Referee #1

General comments:

1. The first requirement of a new routine allowing the calculation of XRD patterns from complex mixtures of disordered/interstratified phyllosilicates is consistency with available routines. Although these routines may not all be easily accessible, some of them are and I think an essential step would be to use to developed program to fit XRD patterns calculated with a different calculation algorithm (for a complex mixed layer) to make sure that ALL structural parameters (for both phases and components) are introduced and considered in the right way in the new algorithm.
   We have included a comparison with output from Sybilla for 13 discrete (pure and mixed-layered) phases and 5 mixtures to confirm PyXRD is producing similar results. We have chosen to restrict the comparison only with this model because of the similar mathematical background, resulting in an almost identical parametrisation. We also think the manuscript would become too bulky by adding more comparisons.

2. In addition I think that refinement of at least one experimental dataset (for example a known mixture of pure clays, or published datasets with determined structural parameters) would be a convincing evidence for the performance of the reported program.
   Although we agree this would be a very good example, we did not include it to prevent the article from becoming too bulky. PyXRD has been used on soil samples in a previous paper (Dumon et al. 2013. Quantitative clay mineralogy of a Vertic Planosol in southwestern Ethiopia: Impact on soil formation hypotheses, Geoderma) showing it can be used for this purpose. We also believe the very similar results obtained in the comparison with Sybilla and the numerous papers using this model serve as an indirect proof for the performance of PyXRD.

3. A second important feature of the multi-specimen approach is its ability to deal with swelling layers which may behave differently depending on experimental data collection conditions: A mono- or bi-hydrated layer in the AD state may for example incorporate 1 or 2 planes of ethylene glycol upon solvation. From the Tables, it seems that all swelling layers are considered as a unique layer type in PyXRD, but precisions are clearly needed with respect to this major and specific aspect. This is especially important for ordered mixed layers (Reichweite parameter >= 1, which are not dealt with in the reported examples) because of the implications on junction probabilities.
   PyXRD indeed supports differently swelling layers by adding several components with different hydration or intercalation states. The weight fractions and junction probabilities for these components can either be set to be identical to each other across the different states or set to different values. This can be done on a parameter-level. This means it is for example possible to have an Illite/Smectite mixed layer represented by two phases, one for the AD and one for the EG state both having two types of smectite. You can then share the weight fraction of illite among the AD and EG state, but have different amounts for the two smectite layer types. We have added some more information in the section on the model structure to clarify this important aspect.

4. Third, I think that the conclusions as to the ability of the proposed refinement algorithm to determine relevant structural (and quantitative) parameters from a single XRD pattern is misleading and should be modified. The main objective of the multi-specimen approach, as described in the abundant devoted literature is NOT to determine better or more accurate parameters but rather to release possible identification ambiguities from different structure models leading, for one of the data collection conditions, to similar XRD patterns. If this was not the case, the multi-specimen approach (and thus the present program) would be

essentially useless, the more common and faster refinement of a single pattern being sufficient… Starting from computed XRD patterns, the ambiguity is easily overcome but this may not be the case when dealing with natural samples. Accordingly, all sections (including conclusion and the abstract) dealing with this aspect should be re-written. I agree with the authors however that refinement of a single pattern may be sufficient once possible identification ambiguities are released (by using the multi-specimen approach).
We did not claim the multi-specimen method was originally intended to determine better or more accurate parameters, but we wanted to see if it is possible to use it in this way. We have re-formulated the hypothesis we are testing and the conclusion to avoid any confusion on this aspect and modified the relevant parts in the manuscript where needed.

5. Finally, reference list is far from being complete. XRD profile modeling has developed significantly over the last two decades because of the increased availability of computing resources and calculation routines. Most of the latter had however been developed and used in the 1970's, and it would be reasonable to cite these pioneering works. It is for example striking that Newmod is cited as one of the available calculation routines with no reference to its original author (R.C. Reynolds, Jr.).
We agree the reference list in the manuscript was a bit lacking. The reason for this was that the complete mathematical deduction is given in the manual. We also feel it is not neccesary to add these deductions in the manuscript, because it would make it too long, and we are not presenting a new development on that front. However, we have added more references in the manuscript including the specific ones mentioned by the reviewer.

Minor remarks:

6. p. 2497: Avoid acronyms in the title
Changes made accordingly
7. p. 2499, l. 1-4: References needed
Changes made accordingly
8. p. 2499, l. 8-11 / p. 2499, l. 21-25: Additional (older) references could be added. In particular reference to the original modeling works of Reynolds (for Newmod), and of the Russian group (Drits and Sakharov) from early 1970's could be included.
Changes made accordingly
9. p. 2500, l. 16-17: References to Meunier and Lanson are not relevant (at least not as presently written) here as they essentially review existing literature.
Changes made accordingly
10. p. 2500, l. 27-28: This statement contradicts the conclusions of the article (as the authors show that it is possible to obtain equally good structural/quantitative determination from a single XRD pattern). From the previous lines, one interest would be to obtain a faster refinement and an improved consistency of the structure models derived from different XRD patterns.
We did not state this is the case, rather we check this hypothesis. We have made some changes to these lines to clarify this.
11. p. 2502, l. 17-18: Probably not necessary to consider ionic species the effect being strongly correlated with thermal motion.
We agree it might not influence the calculation, but include this information anyway.
12. p. 2503, l. 10-12: This is wrong as Newmod also includes uniform (or custom) distribution, MLM2C/3C Ergun's distribution, …
We fail to see what the reviewer is considering wrong here. The lines he refers to are dealing with distributions implemented in PyXRD, not in other models. We are not stating the other models only have these implementations.
13. p. 2503, l. 18: Projection is along c* not c
Changes made accordingly
14. p. 2504, l. 12-14: This possible constrains appear similar to those that were considered

inadequate for other programs (see for example p. 2500, l. 8-11)
We did not claim these constrains are inadequate, we claim the models are not constrained well enough for an automatic refinement without them. We have made changes to clarify our statement.

15. p. 2505, l. 9-20: I am not sure all acronyms are necessary especially as they are (very) seldom used in the rest of the article. Remove acronyms.
Changes made accordingly
16. p. 2507, l. 5: Why not just consider statistical counting noise [sqrt(I0)]
Because the patterns considered are not actual measured patterns and are not expressed on an absolute scale (often with decimal values instead of integer counts). We agree this would be the logical approach for real life data.
17. 2507, l. 8: Such a noise level corresponds to ∼40000 counts which is seldom achieved experimentally on mixed layers.
If you would consider statistical counting noise this would indeed translate to this level of counts. However we also wanted to include detector noise (which we acknowledge has improved greatly with recent detectors, but can still be an issue for older XRD equipment) and noise resulting from sample fluorescence, which used to be a problem for us when working on Fe-rich samples with a Cu X-ray source. Modern detector technology can now largely overcome this problem. We have made some small modifications to clarify the motivation behind this.
18. p. 2508, l. 10-11: Systematic discrepancies should be described and an explanation sought.
Changes made to accordingly.
19. p. 2509, l. 1-2: Meaning unclear.
Changes made to clarify.
20. p. 2510, l. 26-27: Again, XRD profile modelling was used before 2010!
We never stated this is the case, and believe it is more relevant to give recent examples, illustrating the currently used approaches and methods.