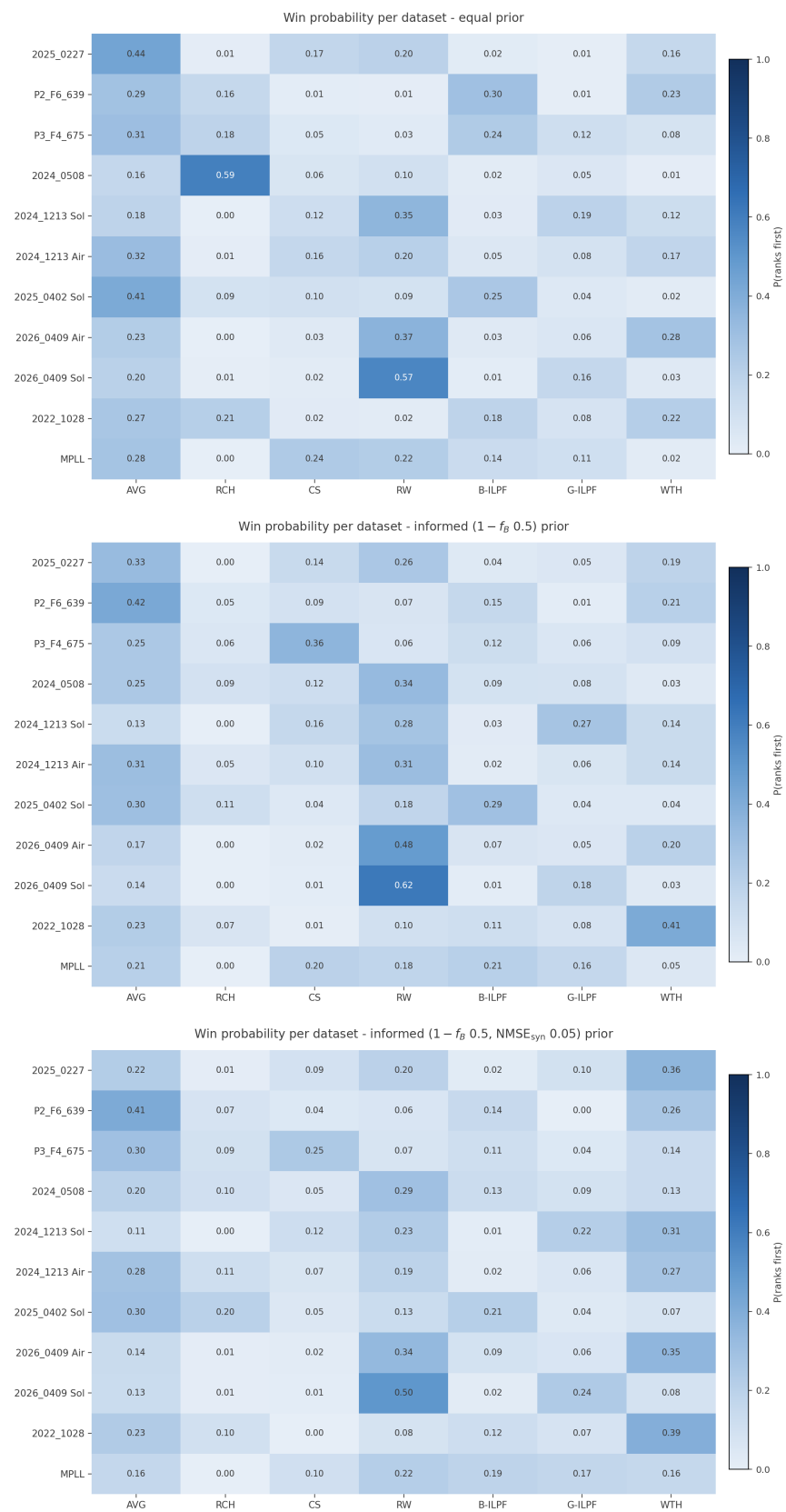


Figure A3. Per-dataset win probability over the weight simplex under three metric-weight priors. Each cell gives the probability that a method (columns) attains the lowest weighted score on an image of that dataset (rows), estimated from 1000 Dirichlet-drawn weight vectors and averaged over the dataset’s images. AVG baseline scores highly and WTH gains across most datasets as the $NMSE_{syn}$ term is down-weighted.

Table A3. Per-dataset breakdown of the aggregate loss over the equator-masked region, resolving the Loss column of Table 3 by dataset. Each cell is the mean $\pm$ standard deviation across the n images of that dataset, using the normalized metrics of Section 2.3.1 and the weight vector $[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] = [0.1, 0.05, 0.05, 0.01, 0.8]$ of Equation (8). The ordering is conditional on the weight vector; see Section 4. Lower values indicate better performance and the best method in each row is highlighted in bold. Method abbreviations are defined in Table 8 and the datasets in Table 9. The summary rows give the unweighted mean and standard deviation over the 11 dataset means, the mean over all 547 images, and the number of datasets in which each method attains the lowest mean loss. The value “# best” shows the number of datasets where this corresponding method was ranked first.

Dataset	n	AVG	B-ILPF	CS	G-ILPF	RCH	RW	WTH
2022_1028	47	0.79 $\pm$ 0.02	0.72 $\pm$ 0.17	0.85 $\pm$ 0.12	0.72 $\pm$ 0.17	1.52 $\pm$ 0.41	0.74 $\pm$ 0.12	0.62 $\pm$ 0.16
2024_0508	50	0.83 $\pm$ 0.32	0.63 $\pm$ 0.10	0.65 $\pm$ 0.09	0.67 $\pm$ 0.14	2.34 $\pm$ 0.02	0.53 $\pm$ 0.05	0.80 $\pm$ 0.22
2024_1213 air	50	0.72 $\pm$ 0.01	0.72 $\pm$ 0.04	0.85 $\pm$ 0.13	0.71 $\pm$ 0.05	1.07 $\pm$ 0.49	0.79 $\pm$ 0.09	0.67 $\pm$ 0.04
2024_1213 sol	50	0.60 $\pm$ 0.03	0.29 $\pm$ 0.02	0.34 $\pm$ 0.04	0.28 $\pm$ 0.02	1.98 $\pm$ 0.35	0.34 $\pm$ 0.04	0.28 $\pm$ 0.02
2025_0227	50	0.71 $\pm$ 0.03	0.57 $\pm$ 0.10	0.66 $\pm$ 0.18	0.55 $\pm$ 0.09	2.06 $\pm$ 0.26	0.54 $\pm$ 0.10	0.44 $\pm$ 0.04
2025_0402	50	0.79 $\pm$ 0.01	0.62 $\pm$ 0.04	0.93 $\pm$ 0.05	0.62 $\pm$ 0.04	0.98 $\pm$ 0.09	0.80 $\pm$ 0.07	0.62 $\pm$ 0.04
2026_0409 air	50	0.68 $\pm$ 0.02	0.45 $\pm$ 0.03	0.67 $\pm$ 0.08	0.46 $\pm$ 0.03	1.72 $\pm$ 0.10	0.50 $\pm$ 0.04	0.44 $\pm$ 0.04
2026_0409 sol	50	0.68 $\pm$ 0.02	0.37 $\pm$ 0.02	0.44 $\pm$ 0.07	0.36 $\pm$ 0.03	1.64 $\pm$ 0.14	0.35 $\pm$ 0.04	0.38 $\pm$ 0.02
P2_F6_639	50	0.68 $\pm$ 0.00	0.96 $\pm$ 0.07	0.76 $\pm$ 0.03	0.92 $\pm$ 0.07	2.15 $\pm$ 0.10	1.03 $\pm$ 0.06	0.88 $\pm$ 0.08
P3_F4_675	50	0.81 $\pm$ 0.01	1.07 $\pm$ 0.07	0.61 $\pm$ 0.00	1.04 $\pm$ 0.06	2.04 $\pm$ 0.11	1.05 $\pm$ 0.07	0.98 $\pm$ 0.08
MPLL	50	0.70 $\pm$ 0.04	0.39 $\pm$ 0.05	0.59 $\pm$ 0.16	0.39 $\pm$ 0.05	1.66 $\pm$ 0.29	0.52 $\pm$ 0.13	0.42 $\pm$ 0.04
Mean (dataset)	11	0.73	0.62	0.67	0.61	1.74	0.65	0.59
SD (dataset)	–	0.07	0.24	0.18	0.24	0.43	0.25	0.22
Mean (image)	547	0.73	0.62	0.67	0.61	1.74	0.65	0.59
images averaged	–	547	547	547	547	547	547	547
# best	–	1	2	1	1	0	2	4

Table A4. Per-dataset win rate under the fixed weight vector of Equation (8). Each cell is the fraction of the n images of that dataset on which the method attains the lowest aggregate loss of all seven methods, showing how consistently a method leads. The highest win rate in each row is highlighted in bold. Method abbreviations are defined in Table 8 and the datasets in Table 9. The summary rows give the unweighted mean win rate over the 11 datasets and the number of datasets in which each method has the highest win rate. Unlike Figure A3, which averages over 1000 Dirichlet-drawn weight vectors, these rates are conditional on the single weight vector $[0.1, 0.05, 0.05, 0.01, 0.8]$. The value “# best” shows the number of datasets where this corresponding method was ranked first.

Dataset	n	AVG	B-ILPF	CS	G-ILPF	RCH	RW	WTH
2022_1028	47	0.17	0.00	0.00	0.00	0.00	0.00	0.83
2024_0508	50	0.00	0.04	0.00	0.00	0.00	0.94	0.02
2024_1213 air	50	0.08	0.00	0.00	0.08	0.00	0.00	0.84
2024_1213 sol	50	0.00	0.00	0.00	0.64	0.00	0.06	0.30
2025_0227	50	0.00	0.00	0.12	0.00	0.00	0.08	0.80
2025_0402	50	0.00	0.72	0.00	0.10	0.00	0.00	0.18
2026_0409 air	50	0.00	0.30	0.00	0.06	0.00	0.00	0.64
2026_0409 sol	50	0.00	0.02	0.00	0.18	0.00	0.76	0.04
P2_F6_639	50	1.00	0.00	0.00	0.00	0.00	0.00	0.00
P3_F4_675	50	0.00	0.00	1.00	0.00	0.00	0.00	0.00
MPLL	50	0.00	0.48	0.00	0.44	0.00	0.06	0.02
Mean (dataset)	11	0.11	0.14	0.10	0.14	0.00	0.17	0.33
# best	–	1	2	1	1	0	2	4

Table A5. Dataset-stratified paired differences relative to WTH. The unit of analysis is the dataset ($N = 11$), not the image. For each contrast the per-image difference $\delta_i = L_{i,WTH} - L_{i,m}$ is formed on the complete-case intersection of the two methods (n_{pair} images, i.e., images for which both methods competed successfully), and these differences are averaged within each dataset to give Δ_d ; the resulting 11 values are the observations entering the test, so each dataset contributes equally. Columns: $\bar{\Delta}$ and SD are the mean and the between-dataset standard deviation of the Δ_d ; the 95% CI is $\bar{\Delta} \pm t_{0.975,10} \text{SD} / \sqrt{11}$ with $t_{0.975,10} = 2.228$; p comes from a two-sided one-sample t -test of $\bar{\Delta} = 0$ on 10 degrees of freedom; p_{Holm} is the Holm–Bonferroni step-down adjustment across the six contrasts, controlling the family-wise error rate at $\alpha = 0.05$; $d_z = \bar{\Delta} / \text{SD}$ is the paired effect size, i.e., the shift expressed in between-dataset standard deviations; Win% (img) is the fraction of the n_{pair} individual images on which WTH attains the lower loss (ties split evenly); and Datasets counts how many of the 11 datasets favour WTH ($\Delta_d < 0$). Negative $\bar{\Delta}$ favours WTH. WTH wins on the majority of individual images in all six comparisons, indicating a consistent but small advantage rather than the absence of one.

Comparison	n_{pair}	$\bar{\Delta}$	SD	95% CI	p	p_{Holm}	d_z	Win% (img)	Datasets
WTH vs. G-ILPF	547	−0.02	0.07	(−0.06, 0.03)	0.446	0.766	−0.24	56%	6/11
WTH vs. B-ILPF	547	−0.02	0.08	(−0.08, 0.03)	0.383	0.766	−0.28	61%	7/11
WTH vs. RW	547	−0.06	0.12	(−0.14, 0.02)	0.147	0.588	−0.47	78%	9/11
WTH vs. CS	547	−0.08	0.21	(−0.21, 0.06)	0.254	0.762	−0.37	72%	8/11
WTH vs. AVG	547	−0.13	0.18	(−0.26, −0.01)	0.038	0.190	−0.72	76%	9/11
WTH vs. RCH	547	−1.15	0.44	(−1.45, −0.85)	<0.001	<0.001	−2.58	100%	11/11

Table A6. Hierarchical (two-level) bootstrap of the aggregate loss. Each replicate resamples the $N = 11$ datasets with replacement and, within every drawn dataset, resamples its images with replacement. The statistic is recomputed on each resample and the procedure repeated $B = 10,000$ times (seed = 0). Mean is the dataset-weighted mean loss. Naive CI (image) is the interval one obtains by pooling all images and treating them as independent, $\bar{L} \pm 1.96s / \sqrt{n}$; it is shown only for contrast and is centred on the image-weighted mean. Hierarchical CI is the 2.5th–97.5th percentile range of the bootstrap distribution. $P(\text{rank } 1)$ is the fraction of the 10,000 replicates in which the method attains the lowest mean, i.e., the probability that it would rank first were the study repeated on a fresh draw of datasets and images. WTH ranks first in 71% of replicates, but no method is separated from the baselines with confidence.

Method	Mean	Naive CI (Image)	Hierarchical CI	$P(\text{Rank } 1)$
WTH	0.59	(0.57, 0.61)	(0.47, 0.72)	0.71
G-ILPF	0.61	(0.59, 0.63)	(0.48, 0.75)	0.09
B-ILPF	0.62	(0.60, 0.64)	(0.49, 0.76)	0.08
RW	0.65	(0.63, 0.67)	(0.52, 0.80)	0.02
CS	0.67	(0.65, 0.69)	(0.57, 0.77)	0.10
AVG	0.73	(0.72, 0.74)	(0.69, 0.77)	0.00
RCH	1.74	(1.70, 1.78)	(1.50, 1.98)	0.00

Table A7. Per-dataset breakdown of the aggregate loss for the methods that model the background across the whole pattern, resolving the Loss column of Table 4 by dataset. The “Fit” column is the iterative parametric fitting of Section 2.6 alone, and each “Fit+” column combines that fitting with the named framework method applied to the residual background. Each cell is the mean $\pm$ standard deviation across the n images of that dataset, using the normalized metrics of Section 2.3.1 and the weight vector $[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] = [0.1, 0.05, 0.05, 0.01, 0.8]$ of Equation (8). The ordering is conditional on the weight vector; see Section 4. Lower values indicate better performance; the best method in each row is highlighted in bold. Method abbreviations are defined in Table 8 and the datasets in Table 9. The loss is evaluated over the equator-masked region, as in Table A3, but these methods are additionally required to represent the equatorial streak, so the two tables should not be read as a like-for-like ranking. RCH is the radial convex hull baseline (Table 8). The summary rows give the unweighted mean and standard deviation over the 11 dataset means, the mean over the successfully processed images (544 of 547 patterns for the Fit-based columns), and the number of datasets in which each method attains the lowest mean loss.

Dataset	n	Fit	Fit+B-ILPF	Fit+G-ILPF	Fit+RW	Fit+WTH	RCH
2022_1028	47	1.92 $\pm$ 0.19	0.87 $\pm$ 0.18	0.87 $\pm$ 0.18	0.87 $\pm$ 0.15	0.83 $\pm$ 0.17	1.52 $\pm$ 0.41

Table A7. Cont.

Dataset	<i>n</i>	Fit	Fit+B-ILPF	Fit+G-ILPF	Fit+RW	Fit+WTH	RCH
2024_0508	50	1.07 ± 0.16	1.04 ± 0.16	1.03 ± 0.15	0.88 ± 0.07	1.04 ± 0.13	2.34 ± 0.02
2024_1213 air	50	1.15 ± 0.15	0.95 ± 0.10	0.92 ± 0.10	1.05 ± 0.11	0.93 ± 0.10	1.07 ± 0.49
2024_1213 sol	50	0.28 ± 0.05	0.34 ± 0.03	0.32 ± 0.03	0.60 ± 0.09	0.33 ± 0.03	1.98 ± 0.35
2025_0227	47	0.78 ± 0.41	0.49 ± 0.11	0.47 ± 0.11	0.68 ± 0.14	0.48 ± 0.12	2.06 ± 0.26
2025_0402	50	2.29 ± 0.10	1.35 ± 0.12	1.36 ± 0.12	1.33 ± 0.07	1.37 ± 0.14	0.98 ± 0.09
2026_0409 air	50	0.78 ± 0.26	0.58 ± 0.06	0.56 ± 0.05	0.82 ± 0.07	0.55 ± 0.06	1.72 ± 0.10
2026_0409 sol	50	0.45 ± 0.09	0.41 ± 0.05	0.40 ± 0.04	0.64 ± 0.06	0.38 ± 0.04	1.64 ± 0.14
P2_F6_639	50	2.37 ± 0.07	1.22 ± 0.08	1.19 ± 0.08	1.09 ± 0.06	1.17 ± 0.09	2.15 ± 0.10
P3_F4_675	50	2.36 ± 0.12	1.30 ± 0.10	1.29 ± 0.10	1.17 ± 0.09	1.28 ± 0.11	2.04 ± 0.11
MPLL	50	0.80 ± 0.45	0.47 ± 0.10	0.46 ± 0.09	0.68 ± 0.14	0.46 ± 0.09	1.66 ± 0.29
Mean (dataset)	11	1.29	0.82	0.81	0.89	0.80	1.74
SD (dataset)	–	0.79	0.38	0.38	0.24	0.38	0.43
Mean (image)	544	1.29	0.82	0.81	0.89	0.80	1.74
images averaged	–	544	544	544	544	544	547
# best	–	1	0	2	4	4	0

Table A8. Per-dataset breakdown of the aggregate loss over the equatorial region, resolving the Loss column of Table 5 by dataset. Unlike Tables A3 and A7, the loss here is computed with a different weight vector: the smoothness term S_{BG} is inapplicable in the equatorial region and is, therefore, excluded, its weight being reallocated to the oversubtraction terms, giving $[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] = [0, 0.1, 0.1, 0.01, 0.8]$ in Equation (8). Fit is the iterative parametric fitting of Section 2.6 without further refinement, the “Fit+” rows combine it with each framework method applied to the residual background, and RCH is the radial convex hull baseline. Each cell is the mean ± standard deviation across the *n* images of that dataset, using the normalized metrics of Section 2.3.1. Lower values indicate better performance; the best method in each row is highlighted in bold. Method abbreviations are defined in Table 8 and the datasets in Table 9. The summary rows give the unweighted mean and standard deviation over the 11 dataset means, the mean over the successfully processed images, and the number of datasets in which each method attains the lowest mean loss. The value “# best” shows the number of datasets where this corresponding method was ranked first.

Dataset	<i>n</i>	Fit	Fit+B-ILPF	Fit+G-ILPF	Fit+RW	Fit+WTH	RCH
2022_1028	47	1.81 ± 0.25	1.40 ± 0.33	1.39 ± 0.30	1.38 ± 0.32	1.41 ± 0.31	2.03 ± 0.14
2024_0508	50	1.20 ± 0.16	1.23 ± 0.15	1.22 ± 0.15	1.30 ± 0.17	1.20 ± 0.15	1.53 ± 0.05
2024_1213 air	50	0.86 ± 0.26	0.47 ± 0.10	0.44 ± 0.06	0.58 ± 0.10	0.45 ± 0.03	1.91 ± 1.80
2024_1213 sol	50	0.54 ± 0.19	0.37 ± 0.06	0.35 ± 0.07	0.57 ± 0.14	0.36 ± 0.09	1.02 ± 0.07
2025_0227	47	1.10 ± 0.30	0.81 ± 0.18	0.73 ± 0.21	1.06 ± 0.21	0.90 ± 0.17	1.17 ± 0.06
2025_0402	50	1.56 ± 0.17	1.28 ± 0.10	1.29 ± 0.13	1.30 ± 0.07	1.26 ± 0.12	1.61 ± 0.14
2026_0409 air	50	1.14 ± 0.22	0.95 ± 0.12	0.89 ± 0.10	0.99 ± 0.13	0.87 ± 0.11	1.97 ± 0.09
2026_0409 sol	50	0.88 ± 0.17	0.64 ± 0.15	0.62 ± 0.13	0.83 ± 0.17	0.60 ± 0.12	1.83 ± 0.08
P2_F6_639	50	2.20 ± 0.04	1.61 ± 0.07	1.54 ± 0.07	1.59 ± 0.09	1.51 ± 0.06	2.12 ± 0.07
P3_F4_675	50	2.16 ± 0.08	1.71 ± 0.09	1.67 ± 0.07	1.64 ± 0.07	1.65 ± 0.07	1.39 ± 0.08
MPLL	50	1.24 ± 0.35	0.79 ± 0.22	0.76 ± 0.23	0.94 ± 0.22	0.85 ± 0.17	6.25 ± 15.44
Mean (dataset)	11	1.34	1.02	0.99	1.11	1.00	2.07
SD (dataset)	–	0.54	0.45	0.45	0.37	0.43	1.43
Mean (image)	–	1.33	1.02	0.99	1.11	1.00	2.07
images averaged	–	544	544	544	544	544	547
# best	–	1	0	4	1	4	1

Table A9. Mean loss by background-removal method, downsampling factor (ds) and smoothing (smooth). A subset of ten images from each dataset is used. The two bottom rows give the mean over all methods for each column and the mean over all methods and downsampling factors for each smoothing setting. The smallest (best) value in each row is shown in bold.

Method	Smooth = 0		Smooth = 1	
	ds = 1	ds = 2	ds = 1	ds = 2
AVG	0.7670	0.7670	0.7670	0.7670
RCH	1.8476	2.1552	1.7187	2.0092
CS	0.8464	0.7067	0.7911	0.6762
RW	0.5761	0.5767	0.6014	0.5756
B-ILPF	0.7361	0.6413	0.6490	0.5793
G-ILPF	0.7558	0.6417	0.5984	0.5743
WTH	0.7465	0.7209	0.5595	0.5758
Mean (all ds)	0.8918		0.8173	

Table A10. Per-dataset change in loss after background subtraction parameter optimization ($\Delta = \text{loss}_{\text{optimized}} - \text{loss}_{\text{defaults}}$). Negative values indicate improvement. The default values used in the evaluation are listed in Table A2; available in the releases following MuscleX version 2.1.0. Since the optimized parameters are selected by minimizing the same loss that is used for the comparison, a reduction primarily demonstrates successful objective-function minimization. It should not be interpreted as evidence that the optimized result is physically more accurate or superior to expert-selected parameters.

Dataset	CS	RW	SB	SG	WTH
2025_0227	−0.0537	−0.0881	−0.5494	−0.2804	−0.0285
P2_F6_639	−0.5295	−0.0759	−0.0049	−0.1513	−0.0005
P3_F4_675	−0.7637	−0.1592	−0.0021	−0.1104	−0.0010
2024_0508	−0.0590	−0.1841	−0.0848	−0.0839	−0.0121
2024_1213 sol	−0.0174	−0.0859	−0.0570	−0.0096	−0.0013
2024_1213 air	−0.2961	−0.3296	−0.0094	−0.0291	−0.0166
2025_0402	−0.0760	−0.1559	−0.0070	−0.1201	−0.0064
2026_0409 air	−0.0386	−0.1670	−0.0092	−0.0730	−0.0333
2026_0409 sol	−0.0222	−0.1508	−0.0048	−0.0165	−0.0029
2022_1028	−0.3632	−0.3385	−0.0143	−0.0403	−0.0022
MPLL	−0.1076	−0.0083	−0.0022	−0.0159	−0.0041
Mean Δ	−0.2115	−0.1585	−0.0677	−0.0846	−0.0099
Change (%)	−24.44	−19.65	−9.98	−12.27	−1.66
Datasets improved	11/11	11/11	11/11	11/11	11/11

Table A11. Sensitivity of the background subtraction quality loss to the optimizer’s starting point. Every method was rerun from $N = 5$ random initial parameter vectors on the same 10 images per dataset for three of the datasets; the spread is taken per image across starts, averaged over images within a dataset. CV is the per-image standard deviation divided by the per-image mean loss, so it is comparable across methods of different loss scale. Because Mean loss, SD, and CV are each averaged separately over images, CV is the average of the per-image ratios rather than the ratio of the Mean loss and SD columns shown.

Method	Starts	Mean Loss	SD	CV
CS	5	0.812	0.069	0.090
RW	5	0.616	0.107	0.160
B-ILPF	5	1.378	0.591	0.360
G-ILPF	5	0.602	0.021	0.029
WTH	5	0.580	0.006	0.010

Table A12. Equator fit of one of the datasets under three background-removal configurations: Fit (iterative 2D background fit), RCH (MuscleX 2D convex hull) and 1d-CH (no 2D removal; 1D convex hull inside the equator fit, the baseline). Values are the mean $\pm$ standard deviation over the compared images. A fit is excluded for one of three reasons: a MuscleX fitting error above 0.2 (err.), a collapsed peak whose area is zero or missing (coll.), or a peak area or intensity ratio that differs from the baseline by more than a factor of 10 (base). The last reason marks a wrong answer; the image is removed from every configuration. The configurations are, therefore, compared on the identical set of N images that all three fitted successfully. Positions are given for the left side of the pattern; areas are averaged over the two sides.

Dataset	Config	Excluded (%)			N	Position (px)		$I_{1,0}$	Area (10^3)	
		Err.	Coll.	Base		$S_{1,0}$	$S_{1,1}$		$I_{1,1}$	$I_{1,1}/I_{1,0}$
P3_F4_675	Fit	12.0	0.0	2.0	40	115.03 ± 2.12	199.23 ± 3.68	65.2 ± 14.4	28.0 ± 6.0	0.437 ± 0.086
	RCH	4.0	0.0	0.0		114.54 ± 1.02	198.40 ± 1.77	53.2 ± 3.8	18.4 ± 2.4	0.350 ± 0.064
	1d-CH	2.0	0.0	0.0		114.52 ± 1.01	198.35 ± 1.76	52.6 ± 4.3	17.5 ± 2.4	0.338 ± 0.069

Table A13. Susceptibility of the background-quality loss to cosmetic post-processing of an already-computed background subtraction. For each of $n = 550$ images (10 images of each dataset–method pair, across the 11 datasets and the evaluated background-removal methods—WTH, G-ILPF, B-ILPF, RW, AVG) we apply one transformation. Offset, clip and rescale act on the residual ($R' = R + \epsilon$, $R' = \max(R, 0)$, $R' = cR$) and the background is then redefined as $B' = I - R'$; blur acts on the background ($B' = G_\sigma * B$) and the residual is redefined as $R' = I - B'$. In every case $R' + B' = I$, i.e., the original image is never altered. The transformed pair is then scored with the loss from Equation (8). The offset ϵ is expressed in units of the per-image baseline τ ($\tau = 0.2 \times$ the mean intensity of I in a 30-px annulus inside $r_{\max}$, floored at 0.05), which is the threshold used by the $1 - f_B$ term. Columns: n , number of images; $\bar{L}_{\text{base}}$, mean loss of the untransformed pair; $\bar{L}_{\text{new}}$, mean loss after the transformation; $\overline{\Delta L} = \bar{L}_{\text{new}} - \bar{L}_{\text{base}}$, mean absolute change; $\overline{\Delta L} [\%]$, mean of the per-image relative change $100 (L_{\text{new}} - L_{\text{base}}) / L_{\text{base}}$; Improved, the fraction of images for which the transformation lowered the loss ($L_{\text{new}} < L_{\text{base}}$). Negative $\overline{\Delta L}$ means the loss was reduced without any genuine improvement of the background estimate. The constant offset and background blurring lower the loss.

Transform	Parameter	n	$\bar{L}_{\text{base}}$	$\bar{L}_{\text{new}}$	$\overline{\Delta L}$	$\overline{\Delta L} [\%]$	Improved
Offset	$\epsilon = 0.05 \tau$	550	0.675	0.666	−0.009	−3.13	48%
	$\epsilon = 0.10 \tau$	550	0.675	0.667	−0.008	−3.80	47%
	$\epsilon = 0.25 \tau$	550	0.675	0.702	+0.027	+1.67	38%
Clip	—	550	0.675	1.323	+0.648	+74.0	10%
Rescale	$c = 0.9$	550	0.675	0.829	+0.154	+20.2	2%
	$c = 0.7$	550	0.675	1.237	+0.562	+76.8	0%
	$c = 0.5$	550	0.675	1.662	+0.987	+136.7	0%
Blur (background)	$\sigma = 1 \text{ px}$	550	0.675	0.671	−0.004	−0.83	84%
	$\sigma = 2 \text{ px}$	550	0.675	0.671	−0.004	−0.85	81%
	$\sigma = 4 \text{ px}$	550	0.675	0.672	−0.003	−0.63	77%

Table A14. Sensitivity of the framework to pattern quality. Twenty frames drawn at random from those of dataset 2025_0227 that were rejected by the visual inclusion criterion of Section 2.5 but in which equatorial reflections remained faintly discernible were processed with the same pipeline as the main comparison. Loss (incl.) reproduces the 2025_0227 row of Table A3, that is the $n = 50$ subselected frames that passed the criterion; Loss (rej.) is the mean $\pm$ st.dev. over the 20 rejected frames; Δ is the difference, positive values indicating worse performance on the low-quality set. Rank gives the position of each method by mean loss on the included set followed by its position on the low-quality set. The five metric columns are the normalized terms of Section 2.3.1 evaluated on the low-quality set and are directly comparable with Table 3; best values are in bold and second-best underlined. The Spearman rank correlation between the two orderings is $\rho = 0.96$; only B-ILPF and CS exchange places. The aggregate loss is not higher on the low-quality set, but this reflects the weighting rather than better subtraction: the residual-baseline term $1 - f_B$, which carries most of the weight, is easier to satisfy when there is less diffuse scattering, whereas the synthetic-feature terms NMSE_{syn} and $f_{\text{neg,syn}}$ are worse than in Table 3.

Method	Loss (Incl.) ($n = 50$)	Loss (Rej.) ($n = 20$)	Δ	Rank (i $\rightarrow$ r)	S_{BG}	$f_{\text{neg,con}}$	$f_{\text{neg,syn}}$ (Low-Quality Set)	NMSE_{syn}	$1 - f_B$
AVG	0.71 $\pm$ 0.03	0.65 $\pm$ 0.05	−0.06	6 $\rightarrow$ 6	0.00 $\pm$ 0.0	4.23 $\pm$ 0.1	1.35 $\pm$ 0.0	0.19 $\pm$ 0.0	0.47 $\pm$ 0.1
B-ILPF	0.57 $\pm$ 0.10	0.62 $\pm$ 0.15	+0.05	4 $\rightarrow$ 5	3.55 $\pm$ 1.1	1.20 $\pm$ 0.3	1.49 $\pm$ 0.1	10.32 $\pm$ 5.6	<u>0.04 $\pm$ 0.0</u>
CS	0.66 $\pm$ 0.18	0.61 $\pm$ 0.15	−0.05	5 $\rightarrow$ 4	2.65 $\pm$ 1.0	1.29 $\pm$ 0.3	0.92 $\pm$ 0.1	4.94 $\pm$ 3.8	0.24 $\pm$ 0.1
G-ILPF	0.55 $\pm$ 0.09	0.59 $\pm$ 0.14	+0.04	3 $\rightarrow$ 3	3.49 $\pm$ 1.1	0.96 $\pm$ 0.2	1.51 $\pm$ 0.0	9.46 $\pm$ 2.8	0.03 $\pm$ 0.0
RCH	2.06 $\pm$ 0.26	1.99 $\pm$ 0.16	−0.07	7 $\rightarrow$ 7	3.79 $\pm$ 1.0	0.06 $\pm$ 0.0	<u>0.80 $\pm$ 0.1</u>	20.36 $\pm$ 10.6	1.70 $\pm$ 0.1
RW	<u>0.54 $\pm$ 0.10</u>	<u>0.54 $\pm$ 0.13</u>	0.00	2 $\rightarrow$ 2	3.11 $\pm$ 1.0	1.70 $\pm$ 0.2	0.84 $\pm$ 0.2	<u>2.58 $\pm$ 2.0</u>	0.10 $\pm$ 0.0
WTH	0.44 $\pm$ 0.04	0.37 $\pm$ 0.05	−0.07	1 $\rightarrow$ 1	<u>1.30 $\pm$ 0.3</u>	<u>0.81 $\pm$ 0.1</u>	0.80 $\pm$ 0.1	3.69 $\pm$ 2.2	0.15 $\pm$ 0.1

References

- Ma, W.; Irving, T.C. Small Angle X-ray Diffraction as a Tool for Structural Characterization of Muscle Disease. *Int. J. Mol. Sci.* **2022**, *23*, 3052. [CrossRef]
- Squire, J.; Al-Khayat, H.; Arnott, S.; Crawshaw, J.; Denny, R.; Diakun, G.; Dover, D.; Forsyth, T.; He, A.; Knupp, C.; et al. New CCP13 Software and the Strategy Behind Further Developments: Stripping and Modelling of Fibre Diffraction Data. *Fibre Diffraction Rev.* **2003**, *11*, 7–19.
- He, B.B. *Two-Dimensional X-Ray Diffraction*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2018.
- Jarvis, R.A. On the Identification of the Convex Hull of a Finite Set of Points in the Plane. *Inf. Process. Lett.* **1973**, *2*, 18–21. [CrossRef]
- Lee, X.; Jiratrakanvong, J.; Nikseresht, G.; Irving, T.; Agam, G. *MuscleX*, Version 1.14.12; Zenodo: Geneva, Switzerland, 2019. Available online: <https://zenodo.org/records/3360909> (accessed on 3 July 2026). [CrossRef]
- Rajkumar, G.; Al-Khayat, H.A.; Eakins, F.; Knupp, C.; Squire, J.M. The CCP13 FibreFix Program Suite: Semi-Automated Analysis of Diffraction Patterns from Non-Crystalline Materials. *J. Appl. Cryst.* **2007**, *40*, 178–184.
- Ivanova, M.I.; Makowski, L. Iterative Low-Pass Filtering for Estimation of the Background in Fiber Diffraction Patterns. *Acta Crystallogr. A* **1998**, *54*, 626–631. [CrossRef]
- Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 4th ed.; Pearson: New York, NY, USA, 2018.
- Hüpfel, M.; Kobitski, A.Y.; Zhang, W.; Nienhaus, G.U. Wavelet-Based Background and Noise Subtraction for Fluorescence Microscopy Images. *Biomed. Opt. Express* **2021**, *12*, 969–980. [CrossRef]
- René de Cotret, L.P.; Siwick, B.J. A General Method for Baseline-Removal in Ultrafast Electron Powder Diffraction Data Using the Dual-Tree Complex Wavelet Transform. *Struct. Dyn.* **2017**, *4*, 044004. [CrossRef]
- Guo, P.; Zheng, X.; Jiang, J.; Yin, S.; Zhang, C.; Yang, B.; Cui, Y.; Yang, T.; Gu, Y.; Li, X.; et al. Polar Coordinate-Based Background Removal Algorithm for 2D X-ray Scattering Data. *Rev. Sci. Instrum.* **2024**, *95*, 125109. [CrossRef]
- Sternberg, S.R. Biomedical Image Processing. *Computer* **1983**, *16*, 22–34. [CrossRef]
- Aranguren Carmona, R.; Agam, G.; Irving, T. Background Subtraction in Diffraction X-ray Images Using Deep CNN. *Electron. Imaging* **2020**, *32*, 344-1–344-6. [CrossRef]
- Jiratrakanvong, J.; Shao, J.; Li, J.; Menendez Alvarez, M.; Li, X.; Das, P.; Nikseresht, G.; Miskin, N.; Huo, R.; Nabon, J.; et al. MuscleX: Data Analysis Software for Fiber Diffraction Patterns from Muscle. *J. Synchrotron Radiat.* **2024**, *31*, 1401–1408. [CrossRef]
- Chandrasekaran, R.; Stubbs, G. Fibre Diffraction. In *International Tables for Crystallography, Volume F: Crystallography of Biological Macromolecules*; Rossmann, M.G., Arnold, E., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; Chapter 19.5, pp. 444–450.

16. Murthy, N.S.; Zero, K.; Grubb, D.T. Full-Pattern Analysis of Two-Dimensional Small-Angle Scattering Data from Oriented Polymers Using Elliptical Coordinates. *Polymer* **1997**, *38*, 1021–1028. [CrossRef]
17. Lan, X.; Nabon, J.; Miskin, N.; Leduc, T.; Huo, R.; Lee, X.; Zhang, E.; Allison N.; Nikseresht, G.; Jiratrakanvong, J.; et al. *MuscleX*, Version 2.1.0; Zenodo: Geneva, Switzerland, 2026. Available online: <https://zenodo.org/records/21709989> (accessed on 30 August 2026). [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.