**Generic Implementation of Bragg Profile in Reverse Monte Carlo Algorithm and Resolution Correction for Modeling the Total Scattering Data**

**Author:** Yuanpeng Zhang1,2, Maksim Eremenko1,3, Victor Krayzman1, Igor Levin1

1Neutron Scattering Division, Oak Ridge National Laboratory (ORNL), Oak Ridge, Tennessee 37831, United States

2Materials Measurement Science Division, National Institute of Standards and Technology (NIST), 100 Bureau Drive, Gaithersburg, Maryland 20899, United States

3

**Abstract:** A generic algorithm has been implemented in the RMCProfile package to include Bragg pattern for fitting total scattering data based on the reverse Monte Carlo methodology. As compared to previous implementation, the new algorithm is realized through tabulating the peak profiles extracted from the unit-cell-based Rietveld/LeBail refinement for Bragg pattern. With such an approach, including Bragg data in the supercell-RMC-based fitting scheme is not limited to the choice of profile function for describing the peak shape. This is beneficial for fitting the total scattering data measured from various instruments, whereby one may have to define specific peak profile function per needed to account for the unique peak shape. Following the same approach, we also implemented a matrix-format resolution correction in both real and reciprocal space for calculating total scattering pattern. Tests have been conducted on data from various instruments and the results prove the robustness of our new implementations.

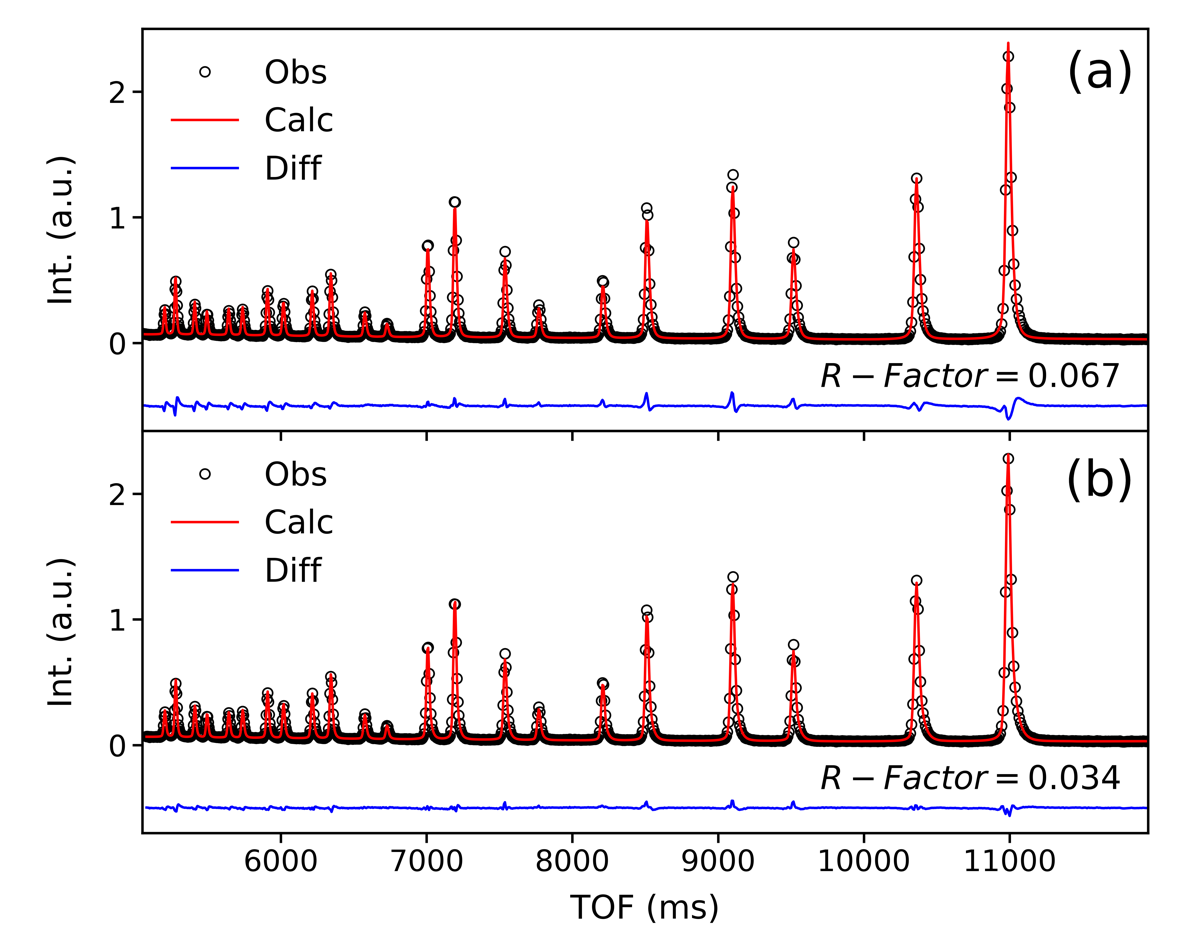
**Introduction**

For data-driven crystal structure modeling techniques such as those based on the reverse Monte Carlo (RMC) algorithm, including more datasets in the fitting is beneficial for improving the reliability of the obtained structural model. Specifically, for total scattering data, it contains both Bragg peaks and diffuse scattering signal. They focus on long-range and short-range ordering, respectively. Therefore, including both in the RMC fitting for total scattering data will comprehensively cover local, to medium and to long-range ordering. In 2007, M. Tucker, et al. introduced an approach for including Bragg data into the supercell-based RMC fitting, implemented in the RMCProfile package. With the functional form of peak profiles implemented in RMCProfile, one needs to conduct a unit-cell-based Rietveld/LeBail refinement for Bragg pattern with GSAS to extract the parameters concerning the chosen peak profile type and background. Then in RMCProfile, peak intensities can be calculated from the given supercell to rebuild the full Bragg pattern. In practice, such an implementation is limited to certain function types available in GSAS. However, for certain instrument, the peak shape of the Bragg pattern cannot be well described by the traditionally used peak profile functions (e.g. a Pseudo-Voigt shape convolved with back-to-back exponential function, usually used for describing asymmetric peaks). For example, for NOMAD time-of-flight (TOF) diffractometer at Spallation Neutron Source (SNS), ORNL, the special peak shape brought in by moderator needs special care, i.e. one may have to define specific peak profile functions per required. To this end, we develop a generic way of implementing Bragg profiles for supercell-RMC based methodologies (here, we use RMCProfile). Following similar approach with M. Tucker, et al., we tabulated the profiles with the help of a Python GUI, instead of implementing the peak profile functions explicitly in RMCProfile. With such an algorithm, to include Bragg data in RMCProfile fitting, users are not restricted to the limited number of peak profile types. But rather one can use any pre-defined function type (with, e.g. the commercially available Topas package) to describe Bragg peak shapes as needed.

Apart from including Bragg data in the fit, the Bragg peak shape can also be important concerning the correction for the resolution effect in modeling total scattering data. Traditionally, the resolution effect is corrected through a parameterized way (e.g. in PDFgui), where one extracts the necessary parameters of the analytical convolution function from fitting the total scattering pattern for a standard sample (e.g. Si). Such an approach relies on the assumption that the broadening and dampening in the Q-space scattering pattern measured for standard sample purely comes from the resolution effect of instrument. This then brings in the same concerns as with describing Bragg peaks for certain instrument, i.e. the resolution effect brought in by special peak shapes may not be described adequately by analytical functions. To this end, a numerical approach for correcting the resolution effect was proposed by M. Eremenko, et al [**Maksim’s JAC paper**]. In this article, we revisit such an approach by pictorially giving a demonstration and detailed explanation about how it works, especially in terms of code implementation in the RMCProfile package. Through fitting the Bragg pattern measured for a standard sample (here, Si is used), one then builds up resolution matrices which will be convolved into the as-calculated total scattering pattern either in real or reciprocal space. Again, with the help of our developed Python GUI, one has the freedom to use any type of user-defined peak profile function to describe Bragg peaks for generating the resolution matrix. With such implementations for Bragg profile and numerical resolution matrix, we were able to, for the first time, successfully conduct a comprehensive RMC fitting for the TOF data measured at NOMAD diffractometer at SNS, ORNL, including wide-range real-space (up 70 Å) and reciprocal space total scattering data together with Bragg pattern.

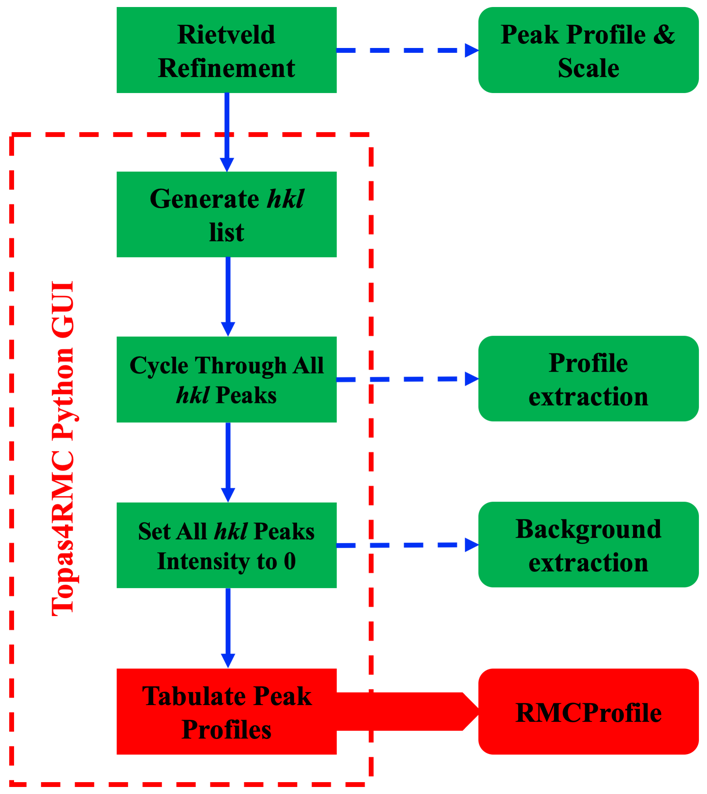
**Profile Implementation**

First in Fig. 1, it is presented a comparison of the Bragg pattern refinement for standard Si sample measured at NOMAD TOF diffractometer at SNS, ORNL, with GSAS type-III function (Fig. 1a) and a specially designed moderator function (Fig. 1b, [**Jue’s JACS paper**]). The Rwp factor labeled out in each figure directly shows the specially designed moderator profile function provides a better description for the Bragg peak shape.



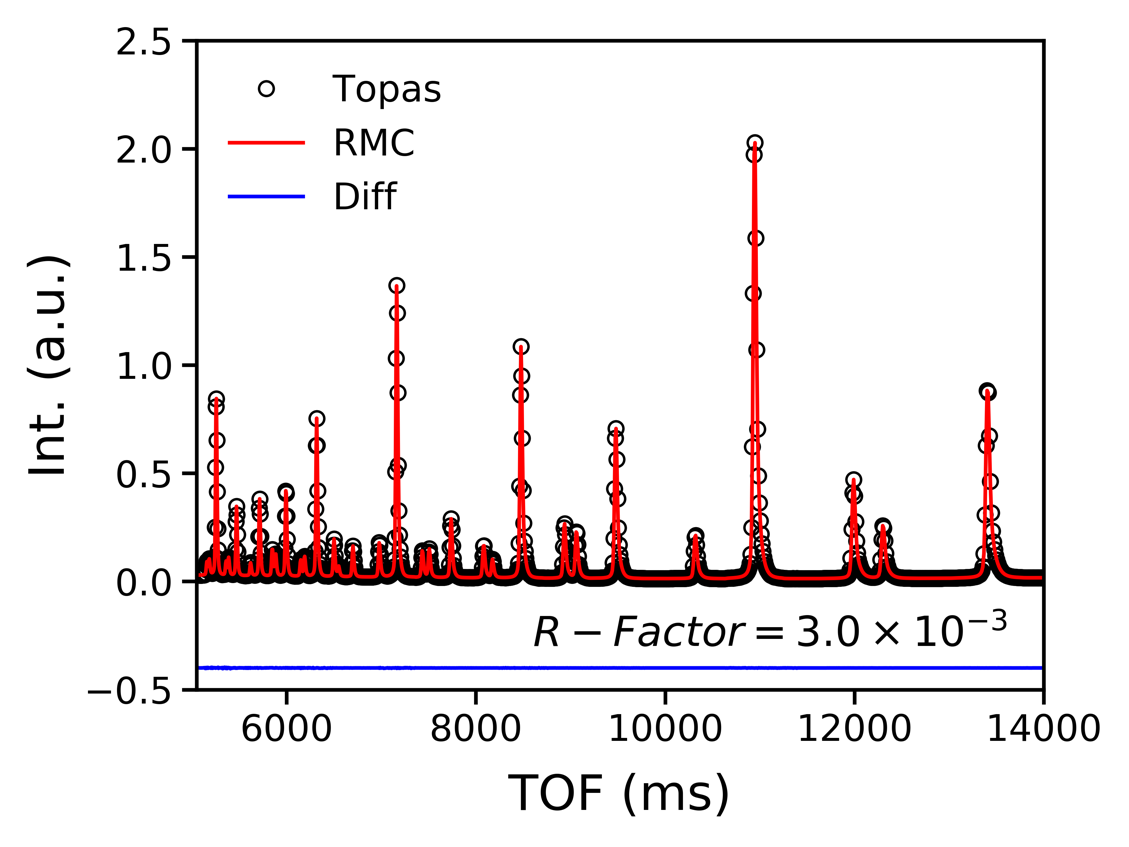
**Figure. 1**. The comparison of Bragg pattern refinement with (a) GSAS III function and (b) a specially designed moderator function, for the data measured at NOMAD TOF diffractometer. The R-Factor labeled out in the figure is calculated following the equation:  (*i* represents data points,  and  for calculated and observed pattern, respectively), which applies across current report.

To include Bragg pattern into the supercell-RMC based total scattering data fitting (here, RMCProfile package is used), we did not explicitly implement the specific function form in RMCProfile. Instead, we created a GUI with Python to extract the peak profile data to a tabulated file. Through such a routine, no matter what format of the peak profile function is defined (e.g. with Topas package, used in this report), one can always use the Python GUI to generate the corresponding tabulated peak profile data files for RMCProfile to read in. Here a diagram is created to demonstrate the workflow of our developed Python GUI for extracting the Bragg peak profiles, as presented in Fig. 2. First, one needs to run a successfully Rietveld refinement to extract the scale factor. Then the Python GUI is responsible for generating the *hkl* list and cycling through every *hkl* in the list, where the corresponding peak intensity is set to 1 and intensities for all the other *hkl*s to 0. In each cycle, Topas will be called to run with the corresponding input file and accordingly the profiles for all hkls will be generated, which is then written into a tabulated file for RMCProfile to read in. As a next step, the GUI will set the intensities of all *hkl*s to 0 for extracting the background.



**Figure. 2**. A diagram illustrating the workflow of extracting the Bragg peak profiles with our developed Python GUI named Topas4RMC.

To demonstrate the applicability of our approach, we carried out a proof-of-principle test. Here RMCProfile was run 0 step with a perfect CeO2 structure model (i.e. all atoms are located at their ideal positions without any distortion). The same perfect structure is then given to Topas with thermal parameters set to 0. In principle, if our routine works as expected, one will get identical result from both, which is indeed what we obtained (within numerical error), as shown in Fig. 3 (refer to the nearly flat blue difference curve).



**Figure. 3**. The comparison of calculated Bragg pattern for the same ideal structure using Topas and RMCProfile.

Here it should be pointed out that numerical uncertainty can be introduced during the output and reading-in of the tabulated profiles, due to the rounding up errors. For the proof-of-principle test, we also calculated the factor introduced by such numerical error, which is presented in Fig. 3. Compared to the factor (0.046, refer to Fig. S1 in the supplementary information, SI) of a Rietveld refinement for the Bragg data of CeO2 measured at NOMAD, the introduced numerical uncertainty can be ignored. Also, it’s worth pointing out that such tabulated routine described here is generally applicable, but not limited to the combination of RMCProfile and Topas packages used in this report.

**Resolution Correction**

In practice when measuring the total scattering data, the limited resolution of instrument will bring in broadening effect to the *Q*-space pattern. Meanwhile, there also exists dampening in the *Q*-space pattern as *Q* increases, compared to the ideal condition of instrument. When Fourier transforming the *Q*-space pattern to real-space, such resolution effects will be transferred. This will bring in dampening and broadening effects in the real-space pattern – corresponding to broadening and dampening in *Q*-space, respectively. Therefore, correction for resolution effect is needed in both spaces when modeling total scattering data, in accordance with practical measurement. Here, the idea of our resolution effect correction is, as usual, to assume that the broadening and dampening effects in the *Q*-space pattern measured for a standard sample (e.g. NIST Si) are purely due to the resolution effect of instrument. Thereby, when measuring samples in practice, one needs to measure a standard sample (here we use NIST Si for demonstration) under the same condition – the same sample environment and container. One then carries out Rietveld or LeBail refinement for the Bragg data of NIST Si to extract the peak profiles, with which a normalized resolution matrix in *Q*-space can be built up and further convoluted into the as-calculated *Q*-space total scattering pattern.

Mathematically, the convolution for correcting the resolution effect can be given as:

 (1)

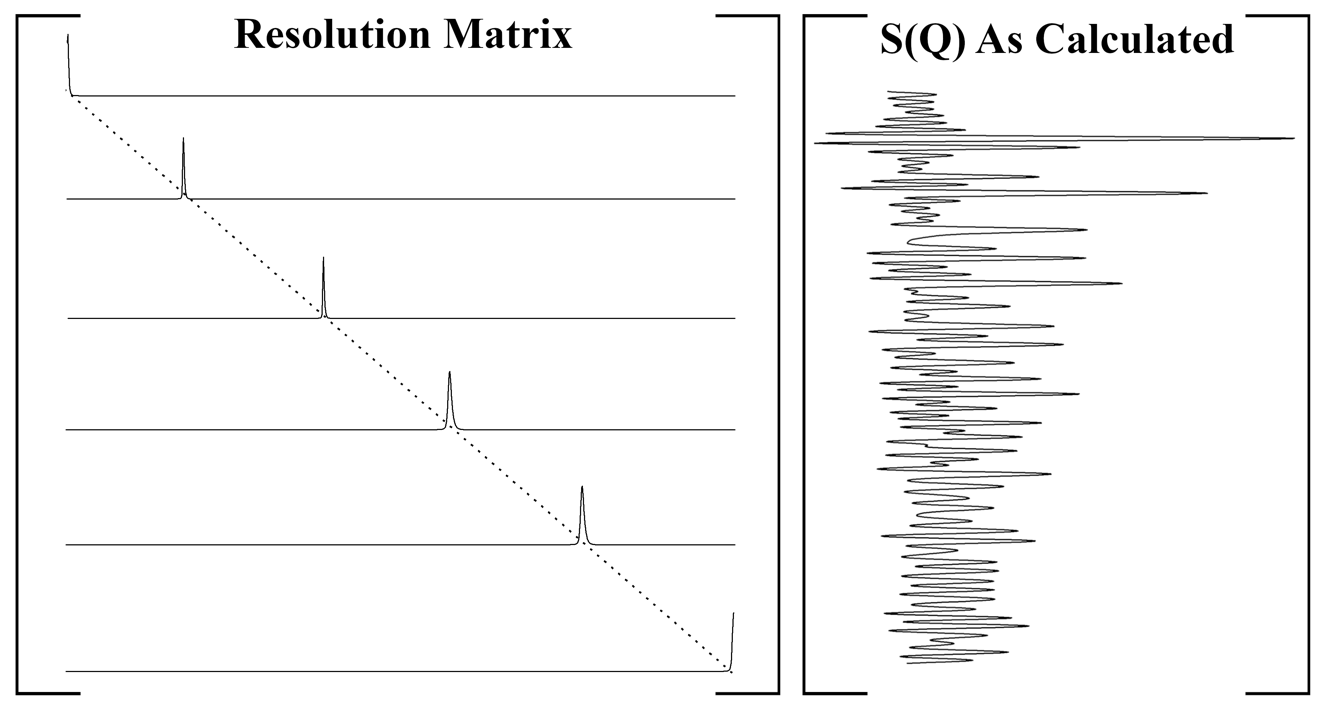
where  is the as-calculated *Q*-space total scattering pattern, which is the Fourier transform () of the as-calculated real-space pattern () for a given structural model. Here the RMCProfile symbols were used to describe the total scattering pattern in both real and reciprocal space (refer to the article by D. Keen for various terminologies used in the community).  represents the convolution function for the resolution effect correction, where one has two parameters – one for *Q* point and the other for *Q* center. This means that in practice, the peak shape may vary as the peak center moves in *Q*-space. Furthermore, the real-space  pattern with the resolution effect corrected can be given as:

 (2)

For practical implementation in RMCProfile, the equations (1) and (2) need to be translated into their matrix forms, which is given as follows:

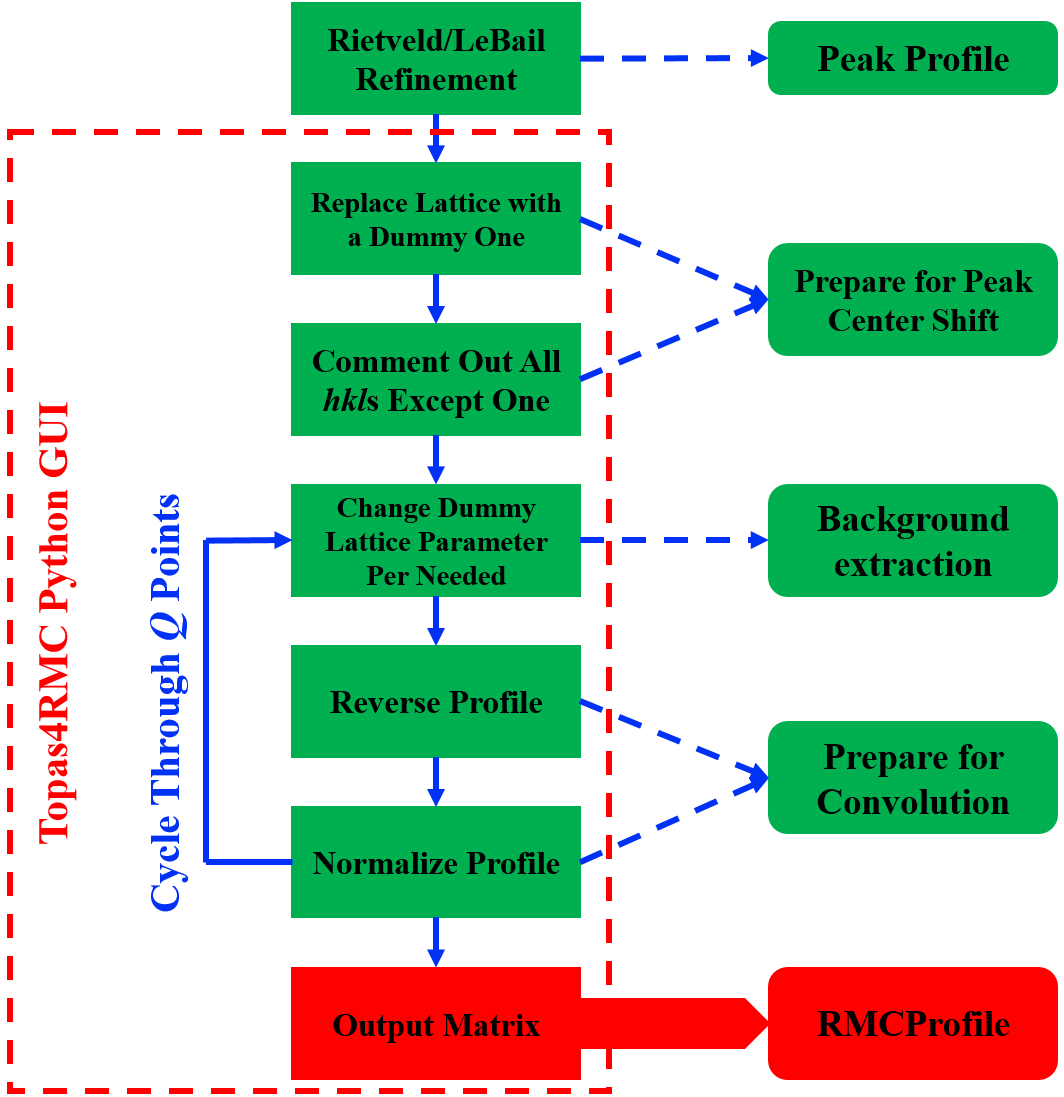
 (3)

Here ‘’ and ‘’ represents column and square matrix, respectively. represents the as-calculated one-dimensional real-space pattern and  represents the resolution matrix generated using the peak profile parameters obtained through fitting the Bragg pattern of a standard sample.  and  represents the Fourier transform matrix from real- to *Q*-space and from *Q*- to real-space, respectively. Finally, column matrix containing the Q- and real-space total scattering pattern with the resolution effect corrected can be obtained – represented by  and , respectively. The idea of such a convolution for correcting the resolution effect in its matrix form is illustrated in Fig. 4.



**Figure. 4**. Illustration of the convolution for correcting the resolution effect, in matrix form.

In practice, when  is calculated on certain *Q*-grid, one needs to calculate the peak profile with each *Q* point in the *Q*-grid respectively as the peak center to build up a resolution matrix. In the matrix, the *i*th row contains the peak profile data corresponding to *i*th Q point as the peak center. Also, the corresponding peak needs to be flipped with respect to the peak center, considering the mathematical definition of convolution. To implement the resolution correction in RMCProfile, our developed Python GUI was also used here to tabulate the resolution matrix. Again, Topas was used for refining the Bragg pattern of NIST Si. After that, one can create a dummy lattice in Topas input file and change the dummy lattice parameters in accordance with the *Q* point as needed. In this way, one can cycle through the whole *Q*-grid to generate the resolution matrix. Detailed flowchart of the resolution matrix generation with our Python GUI is presented in Fig. 5.



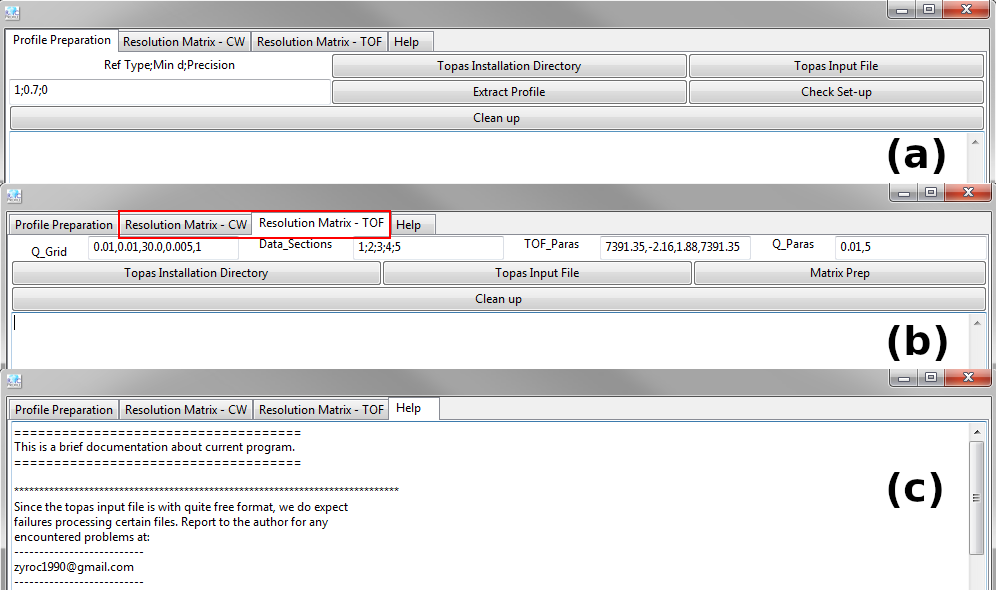
**Figure. 5**. A diagram illustrating the workflow of tabulating the resolution matrix with our developed Python GUI named Topas4RMC.

For spallation neutron sources, typically *Q*-space total scattering data is obtained by combining the pattern from multiple banks. For example, NOMAD diffractometer at SNS has 6 banks and Polaris at ISIS neutron source has 5 banks, etc. Since each bank has different resolution, different weights should be assigned accordingly when combing the resolution matrix from various banks. Mathematically, one has:

 (4)

where  represents the *Q*-space total scattering pattern of for ideal situation (i.e. no resolution effect),  represents the resolution effect of bank-*i* to be convolved into , and  represents the weight assigned to bank-*i* when combining the pattern from  banks to obtain the overall . Here the first line in equation-(4) corresponds to the experimental measurement, where one cannot decouple the terms inside the square bracket. The second line corresponds to theoretical calculation, where one can obtain the ideal  given a structural model. The summation term then refers to the combination of resolution matrix with the correspondingly assigned weight for each bank. In practice, different weight schemes may be used by different diffractometers. From equation-(4), it can be seen that the weight scheme used for combining the resolution matrices should be consistent with that used by the diffractometer where the experimental data is collected.

Graphical User Interface



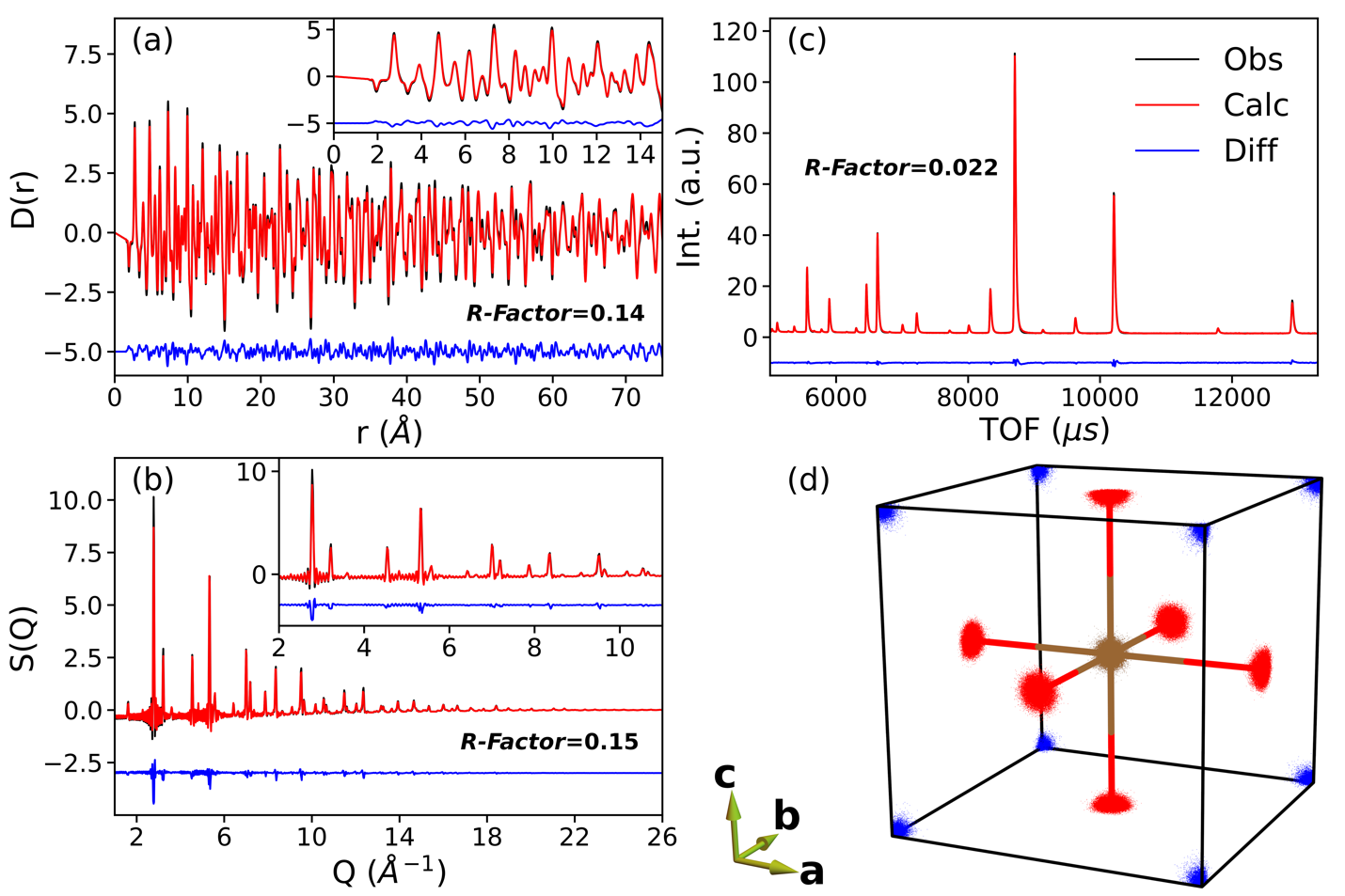
**Figure. 6**. The Python GUI developed for generating the tabulated Bragg profile (panel-a) and resolution matrix (panel-b, for constant wavelength and time-of-flight facilities). A help tab is also created (panel-c) to include guidance to use the software.

For current report, the Topas package was used to prepare the Bragg profiles and resolution matrix, since with Topas one can easily define the profile functions using macros. This is especially important for describing special peak shapes of, e.g. data measured at NOMAD diffractometer. The Python GUI designed for preparing the tabulated data to work with RMCProfile, based on the flowchart illustrated in Fig. 2 and Fig. 5, is shown in Fig. 6. Here it should be mentioned that the working principle presented in Fig. 2 and Fig. 6 is not limited to Topas package, but instead it can be extended per Rietveld package concerned.

The GUI is created using the wxpython framework with Python (>=3.6.0), which requires several other modules to run: numpy (>=1.16.4), scipy (>=1.2.1) and matplotlib (>=3.1.0). Detailed instructions about how to use the GUI is implemented in the ‘help’ tab as shown in Fig. 6 (c).

**Results and Discussion**

To test our implementations, we first conducted a comprehensive fit involving multiple datasets for the SrTiO3 (STO) data measured at Polaris neutron TOF diffractometer at ISIS neutron source. STO is with a cubic structure (space group: *Pm-3m*) and for the fitting with RMCProfile, a  supercell was used. The tabulated Bragg profile and resolution matrix was used for fitting the Bragg and total scattering pattern (both real and reciprocal space), respectively. The system contains 320,000 atoms, and ~9,000,000 moves (~28 moves/atom) were generated among which ~4,000,000 moves (~12 moves/atom) were accepted. The fitting results are presented in Fig. 6, from which one can observe a reasonably good fit for G(r), S(Q) and Bragg datasets. The obtained supercell configuration was then projected back to the unit cell (with *P1* symmetry), from which one can observe a distribution of atoms around their ideal crystallographic positions (see Fig. 7d). Here for Sr (blue dots in Fig. 7d) and Ti (brown dots in Fig. 7d) atoms, the distribution is observed to be uniform, which is consistent with the cubic symmetry of STO. For O atoms, slight anisotropy can be observed. To further quantitatively characterize the spatial distribution of atoms, thermal ellipsoid analysis was conducted, the result of which is presented in Table. 1. The nearly identical value of U11, U22 and U33 confirms the isotropic distribution of Sr and Ti atoms. For O atoms, U11 (U22 or U33) for O-3 (O-4 or O-5) is smaller than the other two, indicating a wider distribution of O-3 (O-4 or O-5) atoms within the *y*-*z* (*x-z* or *x-y*) plane. This is

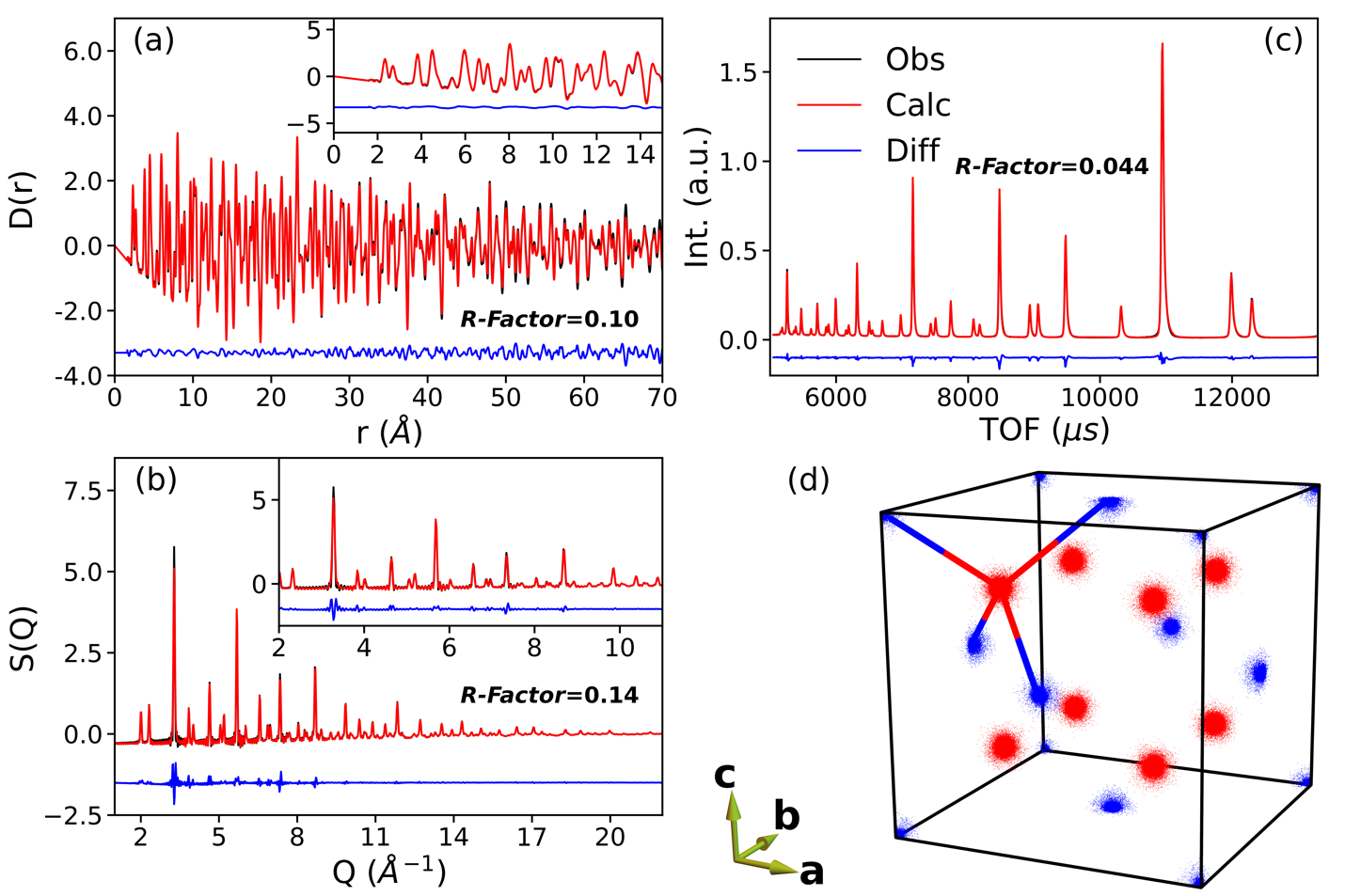


**Figure. 7**. RMC fitting results for (a) *D(r)*, (b) *S(Q)* and (c) Bragg data. Insets in (a) and (b) are the corresponding zoom-in of the low-*r* and low-*Q* part, respectively. (d) STO unit cell with atoms from the RMC supercell projected back. Here, Sr, Ti and O atoms are represented with blue, brown and red color, respectively.

**Table 1**. Average atomic coordinates and their corresponding thermal parameters for STO. Since the principal axes of the thermal ellipsoid are almost along the main axes (*a*, *b* and *c*), the small values of U12, U13 and U23 are not presented here (refer to Table S1 in the SI).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *x* | *y* | *z* | U11 | U22 | U33 |
| Sr-1 | 0.0011 | -0.0007 | 0.0000 | 0.0056 | 0.0057 | 0.0058 |
| Ti-2 | 0.5002 | 0.5000 | 0.4999 | 0.0039 | 0.0041 | 0.0042 |
| O-3 | 0.0005 | 0.5005 | 0.5000 | 0.0033 | 0.0082 | 0.0083 |
| O-4 | 0.4993 | -0.0002 | 0.5003 | 0.0080 | 0.0034 | 0.0084 |
| O-5 | 0.4992 | 0.5005 | -0.0002 | 0.0081 | 0.0083 | 0.0035 |

the natural result of the strong Ti-O bonding (see Fig. 7d), which forms the local TiO6 octahedron. Further analysis containing the projection of atomic positions onto Cartesian coordinate planes (*x-y*, *x-z* and *y-z*) and column projection of atomic displacement can be found in the SI (Fig. S2 – Fig. S9).



**Figure. 8**. RMC fitting results for (a) *D(r)*, (b) *S(Q)* and (c) Bragg data of bulk ceria. Insets in (a) and (b) are the corresponding zoom-in of the low-*r* and low-*Q* part, respectively. (d) CeO2 unit cell with atoms from the RMC supercell projected back. Here, Ce and O atoms are represented with blue and red color, respectively.

Another test was carried out for the datasets of bulk ceria measured at NOMAD diffractometer, SNS, ORNL. Ceria is with a cubic structure (space group: *Fm3m*) and for the fitting with RMCProfile, a  supercell was used. As for STO case, the tabulated Bragg profile and resolution matrix was also used here for fitting the Bragg and total scattering pattern (both real and reciprocal space), respectively. The system contains 324,000 atoms, and ~12,000,000 moves (~37 moves/atom) were generated among which ~3,600,000 moves (~11 moves/atom) were accepted. The fitting results for D(r), S(Q) and Bragg data are presented in Fig. 8 (a), (b) and (c), respectively. The overall fitting quality is reasonably good (refer to Fig. 8 for the calculated *R*-factor). Again, the obtained supercell from RMCProfile was projected back to the unit cell for further inspection of the reasonability of the configuration (see Fig. 8d). Uniform distribution of Ce atoms around their crystallographic positions can be observed. As for O atoms, in contrast to those in STO, the tetragonal bonding environment of O atoms in ceria tends to give an isotropic distribution, which agrees well with the results presented in Fig. 8 (d).

**Table 2**. Average atomic coordinates and their corresponding thermal parameters for bulk ceria. Since the principal axes of the thermal ellipsoid are almost along the main axes (*a*, *b* and *c*), the small values of U12, U13 and U23 are not presented here (refer to Table S3 in the SI).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *x* | *y* | *z* | U11 | U22 | U33 |
| Ce-1 | -0.0002 | 0.0001 | 0.0001 | 0.0056 | 0.0056 | 0.0055 |
| O-2 | 0.7499 | 0.7501 | 0.7503 | 0.0076 | 0.0075 | 0.0076 |
| O-3 | 0.7499 | 0.7503 | 0.2495 | 0.0076 | 0.0078 | 0.0076 |
| Ce-4 | 0.0000 | 0.5000 | 0.4999 | 0.0054 | 0.0056 | 0.0055 |
| O-5 | 0.7499 | 0.2497 | 0.2500 | 0.0076 | 0.0078 | 0.0076 |
| O-6 | 0.7504 | 0.2500 | 0.7498 | 0.0077 | 0.0078 | 0.0079 |
| Ce-7 | 0.5000 | 0.0002 | 0.4999 | 0.0056 | 0.0056 | 0.0056 |
| O-8 | 0.2504 | 0.7500 | 0.2501 | 0.0076 | 0.0077 | 0.0075 |
| O-9 | 0.2497 | 0.7497 | 0.7501 | 0.0078 | 0.0077 | 0.0077 |
| Ce-10 | 0.5002 | 0.5000 | 0.0001 | 0.0057 | 0.0055 | 0.0056 |
| O-11 | 0.2498 | 0.2502 | 0.7496 | 0.0077 | 0.0077 | 0.0076 |
| O-12 | 0.2501 | 0.2500 | 0.2503 | 0.0075 | 0.0077 | 0.0078 |

Quantitative results of the average coordinates of Ce and O atoms in the collapsed unit cell, together with the corresponding thermal parameters obtained from the thermal ellipsoid analysis are presented in Table. 2. The isotropic distribution of both Ce and O atoms is further confirmed by the numerical values of U11, U22 and U33. Further analysis containing the projection of atomic positions onto Cartesian coordinate planes (*x-y*, *x-z* and *y-z*) and column projection of atomic displacement can be found in the SI (Fig. S10 – Fig. S14).

**Summary**

Here in this report, we for the first time demonstrate a generic algorithm for incorporating Bragg pattern and coping with the resolution effect (in both real and reciprocal space) for the supercell-RMC-based fitting of total scattering data. Such an algorithm is implemented in RMCProfile package, with a Python GUI specifically designed for generating the tabulated data (Bragg profiles and resolution matrix) to realize our algorithm in practice. The RMCProfile package was then used to carry out comprehensive fits including multiple datasets – STO data measured at Polaris diffractometer and bulk ceria data measured at NOMAD diffractometer. Both of the two test cases show reasonably good fit quality. The obtained structural configurations were carefully examined with various approaches to guarantee the reasonability of the fitting. The results here demonstrate the general applicability of our methodology, which, in fact, is not limited to the combination of RMCProfile and Topas used in this report.

**Acknowledgement**

**References**