Authors' response

The authors would again like to thank both referees for their time and effort in reviewing our manuscript. Both referees made good and helpful points, focusing mostly on the new sections in the manuscript which has helped to further refine them. Additional attention has been paid to eliminate lingering mistakes as suggested by referee #1. Below is an overview of the comments and remarks made by the referees followed by our response and actions in italic.

Comments from referees

Referee #1

General comments:

1 Additional data provided to support the ability of the provided calculation routine to reproduce calculations performed with other simulation programs is far from being convincing. With provided illustrations it is difficult to assess the reliability of calculations for mixed layers. Results reported in Table 2 do not suggest major discrepancy for the mixing function.
On the other hand, intensity distributions calculated for periodic structures (illite and talc) show significant differences with those calculated with Sybilla. These differences reach 8% for illite 005 peak and 13% for talc 001. From the authors' words, the origin of these differences is unknown (l. 297-298)!!! The proposed explanation (different scattering factors for the ionic species) appears unlikely as to my knowledge both programs use scattering factors for partially ionized species (Si2+, Al1.5+, …) thus minimizing differences related to these parameters. In addition, differences in scattering factors should influence calculated XRD patterns in a systematic manner as a function of diffraction angle, which is not the case in the two comparisons shown in Figure 3.
In my opinion reported differences ARE significant and a convincing explanation should be sought. If all parameters are identical, identical results must be obtained before going any further.
As suggested by referee #2, we have included the weighted residual error Rwp as a further indication of the goodness-of-fit. We have also changed the layout of the relevant figures. We agree there were some peculiar differences for the illite, talc and chlorite phases and were too keen on dismissing this as a result of differences in scattering factors. To sort this out, we made another thorough check and found some explanations for these differences:

- *The Soller slit sizes in the Sybilla projects were set to 2.3° as a default value, while in PyXRD's goniometer setup this was 2.5°. Changing Soller slits to 2.3° in both programs reduced the residual error for all comparisons, however not to the extent that it solved the misfit for illite, talc and chlorite phases.*
- *The K^{1+} cation in the interlayer of illite components was erroneously set to a neutral ion in PyXRD. Changing this to K^{1+} solved most of the observed misfit, reducing the Rp factor from 1.5 % to 0.9 % and the Rwp factor from 2.1 % to 1.3 %.*
- *Since both chlorite and talc are Mg-bearing phases, we suspected the observed differences might be related to this cation. In PyXRD the Mg atom (just like O, Si, Al, Fe and other atoms in the silicate lattice) is assumed to be 50% ionized, with other words it is set as Mg^{1+}. Changing this to a fully ionized Mg^{2+} reduced the Rp factor from 5.7 % to 0.8% and the Rwp factor from 8.6 % to 1.0 %. We have communicated this to D. McCarty (Chevron ETC) in order to figure out if this is an intentional choice or not. We believe it to be more correct to assume partially (as in 50%) ionized states, but for the sake of comparison have adjusted the structural model in PyXRD for chlorite and talc layers so they match with the current structures in Sybilla. This information has been added in lines 278-281.*

In short, after making the mentioned changes, we were able to match the output from Sybilla© for all phases consistently.

2 The authors declined providing an example with a known natural mixture not to make their Ms too long. It is a pity but it is their choice. I think however that such an example would be much more useful than Figures 5-7 that all show the same thing (one is probably enough, and would even fit in a Supplementary data).
Figures 5 and 7 showcase the (sometimes limited) success of the implemented refinement algorithm visually. Although the figures might seem similar at first glance, we still believe they can help to make our observations in the manuscript more clear. We do believe that a

comparison is very useful indeed and we are working on this, but it will not be part of this paper.

3 Overall, the conclusions as to the ability of the proposed refinement algorithm to determine relevant structural (and quantitative) parameters from a single XRD pattern has been modified. The conclusion still remains essentially unchanged in the Abstract and in the Summary however, and this should be modified.
Changes made accordingly - lines 467-473.

4 In addition, I think that modifications brought to the Introduction section make the resulting text messy and imprecise. I do NOT agree that added sections help "to avoid any confusion".

4.1 The first reason for this is the emphasis set by the authors on QPA (quantitative phase analysis) of clay-bearing samples which appears to be their main objective (l. 50-51). I have no problem with this emphasis, but an obvious question one may ask is the validity of the sole analysis of oriented clay-size fractions for this purpose, compared to the analysis of bulk samples.
We do not mention or suggest one should only analyse oriented clay-size fractions. On the contrary, we explicitly wrote "two complementary methods exist to analyse clay minerals using X-ray diffraction.", and continue to refer to powder X-ray diffraction which can be performed on either bulk samples or sub-fractions. We do not even mention 'clay fraction' except when referring to literature.

4.2 "Next, the authors claim that computation of XRD data corresponding to a structure model is the best solution for the proposed QPA, which is correct. What is not is that "structural and compositional information" is not a by-product of these calculations as suggested (l. 59-60) but the primary information derived from these calculations. [...]"
We do not want to suggest the structural and compositional information is a by-product of the calculations. We have re-phrased this sentence to make this more clear.

4.3 With this respect, the description of the multi-specimen method as dedicated "originally […] for the qualitative evaluation of XRD patterns"(l. 102-103) is again misleading. The objective of this method was (and still is) the determination of structural models for investigated samples. Consistently, the first step of these models is to determine the nature of the different phases (periodic and/or interstratified) present in the sample (and the multi-specimen approach is especially well-suited for this purpose to remove possible ambiguities arising from the analysis of a single pattern). The next step is to refine the composition (chemical composition, proportion of the different layer types and their stacking mode). Finally, relative proportions of the different phases are refined to fit the data. Collation of structural models derived from different specimens of the samples (including relative proportions of the different phases) allows assessing the validity of the proposed models.
We agree this was misleading, as we did not mean to imply the multi-specimen method was originally a qualitative evaluation. We meant to say the different treatments (saturations, glycolation, heating, etc.) are used as such in qualitative evaluations. We have made changes to make this clear.

4.4 Finally, wording used throughout the Ms brings additional confusion. For example "programs" and "models" are used to describe the same thing (l. 72, 73). On the other hand "model" is used to describe the calculation routines (l. abstract, 73, 81, 318, …), structural models (l. 101, 110, 114, 343, 345, …), specific parameters of the structural model such as the set of junction probabilities (l. 276, 287, …).
In computational science 'model' and 'program' are indeed often used interchangeably when referring to the actual software, while the underlying mathematical concepts are also referred to as being a model. We have scanned the manuscript and removed this ambiguity by using the word 'program' when referring to a specific model implementation (e.g. PyXRD, Sybilla, NEWMOD), and 'model' when referring to the

underlying concepts, or part of a program (e.g. probability models).

5 More generally, I think that the text, and specifically the Abstract, Introduction, Summary, and Conclusions sections should be extensively edited aiming at a precise and unambiguous writing. I am giving several examples of ambiguous, erroneous, imprecise sections below, but it is the author's responsibility to produce a high-quality Ms. With this respect, reading the Ms before submission could avoid a reference list with references not cited in the text (Molina Ballesteros for example), not sorted (the 1st 4 citations), incorrect figure captions (Captions of Figure 6 & 7 should refer to Figure 5 not 3), not to mention typos.
We apologize for submitting a revised manuscript with some obvious errors we overlooked, and would like to thank the referee for pointing them out. Changes were made accordingly.

Minor remarks:

1 l. 1: Remove comma after 'disordered"
Changes made accordingly.

2 l. 15: "calculation routine" rather than "model". More generally the authors should try to make a consistent use of "model", "program", … throughout the Ms to avoid ambiguities.
Changes made accordingly. Also see general remarks.

3 l. 16: "formalism" rather than "background"
Changes made accordingly.

4 l. 19: allows modelling
Changes made accordingly.

5 l. 24: "one refinement…and another employing…"
Changes made accordingly.

6 l. 28: all minerals present
Changes made accordingly.

7 l. 65: Modelling preferred orientation is one of the key to an accurate QPA from randomly oriented powders. Assumption of a perfectly random orientation is almost never made.
Changes made accordingly.

8 l. 67: "occurs mainly" as again orientation is not perfect, the deviation being accounted for by the sigma-star parameter.
Changes made accordingly.

9 l. 71: Only part of the stacking (dis)order (that along the c* axis) is accounted for.
Changes made accordingly.

10 l. 74-77: I understand that this is just a list of examples, but modelling of XRD patterns from interstratified samples has been developed in the late 1960's and early 1970's by several groups. It would be a good idea to acknowledge these pioneering works. More recently codes such as DiffaX have been used also for clay minerals.
Changes made accordingly.

11 l. 95 "limits, or by choosing starting values close…"
Changes made accordingly.

12 l. 106-107: Incorrect assertion. Hydration heterogeneity has been shown to be intrinsic to "swelling" clay minerals and not to directly depend on layer charge.
Several papers have been published indicating a relation between the amount and position of layer charges and the hydration behaviour of swelling minerals. Indeed, differences in layer charges are usually distributed heterogeneously within the crystallites, leading to a heterogeneous swelling behaviour. This is a type of disorder, can thus be modelled and can be used to identify different types of swelling minerals. We have re-phrased this section to make this more clear and added references to a few recent contributions in this context.

13 l. 132-133: "multispecimen"
Changes made accordingly.

14 l. 220-222: incomplete sentence
Changes made accordingly.

15 l. 263: “13 … phases”
Changes made accordingly.

16 l. 266: “this paper’s supplement” or “supplementary data”
Changes made accordingly.

17 l. 397-399: If relative proportions of the different smectite layers have a limited influence on calculated XRD patterns, additional iterations will not improve the convergence, that will remain hampered by this low sensitivity, and thus by the related large esd’s on these parameters.
We agree that the low sensitivity is the underlying problem, although the parameter evolution plots did show a partial convergence which might further improve with additional iterations. However, this is somewhat speculative and not very relevant, so the sentence has been removed.

18 l. 457: meaning unclear. What is an experiment? A sample? An additional XRD pattern from the same sample?
Changes made to clarify.

19 l. 457-462: I don’t agree. Listed parameters (essentially occupation factors or event atomic coordinates for interlayer species) will always be adjustable parameters, as in all fits to XRD data. The multispecimen approach never aimed at improving this aspects but rather at providing additional constraints to the junction probability parameters.
We are not stating these parameters are not adjustable, just that a user needs to be careful in choosing which parameters are kept identical across treatments or 'specimens'. We have re-phrased the paragraph to clear this up.

20 l. 490-491: references should formatted, sorted, and the presence of relevant in-text citations checked.
Changes made accordingly.

21 Table 1: The set of parameters listed for phase #8 is not sufficient to describe it.
This probability model is restricted in PyXRD, these restrictions have been added to the table.

22 Figures 6-7: Captions should be modified to take into account additional figures 3 & 4.
Changes made accordingly.

Referee #2

General comments:

3.1 Model validation using discrete phases

1. The title of this paragraph should be changed because this paragraph deals with the comparison of calculated XRD patterns from PyXRD and Sybilla © . In addition, the comparison does not only consider discrete phases, comparison of calculated mixed layer minerals is also performed. The title could be: “Comparison between Sybilla © and PyXRD results: calculated 00l reflections for single discrete clays and MLMs”.
   We agree that the title was not very well suited. We have changed it to “Comparison between Sybilla© and PyXRD results: calculated 00l reflections for single discrete and mixed-layer phyllosilicates”
2. Among the clay minerals which were chosen for this comparison, the MLMs and their structural characteristics should be modified in order to be more relevant. Indeed, authors have used the default setting of the different illite-smectite MLMs that are proposed in Sybilla © and in some cases studied, the interest of these examples is still limited. For instance, XRD calculated pattern of R1 illite-smectite (Wi = 0.6; Pss = 0.5) is very closed to the R0 illite-smectite (Wi = 0.6; Pss = 0.4) and also does not largely differ from the R0 illite-smectite chosen (Wi = 0.5; Pss = 0.5). In addition, the choice of the default setting from Sybilla © is also limited for MLMs with higher Reichweite (R2 and R3) because of the too low mean CSDS of 4 that limits the influence of the stacking order (R3 or R0 MLMs illite-smectite have quasi-similar XRD calculated profiles with mean CSDS of 3). Thus the parameters of the MLMs chosen need to be modified and I suggest presenting R1, R2 and R3 MLMs illite-smectite with mean CSDS of 10 or more with MPDO (maximum possible degree of ordering) for R1.
   For the higher Reichweite minerals we have added comparison MPDO models with a higher mean CSDS value of 15.
3. In addition, I would suggest exposing the calculated XRD patterns for these MLMs in a figure. The XRD patterns calculated with PyXRD (solid line) could be plotted over the XRD patterns calculated using Sybilla © (crosses) with the difference plot below. The figure 4 presented in the revised manuscript could be also improved (I do not obtain with Sybilla © the same calculated XRD pattern for the R2 illite-smectite MLM presented in figure 4 that motivate the presentation of calculated XRD patterns).
   Figures have been adjusted accordingly. We have checked the output for the R2 illite-smectite MLM and obtained the same pattern in Sybilla.
4. In the Table 1, the authors give the residual error (Rp) obtained between XRD patterns calculated with PyXRD and Sybilla © . Although quite low, they remain significant, particularly for the talc with a Rp equal to 6.4%. This value is higher than Rp obtained for the comparison of experimental and calculated XRD profiles on natural samples and consequently need more explanation in the text or a new calculation. The explanation in the manuscript l294-298 does not support this difference otherwise such difference should be present on each XRD calculation exposed.
   See answer to referee's general comment #1.
5. In addition, I suggest to the authors to also calculate the Rwp (weighted profile R-factors; Howard and Preston, 1989) that could be more relevant in the present case.
   Howard, S.A. and Preston, K.D. (1989) Profile fitting of powder diffraction patterns. In D.L. Bish and J.E. Post, Eds., Modern Powder Diffraction, 20, p. 217–275. Reviews in Mineralogy, Mineralogical Society of America, Chantilly, Virginia.
   Changes made accordingly.

3.2 Model validation using assemblages

6. As for the part 3.1, I would like to suggest a title. The title could be: “Comparison between Sybilla © and PyXRD results: calculated 00l reflections for mixtures of discrete clays and MLMs”.
   We again agree that the title was not very well suited. We have changed it to “Comparison between Sybilla© and PyXRD results: calculated 00l reflections for mixtures of discrete and mixed-layer phyllosilicates”
7. The authors mention that Sybilla© “does not have an easy way to calculate a pattern for a certain mixture of phases (or in any case this feature is unknown to the authors), while PyXRD does” (l305-306).This is possible using Sybilla© however, I agree with the authors that an easy way does not exist. […] Using this procedure, the authors should directly compare the output from Sybilla© with the output of PyXRD for mixture and improve the validation test for the weight fractions.
   Upon trying this procedure (exporting a certain mixture as an XML file and re-importing it), we noticed Sybilla did not reproduce the same pattern after the import, even without changes to the exported XML file. Significant differences up to 5% Rp were observed. We also noticed that the mean CSDS value in the XML file was not the same as the value entered in the user interface. Correcting this value in the XML file eliminated a large part of the difference, but not everything. We have communicated this problem to D. McCarty (Chevron ETC) in order to resolve this problem. As a result, we have still resorted to manually adjusting the weight fractions so they are close enough to the values in PyXRD.
8. In addition, I suggest to the authors to give in Table 2 the Rp and Rwp values between the XRD patterns of mixtures calculated with Sybilla© and PyXRD.
   Changes made accordingly.
9. Moreover, I suggest adding a figure that show the comparison between the calculated XRD patterns from the two software as mentioned above for the part 3.1.
   Changes made accordingly.

Minor remarks:

1 L31-32: “... might be sufficient”, precise for the cases studied.
Sentence was dropped as it is speculative.

2 L34: remove “very” before useful
Changes made accordingly.

3 L101-102: add references, other authors have used this approach or precise “and reference therein”
Changes made accordingly.

4 L118: “... or at least more robust)...”. This part of the sentence is not necessary. I suggest to remove it.
Changes made accordingly.

5 L130: replace “model” by software
See comment #4.4 by referee #1.

6 L251: add a space between strategy and algorithm
Changes made accordingly.

7 L291: replace “diffrerent” by different
Changes made accordingly.

8 L384-385: “these results...excellent results”. I suggest to remove this sentence that is not necessary.
Changes made accordingly.

9 L398: remove a space after “strategy”
See comment #17 by referee #1.

10 L460-462: “If...more parameters”. This sentence is unclear. Remove this sentence or

precise the idea.
Changes made accordingly.