PyXRD v0.6.7: a free and open source program to quantify disordered phyllosilicates using multispecimen X-ray diffraction profile fitting.

Mathijs Dumon[1*], Eric Van Ranst[1]

[1] Department of Geology and Soil Science (WE13), Ghent University, Krijgslaan 281/S8, B-9000 Ghent, Belgium

* Corresponding author:
Department of Geology and Soil Science (WE13)
Faculty of Sciences, Ghent University, Krijgslaan 281/S8, B-9000 Gent, Belgium
Phone: +32 (0)9 264 46 17
E-mail: mathijs.dumon@ugent.be

Abstract

This paper presents a free and open-source program called PyXRD (short for Python X-ray diffraction) to improve the quantification of complex, poly-phasic mixed-layer phyllosilicate assemblages. The validity of the program was checked by comparing its output with Sybilla v2.2.2, which shares the same mathematical formalism. The novelty of this program is the ab initio incorporation of the multispecimen method, making it possible to share phases and (a selection of) their parameters across multiple specimens. PyXRD thus allows modelling multiple specimens side by side, and this approach speeds up the manual refinement process significantly. To check the hypothesis that this multispecimen set-up – as it effectively reduces the number of parameters and increases the number of observations – can also improve automatic parameter refinements, we calculated X-ray diffraction patterns for four theoretical mineral assemblages. These patterns were then used as input for one refinement employing the multispecimen set-up and one employing the single-pattern set-ups. For all of the assemblages, PyXRD was able to reproduce or approximate the input parameters with the multispecimen approach. Diverging solutions only occurred in single-pattern set-ups which do not contain enough information to discern all minerals present (e.g. patterns of heated samples). Assuming a correct qualitative interpretation was made and a single pattern exists in which all phases are sufficiently discernible, the obtained results indicate a good quantification can often be obtained with just that pattern. However, these results from theoretical experiments cannot automatically be extrapolated to all real-life experiments. In any case, PyXRD has proven to be useful when X-ray diffraction patterns are modelled for complex mineral assemblages containing mixed-layer phyllosilicates with a multispecimen approach.

Keywords: quantitative, XRD, clay mineralogy, PyXRD, multispecimen

1. Introduction

Clay minerals (i.e. phyllosilicates) are among the most difficult minerals to study in detail due to their inherent chemical and structural variability (Środoń, 2006; Velde and Meunier, 2008; Hubert et al., 2012). Nonetheless, these minerals are one of the most abundant constituents of the Earth's upper crust, and have an important influence on various physical (e.g. plasticity, shear strength, porosity) and chemical (e.g. buffering and exchange capacities, pH, electrical conductivity) properties (Agbenin and Tiessen, 1995; Vernik and Liu, 1997; Righi et al., 1999; Wen and Aydin, 2003; Lado and Ben-Hur, 2004; Caner et al., 2010). Phyllosilicates are also very reactive phases responding quickly to changes in their environment (Pai et al., 2004; Meunier, 2007; Velde and Meunier, 2008; Cornelis et al., 2014).

Therefore, quantitative information on the mineralogical composition of clay-bearing samples is an important step in characterizing and understanding them. Different techniques can be used to quantify clay minerals, but those using X-ray diffraction are the most abundant and have proven to be the most reliable (Plançon, 1981; Reynolds Jr., 1985; Drits and Tchoubar, 1990; Righi et al., 1999; Sakharov et al., 1999a; Środoń, 2006; Hubert et al., 2009, 2012; Ufer et al., 2012a; b; Viennet et al., 2015). Programs calculating X-ray diffraction patterns usually provide the highest level of detail because the input for such models can be considered an approximation of the real structure of the minerals (e.g. layer structures, composition, stacking parameters, interlayer composition, orientation). As such, this approach does not only yield quantitative data, but also structural and compositional information. However, this also means a large number of variables are involved, some of which are very difficult to predict or estimate in advance. In combination with the complex, poly-phasic nature of many natural samples, it is a challenge to create software that allows for the quantification of clay minerals.

Two complementary methods exist to analyse clay minerals using X-ray diffraction. One uses powder samples, for which the orientation of crystallites is considered to be approximately random, and the other uses oriented samples, in which the orientation of crystallites occurs mainly along a plane of preferred orientation. Originally, powder X-ray diffraction was and still is used to determine crystal structures for unknown phases (not just phyllosilicates), which then developed into quantitative analysis. However, for disordered structures like mixed-layered clay minerals, powder patterns are often difficult to interpret. In such cases oriented patterns can be used to focus on the stacking (dis)order along the c* axis. Since the 1970's, several computer programs have been developed to calculate X-ray diffraction patterns for (disordered) clay minerals (Kakinoki & Komura 1965; Reynolds, 1967; Ergun, 1970; Sakharov & Drits, 1973; Drits & Sakharov, 1976; Plançon, 1981). Examples of commonly used programs are the NEWMOD©-family (Reynolds Jr., 1985; Pevear and Schuette, 1993; Reynolds Jr. and Reynolds III, 1996; Yuan and Bish, 2010), MLM2C/3C and derivatives (Plançon and Drits, 2000), Sybilla (Aplin et al., 2006; McCarty, 2015), DIFFaX (Treacy et al., 1991) and BGMN (Ufer et al., 2012a; b). Some of these programs (e.g. DIFFaX, BGMN, Wildfire, Sybilla 3D) are able to calculate X-ray diffraction patterns for random powder diffraction patterns, while others (NEWMOD©, MLM2C, MLM3C, Sybilla) focus only on calculating one-dimensional (00l) patterns.

Another aspect to consider is the ability of these programs to automatically refine parameters. For instance, the last version of NEWMOD© uses a simple linear least-squares algorithm, Sybilla makes use of a genetic algorithm, and BGMN has a custom least-squares algorithm. In essence, all of these algorithms try to find a solution by minimizing a target function, usually a measure for the difference between the calculated and observed data. This difference is usually defined as the sum of the squares of the

errors or as the pattern's Rp factor (Toby, 2006). A linear or ordinary least-squares algorithm works well when there is a well-defined global minimum and the target function is relatively smooth. However, for more complex cases this is often not the case, and as a result an ordinary least-squares might not converge at all. Algorithms using a more stochastic approach, like genetic algorithms, can partly overcome problems related to target function smoothness or poorly defined minima (also see section 2.4). Nonetheless, any algorithm will require some guidance e.g. by not releasing all parameters for automatic refinement at once, by adjusting some parameters manually, by setting upper and lower limits or by choosing starting values close enough to the actual solution. The reason is that models describing X-ray diffraction by disordered layered minerals can not always be constrained adequately, and a successful quantification is still very dependent on the skill of the individual modeller. As a result, most published quantifications of complex mixed-layer assemblages employ a time-consuming trial-and-error approach at some point in the modelling process.

Several authors used a 'multispecimen approach' to further constrain their models (Drits, 1997; Sakharov et al., 1999a; b; Hubert et al., 2012 and references therein). This approach involves recording multiple 'specimens' or patterns (e.g. air-dried, glycolated, heat treatments) of the same sample and creating a structural model that can explain the observed features for all patterns. The reason for doing this is that swelling layers (like smectite and vermiculite layers) will expand or contract in response to these treatments. The level of expansion or contraction can be related to layer charges, and helps in discerning the different swelling phases present and understanding their stacking (dis)order (Ferrage et al. 2005a,b; Michot et al., 2005; Ferrage et al. 2007; Dazas et al., 2015). In short, this approach allows to determine the structure and type of (mixed-layer) clay minerals present in the sample with higher certainty. However, today not a single

program allows for a side-by-side calculation of these patterns. Because of this, modellers are still forced to refine their model parameters on one specimen and then check if the solution also explains the other observations. As long as a manual trial-and-error refinement process is used, this does not pose too many practical problems aside from the time needed. In theory however, a program able to integrate all the observations and calculate patterns for them could lead to better automatic parameter refinements, a hypothesis tested in this paper using theoretical assemblages.

The program presented in this paper, called PyXRD (short for Python X-ray Diffraction), was designed with this multispecimen approach in mind. It (selectively) shares phase parameters across specimens and keeps phase quantities identical in each specimen, thus reducing the number of parameters while at the same time increasing the number of observations. Other design goals for PyXRD were (i) to have an easy-to-use interface, (ii) to be an open program allowing as many aspects of the input to be changed as possible, (iii) to provide a means for automatic parameter refinement, and (iv) to provide an open-source program for others, allowing them to use the software freely and make improvements where they see fit.

This paper illustrates the general structure of this program and presents the results from a comparison between PyXRD and Sybilla v2.2.2 and between automatic parameter refinements for several theoretical mineral assemblages, with and without the use of the multispecimen approach. The software manual contains more detailed information about the numerical solutions used for calculating the X-ray diffraction patterns and a guided example on how to create projects using the graphical user interface (GUI).

2. Materials and methods

2.1 Model implementation and licence

PyXRD is written in Python 2.7 and uses a number of open-source third-party modules. The GUI utilizes PyGTK as widget toolkit and has an internal model-view-controller framework. To improve calculation speed, PyXRD makes use of the NumPy and SciPy libraries. NumPy provides multidimensional array objects and many related routines for manipulating them, while SciPy provides more complex mathematical and scientific algorithms built on top of NumPy (Jones et al., 2001; van der Walt et al., 2011). The Matplotlib library is used for plotting patterns and data (Hunter, 2007). Finally, the Distributed Evolutionary Algorithms for Python (DEAP) library is used to harness to power of evolutionary algorithms to automatically refine parameters (Fortin et al., 2012).

PyXRD is released under a BSD licence, except for the mvc module which, as it is a derived work from the gtkmvc project, is licensed as GNU LGPL v2.

2.2 Program data structure

PyXRD is implemented according to a model-view-controller (mvc) paradigm separating data and calculations from GUI-related aspects. In the following section, an overview is given of the most important objects found in the data layer and their associations. More details can also be found in the manual and the source code documentation.

Project object

The user interface of PyXRD can create (or load) a single *Project* object. It is a container object grouping lists of *AtomType*, *Phase*, *Specimen* and *Mixture* objects together. These are the four top-level objects which are used to calculate X-ray diffraction patterns. Their associations are shown schematically in figure 1. The purpose of each of them will be

explained in more detail below.

AtomType object

The *AtomType* object is the most basic building block. This object bundles all the physical constants (e.g. charge, atomic weight, scattering factors) for a single ion (e.g. Fe^{2+}, Fe^{3+}) or for a molecule (e.g. H_2O and ethylene glycol) small enough to be considered having a spherical electron cloud. When a new project is created, a default list of these *AtomType* objects is loaded, using the atomic scattering factors as published by Waasmaier and Kirfel (1995).

Phase and Component objects

Phase objects contain all the information needed to calculate a one-dimensional X-ray diffraction pattern of a (mixed-layer) mineral. A *Phase* combines (i) a *Probability* object, (ii) an object describing the coherent scattering domain size (*CSDS*), and (iii) one or more *Component* objects which contain information about the structure of the different types of layers in the *Phase*.

The *Probability* object describes how these layers are stacked by means of Markovian statistics and the Reichweite concept (Drits and Tchoubar, 1990). Currently PyXRD has implemented probability models for R values ranging from 0 to 3. For each combination of Reichweite and number of components there are a number of independent parameters required to calculated the remaining parameters, which describe the stacking order or disorder. The values of these independent parameters can be based on another phase with the same combination of Reichweite and number of components. For example, this means it is possible to share the illite content in an illite/smectite mixed layer across its AD and EG phase, but have different weight fractions (or junction probabilities) for the different types of smectite in those phases. For a complete explanation on how these calculations

work and which parameters were chosen to be independent we refer to the manual.

The *CSDS* object describes what type of coherent scattering domain size distribution should be used and contains the necessary parameter values to describe this distribution (e.g. average CSDS). Two types of CSDS distributions are currently implemented: a generic log-normal distribution and a log-normal distribution in which the distribution constants published in Drits et al. (1997) are employed and the average CSDS is the only remaining unknown variable. Each phase also has a σ^* factor which makes it possible to correct for incomplete preferred orientation (Reynolds Jr., 1986; Dohrmann et al., 2009).

The *Component* object describes the size, structure, composition and (variation in) basal spacing of a single layer type in that phase. A *Component* contains two lists that combine an *AtomType* from the project with its (projected) coordinate along the c*-axis (also known as the z coordinate) and the number of projected ions of that type at that coordinate. The first list involves atoms in the silicate lattice, while the other list describes the (variable) interlayer space. With this approach, the silicate structure can be shared between different phases (e.g. AD and EG states), while the interlayer contents may still be different.

Specimen objects

Specimen objects provide all the information regarding the experimental data (the actual measurements, sample size, etc.) and the *Goniometer* set-up (radius, slit sizes, etc.). They do not hold a direct reference to phases, but are linked with them through *Mixture* objects.

Mixture objects

Mixture objects are the starting point for the actual calculations as they link phases and specimens together. In the user interface, a table can be created by adding just as many rows as there are *Phases* and just as many columns as there are *Specimens*. In the column headers, there are slots where the user can select the *Specimen*. Similarly, the

user can select the corresponding *Phase* in each cell of the table. This enables the user to select different states of smectite for an AD and an EG *Specimen* (see figure 2 for a screenshot of the GUI), while keeping unaffected *Phases*, (e.g. kaolinites, micas and chlorites) unchanged.

Once a *Mixture* is created in this way, a number of parameters are available for automatic refinement (e.g. weight fractions from the *Probability* object, the average CSDS, etc.). In a refinement dialog, the user can select which parameters s/he would like to improve and the minimum and maximum values between which the ideal value should lie. A number of different refinement methods are also available - some of them more complex or specialized than others. Yet, as a complete description of all methods is beyond the scope of this article, only the algorithm used for the refinements will be explained in detail below.

2.3 Numerical calculations

The X-ray diffraction patterns are calculated using the matrix formalism, for which a very good summary can be found in Drits and Tchoubar (1990). Later developments incorporated can be found in Drits et al. (1997) and Plançon (2001). Since the complete mathematical deduction followed for PyXRD is rather long, in itself does not contain new developments, and is not the aim of this paper it is not included here. However, an overview of the mathematical deductions and calculations, as they are implemented in the 'calculations' module, can be found in the online manual (http://users.ugent.be/~madumon/pyxrd/Manual.pdf) or in the manual included in this article's supplement.

To improve calculation speed, programs can make use of multi-threading, spreading the load from the different threads evenly over the different cores in a multi-core CPU. However, multi-threading is not very effective in Python because of the Global Interpreter Lock (GIL). This lock can only be obtained by a single active thread, while the others have

to wait for it to be released again. So instead of multi-threading, PyXRD uses multi-processing, which creates a new python interpreter for each process, circumventing the GIL problem. The downside is that processes, unlike threads, do not share memory. Therefore, each process needs to be given all the data required to run the calculation. This is achieved by isolating the calculation functions from objects and by extracting the required data from the objects described in the previous section. As a result, the data exchanged between processes is reduced to a minimum.

This approach also makes it possible to run PyXRD refinements effectively on high-performance computing (HPC) clusters. The experiments presented in this paper were run on the HPC clusters of the Stevin Supercomputer Infrastructure at Ghent University. The main reason to run these experiments on an HPC cluster was the large number of replicates (50) involved in this work and the practical aspect of not having to dedicate a separate PC. This does not mean refinements take too long on a regular PC; e.g. a refinement with a dozen parameters finishes in less then 15 minutes on a 64-bit quad-core 3.10 GHz Intel® Core i5-2400 system. The setup on an HPC cluster is also more cumbersome as it requires the user to get PyXRD to work on the cluster and submit a working job script for his/her refinement from a command line. Running a refinement on a local PC is much easier as the refinement algorithm (see below) and its parameters can be selected and run from the GUI itself.

2.4 Refinement algorithm

PyXRD supports several refinement algorithms, but for more complex problems involving several parameters, the genetic algorithms or evolutionary strategies are found to be most reliable. PyXRD implements several evolutionary strategies, among which are a Covariance Matrix Adaptation Evolutionary Strategy (Hansen and Ostermeier, 2001) and a (multiple) Particle Swarm Optimization (Blackwell et al., 2008). While the Particle Swarm

Optimization is effective at searching the parameter space for minima, being able to escape local minima easily, it can take a lot of function calls for it to converge. On the other hand, the Covariance Matrix Adaptation Evolutionary Strategy is much more effective for local searches, but does get stuck in local minima more easily. Therefore, PyXRD also implements a Particle Swarm Covariance Matrix Adaptation Evolutionary Strategy algorithm which extends the Covariance Matrix Adaptation Evolutionary Strategy with collaborative concepts from a Particle Swarm Optimization (Muller et al., 2009), making it the more robust choice. This Particle Swarm Covariance Matrix Adaptation Evolutionary Strategy was also used for the experiments presented below.

3. Results

In the following sections PyXRD's output is compared with Sybilla's output. In the first section, single phases are tested to check the implementation of the model. In the second section a number of assemblages are tested to check if the obtained weight fractions are correct. In the last section a comparison is made between single- versus multispecimen refinements.

3.1 Comparison between Sybilla and PyXRD results: calculated *00l* reflections for single discrete and mixed-layer phyllosilicates

In total, 13 phases were tested. An overview of these phases with their most important structural parameters are given in table 1. The original Sybilla projects, the produced patterns and the PyXRD projects used can be found in this paper's supplement. All patterns were calculated using a fixed σ* value of 12, a sample length of 1.25 cm, a goniometer radius of 17.3 cm, a divergence slit of 0.5°, Soller slits of 2.3° and an angular range of 2°-52° 2θ with 1000 steps (step size of 0.05° 2θ). The z* coordinates of the atoms were set to match with those in Sybilla, as were the scattering factors, the unit cell

dimension in the z direction, the octahedral iron content (for illite, chlorite and smectite components), the interlayer water, ethylene glycol and cation contents (for smectite and illite components) and the average coherent scattering domain size. The probability parameters were entered as such that identical P and W matrices were obtained. For most of the phases this meant identical parameters could be entered. Only for the R2 illite/smectite with two components 2 additional parameters were entered in comparison with Sybilla, which has a more restricted probability model for this combination of Reichweite and components. These parameters are the junction probabilities P_{21} (fixed at 1.0 in Sybilla) and P_{221} (fixed at 0.0 in Sybilla). A complete deduction on how the entered probabilities and weight fractions are used to calculate the unknown weight and probability fractions is present in the PyXRD manual. Sybilla uses scattering factors for the atoms in the silicate lattice assuming 50% ionization, with the exception of Mg which is fully ionized (D. McCarty, 2015). The scattering factors used in PyXRD for this study have been set to match this.

The kaolinite, illite, talc and chlorite phases are composed of a single component. As such, these are testing the basic aspects of the model such as the orientation factor σ*, the calculation of the coherent scattering domain size and the calculation of the atomic scattering and structure factors. To test whether PyXRD can handle different sample states correctly, an R0 two-component smectite in air-dried and glycolated state is modelled as well. To further test the implementation of the matrix algorithm for mixed-layer phases, and the related probability models, a number of illite/smectite phases were used. In total seven phases were tested, four of which are two-component illite/smectite phases with Reichweite values varying from 0 to 3 and three of which are three-component illite/smectite phases with Reichweite values varying from 0 to 2. The different smectite components have different hydration states, i.e. the first component always has 2 planes

of water (AD state) or 2 planes of ethylene glycol molecules (EG state) in its interlayer space while the second component has only a single layer of water or ethylene glycol molecules. For these illite/smectites two variants were calculated: one with a low CSDS not at maximum possible degree of ordering (MPDO) and one with a higher CSDS at MPDO.

Table 1 contains the Rp factor obtained for these test cases. A few of these patterns are presented in figures 3 and 4. From them and from the Rp and Rwp factors, it is clear PyXRD can produce patterns almost identical to those produced by Sybilla. The small deviations can probably be explained by different physical constants (e.g. atomic scattering factors), although it is impossible to know exactly.

3.2 Comparison between Sybilla and PyXRD results: calculated *00l* reflections for mixtures of discrete and mixed-layer phyllosilicates

To further validate the model, five patterns were produced in PyXRD for mixtures of increasing complexity. These patterns were imported in Sybilla and modelled using the same phases and the same parameters. This should allow to validate whether the weight fractions in PyXRD can also be obtained by Sybilla. The entered and obtained weight fractions and the corresponding Rp and Rwp factors are presented in table 2. Figure 5 shows the comparison between the calculated patterns for mixture 5 from Sybilla and PyXRD. The used phases are largely identical to the phases used in the previous validation, except for the addition of a few phases for which details are also given in table 2. The input files for PyXRD and Sybilla are included in this paper's supplement.

As can be observed, the weight fractions in PyXRD and Sybilla are reasonably close to each other, with all of the deviations being smaller then 0.5 wt%. Such differences are not to be considered significant for a real sample, but when using 'ideal' theoretical phases

they do indicate there are differences between Sybilla and PyXRD. In order to check if some phases contribute more or less then others to the whole-pattern misfit, Rp and Rwp factors are calculated for angular ranges corresponding to first-order reflections for the phases in mixtures 1, 2 and 3 (table 3). This was not done for mixtures 4 and 5 due to the presence of too many overlapping peaks, making statements about the contribution of separate phases to the total misfit difficult. The Rp and Rwp factors obtained in this way are all of the same order of magnitude. Therefore each phase must be contributing more or less equally towards the whole-pattern misfit. The remaining differences between the patterns can be explained by small differences in physical constants, while the remaining differences in wt% can be explained by differences in unit cell dimensions.

3.3 Multispecimen tests

3.3.1 Assemblage setup

In total, four theoretical mineral assemblages were tested (table 4):

Assemblage 1 is a very simple test because of the absence of overlapping and similar phases. Its main purpose was to see whether the program and, more importantly, the selected refinement strategy, can produce a reliable result. The assemblage consists of equal amounts of a discrete kaolinite, a discrete illite and an R0 illite/smectite with only 10% illite layers

Assemblage 2 is more complex, comprising six different phases: a discrete illite, a discrete kaolinite, an R0 illite/smectite with 65% illite layers, an R0 kaolinite/smectite with 80% kaolinite layers, a smectite and a poorly-crystalline chlorite. The idea behind this assemblage was to mimic phases encountered in some soils. The poorly-crystalline chlorite component can be interpreted as a small amount of hydroxy-interlayered smectite (or vermiculite) and is not to be considered a primary trioctahedral chlorite, while the

kaolinite/smectite represents a neoformed, defective kaolinite or smectite. This kind of phase has been reported a number of times, usually in finer clay fractions (≤ 0.2 μm) of certain soils (Hubert et al., 2009, 2012; Ryan and Huertas, 2009; Dumon et al., 2014). The different phases are also present in different quantities, with the illite-bearing phases each contributing 25.0 wt%, the smectite taking up 20.0 wt%, the kaolinite phases each accounting for 12.5 wt% and the chlorite being a minor phase with only 5.0 wt%.

Assemblage 3 is composed of 30% discrete illite, 35% kaolinite, 20% high-charge smectite (vermiculite-like) and 15% low-charge smectite. The main idea behind this test assemblage was to see whether the presence of high-charge and low-charge phases (which in this case produced similar patterns under AD and heated conditions, but different patterns under EG conditions) has an influence on the refinement and the quantification in the different set-ups.

Test patterns for assemblage 4 were calculated with 35% well-crystallized kaolinite (with a high average CSDS), 15% poorly-crystallized kaolinite (with a low average CSDS) and 50% of an R0 illite/smectite with 98% of illite layers. However, these patterns were not modelled with the same structural models. Instead of two different kaolinites, a single kaolinite was added, and instead of an illite/smectite, a discrete illite was used. As such, the influence of a simplified model input could be checked, which is a common error in real-life uses (e.g. due to misinterpretation).

After the necessary phases and their parameters were set up, a calculated pattern was generated from 2 to 50 °2θ with a 0.02° step size, saved and re-imported as experimental data. Random noise was also added to these patterns, using the following formula:

$$I_n = I_o \cdot \left(1 + \left(X - 0.5\right) \cdot f_n\right)$$

where I_n is the intensity with noise, I_o the original intensity, X a random fraction between 0 and 1 and f_n the noise factor, which was set to 0.01. This results in a random deviation of

at most 0.5% above or below the original intensity. This is a high noise level when only considering statistical counting noise, however, these noise levels can be obtained on iron-rich samples when working with a Cu X-ray source due to iron fluorescence. Energy dispersive detectors can eliminate most of this noise nowadays, but it can still be a problem on older equipment, hence it is included here.

For assemblages 1 and 2, both the smooth and noisy patterns were used in separate refinements to assess the influence of this treatment. For assemblages 3 and 4, only the noisy patterns were used, because the previous two experiments showed little influence of the noise on the final results (see below).

Since evolutionary refinement strategies have a stochastic component, each refinement will be different, even if starting and boundary conditions are identical. Nonetheless, the starting point may also have an influence on the final result. To average out these differences and to check if the final output is reproducible, 50 random starting points were sampled so that a normal distribution over the parameter space was obtained. For each of these points a refinement was started. At the end of these refinements, average parameter values and their standard deviations were calculated for these 50 iterations. Additionally, the model kept track of the best solution found at each generation in each iteration, allowing us to create parameter evolution plots.

3.3.2 Assemblage 1

An overview of the obtained average parameter values and standard deviations for assemblage 1 can be found in tables 5 and 6. Parameter evolution plots for two selected parameters (the average CSDS and the fraction of illite layers in the illite/smectite) are also shown in figure 6. Most parameters are determined accurately and with very high precision. The difference between noisy patterns and smooth patterns is marginal, and no

difference can be observed between the runs where multiple specimens are combined and those where only a single specimen was used for refinement. As a result of this, the obtained weight fractions for the three phases are also very accurate. The obtained level of accuracy is not a realistic level for natural samples, but stems from the simplicity of this set-up. For the runs using the noisy patterns, a very small (and systematic) deviation in the obtained weight fractions can be observed. This is probably the result of the added noise, since the deviation is not present for runs using the smooth patterns.

3.3.3 Assemblage 2

An overview of the obtained average parameter values and standard deviations for assemblage 2 can be found in tables 7 and 8. As was the case in the previous assemblage, no significant difference can be observed between runs that use smooth patterns and those that use noisy ones. Both types produced similar parameter accuracies and precisions. Overall, the results are less accurate and precise compared to assemblage 1, but still very good. Most notably, the weight fractions of the smectite layer types in the kaolinite/smectite show a much larger imprecision. This is also the case in the parameter evolution plots (figure 7) for these fractions. An explanation can be found in the sensitivity of these parameters: since the kaolinite fraction in this mixed-layer is relatively high (80%), the relative amounts of the different types of smectite layers do not have such a large influence on the calculated pattern. Some differences are also noticeable between runs that combine multiple specimens and those where only heated patterns were used. For the latter, the imprecision on the weight fractions for the illite, illite/smectite and smectite phases is significantly larger compared to the other runs. This is to be expected, as heating collapses swelling layers, causing significant peak overlap with the illite peaks. Despite this overlap, it was still possible to obtain accurate and precise averages for the

other parameters, comparable to the other runs.

3.3.4 Assemblage 3

An overview of the obtained average parameter values and standard deviations for assemblage 3 can be found in table 9. With this assemblage, the combined set-up and the set-up using only the EG pattern both resulted in the same performance, giving accurate and precise parameter values. The set-up with AD or heated patterns, on the other hand, led to inaccurate and imprecise results, especially when the weight fractions are taken into account. Finally, it can also be observed that the weight fractions and parameter values of phases that were unaffected by the treatments (i.e. kaolinite and illite) are more accurate and precise in these set-ups. It is mainly for the overlapping phases (i.e. smectites) that the errors occur.

Figure 8 shows the parameter plots for the multispecimen set-up and the AD set-up for a few selected parameters. This figure illustrates the divergent nature of some parameters in the AD set-up very well, while it is clear that the combined set-up does not suffer from this as it has access to the EG pattern as well.

The outcome of this experiment is in line with our expectations, as only the EG pattern contains enough information to distinguish these two smectites from each other. When the EG pattern is absent, the results become divergent, resulting in the high imprecision observed for the AD and heated pattern set-ups.

3.3.5 Assemblage 4

An overview of the obtained average parameter values and standard deviations for assemblage 4 can be found in table 10. In this set-up, we intentionally misidentified a mixed-layer illite/smectite as an illite and overlooked the presence of two populations of kaolinite instead of one. Nevertheless the flawed structural model is able to give us decent

parameter accuracies. These kinds of 'mistakes' are quite common in the real-life use of this kind of program, and apparently do not matter too much either, as long as they are related to natural inhomogeneities. In contrast, a model based on a completely wrong interpretation will never yield any good output, and will result in a very obvious mismatch between the calculated and observed patterns. Even in this assemblage, the (residual) XRD patterns (figure 9) show a clear mismatch for these phases. An observant user should notice this and as such be able to identify wrong and/or missing phases.

3.3.6 Summary

For all four assemblages, PyXRD has been able to reproduce the input parameters or at least approximate them with the multispecimen approach. The only complications occur when single patterns are used which do not contain enough information on their own (in most cases heated patterns).

The results for these theoretical assemblages seem to suggest that the multi-specimen approach does not add a lot of constraints to the mathematical model. Instead, it appears far more important to correctly identify the phases using multiple specimens than to use these for the parameter refinement. As a result, once the phases are correctly identified, a good quantification can often be obtained with only a single pattern if all phases can be sufficiently discerned from one another in that state. For most natural samples, this could imply that it is sufficient to model the EG and/or the AD pattern. Indeed, many papers presenting modelled X-ray diffraction patterns of phyllosilicates only use the AD and/or EG patterns (Plançon and Roux, 2010; Hubert et al., 2012; Ufer et al., 2012a; Dumon et al., 2014). However, it is important to realize that these results from theoretical experiments cannot be extrapolated automatically to all real-life modelling experiments.

In this context, one needs to understand how realistic it is to share some of the parameters between the different specimens during the refinement. Some of them are rather difficult or

impossible to control from measurement to measurement. For example, the number of water or ethylene glycol planes intercalated into smectite bearing phases is not only dependent on layer charge and the saturating cation, but also on the ambient conditions (i.e. temperature and relative humidity) (Tamura et al., 2000). Because of this, a lot of the parameters cannot and should not be shared, and the advantage of having added more observations is partially lost.

5. Conclusion

In this paper we have presented PyXRD, a new, free and open-source program to perform a (semi-)quantitative analysis of disordered layered minerals using multispecimen X-ray diffraction profile fitting. It is the authors' sincere hope that others will pick up on the program and improve it. The novelty of this program lies specifically in the *ab initio* incorporation of the multispecimen method, making it possible to share phases and (a selection of) their parameters across multiple specimens. This allows to model several specimens side-by-side, and is an important step forward. In theory, this could also help in further constraining the mathematical model and thus improving the automatic parameter refinement results (Sakharov et al., 1999a; Meunier, 2005; Lanson, 2011). However, results from theoretical experiments indicate that a multispecimen refinement setup is not always required to obtain good parameter estimates. Finally, it remains of paramount importance to use the multispecimen method to obtain a correct identification, as without it no meaningful quantification can ever be obtained. We can conclude that PyXRD has proven to be very useful when X-ray diffraction patterns for complex mineral assemblages containing (mixed-layer) phyllosilicates are modelled with a multispecimen approach.

6. Code availability

The source code for PyXRD can be found online at https://github.com/mathijs-

dumon/PyXRD, together with installation instructions and a manual with detailed information regarding the calculations and a step-by-step example on how to use the user interface.

7. Acknowledgements

This research was funded by the project G028714N of the Fund for Scientific Research (Flanders). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, the Hercules Foundation and the Flemish Government (Department EWI), for which we are very grateful.

8. References

Agbenin, J.O., and Tiessen, H., 1995. Soil properties and their variations on two contiguous hill slopes in Northeast Brazil. Catena 24, 147–161.

Aplin, A.C., Matenaar, I.F., McCarty, D.K., and van der Pluijm, B.A., 2006. Influence of mechanical compaction and clay mineral diagenesis on the microfabric and pore-scale properties of deep-water Gulf of Mexico mudstones. Clays Clay Miner 54, 500–514.

Blackwell, T., Branke, J., and Li, X., 2008. Particle Swarms for Dynamic Optimization Problems. p. 193–217. *In* Blum, C., Merkle, D. (eds.), Swarm Intelligence. Natural Computing Series. Springer Berlin Heidelberg.

Caner, L., Hubert, F., Moni, C., and Chenu, C., 2010. Impact of clay mineralogy on stabilisation of organic matter in the clay fraction of a Neo-Luvisol and a Cambisol. p. 73–76. In 19th World Congress of Soil Science: Soil Solutions for a Changing World. Brisbane, Australia.

McCarty, D., (2015, October 1). Personal communication - Sybilla, Computer software version 2.2.2 © Chevron Technology Company, a division of Chevron, U.S.A. Inc., San Ramon, CA.

Cornelis, J.-T., Weis, D., Lavkulich, L., Vermeire, M.-L., Delvaux, B., and Barling, J., 2014. Silicon isotopes record dissolution and re-precipitation of pedogenic clay minerals in a podzolic soil chronosequence. Geoderma 235–236, 19–29.

Dazas, B., Lanson, B., Delville, A., Robert, J.-L., Komarneni, S., Michot, L.J., and Ferrage, E., 2015, Influence of Tetrahedral Layer Charge on the Organization of Interlayer Water and Ions in Synthetic Na-Saturated Smectites: The Journal of Physical Chemistry C, v. 119, no. 8, p. 4158–4172, doi: 10.1021/jp5123322.

Dohrmann, R., Rüping, K.B., Kleber, M., Ufer, K., and Jahn, R., 2009. Variation of

preferred orientation in oriented clay mounts as a result of sample preparation and composition. Clays Clay Miner 57, 686–694.

Drits, V.A., 1997. Mixed-layer minerals. p. 153–190. *In* Merlino, S. (ed.), Modular Aspects of Minerals. EMU Notes in Mineralogy. Eötvös University Press, Budapest.

Drits, V., Środoń, J., and Eberl, D.D., 1997. XRD Measurement of mean crystallite thickness of illite and illite/smectite: reappraisal of the Kubler Index and the Scherrer Equation. Clays Clay Miner 45, 461–475.

Drits, V.A., and Sakharov,, B.A., 1976. X-ray structure analysis of interstratified minerals. Nauka, Moscow, 225 pp (in Russian).

Drits, V.A., and Tchoubar, C., 1990. X-Ray Diffraction by Disordered Lamellar Structures: Theory and Applications to Microdivided Silicates and Carbons. Springer-Verlag, Berlin, Germany.

Dumon, M., Tolossa, A.R., Capon, B., Detavernier, C., and Van Ranst, E., 2014. Quantitative clay mineralogy of a Vertic Planosol in southwestern Ethiopia: Impact on soil formation hypotheses. Geoderma 214-215, 184–196.

Ergun, S. (1970). X-ray scattering by very defective lattices. Phys. Rev. B, 131, 3371-3380.

Ferrage, E., Lanson, B., Sakharov, B.A., Geoffroy, N., Jacquot, E., and Drits, V.A., 2007. Investigation of dioctahedral smectite hydration properties by modeling of X-ray diffraction profiles: Influence of layer charge and charge location. American Mineralogist 92, 1731–1743.µ

Ferrage, E., Lanson, B., Malikova, N., Plancon, A., Sakharov, B.A., and Drits, V.A., 2005a. New Insights on the Distribution of Interlayer Water in Bi-Hydrated Smectite from X-ray Diffraction Profile Modeling of 00l Reflections. Chem Mater 17, 3499–3512.

Ferrage, E., Lanson, B., Sakharov, B.A., and Drits, V.A., 2005b. Investigation of smectite

hydration properties by modeling experimental X-ray diffraction patterns: Part I. Montmorillonite hydration properties. American Mineralogist 90, 1358–1374.

Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A., Parizeau, M., and Gagné, C., 2012. DEAP: Evolutionary Algorithms Made Easy. JMLR 13, 2171–2175.

Hansen, N., and Ostermeier, A., 2001. Completely Derandomized Self-Adaptation in Evolution Strategies. Evol. Comput. 9, 159–195.

Hubert, F., Caner, L., Meunier, A., and Ferrage, E., 2012. Unraveling complex <2μm clay mineralogy from soils using X-ray diffraction profile modeling on particle-size sub-fractions: Implications for soil pedogenesis and reactivity. Am Mineral 97, 384–398.

Hubert, F., Caner, L., Meunier, A., and Lanson, B., 2009. Advances in characterization of soil clay mineralogy using X-ray diffraction: from decomposition to profile fitting. European Journal of Soil Science 60, 1093–1105.

Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering 9, 90–95.

Jones, E., Oliphant, T., and Peterson, P., 2001. SciPy: Open Source Scientific Tools for Python.

Kakinoki, J. and Komura, Y. (1965) Diffraction by a one-dimensionally disordered crystal. I: The intensity equation. Acta Cryst., 19, 137-147.

Lado, M., and Ben-Hur, M., 2004. Soil mineralogy effects on seal formation, runoff and soil loss. Appl. Clay Sci. 24, 209–224.

Lanson, B., 2011. Modelling of X-ray diffraction profiles: Investigation of defective lamellar structure crystal chemistry. p. 151–202. *In* Brigatti, M.F., Mottana, A. (eds.), Layered mineral structures and their Application in Advanced Technologies. EMU Notes in Mineralogy.

Meunier, A., 2005. Clays. Springer, Berlin, Germany.

583 Meunier, A., 2007. Soil Hydroxy-Interlayerd minerals: a re-interpretation of their
584 crystallochemical properties. Clays Clay Miner 55, 380–388.
585 Michot, L.J., Bihannic, I., Pelletier, M., Rinnert, E., and Robert, J.-L., 2005, Hydration and
586 swelling of synthetic Na-saponites: Influence of layer charge: American Mineralogist,
587 v. 90, no. 1, p. 166–172.
588 Muller, C.L., Baumgartner, B., and Sbalzarini, I., 2009. Particle Swarm CMA Evolution
589 Strategy for the optimization of multi-funnel landscapes. p. 2685–2692. *In* IEEE
590 Congress on Evolutionary Computation, 2009. CEC ’09.
591 Pai, C.W., Wang, M.K., King, H.B., Chiu, C.Y., and Hwong, J.-L., 2004. Hydroxy-
592 interlayered minerals of forest soils in A-Li Mountain, Taiwan. Geoderma 123, 245–
593 255.
594 Pevear, D.R., and Schuette, J.F., 1993. Inverting the NEWMOD© X-ray Diffraction Forward
595 Model for Clay Minerals Using Genetic Algorithms. p. 19–42. *In* Computer
596 Applications to X-ray Powder Diffraction Analysis of Clay Minerals Vol. 5. Clay
597 Minerals Society, Boulder, Colorado.
598 Plançon A., 1981. Diffraction by layer structures containing different kinds of layers and
599 stacking faults. Journal of Applied Crystallography, 14, 300-304.
600 Plançon, A., 2001. Order-disorder in clay mineral structures. Clay Miner 36, 1–14.
601 Plançon, A., and Drits, V.A., 2000. Phase analysis of clays using an expert system and
602 calculation programs for x-ray diffraction by two- and three-component mixed layer
603 minerals. Clays Clay Miner 48, 57–62.
604 Plançon, A., and Roux, J., 2010. Software for the assisted determination of the structural
605 parameters of mixed-layer phyllosilicates. European Journal of Mineralogy 22, 733–
606 740.
607 Reynolds, R.C., 1967. Interstratified clay systems: calculation of total one-dimensional

608 diffraction functions. American Mineralogist, 52, 661–672.
609 Reynolds Jr, R.C., 1985. NEWMOD: a computer program for the calculation of one-
610 dimensional diffraction patterns of mixed-layer clays, R.C. Reynolds, Hanover, NH,
611 USA
612 Reynolds Jr, R.C., 1986. The Lorentz-Polarisation factor and preferred orientation in
613 oriented clay aggregates. Clays Clay Miner 34, 359–367.
614 Reynolds Jr, R.C., and Reynolds III, R.C., 1996. NEWMOD-for-Windows. The Calculation
615 of One-Dimensional X-ray Diffraction Patterns of Mixed Layered Clay
616 Minerals. Computer program, RC Reynolds, Hanover, NH, USA.
617 Righi, D., Terribile, F., and Petit, S., 1999. Pedogenic formation of kaolinite-smectite mixed
618 layers in a soil toposequence developed from basaltic parent material in Sardinia
619 (Italy). Clays Clay Miner 47, 505–514.
620 Ryan, P.C., and Huertas, F.J., 2009. The temporal evolution of pedogenic Fe–smectite to
621 Fe–kaolin via interstratified kaolin–smectite in a moist tropical soil chronosequence.
622 Geoderma 151, 1–15.
623 Sakharov, B. A. and Drits, V. A., 1973. Mixed-layer kaolinite-montmorillonite: a comparison
624 of observed and calculated diffraction patterns. Clays Clay Miner 21, 15-17.
625 Sakharov, B.A., Lindgreen, H., Salyn, A., and Drits, V.A., 1999a. Determination of illite-
626 smectite structures using multispecimen x-ray diffraction profile fitting. Clays Clay
627 Miner 47, 555–566.
628 Sakharov, B.A., Lindgreen, H., Salyn, A.L., and Drits, V.A., 1999b. Mixed-layer kaolinite-
629 illite-vermiculite in North Sea shales. Clay Miner 34, 333–344.
630 Środoń, J., 2006. Identification and Quantitative Analysis of Clay Minerals. p. 765–787. *In*
631 Handbook of Clay Science. Developments in Clay Science. Elsevier, USA.
632 Treacy, M. M. J., Newsam, J. M. & Deem, M. W. (1991). A General Recursion Method for

633 Calculating Diffracted Intensities from Crystals Containing Planar Faults. P Roy Soc
634 Lond A, 433, 499–520.
635 Tamura, K., Yamada, H., and Nakazawa, H., 2000. Stepwise Hydration of High-Quality
636 Synthetic Smectite with Various Cations. Clays Clay Miner 48, 400–404.
637 Toby, B.H., 2006. R factors in Rietveld analysis: How good is good enough? Powder Diffr
638 21, 67-70.
639 Ufer, K., Kleeberg, R., Bergmann, J., and Dohrmann, R., 2012a. Rietveld refinement of
640 disordered illite-smectite mixed-layer structures by a recursive algorithm. I: One-
641 dimensional patterns. Clays Clay Miner 60, 507–534.
642 Ufer, K., Kleeberg, R., Bergmann, J., and Dohrmann, R., 2012b. Rietveld refinement of
643 disordered illite-smectite mixed-layer structures by a recursive algorithm. II: Powder-
644 pattern refinement and quantitative phase analysis. Clays Clay Miner 60, 535–552.
645 Velde, B.B., and Meunier, A., 2008. The Origin of Clay Minerals in Soils and Weathered
646 Rocks. Springer-Verlag, Berlin, Germany.
647 Vernik, L., and Liu, X., 1997. Velocity anisotropy in shales: A petrophysical study:
648 Geophysics, 62, 521–532.
649 Viennet, J.C. Hubert, F., Ferrage, E., Tertre, E., Legout, A. and Turpault, M.P., 2015.
650 Investigation of clay mineralogy in a temperate acidic soil of a forest using X-ray
651 diffraction profile modelling: Beyond the HIS and HIV description. Geoderma 241–
652 242, 75–86.
653 Waasmaier, D., and Kirfel, A., 1995. New analytical scattering-factor functions for free
654 atoms and ions. Acta Crystallographica Section A Foundations of Crystallography 51,
655 416–431.
656 Wen, B.P., and Aydin, A., 2003. Microstructural study of a natural slip zone: quantification
657 and deformation history. Engineering Geology, 68, 289–317.

Van der Walt, S., Colbert, S.C., and Varoquaux, G., 2011. The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science & Engineering 13, 22–30.

Yuan, H., and Bish, D.L., 2010. NEWMOD+, a new version of the NEWMOD program for interpreting x-ray powder diffraction patterns from interstratified clay minerals. Clays Clay Miner 58, 318–326.

Table 1: Overview of the discrete phases used to compare the output from PyXRD with the output of Sybilla (R is Reichweite, G is the number of components, $\overline{N}$ is the average CSDS, AD is air-dry, EG is ethylene-glycol, relevant probability (P) and weight (W) factors are given, Rp and Rwp are the unweighted and weighted residual errors of the patterns respectively).

	Phase	R	G	State	$\overline{N}$	P and W factors			Rp	Rwp
1	Kaolinite	-	1	-	20	-			0.7	0.9
2	Illite	-	1	-	20	-			0.9	1.3
3	Talc	-	1	-	20	-			0.8	1.0
4	Chlorite	-	1	-	20	-			0.8	1.0
5	Illite/Smectite	0	2	AD	4	$W_1 = 0.5$			1.0	1.6
6	Illite/Smectite	1	2	AD	4	$W_1 = 0.6$ $P_{22} = 0.5$			1.0	1.6
6b	Illite/Smectite	1	2	AD	15	$W_1 = 0.6$ $P_{22} = 0.0$			0.7	1.5
7	Illite/Smectite	2	2	AD	4	$W_1 = 0.6$ $P_{21} = 1.0$	$P_{112} = 0.5$ $P_{221} = 0.0$		1.4	2.1
7b	Illite/Smectite	2	2	AD	15	$W_1 = 0.6$ $P_{21} = 1.0$	$P_{112} = 1.0$ $P_{221} = 0.0$		0.7	1.5
8	Illite/Smectite	3	2	AD	4	$W_1 = 0.9$ $P_{2112} = 0.5$	$P_{22}=0$ $P_{212}=0$		1.9	2.3
8b	Illite/Smectite	3	2	AD	15	$W_1 = 0.9$ $P_{2112} = 0.0$	$P_{22}=0$ $P_{212}=0$		0.5	1.0
9	Illite/Smectite	0	3	AD	4	$W_1 = 0.33$ $W_2 / (W_2+W_3) = 0.5$			1.1	1.7
10	Illite/Smectite	1	3	AD	4	$W_1 = 0.5$ $W_2 / (W_2+W_3) = 0.8$	$P_{11} = 0.85$ $(W_{22}+W_{23}) / (W_{22}+W_{23}+W_{32}+W_{33}) = 0.85$	$W_{22} / (W_{22}+W_{23}) = 0.8$ $W_{32} / (W_{32}+W_{33}) = 0.7$	1.1	1.7
10b	Illite/Smectite	1	3	AD	15	$W_1 = 0.5$ $W_2 / (W_2+W_3) = 0.8$	$P_{11} = 0.0$ $(W_{22}+W_{23}) / (W_{22}+W_{23}+W_{32}+W_{33}) = 0.85$	$W_{22} / (W_{22}+W_{23}) = 0.8$ $W_{32} / (W_{32}+W_{33}) = 0.7$	1.0	1.7
11	Illite/Smectite	2	3	AD	4	$W_1 = 0.8$ $W_2 / (W_2+W_3) = 0.5$	$P_{x1x} = 0.5$ $(W_{212}+W_{213}) / (W_{212}+W_{213}+W_{312}+W_{313}) = 0.5$	$W_{212} / (W_{212}+W_{213}) = 0.5$ $W_{312} / (W_{312}+W_{313}) = 0.5$	1.7	2.2
11b	Illite/Smectite	2	3	AD	15	$W_1 = 0.8$ $W_2 / (W_2+W_3) = 0.5$	$P_{x1x} = 0.0$ $(W_{212}+W_{213}) / (W_{212}+W_{213}+W_{312}+W_{313}) = 0.5$	$W_{212} / (W_{212}+W_{213}) = 0.5$ $W_{312} / (W_{312}+W_{313}) = 0.5$	0.5	1.0
12	Smectite	0	2	AD	4	$W_1 = 0.7$			1.3	1.7
13	Smectite	0	2	EG	4	$W_1 = 0.7$			1.0	1.2

Table 2: Overview of the test mixtures used to compare the weight fraction output from PyXRD with the output of Sybilla, with details for the different phases (R is Reichweite, $\overline{N}$ is the average CSDS, d_{001} is the basal spacing, relevant probability (P) and weight (W) factors are given).

Mixture	Rp	Rwp	Phases	PyXRD wt%	Sybilla wt%	Phase characteristics
1	0.8	1.0	Kaolinite	70.0	69.9	As in table 1
			Illite	30.0	30.1	As in table 1
2	1.1	1.5	Kaolinite	20.0	20.0	As in table 1
			Illite	30.0	30.1	As in table 1
			IS R0	10.0	9.5	As in table 1
			SSS R0	40.0	40.5	$\overline{N}$ = 4; W_1=0.8; $W_2/(W_2+W_3)$=0.8
3	1.0	1.4	Kaolinite	10.0	10.0	As in table 1
			Illite	25.0	25.1	As in table 1
			Chlorite	20.0	20.1	As in table 1
			IS R0	15.0	14.7	As in table 1
			CS R1	10.0	10.1	$\overline{N}$ = 10; W_1=0.5; P_{11} = 0.1
			SSS R0	20.0	20.0	$\overline{N}$ = 4; W_1=0.8; $W_2/(W_2+W_3)$=0.8
4	1.3	1.8	ISS R0	15.0	15.2	As in table 1
			CSS R0	5.0	5.0	$\overline{N}$ = 5; W_1=0.4; $W_2/(W_2+W_3)$=0.9
			Chlorite	5.0	5.0	As in table 1
			Illite	15.0	14.9	As in table 1
			Kaolinite 1	15.0	14.9	As in table 1
			Kaolinite 2	25.0	25.0	$\overline{N}$ = 6; d_{001} = 0.718 nm
			IS R1	10.0	10.0	As in table 1
			CS R1	10.0	10.1	$\overline{N}$ = 10; W_1=0.5; P_{11} = 0.1
5	1.7	2.4	ISS R0	10.0	10.0	As in table 1
			CSS R0	10.0	10.0	$\overline{N}$ = 5; W_1=0.4; $W_2/(W_2+W_3)$=0.9
			Chlorite	10.0	10.0	As in table 1
			Illite	10.0	9.9	As in table 1
			Kaolinite 1	10.0	10.0	As in table 1
			Kaolinite 2	10.0	10.0	$\overline{N}$ = 6; d_{001} = 0.718 nm
			IS R1	10.0	10.1	As in table 1
			CS R1	10.0	9.9	$\overline{N}$ = 10; W1=0.5; P_{11} = 0.1
			SS R0	10.0	9.9	$\overline{N}$ = 4; W1=0.7
			KSS R0	10.0	10.0	$\overline{N}$ = 7; W1=0.6; W2/(W2+W3)=0.8

Table 3: Calculated Rp and Rwp values for selected intervals of mixtures 1, 2 and 3 calculated in PyXRD and Sybilla. (Rp and Rwp are the unweighted and weighted residual errors of the selected intervals respectively)

Angular range (°2θ)	Dominant phase(s)	Approximate d-spacing	Rp (%)	Rwp (%)
Mixture 1				
8.2-9.2	Illite	~ 1.0 nm	1.1	1.1
11.25-13.25	Kaolinite	~ 0.72 nm	0.6	0.9
Mixture 2				
4.5-7.0	IS R0 & SSS R0	~ 1.5 nm	0.9	1.1
8.2-9.2	Illite	~ 1.0 nm	0.5	0.6
11.25-13.25	Kaolinite	~ 0.72 nm	0.7	1.0
Mixture 3				
2.5-4.0	CS R1	~ 2.9 nm	0.8	0.8
4.5-7.0	IS R0, SSS R0 & CS R1	~ 1.5 nm	1.1	1.3
8.2-9.2	Illite	~ 1.0 nm	0.5	0.6
11.25-13.25	Kaolinite	~ 0.72 nm	0.5	0.7

Table 4: Overview of the different test assemblages for the comparison between multispecimen and single pattern refinements and the type of refined patterns.

	Assemblage	Smooth pattern?	Noisy pattern?
1	33.3% Kaolinite 33.3% Illite 33.3% Illite/Smectite (10/90) R0	yes	yes
2	25.0% Illite 25.0% Illite/Smectite (65/35) R0 20.0% Smectite 12.5% Kaolinite 12.5% Kaolinite/Smectite (80/20) R0 5.0% Chlorite	yes	yes
3	35.0% Kaolinite 30.0% Illite 15.0% High-charge smectite 20.0% Low-charge smectite	no	yes
4	35.0% Kaolinite (CSDS = 20) 15.0% Kaolinite (CSDS = 6) 50.0% Illite/Smectite (98/2) R0	no	yes

Table 5: Overview of the means and standard deviations for weight fractions and refined parameters for assemblage 1 using smooth patterns.

Assemblage #1 - smooth patterns					Multiple specimens (n=50)	Only AD (n=50)	Only EG (n=50)	Only 350 heated (n=50)
Phase	Property name	True value	Range		Obtained value	Obtained value	Obtained value	Obtained value
			Min.	Max.	μ ± σ	μ ± σ	μ ± σ	μ ± σ
Kaolinite	**wt%**	33.3	–	–	33.3 ± 0.00	33.3 ± 0.00	33.3 ± 0.00	33.3 ± 0.00
	T	10.0	8.0	20.0	10.0 ± 0.00	10.0 ± 0.00	10.0 ± 0.00	10.0 ± 0.00
Illite	**wt%**	33.3	–	–	33.3 ± 0.00	33.3 ± 0.00	33.3 ± 0.00	33.3 ± 0.00
	T	10.0	8.0	20.0	10.0 ± 0.00	10.0 ± 0.00	10.0 ± 0.00	10.0 ± 0.00
Illite/Smectite R0	**wt%**	33.3	–	–	33.3 ± 0.00	33.3 ± 0.00	33.3 ± 0.00	33.3 ± 0.00
	T	5.0	3.0	10.0	5.0 ± 0.00	5.0 ± 0.00	5.0 ± 0.00	5.0 ± 0.00
	Illite content	0.1	0.0	1.0	0.10 ± 0.00	0.10 ± 0.00	0.10 ± 0.00	0.10 ± 0.00
	2wat / (2wat + 1wat)	0.5	0.0	1.0	0.50 ± 0.00	0.50 ± 0.00	–	–
	2 glv / (2glv + 1glv)	0.5	0.0	1.0	0.50 ± 0.00	–	0.50 ± 0.00	–
	0 glv / (0glv + 1glv)	1.0	0.0	1.0	1.00 ± 0.00	–	–	1.00 ± 0.00

Table 6: Overview of the means and standard deviations of weight fractions and refined parameters for assemblage 1 using noisy patterns.

Assemblage #1 - Noisy patterns					**Multiple specimens** (n=50)	**Only AD** (n=50)	**Only EG** (n=50)	**Only 350°C** (n=50)
Phase	**Property name**	**True value**	**Range**		**Obtained value**	**Obtained value**	**Obtained value**	**Obtained value**
			Min.	**Max.**	**μ ± σ**	**μ ± σ**	**μ ± σ**	**μ ± σ**
Kaolinite	**wt%**	33.3	–	–	33.4 ± 0.0	33.4 ± 0.0	33.4 ± 0.0	33.4 ± 0.0
	T	10.0	8.0	20.0	10.0 ± 0.0	10.0 ± 0.0	10.0 ± 0.0	10.1 ± 0.0
Illite	**wt%**	33.3	–	–	33.4 ± 0.0	33.4 ± 0.0	33.3 ± 0.0	33.5 ± 0.0
	T	10.0	8.0	20.0	10.0 ± 0.0	10.0 ± 0.0	10.1 ± 0.0	10.0 ± 0.0
Illite/Smectite R0	**wt%**	33.3	–	–	33.2 ± 0.0	33.2 ± 0.0	33.2 ± 0.0	33.1 ± 0.0
	T	5.0	3.0	10.0	5.0 ± 0.0	4.9 ± 0.0	5.0 ± 0.0	5.0 ± 0.0
	Illite content	0.1	0.0	1.0	0.10 ± 0.00	0.09 ± 0.00	0.10 ± 0.00	0.10 ± 0.00
	2wat / (2wat + 1wat)	0.5	0.0	1.0	0.50 ± 0.00	0.49 ± 0.00	–	–
	2 gly / (2gly + 1gly)	0.5	0.0	1.0	0.50 ± 0.00	–	0.50 ± 0.00	–
	0 gly / (0gly + 1gly)	1.0	0.0	1.0	1.00 ± 0.00	–	–	1.00 ± 0.00

Table 7: Overview of the means and standard deviations of weight fractions and refined parameters for assemblage 2 using smooth patterns.

Assemblage #2 – Smooth patterns					**Multiple specimens** (n=50)	**Only AD** (n=50)	**Only EG** (n=50)	**Only 350°C** (n=50)
Phase	**Property name**	**True value**	**Range**		**Obtained value**	**Obtained value**	**Obtained value**	**Obtained value**
			Min.	**Max.**	**μ ± σ**	**μ ± σ**	**μ ± σ**	**μ ± σ**
Illite	**wt%**	25.0	–	–	25.0 ± 0.1	25.0 ± 0.1	25.0 ± 0.0	25.4 ± 0.71
	T	13.0	10.0	30.0	13.0 ± 0.1	13.0 ± 0.0	13.0 ± 0.0	12.9 ± 0.2
Illite/Smectite R0	**wt%**	25.0	–	–	24.9 ± 0.2	25.0 ± 0.1	25.0 ± 0.0	24.8 ± 0.3
	T	5.0	3.0	10.0	5.1 ± 0.1	5.0 ± 0.0	5.0 ± 0.0	5.0 ± 0.1
	Illite content	0.65	0.5	1.0	0.65 ± 0.00	0.65 ± 0.00	0.65 ± 0.00	0.64 ± 0.03
	2wat / (2wat + 1wat)	0.7	0.0	1.0	0.70 ± 0.01	0.70 ± 0.00	–	–
	2 glv / (2glv + 1glv)	0.7	0.0	1.0	0.71 ± 0.02	–	0.70 ± 0.00	–
	0 glv / (0glv + 1glv)	1.0	0.8	1.0	0.96 ± 0.03	–	–	0.99 ± 0.01
Kaolinite	**wt%**	12.5	–	–	12.5 ± 0.0	12.5 ± 0.0	12.5 ± 0.0	12.5 ± 0.0
	T	20.0	10.0	30.0	20.1 ± 0.1	20.0 ± 0.0	20.0 ± 0.0	20.1 ± 0.1
Kaolinite/Smectite R0	**wt%**	12.5	–	–	12.7 ± 0.2	12.5 ± 0.1	12.5 ± 0.0	12.9 ± 0.2
	T	3.0	3.0	10.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0
	Kaolinite content	0.80	0.7	1.0	0.80 ± 0.01	0.80 ± 0.00	0.80 ± 0.00	0.79 ± 0.00
	2wat / (2wat + 1wat)	0.25	0.0	0.6	0.26 ± 0.11	0.25 ± 0.02	–	–
	2 glv / (2glv + 1glv)	0.50	0.0	0.6	0.44 ± 0.10	–	0.50 ± 0.01	–
	0 glv / (0glv + 1glv)	1.00	0.8	1.0	0.93 ± 0.05	–	–	0.93 ± 0.04
Smectite	**wt%**	20.0	–	–	19.9 ± 0.1	20.0 ± 0.1	20.0 ± 0.0	19.6 ± 0.7
	T	3.0	3.0	10.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0
	2wat / (2wat + 1wat)	0.60	0.5	1.0	0.60 ± 0.00	0.60 ± 0.00	–	–
	2 glv / (2glv + 1glv)	0.90	0.5	1.0	0.90 ± 0.00	–	0.90 ± 0.00	–
	0 glv / (0glv + 1glv)	0.90	0.8	1.0	0.92 ± 0.01	–	–	0.90 ± 0.01
Chlorite	**wt%**	5.0	–	–	5.0 ± 0.0	5.0 ± 0.1	5.0 ± 0.0	5.0 ± 0.0
	T	5.0	3	10	5.0 ± 0.0	5.0 ± 0.0	5.0 ± 0.0	5.0 ± 0.0
	$\partial d_{001} \cdot 10^3$	5.0	1.0	10.0	5.0 ± 0.1	5.0 ± 0.1	5.0 ± 0.0	5.1 ± 0.1

Table 8: Overview of the means and standard deviations of weight fractions and refined parameters for assemblage 2 using noisy patterns.

Assemblage #2 – Noisy patterns					**Multiple specimens** (n=50)	**Only AD** (n=50)	**Only EG** (n=50)	**Only 350°C** (n=50)
Phase	**Property name**	**True value**	**Range**		**Obtained value**	**Obtained value**	**Obtained value**	**Obtained value**
			Min.	**Max.**	**μ ± σ**	**μ ± σ**	**μ ± σ**	**μ ± σ**
Illite	**wt%**	25.0	–	–	25.1 ± 0.2	25.2 ± 0.1	25.3 ± 0.1	24.8 ± 1.5
	T	13.0	10.0	30.0	13.1 ± 0.1	13.2 ± 0.0	12.9 ± 0.0	13.2 ± 0.3
Illite/Smectite R0	**wt%**	25.0	–	–	24.6 ± 0.4	25.8 ± 0.2	24.7 ± 0.1	25.8 ± 1.9
	T	5.0	3.0	10.0	5.0 ± 0.1	5.2 ± 0.0	4.9 ± 0.0	5.0 ± 0.4
	Illite content	0.65	0.5	1.0	0.64 ± 0.01	0.65 ± 0.00	0.65 ± 0.00	0.64 ± 0.04
	2wat / (2wat + 1wat)	0.7	0.0	1.0	0.67 ± 0.02	0.70 ± 0.01	–	–
	2 glv / (2glv + 1glv)	0.7	0.0	1.0	0.68 ± 0.01	–	0.67 ± 0.00	–
	0 glv / (0glv + 1glv)	1.0	0.8	1.0	0.96 ± 0.02	–	–	0.96 ± 0.03
Kaolinite	**wt%**	12.5	–	–	12.5 ± 0.0	12.3 ± 0.0	12.5 ± 0.0	12.6 ± 0.1
	T	20.0	10.0	30.0	20.1 ± 0.1	20.1 ± 0.0	20.1 ± 0.0	20.0 ± 0.0
Kaolinite/Smectite R0	**wt%**	12.5	–	–	12.8 ± 0.4	12.1 ± 0.2	12.4 ± 0.1	12.5 ± 0.1
	T	3.0	3.0	10.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0
	Kaolinite content	0.80	0.7	1.0	0.80 ± 0.01	0.81 ± 0.01	0.81 ± 0.00	0.82 ± 0.00
	2wat / (2wat + 1wat)	0.25	0.0	0.6	0.30 ± 0.11	0.34 ± 0.03	–	–
	2 glv / (2glv + 1glv)	0.50	0.0	0.6	0.47 ± 0.10	–	0.54 ± 0.02	–
	0 glv / (0glv + 1glv)	1.00	0.8	1.0	0.91 ± 0.05	–	–	0.94 ± 0.04
Smectite	**wt%**	20.0	–	–	20.1 ± 0.2	19.6 ± 0.2	20.2 ± 0.1	19.5 ± 3.4
	T	3.0	3.0	10.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.0	3.0 ± 0.2
	2wat / (2wat + 1wat)	0.60	0.5	1.0	0.60 ± 0.01	0.60 ± 0.00	–	–
	2 glv / (2glv + 1glv)	0.90	0.5	1.0	0.90 ± 0.01	–	0.90 ± 0.00	–
	0 glv / (0glv + 1glv)	0.90	0.8	1.0	0.92 ± 0.01	–	–	0.91 ± 0.02
Chlorite	**wt%**	5.0	–	–	5.0 ± 0.0	5.1 ± 0.1	4.9 ± 0.0	4.9 ± 0.1
	T	5.0	3	10	5.1 ± 0.0	5.2 ± 0.1	5.2 ± 0.0	5.0 ± 0.0
	$\partial d_{001} \cdot 10^3$	5.0	1.0	10.0	5.2 ± 0.3	5.5 ± 0.3	4.5 ± 0.2	5.4 ± 0.3

Table 9: Overview of the means and standard deviations of weight fractions and refined parameters for assemblage 3.

Assemblage #3 – Noisy patterns					**Multiple specimens** (n=50)	**Only AD** (n=50)	**Only EG** (n=50)	**Only 350°C** (n=50)
Phase	**Property name**	**True value**	**Range**		**Obtained value**	**Obtained value**	**Obtained value**	**Obtained value**
			Min.	**Max.**	**μ ± σ**	**μ ± σ**	**μ ± σ**	**μ ± σ**
Kaolinite	**wt%**	35.0	–	–	35.0 ± 0.0	35.3 ± 0.6	34.7 ± 0.0	34.9 ± 0.0
	T	18.0	5	40	18.0 ± 0.0	18.0 ± 0.2	18.0 ± 0.0	18.0 ± 0.0
Illite	**wt%**	30.0	–	–	30.0 ± 0.0	30.0 ± 0.8	30.1 ± 0.1	29.0 ± 0.1
	T	25.0	5	40	25.0 ± 0.0	25.5 ± 0.1	24.8 ± 0.0	25.2 ± 0.12
High-charge smectite	**wt%**	15.0	–	–	15.1 ± 0.0	16.9 ± 5.9	15.8 ± 0.1	16.0 ± 0.1
	T	10.0	5	40	10.0 ± 0.0	11.3 ± 5.2	10.0 ± 0.0	10.0 ± 0.1
	HC / (HC + LC)	0.90	0.50	1.00	0.90 ± 0.00	0.87 ± 0.06	0.90 ± 0.00	–
Low-charge smectite	**wt %**	20.0	–	–	19.9 ± 0.0	17.8 ± 7.2	19.4 ± 0.2	20.3 ± 0.2
	T	10.0	5	40	10.0 ± 0.0	12.2 ± 7.4	10.0 ± 0.0	10.2 ± 0.1
	LC / (LC + HC)	0.80	0.50	1.00	0.80 ± 0.00	0.83 ± 0.06	0.80 ± 0.00	–

Table 10: Overview of the means and standard deviations of weight fractions and refined parameters for assemblage 4.

	Assemblage #4 – Noisy patterns				**Multiple specimens** (n=50)	**Only AD** (n=50)	**Only EG** (n=50)	**Only 350°C** (n=50)
Phase	**Property name**	**True value**	**Range**		**Obtained value**	**Obtained value**	**Obtained value**	**Obtained value**
			Min.	**Max.**	**μ ± σ**	**μ ± σ**	**μ ± σ**	**μ ± σ**
Kaolinite	**wt%**	50.0	–	–	49.7 ± 0.1	49.3 ± 0.0	50.3 ± 0.2	49.3 ± 0.1
	T	15.8	5	40	15.2 ± 0.1	15.2 ± 0.0	15.2 ± 0.0	15.6 ± 0.0
Illite	**wt%**	50.0	–	–	50.3 ± 0.1	50.7 ± 0.0	49.7 ± 0.2	50.7 ± 0.1
	T	30.0	5	40	21.2 ± 0.0	18.8 ± 0.0	22.7 ± 0.1	28.0 ± 0.0
	Oct. Fe^{3+} / Oct. Al^{3+}	0.125	0	0.5	0.133 ± 0.000	0.126 ± 0.002	0.151 ± 0.001	0.139 ± 0.001
	K content	1.50	0	2	1.52 ± 0.01	1.49 ± 0.00	1.52 ± 0.01	1.44 ± 0.00

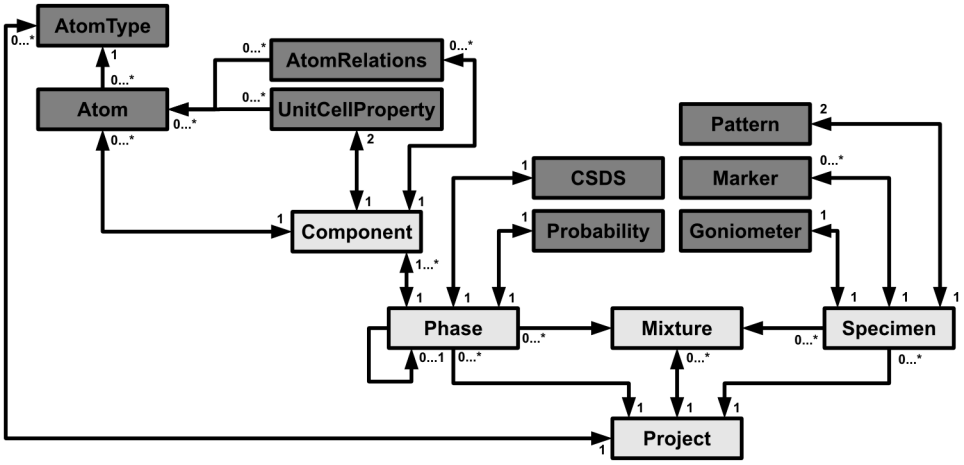

Figure 1: Schematic overview of the most important objects in PyXRD and their relations. Arrows indicate 'is referenced x times by' relations and the numbers indicate the multiplicity of that relation (e.g. Project holds 0 or more references to AtomType).

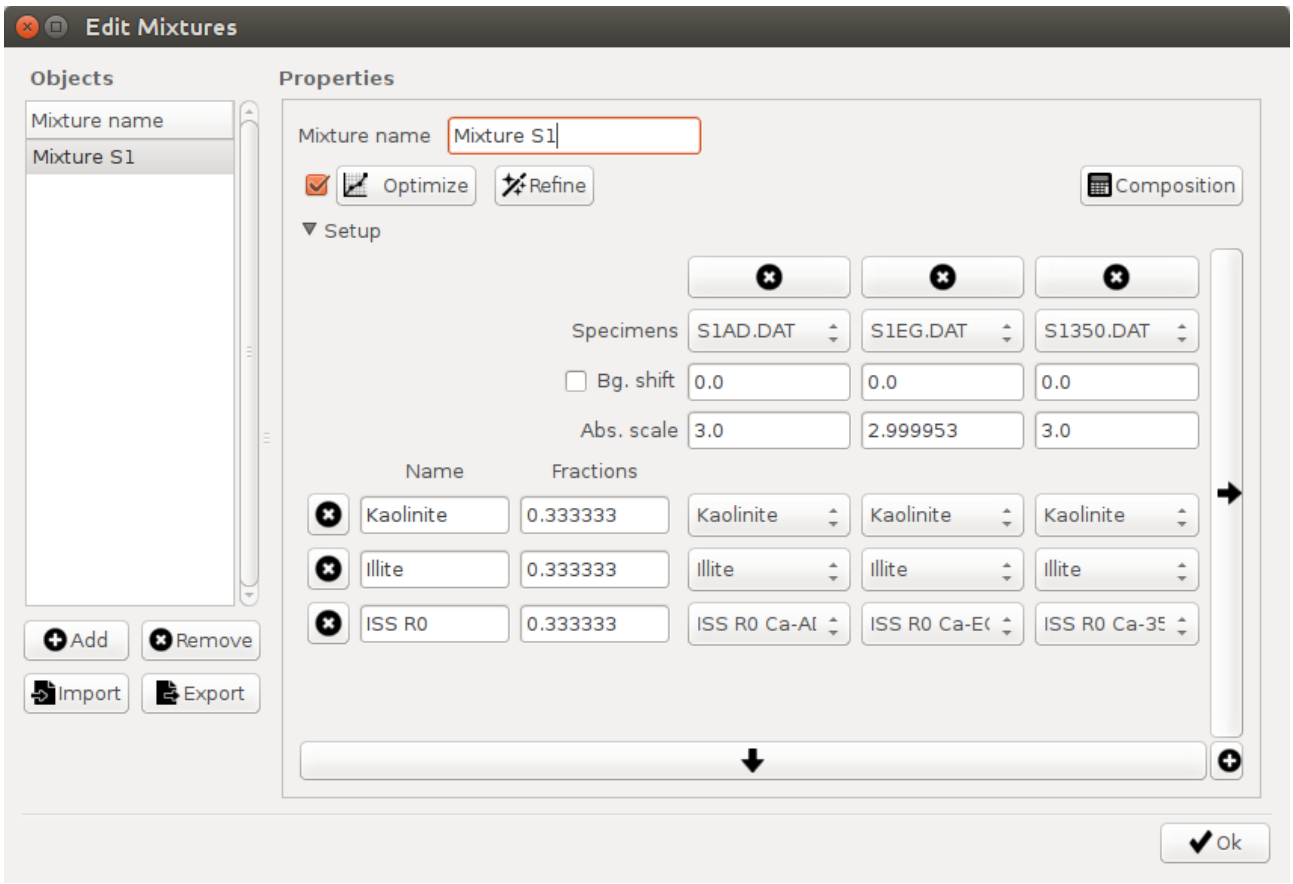

Figure 2: screenshot showing the 'Edit mixtures' dialog where a user can link different phases (Kaolinite, Illite, ISS R0 Ca-AD, ...) with the corresponding specimens (S1AD.dat, S1EG.dat, …).

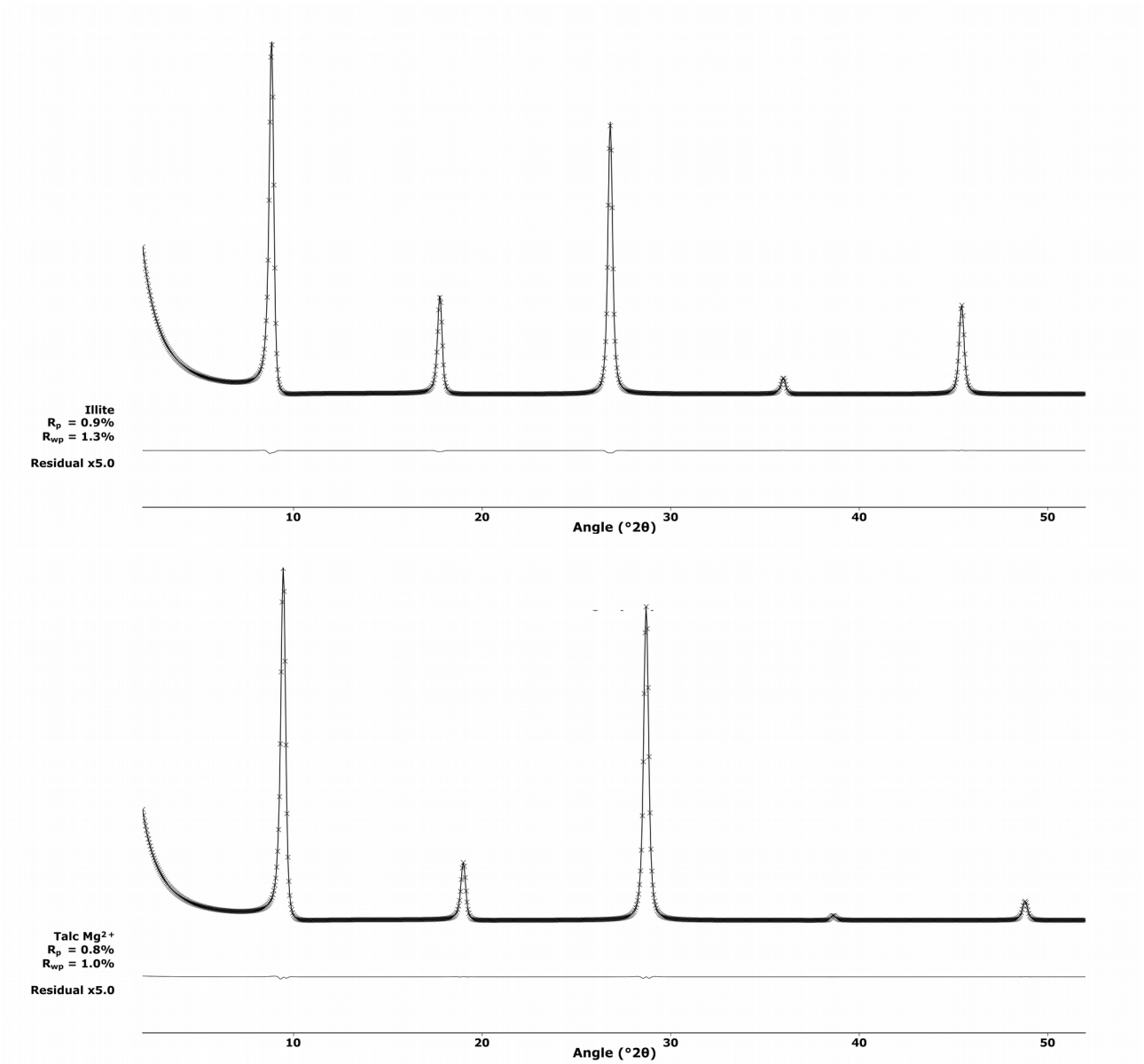

Figure 3 – Calculated patterns for discrete illite (top) and talc (bottom), showing nearly identical output for PyXRD (solid line) and Sybilla (crosses). For clarity the residual patterns are scaled to 5x their original intensity.

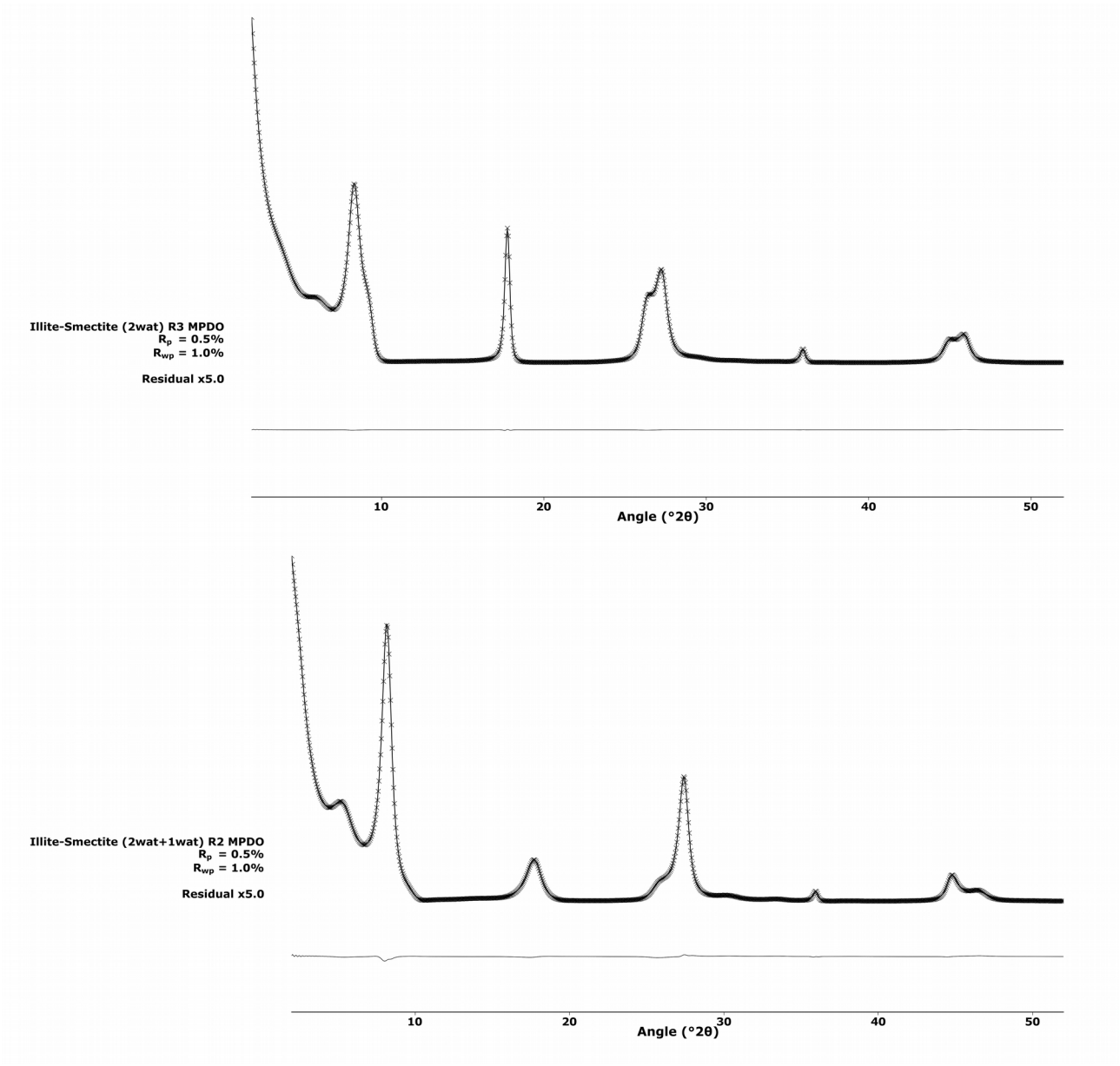

Figure 4 – Calculated patterns for illite-smectite (IS) R3 (top) and illite-smectite (ISS) R2 (bottom; both have MPDO), showing nearly identical output for PyXRD (solid line) and Sybilla (crosses). For clarity the residual patterns are scaled to 5x their original intensity.

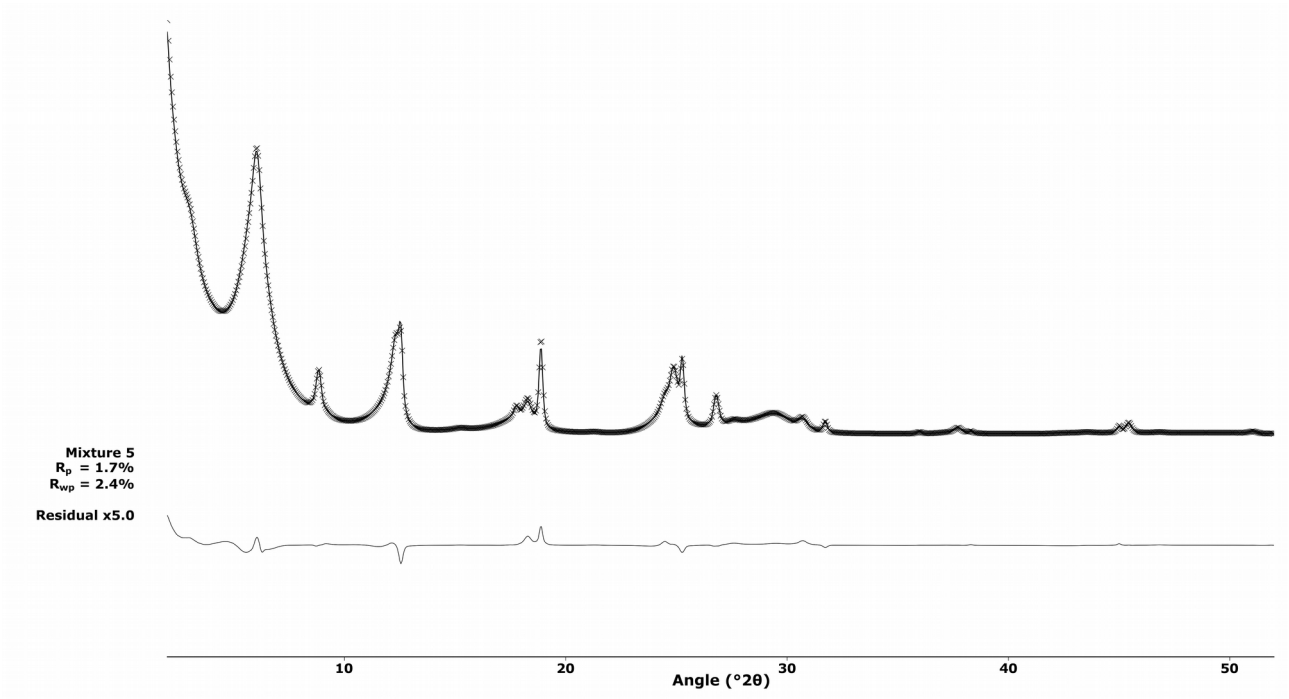

Figure 5 – Calculated patterns for mixture 5 for PyXRD (solid line) and Sybilla (crosses). For clarity the residual pattern is scaled to 5x its original intensity.

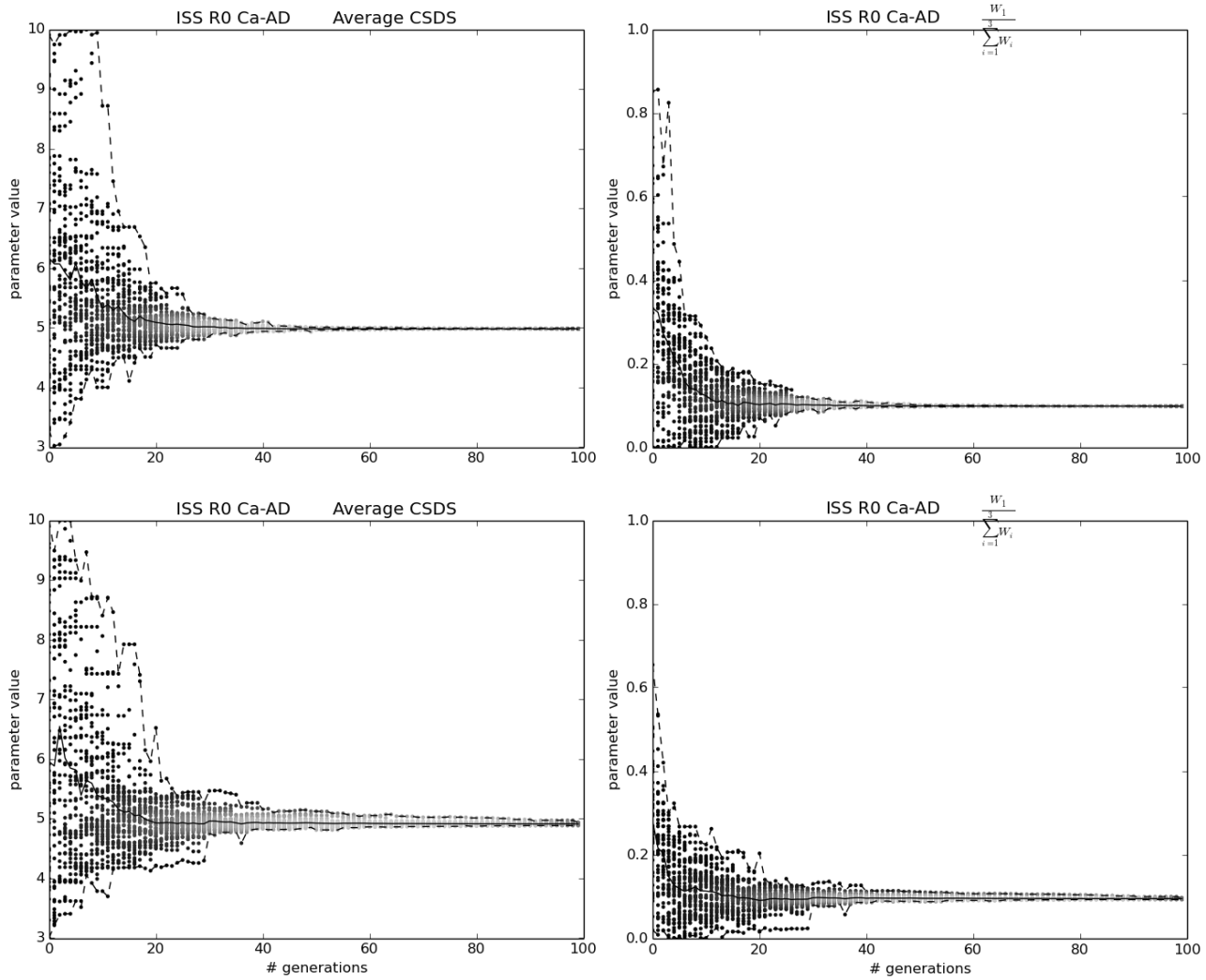

Figure 6: Parameter evolution plots (left: average CSDS; right: illite content) for the noisy patterns of assemblage 1 for the multispecimen run (top plots) and the isolated AD run (bottom plots). Minimum and maximum values during the refinement are indicated with dashed lines, iterations' best solutions at each generation indicated by dots and average solution with a solid line. The higher the density of the dots, the lighter they are colored.

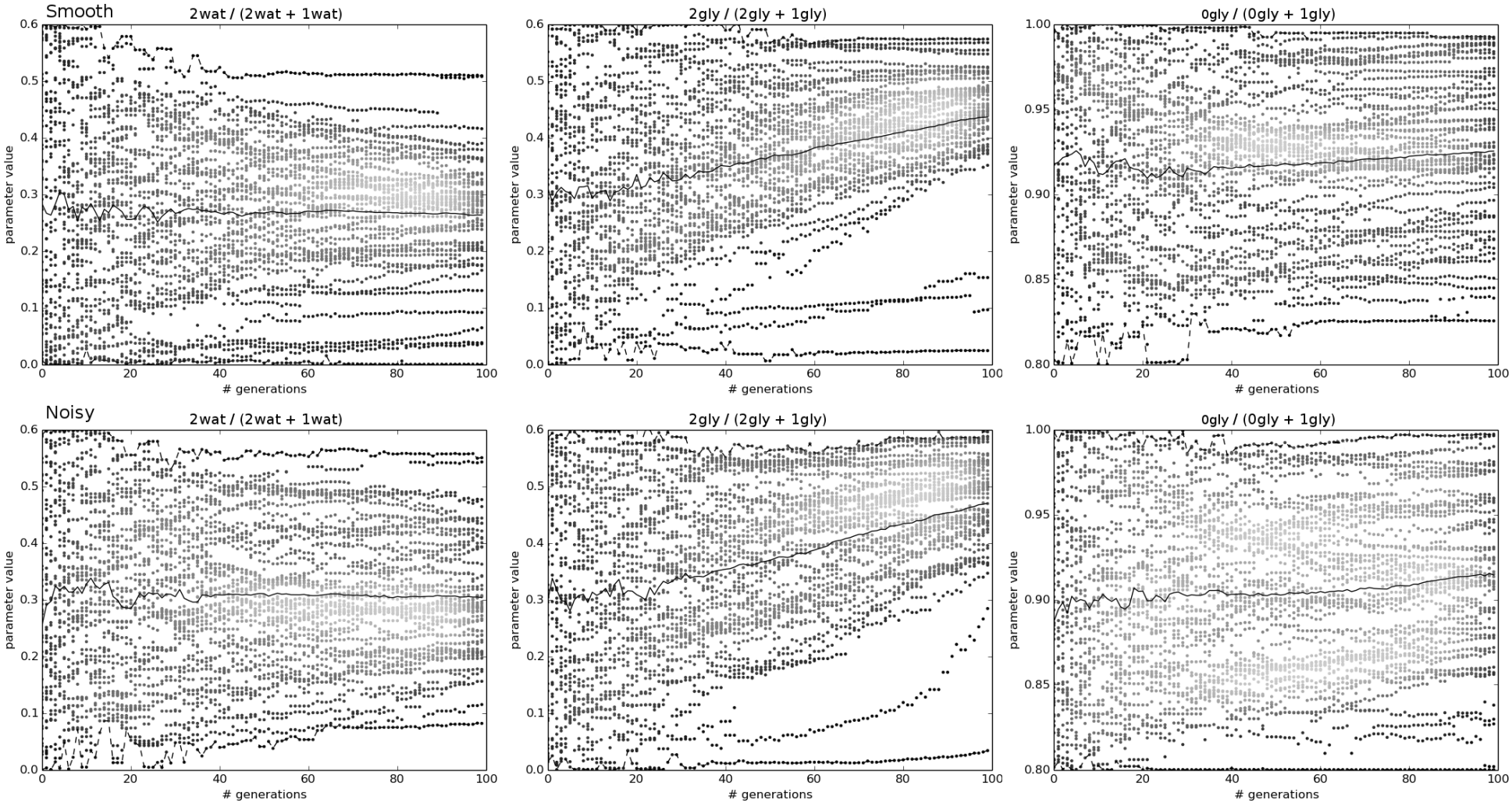

Figure 7: Parameter evolution plots for the smectite fractions in the kaolinite-smectite mixed layer of assemblage 2 using the multispecimen setup. Plots for the smooth patterns are in the top row, for noisy patterns in the bottom row. Legend as in Figure 6.

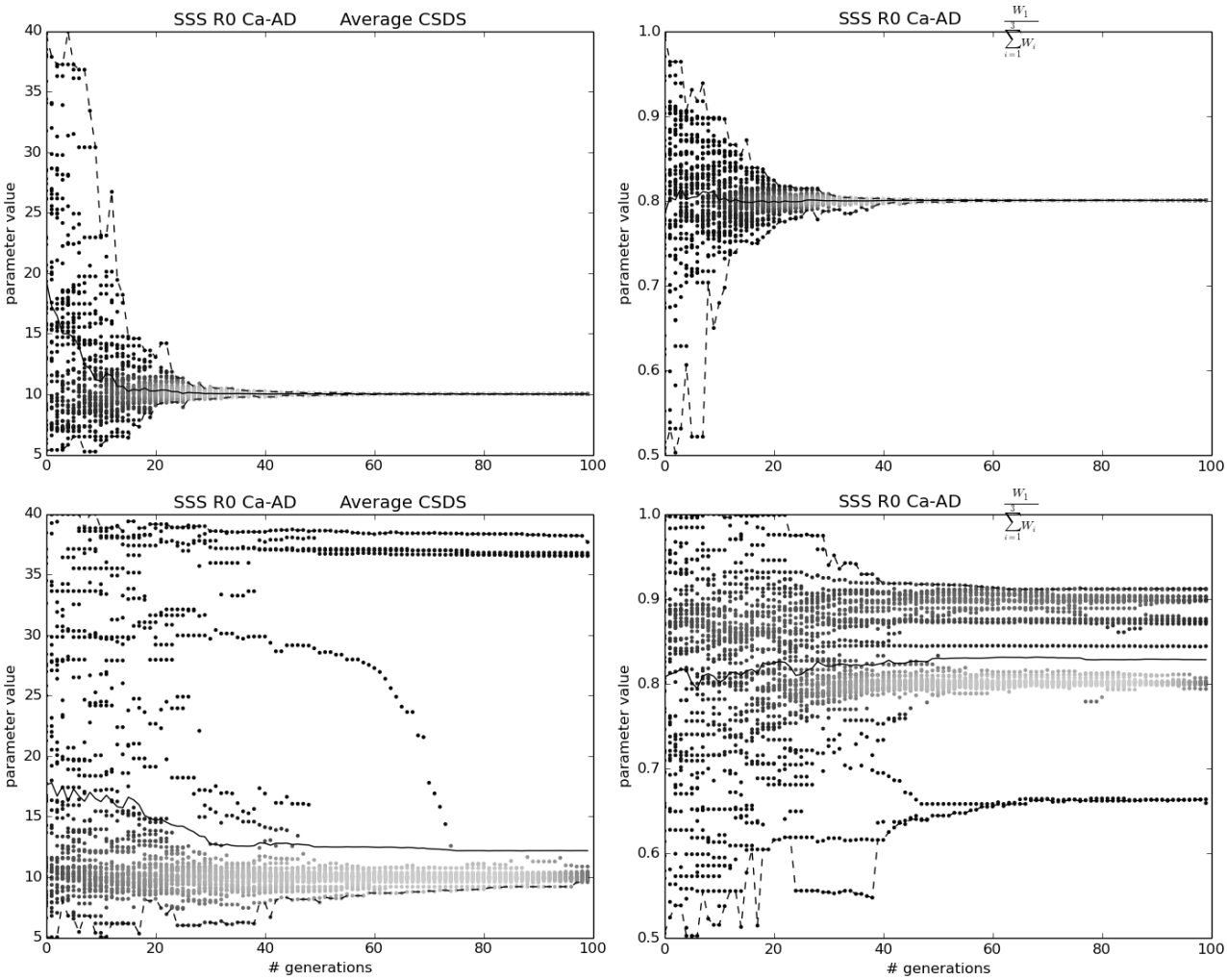

Figure 8: Parameter evolution plots for the low-charge smectite in assemblage 3. Plots for the multispecimen setup are in the top row, for the AD single pattern setup in the bottom row. Legend as in Figure 6.

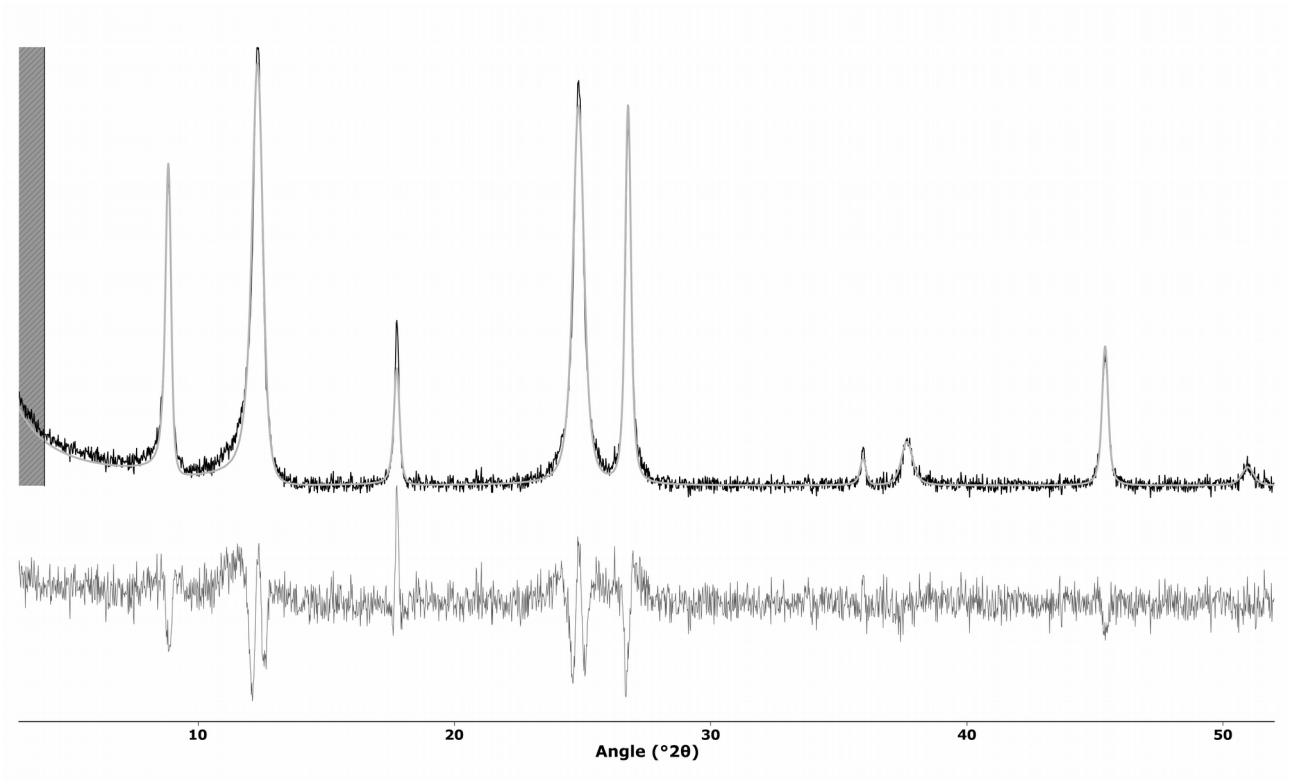

Figure 9: The input (black solid line) and refined (grey solid line) AD pattern and their difference (grey solid line at the bottom) for the multispecimen setup of assemblage 4. An observant user should see the mismatches in the patterns and realize his model needs improvement.