



AUTOMATED FULL-PATTERN SUMMATION OF X-RAY POWDER DIFFRACTION DATA FOR HIGH-THROUGHPUT QUANTIFICATION OF CLAY-BEARING MIXTURES

BENJAMIN M. BUTLER¹* AND STEPHEN HILLIER^{1,2}

¹The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, UK

²Department of Soil and Environment, Swedish University of Agricultural Sciences (SLU), SE-75007 Uppsala, Sweden

Abstract—X-ray powder diffraction (XRPD) is found consistently to be the most accurate analytical technique for quantitative analysis of clay-bearing mixtures based on results from round-robin competitions such as the Reynolds Cup (RC). A range of computationally intensive approaches can be used to quantify phase concentrations from XRPD data, of which the ‘full-pattern summation of prior measured standards’ (FPS) has proven accurate and parsimonious. Despite its proven utility, the approach often requires time-consuming selection of appropriate pure reference patterns to use for a given sample. As such, applying FPS to large and mineralogically diverse datasets is challenging. In the present work, the accuracy of an automated FPS algorithm implemented within the *powdR* package for the R Language and Environment for Statistical Computing was tested on a set of 27 samples from nine RC contests. The samples represent challenging and diverse clay-bearing mixtures with known concentrations, with the added advantage of allowing the accuracy of the algorithm to be compared with results submitted to previous contests. When supplied with a library of 201 reference patterns representing a comprehensive range of phases that may be encountered in natural clay-bearing mixtures, the algorithm selected appropriate phases and achieved a mean absolute bias of 0.57% for non-clay minerals ($n = 275$), 2.37% for clay minerals ($n = 120$), and 4.43% for amorphous phases ($n = 14$). This accuracy would be sufficient for top-3 placings in all nine RC contests held to date (RC1 = 2nd, RC2 = 2nd, RC3 = 1st, RC4 = 2nd, RC5 = 1st, RC6 = 3rd, RC7 = 3rd, RC8 = 1st, RC9 = 2nd). The comparatively low values of absolute bias in combination with the competitive placings in all RC contests tested is particularly promising for the future of automated quantitative phase analyses by XRPD.

Keywords—Full pattern fitting · Full-pattern summation · Quantitative mineralogy · Reynolds Cup · X-ray powder diffraction

INTRODUCTION

X-ray powder diffraction is a widely applied analytical technique in the study of mineral mixtures, both for the qualitative identification of crystalline components and the quantitative determination of their concentrations. Quantitative mineralogy from XRPD data has a long history, dating back to the early 20th Century (Navias 1925; Clark and Reynolds 1936). Since these early examples, advances in instrumentation, databases (ICDD 2016; Gates-Rector and Blanton 2019), sample-preparation methods (Hillier 1999), and software (Rietveld 1969; Bergmann et al. 1998; Chipera and Bish 2002; Eberl 2003; Doebelin and Kleeberg 2015) have now made obtaining accurate quantitative analysis of even very challenging mixtures from XRPD data possible (Raven and Self 2017).

Modern instrumentation and sample-preparation methods can also result in the accumulation of large, high-throughput datasets containing hundreds to thousands of reproducible diffractograms – each representing a precise mineralogical signature of a sample (Woodruff et al. 2009; Butler et al. 2018, 2020). High-throughput datasets with limited mineralogical variation and primarily ordered crystalline phases can be

quantified readily using the now widely adopted Rietveld approach (Rietveld 1969). With increasing mineralogical diversity of a dataset along with the presence of disordered (e.g. clay minerals) and amorphous phases (e.g. volcanic glass or soil organic matter), however, the process of identifying and quantifying components in large numbers of samples can become a challenging and particularly time-consuming undertaking. These challenges create a need for an approach that can move towards automation of mineral identification and quantification of high-throughput, mineralogically diverse datasets containing clay-bearing samples, whilst maintaining good accuracy.

Round robin competitions such as the Reynolds Cup challenge participants to quantify complex mixtures containing a wide variety of clay minerals, with an overall goal of stimulating improvements in analytical techniques for characterization of clay-bearing mixtures. Of the many available analytical techniques, XRPD analysis of bulk powders (i.e. randomly oriented milled samples) is usually the primary technique for quantifying RC samples (Omotoso et al. 2006; Raven and Self 2017), but is used in combination with auxiliary analyses to complement the precision of the initial stage of mineral identification. Frequently, these auxiliary techniques often include XRPD measurements of the clay fraction (<2 μm) of oriented specimens subject to glycolation and subsequent heat treatments, along with total bulk-sample elemental analysis to cross check the feasibility of the quantitative mineralogical results. Since 2002 the nine biennial RC contests have promoted the

* E-mail address of corresponding author: Benjamin. Butler@hutton.ac.uk
DOI: 10.1007/s42860-020-00105-6
© The Clay Minerals Society 2021

advancement of protocols for quantitative mineralogy, from which a range of approaches have proven accurate, all of which rely on XRPD as the primary tool for quantification.

One approach for quantifying RC samples that has been placed in the top three at each of the nine contests (2002–2018) is the full-pattern summation (FPS) of prior measured standards (Smith et al. 1987; Chipera and Bish 2002; Eberl 2003; Vogt et al. 2002; Omotoso et al. 2006; Raven and Self 2017). This approach is based upon the principle that an observed XRPD measurement is the sum of individual crystalline and amorphous components within the sample, including instrument-dependent contributions (Smith et al. 1987; Chipera and Bish 2002). Full-pattern summation utilizes a reference library of pure diffraction patterns ('standards'/'reference patterns') which are preferably measured on the same instrument used to run the unknowns in order to best match the instrument-dependent variation in both the sample and reference library data (Chipera and Bish 2002; Omotoso et al. 2006; Eberl 2003). Upon optimizing an observed pattern based on the sum of contributions from the appropriate pure standards, all of these methods derive phase concentrations using Reference Intensity Ratios [RIRs; Hillier (2000)], which describe the diffraction intensity associated with a given phase relative to that of a standard (usually corundum, Al_2O_3). It is worth noting that the way in which the RIRs are described, derived, and formulated varies from one implementation to another. All of these recent FPS approaches also include the background in the reference library patterns on the assumption that background effects, such as those due to fluorescence, are also additive.

For the present investigation, the hypothesis was that, given a comprehensive reference library that can cover most, if not all, of the minerals that may be encountered in a given set of samples, the FPS approach can be automated to provide both identification and quantification of phases in mineralogically diverse datasets. The algorithm presented for doing so was implemented in version 1.2.3 of the *powdR* package (Butler and Hillier 2020; Butler and Hillier 2021) for the *R* Language and Environment for Statistical Computing (R Core Team 2020). Implementation in *R* implies that the software is open source and multi-platform. The automated algorithm uses a single bulk XRPD measurement in combination with a comprehensive reference library to identify and quantify the concentrations of non-clay, clay, and amorphous components. Here, a mineralogically diverse dataset comprising 27 RC samples, three from each of the previous nine contests (RC1 to RC9), has been utilized. The RC samples were considered most suitable for testing the accuracy of the automated approach for several reasons: (1) the dataset exhibits substantial mineralogical diversity; (2) implicit within this diversity is the presence of clay minerals and occasional amorphous phases; (3) all samples were prepared rigorously by independent laboratories; and (4) the availability of anonymous results for each contest allowed comparison of the accuracy relative to all other participants.

MATERIALS AND METHODS

X-ray Powder Diffraction

Sample preparation and measurement. Samples from RC1 to RC9 were available based on the participation of Stephen Hillier in all previous Reynolds Cup contests. For RC1 and RC2, samples were spiked with a known weight percentage of an internal standard (~20% corundum), whereas for RC3 through to RC9, the sample-preparation protocol was changed and all samples were prepared without addition of an internal standard.

Each of the 27 RC samples was prepared for XRPD as received by McCrone milling 3 g of sample for 12 min in ethanol and spray drying the resulting slurry to obtain a random powder specimen as described by Hillier (1999) and demonstrated by Kleeberg et al. (2008). This preparation was done at the time of each of the respective Reynolds Cups, so over a 16 year period (2002–2018). To enable further the detection of trace-mineral phases, very high quality diffraction data were recorded by scanning over the range $4\text{--}70^\circ 2\theta$ on a Bruker D8 using Ni-filtered $\text{Cu K}\alpha$ radiation, fixed divergence slits, and a Lynxeye XE detector, with counts recorded for 16 s per $0.0195^\circ 2\theta$ step yielding scan times of 16 h. These scans were already available for RC7 to RC9, but for RC1 to RC6 the spray-dried specimens were retrieved from their storage (8–18 y) in capped glass vials and re-run on the D8 diffractometer, which was not available in the authors' laboratory prior to RC7 (2014).

Reference library preparation. A reference library of standard XRPD patterns of pure minerals has been compiled in Stephen Hillier's laboratory over a period of time from specimens of pure minerals obtained from various mineral collections or purchased, such as from the Source Clays Repository of The Clay Minerals Society (Costanzo and Guggenheim 2001). The purity of the minerals was assessed mainly by XRPD data, and for many samples – especially the clay minerals – the best purity was obtained by picking or by size-fractionation procedures. Inevitably, small impurities remained in many samples, e.g. quartz is a ubiquitous contaminant of most clays, even very fine-size clay fractions. Where required, remaining impurities were, therefore, removed electronically by subtraction of the whole pattern of the pure phase impurity. All such treated patterns were scaled to a maximum intensity of 10,000 counts prior to determination of a full-pattern RIR from a mixture of the pure mineral (plus any impurities) with corundum as an added internal standard, for which the weight fractions were known. Further details of this procedure will be presented elsewhere. All standards were run under the same diffractometer conditions as the unknowns, except that the recording time per $0.0195^\circ 2\theta$ step for library standards was just 2 s. All backgrounds of the standards and samples were retained throughout.

The full reference library available for this investigation included 201 diffractograms of pure standards designed to cover most components associated with geologic, soil, and sediment samples. Of these, 76 were clay mineral/phylosilicate reference patterns, 116 were non-clay, and nine were amorphous,

nanocrystalline or paracrystalline (allophane, ferrihydrite, glass, obsidian, opal-CT, opal-A, aluminosilicate gel, organic matter, and graphite). Many of the library entries are for the same mineral, e.g. the library as used contains patterns for seven different specimens of kaolinite. Since RC5 was organized by Hillier, this library contains patterns for exactly the same mineral specimens that were used for the preparation of the RC5 samples. Given this, the current testing of the automated algorithm for RC5 samples represents a ‘best case scenario.’ In all other cases, the minerals in the library are not necessarily from the same source as the minerals in the unknown Reynolds Cup samples, though some may be when RC organizers have used widely available materials such as those from the Source Clays Repository of The Clay Minerals Society.

Mineral standards used to create this reference library were also associated with the top-three place finishes in all RC contests except for RC5 (organized by Hillier). The key component tested here was, therefore, the ability of the present algorithm to pre-select the appropriate phases from a large and comprehensive reference library for subsequent automatic quantification.

Automated Full-Pattern Summation

The automated full-pattern summation algorithm and its source code are freely available as the `afps` function in version 1.2.3 of the `powdR` package (Butler and Hillier 2020; Butler and Hillier 2021) for the *R* language and environment for statistical computing (R Core Team 2020), and is hosted on the Comprehensive R Archive Network (<https://cran.r-project.org/package=powdR>). Detailed descriptions of `afps` arguments and their usage are provided in the `powdR` documentation. A flowchart detailing the use of `afps` in the present study is provided in Fig. 1, and relevant arguments summarized in Table 1. More detailed descriptions of the key steps outlined in Fig. 1 are provided in subsequent sections.

Step 1: Sample alignment. Previous in-house experience with full-pattern summation has highlighted the importance of aligning the sample diffractogram along the 2θ axis to that of a calibrated pattern in order to correct for common experimental aberrations associated with the collection of XRPD data (Butler et al. 2019). The discrete nature of XRPD peaks means that seemingly small misalignments can have particularly detrimental effects on data analysis along with the accuracy of phase identification and quantification.

The automated alignment of a sample via the `afps` algorithm used here requires selection of a phase present within it to use as an internal standard (`std` argument; Table 1). For RC1 and RC2, samples were prepared with corundum as the internal standard which was, therefore, used for this alignment. For RC3–RC9, samples were prepared for XRPD analysis without an internal standard, and hence the ‘internal standard’ for alignment was chosen simply as a component of the mineral mixture with sharp, well characterized diffraction features for use as internal *d*-spacing standard. The designated ‘internal standard’ for each sample was then used by `afps` to align the diffractogram along the 2θ axis and hence correct for common experimental aberrations such as sample displacement.

Alignment of the sample to the chosen standard is achieved by maximizing the Pearson correlation via one-dimensional optimization (Brent 1971) within a fixed limit of positive and negative 2θ shifts defined by the `align` argument (Table 1).

With respect to many geologic and environmental samples, the omnipresent mineral quartz can act as a suitable internal standard for the large majority of cases. In the absence of quartz, any well characterized non-clay mineral with few overlapping peaks may be suitable (e.g. dolomite, calcite, anhydrite) providing it is present within the sample(s) being considered and that any solid solutions are represented appropriately by a standard in the reference library. For the present dataset, visual inspection of each of the samples identified that quartz reference patterns would be suitable internal standards in 18 cases. In the remaining three cases where a suitable quartz signal was not observable, internal standards of fluorite (RC5-2), anhydrite (RC7-1), and dolomite (RC9-3) were selected. For all samples presented here, the `align` argument was set to $0.1^\circ 2\theta$.

Step 2: Phase selection with non-negative least squares.

For high-throughput datasets that may display substantial mineralogical variation, a comprehensive reference library is necessary that can cover most, if not all, of the non-clay, clay, and amorphous phases that may exist within the sample set. In such cases it is reasonable to expect that reference libraries containing >100 reference patterns would be required, in which case it becomes impractical to optimize so many variables at once – both in terms of accuracy and time. For this reason, phase selection on a sample-by-sample basis is a key component of automated full-pattern summation.

The `afps` algorithm applied here uses non-negative least squares (NNLS) to identify quickly the phases that can be removed from the reference library. Functionality for NNLS in *R* is provided by the NNLS package (Mullen and van Stokkum 2012), which is based on the *FORTRAN* code of Lawson and Hanson (1995). Application of NNLS facilitates rapid identification of phases in the reference library that probably exhibit no contribution to the observed pattern via derivation of coefficients equal to zero, which are thus omitted from the process (Fig. 1).

Step 3: Minimization of an objective function. As outlined by Chipera and Bish (2002), a range of functions can be minimized for full-pattern summation. Choosing an appropriate function for minimization is key to accurate quantitative analysis via this approach. For mixtures containing clay minerals, non-clay minerals, and amorphous phases, past experience at the James Hutton Institute has shown that the minimization of R_{wp} (Bish and Post 1989), defined as:

$$R_{wp} = \left(\frac{\sum [I_m^{-1} \times (I_m - I_c)^2]}{\sum [I_m^{-1} \times I_m^2]} \right)^{\frac{1}{2}} \quad (1)$$

often results in the most accurate quantitative results (I_m and I_c are vectors of measured and calculated intensities, respectively). Indeed, the R_{wp} statistic is noted typically as one of several

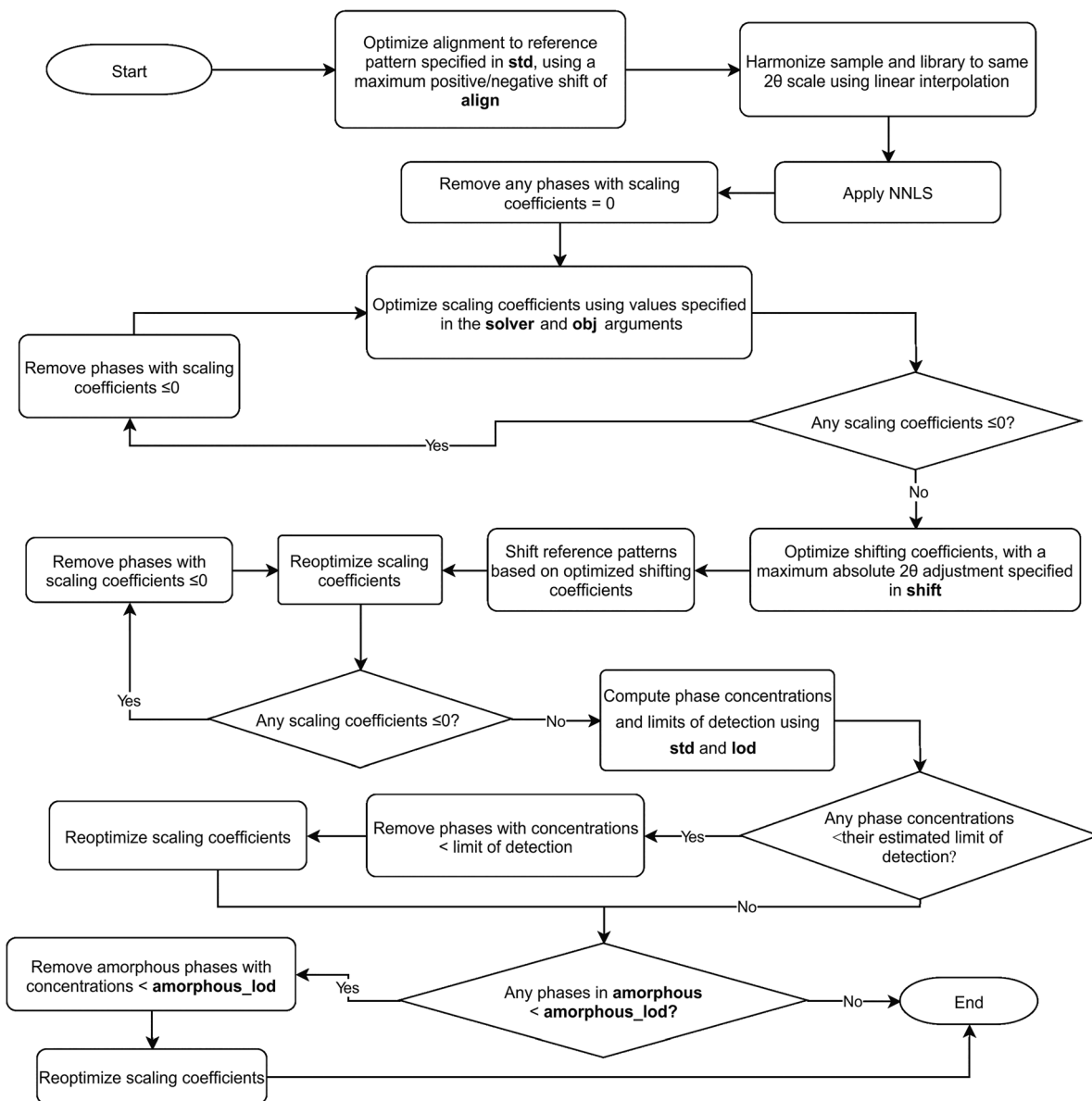


Fig. 1. Flowchart detailing the stages implemented within the *afps* algorithm as applied here. Arguments of *afps* are represented in bold, with further details on their definitions and values provided in Table 1. NNLS = non-negative least squares

performance parameters in Rietveld refinements (Toby 2006), and by weighting the count intensities via the I_m^{-1} terms, results in calculated patterns that prioritize the fitting of regions near to the tails of peaks (Bish and Post 1989). This attribute has beneficial effects when handling the diffuse diffraction signal of poorly ordered and amorphous phases that are encountered commonly in RC samples. Hence, the R_{wp} was used as the objective function for all RC samples presented here, and was minimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Broyden 1970; Fletcher 1970; Goldfarb 1970; Shanno 1970).

The initial optimization of R_{wp} in the *afps* algorithm is applied across all reference patterns that remain after NNLS (Fig.

1). The BFGS optimization routine does not constrain coefficients to positive values; thus any reference patterns that have negative coefficients after optimization are removed from the process, and R_{wp} re-optimized until no negative values remain (Fig. 1).

Step 4: Shifting of reference library patterns. Further to linear alignment of the sample pattern to a reference pattern (Step 1), small additional 2θ shifts applied to each reference pattern in the fitting process of the *XRPDBULK* program (Hillier 2015, 2018) have been found to yield more accurate results. Hence, a second alignment step implemented in the *afps* algorithm seeks to apply small 2θ corrections to the reference patterns relative to the sample pattern to account for

Table 1. Adjustable arguments for the `afps` algorithm (Fig. 1) applied here along with associated descriptions and values used. Further arguments not relevant to the use of `afps` presented here are described in the `powdR` package documentation available at <https://cran.r-project.org/package=powdR>

Argument name	Description	Value(s) used
<code>std</code>	A unique identifier of the reference pattern to use as the internal standard. This internal standard needs to be a phase present in the sample, and is used for sample alignment and computation of the limits of detection for clay and non-clay phases.	Quartz for all RC samples except RC1 and RC2 (corundum), RC5-2 (fluorite), RC7-1 (anhydrite), and RC9-3 (dolomite).
<code>align</code>	A numeric parameter defining the maximum (positive and negative) 2θ shifts applied during alignment of the sample relative to the reference pattern of the internal standard specified in the <code>std</code> argument.	$0.1^\circ 2\theta$.
<code>shift</code>	A numeric parameter defining the maximum (positive and negative) 2θ shifts that can be applied to each reference pattern after optimization of its shifting coefficient relative to the sample.	$0.5^\circ 2\theta$.
<code>lod</code>	A numeric parameter (units of wt.%) defining the limit of detection of the internal standard specified in the <code>std</code> argument, from which all other limits of detection for clay and non-clay phases are estimated via Eq 3.	0.15% assumed for quartz (RIR = 5.68), from which LODs were computed for corundum (RIR = 1), fluorite (RIR = 4.9), anhydrite (RIR = 3.05), and dolomite (RIR = 2.31) ‘internal standards’ via Eq 3.
<code>amorphous</code>	A string of unique identifiers used to define which phases in the reference library should be treated as amorphous.	Allophane, ferrihydrite, glass, obsidian, opal-CT, opal-A, aluminosilicate gel, organic matter, and graphite.
<code>amorphous_lod</code>	A numeric parameter (units of wt.%) below which any phases specified in the <code>amorphous</code> argument are considered below the limit of detection and hence excluded.	2%.

additional small variations such as uncorrected sample displacement errors that may be present in the reference library. More specifically, after an initial optimization of the scaling coefficients, the objective function (Eqn 1) is again minimized by optimizing a shifting coefficient for each remaining reference pattern, which specifies its positive or negative adjustment along the 2θ axis. During the optimization of the shifting coefficients, the scaling coefficients are fixed and reference patterns are maintained on the same 2θ axis interval using cubic spline interpolation. If the absolute value of any optimized shifting coefficient exceeds the value specified in the `shift` argument, its shifting coefficient is reset to zero. Reference patterns are then shifted by the derived shifting coefficients, with cubic spline interpolation again used to ensure that they all remain on the original 2θ axis interval, and the scaling coefficients re-optimized (Step 3). In all cases presented here, the `shift` argument was set to $0.5^\circ 2\theta$.

Step 5: Quantification and limit of detection estimation.

At this stage, appropriate phases were assumed to have been selected from the reference library and the fitted pattern was assumed to be reasonable, from which a reasonably accurate estimation of phase concentrations can be obtained using the RIRs. Because all RC samples presented here were prepared without an internal standard, all detectable phases within the mixture were assumed to be identifiable and that their concentrations summed to 100 wt.%. As such, phase concentrations (X) were computed by:

$$X = \frac{s/\text{RIR}}{\sum(s/\text{RIR})} \times 100 \quad (2)$$

where s and RIR denote vectors of the scaling coefficients (i.e. the parameters derived from NNLS and optimization of R_{wp}) and RIRs covering all remaining phases, respectively.

At this point in the process, some phases, estimated to have very small concentrations, may, inevitably, have been selected in error. Whilst all phases below a defined limit (e.g. 0.1%) could be simply excluded, such an approach would not account for the way in which different phases diffract X-rays with different power (reflected in the RIRs) and, hence, have different limits of detection (LOD) (Hillier 2003). For example, a strong diffractor such as quartz, with a RIR (relative to corundum) of ~ 5.7 , would have a smaller LOD than a weak diffractor such as muscovite (RIR ≈ 0.5). Thus the `afps` algorithm uses the RIRs to derive sensible estimations of LODs for all remaining phases via:

$$\text{LOD} = \text{LOD}_{\text{std}} \times \frac{\text{RIR}_{\text{std}}}{\text{RIR}} \quad (3)$$

where LOD_{std} is the LOD of the internal standard (defined by the `lod` and `std` arguments of the `afps` algorithm; Table 1), RIR_{std} is the RIR of the internal standard, and RIR is a vector of RIRs for all remaining phases. Upon calculating the LODs, all clay and non-clay phases below their respective LOD are removed from the process. For all RC samples presented here,

the `lod` argument was based on the assumption that the LOD of quartz (RIR = 5.68) was 0.15%, from which the LODs of other internal standards used (fluorite, RIR = 4.9; anhydrite, RIR = 3.0; and dolomite, RIR = 2.3) were estimated via Eq 3. Note that the actual value for the LOD of any phase used as reference can be calculated if a sample is spiked with a known weight fraction of another phase as outlined by Hillier (2003), but the simplified approach presented here is based on an arbitrary but realistic LOD for quartz.

Amorphous phases need to be treated slightly differently in the `afps` algorithm to account for the way in which their diffusely scattered signal can be difficult to detect in XRPD data, deeming the approach of Eq 3 as inappropriate. For this reason the amorphous phases defined by the `amorphous` argument are retained unless their estimated concentrations are lower than the value specified in the `amorphous_lod` argument (Table 1). For all RC samples presented here, nine phases in the reference library were defined in the `amorphous` argument (allophane, ferrihydrite, glass, obsidian, opal-CT, opal-A, aluminosilicate gel, organic matter, and graphite), and the `amorphous_lod` argument set to 2%.

After omission of phases based on LODs, a final re-optimization of R_{wp} was applied until no negative parameters remained (Fig. 1). At this point the fitting process was considered complete, and final concentrations computed via Eq 2 in units of wt.%.

Computation

Application of the `afps` algorithm to each of the 27 RC samples was carried out in `powdR` version 1.2.3 (Butler and Hillier 2020; Butler and Hillier 2021) on a Windows 10 machine equipped with an Intel® Core™ i7-6600U CPU @ 2.60 GHz. Computation time averaged ~1 h per sample. For faster computation time of a batch of samples, the `afps` algorithm can be used in combination with the `foreach` and `doParallel R` packages for parallel processing across multiple cores (Microsoft and Weston 2017, 2018).

With the exception of the specification of different ‘internal standards’ (used for alignment in this case) and limits of detection, the same parameters were used for all arguments of the `afps` algorithm for all 27 samples (Table 1). Further, no visual inspection or amendments to the output were carried out. Whilst it is always recommended to inspect visually outputs from full-pattern summation, the aim of the present study was to test whether an entirely automated approach to quantifying mineralogically diverse samples could yield accurate results.

Pre-requisites for Automated Full-Pattern Summation

Although running the `afps` algorithm is relatively simple, accurate quantification from it is ultimately facilitated by the combination of reproducible diffraction data and a comprehensive reference library that can account for all of the phases present within a given dataset. The quality of the XRPD data, both for the sample and the reference library, relates particularly to the potential effects of particle statistics and preferred orientation. Preferred orientation of minerals with prominent cleavage planes can be eliminated during sample preparation

using techniques such as spray drying (Hillier 1999; Kleeberg et al. 2008), and the reproducibility of diffraction data as a result is a major advantage to methods using full-pattern summation of prior measured standards as presented here.

Although not tested in the present study, the size of the library can prove influential on the accuracy of the final output and the speed with which it can be obtained. More specifically, whilst larger libraries (hundreds of reference patterns) promote the selection of appropriate phases, it would be recommended that they are customized for a given dataset based on the minerals that are likely to be encountered. The presence of additional phases that would not be encountered within the samples simply acts to slow down the computation and/or increase the chance of misidentification. Aside from misidentifications, in some cases visual inspection of the output may identify that a sample contains a mineral that is not present within the library. This would require the user to source a suitable reference mineral and add it to the library using protocols outlined above. In either case the incidence of misidentified and/or unidentified phases can be assessed quickly via visual inspection of the outputs and residuals, which would always be recommended.

Reynolds Cup Accuracy Determination

The accuracy of the automated algorithm for all RC samples was assessed based on absolute bias (in wt.%) for all phases. In order to allow direct comparison with previous RC contestants (i.e. to derive comparative contest placings), these absolute bias values were summed to produce an overall score using the procedures applied in the judging of each previous contest. For RC1–RC3 placings were determined based on the sum of absolute bias for all known phases. For RC4–RC9 placings were determined based on the sum of bias for all known phases plus the summed weight percentages of any misidentified phases (i.e. phases not present within the sample). The mineralogical groupings used for each RC contest are provided in Tables S1–S9 in the Supplementary Material.

RESULTS

Phase Selection and Overall Accuracy

As outlined above and illustrated in Fig. 1, the `afps` algorithm involves several steps that reduce the full reference library to an appropriate subset for each sample. These include the application of NNLS, removal of negative coefficients during optimization, and exclusion of phases estimated to be below the limit of detection. The number of phases remaining at various points in the `afps` process for the 27 RC samples is summarized in Fig. 2. From an initial library containing 201 reference patterns, application of NNLS and the associated exclusion of any phase with a parameter equal to zero (Fig. 1) yielded a reduced library containing a mean of 52 patterns. Subsequent optimization and removal of negative coefficients resulted in the removal of another seven patterns, on average. Shifting and reoptimizing the scaling coefficients until no negatives remained resulted in removal of a further three patterns, on average. Estimation of LODs and the associated

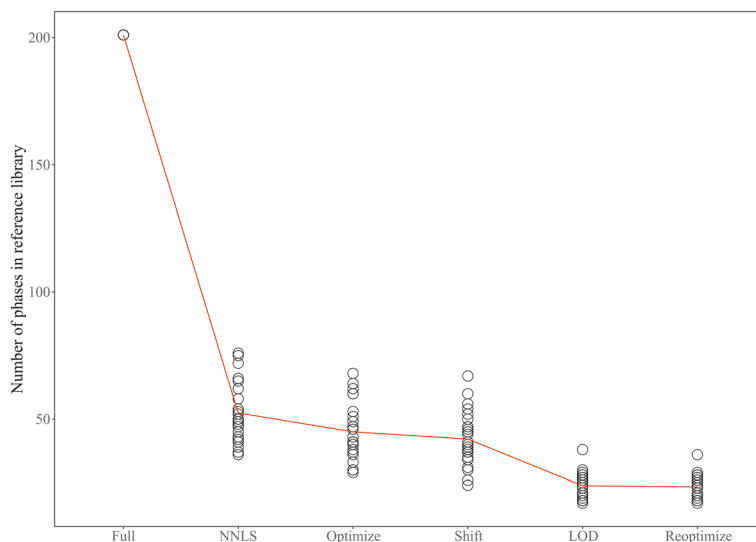


Fig. 2. Phase selection at various stages of the afps algorithm for all 27 RC samples. Red line represents the mean at each stage

removal of phases below their respective LOD (including any amorphous phases estimated to be below the `amorphous_lod` argument; Table 1) resulted in a further 43% reduction to the library, with a mean of 24 remaining phases. Lastly, re-optimization and the associated removal of phases with negative coefficients resulted in a final selection of 23 reference patterns, on average.

The resulting final phase selections across the 27 RC samples presented here covers 151 reference patterns from the full library, representative of 61 correctly identified (i.e. present within the sample and the afps output) clay/non-clay/amorphous groups (Table 2). This large number of reference patterns selected across the dataset illustrates the mineralogical diversity of the Reynolds Cup samples, whilst the relatively small mean number of reference patterns in the final selection from the afps algorithm indicates selectivity. The appropriateness of the final selection for each sample determines ultimately the quality of the fits and, therefore, the accuracy of the resulting quantification.

All correctly identified, misidentified (i.e. not present within the sample but present within the afps output), and unidentified (i.e. present within the samples but not present within the afps output) phases encountered across the 27 RC samples are summarized in Fig. 3. Misidentified phases (Table 3) are scattered on the vertical at the intercept $x = 0$, whilst unidentified phases (Table 4) are scattered on the horizontal at the intercept $y = 0$. In summarizing all data displayed in Fig. 3, the mean absolute bias for non-clay, clay, and amorphous phases equates to 0.57% ($n = 275$), 2.37% ($n = 120$), and 4.43% ($n = 14$), respectively. Further exploration of these results is provided below according to the correctly identified, misidentified, and unidentified groupings.

By comparing the overall accuracy for each contest to the anonymous results of all participants, the accuracy of the afps outputs presented here would have been sufficient for the following placings: RC1 = 2nd/15; RC2 = 2nd/35; RC3 = 1st/39; RC4 = 2nd/44; RC5 = 1st/64; RC6 = 3rd/63; RC7 = 3rd/

68; RC8 = 1st/70; RC9 = 2nd/74 (Table 5). Given that the world's leading mineralogists and laboratories are amongst the participants of each Reynolds Cup (based on the published top named finishers; www.clays.org/Reynolds.html), the competitive accuracy of the afps algorithm in combination with the small values of absolute bias together indicate that the approach can derive accurate quantitative mineralogical analysis from a single random powder XRPD measurement when provided with a suitable reference library.

Correctly Identified Phases

The absolute bias of all correctly identified phases is summarized in Table 2. The mean absolute bias of non-clay constituents was 0.55% across the 0.20% to 45.70% known concentration range ($n = 222$). In contrast, the mean absolute bias of the correctly identified clay constituents was 2.18% ($n = 102$) across the 1.00–40.20% known concentration range, with that of the amorphous constituents being even higher at 3.74% across the 6.90–18.27% known concentration range ($n = 6$).

Misidentified Phases

Across the 27 RC samples tested, the sum of misidentified phases averaged 3.13% per sample. All misidentified phases are presented in Tables S1–S9, and are summarized in Table 3. The majority of misidentified phases were non-clay minerals, with a mean misidentified concentration of 0.68% ($n = 24$). The number of misidentified clay minerals across the 27 samples was smaller than for non-clay minerals, but with a notably larger mean misidentified concentration of 4.33% ($n = 11$). Three cases of misidentified amorphous phases were found, all in RC1 samples, with a mean misidentified concentration of 6.76%.

Unidentified Phases

Across the 27 RC samples, there was a total of 41 cases where phases present within the samples remained unidentified in the outputs from the afps algorithm, with the sum of unidentified phases averaging 2.77% per sample. All

Table 2. Summary of the accuracy for correctly identified phases present in the RC samples

Phase	<i>n</i>	Min.–Max. (known,%)	Mean absolute bias (%)
<u>Non-clay</u>			
Alunite	1	10.7	0.0
Amphibole	2	3.9–4.0	0.6
Anatase	12	0.2–3.3	0.1
Anhydrite	3	5.2–22.7	0.8
Apatite	4	1.1–4.0	0.3
Aragonite	3	3.0–5.0	0.5
Barite	6	2.0–5.8	0.2
Bassanite	1	1.1	0.2
Bimessite	1	2.0	0.9
Boehmite	2	7.9–8.3	0.8
Calcite	16	2.1–18.6	0.3
Celestine	1	1.3	0.1
Cristobalite	1	3.9	2.6
Dolomite/ankerite	13	2.0–21.9	0.3
Epidote	1	4.0	0.7
Fluorite	4	1.9–3.9	0.6
Gibbsite	4	2.0–10.7	0.5
Goethite	7	1.6–7.8	1.5
Gypsum	6	1.3–21.8	1.1
Halite	7	1.5–5.0	0.8
Hematite	10	1.0–9.4	0.4
Huntite	1	5.0	0.1
Ilmenite	1	3.5	0.3
K-feldspar	17	1.8–11.4	0.6
Magnesite	6	2.6–5.7	0.3
Magnetite	6	0.2–6.4	1.2
Olivine	1	7.0	1.3
Opal CT	2	5.0–35.6	1.8
Plagioclase	19	2.0–10.7	0.3
Pyrite	10	0.8–3.0	0.3
Pyroxene	1	6.7	0.2
Quartz	26	2.1–45.7	0.8
Rutile	6	1.0–3.8	0.3
Siderite	8	1.0–6.6	0.6
Sodalite	1	2.5	0.5
Spinel	1	5.1	0.8
Sulphur	1	2.6	1.8
Topaz	1	10.8	0.2
Tourmaline	1	1.8	0.4
Vanadinite	1	0.8	0.2
Witherite	1	6.4	0.4
Zeolite	2	2.2–10.2	0.8
Zircon	4	0.2–2.9	0.1
<u>Clay</u>			
Chlorite (trioct.)	17	3.0–14.2	1.0
Dickite	1	4.0	0.6
Halloysite	4	7.2–22.6	1.4

Table 2. (continued)

Phase	<i>n</i>	Min.–Max. (known,%)	Mean absolute bias (%)
Kaolin group*	7	8.0–25.3	0.7
Kaolinite	12	6.6–22.5	1.1
Mica (dioct.)	9	2.4–22.1	3.3
Mica (trioct.)	5	1.2–19.7	1.6
Mixed-layer (dioct.)	6	4.6–14.3	5.2
Palygorskite	2	7.5–15.2	1.3
Sepiolite	2	10.0–18.7	1.6
Serpentine	2	7.7–22.8	0.5
Smectite (dioct.)	8	2.4–40.2	4.4
Smectite (trioct.)	4	1.0–27.0	5.0
Talc	4	5.0–11.8	0.4
Total dioct. 2:1 mica or smectite*	12	6.1–33.0	2.9
Total trioct. 2:1 mica or smectite*	3	6.1–28.9	1.4
Vermiculite (trioct.)	3	4.3–11.2	3.0
Amorphous			
Amorphous group	6	6.9–18.3	3.7

*denotes clay mineral groupings that were used only in RC1–RC4.

unidentified phases are presented in Tables S1–S9, and are summarized in Table 4. As found for misidentifications, the majority of unidentified phases were non-clay minerals ($n = 29$), with a mean known concentration of 0.63%. Three of these unidentified non-clay minerals were not present within the reference library (cryolite, nahcolite, and vivianite; Table 4). In contrast to the number of unidentified non-clay minerals, only seven cases of unidentified clay minerals were identified across the 27 samples, with a mean known concentration of 2.18%. Further to non-clay and clay minerals, five cases emerged where

amorphous phases within the sample were not identified by afps, with a mean known concentration of 4.37%.

DISCUSSION

Reynolds Cup samples are prepared to be challenging mixtures to quantify – mainly due to the diversity of clay minerals that they may contain along with a relatively detailed clay-mineral classification system that is applied to the results (Raven and Self 2017). As would be expected, the accuracy of

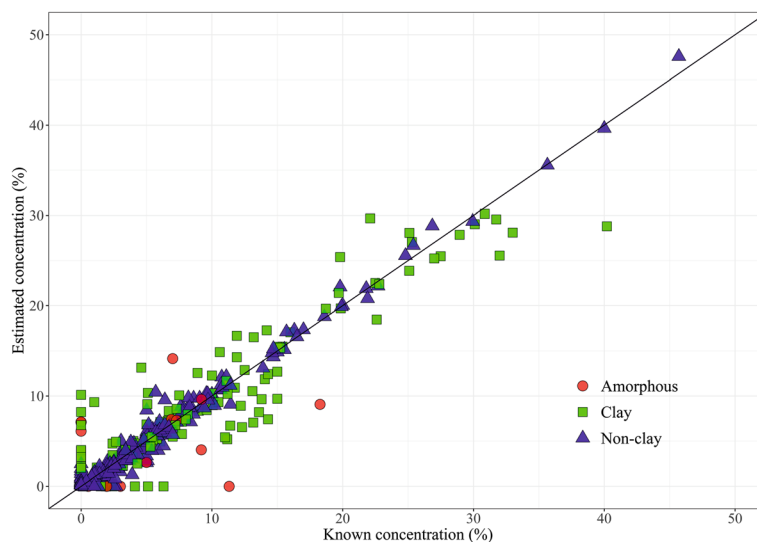


Fig. 3. Known concentrations of all phases (classified according to amorphous, clay, and non-clay groupings) from all 27 RC samples plotted against the estimated concentrations from the afps algorithm

Table 3. Summary of the misidentified phases present in the outputs from a f p s

Phase	<i>n</i>	Mean misidentified concentration (%)
<u>Non-clay</u>		
Anhydrite	2	0.4
Birnessite	1	0.4
Corundum	2	1.2
Erionite	1	0.4
Garnet	2	0.5
Huntite	1	0.9
Ilmenite	1	0.5
Magnesite	3	1.5
Magnetite	1	0.6
Plagioclase	1	2.5
Siderite	2	0.5
Sodalite	1	0.3
Sphalerite	4	0.2
Sylvite	1	0.1
Witherite	1	0.2
<u>Clay</u>		
Mixed-layer (trioct.)	2	2.8
Palygorskite	2	2.5
Sepiolite	3	6.0
Serpentine	1	2.2
Smectite (dioct.)	1	1.9
Smectite (trioct.)	2	7.5
<u>Amorphous</u>		
Amorphous group	3	6.8

the a f p s outputs shows a notable difference between the non-clay and clay-mineral groupings (Fig. 3), with the mean absolute bias of all non-clay minerals being ~4.2 times less than that of clay minerals.

The inaccuracy of clay-mineral quantification relative to non-clay minerals may reflect, in part, the difficulty in identifying correctly clay minerals from a single bulk XRPD measurement, as the a f p s algorithm seeks to do, due to the way in which many clay minerals have many similar features in their bulk diffraction patterns. The even greater inaccuracy with respect to amorphous phases would also be expected given the often ambiguous ‘background’ signal associated with phases of this type which often lack any coherent Bragg diffraction. In relation to clay-mineral and amorphous-phase identification, the most successful participants of the Reynolds Cup used oriented specimens and successive treatments of these to identify precisely the clay-mineral components in a sample, whilst total elemental analysis of bulk samples was also often used to identify more accurately the nature of X-ray amorphous phases (Omotoso et al. 2006; Raven and Self

Table 4. Summary of the accuracy for unidentified phases present in the RC samples but not in the respective a f p s output

Phase	<i>n</i>	Mean known concentration (%)
<u>Non-clay</u>		
Amphibole	2	0.8
Anatase	4	0.1
Arcanite	1	2.7
Calcite	3	0.1
Cassiterite	1	0.3
Cryolite*	1	0.5
Dolomite/ankerite	1	0.3
Goethite	2	0.1
Ilmenite	2	0.4
K-feldspar	1	2.1
Magnesite	1	0.5
Nahcolite*	1	2.5
Plagioclase	1	0.5
Rutile	1	0.1
Stilbite	1	1.3
Sylvite	1	0.3
Tourmaline	4	1.0
Vivianite*	1	0.3
<u>Clay</u>		
Chlorite	2	1.7
Mica (dioct.)	1	0.1
Mixed-layer (dioct.)	1	2.4
Mixed-layer (trioct.)	1	4.1
Smectite (dioct.)	1	5.1
Talc	1	0.1
<u>Amorphous</u>		
Amorphous group	5	4.4

*denotes phases that were not present within the reference library.

2017). The finding that the a f p s algorithm produces highly competitive results compared to other RC participants, even without these additional analyses, is particularly promising for the future of automated quantitative phase analysis by XRPD.

In addition to the bias associated with correctly identified phases, an important aspect of the results is the presence of misidentified phases (Tables 3, 5, and S1–S9), which can easily compromise the accuracy of quantitative analysis. Misidentified non-clay minerals were present in 21 of the 27 samples (Table 5), with an overall mean of 1.37% per sample. Misidentified clay minerals were present in 11 of the 27 samples (Table 5), with an overall mean of 1.76% per sample. The slightly higher concentrations for misidentified clay minerals compared to non-clay minerals again reflects the challenging nature of their identification from bulk XRPD data alone.

Table 5 Accuracy of automated full-pattern summation relative to known weights of the samples, summarized as the sum of bias for non-clay and clay phases along with misidentified non-clay and clay minerals. Further details for each RC contest are provided in Tables S1–S9

Sample	Absolute bias		Misidentified		Total bias	Total misidentified	Placing/No. of participants
	Non-clay	Clay	Non-clay	Clay			
RC1-1	3.75	8.93	7.08	—	12.68	7.08	2 nd /15
RC1-2	3.37	6.59	7.12	—	9.96	7.12	
RC1-3	6.65	5.20	7.37	—	11.85	7.37	
RC2-1	1.20	2.96	0.47	—	4.16	0.47	2 nd /35
RC2-2	2.46	1.73	—	—	4.19	—	
RC2-3	3.51	4.19	0.61	—	7.64	0.61	
RC3-1	4.14	5.04	1.32	3.89	9.19	5.21	1 st /37
RC3-2	3.83	6.73	1.70	6.77	10.56	8.47	
RC3-3	6.07	3.32	0.26	—	9.40	0.26	
RC4-1	18.44	2.29	—	4.00	20.73	4.00	2 nd /43
RC4-2	6.15	4.55	—	—	10.70	—	
RC4-3	5.06	2.93	1.67	2.97	7.99	4.64	
RC5-1	9.87	19.43	1.84	—	29.30	1.84	1 st /63
RC5-2	3.59	14.14	—	1.85	17.74	1.85	
RC5-3	14.34	14.30	2.48	2.37	28.64	4.85	
RC6-1	5.12	15.62	1.96	—	20.74	1.96	3 rd /62
RC6-2	4.61	8.69	0.46	8.22	13.30	8.68	
RC6-3	21.07	9.07	0.13	2.15	30.14	2.28	
RC7-1	6.08	7.16	0.19	—	13.24	0.19	3 rd /67
RC7-2	7.64	23.77	—	10.13	31.41	10.13	
RC7-3	5.50	5.27	0.50	—	10.77	0.50	
RC8-1	4.08	5.78	0.21	—	9.86	0.21	1 st /69
RC8-2	4.99	11.00	0.44	—	15.99	0.44	
RC8-3	3.00	4.75	0.27	—	7.75	0.27	
RC9-1	14.94	20.56	0.68	—	35.50	0.68	2 nd /73
RC9-2	4.61	15.17	—	2.02	19.78	2.02	
RC9-3	7.25	8.07	0.15	3.22	15.32	3.37	

The presence of ~7% misidentified amorphous material in each of the 3 RC1 samples (Table S1) highlights the care needed when quantifying amorphous phases from bulk XRPD data alone, particularly since the reason for this consistent misidentification in RC1 samples remains unclear. Of the clay-mineral misidentifications, two stand out as being particularly high, relating to samples RC6-2 and RC7-2 (Tables 5, S6, and S7). The RC6-2 misidentification relates to selection of a trioctahedral smectite reference pattern (8.22%; Table S6), which, based on the true sample composition, is probably a misidentification against the dioctahedral smectite it contains. The RC7-2 misidentification relates to the selection of a sepiolite reference pattern (10.13%; Table S7), again probably instead of the dioctahedral smectite.

When using the *aFPS* algorithm, ultimately a balance must be struck between the number of misidentified phases and the number of unidentified phases, which is controlled largely by the appropriateness of the reference library and value that the user specifies in the *lod* argument (Table 1). With respect to *lod*, setting this parameter to a very small value or zero would promote increased numbers of misidentifications but may decrease the number of unidentified phases. The reverse applies if the *lod* value is excessively high. Thus, in the present case, the approximate balance of both misidentified and unidentified phases (3.5% and 2.6% per sample on average, respectively) may represent a reasonable compromise, based on the assumption that the LOD for quartz in all samples would be 0.15% (Table 1). Further reduction of the incidence of misidentified and unidentified phases

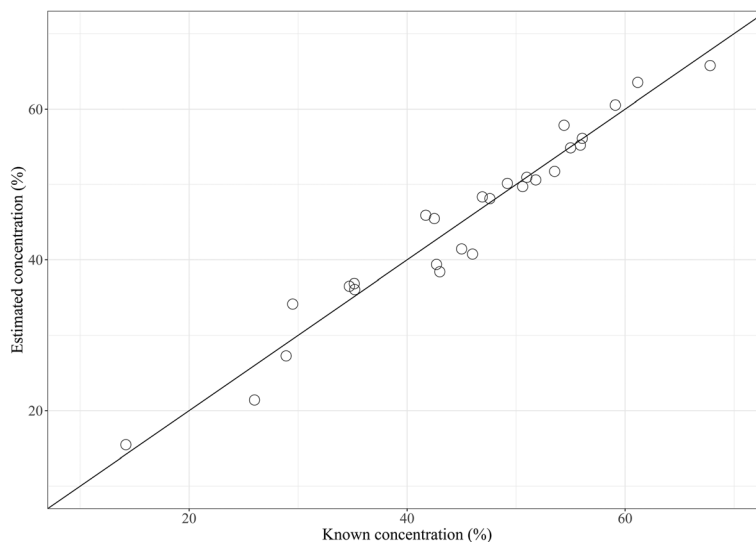


Fig. 4. Known concentration of total clay minerals for all 27 RC samples plotted against that estimated from the *a*fps algorithm

would almost certainly be achieved by visual inspection of the results, which, as outlined above, was not undertaken in this study.

As previously mentioned, the clay-mineral classification of RC samples is relatively detailed, creating a challenge for the selection of appropriate reference patterns via an automated approach. If instead, a very coarse description of clay is applied, i.e. total clay minerals, the accuracy of the results can be improved (Fig. 4). Comparing the total clay mineral concentrations estimated by *a*fps (including misidentified clay minerals) to the known total concentrations, the mean absolute bias across the 27 samples reduces to 2.13% in the known range of 14.20% to 67.80% total clay. It is, therefore, worth emphasizing that if a completely automated approach is going to be applied, the user may wish to adjust the clay-mineral classification system to best reflect the limitations of the method - with coarser descriptions providing greater accuracy in terms of total clay or related clay groupings at the compromise of detail.

Viewed as a whole, the accuracy of the *a*fps algorithm presented here is promising, particularly as the results are derived from a single bulk XRPD measurement. This relative simplicity is important in the case of high-throughput datasets because additional mineralogical (i.e. clay fractions) and geochemical (e.g. total elemental) analyses are undoubtedly a time-consuming and expensive undertaking. Whilst one would not expect an automated approach to exceed the accuracy of that achieved from multiple forms of analyses combined with expert input, the present data illustrate that accurate results can still be obtained. For high-throughput cases, some accuracy will inevitably need to be compromised in order to quantify mineral concentrations in hundreds or thousands of samples. Expert input is still no doubt necessary in such high-throughput cases, and although not included within this investigation (outputs from the *a*fps algorithm were not inspected or altered in any way), would probably act to enhance the accuracy of automated approaches. The most effective form of expert input is visual inspection of fitted patterns and their residuals relative to the original measurement (Butler and

Hillier, 2021). Such inspection allows a trained user to identify phases that are missing from the analysis, or those that should be removed.

Applicability to Natural Samples

Whilst RC samples are prepared to represent naturally occurring clay-bearing mixtures, the challenging nature of sourcing pure clay mineral standards increases the likelihood that the standards in the reference library match exactly those used to prepare the samples – resulting in artificially enhanced accuracy compared to the quantification of natural samples. To assess for the occurrence of exact matching in the present study, available information on how RC samples were prepared was collated and contest organisers contacted where sufficient information was not available. Based on this information, the majority (50–92%) of reference patterns in the library supplied to the *a*fps algorithm were not used to prepare the RC samples for each contest (Table 6), with the exception of RC5, which was organized by Stephen Hillier.

Table 6. Summary of the clay mineral standards used to prepare the samples for each RC contest that are not present as reference patterns in the library supplied to the *a*fps algorithm in this study

Contest	<i>n</i>	%
RC1	3/4	75%
RC2	5/7	71%
RC3	5/10	50%
RC4	10/13	77%
RC5	1/13	8%
RC6	6/12	50%
RC7	8/12	67%
RC8	11/13	85%
RC9	12/15	92%

The general absence of exact matching between reference library standards and RC sample constituents presented here indicates, therefore, that the approach should be suitable for natural clay-bearing samples if appropriate mineral standards can be sourced.

Natural clay-bearing mixtures may contain more complex clay minerals than those in RC samples, especially in relation to interstratified clay minerals that are more difficult to isolate as pure phases for use in round-robin contests. Furthermore, phases with a broad solid solution series may make it difficult to cover the whole range of each series without a large library of standards specially designed to do so. That said, neither of these issues is unsurmountable. The only way to gauge the likely accuracy of any form of quantitative mineralogical analysis on natural samples is indirectly, however, e.g. by comparing the measured bulk chemical composition to a bulk chemical composition generated from the mineralogical analysis by assuming, or obtaining, chemical compositions of the respective minerals quantified in any given sample. This approach was used, for example, by Casetou-Gustafson et al. (2018) for soils quantified by the FPS approach using the same standard pattern library. Future testing of the `afps` algorithm will seek to assess its accuracy when applied to natural samples, but such assessments can never be as direct as those obtained from application to round robin samples where accuracy can be assessed precisely by comparison to the known mineralogical compositions.

CONCLUSIONS

An open source, automated, full-pattern summation algorithm has been shown to quantify accurately mineral concentrations in complex clay-bearing mixtures from the previous nine Reynolds Cup contests. The accuracy of the automated results would have been sufficient for the top three placings in all RC contests tested (RC1 = 2nd, RC2 = 2nd, RC3 = 1st; RC4 = 2nd, RC5 = 1st; RC6 = 3rd, RC7 = 3rd, RC8 = 1st; RC9 = 2nd). Non-clay minerals were quantified with a mean absolute bias of 0.57%, whilst that of the clay minerals was higher at 2.37%, and for amorphous phases was 4.43%. In some cases the incorrect identification of clay minerals was a key component of the overall bias; when comparing total clay content, however, the automated algorithm yielded very accurate values, suggesting that careful consideration should be given to the level of clay identification that can be expected of automated approaches based on a single bulk XRPD measurement. The detection and quantification of amorphous phases remains difficult from bulk XRPD data alone, especially when mixed into complex mineral assemblages. Although many 'X-ray amorphous' phases have quite distinctive features in their scattering/diffraction patterns, others can look very alike, and therefore manual inspection of `afps` outputs in combination with auxiliary analysis (e.g. total element analysis) remains beneficial and recommended for enhanced accuracy. The results are ultimately promising, and the proven accuracy justifies the potential for further application to high-

throughput XRPD datasets. Future testing of the algorithm's accuracy on natural samples via the use of total elemental analysis will act to assess further its performance and applicability for high-throughput mineral quantification of soils and sediments.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s42860-020-00105-6>.

ACKNOWLEDGMENTS

This work was supported by a Macaulay Development Trust Fellowship, United Kingdom, Grant No. MDT-50. The support of the Scottish Government's Rural and Environment Science and Analytical Services Division (RESAS) is also gratefully acknowledged. The authors thank the three anonymous reviewers and the Editorial Board for their useful comments which helped to improve this paper.

Funding

Funding sources are as stated in the Acknowledgments.

Compliance with Ethical Statements

Conflict of Interest

The authors declare that they have no conflict of interest.

REFERENCES

- Bergmann, J., Friedel, P., & Kleeberg, R. (1998). BGMN – a new fundamental parameters based Rietveld program for laboratory X-ray sources, its use in quantitative analysis and structure investigations. *CPD Newsletter*, 20.
- Bish, D. & Post, J. (editors) (1989). *Modern Powder Diffraction*. Reviews in Mineralogy, 20 Mineralogical Society of America, Chantilly, Virginia, USA.
- Butler, B. & Hillier, S. (2020). *powdR: Full Pattern Summation of X-Ray Powder Diffraction Data*. R package version 1.2.3. URL: <https://CRAN.R-project.org/package=powdR>
- Butler, B. M. & Hillier, S. (2021). *powdR: An R package for quantitative mineralogy using full pattern summation of X-ray powder diffraction data*. *Computers and Geosciences*, 107, 104662.
- Brent, R. P. (1971). An algorithm with guaranteed convergence for finding a zero of a function. *The Computer Journal*, 14, 422–425.
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms I. General considerations. *IMA Journal of Applied Mathematics*, 6, 76–90.
- Butler, B. M., O'Rourke, S. M., & Hillier, S. (2018). Using rule-based regression models to predict and interpret soil properties from X-ray powder diffraction data. *Geoderma*, 329, 43–53.
- Butler, B. M., Sila, A. M., Shepherd, K. D., Nyambura, M., Gilmore, C. J., Kourkoumelis, N., & Hillier, S. (2019). Pre-treatment of soil X-ray powder diffraction data for cluster analysis. *Geoderma*, 337, 413–424.
- Butler, B. M., Palarea-Albaladejo, J., Shepherd, K. D., Nyambura, K. M., Towett, E. K., Sila, A. M., & Hillier, S. (2020). Mineral–nutrient relationships in African soils assessed using cluster analysis of X-ray powder diffraction patterns and compositional methods. *Geoderma*, 375, 114474.
- Casetou-Gustafson, S., Hillier, S., Akselsson, C., Simonsson, M., Stendahl, J., & Olsson, B. A. (2018). Comparison of measured

- (XRPD) and modeled (A2M) soil mineralogies: A study of some Swedish forest soils in the context of weathering rate predictions. *Geoderma*, 310, 77–88.
- Chipera, S. J., & Bish, D. L. (2002). FULLPAT: A full-pattern quantitative analysis program for X-ray powder diffraction using measured and calculated patterns. *Journal of Applied Crystallography*, 35, 744–749.
- Clark, G. L., & Reynolds, D. H. (1936). Quantitative analysis of mine dusts: an X-ray diffraction method. *Industrial & Engineering Chemistry Analytical Edition*, 8, 36–40.
- Costanzo, P. A., & Guggenheim, S. (2001). Baseline studies of the Clay Minerals Society Source Clays: preface. *Clays and Clay Minerals*, 49, 371–371.
- Doebelin, N., & Kleeberg, R. (2015). Profex: a graphical user interface for the Rietveld refinement program BGMN. *Journal of Applied Crystallography*, 48, 1573–1580.
- Eberl, D. D. (2003). User's guide to ROCKJOCK – A program for determining quantitative mineralogy from powder X-ray diffraction data. Technical report, USGS, Boulder, Colorado, USA.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13, 317–322.
- Gates-Rector, S., & Blanton, T. (2019). The Powder Diffraction File: a quality materials characterization database. *Powder Diffraction*, 34, 352–360.
- Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24, 23–26.
- Hillier, S. (1999). Use of an air brush to spray dry samples for X-ray powder diffraction. *Clay Minerals*, 34, 127–135.
- Hillier, S. (2000). Accurate quantitative analysis of clay and other minerals in sandstones by XRD: comparison of a Rietveld and a reference intensity ratio (RIR) method and the importance of sample preparation. *Clay Minerals*, 35, 291–302.
- Hillier, S. (2003). Quantitative Analysis of Clay and other Minerals in Sandstones by X-Ray Powder Diffraction (XRPD). *Clay Mineral Cements in Sandstones*, 34, 213–251.
- Hillier, S. (2015). X-ray powder diffraction full-pattern summation methods for quantitative analysis of clay bearing samples. In *Euroclay 2015 Programme and Abstracts*, page 174.
- Hillier, S. (2018). Quantitative analysis of clay minerals and poorly ordered phases by prior determined X-ray diffraction full pattern fitting: procedures and prospects. In *9th Mid-European Clay Conference Book*, page 6.
- ICDD (2016). *PDF-4+ 2016 (Database)*. International Center for Diffraction Data, Newtown Square, PA, USA.
- Kleeberg, R., Monecke, T., & Hillier, S. (2008). Preferred orientation of mineral grains in sample mounts for quantitative XRD measurements: How random are powder samples? *Clays and Clay Minerals*, 56, 404–415.
- Lawson, C. L. & Hanson, R. J. (1995). *Solving least squares problems*, volume 15. Siam.
- Microsoft & Weston, S. (2017). *foreach: Provides Foreach Looping Construct for R*. R package version 1.4.4. URL: <https://CRAN.R-project.org/package=foreach>
- Microsoft & Weston, S. (2018). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.14. <https://CRAN.R-project.org/package=doParallel>
- Mullen, K. M. & van Stokkum, I. H. M. (2012). *nls: The Lawson-Hanson algorithm for non-negative least squares (NLS)*. R package version 1.4. <https://CRAN.R-project.org/package=nls>
- Navias, L. (1925). Quantitative determination of the development of mullite in fired clays by an X-ray method. *Journal of the American Ceramic Society*, 8, 296–302.
- Omotoso, O., McCarty, D. K., Hillier, S., & Kleeberg, R. (2006). Some successful approaches to quantitative mineral analysis as revealed by the 3rd Reynolds Cup contest. *Clays and Clay Minerals*, 54, 748–760.
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raven, M. D., & Self, P. G. (2017). Outcomes of 12 years of the Reynolds Cup quantitative minerals analysis round robin. *Clays and Clay Minerals*, 65, 122.
- Rietveld, H. M. (1969). A profile refinement method for nuclear and magnetic structures. *Journal of Applied Crystallography*, 2, 65–71.
- Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, 24, 647–656.
- Smith, D. K., Johnson, G. G., Scheible, A., Wims, A. M., Johnson, J. L., & Ullmann, G. (1987). Quantitative X-ray powder diffraction method using the full diffraction pattern. *Powder Diffraction*, 2, 73–77.
- Toby, B. H. (2006). R factors in Rietveld analysis: How good is good enough? *Powder Diffraction*, 21, 67–70.
- Vogt, C., Lauterjung, J., & Fischer, R. X. (2002). Investigation of the clay fraction (<2 μm) of The Clay Minerals Society reference clays. *Clays and Clay Minerals*, 50, 388–400.
- Woodruff, L. G., Cannon, W. F., Eberl, D. D., Smith, D. B., Kilburn, J. E., Horton, J. D., Garrett, R. G., & Klassen, R. A. (2009). Continental-scale patterns in soil geochemistry and mineralogy: results from two transects across the United States and Canada. *Applied Geochemistry*, 24, 1369–1381.

(Received 17 June 2020; revised 30 October 2020; AE: Peter Ryan)