

# MetaboliteDetector – Development Version

MD-Version: 2.1.N-2012-02-16

Author: Christian Nieke

Date: 27.02.12

## About this document

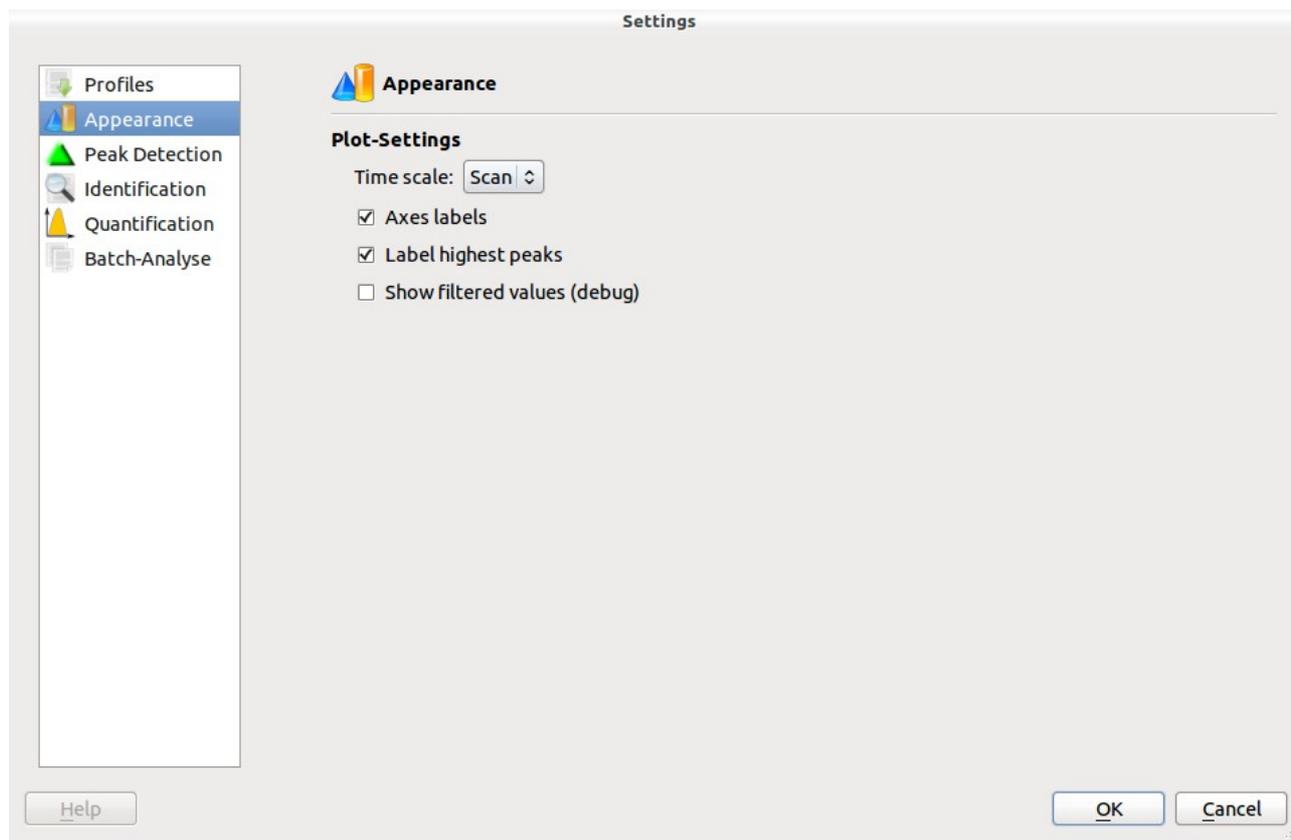
This document is intended to explain the new features in the development version of MetaboliteDetector and their respective parameters.

I will update this document whenever I see the need, so if you have anything you want added, changed or explained, give me a hint where it still lacks:

`c.nieke@tu-bs.de`

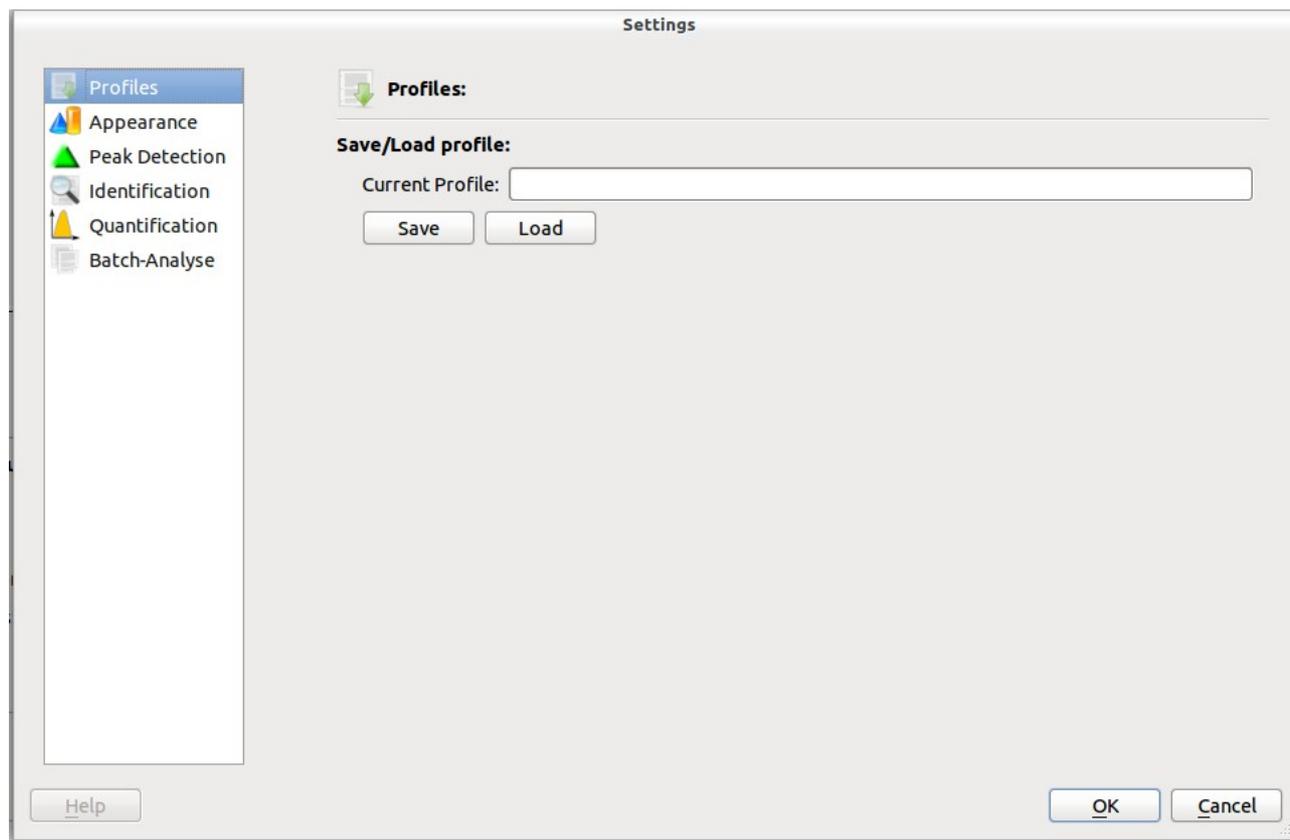
## Settings

### Appearance



TODO

## Profile



This dialog allows to save your settings in a profile or load them from one. This new feature can be used for convenience, but is also intended to ease documentation.

We recommend using one profile for each of the main steps of the workflow:

- **RI-Calibration:**  
Use low-sensitivity settings for the detection of compounds in your reference chromatogram.
- **Compound Detection:**  
Use high sensitivity settings for the second detection step during the RI-Calibration, which will detect compounds in all sample files that are calibrated. Also use these settings for any further detection step.
- **Batch quantification:**  
Use these settings during the batch quantification.

## Peak-Detection

The screenshot shows the 'Settings' dialog for Peak Detection. The sidebar on the left lists: Profiles, Appearance, Peak Detection (highlighted), Identification, Quantification, and Batch-Analyse. The main content area is titled 'Peak Detection:' and includes a green triangle icon. Under 'SIC Peak Detection', there is a 'Detection Range (scan)' section with 'Until scan: 8000.00' and a 'Reset' button. Below that is a dropdown menu for 'Select method for peak detection:' set to 'Matched-GSD (new)'. The 'Matched-GSD Settings:' section contains four input fields: 'Peak-width (scans): 12', 'Filter-Threshold: 3000.00', 'Required Peak Quality (%): 50', and 'Required SignalToNoise: 0.00'. There is also an unchecked checkbox for 'Additional run with 80% of peak width'. The 'Deconvolution / Clustering' section includes 'Pre-Clustering' set to 'Faster', 'Cluster threshold' set to '200.00', 'Required intensity (% of base peak)' set to '0.00', and 'Required number of peaks' set to '10'. An unchecked checkbox for 'Enable baseline adjustment' is also present. At the bottom of the dialog are 'Help', 'OK', and 'Cancel' buttons.

This dialog is used to define the settings for the peak detection, i.e. finding peaks in the single ion chromatograms (SICs) and combining them into one peak of the total ion chromatogram (TIC).

The combination step is often referred to as „deconvolution“, however we use a clustering approach here.

### *Detection Range:*

Allows you to limit the range where the detection will take place i.e. you can „cut off“ the end of the chromatogram.

### *Method for peak detection:*

Select one of the two methods for peak detection. You can read about the old, unchanged version in the wiki. This document will explain the new detection method, based on:

"Danielsson, Rolf et al., Analytica Chimica Acta 454 (2002), 167-184 Matched filtering with background suppression for improved quality of base peak chromatograms and mass spectra in liquid chromatography–mass spectrometry “

### *Peak width:*

Approximate the width of your average peak in scans. This parameter might change quite drastically, depending on your scan rate.

Hint: You can change the time scale in the „Appearance“ dialog to scans for a better guess.

### *Filter-Threshold:*

The threshold that must be passed for a SIC peak to be detected. This is a rather artificial value, generally speaking, the higher the value, the more peaks will be filtered out.

Note that this filtering is based on the SIC, so even with a high value for threshold, a TIC peak might still be found, however it might have a very thin spectrum. If the value is set too low however, this will cause noise to be identified as peaks and also induce noise into the spectra of your peaks.

Hint: In the „Appearance“ dialog, you can activate the option „Show filtered values“. This will add an additional black curve to the TIC-Plot, showing the filtered values of the currently selected curve. The threshold corresponds to those filtered values.

Hint: It is usually a good idea, to have too much fals-positive peaks rather than too few. You can filter them out during the batch-quantification using the reproducibility restriction.

### *Required Peak Quality:*

A filter applied on SIC peaks. A peak will be ignored, if it's quality is below this value. The quality is calculated based on the derivative of the peak, which should be positive before the maximum, and negativ afterwards for a high value.

Hint: You can use this to avoid zick-zack-lines being detected as peaks. You should not use it unless necessary.

### *Required Signal-to-Noise:*

A filter applied on SIC peaks. A peak will be ignored, if it's S/N is below this value.

Hint: You should not use it unless necessary.

### *Additional run with 80% peak width:*

If peak width differs strongly between peaks, having just one value does not allow to find all peaks. Activate this feature to perform an additional peak detection run on each SIC with a lower peak width setting. This should help to better find smaller peaks as well.

This feature might be extended in the future to allow more configurability.

### *Pre-Clustering:*

You can speed up the clustering step (see below) by activating the pre-clustering. This will use a faster approximation for the first clustering steps, which might however change the results.

Hint: The option „faster“ is usually safe to use, „fastest“ should only be used if necessary.

### *Cluster Threshold:*

To combine SIC peaks into a TIC peak, we use a hierachic clustering, based on the distance of the SIC peaks maxima (i.e. peaks cluster together if their maxima are near together). In the beginning, each cluster represents one SIC peak (unless pre-clustering was performed). In an iterative process, the clusters with the smalles distance in the dendrogram tree will be combined, if the distance is below the *Cluster Threshold*, until no more clusters match this criteria.

In short: A higher value merges more SIC peaks to a TIC peak. Use lower values if your peaks are badly seperated.

### *Required intensity:*

Don't use this.

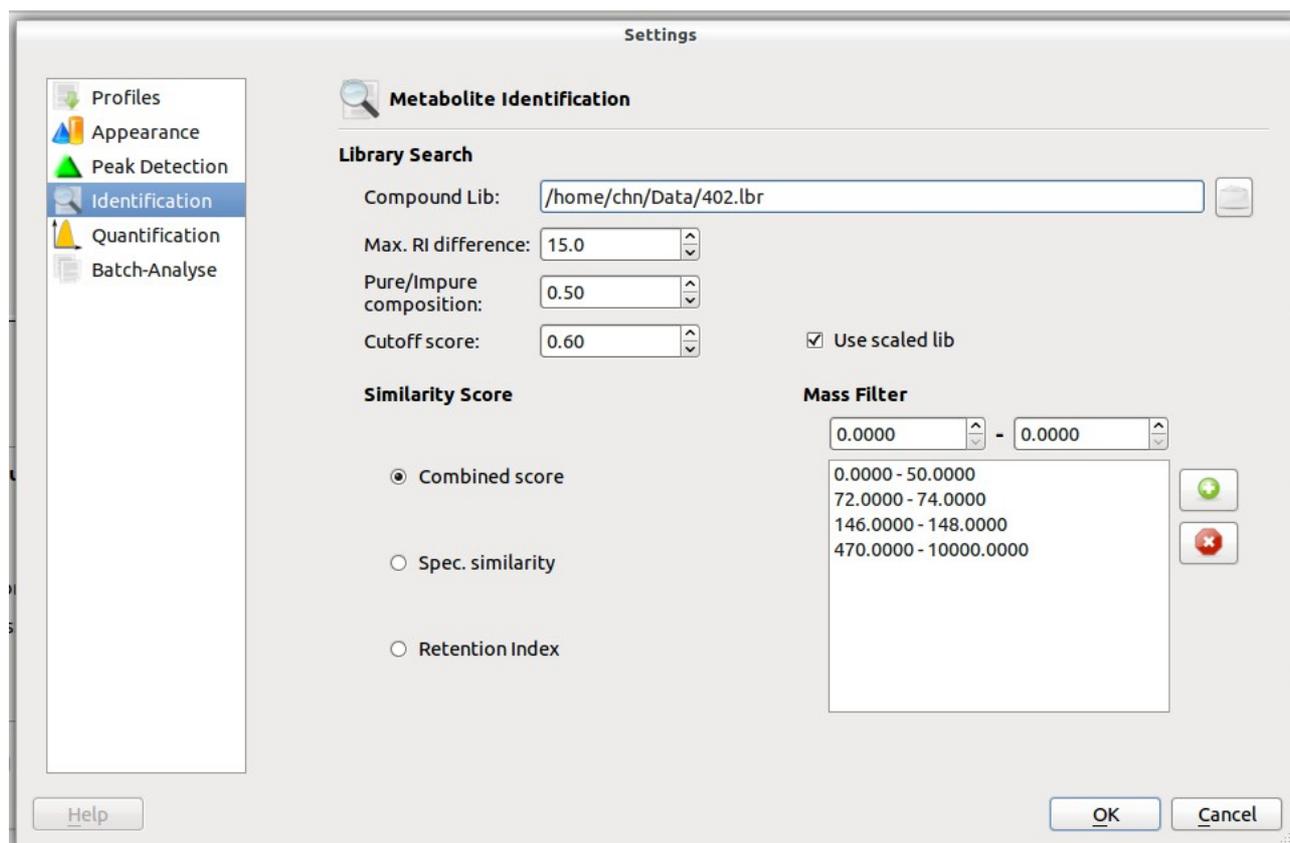
### *Required number of peaks:*

A TIC peak will be ignored, if it does not contain at least this many SIC peaks. Note that the number of SIC peaks also corresponds to the number of entries in the spectrum.

*Enable baseline adjustment:*

Don't use this, unless you have accurate masses.

# Identification



The identification step matches the detected peaks to the library.

## *Compound library:*

Choose your compound library here. The file must be in the .lbr format. If you have a library in .msl format, you can convert it via File->Import->Import MSL library

## *Max RI difference:*

During the identification, each detected TIC peak will be compared to each library compound. To speed this up, library compounds with a RI-difference to the peak greater than this value will be ignored.

Hint: If the library you use was created with the device you used for your measurement, this value can be rather low. If the library was created on a different device, you might have to set the difference quite high.

## *Pure/Impure Composition:*

Leave this at 0.50

## *Cutoff score:*

The minimal score required to consider a peak „identified“.

## *Use scaled lib:*

Activate this feature.

### *Similarity score:*

Define if the identification should be based on spectrum or RI similarity or both.

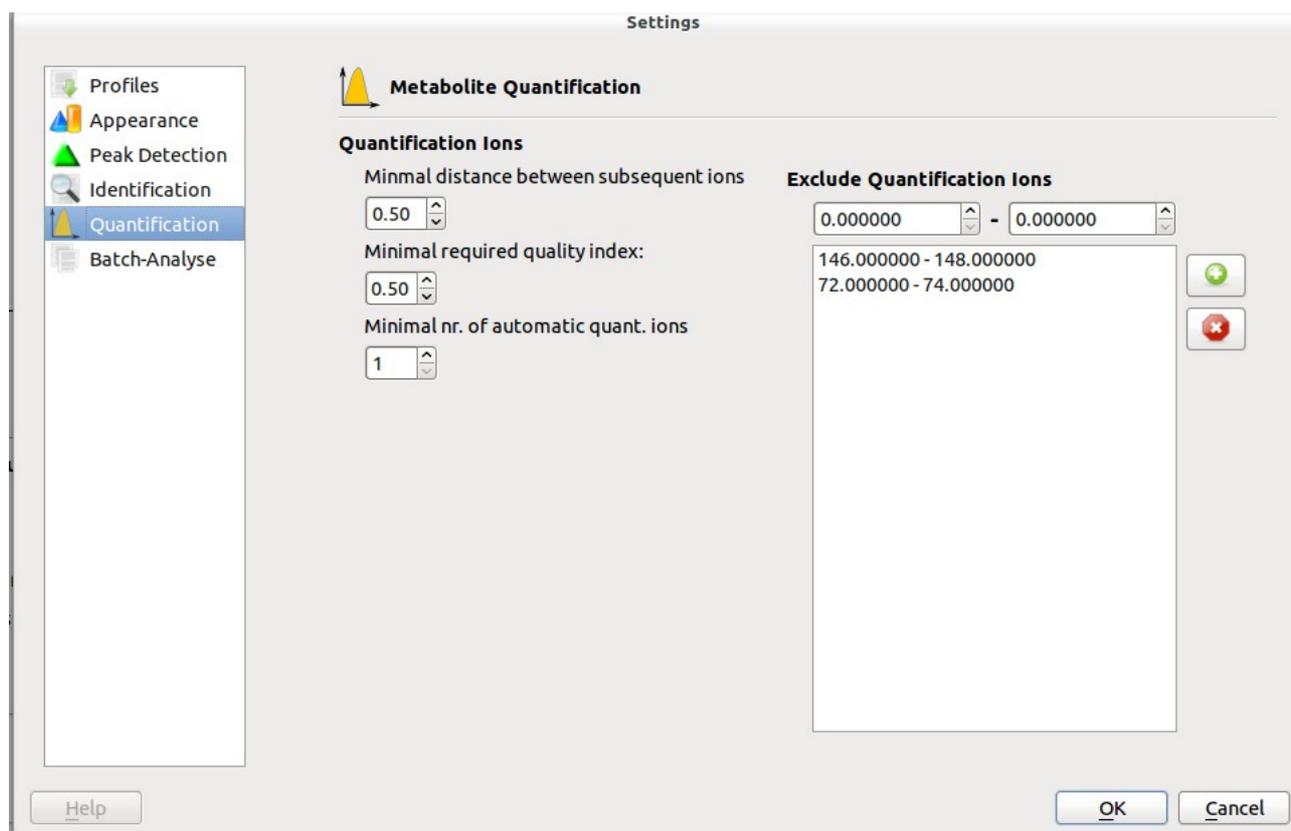
We strongly recommend using both.

### *Mass filter:*

You may define ranges of  $m/z$  here, that will be ignored during the calculation of spectrum similarity.

Hint: You could use this to filter  $m/z$  values below or above the detection range of your device, so that „missing“ those values in the peak while they are contained in the library compound won't lead to a bad scoring.

# Quantification



This dialog defines the settings used for quantification, which is mainly relevant during the batch quantification step.

## *Minimal distance between subsequent ions:*

Leave this at 0.50

## *Minimal required quality index:*

Leave this at 0.50.

## *Minimal nr of automatic quant. Ions:*

If the library does not specify a quantification ion for a compound, how many ions should automatically be detected.

## *Exclude Quantification Ions:*

Specify ions you don't want used as automatic quantification ions.

Note: If a quantification ion was set manually in the library, it will not be excluded by this filter.

## Batch Quantification

The screenshot shows the 'Settings' dialog for 'Batch-Analysis'. The sidebar on the left lists: Profiles, Appearance, Peak Detection, Identification, Quantification, and Batch-Analysis (highlighted). The main content area is titled 'Settings' and contains the following sections:

- Compound Matching Between Samples**
  - ΔRI: 10.0
  - Pure/Impure: 0.50
  - Scoring method: RI + Spec
  - Req. Score: 0.70
- Batch Filter**
  - Compound reproducibility: 0.50
- Quantification**
  - Mass filter in use
  - Extended SIC scan
  - Redetect all quant. ions
- Extended SIC settings**
  - ΔRI: 10
  - Min. reproducibility to perform ext. SIC scan: 0.50

Buttons at the bottom: Help, OK, Cancel.

This dialog handles settings specific to the batch quantification. The batch quantification is, however, also dependant on the settings for identification and quantification, so we will give a short overview on the workflow to explain how to configure a batch quantification:

- Peak detection will **NOT** be performed during the batch quantification step. It is based on the compound already detected during RI-calibration or subsequent compound detection steps.
- The batch quantification starts with a compound matching between samples, i.e. MD will take the unidentified compounds from each sample file and compare them to the compounds of other samples within the same group. This is used to match the compounds within different samples. You can define the criteria for this matching in the Batch-Analysis dialog.
- After the compounds of the samples within a group are matched to one peak representing them, this combined peak will be identified using the settings of the „Identification“ dialog.
- After the compounds are identified, they will be quantified, using the settings from the „Quantification“ dialog.

### *Compound matching between samples:*

See the „Identification“ dialog for these settings.

### *Compound Reproducibility:*

Percentage of samples a compound must be found in within a group. So 0.5 means a compound must be present in at least half the samples of a group or it will be dismissed.

Hint: This is a good way to remove false-positive peaks.

### *Redetect all quant. ions:*

Activate this to ignore the manually set quantification ions and automatically detect new ones.

### *Extended SIC scan:*

You can activate an extended SIC scan. This will try to quantify a compound, that was missing in a few samples of the group, using more relaxed requirements.

For the extended SIC scan to be performed, the compound must be present in a percentage of the samples of the group and the search will be limited to the given RI difference.

Hint: Be careful when using this settings, since it has a high chance to create incorrect quantification values, rather than returning 0. This could hide problems caused in any of the former steps, which should preferably be solved there.