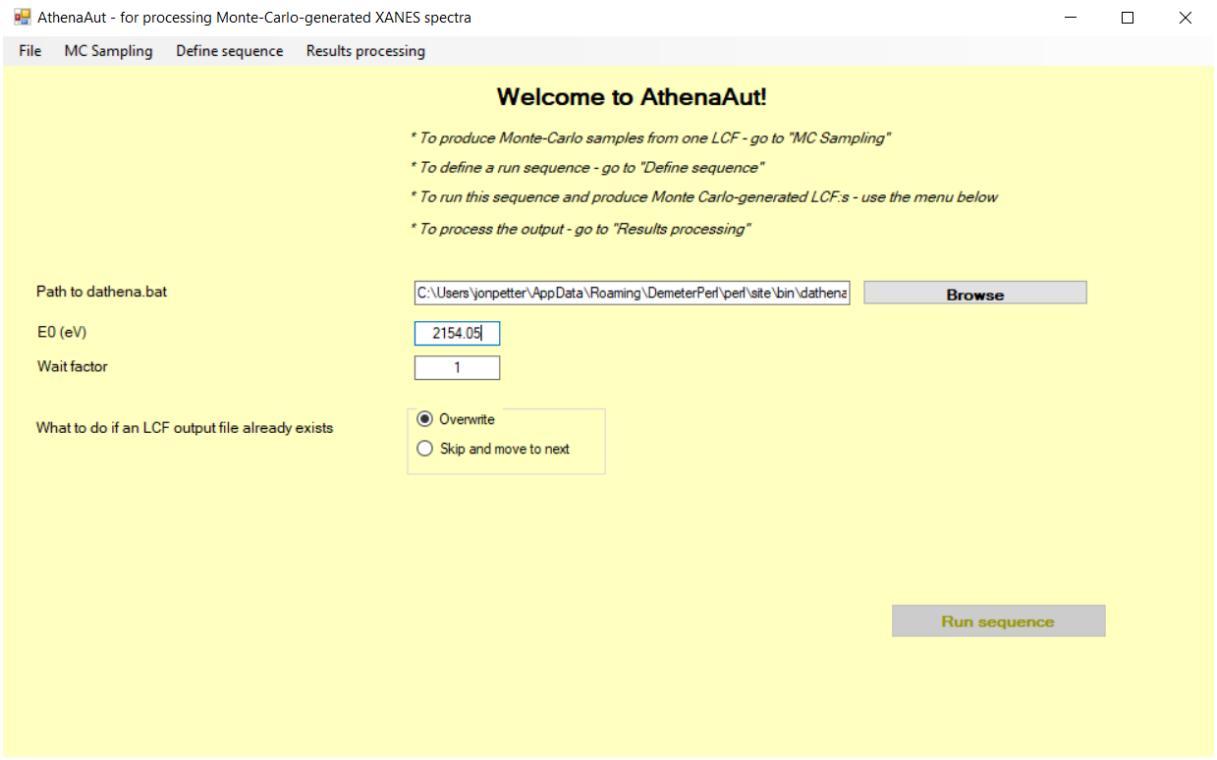

AthenaAut

Uncertainty analysis of linear combination fits from Athena

A manual

Jon Petter Gustafsson

Department of Soil and Environment, Swedish University of Agricultural Sciences, Uppsala, Sweden

e-mail: jon-petter.gustafsson@slu.se

April 2020

Contents

What does AthenaAut?

AthenaAut is used for uncertainty analysis of a linear combination fit that has already been performed in Athena, and saved as an .lcf file with normalized absorption intensities for a sample and for a fit. AthenaAut can consider uncertainties both in the normalization and in the acquired energy; however, other factors such as noise are not considered at present. AthenaAut performs the following tasks:

a) it produces a number of Monte-Carlo generated spectral variants (as normalized absorption intensities) and save these as .CSV files for later input to Athena
b) for each variant it runs Athena externally (using simulated mouse clicks) to produce outputs in the form of new .lcf files
c) it reads these outputs, exports them to Excel, groups them and provides averages and standard deviations of the outputs.

Consistent with the above, the user will typically apply the following workflow for the uncertainty analysis of an LCF:

1. Monte-Carlo-type generation of spectral variants using the menu “MC Sampling”
2. Defining a sequence of Athena runs using the menu “Define sequence”
3. Calling Athena for all the runs using the main menu
4. Retrieving the results and the statistics using the menu “Results processing”

What you need to do before using AthenaAut

1. Computer settings

- Windows 8 or later is required. Whether it is possible to run AthenaAut in a VM for Mac, for example, is not anything I can answer, users interested in this need to explore this possibility on their own, alternatively modify and compile the source code in another OS themselves.

- To run AthenaAut successfully you are advised to use a computer with relatively high memory capacity and processor speed. Further, for the best result, do not run other programs in the background, and remove unnecessary programs from Autostart.

- You will need to adjust your screen resolution to 1920×1080 and to change the size of "text, apps and other items" to 125% (these settings are usually found among the Display settings)

- To be able to use the features for "Results processing" you need to have Microsoft Excel 15.0 or later installed. Further, AthenaAut stores files in your Documents folder, which needs to be available.

- AthenaAut requires .NET Framework 4.5 or later, if this is not installed on your computer, AthenaAut will not start.

- To be able to group the standards, you will need the text file "Standardgroups.txt", which should accompany the main AthenaAut program. The text file should be present in the My Documents folder.

2. Settings in Athena

- Athena versions: 0.9.025 or 0.9.026. Please note that the software was developed with 0.9.025, so far I have rather limited experience with 0.9.026, and none at all with earlier versions (0.9.024 and older). Therefore I recommend use of 0.9.025 for the best possible result. To download ver. 0.9.025, please follow this link: https://bruceravel.github.io/demeter/#old

- Check that your version of Athena can import a scan successfully. If not you may need to enable one of the plugins (accessible from the "Plugin registry")

- Some of the LCF parameters need to be changed under "Preferences". First, expand "lcf" under "Parameters". First, "nstan" (number of rows that will be displayed in the list of standards…) should be changed to 20. Second, under "unity", untick the box for "unity" to make sure that the weights are not forced to sum to one. Finally, you may want to change the value of "emin" (default lower bound of a fit in energy), I

normally use -10 instead of -20. After each change, click “Apply and save”. If Athena freezes when you do this, exit the program and enter again – normally Athena will have completed the change you made.

- Before running a sequence of input files for Athena (by clicking Run on the main menu) you will need to prepare Athena so that it knows what columns to import and that it expects normalized scans in the input. To do this, import a .CSV scan into Athena that you have generated with Monte-Carlo sampling (see next section, “Monte-Carlo sampling”). Athena will then display the “Column selection” menu. On this menu, the energy should be imported from column no. 1, whereas the Numerator is imported from column no. 2 (leave the box for “Denominator” unticked). The alternative for “Data type” should be “norm(E)” while “Energy units” should be “eV”. No reference channel should be imported. Then click OK to import the scan. Athena will then remember these settings until you import the next scan.

- To run Athena from AthenaAut you should have prepared Athena project files that contain the standards you include in the LCF. It is important that all the standards are ticked (see Fig. 1)

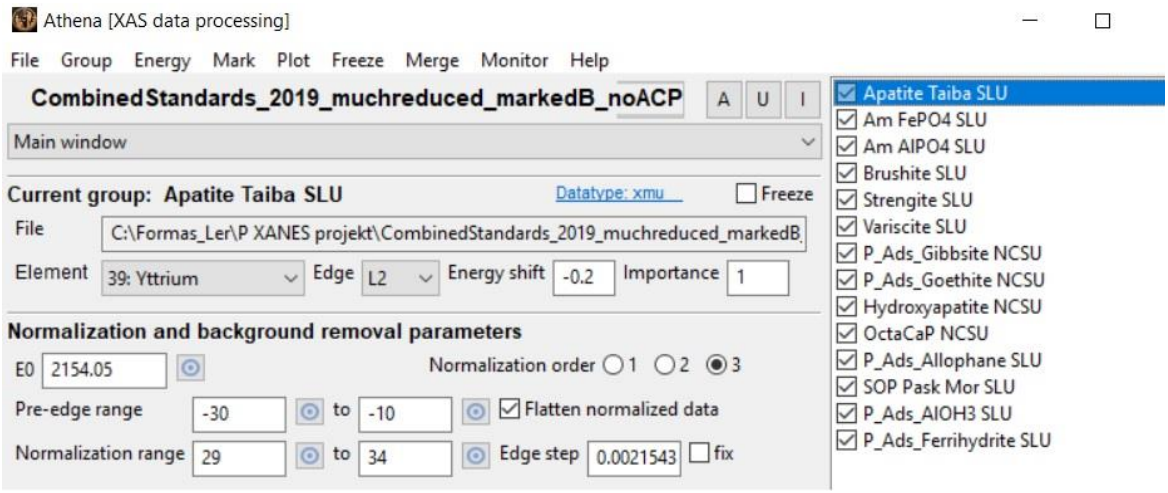

Fig. 1. This is what an Athena project file with standards should look like in the main window of Athena.

3. Preparing LCF files

- When you have completed an LCF in Athena, use the option “Save fit as column data”. This will produce an ASCII text file with the suffix .lcf, which is the file that AthenaAut reads.

4. Compilation and availability of the executable .EXE file

VB.NET source code files are available for those who want to modify or recompile the code. If you have Microsoft Visual Studio 2012 or later you can incorporate the code project files into a new Visual Basic .NET project and modify/compile the code yourself.

Monte-Carlo sampling

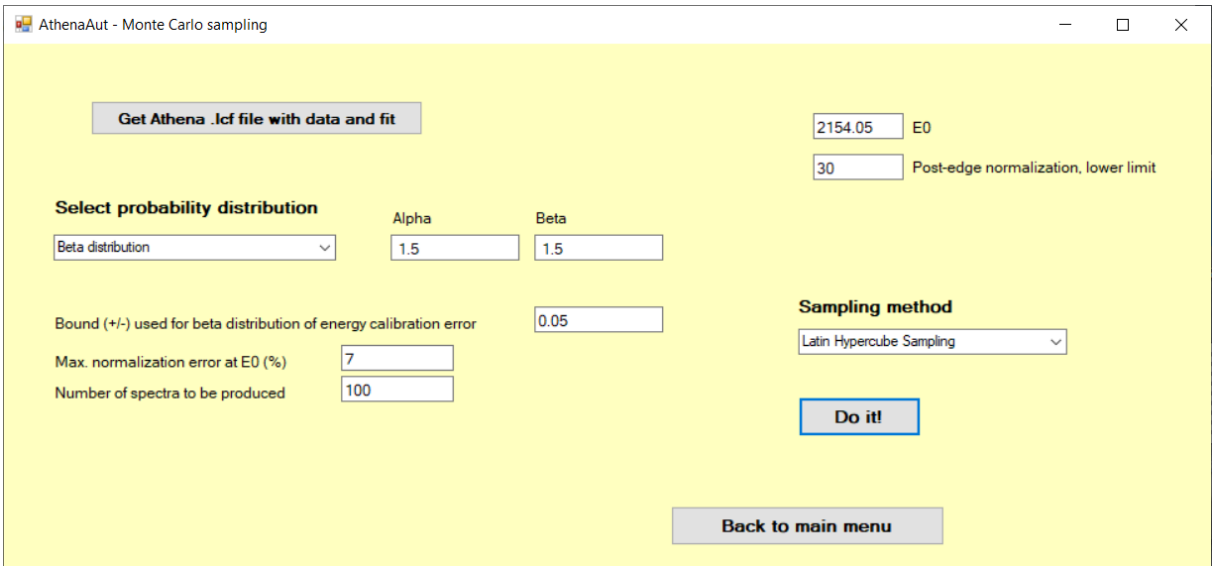

Fig. 2. The Monte Carlo sampling menu in AthenaAut.

All random sampling for the uncertainty analysis is made in a single menu (Fig. 2), which is accessible by selecting “MC Sampling” from the AthenaAut main menu. Two sampling methods are available, classical Monte Carlo sampling and Latin Hypercube sampling. The latter method is used by default. The sampling is made from a probability distribution. Currently the user can choose between the Beta, the Rectangular and the Normal distributions. The Beta distribution is the default. When this is selected, the values for the α and β values in the Beta distribution also need to be set. The default values are 1.5 and 1.5.

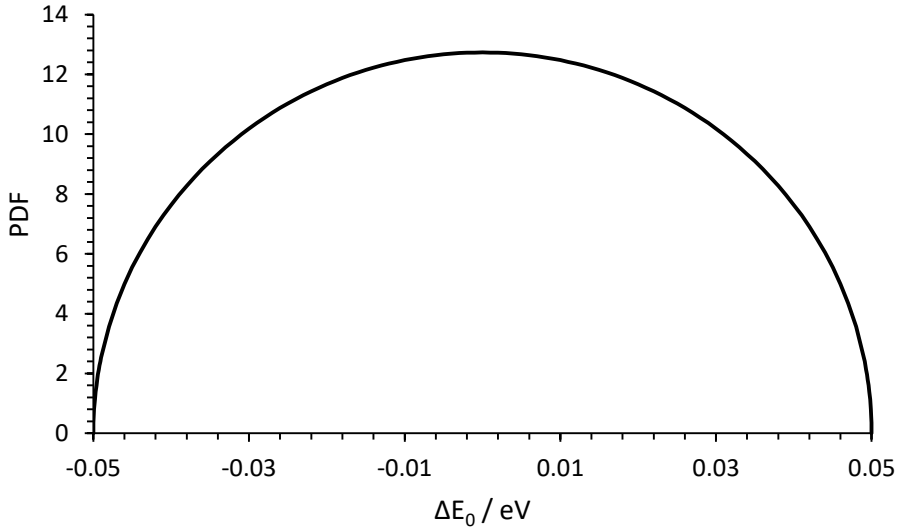

Fig. 3. Probability density function of the Beta distribution used for linear hypercube sampling of the misalignment of the energy calibration, with the shape parameters fixed at $\alpha = \beta = 1.5$. ΔE_0 is defined as the difference between the assigned energy and the real energy.

AthenaAut assigns uncertainties in two input parameters, the calibrated energy (the "energy calibration error" on the menu), and normalization ("max. normalization error at E0"). As an example, in Fig. 3, the shape of a Beta(1.5,1.5) distribution, with a maximum energy calibration error of 0.05 eV, is shown. For the error in normalization, AthenaAut assumes that the maximum error occurs at and below E_0. Above E_0, the error decreases linearly until a value of 0 at the lower limit of the post-edge normalization range (which is +30 eV relative to E_0 in the text box of Fig. 2).

Typically, the following steps are followed when performing a Monte-Carlo sampling:

1. The normalized spectrum is read from the output of an LCF analysis in Athena, to ensure that the result is directly comparable to the Athena output result without uncertainty analysis. Therefore to import your spectrum, click "Get Athena .lcf file with data and fit".
2. Decide on appropriate values on the errors, E_0 and the lower limit of the post edge normalization range and edit the text boxes accordingly.
3. Select the number of spectral variants to be produced. The default value is 100.
4. Click **Do it!** The menu will freeze for a few seconds (this is necessary to produce unique new sets of randomized values).
5. A dialog box will appear, where you select the folder and the name of the spectra. AthenaAut will add the characters "_*x*" to the file name you decide, where *x* is the number of the scan. When finished, AthenaAut will display a Finished message.

Define a sequence of runs

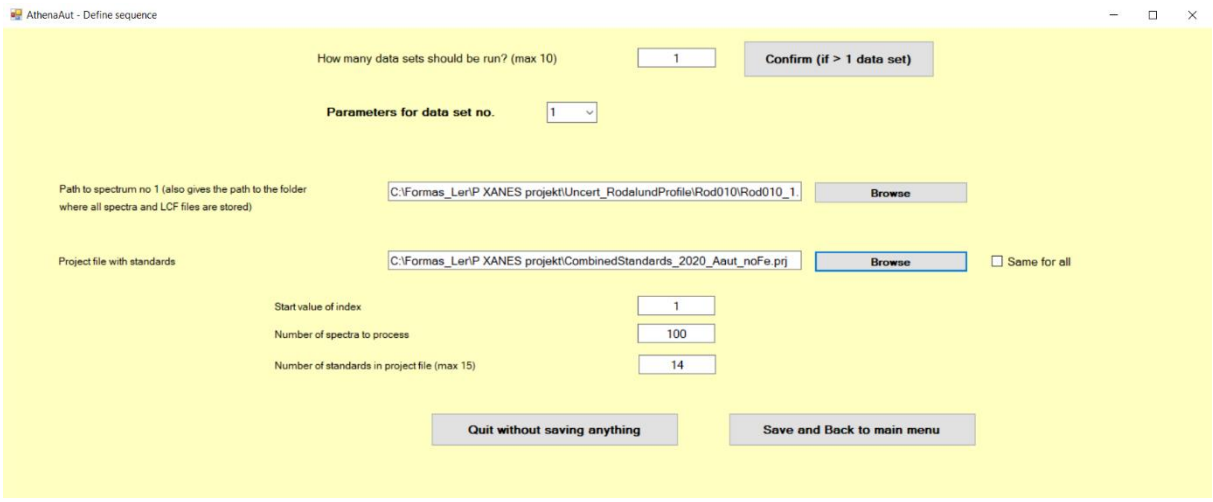

Fig. 4. The Define sequence menu in AthenaAut.

By defining a sequence you can let AthenaAut run a number of different sets of runs, one after the other. To add a new sequence (and thus to define which spectra to run), go to the Define sequence menu (Fig. 4). First, decide on how many sets you would like AthenaAut to run (max. 10). If this number is greater than 1, click the **Confirm** button. With the **Parameters for data set no.** combo box you can scroll between the sets. For each set you will need to provide:

- the path to the first spectrum. Click **Browse** and locate it. Normally it is the one that ends with "_1.CSV" in its file name.
- the path to the Athena project file for the standards (see *Preparing LCF files* above)
- "The start value of index". This is the first scan in a set of scans, the default value is 1
- the number of spectra to process, and
- the number of standards in the Athena project file where you keep all the standards. This number is important to state correctly, as AthenaAut will calculate the execution time for Athena based on this number.

When finished, click **Save and Back to main menu.**

Running a sequence

To run the sequence that you defined, you go back to the main menu (see figure on first page). On this menu, a few parameters have to be defined before you can run.

- First, make sure that you have supplied the path to the Athena executable "dathena.bat". To avoid having to supply this more than once, click the button "Save as default settings" as soon as you have found the file.

- Second, enter the value of E_0 that you want to use in the LCF. AthenaAut will always change the E_0 in the Athena interface to this value when running.

- The **wait factor** is an important parameter that you will need optimize for your computer. The wait factor defines the time that AthenaAut will wait until all combinations of the LCF in Athena have been completed. Values between 0.8 and 1.5 seem realistic, depending on the speed of your computer (higher values should be used for slower computers). If the wait factor is too low, AthenaAut will go on before Athena is finished, which will result in an error message, which in turn will force AthenaAut to try again with a 60 s longer wait time. On the other hand, if you use a high value of the wait factor, AthenaAut will wait unnecessarily long for Athena to complete. When you set this factor, please remember that the execution time may vary somewhat depending on the phase composition of the sample.

- You can also choose whether to overwrite an LCF file that already exists, or to skip and jump to the next one.

- When you think that everything is correct, click **Run sequence**, and wait until everything is finished! ***Advice:*** if you have a mouse connected to the computer, remove it before, or immediately after, you click Run sequence. This is because if you accidentally move the mouse, even a tiny bit, this might screw up the simulated mouse clicks in AthenaAut, which often causes Athena to crash.

Early exits

Often you may want to escape from a sequence early before AthenaAut has completed all runs in a sequence, or if AthenaAut has got stuck. It is very important to do this using the procedure below. ***If you don't, AthenaAut may still be running and the mouse clicks it generates can create havoc on your computer!***

<u>Recommended procedure:</u>

- Wait until Athena has started calculating combinatorial fits for the next sample
- Then click on X in the upper right corner of the black background in Athena, to close Athena.
- Then as soon as possible after this, click on X in the upper right corner in AthenaAut – this will close the AthenaAut form as well as all its background threads.

Retrieving the results

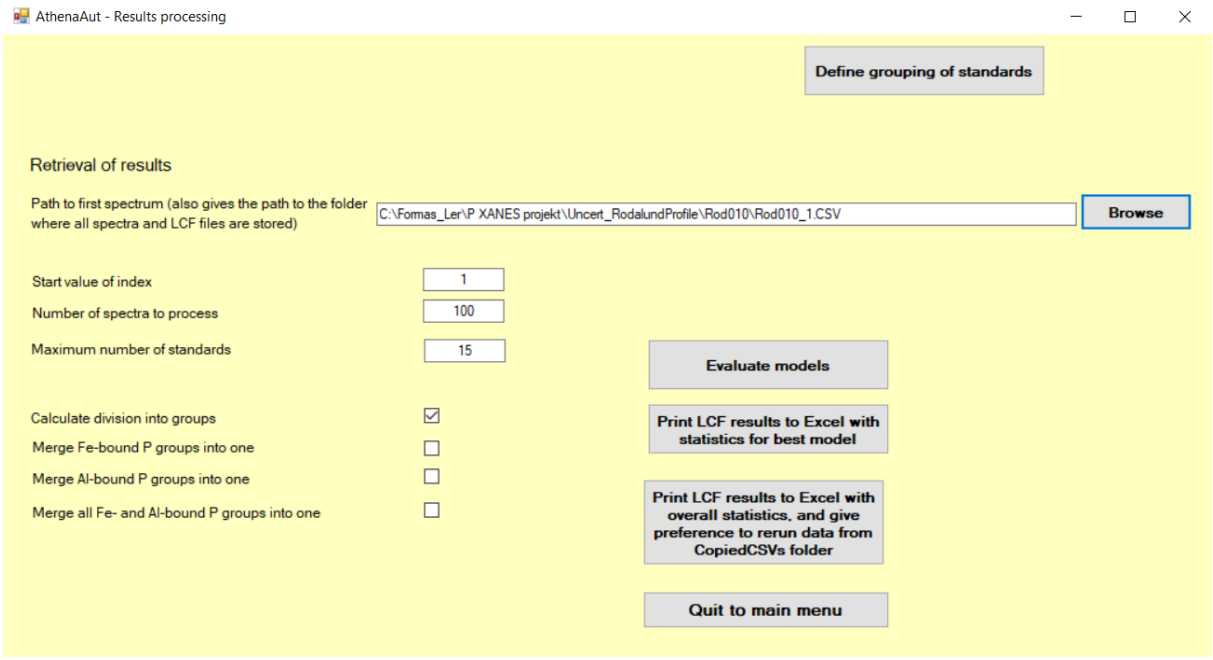

Fig. 5. The Results processing menu.

When you are finished with running a sequence, go to the **Results processing** menu (Fig. 5), where you will choose how to present the results as described below. Here you can also review how the different standards are grouped by clicking **Define grouping of standards.** The rationale behind the options on this page is described in section 2.3 of the paper (Gustafsson et al. 2020). A typical workflow through this menu is:

- Select the path to the first spectrum in a set of spectral variants, for which you have run AthenaAut. You do this by clicking **Browse**. The spectrum name should end with "_1.CSV" as in the example in Fig. 5.
- Review, and if necessary change, the **Start value of index** (normally 1) and the **Number of spectra to process**.
- The **Maximum number of standards** is just a number that should be equal to or larger than the number of standards in your Athena project file with all the standards, and is used to dimension the array for the Excel export.
- The first step in the data treatment, and grouping of your results, is to use the **Evaluate models** button. Depending on whether the boxes on the lower left are ticked or unticked, this will produce a report on the statistics of the groups. Usually I only accept models that cover 50% or more of the spectral variants.
- When you have decided on a model that fits your data, review the detailed results in Excel by clicking **Print LCF results to Excel with results for best model**. Usually, you will see that there are a number of spectra that your model does not cover, as it does not contain the same composition of groups.

- To get improved statistics for the spectra not covered by your group division, you can rerun these in AthenaAut, restricting the standards to those covered by your groups. You can do this by copying these input .CSV files to a separate folder (i.e. click **Copy CSV files not used by best model**).
- Go back to the main menu and define a new sequence with these copied CSV files. For this you will also need to create a new project file in Athena (see "Settings in Athena" above, which contains only these standards that are covered by the groups that you have defined.´
- Repeat the run (this should be much quicker the second time both because of the smaller number of spectra and because the smaller number of standards in the Athena project file).
- Go back to the Results processing menu. Now you can get the final results/statistics by clicking **Print LCF results to Excel with overall statistics, and give preference to rerun data from CopiedCSVs folder**.

Reference:

Gustafsson, J.P., Braun, S., Tuyishime, J.M.R., Adediran, G.A., Warrinnier, R., Hesterberg, D. 2020. A probabilistic approach to phosphorus speciation of soils using P K-edge XANES spectroscopy with linear combination fitting. *Soil Systems.*