nmfMapping: a cloud-based web application for non-negative matrix factorization of powder diffraction and pair distribution function datasets

Zachary Thatcher,[a]* Chia-Hao Liu,[a] Long Yang,[a] Brennan C. McBride,[b] Gia Thinh Tran,[b] Allison Wustrow,[b] Martin A. Karlsen,[c] James R. Neilson,[b] Dorthe B. Ravnsbæk[c] and Simon J. L. Billinge[a,d]*

[a]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027, USA, [b]Department of Chemistry, Colorado State University, Fort Collins, CO 80523-1872, USA, [c]Department of Chemistry, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark, and [d]Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY 11973, USA. *Correspondence e-mail: zachary.alan.thatcher@gmail.com, sb2896@columbia.edu

A cloud-hosted web-based software application, nmfMapping, for carrying out a non-negative matrix factorization of a set of powder diffraction or atomic pair distribution function datasets is described. This application allows structure scientists to find trends rapidly in sets of related data such as from in situ and operando diffraction experiments. The application is easy to use and does not require any programming expertise. It is available at https://pdfitc.org/.

1. Introduction

Modern hardware and software at X-ray synchrotrons allow scientists to study materials in ground-breaking detail (Maier et al., 2007; Chupas et al., 2007), allowing high-throughput combinatorial, in situ and operando experiments that produce unprecedented amounts of data on previously inaccessible systems.

To extract information the data need to be reduced to a scientifically meaningful function, such as a 1D powder diffraction pattern, and then interpreted, for example using structural fitting such as Rietveld analysis (Rietveld, 1969) or atomic pair distribution function (PDF) analysis (Egami & Billinge, 2012; Farrow et al., 2007; Tucker et al., 2007). This can be time consuming and labour intensive. It often involves fitting the data from a single diffraction pattern at a time, as opposed to fitting cross-pattern trends, so called parametric refinement (Stinton & Evans, 2007). It can often be difficult to extract trends in the data if they are not expected a priori, leading to missed insights. For example, the novel nature of modern experiments can often result in one or more unexpected structures. Whilst tools are emerging that can automatically identify constituent structural signals (Yang et al., 2020), they are limited to cases where the structure is in the databases they are using. In cases where novel materials are being investigated, it is highly valuable to use structure-agnostic ways of interrogating large sets of data to extract meaningful signals.

Non-negative matrix factorization (NMF) (Paatero & Tapper, 1994) has been a particularly useful, reliable and intuitive unsupervised machine learning (ML) approach for analytically reducing large datasets of physical signals to reveal trends (Brunet et al., 2004; Pauca et al., 2006). More specifically, it can yield isolated structural signals and their
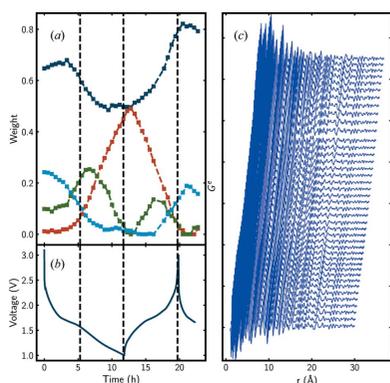
relative presence for both PDF (Liu *et al.*, 2021; Geddes *et al.*, 2019) and powder diffraction (Long *et al.*, 2009) data.

Although NMF modules are present in, for example, *scikit learn* (Pedregosa *et al.*, 2011), there are impediments to their widespread adoption. They are not easy to use for non-programming/non-mathematician experts. Also, even when a user is familiar with Python scripting, it may not be possible or straightforward to install such libraries at any given synchrotron facility during the allocated beamtime when such a rapid analysis of trends in data is particularly valuable. To address these issues we have developed a cloud-hosted web-based application (app), described herein, for carrying out NMF decomposition of powder diffraction and PDF datasets. This app, called *nmfMapping*, has been implemented as a service on the *PDFitc* cloud platform at https://pdfitc.org/ (Yang *et al.*, 2021).

2. NMF

NMF has been described in depth previously in both a mathematically general sense (Lee & Seung, 1999; Berry *et al.*, 2007) and in regard to diffraction experiments (Liu *et al.*, 2021; Long *et al.*, 2009; Geddes *et al.*, 2019). We reproduce here a brief discussion on the use of NMF on diffraction data, for the sake of context.

As a matrix decomposition algorithm, NMF is a method to find and track a reduced set of basis components that can approximate the full matrix to a desired precision. In this respect, NMF is similar to principal component analysis (PCA) (Lee & Seung, 1999), but with constraints applied to the decomposition that the components and the weights must be positive. Since many experimental signals (such as intensity) are positive, and the compound signals are sums of such signals (*e.g.* the diffraction signals from each chemical phase in a multi-phase sample), the enforced positivity constraints mean that NMF decomposition is more likely than PCA to find solutions that are physically meaningful. A drawback to NMF is that the optimization is non-convex and finding the global minimum is not assured. However, in practice proper initialization seems to make NMF results consistent for diffraction and PDF data (Liu *et al.*, 2021; Long *et al.*, 2009) where its application looks very promising (Todd *et al.*, 2020). We note that it is also possible to include actual PDFs in a constrained NMF decomposition (Maffettone *et al.*, 2021) to push the results further in the direction of physical components. This has not been implemented in the current version of *nmfMapping* as it complicates the workflow for the user whilst introducing bias, and we have found that in most cases the simple and less biased NMF decomposition is already quite successful at returning physically reasonable components. As studies of NMF applied to structural studies develop we may revisit this in the future.

To use NMF to find trends in sets of experimentally measured 1D signals we first create a matrix where each column is a signal and each signal is measured on an identical grid of points, *e.g.* 2θ for powder data or r for PDF data, that form the rows. The decomposition can then be done using standard methods (Pedregosa *et al.*, 2011). Our *nmfMapping@PDFitc* implementation uses standard methods but in a user-friendly web environment, with minimal preprocessing suited to PDF and powder diffraction data.

3. *PDFitc* interface

PDFitc (https://pdfitc.org/) (Yang *et al.*, 2021) is a cloud-based platform containing tools that can, given a user-uploaded measured PDF, automatically find candidate structures for subsequent fitting (the *structureMining* app; Yang *et al.*, 2020), predict the space group of the structure that produced the PDF (the *spacegroupMining* app; Liu *et al.*, 2019) and find the similarity between PDFs in a set (*similarityMapping* app; Yang *et al.*, 2021). It has an easy to navigate interface that can be utilized on all of the popular web browsers (*e.g.* Chrome, Firefox, Safari *etc.*). Users can log in to the website using Google, GitHub and ORCID credentials to upload data for processing. The website can even be utilized on a computer without data, since logging in connects users to their previous results, as well as to results other users may have explicitly shared with them (by default all data can only be seen by the user who uploaded them). The goal of the website is to provide PDF analysis tools without the need to install and run software on the user's local computer. This paper describes the addition of a new app on the *PDFitc* platform for carrying out NMF decomposition of a set of PDF or powder diffraction spectra collected as a related set, such as time, temperature or spatially resolved series of spectra.

4. The *nmfMapping* app interface

Here we describe the new NMF decomposition functionality in the *nmfMapping* app. We illustrate this by going through the process of accessing the website and performing an NMF decomposition on a synthetic (calculated) PDF dataset from Liu *et al.* (2021), and then rerunning the analysis on a smart phone.

On going to https://pdfitc.org the user will see the home page of the *PDFitc* website (Fig. 1).

Currently, including *nmfMapping*, *PDFitc* presents four application tiles (*structureMining*, *spacegroupMining*, *similarityMapping* and *nmfMapping*) to the user. Upon selecting the *nmfMapping* tile, the user will be redirected to the login page. Logging in is required to use the services. It also allows the user to revisit previous analyses, as well as any analyses that have been explicitly shared by another user, through their account icon in the top right-hand corner of the screen, as will be shown below. Note that a different user profile is associated with logins from each authentication service (ORCID, GitHub or Google) so it is important to pick one service and use it consistently. After logging in, the user is directed to the *nmfMapping* landing page (Fig. 2).

To use the app the user uploads a list of either powder diffraction or PDF data, inputs a small number of configuration parameters and clicks 'Submit'. The app will do the
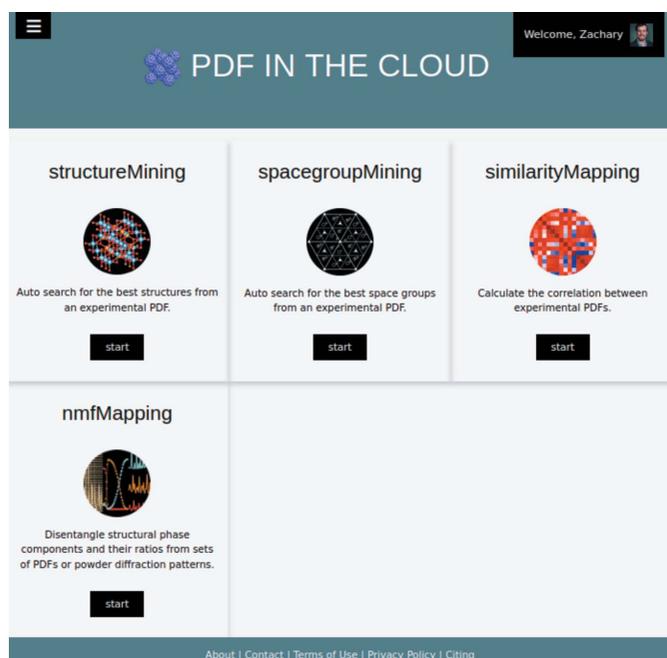
Figure 1
A screenshot of the https://pdfitc.org homepage. Each tile may be clicked to access the specified app. If not logged in, the user will be directed to a login page before continuing to the app.

Figure 2
A screenshot of the *nmfMapping@PDFitc* landing page. Users upload their datasets via the Browse button and can optionally specify parameters such as the number of NMF components or data ranges. After specifying whether the uploaded data are experimental or simulated, the computation is started by clicking Submit.

decomposition and return the results, as described below. There is no well defined minimum number of PDFs in the set for NMF to work well. However, we have found that at least four times the number of components can give satisfactory results. There is also no upper limit, though the larger the set size the slower the decomposition will be.

The app can currently accept `.gr` format files that are output by standard PDF analysis programs such as *PDFgetX2* (Qiu *et al.*, 2004) and *PDFgetX3* (Juhás *et al.*, 2013), `.chi` files that are output by integration programs such as *pyFAI* (Ashiotis *et al.*, 2015) or *Fit2D* (Hammersley, 2016) and general multi-column text files such as `xy` or `xye` format files. These files may also contain header information above the columns of data. The negative PDF values are handled automatically by the program with no pre-processing required of the user.

In terms of the configuration parameters, the most important is the 'Number of components'. Setting this to the wrong number leads to poor or non-physical results. This value may be left blank at first and *nmfMapping* will attempt to select the correct number of signals automatically. The automatic approach considers the amount by which the frob norm loss is lowered by the addition of an extra contribution. If the loss matrix is defined as $\mathbf{L} = \mathbf{V} - \mathbf{WH}$ where $\mathbf{V}$ is the original matrix and $\mathbf{WH}$ is the approximation to it from the NMF reconstruction, the frob norm loss is then the Frobenius norm of the loss matrix, F_n,

$$F_n = \|\mathbf{L}\|_F, \qquad (1)$$
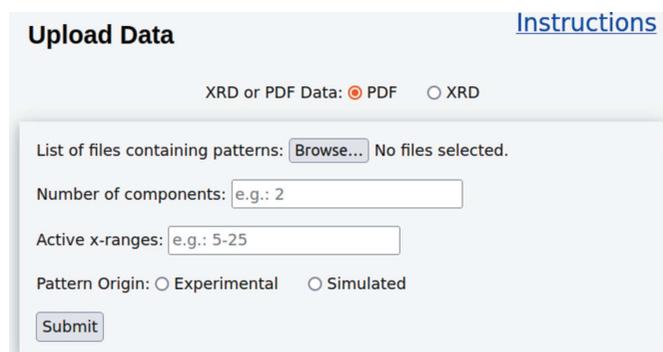
where the Frobenius norm is defined in the usual way,

$$F_n = \left(\sum_{i=1}^{m} \sum_{j=1}^{k} |\ell_{ij}|^2 \right)^{1/2}, \qquad (2)$$

with ℓ_{ij} being the (i,j)th element of the loss matrix. The frob norm loss is our measure of the accuracy of the reconstruction for n components. As a starting point for choosing the number of components for any given situation, the algorithm computes the Frob norm loss for different n, and when the improvement in the accuracy for the addition of a new component, $(F_{n+1} - F_n)/F_n$, is below a minimum threshold (currently 0.1%), we choose n as the number of components. If this threshold is not met, as can happen for real data with noise, an upper limit of ten components is arbitrarily imposed to get a first look at the reconstruction error graph which is returned to the user for manual consideration (elaborated on below). The decomposition can then be rerun with the user-selected number of components after making the appropriate input to the 'Number of components' box.

The number of components is an important parameter to control for good scientific results. The best number roughly corresponds to the number of PDFs/powder patterns it takes to explain all the signals in the data. For example, a sputtered film that contains overlapping patches of three metals would probably need three components, being the three chemical components. On the other hand a balanced chemical reaction where chemicals A and B transform to products C and D will require two components, even though there are four chemical components, namely the 'reagent signal' and the 'product signal'. If the composition of the sample does not respect the balanced synthesis equation, for example if not all of the A chemical is consumed in the reaction, the NMF will probably require three components, namely the PDF/powder signals from the reactants, products and A. Tuning the threshold on real data is discussed further in the *Examples* section. There is often an 'elbow' in the reconstruction error versus component number graph, with it showing a distinct flattening. The point of flattening is often a good first guess for the correct number of components (Liu *et al.*, 2021; Maffet-
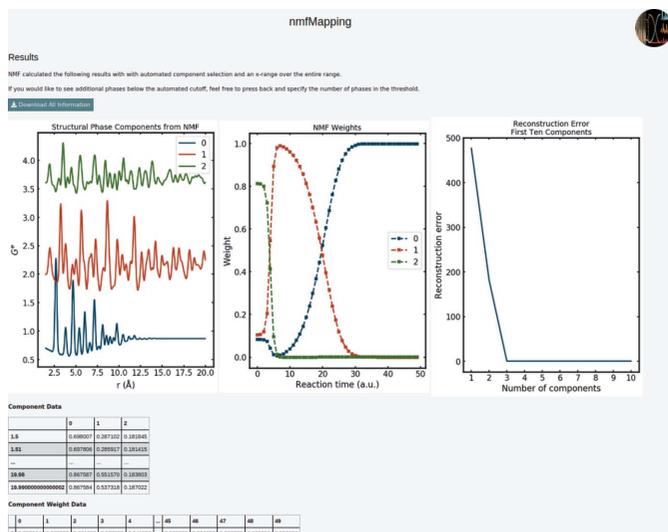
Figure 3
A screenshot of the NMF results page after completing a decomposition. The screenshot is from the decomposition described in detail by Liu *et al.* (2021). The first panel shows the NMF components (in this case, three). The center panel shows the weights that were used to recreate each uploaded PDF in the set from these three components. The right-hand panel shows the reconstruction error plotted versus the number of components and is useful for making decisions on how many components to try. For these simulated and noiseless data all the error is eliminated after three components. The tables give a snapshot summary of the results which may be downloaded in full to the user's local computer by clicking the Download button.

tone *et al.*, 2021; Long *et al.*, 2009) and we have found this also to be the case here.

After completing the decomposition, *nmfMapping@PDFitc* returns three graphs and related tables (Fig. 3). The information in the panels is described in the caption to Fig. 3. They show plots of the NMF extracted components and the weights needed to reconstruct each of the uploaded datasets. The data in the set are ordered alphanumerically by filename, so if the order is scientifically meaningful the files need to be named appropriately. In the third panel the 'Reconstruction Error' may be used to determine the correct number of components to use in the NMF decomposition. For example, in the current case, which is simulated data described by Liu *et al.* (2021), the graph shows a steep decrease in error reduction up to three components, where three is the best number of components in this case. A number of data points from the graph are also provided in tables on the web page beneath the graphs, allowing close inspection of the data when iterating.

All the graphs and tables can be downloaded by the user in a compressed file, along with files containing the individual constituent components. The output component files are in a format where they can be further processed by directly loading to most X-ray diffraction (XRD) or PDF fitting software, including the *structureMining* (Yang *et al.*, 2020) package also located on *PDFitc* and the widely used *PDFgui* (Farrow *et al.*, 2007) structure refinement software.

Fig. 4 emphasizes the accessibility of the *nmfMapping@PDFitc* app. After logging in and performing the analysis above on a PC where the data were located, a smart
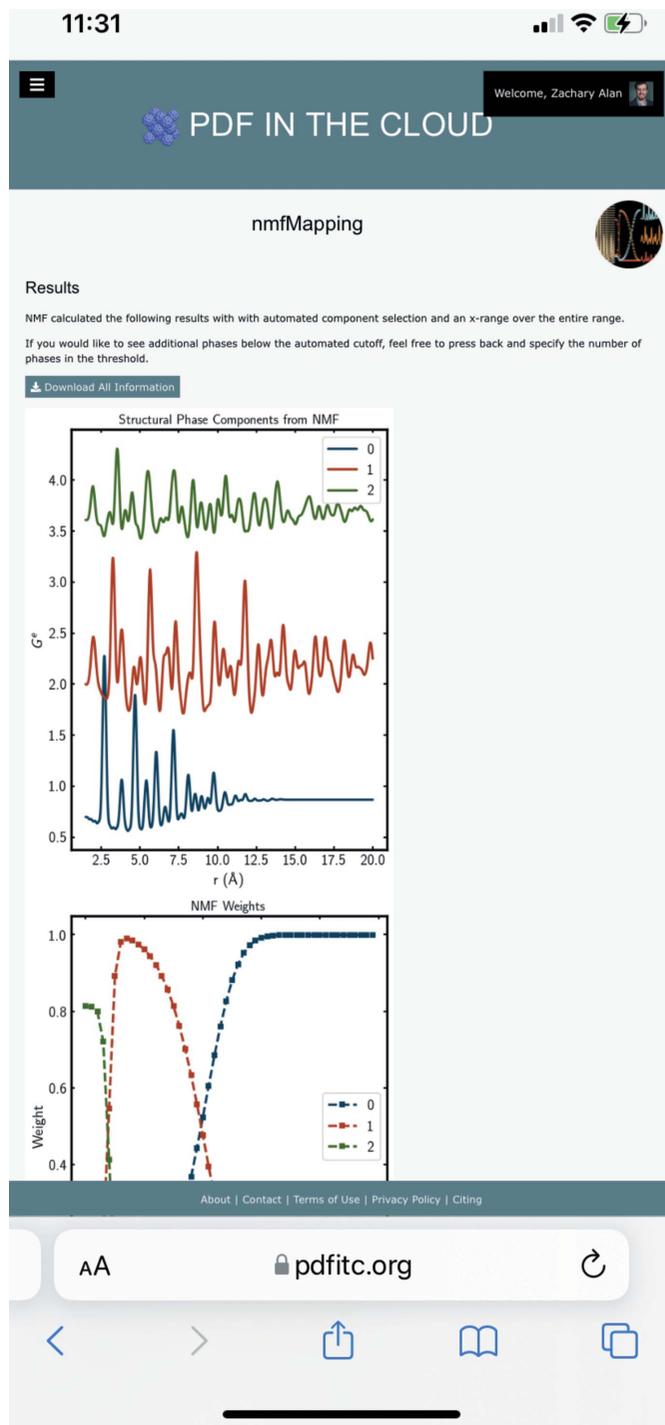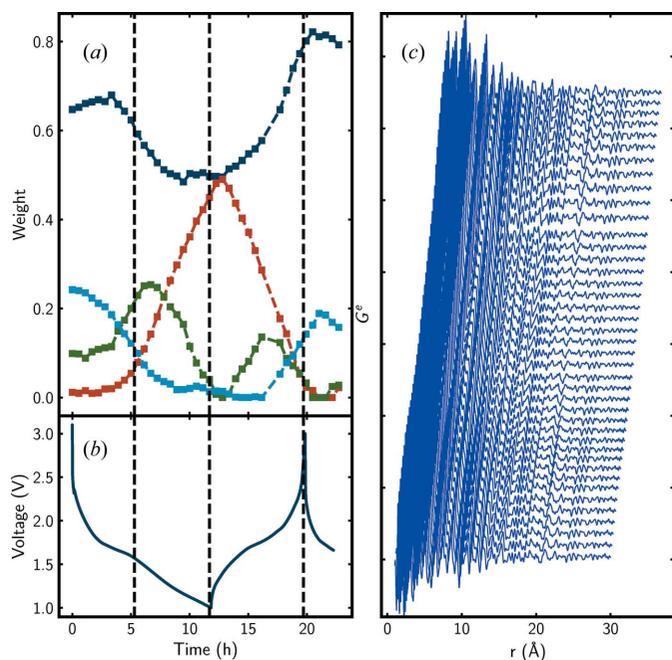

Figure 4
The NMF results page after rerunning the analysis from the Safari browser on a smart phone.

phone was utilized to log in to the same account. By going to the 'Activities' web page on the phone and clicking the Play button, the analysis results were recreated without requiring that the data be on the smart phone.

5. Examples

In this section we present some real-world examples of *in situ* experimental data run through the *PDFitc* app. In the first
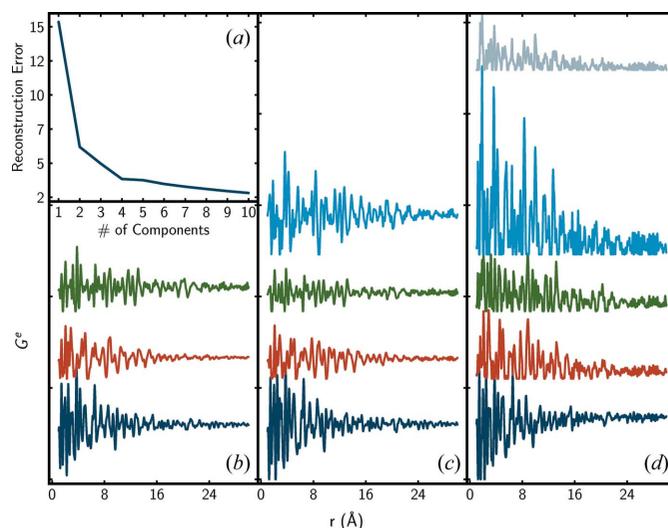
Figure 5
(*a*) Weights of the NMF components versus time obtained from *nmfMapping@PDFitc* and (*b*) the galvanostatic charge–discharge curve at the same time points. Vertical dashed lines are drawn at significant time points, *i.e.* first a change in slope of the galvanostatic curve during discharge, then the point where the cell was switched from discharge to charge, and thirdly the point where it was switched back again to a discharge cycle. The NMF analysis reveals that an intermediate phase (shown in green) begins to transform into an Li-rich phase, represented by the red component, precisely where the voltage starts changing slope during discharge. (*c*) A waterfall plot of the PDF dataset reconstructed with NMF weights from panel (*a*) and components shown in Fig. 6(*c*).

example we consider structural changes in an *operando* battery experiment, and in the second we discuss analysis of a set of diffraction patterns collected while tracking a chemical synthesis reaction.

5.1. *Operando* PDF experiment for a rechargeable Li-ion battery

In this example we consider data from an experiment characterizing nanostructured and disordered cathode battery structures via PDF analysis of *operando* experiments. Such experiments benefit greatly from use of an NMF-type analysis because of the high complexity of the battery electrodes, with a large number of unknown chemical components under dynamic conditions, and the large numbers of datasets (Christensen & Ravnsbæk, 2021).

In this particular example, *nmfMapping@PDFitc* was used to gain a quick preliminary insight into the structural shifts present in a nanocrystalline TiO_2–bronze cathode prior to a more detailed PDF fitting analysis. The X-ray data presented here were collected in transmission on a Perkin–Elmer area detector (XRD1621) with a wavelength of 0.20736 Å on beamline P02.1 at PETRA III, DESY, Hamburg, Germany. The battery was assembled in an AMPIX electrochemical cell configuration (Borkiewicz *et al.*, 2012). The battery stack consisted of a metallic lithium anode, a glass fiber separator

Figure 6
(*a*) Reconstruction error versus number of NMF components. (*b*)–(*d*) The resulting NMF components when the number of components specified in the decomposition is (*b*) three, (*c*) four and (*d*) five in *nmfMapping@PDFitc*.

wetted with electrolyte, and a cathode pellet made from active material (TiO_2), poly(vinylidene fluoride) (PVDF) binder and conductive carbon. The diffraction data were processed with *DAWN 2* (Filik *et al.*, 2017) and *PDFgetX3* (Juhás *et al.*, 2013) to get the PDFs. More information regarding the capture and analysis of the battery cathode data, and a scientific discussion of the results, will be the subject of a future publication.

The progress of the battery discharge is tracked using galvanostatic voltage measurements [Fig. 5(*b*)]. The voltage over time graph reveals two regimes of different slope for the initial discharge. The goal of the PDF measurements is to uncover any structural origin of this galvanostatic behavior. The NMF phase tracking is also shown in Fig. 5(*a*), lined up with the galvanostatic response for comparison. The NMF clearly reveals that this voltage behavior is at least partially due to structural variations. These variations are difficult to discern from the PDF but found automatically by the NMF analysis.

For illustrative purposes the dataset reconstruction, strictly utilizing the NMF components from Fig. 6(*c*) and weights from Fig. 5(*a*), is shown in Fig. 5(*c*).

In order to give a quick check on the physical reasonableness of the *nmfMapping@PDFitc* output, the analysis can be rerun repeatedly in *nmfMapping@PDFitc* with different numbers of components specified. The results of repeated NMF runs (which can typically be done within a matter of minutes, as was the case here) show that attempting to track five or more signals results in NMF components that do not resemble measured PDFs, in this case with signals that clip at high and low values [Fig. 6(*d*)].

In this case four components appear to give the best results [Fig. 6(*c*)], on the basis that this is the largest number of components that appear to be undistorted PDF-like signals. In addition, Liu *et al.* (2021) and Long *et al.* (2009) have shown that limiting the number of components to where the

reconstruction error begins to flatten is a good heuristic for deciding where additional components lead to non-physical results, and in Fig. 6(a) we see that this occurs with four components. Satisfyingly, this result is also supported by chemical intuition. Two components are expected for the charged (Li-poor) state (dark blue and light blue). At least one component (red) is expected for the discharged (Li-rich) state. What pops out unexpectedly from this analysis is an intermediate, partially lithiated, phase (green). In general, it is important to assess the chemical validity of any NMF results, but this illustrates the utility of this approach.

5.2. Inorganic chemical synthesis reaction

In this experiment a mixture of $LiMnO_2$ and LaOCl powders was reacted at a range of temperatures in a gradient furnace (O'Nolan *et al.*, 2020) (see below for additional details) for 40 min, then cooled to room temperature. Diffraction patterns corresponding to different reaction temperatures were then measured in transmission mode on the 28-ID-2 beamline at the National Synchrotron Light Source II, Brookhaven, New York, USA. The data were integrated using *GSAS-II* (Toby & Von Dreele, 2013) before being passed to the *nmfMapping@PDFitc* app. For more details of the experimental setup, see Wustrow *et al.* (2021).

The dataset contains 25 diffraction patterns that come from a continuum of temperatures from 300°C up to 750°C. The gradient furnace (O'Nolan *et al.*, 2020) produces a more or less linear dependence of temperature versus position along the rod-shaped sample. The ability to measure reaction products from reactions at different temperatures as a function of position along the rod allows rapid acquisition of many temperatures as each data point takes just seconds to collect. Though in this test case there are just a manageable 25 XRD patterns, it is an example of an experimental configuration with the potential to produce hundreds or even thousands of patterns for a single experiment. This is a good illustration of how experimental innovation can drive the production of larger datasets, which require more sophisticated processing such as NMF analysis.

One of the main goals of these experiments is to obtain the temperature dependence of the compositional products, which provides insight into different reaction pathways available to a system (Martinolich & Neilson, 2017). Traditionally this would have been done by carrying out a sequence of Rietveld refinements (Rietveld, 1969) on the datasets, which is a time-consuming and difficult task, not least because the number and type of chemical phases at each temperature are not known *a priori*. Fig. 7(a) shows the results obtained by *nmfMapping@PDFitc* and these can be compared with those obtained by carrying out a multi-phase Rietveld refinement on each of the 25 datasets [Fig. 7(b)].

We found we got the best results when we ran *nmfMapping@PDFitc* with four components. However, the chemistry suggested that there may be six chemical species. By carrying out Rietveld refinements on the NMF components we could attribute mixtures of multiple chemical species to each
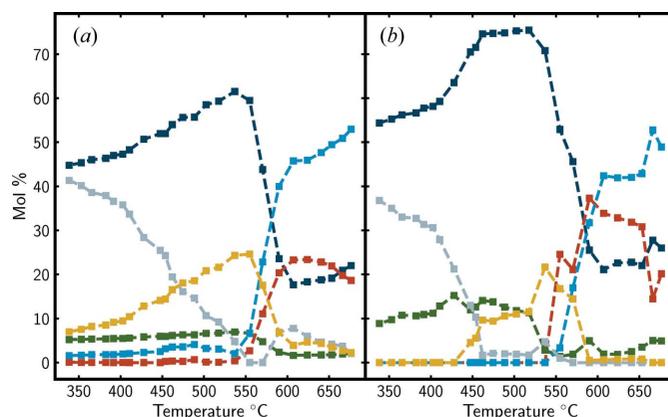

Figure 7
Comparison of mol % calculations when performed on (a) NMF component ratios and (b) sequential Rietveld fits on every dataset. The color of the line in each graph corresponds to a specific chemical compound, with the main species at low temperature being LaOCl (dark blue) + tetragonal $LiMnO_2$ (gray) + Mn_3O_4 (green). At intermediate temperatures orthorhombic $LiMnO_2$ (gold) is formed, and at high temperature the main products are $LaMnO_3$ (light blue) + LiCl (red), with a significant amount of left over LaOCl.

component. The plot in Fig. 7(a) was then obtained by multiplying the NMF component weights by the chemical weights in each NMF component to obtain the evolution of chemical species with temperature. The agreement with the chemical mole percentages obtained by carrying out laborious multi-phase Rietveld refinements on each of the datasets is very good. This shows that the NMF decomposition yields chemically sensible components, despite not using any chemical information to carry out the decomposition.

6. Conclusions

We have reported the *nmfMapping@PDFitc* app, which makes NMF more easily accessible to diffraction experimentalists, both on a beamline and at their home institution. This tool provides users with a fast, reliable and intuitive way to gain physically meaningful information about large datasets rapidly. To use the app, the user can simply create and log in to their own account on the *PDFitc* platform at https://pdfitc.org and upload their set of data.

End users should remember that NMF is a purely mathematical construct. It is important to exercise good sense and physical intuition when interpreting scientific results from an NMF. However, it is emerging as a powerful approach for rapidly assessing and extracting insight from large sets of related PDF and diffraction data, and we hope that *nmfMapping@PDFitc* can be a useful tool to help in its application.

research papers

References

Ashiotis, G., Deschildre, A., Nawaz, Z., Wright, J. P., Karkoulis, D., Picca, F. E. & Kieffer, J. (2015). *J. Appl. Cryst.* **48**, 510–519.

Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P. & Plemmons, R. J. (2007). *Comput. Stat. Data Anal.* **52**, 155–173.

Borkiewicz, O. J., Shyam, B., Wiaderek, K. M., Kurtz, C., Chupas, P. J. & Chapman, K. W. (2012). *J. Appl. Cryst.* **45**, 1261–1269.

Brunet, J.-P., Tamayo, P., Golub, T. R. & Mesirov, J. P. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 4164–4169.

Christensen, C. K. & Ravnsbæk, D. B. (2021). *J. Phys. Energy*, **3**, 031002.

Chupas, P. J., Chapman, K. W. & Lee, P. L. (2007). *J. Appl. Cryst.* **40**, 463–470.

Egami, T. & Billinge, S. J. L. (2012). *Underneath the Bragg peaks: Structural Analysis of Complex Materials. Pergamon Materials Series*, Vol. 16, 2nd ed. Amsterdam: Elsevier.

Farrow, C. L., Juhas, P., Liu, J. W., Bryndin, D., Božin, E. S., Bloch, J., Proffen, T. & Billinge, S. J. L. (2007). *J. Phys. Condens. Matter*, **19**, 335219.

Filik, J., Ashton, A. W., Chang, P. C. Y., Chater, P. A., Day, S. J., Drakopoulos, M., Gerring, M. W., Hart, M. L., Magdysyuk, O. V., Michalik, S., Smith, A., Tang, C. C., Terrill, N. J., Wharmby, M. T. & Wilhelm, H. (2017). *J. Appl. Cryst.* **50**, 959–966.

Geddes, H. S., Blade, H., McCabe, J. F., Hughes, L. P. & Goodwin, A. L. (2019). *Chem. Commun.* **55**, 13346–13349.

Hammersley, A. P. (2016). *J. Appl. Cryst.* **49**, 646–652.

Juhás, P., Davis, T., Farrow, C. L. & Billinge, S. J. L. (2013). *J. Appl. Cryst.* **46**, 560–566.

Lee, D. D. & Seung, H. S. (1999). *Nature*, **401**, 788–791.

Liu, C.-H., Tao, Y., Hsu, D., Du, Q. & Billinge, S. J. L. (2019). *Acta Cryst.* A**75**, 633–643.

Liu, C.-H., Wright, C. J., Gu, R., Bandi, S., Wustrow, A., Todd, P. K., O'Nolan, D., Beauvais, M. L., Neilson, J. R., Chupas, P. J., Chapman, K. W. & Billinge, S. J. L. (2021). *J. Appl. Cryst.* **54**, 768–775.

Long, C. J., Bunker, D., Li, X., Karen, V. L. & Takeuchi, I. (2009). *Rev. Sci. Instrum.* **80**, 103902.

Maffettone, P. M., Daly, A. C. & Olds, D. (2021). *Appl. Phys. Rev.* **8**, 041410.

Maier, W. F., Stöwe, K. & Sieg, S. (2007). *Angew. Chem. Int. Ed.* **46**, 6016–6067.

Martinolich, A. J. & Neilson, J. R. (2017). *Chem. Mater.* **29**, 479–489.

O'Nolan, D., Huang, G., Kamm, G. E., Grenier, A., Liu, C.-H., Todd, P. K., Wustrow, A., Thinh Tran, G., Montiel, D., Neilson, J. R., Billinge, S. J. L., Chupas, P. J., Thornton, K. S. & Chapman, K. W. (2020). *J. Appl. Cryst.* **53**, 662–670.

Paatero, P. & Tapper, U. (1994). *Environmetrics*, **5**, 111–126.

Pauca, V. P., Piper, J. & Plemmons, R. J. (2006). *Linear Algebra Appl.* **416**, 29–47.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). *J. Mach. Learn. Res.* **12**, 2825–2830.

Qiu, X., Thompson, J. W. & Billinge, S. J. L. (2004). *J. Appl. Cryst.* **37**, 678.

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

Stinton, G. W. & Evans, J. S. O. (2007). *J. Appl. Cryst.* **40**, 87–95.

Toby, B. H. & Von Dreele, R. B. (2013). *J. Appl. Cryst.* **46**, 544–549.

Todd, P. K., Wustrow, A., McAuliffe, R. D., McDermott, M. J., Tran, G. T., McBride, B. C., Boeding, E. D., O'Nolan, D., Liu, C.-H., Dwaraknath, S. S., Chapman, K. W., Billinge, S. J. L., Persson, K. A., Huq, A., Veith, G. M. & Neilson, J. R. (2020). *Inorg. Chem.* **59**, 13639–13650.

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

Wustrow, A., Huang, G., McDermott, M. J., O'Nolan, D., Liu, C.-H., Tran, G. T., McBride, B. C., Dwaraknath, S. S., Chapman, K. W., Billinge, S. J. L., Persson, K. A., Thornton, K. & Neilson, J. R. (2021). *Chem. Mater.* **33**, 3692–3701.

Yang, L., Culbertson, E. A., Thomas, N. K., Vuong, H. T., Kjær, E. T. S., Jensen, K. M. Ø., Tucker, M. G. & Billinge, S. J. L. (2021). *Acta Cryst.* A**77**, 2–6.

Yang, L., Juhás, P., Terban, M. W., Tucker, M. G. & Billinge, S. J. L. (2020). *Acta Cryst.* A**76**, 395–409.