JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Faster and lower-dose X-ray reflectivity measurements enabled by physics-informed modeling and artificial intelligence co-refinement

David Mareček,[a] Julian Oberreiter,[a] Andrew Nelson[b] and Stefan Kowarik[a]*

[a]Physikalische und Theoretische Chemie, Universität Graz, Heinrichstraße 28, Graz, 8010, Austria, and [b]ANSTO, Locked Bag 2001, Kirrawee DC, NSW 2232, Australia. *Correspondence e-mail: stefan.kowarik@uni-graz.at

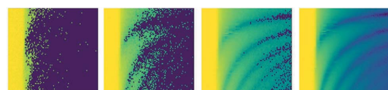
Received 10 February 2022
Accepted 11 August 2022

Edited by A. Barty, DESY, Hamburg, Germany

Keywords: neutron reflectivity; X-ray reflectivity; neural networks; co-refinement; *in situ* measurement.

Supporting information: this article has supporting information at journals.iucr.org/j

An approach is presented for analysis of real-time X-ray reflectivity (XRR) process data not just as a function of the magnitude of the reciprocal-space vector q, as is commonly done, but as a function of both q and time. The real-space structures extracted from the XRR curves are restricted to be solutions of a physics-informed growth model and use state-of-the-art convolutional neural networks (CNNs) and differential evolution fitting to co-refine multiple time-dependent XRR curves $R(q,t)$ of a thin film growth experiment. Thereby it becomes possible to correctly analyze XRR data with a fidelity corresponding to standard fits of individual XRR curves, even if they are sparsely sampled, with a sevenfold reduction of XRR data points, or if the data are noisy due to a 200-fold reduction in counting times. The approach of using a CNN analysis and of including prior information through a kinetic model is not limited to growth studies but can be easily extended to other kinetic X-ray or neutron reflectivity data to enable faster measurements with less beam damage.



OPEN ACCESS

Published under a CC BY 4.0 licence

1. Introduction

X-ray reflectivity (XRR) and neutron reflectivity (NR) are well known techniques for studying ultra-thin layers (Tolan, 1999; Holý *et al.*, 1999; Braslau *et al.*, 1988; Skoda *et al.*, 2017; Russell, 1990; Kowarik *et al.*, 2006; Pietsch *et al.*, 2004). Due to the penetrating nature of X-rays and neutrons, the measurements can be taken in various environments, such as in a vacuum, or in liquids or gases. This has led to numerous applications in many fields that profit from *in situ* and real-time measurements, such as the solid–electrolyte interface growth during the process of lithiation in lithium-ion batteries (Liu *et al.*, 2016; Cao *et al.*, 2016) and studies of layer-by-layer growth (Kowarik *et al.*, 2006; Hanke *et al.*, 2012; Joress *et al.*, 2018; Krause *et al.*, 2019; Bommel *et al.*, 2014). Real-time measurements usually consist of repeated reflectivity scans at defined time intervals, and each measurement results in a single reflectivity curve that represents the growth process at a certain point in time. From these scans, structural film properties including thickness, roughness and scattering length density (SLD) can be extracted as a function of time t, although the analysis is non-trivial and fitting the reflectivity $R(q,t)$ only as a function of q wastes information available from data at adjacent time steps. An important challenge for XRR and NR curves is the reconstruction of real-space sample structures from reciprocal-space scattering data and the reliability of this procedure. Using the Parratt formalism (Parratt, 1954) to calculate reflectivity curves allows one to easily obtain the shape of the reflectivity curve for a given thin

layer structure. However, the inverse reconstruction of sample parameters from an XRR curve is highly non-trivial (López García & Rivero, 2021) and multiple non-unique solutions are possible. Results are usually obtained by iterative least mean squares (LMS) fitting algorithms or advanced genetic algorithms, which are available in various XRR and NR programs (Nelson, 2006; Nelson & Prescott, 2019; Björck & Andersson, 2007; Danauskas *et al.*, 2008; Lazzari, 2002).

In contrast to the traditional, manual fitting processes, artificial intelligence techniques in the form of artificial neural networks (NNs) (Bishop, 1994) have become available recently for fast, accurate fits of XRR or NR curves without human input (Greco *et al.*, 2019, 2021; Liu *et al.*, 2019; Park *et al.*, 2017; Vecsei *et al.*, 2019; Chen *et al.*, 2021; Kim & Lee, 2021; Mironov *et al.*, 2021; Doucet *et al.*, 2021; Andrejevic *et al.*, 2022; Loaiza & Raza, 2021). The convolutional neural network (CNN) approach for inverting XRR data has become popular due to better computer performance resulting from relatively cheap modern graphics and central processing units (GPUs and CPUs), as well as the accessibility in recent years of free CNN programming frameworks such as *TensorFlow* (https://www.tensorflow.org/) and *PyTorch* (https://pytorch.org/). A drawback of the CNN approach is that it needs a large volume of training data for CNN training; however, in the case of XRR and NR, these training data can be easily simulated. A CNN trained on synthetic data can then be applied to experimental data including time series of XRR curves. Owing to the comparatively large number of data points, time-series measurements can profit from automated CNN analysis and co-refinement.

In real-time growth, most of the parameters of a thin film structure change continuously and are dependent on each other; for example, many individual thickness parameters for each curve can be replaced by a growth rate. With a corresponding growth model, we can re-parametrize (Campbell *et al.*, 2018) the structural models of the thin film at each time step with very few parameters rather than using individual parameter sets for each time step. Moreover, the XRR fits of individual time steps are lacking the improvement of parameter estimates that co-refinement (McCluskey *et al.*, 2019) with previous or following curves can bring, and therefore the prediction is noisier than one obtained by using an underlying growth model. In this paper we report an implementation of a growth model, whose parameters are determined with the *refnx* software (Nelson & Prescott, 2019) in a growth model co-refinement of the experimental growth data. Secondly, we trained a CNN with synthetic data generated via the growth model and then used the CNN to predict the structural evolution from the experimental growth data. We show that we can significantly speed up the measurement and reduce the X-ray exposure of the sample through growth model co-refinement and CNN approaches, because they allow one to only use sparse sampling and shorten the exposure times. This achievement is enabled through (i) the incorporation of prior knowledge in the form of a growth model, (ii) re-parametrization of the fit problem, and (iii) co-refinement or CNN image recognition of multiple XRR curves at once.

2. Methods

2.1. Growth model

Using a (rate-equation) model for kinetic processes is crucial for our goal of co-refinement of real-time XRR data sets, and for our example of crystalline thin film deposition a growth model is needed. Crystal growth of highly perfect crystalline layers ideally proceeds in a layer-by-layer fashion, but also deviations from a layer-by-layer type of growth and roughening are common. Numerous models have been developed over recent decades which can describe this layer-by-layer growth of thin films (Cohen *et al.*, 1989; Braun *et al.*, 2003; Trofimov *et al.*, 1997). These models are quite general and can all fit a variety of growth scenarios from layer-by-layer growth to layer-plus-island growth or rough 3D growth. Here we employ a model used by Woll *et al.* (2011) for organic molecular thin films, even though we note that the above-mentioned models could also be used. This model generates the surface roughness and thickness evolutions with realistic behavior, such as oscillating roughness for layer-by-layer growth. The growth rate in the model is described via the parameter G_n for the nth layer. A parameter $\theta_{n,\rm cr}$ defines the critical layer coverage of the nth layer, where the n+1-th layer starts to nucleate on the top of the nth layer. Once nucleated, the feeding zone ξ_n then is the size of the zone on top of the nth layer, where molecules contribute to the growth of the n+1-th layer. Molecules outside the feeding zone will diffuse over the edge into the nth layer.

$$\frac{{\rm d}\theta_n}{{\rm d}t} = \begin{cases} R_1(1-\theta_1) + R_{n>1}(\theta_1 - \xi_1), & n = 1, \\ R_{n>1}(\xi_{n-1} - \xi_n), & n > 1, \end{cases} \tag{1}$$

where R_n is the growth rate of the nth layer and the feeding zone size is given as

$$\xi_n = \begin{cases} 0, & \theta_n < \theta_{n,\rm cr}, \\ 1 - \exp(-\{[-\ln(1-\theta_n)]^{1/2} - [-\ln(1-\theta_{n,\rm cr})]^{1/2}\}^2). \end{cases} \tag{2}$$

On the basis of this model, the thickness and roughness of the growing thin film can be calculated at each moment in time, and together with the SLD, used to generate the corresponding XRR curves. The SLD parameter is set as a fixed value in our model even though different structural phases in the first monolayers slightly modify the SLD (Kowarik *et al.*, 2006; Frank *et al.*, 2014). Here, we intentionally use a simple model of a fixed SLD for thin films and describe partially filled, growing layers via an oscillating film roughness during initial layer-by-layer growth. With the layer-by-layer growth model, the evolution of the thin film can be defined with ten parameters. Four growth rate parameters G_n were used to account for different sticking coefficients on the substrate and the first layers, where n = 1, 2, 3 stands for the growth rate of the first, second and third layers, while G_4 is used as the growth rate for all the following layers. The next parameter to be defined is $\theta_{\rm cr}$ for each layer. To reduce the parameter space we calculated $\theta_{n,\rm cr}$ from the following equation:

$$\theta_{n,\mathrm{cr}} = \begin{cases} a\dfrac{0.5\tanh[-0.5(t+b)]+0.5}{0.5\tanh(-0.5b)+0.5}, & \theta_{n,\mathrm{cr}} \geq c, \\ (c-f)\exp[-d(t-t_\mathrm{c})]+f, & \theta_{n,\mathrm{cr}} < c, \end{cases} \quad (3)$$

where a and b parametrize the time evolution of $\theta_{n,\mathrm{cr}}$ with growth time t. When the $\theta_{n,\mathrm{cr}}$ value reaches a threshold, represented by a crossover parameter c, the function for $\theta_{n,\mathrm{cr}}$ is switched to an exponentially decaying function. Here, the parameter d is the steepness of exponential decay, f is the asymptotic value for large t, and t_c is the time when the $\theta_{n,\mathrm{cr}}$ value reaches the crossover value. All ten parameters were randomly but uniformly distributed to generate a synthetic data set for NN training and used as training labels. Note that the above growth model and parametrization of θ_cr restrict the possible growth scenarios we can fit and might lead to wrong fit results if the real scenario is not covered. However, the parameter ranges are chosen to be wide enough to include 3D growth and roughening from the very first layer, to layer-by layer growth followed by roughening as well as layer-by-layer growth with continuing roughness oscillations over the whole thickness range. The ten parameters are collectively denoted as growth model parameters in the following.

2.2. XRR curve simulation

We simulated XRR layer-by-layer growth using an adaptation of the optical matrix method (Abelès, 1950; Heavens, 1960), which is computationally more efficient than the recursive Parratt formalism (Parratt, 1954). For this purpose, parts of the *Refl1D* (Kienzle *et al.*, 2022) source code were used. We assumed that our thin film sample consists of three thin film layers: two for the substrate (silicon and native oxide) plus the deposited thin film. The properties of Si/SiO_2 substrates are known from our prior work and thus we used a constant roughness for the SiO_2 and Si layers of 1 and 2.5 Å, respectively, and a native oxide thickness of 10 Å. Furthermore, the SLDs of those layers were assumed to be constant with values of 17.8×10^{-6} and 20×10^{-6} Å^{-2}, respectively (Greco *et al.*, 2019). The XRR curves were simulated in a q range between 0.01 and 0.14 Å^{-1} at 109 equally spaced points.

For the following, we appended all individual XRR curves corresponding to a given growth scenario into a single XRR time-series matrix $R(q, t)$. Due to the shape of the experimental diindenoperylene (DIP) data, $R(q, t)$ contains 109 q points and 80 time steps; we also created a synthetic data set with the same dimensions. Each one of the curves corresponds to a certain thickness, roughness and SLD of the film layer during the growth. The fixed SLD and the evolution of the thickness d(t) and roughness $\sigma(t)$ are later on in the text collectively referred to as a thin film growth scenario. The growth of the simulated thin film started from a thickness of 0 Å and the final thickness was varied from 180 to 450 Å using the growth rate parameters (G_1, G_2, G_3, G_4). The final roughness of the film was allowed to vary from 8 to 40 Å. The SLD was set in a range from 7×10^{-6} to 18.5×10^{-6} Å^{-2}. We simulated 25 000 $R(q, t)$ data sets to get a sufficient quantity of data for the CNN training. The simulated $R(q, t)$ data were split into training and validation data sets, where the validation data set contained 10% of the generated data, *i.e.* 2500 simulated data sets $R(q, t)$. The data generation took 90 min on 20 independent CPU threads; this procedure can be faster if *refnx* is used. We took the natural logarithm of all generated data sets and then normalized them to unity before using them as an input for the CNN. This method of normalization is widely used, including in previous work by us; however, we note that there are alternatives for normalizing reflectivity data (Loaiza & Raza, 2021; Doucet *et al.*, 2021; Kim *et al.*, 2021). In this way, we reduced the value spread of the input data set, because the CNN training is faster for input and

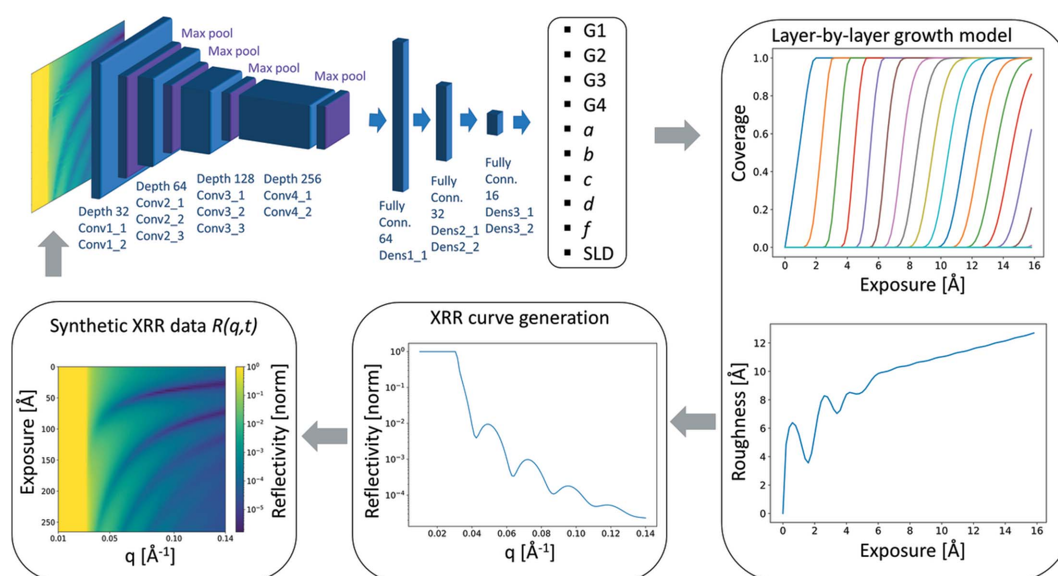


Figure 1
From a time-dependent XRR data set $R(q, t)$ a CNN predicts ten parameters of a time-dependent layer-by-layer growth model. From these parameters, the thin film growth scenario with thickness and roughness evolution is calculated. For different growth scenarios, we create simulated XRR and synthetic $R(q, t)$ data sets that can be used to train the CNN before applying it to real data.

output data in the range of 0 to 1. Therefore, all growth model parameters which are the output of the CNN were also normalized to unity.

2.3. Neural network

We utilized 2D CNNs, commonly used for image recognition, to predict growth model parameters from the 2D data set $R(q, t)$. We employed a smaller version of the standard VGG16 CNN which is a common architecture used for image recognition (Simonyan & Zisserman, 2015). In this architecture, 2D data are processed from a set of input neurons to a set of output neurons through hidden layers of neurons. The input layer of the CNN takes the $R(q, t)$ data. The architecture of our CNN is graphically depicted in Fig. 1. As an activation function, a simple ReLU (rectified linear function) unit was chosen for all layers. The optimization algorithm employed in this work is adaptive moment estimation (ADAM) (Kingma & Ba, 2015). Our code was written in Python 3.7 with the use of the *TensorFlow* software library. The CNN was trained for 3000 epochs (4 h) on a GPU (Nvidia GeForce RTX 2080 Ti) and Intel Core i5-9600K CPU.

2.4. Experimental data

As an experimental example, we used the previously published (Kowarik *et al.*, 2006) real-time XRR data set related to growth of the organic semiconductor molecule DIP via organic molecular beam deposition on an Si/SiO_2 wafer at 403 K substrate temperature. The experiments were performed at beamline ID10B at the ESRF in Grenoble, France, in a small ultra-high vacuum chamber, equipped with a Be window, effusion cells and thickness monitor, at X-ray wavelengths of 0.903 Å. The original experimental data set was measured with 52 q points per curve and interpolated to 109 points per curve in the same q range. For our CNN approach, the experimentally obtained $R(q, t)$ was used as CNN input, from which the CNN predicted growth model parameters and the corresponding thin film growth scenario. For the CNN performance evaluation on the experimental data, the thin film growth scenario determined by the model was compared with a manual LMS fit using the *Parratt32* software (Goedel *et al.*, 1999). The output of *Parratt32* is the thickness, roughness and SLD for every curve, which is more than 240 parameters. The only varying parameters were parameters corresponding to the thin DIP layer, and the q range was the same as for the CNN, from 0.01 to 0.14 $Å^{-1}$. *Refnx* (Nelson & Prescott, 2019) was used to co-refine all XRR curves, with the growth model being incorporated into the analysis, reducing the number of fit parameters to 10. In this last approach the varying growth model parameters are used to calculate the thickness/roughness of the system at each of the time steps.

3. Results

We used our CNN to predict the growth model parameters from the experimental $R(q, t)$ shown in Fig. 2(a). Then these parameters were used to generate a new fitted $R_{CNN}(q, t)$ [Fig. 2(b)], and we compared the two $R(q, t)$'s by calculating a χ^2 value as a measure of goodness of fit, with χ^2 defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{\left\{\log[R_{\rm fit}(q,t)]_i - \log[R_{\rm exp}(q,t)]_i\right\}^2}{\log[R_{\rm exp}(q,t)]_i}. \qquad (4)$$

From the comparison in Figs. 2(a) and 2(b), it is obvious that the Kiessig fringes in the experiment and in $R_{CNN}(q, t)$ are at the same position. Moreover, the damping of oscillations increases at higher q and t due to increased roughness in both $R(q, t)$'s. The reflectivity range and minimum reflectivity values agree very well in the two images. The χ^2 deviation between the experiment and $R_{CNN}(q, t)$ was calculated as 5.0. For the *Parratt32* fit where we fit curves individually (individual fit), we found $\chi^2_{\rm Parratt}$ = 1.7. Finally, with an automatic growth model co-refinement using *refnx* we obtained $\chi^2_{\rm co\text{-}refine}$ = 9.2. The fitted XRR curves $R_{\rm Parratt}(q, t)$ and $R_{\rm co\text{-}refine}(q, t)$ are shown in Figs. 2(c) and 2(d) and also are very similar to the experimental data. All the χ^2 errors were calculated by the same method, without use of *Parratt32* or *refnx* implemented error calculations. The individual *Parratt32* curve fit adapts to spurious experimental noise features (red circle) which partly explains the lower χ^2 value of the individual curve *Parratt32* fit compared with the CNN fit. The results are plotted as a function of q and exposure. The exposure is tj, where t is the time and j is the molecular flux. In an experimental setup, the molecular flux does not need to be always constant and it is straightforward to implement these changes when the exposure is used as the axis. Note that the *Parratt32* manual fit is a slow process (hours) where the user selects bounds and starting points in the fit of thickness, roughness and SLD of the DIP layer for every curve. The growth model co-refinement approach is significantly faster (30 min) than the *Parratt32*

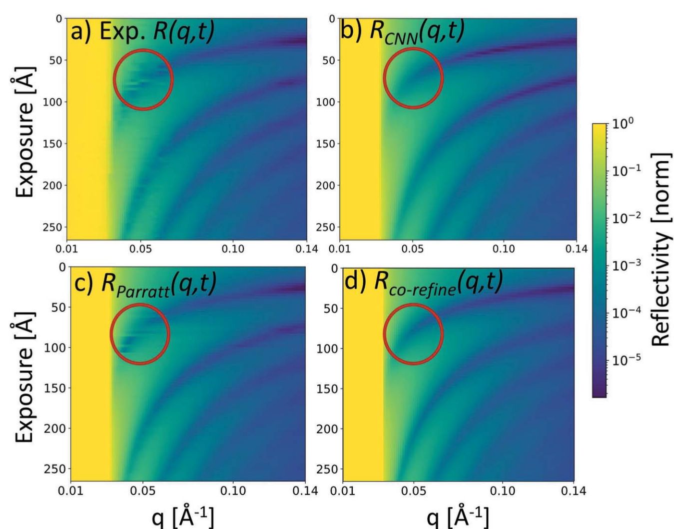


Figure 2
The experimental $R(q, t)$ in (a) is compared with different fitted XRR curves. (b) shows the CNN prediction with a χ^2 deviation value of 5.0; a manual *Parratt32* fit (c) has a lower deviation of $\chi^2_{\rm Parratt}$ = 1.7 as it also fits experimental noise (red circle). The growth model co-refinement (d) yields $\chi^2_{\rm co\text{-}refine}$ = 9.2.

manual fit but still much slower than a CNN prediction, which takes up to 55 ms, when not counting the 5.5 h of CNN data generation and training time.

Besides the comparison of raw experimental data and the various fits of $R(q, t)$, we also evaluated the thin film growth scenarios obtained with the CNN, independent $R(q)$ fits and the growth model co-refinement [Figs. 3(*a*)–3(*c*)]. In the plots of the thickness evolution during growth [Fig. 3(*a*)] and the SLD [Fig. 3(*b*)] the CNN prediction, independent fit and growth model co-refinement closely align. The roughness comparison in Fig. 3(*c*) shows that all fits are roughly comparable with some deviations for the growth model co-refinement method. The oscillations of the roughness visible in the CNN fit at the beginning of the growth are due to the layer-by-layer growth as filled layers are smoother and half-filled layers during layer-by-layer growth are rougher. Overall, the growth model based on the CNN predictions and the growth model co-refinement can explain the real-time XRR curves with far fewer parameters and produce results that come close to manual fits in terms of the χ^2 metric of the fit quality and match well the results of the thin film growth scenarios.

3.1. Reducing the number of data points in a measurement – sparse sampling

In the following, we demonstrate the high performance of the CNN and the growth model co-refinement approach on a sparsely sampled data set with a very low quantity of data. The $R(q, t)$ of DIP has been reduced using a dropout layer from *Keras* (Ketkar, 2017) to remove a certain percentage of the experimental data points by setting them to zero. These points can therefore be skipped in a faster measurement. For the CNN to work with reduced data sets, it is necessary to further train the CNN so that it learns to ignore unavailable data points. We retrained the previously optimized CNN with a new data set of 120 000 $R(q, t)$'s which had a varying degree of dropout reduction from 0 to 99.9%. The CNN was retrained for 30 epochs without any sign of overfitting. The experimental $R(q, t)$ with a reduced number of data points was then used as an input for the retrained CNN.

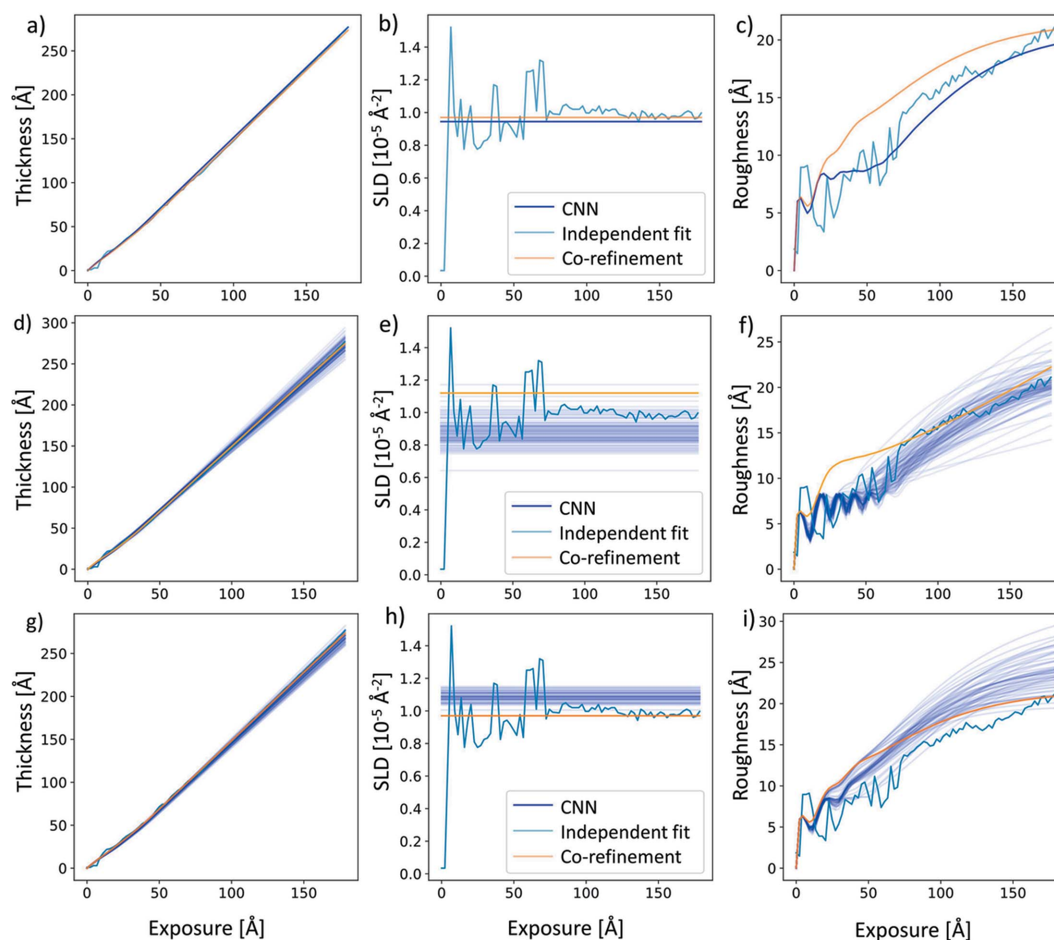


Figure 3
(*a*) Thickness, (*b*) SLD and (*c*) roughness predictions derived from the experimental XRR data $R(q, t)$ in Fig. 2 agree reasonably well between standard individual curve fits, growth model based CNN and co-refinement results. (*d*)–(*f*) For a reduced data set of only 200 reflectivity points in $R(q, t)$, the CNN and the growth model co-refinement fit still fit the data set correctly. A set of CNN prediction curves are shown for different, random selections of 200 points out of the full data set, giving an indication of the CNN prediction variability. (*g*)–(*i*) For noisy data, the growth model co-refinement needs a count rate of only ~1000 counts in the total reflection region to allow for adequate reconstruction, while the CNN can fit data even with ~300 counts in the total reflection region. Again, CNN predictions of a set of curves with different simulated photon shot noise are shown to indicate the variability of CNN predictions.

To demonstrate the accuracy of predictions with reduced data sets, we compare the CNN prediction and growth model co-refinement from only 200 $R(q, t)$ data points and the baseline individual curve fit from the full data set in Figs. 3(*d*)–3(*f*). We find that the growth model predictions/fits are close to the benchmark of individual curve fits of the full data set. Note that the CNN predictions are shown as a set of curves, each of which corresponds to a different set of randomly selected sets of 200 data points. This illustrates the fairly narrow spread of the predicted results and demonstrates that the result does not strongly depend on a specific set of $R(q, t)$ points. The growth model co-refinement is also able to fit reduced $R(q, t)$ XRR data sets down to 200 reflectivity points and is very similar to CNN performance.

In Figs. 4(*a*)–4(*c*) we show how the fit results degrade when systematically reducing the number of XRR data points entering the fit. This allows us to quantify the gains in speed and/or reduction in X-ray dose that are possible by reducing the XRR data-set size. We compare independent fits of all the individual curves using the batch fit mode of *refnx*, a growth model co-refinement of all curves at once and the CNN analysis of all curves. As three figures of merit, we choose the mean absolute error (MAE) for thickness $d(t)$, density SLD(t) and roughness $\sigma(t)$ in Figs. 4(*a*)–4(*c*) calculated for each of the above three analysis strategies in relation to our manual fit of the whole data set. A reasonable threshold for an acceptable MAE was set to the value indicated by the horizontal lines in Fig. 4, which corresponds to a doubling of the lowest MAE value for the full data set. We find that the growth model co-refinement and CNN predictions provide almost identical results as the error stays below our limit down to 200 data points, meaning that each XRR curve contains only two or three data points. The independent fit of each XRR curve by itself starts to fail earlier, at 1700–2200 points in the data set. This is understandable, because two to three data points do not constitute an XRR curve that can be fitted, and between 20 and 30 points are needed for fitting each of the 80 time steps of the growth. Also note that, while the growth model co-refinement and the CNN both perform well for sparsely sampled data sets, the batch fit takes 30 min to fit the data set and CNN can make a prediction in ∼55 ms. This faster prediction time from a CNN enables us to randomly select 100

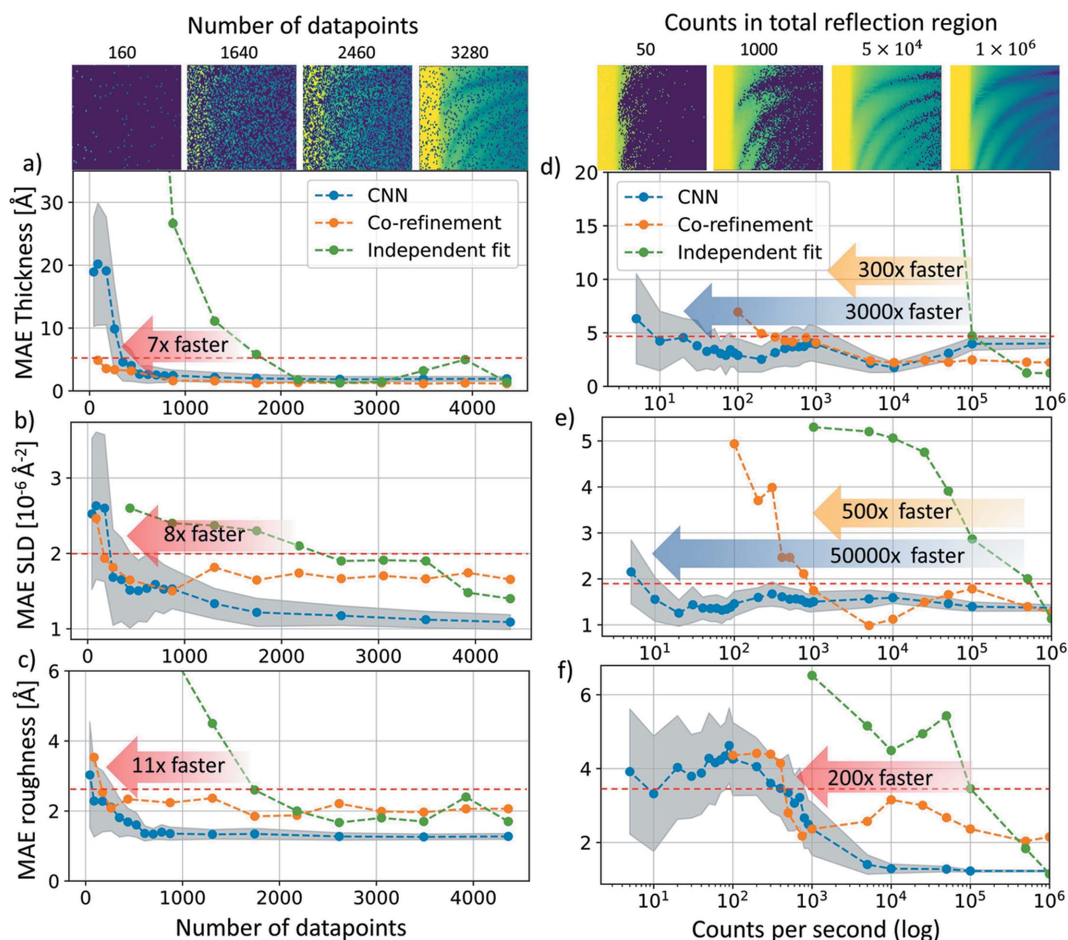


Figure 4

(*a*)–(*c*) The mean absolute error of the thickness, roughness and SLD increases when fewer data points are available for fitting. The CNN and the growth model co-refinement can analyze sparsely sampled data with a 6–11× lower number of measured reflectivity values compared with the individual fit. Panels (*d*)–(*f*) illustrate the effect of simulating increasing amounts of photon shot noise for the $R(q, t)$ data, which would occur for shorter integration times or weaker X-ray sources. The growth model co-refinement is accurate for noise levels more than two orders of magnitude higher than for the independent fit of individual XRR curves. The CNN can predict correct parameters with even higher noise levels than the co-refinement approach.

Table 1
Limits of data reduction and artificial noise for an acceptable performance of the individual curve fit, growth model co-refinement and CNN predictions.

	Sparse data set Minimum number of data points gain: x times			Noisy data Minimum count rate at total reflectivity edge gain: x times		
	Thickness	SLD	Roughness	Thickness	SLD	Roughness
Benchmark: individual fit	1700	1700	2200	1×10^5	1×10^5	5×10^5
Growth model co-refinement	200/8×	200/8×	200/11×	300/300×	500/200×	1000/500×
CNN	240/7×	200/8×	200/11×	30/3000×	500/200×	10/5 × 10^4×

different sets for each point in the curve and run the CNN prediction 100 times, so that we can give a confidence interval for the MAE in the CNN analysis.

3.2. Reduced counting time or lower X-ray flux

As a second strategy to speed up the measurement and reduce the X-ray dose, we also investigated the effects of reduced count rate per data point instead of reducing the number of data points. Through a shorter integration time, or alternatively lower incident X-ray flux, the photon shot noise increases. This can be modeled by replacing the count rate $R(q,t)$ at a given q point with a random value of counted photons x selected according to a Poissonian probability distribution:

$$\text{probability}\,(x) = \frac{[CR(q,t)]^x \exp[CR(q,t)]}{x!}. \qquad (5)$$

Here, C is the maximum number of counts in the total reflection region. Since the normalized XRR intensities $R(q,t)$ only range from 0 to 1, they must be scaled by the count rate C before calculating the photon shot noise. This assumes constant integration times and disregards the footprint correction needed for small sample sizes, but both these effects can be easily accounted for if needed in a particular experiment. A reflectivity value $R(q,t)$ is then replaced by $x(q,t)/C$. We introduced the Poisson-distributed noise into our training data and created a data set of 500 000 $R(q,t)$ in the range of C from 5 to 1×10^8. With the new noisy data set, we retrained our CNN for 20 epochs.

Next, we introduced the Poisson-distributed noise into the experimental data and generated $R(q,t)$ where the range of C was from 5 to 10^6 counts in the total reflection region. We used the retrained CNN to predict the thin film growth scenario from the noisy experimental $R(q,t)$, finding good CNN prediction performance even for extremely low X-ray flux. Next, we fitted noisy data with the growth model co-refinement of all curves. In Figs. 3(*g*)–3(*i*) we compare CNN prediction from noisy data with $C = 300$ and the growth model co-refinement with $C = 1000$ counts at the total reflectivity edge. Note that the 100 CNN prediction curves show each result from a different realization of shot noise to illustrate the variance in CNN predictions introduced by shot noise. The results are compared with our independent fit of the experimental $R(q,t)$ without added noise as reference. We get reasonable results for thickness, SLD and roughness, but note that the CNN prediction with only 300 counts overestimates the final roughness. This is not surprising, because the data at high q/low reflectivity will be significantly degraded because of the reduced count rate. For example, if the direct beam rate is 1000 Hz, and the background is 1 Hz, then one will not be able to go below $R = 0.001$.

In Figs. 4(*d*)–4(*f*) we quantify the performance of the CNN fit, growth model co-refinement and independent curve fits on data with different levels of noise. The point of reference is our manual fit of the full experimental data set without added noise. Again, the individual curve fit performs worst, and introduction of a physics-based growth model in either the co-refinement or the CNN fit significantly increases the robustness to noise and lowers the MAE. The growth model co-refinement can fit noisier data with a 200× to 500× reduction in counting time/X-ray flux over the individual curve fit of noisy data. The exact gain depends on the parameter, with roughness being the hardest to predict, offering only a 200× gain, and thickness being the easiest to predict and offering a 500× gain. Interestingly, the CNN outperforms both the individual curve fit and the growth model co-refinement, and can predict thickness from 3000× noisier data than the individual fit and SLD from up to 50 000× noisier data than the individual fit. In Table 1, we summarize the speed-up provided by the growth model implementation into the fitting algorithms. We used the same range in which the thin film parameters were allowed to vary for the growth model co-refinement and CNN. For the independent fit of each individual curve, we were not able to use exactly the same parameters but we used comparable thickness, roughness and SLD parameter ranges. Performance gains through using re-parametrization and co-refinement with the CNN are shown to be greater than sevenfold for data reduction and more than 200-fold for noisy data sets from shorter integration times or lower flux.

4. Discussion

The approach presented here enables a more than two orders of magnitude reduction in X-ray/neutron photon dose and faster experiments due to the implementation of a physics-based growth model and analysis of all real-time XRR data sets at once. It has already been shown that the prediction of CNN for a single noisy XRR curve is possible (Greco *et al.*, 2021) and an analysis of intentionally high-noise NR was published recently (Aoki *et al.*, 2021; Doucet *et al.*, 2021). All these studies show that an NN can perform well on noisy XRR or NR data. However, the re-parametrization and co-refinement of multiple XRR curves used here enable us to

achieve correct prediction from an even lower quantity of data than reported before. This performance increase can be obtained either by the implementation of the growth model into a co-refinement approach or by using a CNN. The growth model co-refinement is better for single experiments because it does not need any training data or CNN training time. However, the analysis of the data set is significantly slower at ~30 min compared with 55 ms with a CNN. The growth model co-refinement could be sped up by implementing a simpler growth model, but our example of a rate equation/differential equation based temporal model is not untypical for processes measured in real time. Moreover, we expect that some further speed-ups may be achievable by improving the ordinary differential equation solving step in Python. For the CNN case, implementing a growth model for training data generation and training of the CNN take 5.5 h, which is slower than a single growth model co-refinement.

The CNN analysis has advantages if a growth study encompasses repeated measurement under varying conditions and necessitates several fits. Also, the CNN predictions lie closer to results from a manual fit in Fig. 3 than growth model co-refinement. Due to its prediction speed, the CNN could be used as a process analytical technique that enables online analysis or even feedback during processing. Lastly, while the performance of growth model co-refinement and the CNN is similar on sparsely sampled data, the CNN has advantages when analyzing noisy data. The choice of analysis strategy therefore depends strongly on the experimental setup and the number of experiments to be analyzed. Using a weak source with low flux or short integration times results in noisy data and would favor the use of a CNN approach. In applications where the motor movement times of a diffractometer are limiting the time resolution, our sparse sampling approach with fewer data points and fewer motor movements is a viable alternative for speeding up measurements. Both growth model co-refinement and the CNN approach perform similarly with respect to the possible reduction in data points needed and fit errors in this case. Note that both sparse sampling approaches presented here can also deal not just with data collected at random q points but also with data collected at one or several fixed q points as a function of time. Examples of such growth studies include measurements of the anti-Bragg oscillations at $q = 1/2q_{\rm Bragg}$, which both our CNN and growth model co-refinement can analyze. However, a single q point measured at 80 time steps is below our lower limit of 200 data points for a reasonable accuracy (see Table 1), so that it is advisable to measure also at other q points beyond the anti-Bragg point, *e.g.* at $q = 1/3q_{\rm Bragg}$ and $1/4q_{\rm Bragg}$ (Kowarik *et al.*, 2009).

The CNN and *refnx* co-refinement approach allow one to implement models not just for growth processes but also for other time-dependent XRR problems. Kinetic models can, for example, be included as prior information in the CNN via a simulation of the corresponding training data with the kinetic model, the parameters of which become training labels. Alternatively, the kinetic model can be directly incorporated into a *refnx* co-refinement. Note that a CNN can learn about physically sensible solutions not just via the choice of training data, as presented here, but also via a careful choice of a physics-informed loss function, as has recently been demonstrated (Raissi *et al.*, 2019). Through their flexibility, the CNN and the growth model co-refinement lend themselves to numerous applications in fields that profit from *in situ* and real-time measurements, such as electrochemical changes at interfaces, but the co-refinement approach is also useful for data acquired *e.g.* with different contrasts in NR.

5. Conclusion

In this article we have shown the possibilities of the NN and growth model fit approaches for fitting *in situ* XRR measurements using co-refinement of multiple XRR curves. Prior knowledge is implemented in the analysis in the form of a kinetic growth model. This model re-parametrizes the structural evolution with far fewer parameters than independent structural film properties at each time step. The CNN and batch fit directly predict growth model parameters without the need for independent single XRR curve fitting. The co-refinement of many XRR curves enables predictions for sparsely sampled data sets that have insufficient information for single curve fits or single curve NN predictions. We tested our CNN and growth model co-refinement approaches on experimentally measured data where we reduced the number of reflectivity points from 4160 measured reflectivity data points to 200 randomly distributed data points that were sufficient to make correct thin film parameter predictions. This significantly undercuts the minimum number of 1700 data points needed for individual curve fitting. By combining fast data analysis and accurate prediction from a sparse data set, we can go beyond the state of the art in XRR measurements and effectively speed up XRR measurements by almost an order of magnitude. This allows one to analyze sensitive samples without beam damage or increase the throughput in experiments.

Moreover, we tested our NN and growth model co-refinement on data with high photon shot noise, as occurs for short data acquisition times or weak X-ray sources. We showed that our CNN can predict correct structural film properties even from extremely noisy data. In comparison with a conventional independent XRR curve fit, the exposure of the sample can be three orders of magnitude lower and the CNN prediction still corresponds to a real thin film growth scenario. This result will allow significantly faster measurements to be made at synchrotron facilities with high flux or enable the use of lower-brightness laboratory sources for real-time measurements.

The supporting information associated with this article is also available in a GitHub repository: https://github.com/kowarik-lab/XRR-model-based-co-refinement-CNN-refnx.

Acknowledgements

We thank A. Greco, V. Starostin, A. Gerlach, A. Hinderhofer and F. Schreiber for helpful discussions.

Funding information

DM acknowledges the University of Graz for financial support.

References

Abelès, F. (1950). *J. Phys. Radium*, **11**, 307–309.
Andrejevic, N., Chen, Z., Nguyen, T., Fan, L., Heiberger, H., Zhou, L. J., Zhao, Y. F., Chang, C. Z., Grutter, A. & Li, M. (2022). *Appl. Phys. Rev.* **9**, 011421.
Aoki, H., Liu, Y. & Yamashita, T. (2021). *Sci. Rep.* **11**, 22711.
Bishop, C. (1994). *Rev. Sci. Instrum.* **65**, 1803–1832.
Björck, M. & Andersson, G. (2007). *J. Appl. Cryst.* **40**, 1174–1178.
Bommel, S., Kleppmann, N., Weber, C., Spranger, H., Schäfer, P., Novak, J., Roth, S. V., Schreiber, F., Klapp, S. H. L. & Kowarik, S. (2014). *Nat. Commun.* **5**, 5388–5395.
Braslau, A., Pershan, P. S., Swislow, G., Ocko, B. M. & Als-Nielsen, J. (1988). *Phys. Rev. A*, **38**, 2457–2470.
Braun, W., Jenichen, B., Kaganer, V. M., Shtukenberg, A. G., Däweritz, L. & Ploog, K. H. (2003). *Surf. Sci.* **525**, 126–136.
Campbell, R. A., Saaka, Y., Shao, Y., Gerelli, Y., Cubitt, R., Nazaruk, E., Matyszewska, D. & Lawrence, M. J. (2018). *J. Colloid Interface Sci.* **531**, 98–108.
Cao, C., Steinrück, H. G., Shyam, B., Stone, K. H. & Toney, M. F. (2016). *Nano Lett.* **16**, 7394–7401.
Chen, Z., Andrejevic, N., Drucker, N. C., Nguyen, T., Xian, R. P., Smidt, T., Wang, Y., Ernstorfer, R., Tennant, D. A., Chan, M. & Li, M. (2021). *Chem. Phys. Rev.* **2**, 031301.
Cohen, P. I., Petrich, G. S., Pukite, P. R., Whaley, G. J. & Arrott, A. S. (1989). *Surf. Sci.* **216**, 222–248.
Danauskas, S. M., Li, D., Meron, M., Lin, B. & Lee, K. Y. C. (2008). *J. Appl. Cryst.* **41**, 1187–1193.
Doucet, M., Archibald, R. K. & Heller, W. T. (2021). *Mach. Learn. Sci. Technol.* **2**, 035001.
Frank, C., Novák, J., Banerjee, R., Gerlach, A., Schreiber, F., Vorobiev, A. & Kowarik, S. (2014). *Phys. Rev. B*, **90**, 045410.
Goedel, W. A., Luap, C., Oeser, R., Lang, P., Braun, C. & Steitz, R. (1999). *Macromolecules*, **32**, 7599–7609.
Greco, A., Starostin, V., Hinderhofer, A., Gerlach, A., Skoda, M. W. A., Kowarik, S. & Schreiber, F. (2021). *Mach. Learn. Sci. Technol.* **2**, 045003.
Greco, A., Starostin, V., Karapanagiotis, C., Hinderhofer, A., Gerlach, A., Pithan, L., Liehr, S., Schreiber, F. & Kowarik, S. (2019). *J. Appl. Cryst.* **52**, 1342–1347.
Hanke, M., Kaganer, V. M., Bierwagen, O., Niehle, M. & Trampert, A. (2012). *Nanoscale Res. Lett.* **7**, 03.
Heavens, O. S. (1960). *Rep. Prog. Phys.* **23**, 1–65.
Holý, V., Ullrich, P. & Baumbach, T. (1999). *High-Resolution X-ray Scattering from Thin Films and Multilayers*. Berlin, Heidelberg: Springer.
Joress, H., Brock, J. D. & Woll, A. R. (2018). *J. Synchrotron Rad.* **25**, 706–716.
Ketkar, N. (2017). *Deep Learning with Python*, pp. 97–111. Berkeley: Apress.
Kienzle, P. A., Krycka, J., Patel, N. & Sahin, I. (2022). *Refl1D*. Version 0.8.15. University of Maryland, College Park, MD, USA. https://refl1d.readthedocs.io/en/latest/getting_started/license.html.
Kim, K. T., Kim, D. O., Kee, J. Y., Seo, I., Choi, Y., Choi, J. W. & Lee, D. R. (2021). *Curr. Appl. Phys.* **30**, 46–52.
Kim, K. T. & Lee, D. R. (2021). *J. Appl. Cryst.* **54**, 1572–1579.
Kingma, D. P. & Ba, J. L. (2015). *arXiv*:1412.6980 [cs.LG].
Kowarik, S., Gerlach, A., Sellner, S., Schreiber, F., Cavalcanti, L. & Konovalov, O. (2006). *Phys. Rev. Lett.* **96**, 125504.
Kowarik, S., Gerlach, A., Skoda, M. W. A., Sellner, S. & Schreiber, F. (2009). *Eur. Phys. J. Spec. Top.* **167**, 11–18.
Krause, B., Abadias, G., Furgeaud, C., Michel, A., Resta, A., Coati, A., Garreau, Y., Vlad, A., Hauschild, D. & Baumbach, T. (2019). *Appl. Mater. Interfaces*, **11**, 39315–39323.
Lazzari, R. (2002). *J. Appl. Cryst.* **35**, 406–421.
Liu, C., Neale, Z. G. & Cao, G. (2016). *Mater. Today*, **19**, 109–123.
Liu, S., Melton, C. N., Venkatakrishnan, S., Pandolfi, R. J., Freychet, G., Kumar, D., Tang, H., Hexemer, A. & Ushizima, D. M. (2019). *MRS Commun.* **9**, 586–592.
Loaiza, J. M. C. & Raza, Z. (2021). *Mach. Learn. Sci. Technol.* **2**, 025034.
López García, J. & Rivero, Á. (2021). *Advances on Links Between Mathematics and Industry*, SxI – Springer for Innovation, Vol. 15, pp. 131–152.. Cham: Springer.
McCluskey, A. R., Sanchez-Fernandez, A., Edler, K. J., Parker, S. C., Jackson, A. J., Campbell, R. A. & Arnold, T. (2019). *Phys. Chem. Chem. Phys.* **21**, 6133–6141.
Mironov, D., Durant, J. H., MacKenzie, R. & Cooper, J. F. K. (2021). *Mach. Learn. Sci. Technol.* **2**, 035006.
Nelson, A. (2006). *J. Appl. Cryst.* **39**, 273–276.
Nelson, A. R. J. & Prescott, S. W. (2019). *J. Appl. Cryst.* **52**, 193–200.
Park, W. B., Chung, J., Jung, J., Sohn, K., Singh, S. P., Pyo, M., Shin, N. & Sohn, K.-S. (2017). *IUCrJ*, **4**, 486–494.
Parratt, L. G. (1954). *Phys. Rev.* **95**, 359–369.
Pietsch, U., Holý, V. & Baumbach, T. (2004). *High-Resolution X-ray Scattering*. New York: Springer.
Raissi, M., Perdikaris, P. & Karniadakis, G. E. (2019). *J. Comput. Phys.* **378**, 686–707.
Russell, T. P. (1990). *Mater. Sci. Rep.* **5**, 171–271.
Simonyan, K. & Zisserman, A. (2015). *arXiv*:1409.1556 [cs.CV].
Skoda, M. W. A., Thomas, B., Hagreen, M., Sebastiani, F. & Pfrang, C. (2017). *RSC Adv.* **7**, 34208–34214.
Tolan, M. (1999). *X-ray Scattering from Soft-Matter Thin Films*. Berlin, Heidelberg: Springer-Verlag.
Trofimov, V. I., Mokerov, V. G. & Shumyankov, A. G. (1997). *Thin Solid Films*, **306**, 105–111.
Vecsei, P. M., Choo, K., Chang, J. & Neupert, T. (2019). *Phys. Rev. B*, **99**, 245120.
Woll, A. R., Desai, T. V. & Engstrom, J. R. (2011). *Phys. Rev. B*, **84**, 075479.