

SoFi history list

fixed bugs and new features

CONSIDERATIONS

The list contains the improvements and fixed bugs per SoFi version. Each version is on a separate page and the changes always refer to the previous version. The history starts with the version 6.5, released mid-2018. The list is not exhaustive, i.e., only major improvements and resolved bugs are reported.

Table of contents

Changes made in 8.0.5.....	3
Changes made in 8.0.4.....	4
Changes made in 8.0.3.1.....	5
Changes made in 6.F6.....	6
Changes made in 6.F.....	7
Changes made in 6.E.....	8
Changes made in 6.D.....	9
Changes made in 6.C.....	10
Changes made in 6.B.....	12
Changes made in 6.A.....	13
Changes made in 6.9.....	14
Changes made in 6.8.....	16
Changes made in 6.7.....	17
Changes made in 6.6.....	18
Changes made in 6.5.....	19

Changes made in 8.0.5

BUGS

- Criteria panel (blue): significance tests were still performed in case the data points were all missing, i.e., all -9.99, resulting in rather large computational times.
- Result panel (red): residual histogram plot was not properly working for the single variables non-pdf case.
- Result panel (red): normalization is applied also before reimporting the data. Necessary in case a constrained profile contains some free variables.
- Explore and Result panel (green and red panel): linear fit was previously reporting 2 decimal digits. Now it shows 3 significant digits.
- PMF input definition and External data (yellow panel): hard-coded direction-dependent wind averaging was not operational for 24 hrs filter data.

IMPROVEMENTS

- Result panel (red): mass-dependent bar plot in the fraction tab has now two options. The user can generate plots that contain bins either by mass or by percentiles.
- Result panel (red): explained variation matrix for single factors or unexplained noisy variation is accessible and can be easily plotted and inspected from the residual tab.
- New policy for SoFi standard: SoFi standard has been downgraded to a very simple version for beginners only with limited analysis, e.g., PMF run over the entire imported data for single PMF runs with α value constraints are mainly supported.

,

Changes made in 8.0.4

BUGS

- PMF import panel (gray): several minor errors (bad indexing or missing waves), in particular for the Aethalometer subpanel are fixed.
- Standalone bootstrap panel (pink): several minor errors (bad indexing or missing waves) are fixed.
- Criteria panel (blue): Criteria list with new types of criteria (image or movie) can now be stored, loaded and criteria can be moved within the list.

IMPROVEMENTS

- Exploration panel (light green): data matrix is accessible as well as scans on user-defined time selections drawing the marquee or entering the edges of the marquee.
 - PMF prep panel (dark green): full indexing freedom, e.g., monthly, hourly, minutes, etc. (consult the manual for all details) is available for definition of classes.
 - PMF options (orange): support rapid downweight of single variables for PMF (be cautious, you have to justify any manual downweight)
 - Criteria panel (blue): criteria from other experiments with different PMF input data (just the list of criteria, not the evaluated scores) can now be imported. This avoids redefining the whole list of criteria each time.
 - Result panel (red): single PMF runs can be easily selected using a dropdown menu on the data subpanel. This allows simple browsing through single PMF runs.
-

Changes made in 6.G

BUGS

- Result panel (blue): explained variation plot was wrong in case the S2N threshold for the differentiation between real and noisy part was set to zero.

Changes made in 8.0.3.1

BUGS



IMPROVEMENTS

- Result panel (red): slider function for the dynamic plot inspection is now accessible also from the exploration panel and for all options of the result scatter plot.
- Result panel (red): correlation coefficients and p-values can be directly added to the correlation images.
- Result panel (red): external information can be included to the pie charts.
- Criteria panel (blue): image and movie option for PMF run inspection is now accessible.
- Import panel (gray): import strategy has been completely restructured and allows to import also .txt, .csv, .dat PMF input and external files apart from .itx files.
- Import panel (gray): subpanel for the efficient treatment of Aethalometer data including the source apportionment of eBC for traffic and biomass burning is operational.
- Bootstrap panel (violet): generalized bootstrap panel for the application of random resampling on data from SoFi panels and user-defined waves from the data browser is operational.

Changes made in 6.F6

BUGS

- Result panel (red): avg. was taking too long, as the entire time-dependent a value matrix was always evaluated.
- Result panel (red): mass-dependent fraction plot is now based on the quantiles and the nb. of occurrences is normalized.
- Criteria panel (blue): completely missing score plots didn't show the -9.99 for missing data.

IMPROVEMENTS

- -

Changes made in 6.F

BUGS

- PMF runs: Empty PMF runs (very rare, most likely an issue in ME2) generated empty averaged results.
- PMF runs: adjustment factors for multi-time were not properly treated in case the exclude CO2 related variables was selected.
- Exploration panel (green panel): option blacklisted variables was only considering blacklisted time-points.
- Criteria panel (blue): Correlation between two factors was giving an error, when a factor was not always present, e.g., in a result file containing changing numbers of factors.
- Criteria panel (blue): T-Test with subsequent p-values for explained variation was disabled.
- Result panel (red): Multi-time: start and end time can now be passed for the proper averaging of points, in case the time series contains gaps (please consult the manual for more details).
- Result panel (red): linear combination of factors was not working properly for the previous version (empty combined factors).
- Result panel (red): averaging didn't always properly treated the median and quantile representation for combining only few PMF runs.
- Result panel (red): evaluation of the residuals was storing the avg for family of residuals rather than the sum.
- Result panel (red): model subtraction plot was not considering the residuals in the denominator when performing the fraction plots.
- Result panel (red): median results were shifted in case variables were blacklisted.

IMPROVEMENTS

- Result panel (red): average results now also reports a values for constraints applied on single variables.
- Result panel (red): the standard deviation of the mean is added to the mean values in for all pie charts.
- Result panel (red): model subtraction plots are now also supporting combinations of factors.
- Result panel (red): mass-dependent fraction plot is now available.
- Result panel (red): externals can be added to the pie charts based on user-selection.

Changes made in 6.E

BUGS

- PMF runs: the rolling mechanism was stopping nb. of days equal to the window length earlier than expected.
- Result panel (red): Non-default resolution for single PMF runs was not showing the median and quantile statistics.
- Result panel (red): O:C and C:H ratios were not properly evaluated for the HR plot for factors higher than 2 and in case of blacklisting the HR family plot experienced a variable shift.
- Result panel (red): linear fit option for scatter plots was sometimes causing an error message.
- Result panel (red): factor combination had a problem in case the names were repeating with different suffixes.
- Result panel (red): pie charts for factors that were not constantly present were overestimated.
- Criteria panel (blue): When deleting a criterion, all score values were not shifted, and the list of criteria had to be re-evaluated.
- Criteria panel (blue): When creating a brand-new list, the criteria were always bold, i.e., as if there were a sub-selection of the data.
- Criteria panel (blue): The criteria options scaled and absolute (scaled) residuals were not properly evaluated.
- Criteria panel (blue): Residual criteria with non-default resolution were generating an error message.
- Criteria panel (blue): p values for factor comparison when using multilinear regression criteria were too high.

IMPROVEMENTS

- -

Changes made in 6.D

BUGS

- Rolling PMF: when performing a rolling PMF over parts of the PMF input, too many PMF runs were performed, as the starting point was artificially set to the begin of the PMF input. The runs were performed correctly, but there were too many and this caused sometimes some out of memory or other unexpected errors.
- AVG. procedure (blue/red): the median option for cycle plots was using the median wave rather than the normal wave. This resulted in rather flat cycle plots, as the median of the median wave was taken.
- Criteria panel (blue): Writing a criterion on a new IGOR file was causing an error.
- Criteria panel (blue): a simplification in the procedure caused the 2nd highest score values to be evaluated only for the unconstrained factor positions. Now it is again based on all factors, independent from being constrained or unconstrained.
- Criteria panel (blue): Selection from the panel “Select time and class” was not properly passed in 6.C, due to a wrong declaration of the strings of this panel.
- Result panel (red): average values on pie charts were based on the fraction wave rather than on the contribution wave.
- Result panel (red): quantiles for factor profiles were not properly evaluated.
- PMF settings (orange): all “disable” and “enable” strings have been uniformed. Make sure that the settings, e.g., a values, pulling, etc. are properly enabled or disabled on the orange panel depending on your needs, i.e., the buttons should be red/green and the string should be visible with “enabled/disabled”.
- PMF settings (orange): pulling equations have been aligned to the hard-coded ME2 notation for mathematical expressions. Please consult the manual (section 6.3.4, p. 56-60) for further assistance on how to set up a pulling equation.

IMPROVEMENTS

- Criteria panel (blue): manual changes on the criteria list, once being popped, are now properly passed to the batch table for the evaluation.
- Criteria panel (blue): moving of criteria positions is available for a better sorting of the criteria list.
- Result panel (red): in normal and scatter plot section the option “variable of sol.” considers time-dependent factor profiles, the case for rolling runs.
- Result panel (red): model subtraction scatter plot is improved with a slider function and the information of the relative contribution of the factors to the selected variables. Users can dynamically inspect the performance of PMF over the various rolling windows.
 - Result panel (blue and red): factor specs wave can be modified from the blue panel (button “factor specs”). New: “factor specs” supports combination of factors. This allows to create for example a POA or OOA factors that can be used for further analysis in the red result panel.

Changes made in 6.C

BUGS

- AVG. procedure (blue panel): non-default resolution for single PMF runs produced an error message when varying nb. of factors in HDF file.
- AVG. procedure (blue panel): for non-default resolutions the error was propagated, rather than averaged. This caused all results based on uncertainties, e.g., Q values to vary from the default resolution result.
- Years selection checkbox (green, blue and red): list of years is now only showing the years present under time series. Before, the previous year was also sometimes erroneously shown.
- Reintroducing m/z 44 related variables was not properly working in case all m/z's below 20 were blacklisted.
- Criteria panel (blue panel): passing nans for left and right hand-side on the marquee selection was not selecting the entire x-wave.
- Criteria panel (blue panel): p-values for significance tests were now drawn on the score plot.
- Criteria panel (blue panel): saving and loading score evaluation was not properly passing the factor profile index wave.
- Criteria panel (blue panel): "-" was not properly interpreted and caused all criteria, where subtractions should have been performed, to be blank.
- Criteria panel (blue panel): strings and variables of the subpanel "time and class" are now independent. Before, they were using the ones from the result panel.
- Criteria panel, bootstrap (pink panel): non-default resolution, e.g., bootstrap base case using hourly or daily resolution had an index problem and was running properly.
- Result panel (blue panel): Marquee selection of PMF runs is now also y-sensitive. Before, it was only x sensitive.
- Result panel (red): adjustment wave for MULTI-TIME was not stored for all PMF runs but only for the first one.
- Result panel (red): averaging index for MULTI-TIME is now stored within the parameter folder.
- Result panel (red): MULTI-TIME is now resampling- and gap-safe.

IMPROVEMENTS

- HDF files are repacked during the PMF runs to save space on the hard disk (visit <https://datalystica.com/sofi> for downloading the latest ME2 folder and don't forget to save your ME2 key first before overwriting the ME2 folder!)
- Qexp calc.: The estimation of Qexp is now $Q_{exp} = n \cdot m - p(n+m)$. Previously, it was approximated by $Q_{exp} = n \cdot m$.
- Result panel (blue panel): grand log file for all PMF-HDF files can be generated here.

- Daily cycle calculation: resolution parameter has been added (SoFi tab/Main settings). Default value is 1 hour that leads to the classical daily cycle from 0 to 24.
- PMF settings (orange panel): 3rd to 5th sensitivity analyses for a-value and pulling equations have been added.
- PMF settings (orange panel): for a-value runs “random” and “sensitivity analysis” the initial a-value can be defined.
- PMF settings (orange panel): nb. of iterations is now independent from the fpeak, a-value and pulling sensitivity analyses. This means that increasing the nb. of iterations creates repeats in addition to the sensitivity analysis.
- PMF settings (orange panel): exact and random fpeak are now available as additional fpeak options.
- Criteria panel (blue): SoFi prevents averaging PMF runs with different nb. of factors. In case a selection PMF runs in an HDF file with multi-factor solutions leads to overlap periods, SoFi considers only the factor solutions with the highest occurrence only.
- Result panel (red): a-value statistics for averaged solution contains now the pdf over time as an image plot.

Changes made in 6.B

BUGS

- Criteria (blue panel): p value for multilinear regression was not properly applied under “marquee selection options”.
- Criteria panel (blue panel): score values were not always properly evaluated when sorting criteria with combined different resolutions, e.g. normal and diurnal cycle.
- Criteria panel (blue panel): fraction of a variable under time series was not properly normalized.
- AVG. procedure (blue panel): when evaluating the quantile statistics for the averaged solution the lowest and highest limits of the current dataset are now first estimated. This prevents saturation effects in the final median contribution plot.
- AVG. procedure (blue panel): quantiles were not evaluated for the profile matrices.
- AVG. procedure (blue panel): nb of non-modeled time points was wrong in case of multi-factor solutions.
- AVG. procedure (blue panel): combination of various HDF files with different time periods, e.g. seasons had an indexing problem.
- Import (blue panel): importing single PMF runs was taking too much time for datasets with a lot of variables, e.g., 1000 variables and more.

IMPROVEMENTS

- Results (red panel): cycle selection, e.g. diurnal or weekly cycle can now be selected also for the correlation matrix.
- Results (red panel): y- and x- list for the correlation matrix is HDF-specific.
- Results (red panel): position of non-modeled time points is shown as blue line to zero in the avg. section.

Changes made in 6.A

BUGS

- Criteria panel (blue result panel): sorting unconstrained factors was not properly performed in case fewer criteria compared to nb. of factors were defined
- Criteria panel (blue result panel): score values for residuals were not considering sparse missing points.
- Criteria panel (blue result panel): score values were not properly evaluated when sorting criteria with combined different resolutions, e.g. normal and diurnal cycle.
- AVG. procedure (blue result panel): when evaluating the quantile statistics for the averaged solution the lowest and highest limits of the current dataset are now first estimated. This prevents saturation effects in the final median contribution plot.
- AVG. procedure (blue result panel): disabling the quantiles statistics also prevented from evaluating the quantiles of the residuals for the single and averaged solutions.
- AVG. procedure (blue result panel): all result waves are now single precision and unnecessary waves are removed during evaluation in order to prevent from going out of memory. However, for big datasets, e.g. ToF-ACSM over many years, you might also want to uncheck certain parts of the solution, e.g. quantiles, explained variation. Skipping the quantiles will save a lot of memory and makes the averaging process fast.

IMPROVEMENTS

- PMF options (orange panel): Multi-time, similar to the multitime ini file form Pentii Paatero is now operational in SoFi Pro. Consult the manual for more details, under 6.2.3.4.
- Criteria panel (blue result panel): options sum, min / max, histogram and combined criterion are operational. Consult the manual for more details on these features.
- Criteria panel (blue result panel): fast scan to identify the amount of non-modelled time-points is 1-2 orders of magnitude faster.
- Criteria panel (blue result panel): inspecting score plots of PMF results with a varying nb. of factors in an HDF file has been improved. The user can select and visualize the scores for a specific nb. of factors (read more in the manual under 7.3.3.8).

Changes made in 6.9

BUGS

- Treating PMF input: missing columns/rows or sparse points is now sensitive also on error matrix.
- Running PMF: possible error messages after user-abort are removed.
- ToF-ACSM: m/z 19 was erroneously evaluated when using the option exclude from run. This caused wrong residuals for m/z 19 in the PMF results.
- PMF options (orange): downweight is based on blacklisted data rather than on entire data/error matrix.
- PMF options (orange): relative downweight of errors has been boosted and problem related to the C-value table solved.

- Result selection panel (blue): Q/Qexp values for PMF runs were zero if user selected settings “do not apply robust value” under general settings.
- Result criteria panel (blue): names of external waves containing spaces weren’t properly Result criteria panle (blue): bad character definition on criteria on parts of data caused abortion of criteria evaluation.
- Result selection panel (blue): overview plots didn’t account for bootstrap repeats. In addition, the unexplained variation information was too small compared to values reported on the fraction plot (e.g. for pie charts on the red panel) due to mismatching averaging procedures.

- Result panel (red): average solution and residuals contained an erroneous zero standard deviation for the x-dimension.
- Result panel (red): for little nb. of factor profiles the time-dependent profile matrix for the 90th % was sometimes higher than max value, due to erroneous interpolation in igor.
- Result panel (red and blue): swapping single factors from the red panel didn’t affect the overview plots on the blue panel.

IMPROVEMENTS

- SoFi license structure adapted for institute-specific SoFi Pro keys and for the future realtime SA (SoFi RT).
- Running PMF: pretreating data before calling ME-2 is boosted for long datasets.
- Constraints: name of constrained anchors is passed to the log file.
- Progress bar: time left (minutes) until calculation is over is also reported below the progress bar.
- Progress bar in SoFi is optional, can be enabled/disabled from the main settings (main SoFi tab).

- Error analysis: robust mode values were first passed through the uncertainties and affected the scaled residuals too. Now they only apply to the Q-values, as also intended during regular PMF analysis.
- Factor specifications: The user can give names to individual factors for a PMF run. When averaging over several HDF files, SoFi will consider all factors by name and will properly average over HDF files with various factor names.
- Averaging and importing PMF results has been completely revisited to properly treat large datasets, e.g. AMS or ToF-ACSM data avoiding also out of memory in Igor.
- Marquee selection on the time series (options in time panel) that was previously only possible based on the PMF input wave can now be also performed on the external waves.
- After averaging the external waves the various resolutions can be enabled/disabled with checkboxes on the graphs.
- HDF files are repacked during the PMF runs to save space on the hard disk.
- Model settings: When defining seed for single variables with no external wave the user can directly enter anchor values into the table avoiding the creation of dummy waves.

- Criteria panel (blue): score evaluation including significance statistics for two regular waves and correlation coefficients can be automatically performed for all factors (constrained and unconstrained). This allows a comparison with other factors and factor stability/robustness of a factor can be inferred. Subsequently, mixed factors can be eventually removed.
- Criteria panel (blue): NaN score values due to missing external points are now receiving -1. This allows the user to select and explore these PMF runs.
- Criteria panel (blue): Factor sorting (for unconstrained factors only) has been improved and is 1-2 orders of magnitude faster.
- Criteria panel (blue): Score selection (previously only with the marquee) can be performed directly on the subpanel by entering score values as boundaries.
- Criteria panel (blue): combined criterion has been added. The possibility to combine two and more criteria into a new one.

- Result panel (red): Sally's triangle can be directly added to the f44 vs f43 scatter plot.
- Result panel (red): When performing rolling PMF runs, the residuals are now based on the time-dependent factor profile matrix rather than on the averaged profiles.

Changes made in 6.8

BUGS

- Result panel (red): modelled fraction points in model subtraction plots are now colored properly.
- Criteria panel (dark blue): multilinear regression option is now working for a single independent variable.
- Criteria panel (dark blue): ratio between two variables of the same factor is now working properly.
- Criteria panel (dark blue): bootstrap option (pink panel) doesn't cause problems on the SoFi Pro key anymore.

IMPROVEMENTS

- Result panel (red): calc. for option "model subtraction" has been boosted and is now ~5 times faster.
- Terms and conditions have been fully revisited

Changes made in 6.7

BUGS

- Unconstrained factors were sorted in the criteria panel if the user highlighted them. But the positions were not passed to the result panel.
- Interquartile range was too small (erroneous normalization) for the relative fraction.
- Criteria panel: score representation wasn't properly visualized when sorted for non-rolling PMF runs
- Profile plot: when using txt symbols for the variables error message 'out of range' popped
- Bootstrap analysis in criteria panel was not working for criteria involving cycles and profiles

IMPROVEMENTS

- Result panel (red): user chooses on which axis (left, right or new left, new right) to plot traces.
- Result panel (red): plot from overview or fraction result tab can be loaded into the standard plot tab, e.g. when external traces should be added to factor traces
- PMF settings panel: user chooses the normalization equations during PMF (default is disabled, over the profiles, over the time series)
- Result panel (red): the residual tab contains full residual (res., abs res., scaled res., abs. scaled res., Q) matrix. The thresholds for the color-code are slider-controlled and allow for a dynamic inspection of the matrix.
- Result panel (red): the avg. solution tab allows under the t-profile tab to pop averaged factor profiles over user-defined temporal regions. The same tab performs a one-way ANOVA test on the time-dependent factor profiles. The test analyses the probability that all profiles are drawn from the same distribution
- all SoFi graphs: SoFi-related marquee tool is now windows-specific
- Result panel (red): user selects which result graphs should remain open
- External panel (yellow), Criteria-panel (blue) and Result panel (red): resolutions: weekly, monthly and yearly have been added

Changes made in 6.6

BUGS

- fraction plot over time produced an error message
- blacklist panel had all years pre-checked and caused some trouble and was responsible for year-related error messages, as all years were blacklisted by default
- Bootstrap runs were not compatible with PMF input containing classes

IMPROVEMENTS

- fast averaging over PMF results allows now to treat up to 5000 PMF/ACSM runs of ~2 month length each (previously it was around 1000 PMF runs). This corresponds to ~100'000 cells for the input data matrix.
- for highly constrained PMF runs the normalization during the PMF iteration was responsible for non-convergent results. Default in SoFi is now to have un-normalized PMF runs. The user can enable the normalization equations anytime in SoFi (consult the manual).
- criteria panel has independent mean and median results for the cycle plots
- criteria panel shows p-values also for the criterion “multilinear regression”

Changes made in 6.5

BUGS

- median values for plots were nan when values were zero. Zero is now passed
- scatter plot for single variables contained a bug when data was blacklisted first in rolling windows run the hdf file h.5 was empty, due to an erroneous stringmatch
- HR plot: elemental ratios were missing
- a-value constraints panel: external wave as template was not working
- a-value constraints panel: constraints of single variables was off, when variables were blacklisted
- first part of criteria-based approach in SoFi Pro revised
- avg over wind data is present in SoFi and operational when averaging and the wavename of the wind direction wave contains 'wind_dir'

IMPROVEMENTS

- criteria-based panel revised (consult the manual)
- normalization over the profiles during the iteration can enabled/disabled