Precise implications for real-space pair distribution function modeling of effects intrinsic to modern time-of-flight neutron diffractometers

Daniel Olds,[a]* Claire N. Saunders,[b] Megan Peters,[c] Thomas Proffen,[a] Joerg Neuefeind[a] and Katharine Page[a]*

[a]Oak Ridge National Laboratory, One Bethel Valley Road, PO Box 2008, MS-6454, Oak Ridge, TN 37831-6454, USA, [b]California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA, and [c]University of Kentucky, Lexington, Kentucky 40506, USA. *Correspondence e-mail: oldsdp@ornl.gov, pagekl@ornl.gov

Total scattering and pair distribution function (PDF) methods allow for detailed study of local atomic order and disorder, including materials for which Rietveld refinements are not traditionally possible (amorphous materials, liquids, glasses and nanoparticles). With the advent of modern neutron time-of-flight (TOF) instrumentation, total scattering studies are capable of producing PDFs with ranges upwards of 100–200 Å, covering the correlation length scales of interest for many materials under study. Despite this, the refinement and subsequent analysis of data are often limited by confounding factors that are not rigorously accounted for in conventional analysis programs. While many of these artifacts are known and recognized by experts in the field, their effects and any associated mitigation strategies largely exist as passed-down 'tribal' knowledge in the community, and have not been concisely demonstrated and compared in a unified presentation. This article aims to explicitly demonstrate, through reviews of previous literature, simulated analysis and real-world case studies, the effects of resolution, binning, bounds, peak shape, peak asymmetry, inconsistent conversion of TOF to d spacing and merging of multiple banks in neutron TOF data as they directly relate to real-space PDF analysis. Suggestions for best practice in analysis of data from modern neutron TOF total scattering instruments when using conventional analysis programs are made, as well as recommendations for improved analysis methods and future instrument design.

1. Introduction

Total scattering and pair distribution function (PDF) analysis methods have been evolving rapidly in recent years (Farrow *et al.*, 2007; Billinge & Levin, 2007; Playford *et al.*, 2014; Egami & Billinge, 2012; Keen & Goodwin, 2015; Mancini & Malavasi, 2015) as the advent of faster computers has allowed for increasingly complex modeling programs such as *RMCProfile* (Tucker *et al.*, 2007), *DEBUSSY* (Cervellino *et al.*, 2015) and *FullRMC* (Aoun, 2016), to address longer length-scale correlations. At the same time, the complexity of many modern material studies often involves characterizing the structural correlations and nanoscale ordering at length scales on the order of tens to hundreds of ångströms (Aksel *et al.*, 2013; Coduri *et al.*, 2013; Hill & Allieta, 2013; Usher *et al.*, 2015; Allieta *et al.*, 2015; Checchia *et al.*, 2015; Liu *et al.*, 2016, 2017; Jiang *et al.*, 2017). To take full advantage of the higher-fidelity models, the need to measure reliable, long length-scale spanning, high-resolution PDFs is greater than ever. Neutron total scattering methods in particular present this capability due to a balance between high Q range and resolution, and the absence of atomic form factor damping of the measured
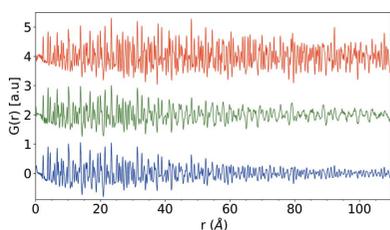
scattering pattern. Furthermore, neutron total scattering is the technique of choice for light atomic species (Li/H/C), particularly in the presence of heavy atoms, and through low absorption in complex sample environments.

Modern instrumentation of many neutron total scattering capable beamlines has been based around the time-of-flight (TOF) architecture of spallation sources, such as the NOMAD (Neuefeind *et al.*, 2012) and POWGEN (Huq *et al.*, 2011) beamlines at the Spallation Neutron Source (SNS) at Oak Ridge National Laboratory (ORNL), the NPDF (Proffen *et al.*, 2002) beamline at Los Alamos National Laboratory (LANL), the GEM (Williams *et al.*, 1997) and POLARIS (Hull *et al.*, 1992) beamlines at the ISIS Neutron Source, and the NOVA (Hattori *et al.*, 2010) beamline at the Japan Proton Accelerator Research Complex (J-PARC). Despite the advantages posed by spallation sources for total scattering techniques, in practice the real-space PDF is rarely used to examine structure beyond the short-range local order (<20 Å).

Previous studies have detailed a wide variety of intrinsic reciprocal-space data artifacts which influence the resultant PDF, including the effects of reciprocal-space constant resolution (Howe *et al.*, 1989; Qiu *et al.*, 2004), finite Q range and Q-dependent resolution (Toby & Egami, 1992; Billinge & Egami, 1993; Chung & Thorpe, 1999; Tucker *et al.*, 2001; Qiu *et al.*, 2004; Olds *et al.*, 2015), Nyquist limits (Farrow *et al.*, 2011), combining spectra (Howe *et al.*, 1989), finite size effects (Jeong *et al.*, 2005; Page *et al.*, 2011; Olds *et al.*, 2015), back-to-back exponential peak shapes (Jeong *et al.*, 2005), and *ad hoc* corrections to optimize a PDF (Peterson *et al.*, 2000, 2003). However, many of these factors can have a similar impact on the PDF, and no previous work has presented a unified review or presentation of the combined repercussions these artifacts have on the resultant PDF.

While methods have been developed to address some of these intrinsic artifacts in the PDF, the details of these mitigation strategies are too often employed solely by experts of the technique, or passed down as communal 'at-the-beamline' knowledge. For example, *MCGRtof* (Tucker *et al.*, 2001) is a Fortran program which calculates the PDF from total scattering data *via* a Monte Carlo approach in order to correct for inherent instrument resolution functions. This program saw little adoption by the greater total scattering community (likely due to relatively long computation times) and, in its current implementation, is unable to address most of the issues presented herein (Tucker, 2017). Given the landscape of newly emerging total scattering beamlines and the complex material challenges they seek to address, it may be time to reassess the reduction and analysis approaches employed by the community, and develop new methods and software to fill the gaps.

This contribution aims to distill and concisely present the cumulative tribal knowledge about artifacts found in neutron TOF-based PDF modeling in a way that is both concise and accessible to novices of the technique. Through simulation, the effects of individual reciprocal-space artifacts are isolated and their effect on the PDF presented. Also discussed are the current methods used in conventional PDF analysis software,

such as *PDFgui* (Farrow *et al.*, 2007), *TOPAS* (Bruker, 2015) and *DISCUS* (Proffen & Neder, 1997) to mitigate or model data aberrations. It is demonstrated that many of these methods are insufficient at high *r* for data from modern TOF-based total scattering instrumentation. Three case studies from neutron TOF instruments are shown (NPDF, NOMAD, POWGEN), which demonstrate real-world examples of the effects of these artifacts on measured data. All simulated and real data presented herein are included as supporting information. Finally, a list of best practices for analyzing neutron PDF data from modern TOF sources is compiled, as well as suggestions for improvements to commonly used analysis programs, and related considerations for future neutron total scattering beamline designs.

2. Simulated study of the effect of data artifacts on the PDF

We begin our presentation with a systematic simulation study of the effects on the PDF due to commonly encountered TOF reciprocal-space artifacts. Although many of these are well known and have been described previously, they are presented here in series to illustrate how similar some of them can be in their alteration of the final PDF and thus emphasize the care in analysis that must be taken when they are present. For these studies, a simulated idealized Si total scattering pattern is utilized. This pattern extends from $Q = 0$ Å^{-1} to $Q = 50$ Å^{-1} with $\Delta Q = 0.01$ Å^{-1}. The form of the PDF employed in this article is defined by Egami & Billinge (2012) as

$$G(r) = 4\pi r[\rho(r) - \rho_0] \qquad (1)$$

where $\rho(r)$ is the atomic pair density and ρ_0 is the average number density. This form of the PDF (often referred to as the reduced pair distribution function) is generated from the measured scattering structure function, $S(Q)$, *via* the form

$$G(r) = \frac{2}{\pi} \int_0^\infty Q[S(Q) - 1] \sin(Qr)\, \mathrm{d}Q. \qquad (2)$$

2.1. The effect of absolute Q resolution and bounds

The PDF is generated, broadly speaking, through the Fourier transform of the measured scattering data (Bragg and diffuse), which has been normalized such that it oscillates about zero at high Q. If the diffuse component is ignored, and the Bragg peaks are treated as delta functions, the PDF would be constructed through the sum of weighted sine functions, with the set of frequencies and amplitudes defined through the positions and amplitudes of the Bragg peaks. The effect of finite broadening of Bragg peaks, therefore, is to introduce a band of frequencies centered about each peak's maxima. The net effect of this is to damp the PDF, which is discussed in detail in §2.2 of this article.

Viewing the measured scattering data through this lens, it becomes evident that the effect of limiting the bounds used in

the generation of the PDF (Q_{min}/Q_{max}) is to reduce the available frequencies employed in the construction of the PDF. Because there are fewer neutron counts at high Q on most instruments, measurement noise tends to increase with Q. Selecting a lower Q_{max} is a commonly employed technique to reduce high-frequency noise in the PDF originating from these components of the data. However, care must be taken in such bound optimization, as limiting the Q_{max} could potentially remove true high-frequency local structure features in the real-space representation of the data. Similarly, this also explains how removing low-Q peaks (*via* the Q_{min}) can result in the absence of slowly oscillating real-space signal (*e.g.* pore structure) from the PDF. Examples of limiting Q_{min}/Q_{max} are shown in Fig. 1.

Beyond the effect of any missing structural signal, abrupt termination of non-zero intensity leads to 'termination ripples'. Most PDF analysis programs offer the ability to model the effects of termination at Q_{max} through a convolution of the calculated PDF with a sinc function. These effects have been demonstrated and explored in detail in many previous studies (Chung & Thorpe, 1999; Farrow & Billinge, 2009; Gagin *et al.*, 2014; Olds *et al.*, 2015). Note that the Q_{max} value present in typical neutron TOF-produced PDFs does not originate from the physical limits of the instrument, but rather the choice of Q_{max} employed in the data reduction process.
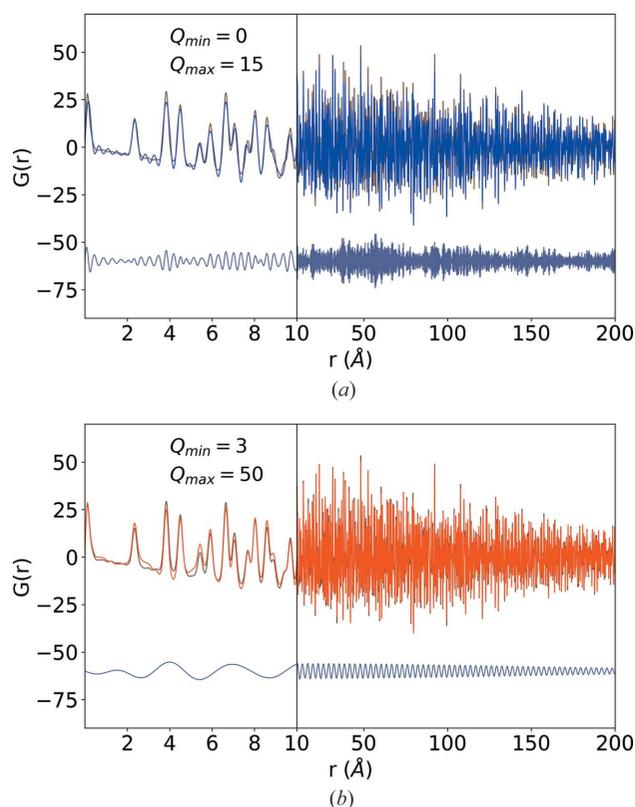
The bandwidth of Q space employed defines the minimum-frequency (Λ_{min}) and maximum-frequency (Λ_{max}) sinusoidal component used in the generation of the PDF, where

$$\Lambda_{max} = \frac{\pi}{Q_{min}} \qquad (3)$$

and

$$\Lambda_{min} = \Delta r = \frac{\pi}{Q_{max}}. \qquad (4)$$

While real-space resolution (Δr) is directly related to Λ_{min} (no resolvable feature can exist finer than Λ_{min}), the effect of the Λ_{max} is more subtle, as it defines a lower limit on the frequency of sinusoidal, single-component contributions (long wavelength) to the PDF. In general, $S(Q)$ can be extrapolated down to $Q = 0$; however, if longer length-scale correlations are present in the material (such as the kind that might be expected in small-angle scattering) caution is required as these simple extrapolations may invalidate fully atomistic models that include long length-scale features (Farrow & Billinge, 2009; Gagin *et al.*, 2014; Olds *et al.*, 2015). A particularly disastrous example of using a Q_{min} value above a Bragg peak is shown in Fig. 1(*b*).

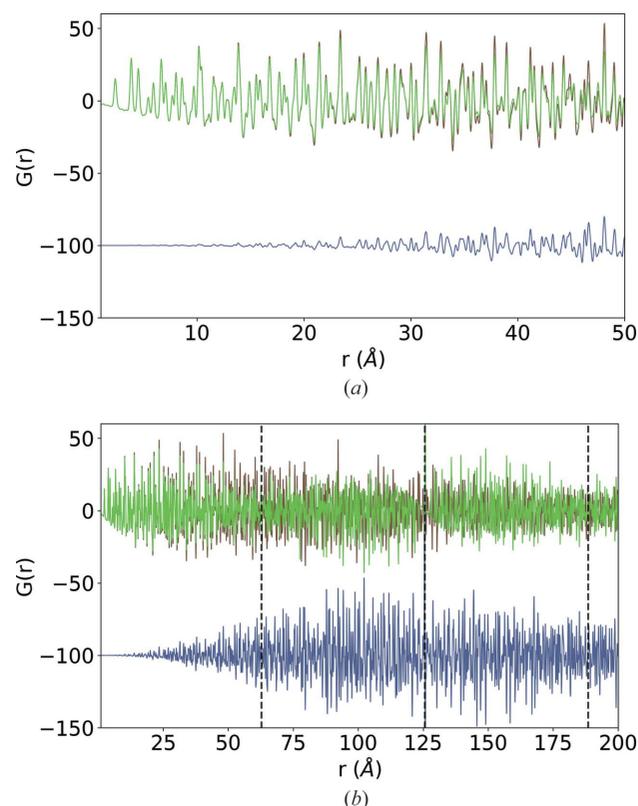During typical TOF neutron total scattering measurements, neutron detection events are measured at a much finer gran-

Figure 1
Demonstration of limiting the employed Q_{min} and Q_{max} on the PDF, when the bin widths are kept constant ($\Delta Q = 0.01$ Å^{-1}). Shown are the ideal $Q_{max} = 50$ Å^{-1} case (brown, all), $Q_{max} = 15$ Å^{-1} (blue) (*a*) and $Q_{min} = 3$ Å^{-1} (orange) (*b*). Differences are offset beneath the PDFs.

Figure 2
Demonstration of the effects of aliasing and resolution loss due to rebinning data coarsely, into widths of $\Delta Q = 0.05$ Å^{-1} (green) compared with $\Delta Q = 0.01$ Å^{-1} (brown), with the difference offset. (*a*) Local changes are subtle, mostly defined by damping. Longer range (*b*), the effects of aliasing are clearly seen, at distances defined by $r_{max} = N \times 62.8$ Å (dashed black lines).

ularity (δQ) than is ultimately employed in the generation of the PDF, on the order of $\delta Q = 10^{-4}$ ms/(0.3 ms < TOF < 16.6 ms) = 3×10^{-4} to 6×10^{-6}. During data reduction, these raw counts are binned linearly in Q space with widths on the order of $\Delta Q = 0.02$ Å^{-1}. Rebinning data into coarser schemes is possible, but the effects of aliasing (Shannon, 1949; Qiu *et al.*, 2004; Farrow *et al.*, 2011) will be present when the PDF is generated to distances larger than

$$L_{\text{max,bin}} = \frac{\pi}{\Delta Q}. \tag{5}$$

Above $L_{\text{max,bin}}$, the calculated PDF will become non-physically periodic, a result of the introduced aliasing (the signal will be reversed and intensity inverted). To demonstrate such an artifact, the data have been rebinned on the scale of $\Delta Q = 0.05$ Å^{-1}, and the PDF calculated out to 200 Å, with the results shown in Fig. 2. In this example, clear aliasing effects are seen at periods of $\pi/0.05$ Å = 62.8 Å, which both confound and repeat (non-structurally) the apparent PDF. Note that, in this example, some damping is also seen due to the subtle loss of resolution, discussed in more detail in §2.2.

Usually, the granularity resolution is much finer than the employed Q binning ($\delta Q \ll \Delta Q$) in order to avoid loss of information. However, since the number of neutrons per bin is proportional to the bin width, smaller ΔQ results in seemingly noisier data in Q space. Although selecting a coarser binning scheme in ΔQ will produce smoother-appearing reciprocal-space data, such rebinning will only result in a PDF with a greater damping rate and more limited length-scale range. Thus, there is no clear benefit for the resultant PDF to rebinning of reciprocal-space data prior to performing the Fourier transformation.

2.2. The effect of symmetric peak shapes

If using a nuclear reactor or X-ray synchrotron source, the characteristic shape of the integrated incident radiation at a reflection point can often be well modeled as a Gaussian or Voigt function (Rietveld, 1969; Finger *et al.*, 1994), with many contributions to the characteristic width given by instrumentation design (guides, slits, collimators, monochromators *etc*). Additional sources of peak broadening, due to factors such as detector precision, beam dispersion and sample thickness, will lead to convolutions of additional Gaussian-like effects on the peak shape. These effects broaden the peak shape, but do not change its inherent characteristics. Some of these effects will have a Q dependence, which is addressed in §2.3. Here, the inherent effects of convolving a constant peak shape with the idealized reciprocal-space data are shown on a resultant real-space PDF.
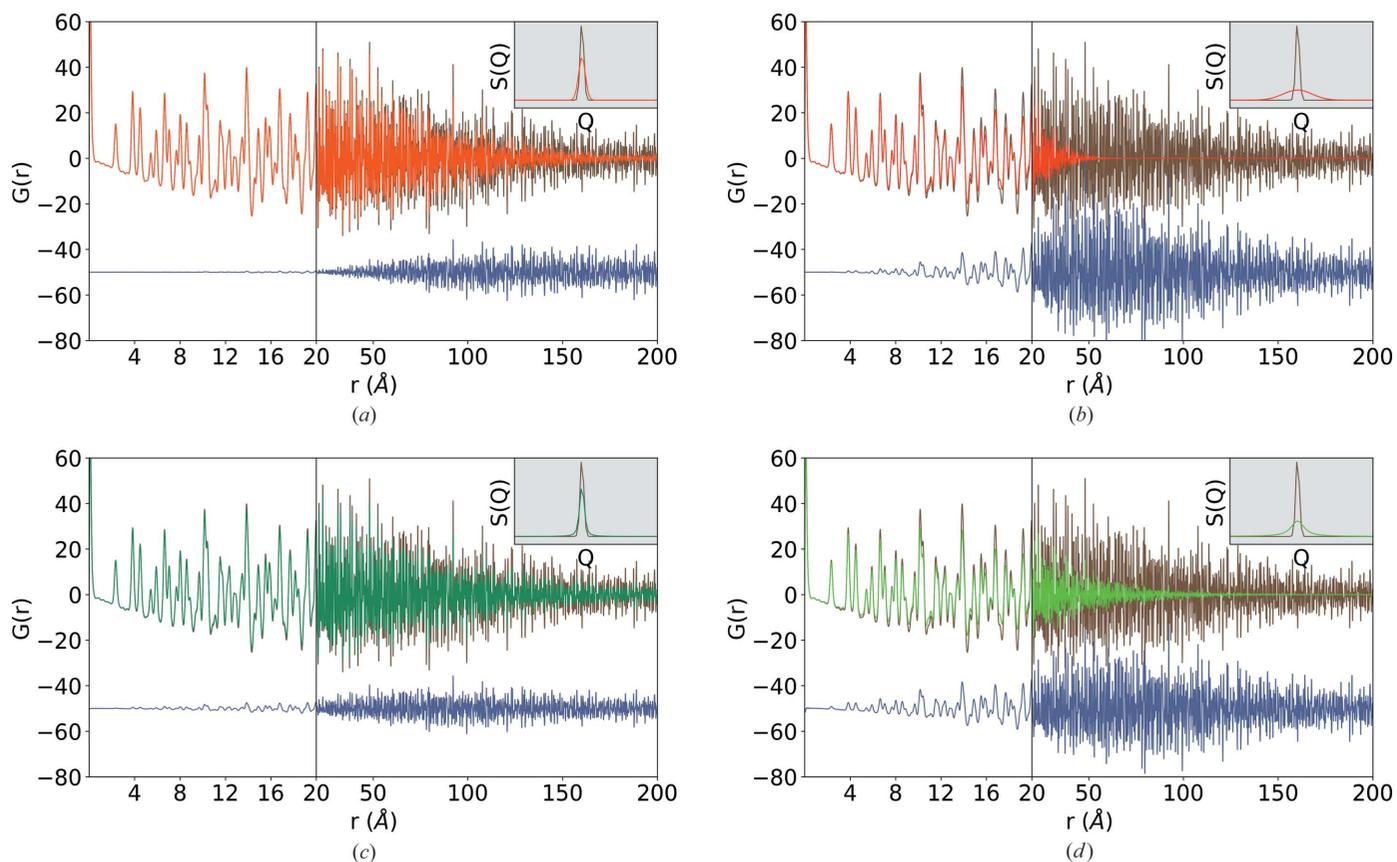

Figure 3
The resultant PDF after convolving different constant peak shapes with the idealized reciprocal-space data, including a Gaussian of width 0.01 Å^{-1} (*a*), a Gaussian of width 0.05 Å^{-1} (*b*), a Lorentzian of width 0.01 Å^{-1} (*c*) and a Lorentzian of width 0.05 Å^{-1} (*d*). For each case, the inset is a sample peak from $Q = 3.8$ Å^{-1}, and data are compared with the unconvolved, idealized case (brown). Differences are offset beneath the PDFs.

Fig. 3 demonstrates the effect of static convolutions for the case of a Gaussian of width 0.01 Å^{-1} [Fig. 3(a), orange], a Gaussian of width 0.05 Å^{-1} [Fig. 3(b), red], a Lorentzian of width 0.01 Å^{-1} [Fig. 3(c), dark green] and a Lorentzian of width 0.05 Å^{-1} [Fig. 3(d), light green]. In each case, the inset figure highlights the effect of the convolution in reciprocal space, compared with the idealized (unconvolved data) at $Q = 3.8$ Å^{-1}. The primary effect is the damping of the high-r real-space PDF as the reciprocal-space peak is broadened. To first order, this can be considered a resolution feature similar to the effect of ΔQ discussed in §2.1. However, as this loss in effective resolution has been framed precisely (as convolutions of known functions), it is possible to explicitly address these effects in real space.

Recall the convolution theorem, which states that when two functions, $j(r)$ and $k(r)$, individually Fourier transform into corresponding functions, $J(Q)$ and $K(Q)$, then the product of these two functions in one space will Fourier transform into the convolution of the corresponding pairs in the other space, as described by

$$j(r) \otimes k(r) = \mathrm{FT}\,[J(Q) \times K(Q)]. \qquad (6)$$

In regards to the examples shown in Figs. 3(a) and 3(b), the reciprocal-space data have been convolved by a Gaussian of width σ_G. The resultant real-space effect will be the original PDF multiplied by the Fourier transform of the convolving Gaussian, resulting in an associated Gaussian scaling factor, G_{damp}, defined as

$$G_{\mathrm{damp}}(r) = \exp^{-(r\sigma_G)^2/2}. \qquad (7)$$

This damping method is demonstrated in Fig. 4(a) and is seen to correctly address this effect.

The Fourier transform of a Lorentzian function is an exponential function, and thus in Figs. 3(c) and 3(d), where the ideal data have been convolved with a Lorentzian function of peak width σ_L, one can see an exponential damping envelope in the PDF, L_{damp}, described as

$$L_{\mathrm{damp}}(r) = \exp^{-(r\sigma_L)/2}. \qquad (8)$$

This convolution is demonstrated in Fig. 4(b) and is seen to correctly address this effect.

Several software packages address the effect of instrument resolution on the PDF by utilizing Gaussian damping effects. For instance in *PDFgui* (Farrow *et al.*, 2007), *DISCUS* (Proffen & Neder, 1997), and *TOPAS* (Bruker, 2015) (encoded intrinsically or *via* defined functions), a so-called Q_{damp} term can be employed when refining to data, where Q_{damp} defines the Gaussian-shaped damping in equation (7). However, the value of this decay constant is typically fit during refinement of a standard data set, and not explicitly based on inherent instrument characteristics. Note that such a Gaussian implementation of Q_{damp} can only precisely account for peak-shape effects that are Gaussian in nature, and employing it to account for alternative peak shapes is inherently an approximation.
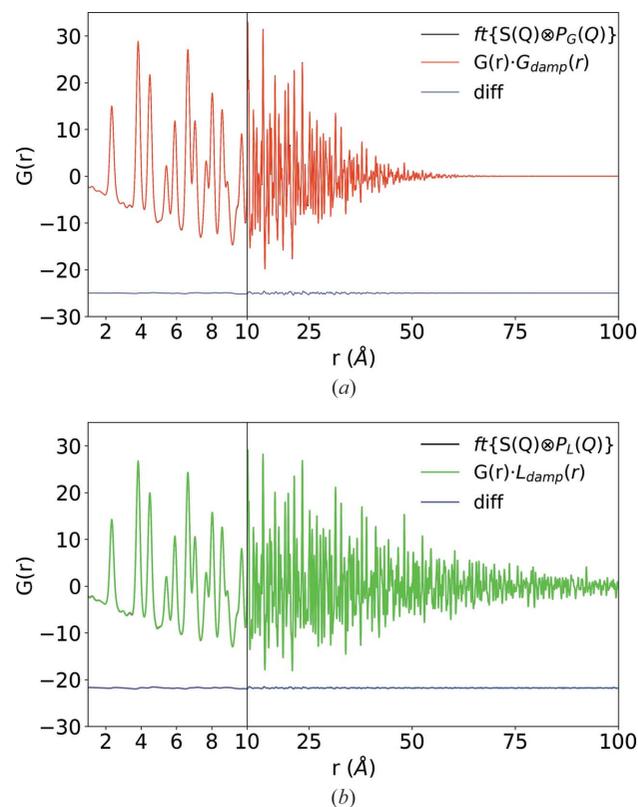

Figure 4
Demonstrating the described method to model the effects of static peak-shape convolutions from reciprocal-space data in real space, *via* equations (7) and (8). The procedure of Gaussian damping is shown in (a), while exponential damping (associated with Lorentzian peak shape) is shown in (b).

2.3. The effect of asymmetric peak shapes

Up until now, the effects presented have been exclusively due to symmetric peak-shape profiles. For TOF neutron sources, the inherent nature of the pulse shape as it is emitted from the moderator will produce asymmetric peak shapes, with a sharp rise at lower TOF (corresponding to the onset of the neutron pulse), followed by a much longer decay tail. To first order, this pulse shape is often treated in reciprocal space as a back-to-back exponential and has been explained in detail previously (Von Dreele *et al.*, 1982; Jeong *et al.*, 2005; Avdeev *et al.*, 2007). Jeong *et al.* have shown that simulated TOF peaks can lead to r-dependent artifacts in PDF data (Jeong *et al.*, 2005). To examine how the degree of asymmetry in a peak profile affects a PDF, it is convenient to define an asymmetric Gaussian function,

$$G_{\alpha,\beta}(x) = \exp[-x^2/(2\alpha^2)]H_c(x) + \exp[-x^2/(2\beta^2)]H_c(-x) \quad (9)$$

$$H_c(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}, \qquad (10)$$

where α and β are different widths and H_c is the Heaviside function, defined as 1 when $x > 0$ and 0 when $x \leq 0$. This peak may be numerically recentered such that the centroid position (at $x = 0$) is defined by half the total integrated area under the curve, $(\pi/2)^{1/2}(\alpha + \beta)$. Although non-standard for fitting TOF

feature articles

reciprocal-space peak shapes (where a back-to-back exponential is typical), by demonstrating the effects of this asymmetric Gaussian peak profile on the PDF, it is possible to isolate those artifacts introduced solely due to characteristic asymmetry.

Here, convolutions of the recentered asymmetric Gaussian function have been applied to the idealized Si pattern for the case of $\alpha = 1, \beta = 5$, and compared against data corresponding to symmetric Gaussian functions with equivalent area, $\alpha = \beta = 3$. The resultant PDFs from these convoluted data sets are shown in Fig. 5(a).

Performing boxcar-style refinements on these PDFs, where the refinement is performed iteratively for a fixed width in r but at increasingly higher $r_{\min}$ and $r_{\max}$ bounds, it is clear that in the case of the asymmetric peak, the as-fit lattice constant [Fig. 5(b), top] displays an r-dependent contraction (consistent with data shifted to higher Q). The overall ability to fit the PDF from the asymmetric case decreases at higher r, as surmised by the increase in the weighted residual, R_{wp} [Fig. 5(b), bottom], defined as

$$R_{\text{wp}} = \frac{\sum w(r)[G_d(r) - G_f(r)]^2}{\sum w(r)G_d(r)} \tag{11}$$

where $G_d(r)$ is the observed PDF and $G_f(r)$ is the calculated model PDF, and $w(r)$ is the associated weighting function (Dinnebier, 2008), taken as unity in these simulations.

Conventional approaches to resolution-dependent damping of the PDF with a Gaussian function are found to poorly fit the asymmetric peak-induced damping behaviour at high r. Despite the overall area under the convolution peaks being equivalent in these two cases, the damping of the PDF in the case of the symmetric peak is occurring faster than in the case of the asymmetric peak. The r-dependent contraction of the PDF becomes significant at higher r values and thus data are not quantitatively reliable above 40 Å in this example. As the artifact introduced to the PDF from asymmetric peak shapes is cumulative in r, low-r data in a measured PDF (below approximately 20 Å for typical measurements) will not be dramatically affected by asymmetric peak shape.

2.4. The effect of Q-dependent resolution

Typically, neutron TOF instrument resolution drops as a function of Q, such that the width of measured diffraction peaks increases as a function of Q. Variation in modern TOF neutron detector bank resolution is often addressed by invoking a $\Delta Q/Q$ or $\Delta d/d$ correction term. Here, the implication of constantly varying reciprocal-space peak widths is
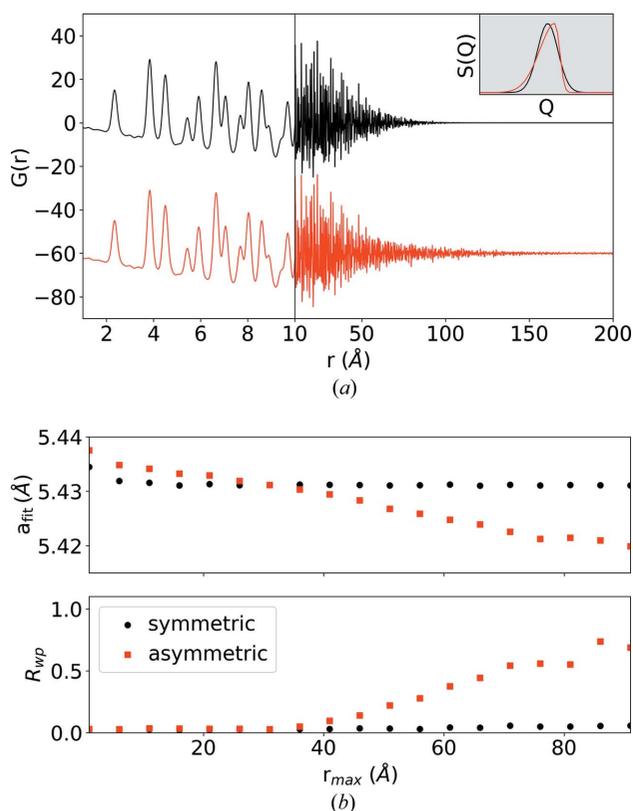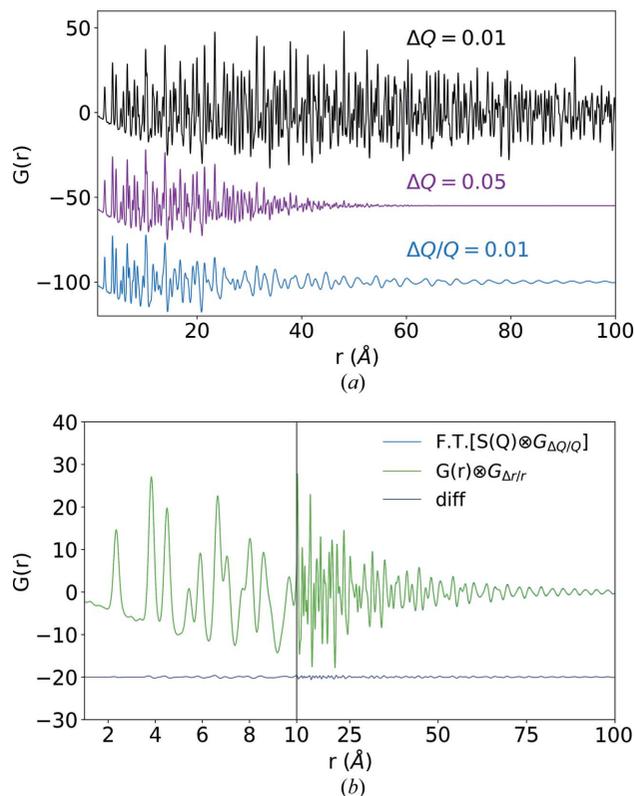

Figure 5
(a) Comparison of PDFs resulting from convolving the ideal pattern with symmetric peak ($\alpha = \beta = 3$, black) and asymmetric peak ($\alpha = 1, \beta = 5$, red) profiles. Also shown (b) are the results of boxcar-style refinements of these two PDFs, with the top panel showing the as-fit lattice constant and the bottom panel showing the resultant R_{wp}.

Figure 6
(a) PDFs from reciprocal-space data which have had Gaussian convolutions applied of constant width ($\Delta Q = 0.01$ Å^{-1}, black, and $\Delta Q = 0.05$ Å^{-1}, purple) and linearly varying width ($\Delta Q/Q = 0.01$, light blue). (b) Demonstration of how the PDF produced from the reciprocal-space data which have been convolved by $\Delta Q/Q = 0.01$ can be correctly modeled through the calculation of the ideal PDF, which is then itself convolved by a Gaussian of corresponding width ($\Delta r/r = 0.01$, green).

demonstrated by convolving the ideal Si data set with a Gaussian of linearly varying width, $\sigma_G = \Delta Q / Q$.

Fig. 6(a) demonstrates the difference in applying constant *versus* Q-dependent resolution functions, as they relate to the PDF. The ideal Si $S(Q)$ pattern was convolved by a Gaussian of either static (black and purple) or linearly changing (blue) width. The resultant PDF from the linearly varying width ($\Delta Q / Q = 0.01$) case is effectively being convolved by a Gaussian with a width varying as $\Delta r / r = 0.01$. This effect is the basis of the so-called 'Q_{broad}' term in PDF refinements, which acts to linearly increase the real-space peak widths as a function of r (Toby & Egami, 1992; Billinge & Egami, 1993; Qiu *et al.*, 2004). Although directly attributed to instrument resolution effects, any linearly dependent peak-broadening effect can be modeled in this way. In Fig. 6(b) the PDF resulting from convolving the reciprocal-space data by $\Delta Q / Q = 0.01$ (blue) is seen to be essentially identical to that resulting from convolving the ideal PDF with $\Delta r / r = 0.01$ (green). Those differences remaining (offset gray) can be attributed to termination errors in the method of the convolution applied to the ideal PDF.

One is cautioned that if Q-dependent broadening displays non-linearity (for instance, if above a threshold the broadening becomes constant), these simple Q_{broad} corrections will no longer address the full effect. This concept is explored in detail in §3.2 of this article.
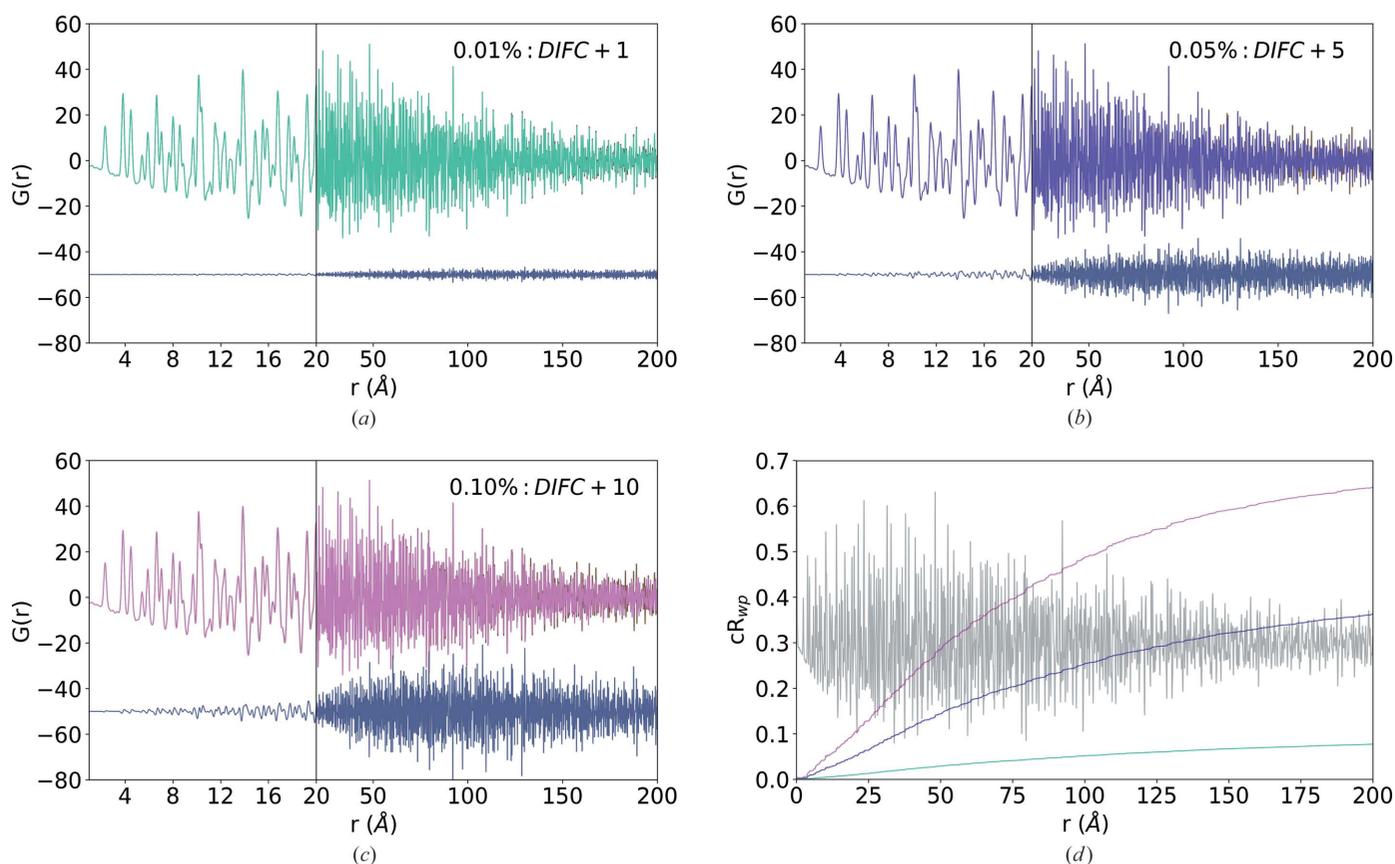
2.5. The effect of inconsistent TOF-to-*d* conversion

The natural units for data from TOF neutron instruments are, explicitly, the measured time of flight (ms). Raw data are 'time-focused' to a single effective energy and then converted from TOF to d (where d is the interplanar distance) most commonly through a quadratic equation, with diffractometer constants (DIFC, DIFA and T0) as defined in the software suite *General Structure Analysis System* (*GSAS*) (Larson & Von Dreele, 1994). Solving the quadratic form of this equation, d values can be expressed in TOF as

$$d = \frac{2\pi}{Q} = \frac{-\text{DIFC} + [\text{DIFC}^2 - 4\,\text{DIFA}(\text{T0} - \text{TOF})]^{1/2}}{2\,\text{DIFA}}.$$

(12)

In the limit where DIFA tends to zero, the relationship simplifies to

$$d = \frac{2\pi}{Q} = \frac{\text{TOF} - \text{T0}}{\text{DIFC}}.$$

(13)

In principle, DIFC is the only constant required to convert TOF data to Q or d spacing. In practice, T0 corrects for any delay in moderator emission time or detector dead time, and DIFA acts as a second-order correction term accounting for shifts in the center of scattering, such as those attributed to sample misalignment and wavelength-dependent absorption.

Figure 7
Resultant PDF from miscalibration in DIFC by 0.01% (*a*), 0.05% (*b*) and 0.10% (*c*), which has been compared with the ideal in each case, with the difference offset beneath the PDFs. Also shown is (*d*) the cumulative weighted residual (cR_{wp}) calculated against the idealized PDF as a function of r for each case.

Daniel Olds *et al.* · The origins of artifacts commonly found in PDF modeling

During Rietveld refinement, it is common to initially refine these terms to a known standard, although there can be cases where these values are refined during the fitting of data. This is because a wide variety of variations between the measured standard and the measured samples can cause subtle changes in the apparent TOF value, such as sample placement, sample thickness relative to calibrant, detector tube drift, changes in detector tube temperature, moderator changes *etc.* The value of DIFC, although physically defined based on the diffractometer geometry, is often numerically fit during calibration procedures on the instrument. If multiple banks of data are being merged, there exists the possibility that inconsistent TOF-to-*d* conversion routines are being used to generate the unified $S(Q)$. During PDF refinements, these diffractometer constants are rarely exposed, forcing the user into a 'what you see is what you get' reality. Here, the effect of slight inaccuracies in the employed DIFC parameter on the resultant PDF is explored. Note that slight variation in DIFA or T0 would have a potentially similar effect.

Fig. 7 demonstrates the effect of overestimating the true DIFC value (which for this idealized data set was approximately 10 000) by values of 1 (Fig. 7*a*), 5 (Fig. 7*b*) and 10 (Fig. 7*c*). Although these are very subtle shifts in the apparent DIFC values (0.01%, 0.05% and 0.1%), one can observe that the net effect on the PDF can be dramatic. The cumulative R_{wp} values are calculated against the ideal PDF, shown in Fig. 7(*d*). The net effect from such a shift in DIFC will be a fixed offset in the as-fit lattice constant. This was verified by performing boxcar-style refinements of the 'miscalibrated' data, in which the as-fit lattice constant was found to be decreased by a scale factor matching the degree of miscalibration (*e.g.* the fit lattice constant is 0.05% smaller than ideal when the DIFC value was 0.05% larger). While such a miscalibration in DIFC would alter the interpretation of the real-space PDF, the effect will be isolated to a fixed apparent offset of the lattice constant, and not introduce any *r*-dependent artifacts in the PDF. Such differences in lattice parameters are often observed in real-*versus* reciprocal-space analyis.

If the degree of miscalibration is not constant in Q, the implication for the PDF is far more complicated and ultimately uncorrectable during analysis. To demonstrate this, the PDF resulting from data which have an incorrect DIFC value of 0.05% above $Q = 5.5\,\text{Å}^{-1}$ (but nominal value below) is calculated. This would be similar to the effect of combining data from two detectors which had inconsistent calibration of DIFC values. The resultant PDFs are shown in Fig. 8(*a*). Fitting this resultant data set in *PDFgui* produced lower-quality fits when compared with the constant offset and ideal cases, as demonstrated by the resultant R_{wp} as a function of r_{max}, Fig. 8(*b*). It is also clear that the effect of this distortion on the apparent lattice parameter is complex and intractable. There are no conventionally available corrections for instrument characteristics in the PDF due to the merging of two inconsistently calibrated 'banks' of data.

Certain types of inconsistent TOF-to-*d* conversion across multiple banks are likely random in nature (*e.g.* tube drift). To simulate such a condition, a modeled effect of random drift in
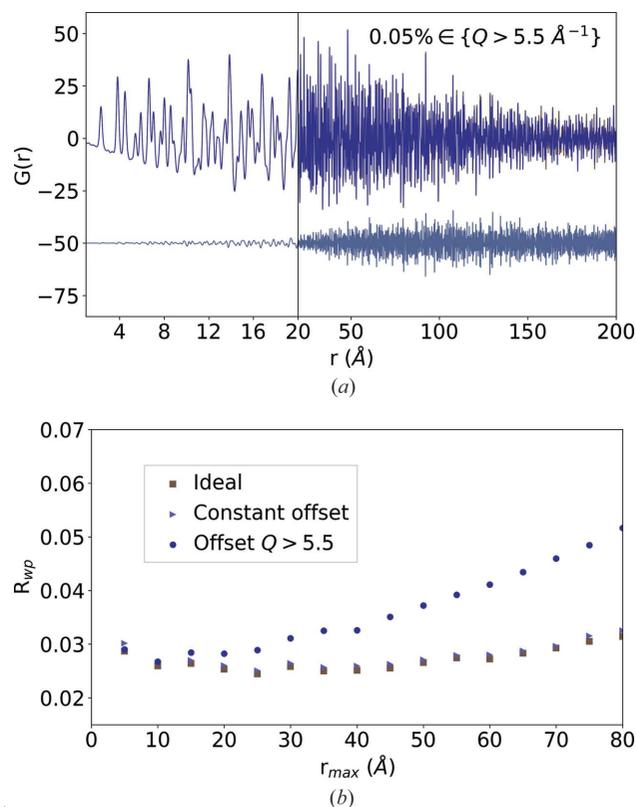

Figure 8
Resultant PDF (*a*) from a 0.05% miscalibration in DIFC [similar to Fig. 7(*b*)] which only occurs above $Q = 5.5\,\text{Å}^{-1}$ compared with the idealized PDF. The difference is plotted as offset. The resultant R_{wp} values from fitting the ideal, constant offset 0.05% (Fig. 7*b*) and this partial offset case are also shown (*b*), where the value is a function of the r_{max} used in the refinement.

collected reciprocal-space data is employed. Beginning with no offset at $Q = 0$, each Q value is assigned a random offset drift which is the sum of the previous random offsets and a uniformly distributed random value between $-\sigma$ and $+\sigma$.

The effect of three such random drift patterns is shown in Fig. 9, which demonstrates the resultant PDF compared with the idealized case. Even though the degree of total offset is much less than the constant offset cases discussed previously, the overall effect is more detrimental to modeling as there is no overarching motif uniformly shifting the data.

Although this degree of random drift in real calibrations would hopefully be avoided through careful instrumentation, calibration and reduction procedures, this demonstrates that even a small deviation, such as the case shown in Fig. 9(*a*), would result in PDFs that cannot be fit well at high r. It is noteworthy that below $r = 20\,\text{Å}$ the PDF is affected very little by this sort of random drift.

3. TOF powder diffraction instrument case studies

In this section, data from a standard reference material [Si-SRM 640 (Mendenhall, 2016)] that have been measured on three modern neutron TOF powder diffractometers, NOMAD (SNS ORNL), POWGEN (SNS ORNL) and NPDF (Lujan Center LANL, no longer openly available) are presented and
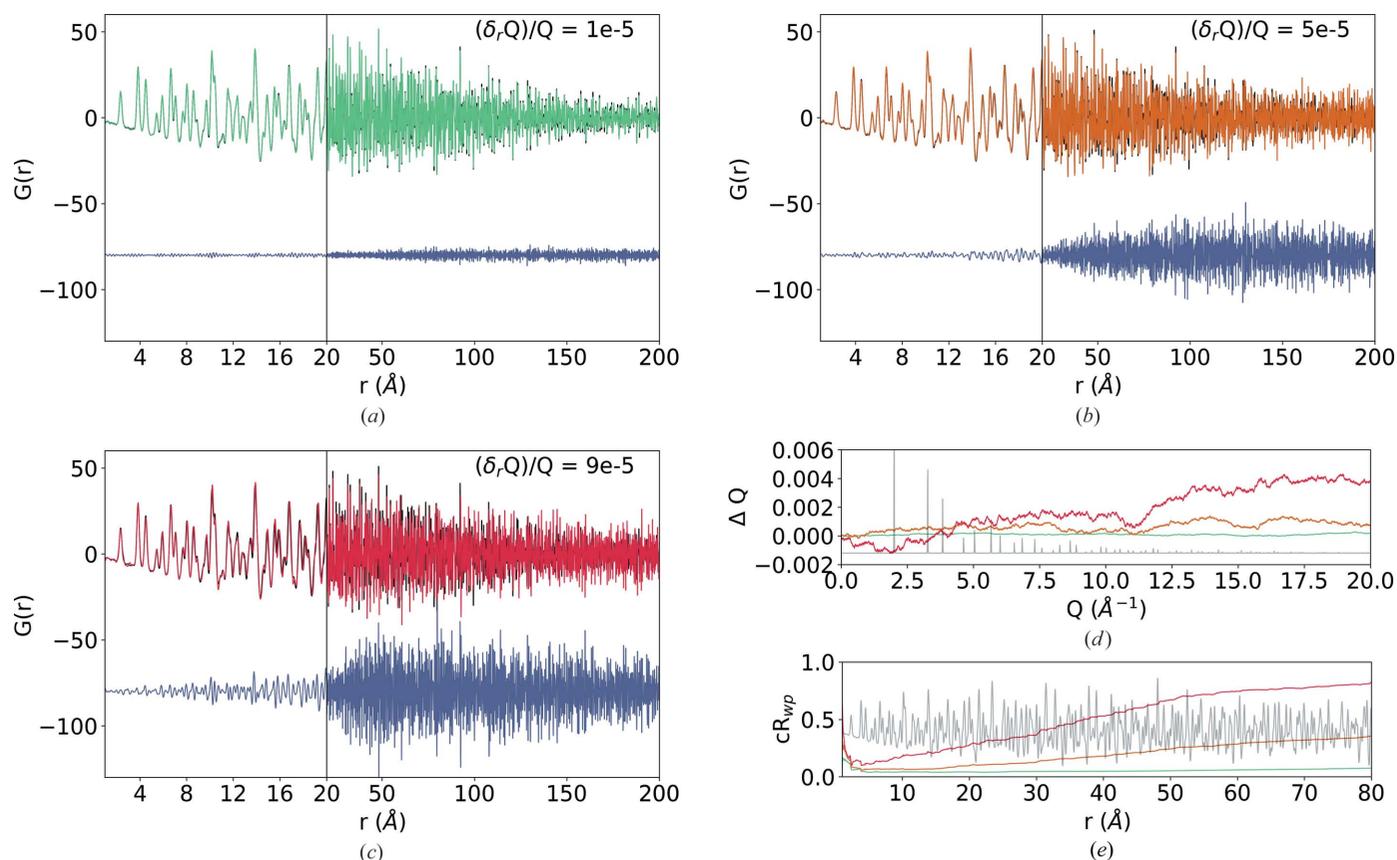
Figure 9
Average PDFs resultant from miscalibrations following a random drift in the Q model, with $\sigma = 0.00001$ (a), 0.00005 (b) and 0.00009 Å^{-1} (c), compared with the ideal PDF in each case, with the difference shown offset below. Also shown are the as-applied random drift in Q values for these three cases (d) and the cumulative R_{wp} from comparing these against the ideal case (e).

explored. The goodness of fit on this standard material using conventional (*PDFgui*) real-space refinement is presented, with a focus on any aberrations found and how they link to the fundamental principles of PDF generation on TOF neutron diffraction instruments.

3.1. NPDF

Although presently not available in an open user program, the NPDF diffractometer at the Lujan Neutron Scattering Center (LANL) (Proffen *et al.*, 2002) was previously a dedicated neutron total scattering beamline for disordered crystalline materials (Qiu *et al.*, 2005; Page *et al.*, 2008; Jeong *et al.*, 2008; Božin *et al.*, 2010; Harrington *et al.*, 2011; Ramezanipour *et al.*, 2012). Four distinct detector banks on NPDF (centered at 45, 90, 119 and 148° 2θ) were typically weighted according to neutron counts and combined to form the merged total scattering structure factor, $S(Q)$.

Data from Si-SRM (640c) were collected on NPDF in a 3/8 inch (9.525 mm)-diameter vanadium sample canister, measured with a beam size of 1×4 cm for 1 h on the NPDF diffractometer. TOF-to-d calibration was performed with *NXproc* (Proffen, 2015), with bank-merging and the generation of the PDF carried out using the *PDFgetN* program (Peterson *et al.*, 2000). The as-fit data and refined lattice

constant are shown in Fig. 10 (details of the refinement can be found in the supporting information).

It is clear that the modeled refinement matches the measured PDF to a high degree; however, there are observable systematic offsets between the as-refined model and the measured PDF. Boxcar-style refinements, performed with a constant width of 5 Å, were completed while refining only the scale, isotropic atomic displacement parameter (U_{iso}) and lattice constant (other parameters were held fixed to values refined from the initial fit). The as-fit lattice constant is shown in Fig. 10(b), which demonstrates an r-dependent contraction of the data. This apparent r-dependent lattice contraction can be fit to the form $a = a_o(1 + r)^p$, for which the values fit were $a_o = 5.43965$ and $p = -0.000263$, shown as the orange line in Fig. 10(b). Although the as-fit lattice constant is gradually approaching the expected average value [shown as a dashed line in Fig. 10(b)], the offset coupled with the r-dependent behavior are evidence of artifacts or miscalibration in the data.

To better understand the origin of this artifact, it is useful to examine the reciprocal-space data from the separate solid-angle detector banks of NPDF, prior to merging the data in *PDFgetN*. The four banks of data from NPDF are shown in Fig. 11 (left), with insets focused on two of the peaks. It is immediately clear that, in this data set, the calibration is more consistent between the four banks at low Q (seen in the peak
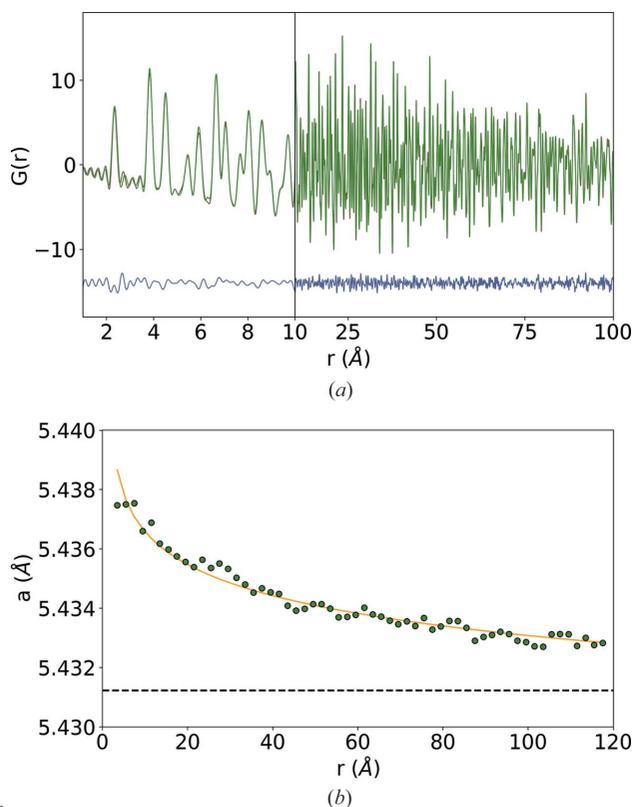
Figure 10
(a) The data and as-fit PDF from Si data gathered on NPDF. (b) Results of boxcar-style refinements with a width of 5 Å, demonstrating an apparent lattice constant contraction as a function of r (dots), fit to $a' = a_0(1 + r)^p$ (orange line), and the as-published value for the lattice constant (dashed).

at 3.37 Å^{-1}) than high Q (seen in the peak at 8.26 Å^{-1}). To further quantify this degree of misalignment, the offset between the measured positions of the 18 largest peaks in the Si data from each bank compared with their known positions is shown in Fig. 11 (right). One can see that the higher-resolution banks (3 and 4) are more consistently calibrated than the lower-resolution banks (1 and 2). When merged

together, not only will this introduce a Q-dependent varying degree of misalignment, it will also introduce a gradually varying peak shape. The lower-Q data, where calibration was more consistent between banks, will merge to more Gaussian-like peak shapes, whereas the higher-Q data will merge to a more asymmetric shape with the peak maxima shifted to the right of the centroid.

As demonstrated through simulation in §2.5, inconsistent TOF-to-d conversion in reciprocal space can be a source of intractable behavior in the real-space PDF. In this case, any number of factors could have led to a miscalibration of the NPDF data, such as incorrect sample positioning, inadequate calibration procedure and short measurement time resulting in low statistics. Because measured data from a set of detectors with different characteristic peak profiles (and apparently, quality of calibration) were merged to generate the PDF shown in Fig. 10(a), the net effect on the PDF is not easy to track down *ex post facto*, and thus it is difficult to derive an analytical correction (in the manner of Q_{damp} or Q_{broad}) which can properly model the observed artifacts in real space. Note that higher-quality data were more typical on NPDF; however, this presented atypical data set (from a short collection time) is included to demonstrate the effects of a miscalibration on the resultant PDF.

3.2. NOMAD

The NOMAD diffractometer at the SNS (ORNL) is a dedicated neutron total scattering instrument (Neuefeind *et al.*, 2012), with unparalleled flux and a wide simultaneously accessible Q range (0.1–50 Å^{-1}). Data from NOMAD's six detector banks (centered at 6, 15, 31, 67, 122 and 154° 2θ) are weighted according to neutron counts and combined to form the merged total scattering structure factor, $S(Q)$. The instrument was developed with a dual purpose: for liquids and amorphous materials (Skinner *et al.*, 2014; Lan *et al.*, 2014; Yang *et al.*, 2016; Wang *et al.*, 2017), and for disordered crystalline materials (Shamblin *et al.*, 2016; Charles *et al.*, 2017;
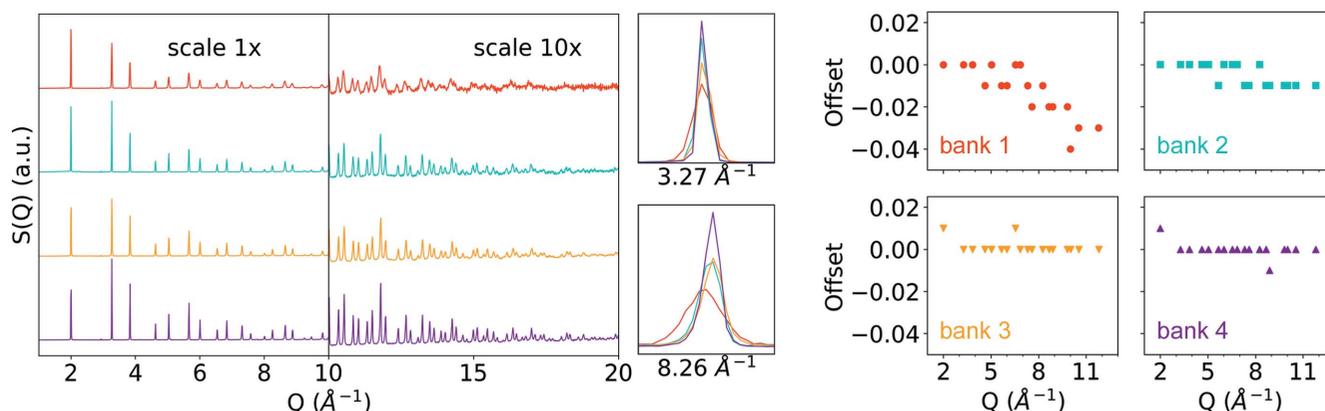

Figure 11
(Left) Reciprocal-space data from NPDF Si standard measurement, separated by bank (bank 1 is red, bank 2 is cyan, bank 3 is orange, bank 4 is purple). In all banks, the binning employed is $\Delta Q = 0.01$ Å^{-1}. (Center) Two example peaks are highlighted at Q = 3.27 and 8.26 Å^{-1}, demonstrating the emergence of a miscalibration effect at higher Q values. (Right) The results of fitting the positions of the 18 largest intensity peaks in the individual banks, relative to expected positions, plotted as a function of their offset in Q.
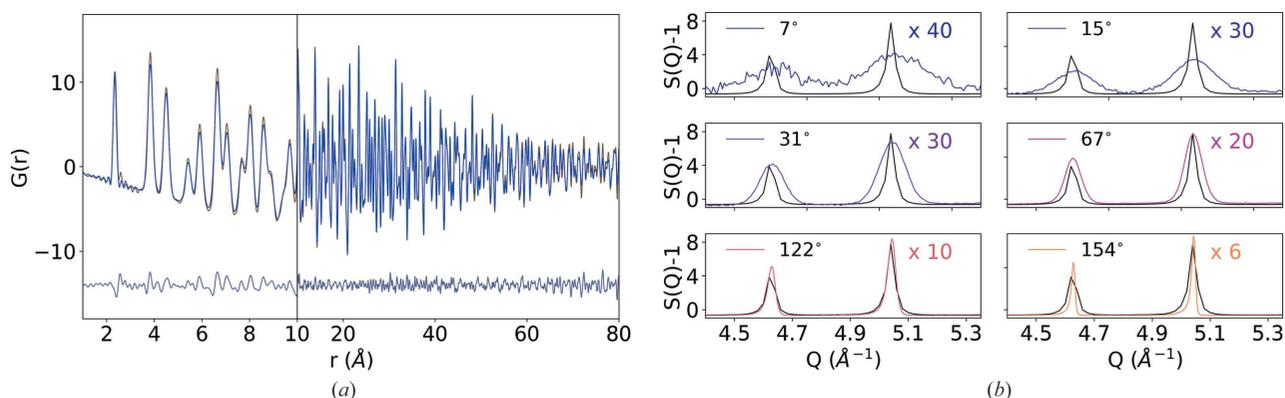
Figure 12
Results of *PDFgui* refinement to Si standard data from NOMAD (*a*) and a sample of the reciprocal-space data from six individual banks (*b*), where the effective combined $S(Q)$ used in the generation of the PDF is plotted in black. Note that, for visual comparison here, the reciprocal-space data from individual banks have been scaled by the value on each subfigure.

Fabini *et al.*, 2017). NOMAD trades some resolution for high flux and a wide Q range, being the lowest-resolution diffractometer of the three case studies here, but is essential for many *in situ* and *in operando* total scattering studies (Wang *et al.*, 2013; Ushakov *et al.*, 2015; Hillis *et al.*, 2016; Olds *et al.*, 2017; Wu *et al.*, 2017).

Data from a Si-SRM (640d) were collected on NOMAD in a 6 mm-diameter vanadium sample canister, with a beam size of diameter 1.5 cm for 6 h. Data were reduced using standard data reduction protocols, which include TOF-to-*d* calibration, merging of banks and generation of the PDF with custom IDL code available at the instrument (Neuefeind *et al.*, 2012). The data, and resultant fit, are shown in Fig. 12(*a*) (detailed refinement parameters can be found in the supporting information).

Compared with NPDF data, the lower average reciprocal-space resolution damps the NOMAD PDF significantly faster, and the overall fit from 0 to 100 Å is markedly worse. However, examining the difference curve between PDFs in Fig. 12(*a*) shows that the majority of the weighted residual occurs above $r = 20$ Å. If the refinement is carried out solely below $r = 20$ Å, the resultant R_{wp} of the refinement is 0.055, as opposed to 0.161. This is an indication that the artifacts which contribute to corrupting the NOMAD data set are, like many of the demonstrated artifacts, cumulative in their effect on the real-space PDF. Just as with the NPDF data set, one can examine the reciprocal-space data in a bank-by-bank manner to better understand the origin of the aberrations. Fig. 12(*b*) shows a small Q range where all six banks overlap, where the as-shown scattered intensity has been rescaled for visual comparison, and plotted concurrently with the merged $S(Q)$ in black in each panel. In each bank, the peak profiles are seen to vary considerably from the final, merged profile. The highest-resolution bank (154°) exhibits significantly finer resolved peaks than the merged data set, while the lower-resolution banks exhibit more broad peak shapes, which also vary as a function of Q.

Whether the data from the banks are merged in an overlapping or non-overlapping (such that no two banks contribute to the same Q value) manner, the inherent peak profile

varies from Gaussian-like at low Q to asymmetric back-to-back exponential-like at high Q. The inability to match the damping profile of the PDF here with the straightforward application of a standard Gaussian 'Q_{damp}' correction is testament to the non-Gaussian nature of the peak shape in some detector banks, as well as differences in resolution between banks.

These NOMAD data confirm what the simulations in §2.5 demonstrate: combining banks of data must be performed very carefully, and differences between banks in reciprocal space can lead to unexpected and cumulative artifacts in real space. As was the case for the NPDF data, deriving an analytical real-space correction to account for the artifacts which have been introduced in reciprocal space is very challenging.

3.3. POWGEN

POWGEN is a third-generation high-resolution powder diffractometer at the Spallation Neutron Source (Huq *et al.*, 2011), which, although not designed for total scattering, has the capability to produce PDFs (Liu *et al.*, 2016). Through the use of a novel guide design and detector placement, POWGEN is able to maintain a much more constant reciprocal-space resolution relative to other cases.

Data from Si-SRM (640d) were collected in a 6 mm-diameter vanadium canister measured with a 1 × 3 cm beam size for 4 h total across two separate histograms with wavelength centers of 0.533 and 4.797 Å, which were then combined to span the desired Q range to generate the PDF. TOF-to-*d* calibration was performed using the *Mantid* software framework (Arnold *et al.*, 2014), with bank-merging and the generation of the PDF carried out using the *PDFgetN* program (Peterson *et al.*, 2000). The resultant total scattering structure function and PDF are shown in Figs. 13(*a*) and 13(*c*), respectively.

The higher resolution available on POWGEN is plainly visible when comparing the reciprocal-space data in Fig. 13(*a*). This results in a reduced damping rate of the associated PDFs, shown in Fig. 13(*d*). It is worth noting that at low r
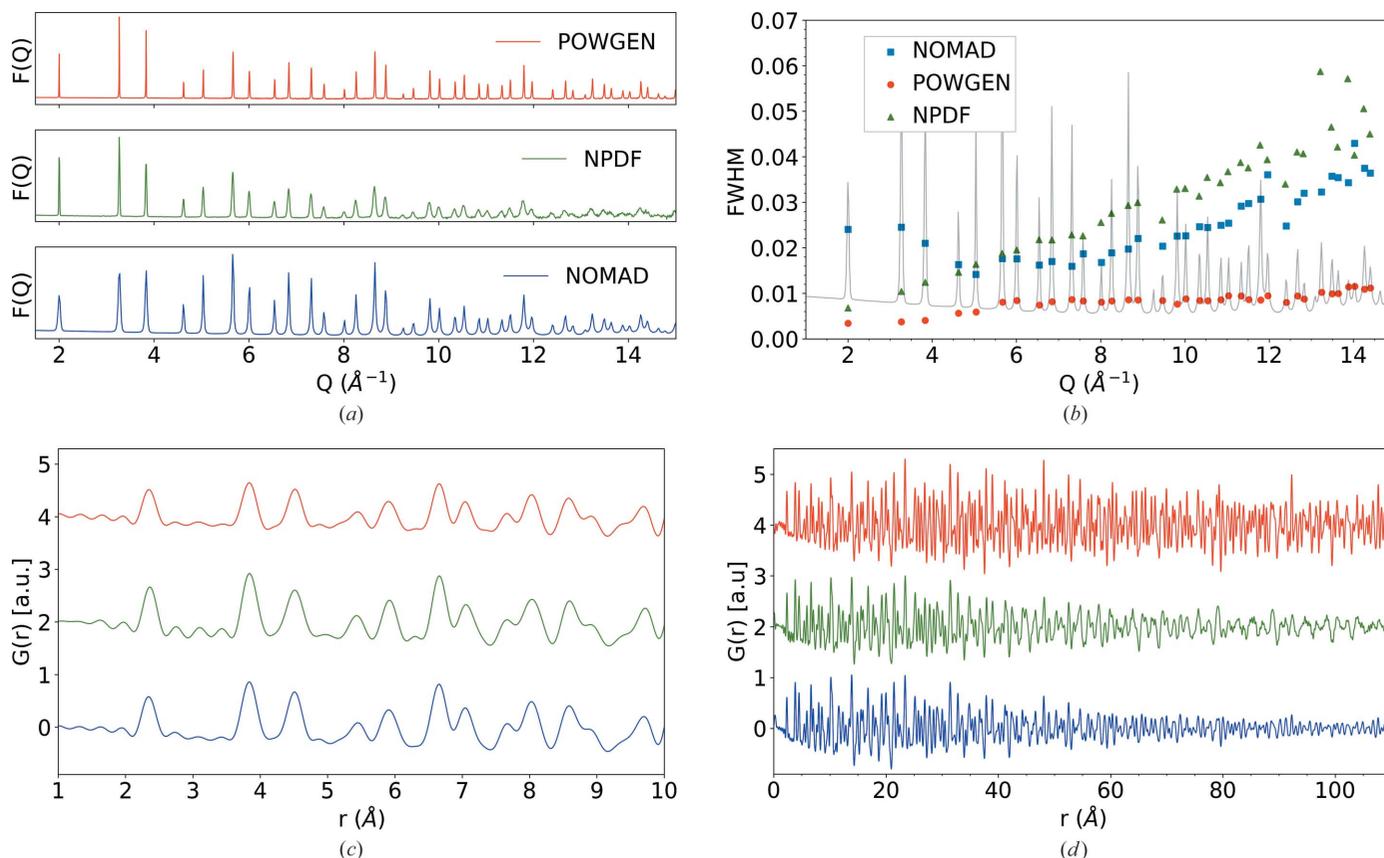
Figure 13
Results of comparing POWGEN, NPDF and NOMAD data from Si showing (*a*) the reduced total scattering structure factor for POWGEN (red), NPDF (green) and NOMAD (blue). (*b*) Performing single-peak fits, the FWHM for the measured total scattering structure factors are compared, overlaying the measured structure-factor data from NOMAD for reference. The associated PDF is shown out to 10 Å (*c*) and 110 Å (*d*) for each instrument.

in Fig. 13(*c*), the PDFs from the three instruments are very similar.

3.4. Comparison of instruments

A summary of the fits from the three instruments is given in the supporting information. To further explore the differences in resolution between NOMAD, NPDF and POWGEN, Gaussian peak fits were performed of the largest, first 18 peaks in the measured pattern up to 18 Å^{-1}. In Fig. 13(*b*), the resultant full width at half-maxima (FWHM) of the as-fit peaks are shown. This plot can help explain a majority of the differences in the PDFs of the three instruments. In the case of POWGEN, not only is a significantly smaller FWHM observed, but the value of the FWHM remains much more consistent over the full reciprocal-space pattern. This results in a PDF which has relatively little damping or real-space broadening present. Although the POWGEN peak-width behavior is not ideal for the generation of PDFs (for which one would desire a constant dQ/Q profile), the overall lower value of the FWHM means there is less of a need for Q_{damp} and Q_{broad} terms in order to model the PDF behavior.

In the NPDF data, significantly more variation in the FWHM of the Si peaks is observed than in the POWGEN data. However, the FWHM behavior follows a much more linear trend. Thus, despite the significant broadening and

damping of the NPDF PDF at high *r*, the trends are readily modeled by the conventional correction of Q_{broad} (which assumes linear dQ/Q).

The NOMAD FWHM displays significant deviations from these two behaviors. The slope of the apparent dQ/Q effect is negative from $Q = 0$ to 5 Å^{-1}, and then increasing. This can be attributed to the onset of the high-resolution detector bank, which begins contributing to the total measured pattern at approximately 5 Å^{-1}. The resultant NOMAD PDF will be far less broadened at high *r*, compared with the NPDF PDF. However, the quality of the refinement of the NOMAD PDF is greatly reduced since the particular reciprocal-space aberrations are not well handled by available corrections. Also, because the overall peak shape of NOMAD reciprocal-space data is more broad than for the POWGEN data, the damping function effect is more severe.

4. Best practices and looking forward

Many details of reciprocal-space data which are directly addressed in Rietveld refinements, such as peak shape and individual bank resolution characteristics, are not commonly addressed by most conventional PDF analysis programs. Yet, as can be seen from the presented simulations and case studies, these artifacts can have important effects on the resultant PDF, primarily at high *r*. The inherent asymmetry

and non-Gaussian nature of the merged reciprocal-space data from modern TOF neutron powder diffraction instruments are not well addressed with the simple models for Gaussian peak shapes and constant dQ/Q behavior commonly employed in fitting PDF data from other sources. Regardless of their origin, asymmetric peak shapes like those shown in §2.3 will always introduce aberrations that will result in a changing lattice constant as a function of real space during data fitting.

Furthermore, the need for accurate calibrations, data merging algorithms and consistent sample positioning cannot be understated, particularly when combining multiple banks of data into a unified reciprocal-space pattern. If improper peak shapes are used in the procedures generating the TOF-to-d conversion (such as using peak centroid positions instead of true peak centers), it is possible that the instrument calibration will be adversely affected. Any misalignment of the sample or other pixel-to-d spacing offsets will produce artifacts in the PDF such as those shown in §2.5, which are intractable and likely irreversible during data analysis in real space. Although not presently possible, future methods similar to those in the *MCGRtof* program (Tucker *et al.*, 2001) may be capable of addressing some or all of these effects and work in this area is ongoing (Tucker, 2017). The data sets available in the supporting information are provided in part with the hope that they will aid the community in developing solutions to the demonstrated problems.

It is important to note that, as seen in both the simulations and the case studies, the PDF is surprisingly resilient in maintaining an accurate representation of local structure up to approximately 20 Å, almost independent of any corrupting or complicating factors in the reciprocal-space data presented here. Also, for *in situ* studies, where a local feature of interest is being tracked or relative comparisons are being made between samples measured under the same conditions, these concerns are somewhat mitigated as many unaccounted for artifacts or aberrations in the data (*e.g.* lattice contractions, damping) will be the same among all the measured data.

What can be said of fits above 20 Å in real space? With available analysis programs, the authors recommend the following best practices for neutron TOF total scattering studies. A known standard should always be measured concurrently with samples of interest, and used to derive any Q_{damp}- and Q_{broad}-like terms, which should then be fixed for subsequent refinements of experimental samples. Furthermore, if the overall goodness of fit to standard data worsens as the r_{max} is increased (as is the case in all examples presented herein), this is a clear indication that data should not be quantitatively refined above that length scale. When possible, the results of Rietveld refinements should be compared with the results of fits to the PDF, with attention paid to the cases where the refined values deviate. If using flexible software such as *TOPAS*, *DISCUS* or *Diffpy-CMI* for PDF refinements, explicit damping and broadening terms can be defined beyond the standard single Gaussian. In initial testing, it has been found that simply allowing the refined damping term to be a sum of two Gaussians, where widths are independent vari-

ables, greatly improves the quality of NOMAD data refinements. A full analysis of best practices for refining NOMAD data is beyond the scope of this article, and will be presented elsewhere.

To be clear, the long-term solution would be to derive precisely how the peak shape of a given instrument directly impacts the as-calculated PDF. These suggested numerical corrections would allow a path forward in the same manner as the currently used Gaussian Q_{damp} and Q_{broad} corrections, without major modifications to available code bases.

Given the average computing power of a desktop computer is roughly 500 times greater than it was the year *PDFfit* was initially published (Proffen & Billinge, 1999), as well as the broad availability of massively parallel computing resources, it is perhaps not unreasonable to re-examine and re-define the approach employed in analysis of total scattering data. With a seemingly small addition, the ability to model Q_{min}, Q_{max}, and ΔQ effects, one could fit PDFs generated from a narrowly defined range in reciprocal space, even though the PDF itself could span as wide a real-space length scale as desired. Although such narrow-bandwidth PDFs would be visually abstracted, this approach would allow the real-space data to be broken apart into different principal components. The net sum of these separate PDFs would still reproduce the conventional PDF. The advantage here is that a narrowly defined band of reciprocal space may be selected where the resolution and peak-shape profile are well defined and consistent, allowing for straightforward application of the analytically defined real-space corrections. In much the same way that Rietveld refinements are performed simultaneously on different banks of reciprocal-space data, PDF refinements could be performed on a single model applied to multiple PDFs (high-resolution, low-resolution, different Q ranges, different degrees of signal to noise *etc*).

It is entirely possible that the most accurate way of modeling total scattering data will ultimately involve, as is presently possible in *RMCProfile* and *TOPAS*, calculating both the real- and reciprocal-space data from a unified model, applying various corrections in whatever space is most appropriate (*e.g.* peak shape applied in reciprocal space, nanoparticle shape effects in real space) and fitting directly to measured data.

The importance of minimizing intractable detector tube drift or variation in TOF-to-d conversion was illustrated in §2.5, and should be considered in future instrument design. Note that wider Q coverage does not necessarily result in higher-quality PDFs, especially if these banks have greatly varying resolution functions. In principle, a complex geometry of detector banks could be accounted for during analysis through the methods just proposed, although solving a problem with a hardware solution may avoid complex mitigation algorithms in software. Conventional requirements for high-quality measurements of diffuse scattering data still very much apply (low and consistent background), and detailed implications for factors such as absorption and multiple scattering must be considered. However, prioritizing constancy in the resolution functions, or keeping this behavior linear, could

feature articles

greatly simplify the modeling of real-space data for neutron TOF-derived PDFs.

References

Aksel, E., Forrester, J. S., Nino, J. C., Page, K., Shoemaker, D. P. & Jones, J. L. (2013). *Phys. Rev. B*, **87**, 104113.
Allieta, M., Scavini, M., Naldoni, A., Coduri, M., Cappelli, S., Oliva, C., Santangelo, S., Triolo, C., Patané, S., Lascialfari, A. & Scagnoli, V. (2015). *Phys. Rev. B*, **92**, 054202.
Aoun, B. (2016). *J. Comput. Chem.* **37**, 1102–1111.
Arnold, O. *et al.* (2014). *Nucl. Instrum. Methods Phys. Res. A*, **764**, 156–166.
Avdeev, M., Jorgensen, J., Short, S. & Von Dreele, R. B. (2007). *J. Appl. Cryst.* **40**, 710–715.
Billinge, S. & Egami, T. (1993). *Phys. Rev. B*, **47**, 14386–14406.
Billinge, S. J. & Levin, I. (2007). *Science*, **316**, 561–565.
Božin, E. S., Malliakas, C. D., Souvatzis, P., Proffen, T., Spaldin, N. A., Kanatzidis, M. G. & Billinge, S. J. (2010). *Science*, **330**, 1660–1663.
Bruker (2015). *TOPAS* V6.0. Bruker AXS, Karlsruhe, Germany.
Cervellino, A., Frison, R., Bertolotti, F. & Guagliardi, A. (2015). *J. Appl. Cryst.* **48**, 2026–2032.
Charles, D. S., Feygenson, M., Page, K., Neuefeind, J., Xu, W. & Teng, X. (2017). *Nat. Commun.* **8**, 15520.
Checchia, S., Scavini, M., Allieta, M., Brunelli, M., Ferrero, C. & Coduri, M. (2015). *Powder Diffr.* **30**(S1), S119–S126.
Chung, J. S. & Thorpe, M. (1999). *Phys. Rev. B*, **59**, 4807–4812.
Coduri, M., Scavini, M., Allieta, M., Brunelli, M. & Ferrero, C. (2013). *Chem. Mater.* **25**, 4278–4289.
Dinnebier, R. E. (2008). *Powder Diffraction: Theory and Practice.* London: Royal Society of Chemistry.
Egami, T. & Billinge, S. J. (2012). *Underneath the Bragg Peaks: Structural Analysis of Complex Materials*, 2nd ed. Amsterdam: Elsevier.
Fabini, D. H., Siaw, T. A., Stoumpos, C. C., Laurita, G., Olds, D., Page, K., Hu, J. G., Kanatzidis, M. G., Han, S. & Seshadri, R. (2017). *J. Am. Chem. Soc.* **139**, 16875–16884.
Farrow, C. L. & Billinge, S. J. L. (2009). *Acta Cryst.* A**65**, 232–239.

Farrow, C., Juhas, P., Liu, J., Bryndin, D., Božin, E., Bloch, J., Proffen, T. & Billinge, S. (2007). *J. Phys. Condens. Matter*, **19**, 335219.
Farrow, C. L., Shaw, M., Kim, H., Juhás, P. & Billinge, S. J. (2011). *Phys. Rev. B*, **84**, 134105.
Finger, L. W., Cox, D. E. & Jephcoat, A. P. (1994). *J. Appl. Cryst.* **27**, 892–900.
Gagin, A., Allen, A. J. & Levin, I. (2014). *J. Appl. Cryst.* **47**, 619–629.
Harrington, R., Hausner, D. B., Xu, W., Bhandari, N., Michel, F. M., Brown, G. E. Jr, Strongin, D. R. & Parise, J. B. (2011). *Environ. Sci. Technol.* **45**, 9883–9890.
Hattori, T., Katayama, Y., Machida, A., Otomo, T. & Suzuya, K. (2010). *J. Phys. Conf. Ser.* **215**, 012024.
Hill, A. H. & Allieta, M. (2013). *Phys. Chem. Chem. Phys.* **15**, 8583–8588.
Hillis, B. G., Losey, B. P., Weng, J., Ghaleb, N., Hou, F. & Martin, J. D. (2016). *Crystals*, **7**, 11.
Howe, M., McGreevy, R. & Howells, W. (1989). *J. Phys. Condens. Matter*, **1**, 3433–3451.
Hull, S., Smith, R., David, W., Hannon, A., Mayers, J. & Cywinski, R. (1992). *Physica B*, **180–181**, 1000–1002.
Huq, A., Hodges, J. P., Gourdon, O. & Heroux, L. (2011). *Z. Kristallogr. Proc.* **1**, 127–135.
Jeong, I.-K., Graf, M. J. & Heffner, R. H. (2005). *J. Appl. Cryst.* **38**, 55–61.
Jeong, I.-K., Lee, J. & Heffner, R. (2008). *Appl. Phys. Lett.* **92**, 172911.
Jiang, B., Grande, T. & Selbach, S. M. (2017). *Chem. Mater.* **29**, 4244–4252.
Keen, D. A. & Goodwin, A. L. (2015). *Nature*, **521**, 303–309.
Lan, S., Wei, X., Zhou, J., Lu, Z., Wu, X., Feygenson, M., Neuefeind, J. & Wang, X.-L. (2014). *Appl. Phys. Lett.* **105**, 201906.
Larson, A. C. & Von Dreele, R. B. (1994). *General Structure Analysis System, GSAS*. Report LAUR 86-748. LANSCE, MS-H805, Los Alamos, New Mexico, USA.
Liu, J., Huq, A., Moorhead-Rosenberg, Z., Manthiram, A. & Page, K. (2016). *Chem. Mater.* **28**, 6817–6821.
Liu, J., Olds, D., Peng, R., Yu, L., Foo, G. S., Qian, S., Keum, J., Guiton, B. S., Wu, Z. & Page, K. (2017). *Chem. Mater.* **29**, 5591–5604.
Mancini, A. & Malavasi, L. (2015). *Chem. Commun.* **51**, 16592–16604.
Mendenhall, M. (2016). Powder diffraction SRMs. https://www-s.nist.gov/srmors/certificates/640D.pdf.
Neuefeind, J., Feygenson, M., Carruth, J., Hoffmann, R. & Chipley, K. K. (2012). *Nucl. Instrum. Methods Phys. Res. B*, **287**, 68–75.
Olds, D., Lawler, K. V., Paecklar, A., Liu, J., Page, K., Peterson, P. F., Forster, P. M. & Neilson, J. R. (2017). *Chem. Mater.* **30**, 296–302.
Olds, D., Wang, H.-W. & Page, K. (2015). *J. Appl. Cryst.* **48**, 1651–1659.
Page, K., Hood, T. C., Proffen, Th. & Neder, R. B. (2011). *J. Appl. Cryst.* **44**, 327–336.
Page, K., Kolodiazhnyi, T., Proffen, T., Cheetham, A. K. & Seshadri, R. (2008). *Phys. Rev. Lett.* **101**, 205502.
Peterson, P. F., Božin, E. S., Proffen, Th. & Billinge, S. J. L. (2003). *J. Appl. Cryst.* **36**, 53–64.
Peterson, P. F., Gutmann, M., Proffen, Th. & Billinge, S. J. L. (2000). *J. Appl. Cryst.* **33**, 1192.
Playford, H. Y., Owen, L. R., Levin, I. & Tucker, M. G. (2014). *Annu. Rev. Mater. Res.* **44**, 429–449.
Proffen, T. (2015). *NXProc: Program for Analysis and Reduction of Time-of-Flight Powder Diffraction Data*. Los Alamos National Laboratory, Los Alamos, New Mexico, USA.
Proffen, T. & Billinge, S. J. L. (1999). *J. Appl. Cryst.* **32**, 572–575.
Proffen, T., Egami, T., Billinge, S., Cheetham, A., Louca, D. & Parise, J. (2002). *Appl. Phys. A Mater. Sci. Process.* **74**, s163–s165.
Proffen, Th. & Neder, R. B. (1997). *J. Appl. Cryst.* **30**, 171–175.
Qiu, X., Božin, E. S., Juhas, P., Proffen, T. & Billinge, S. J. (2004). *J. Appl. Cryst.* **37**, 110–116.

Qiu, X., Proffen, T., Mitchell, J. & Billinge, S. (2005). *Phys. Rev. Lett.* **94**, 177203.

Ramezanipour, F., Greedan, J. E., Cranswick, L. M., Garlea, V. O., Donaberger, R. L. & Siewenie, J. (2012). *J. Am. Chem. Soc.* **134**, 3215–3227.

Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.

Shamblin, J., Feygenson, M., Neuefeind, J., Tracy, C. L., Zhang, F., Finkeldei, S., Bosbach, D., Zhou, H., Ewing, R. C. & Lang, M. (2016). *Nat. Mater.* **15**, 507–511.

Shannon, C. E. (1949). *Proc. IRE*, **37**, 10–21.

Skinner, L., Benmore, C. J., Weber, J., Du, J., Neuefeind, J., Tumber, S. & Parise, J. B. (2014). *Phys. Rev. Lett.* **112**, 157801.

Toby, B. H. & Egami, T. (1992). *Acta Cryst.* A**48**, 336–346.

Tucker, M. (2017). Personal communication.

Tucker, M. G., Dove, M. T. & Keen, D. A. (2001). *J. Appl. Cryst.* **34**, 780–782.

Tucker, M. G., Keen, D. A., Dove, M. T., Goodwin, A. L. & Hui, Q. (2007). *J. Phys. Condens. Matter*, **19**, 335218.

Ushakov, S. V., Navrotsky, A., Weber, R. J. & Neuefeind, J. C. (2015). *J. Am. Ceram. Soc.* **98**, 3381–3388.

Usher, T.-M., Levin, I., Daniels, J. E. & Jones, J. L. (2015). *Sci. Rep.* **5**, 14678.

Von Dreele, R. B., Jorgensen, J. D. & Windsor, C. G. (1982). *J. Appl. Cryst.* **15**, 581–589.

Wang, H.-W., Daemen, L. L., Cheshire, M. C., Kidder, M. K., Stack, A. G., Allard, L. F., Neuefeind, J., Olds, D., Liu, J. & Page, K. (2017). *Chem. Commun.* **53**, 2942–2945.

Wang, H.-W., Wesolowski, D. J., Proffen, T. E., Vlcek, L., Wang, W., Allard, L. F., Kolesnikov, A. I., Feygenson, M., Anovitz, L. M. & Paul, R. L. (2013). *J. Am. Chem. Soc.* **135**, 6885–6895.

Williams, W., Ibberson, R., Day, P. & Enderby, J. (1997). *Physica B*, **241–243**, 234–236.

Wu, Z., Lan, S., Wei, X., Olds, D., Page, K., Shen, B. & Wang, X.-L. (2017). *Physica B*, doi: 10.1016/j.physb.2017.12.030.

Yang, K., Cai, Z., Tyagi, M., Feygenson, M., Neuefeind, J. C., Moore, J. S. & Zhang, Y. (2016). *Chem. Mater.* **28**, 3227–3233.