A Short Practical Introduction to TBCAT_S

Günther Meinrath (2006)

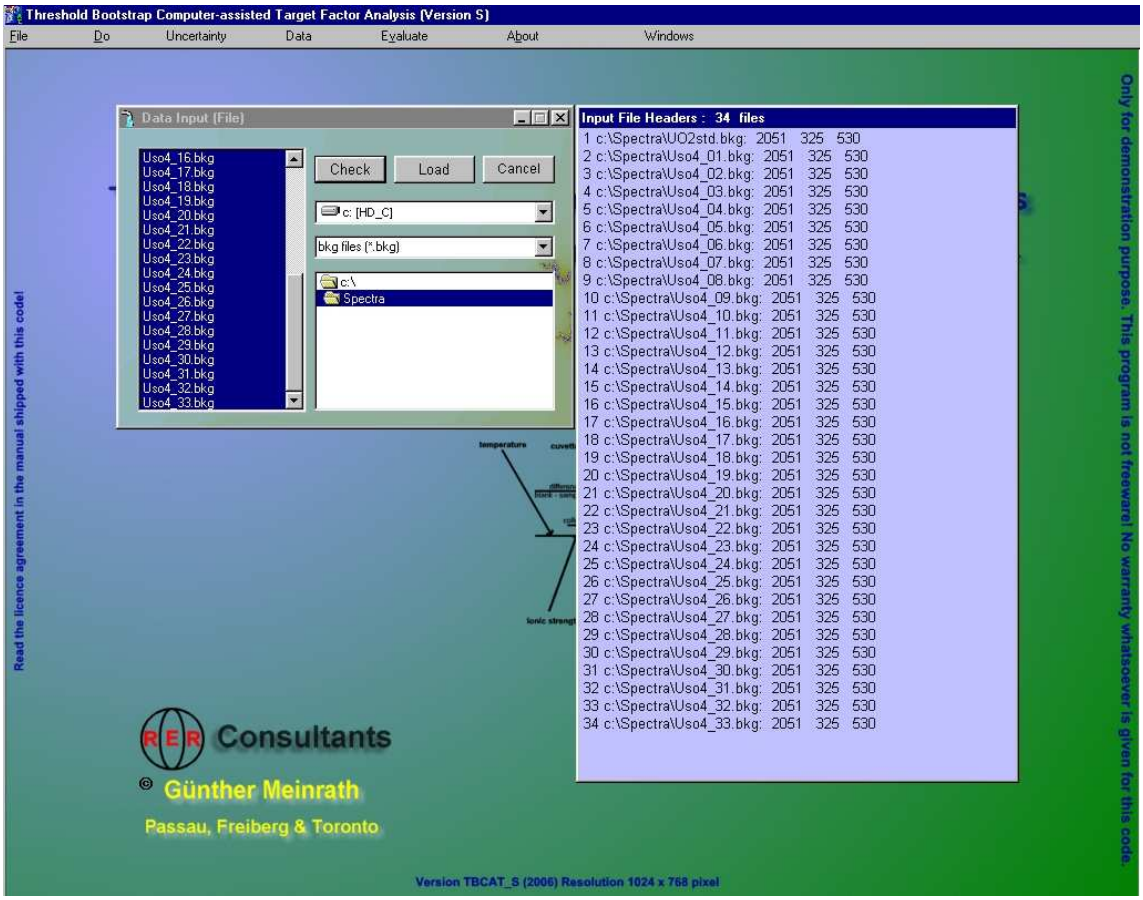
RER Consultants

Provided as an addendum to 'Quality Assurance in Chemistry and Environmental Science - Metrology from pH Measurement to Nuclear Waste Disposal' (Springer-Verlag Heidelberg FRG) by G. Meinrath & S. Schneider

These pages give a short guide into the application of TBCAT_S. It is necessary that the manual has been studied and the numerical concepts are at least basically understood. It is further assumed that TBCAT_S has been installed on a suitable computer and the sub-folders of the "TBCAT_S" directory are saved on a hard disk (remove write protection if necessary!).

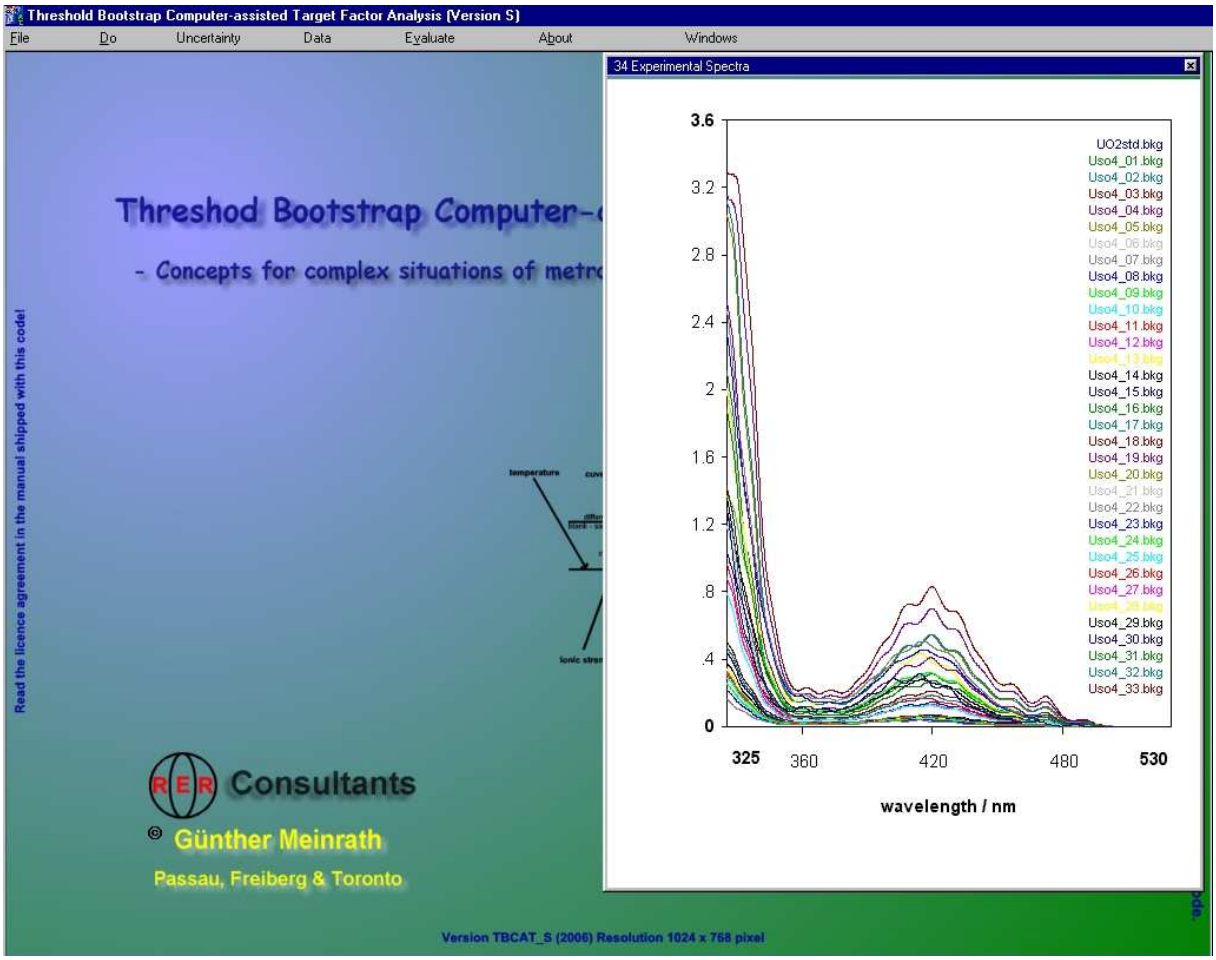
The folder "Spectra_Evaluation" holds the output of an interpretation of 34 UV-Vis absorption spectra by 75 TBCAT_S cycles. The metal ion is UO_2^{2+}, the ligands are OH^- (pH) and sulfate. There are 33 spectra collected at different total uranium concentrations, different SO_4^{2-} content and slightly varying pH. Ionic strength is assumed to be 0.1 M. These spectra have been discussed in Ref. [1]. There are 33 spectra plus a pure UO_2^{2+} spectrum: UO2std.bkg. This spectrum is used as a reference. It has to appear first in the list of spectra loaded to TBCAT_S. The measurement uncertainty values are default values of TBCAT_S.

The total of 34 spectra are loaded into TBCAT_S by selecting all spectra. It is essential that the spectra holding the 'pure component information' appears at first place in the list.



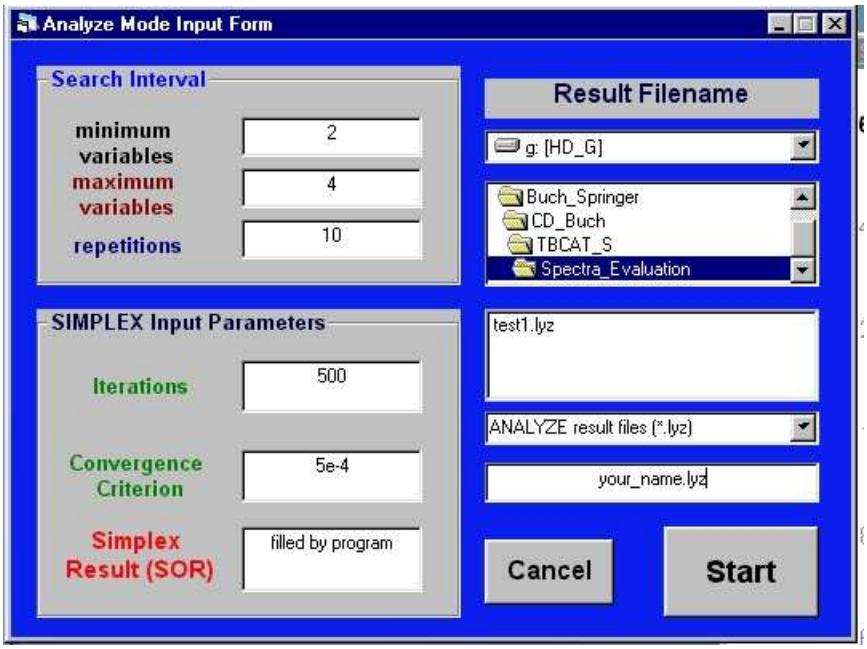
The light blue window on the right side appears if the 'Check' button is pressed. It gives some information of the spectra headers. Its purpose is to ensure that the correct spectra are selected and the basic properties of the spectra are identical. After clicking the 'Load' button, the spectra are loaded. This process takes some time, because a series of checks are done with the spectral data.

The spectral information is presented as a diagram as shown in the next figure. Note that these spectra are already background-corrected (extension is *.bkg). The wavelength range is set to 340 nm and 530 nm (the wavelength range differs from the values to be seen in the figures).



The first step in the analysis of the information provided by these spectra is the ANALYZE process. ANALYZE (in the DO menu) is searching for potential numerical interpretations (e.g. number of species giving rise to these spectra) on a purely numerical plausibility level. Negative absorption, misfit between observed and (re)calculated spectra, negative concentrations are criteria to eliminate certain interpretations. The process is based on factor analysis. Details are found in the manual. The following graph shows the ANALYZE window with reasonable input values. The file name of the file resulting from the process can be chosen freely. If the file name already has been used, a warning will pop up.

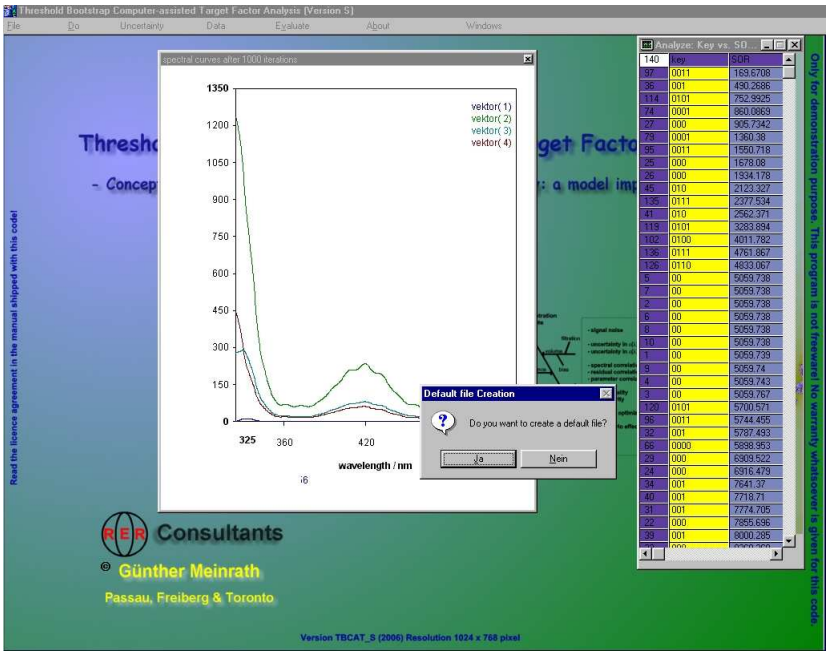
The working horse for fitting curves to data is the SIMPLEX. It is a reasonably fast non-linear fitting algorithms. It works without derivative information - which cannot provided easily for data like that treated here. The SIMPLEX requires some starting input which is generated by TBCAT_S. The user has to provide the maximum number of iterations and a convergence criterion which should be in the range $5 \cdot 10^{-3}$ to $5 \cdot 10^{-5}$. The convergence criterion at occasions has an influence on the result - which, however, is very difficult to assess.



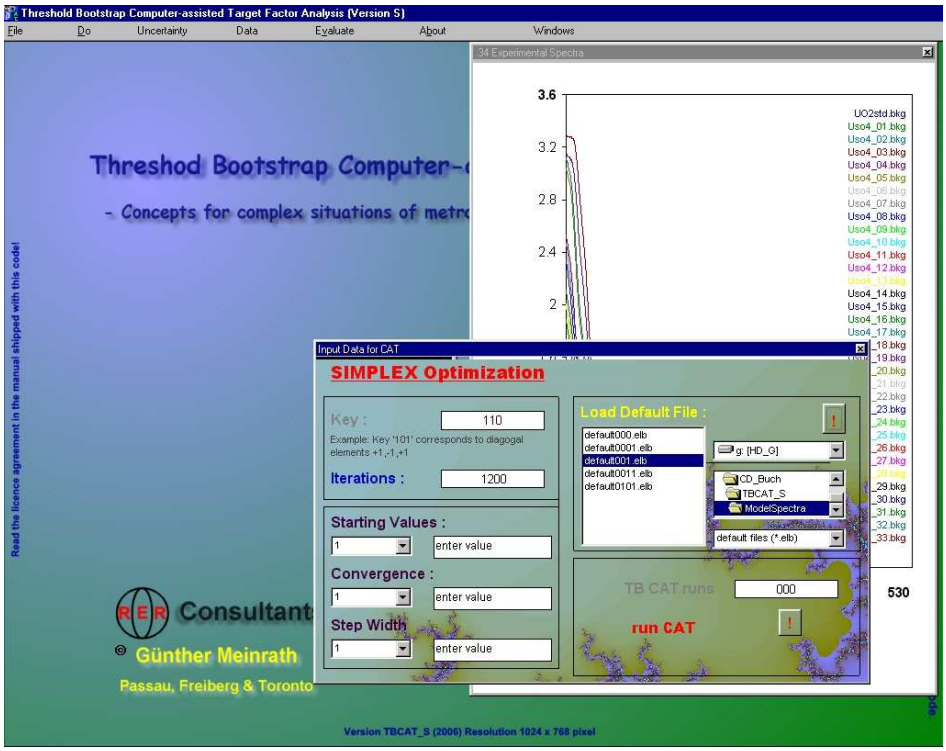
Upon clicking the START button, the process runs automatically. The spectral data are analysed on the possible rank of the data matrix and the compatibility with certain general boundary conditions. The result is expressed by a sum of residuals (SOR). Several criteria apply. The user can influence these criteria under the "Evaluate" menu item "Penalties". The criteria, however, should be modified by experienced users only.

ANALYZE may take considerable time. The total number of runs necessary to test all possible permutations with some repetitions (at least 10 - the SIMPLEX starting values are selected randomly and a minimum sample size is necessary) is time-demanding.

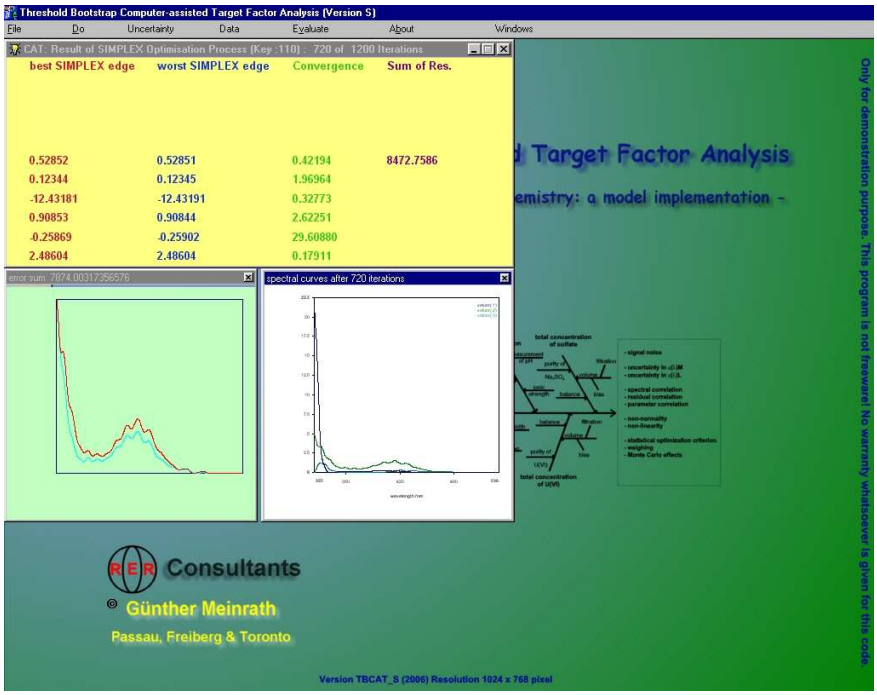
After finishing, the keys and the respective SORs are sorted. Upon clicking on a field in the "ANLYZE: key vs. SOR" table a window pops up allowing to save the respective starting values having given rise to the SOR value in a default file. It is reasonable to keep the suggested file name. It is reasonable to save several default files corresponding to those keys with lowest SOR values. It is necessary to save different keys. It happens regularly that a set of spectra can be nicely interpreted with a certain key independent of the starting values. Then, all repetitions have the same, low SOR. It would be useless to save each of them. The default files will be helpful later because they provide a 'one click' access to reasonable test data in the search for quantitative interpretations. The folder "Spectra_Evaluation" already has some default files for illustration. Do not hesitate to overwrite them.



The next step is the search for suitable chemical interpretation within the numerically acceptable keys. There are, usually, two to five candidates. This analysis is done by the CAT item of the DO menu. The following graph shows its window. By selecting a default file (note: the directories in the graph will most probably differ from installation to installation) , loading it by clicking the red "!"-button, the content of the respective default file is transferred to the respective fields.



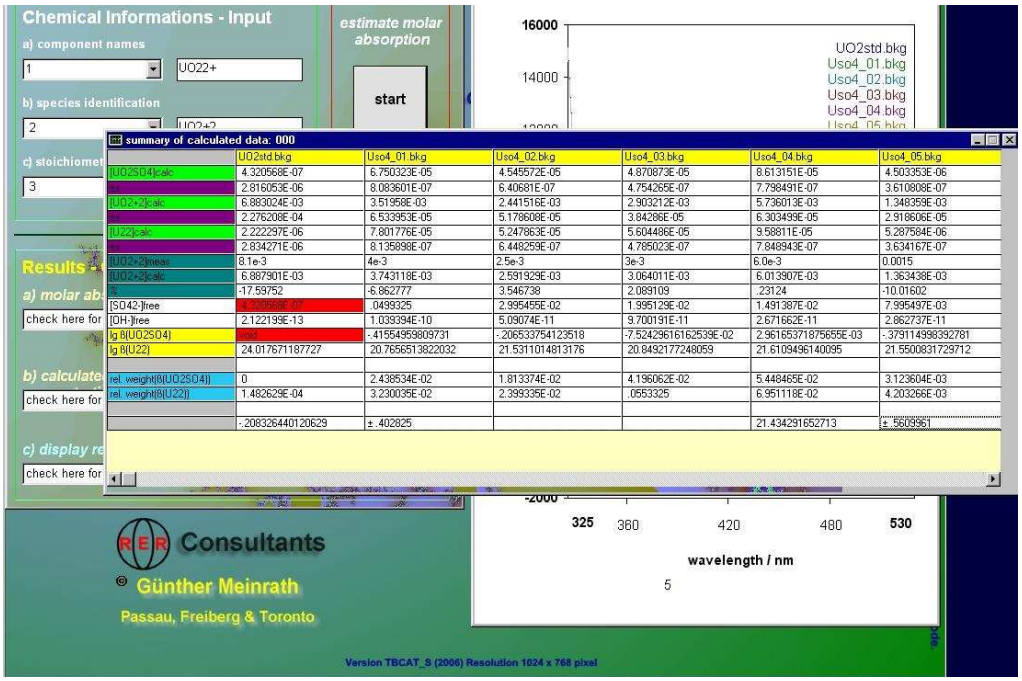
Clicking the 'run CAT' button starts the procedure. CAT is only looking for suitable spectral curves of the single components. The top yellow window gives some informations on the convergence process of the SIMPLEX. Important are the values under the 'Convergence' header and the value of the SOR. Furthermore, two graphics windows appear.



The green window shows the spectrum with known spectral information (here the UO2std.bkg file) with the red spectrum giving the best approximation CAT can obtain. The right side window shows all three (the length of the key "110" is three - therefore three species are assumed by CAT).

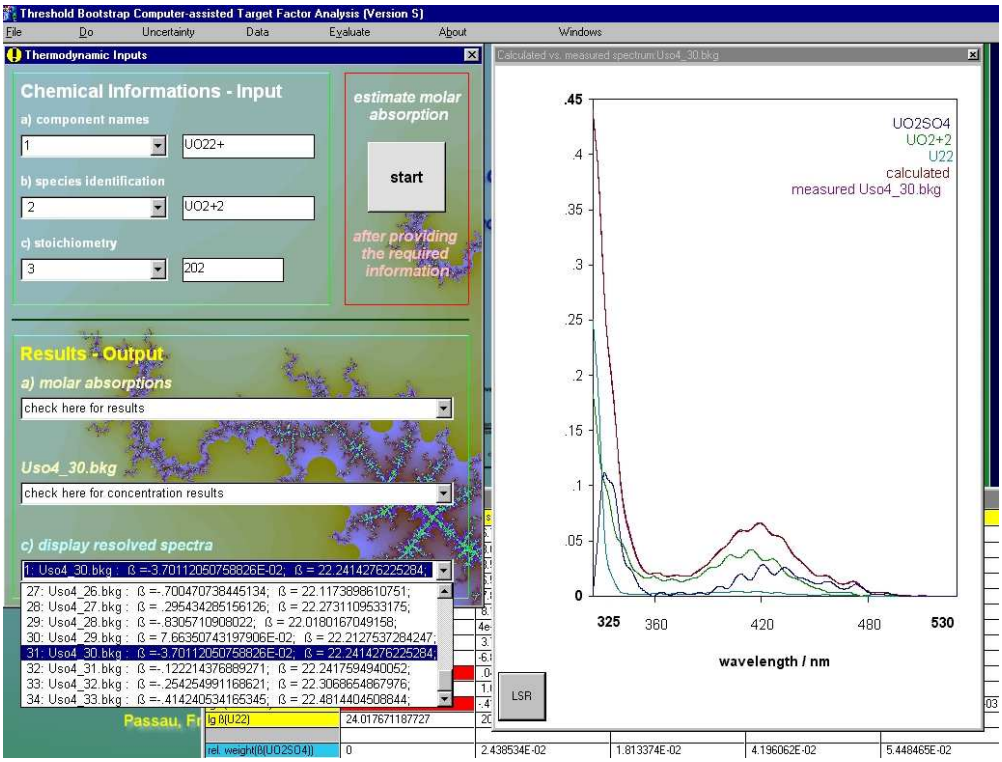
After termination of the CAT process (which often may end in a 'fatal error' message) the results are given. The graphical presentation of the single component spectral curves occasionally allows a general judgement on the feasibility of the interpretation and even a first guess on the species giving rise to the spectrum.

The scaling of these potential single components is done by the 'Molar absorption" item under the DO menu. Here, the chemist must provide good guesses on the possible species in solution and their likely stoichiometric composition. The solutions in this example have three components: UO_2^{2+}, SO_4^{2-}, and OH^- (expressed as pH). The component with the known spectral information (here: UO_2^{2+}) has always the second position. Its stoichiometric composition is 100 (one UO_2^{2+}, no SO_4^{2-} and no OH^-), while $(UO_2)_2(OH)_2^{2+}$ has stoichiometric composition 202, while UO_2SO_4 is represented by 110. These values have to be entered in the drop-down fields of the "Chemical Information - Input" frame. The process starts with the big 'Start' button. An SIMPLEX procedure starts trying to scale the single component spectra that all spectra can be interpreted with minimum deviation between known total concentration(s) and minimum spectral deviations etc. The result is given graphically and as a table, as shown in the next graph.

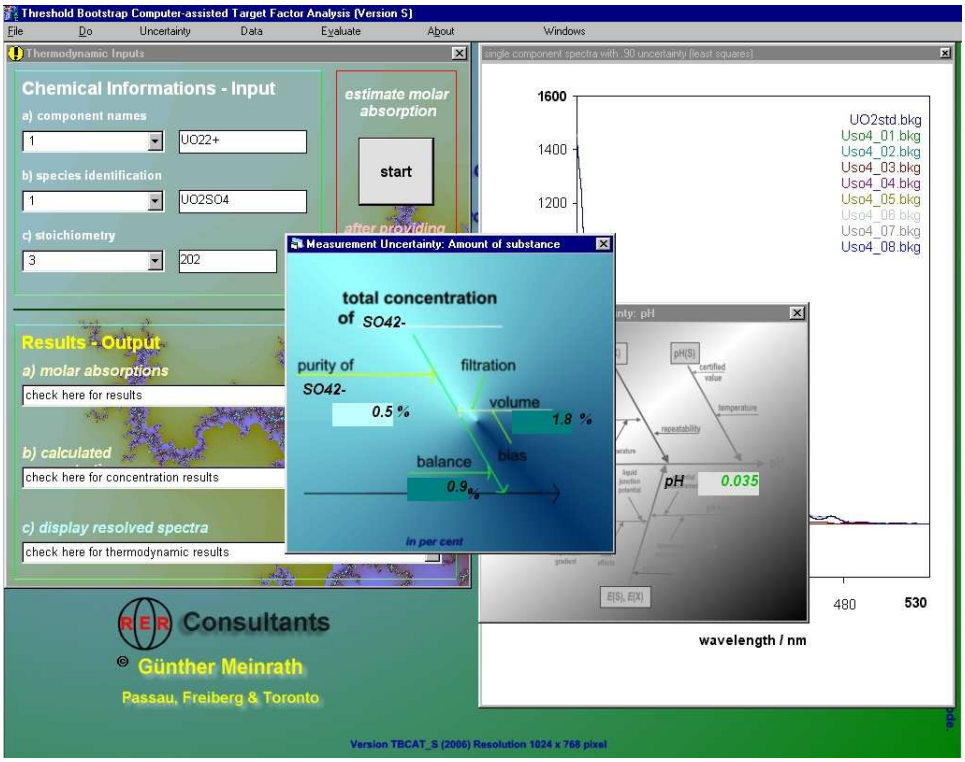


Red fields indicate physically impossible values (commonly negative concentrations or absorptions). Computers do not like the numerical value 'zero'. Small negative values are not significant - the meaning generally is 'zero'. Larger negative values, however, render an interpretation as questionable - or the input data as inconsistent. In the bottom line of the "Summary of calculated data" table, the formation constants and a first estimate of the variability are given. Some other information of interest characterising the fitting process and its result are given in this table.

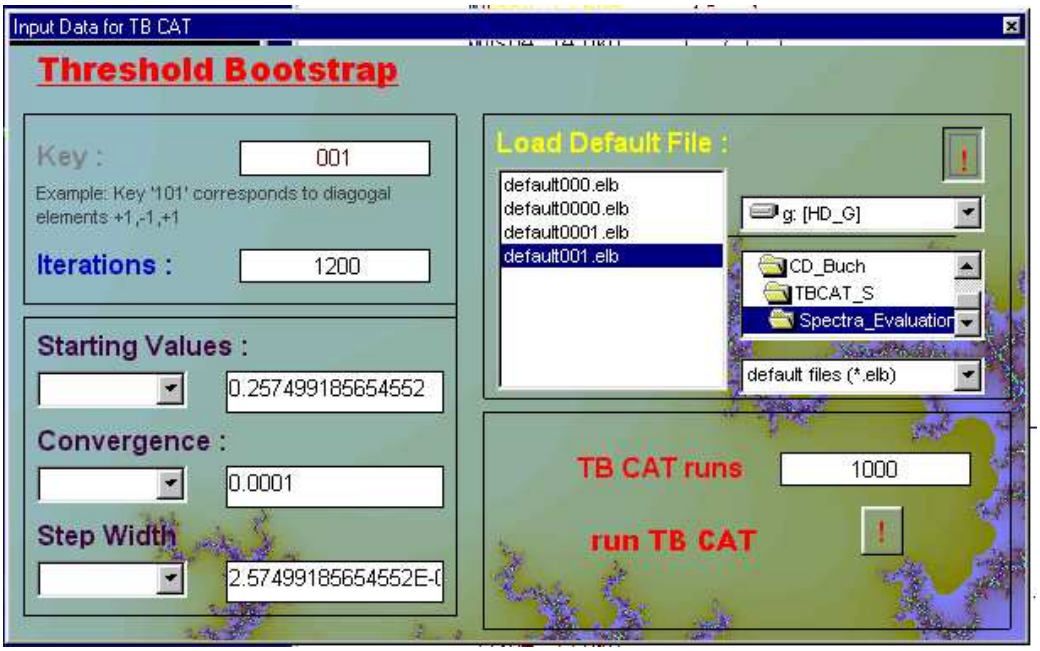
The 'Results - Output' section of the 'Thermodynamic Inputs' frame summarises further information. Selecting a spectrum under the 'display resolved spectra' drop-down list draws the spectrum, the contributions of the single components and the measured and calculated spectrum. This is shown in the following graph. The 'LSR' button performs a more detailed least-square analysis on basis of the QR algorithm. These calculations are quite time-consuming. The graphical data can be saved by selecting the SAVE option in the 'File' menu. The same is true for the summary table. The information is saved as ASCII file.



If an acceptable (and, hopefully, unique) interpretation is identified, the TBCAT process may start. TBCAT uses computer-intensive statistics to simulate the measurement process numerically. The result is a first estimate of the complete measurement uncertainty budget of the measurement process. The uncertainties for different influence quantities can be entered The uncertainty input frames are accessed via the 'Uncertainty' menu item.



TBCAT requests the fitting information and the chemical information successively. In addition the 'TB CAT runs' item in the respective frame is enabled. A minimum of 1000 repetitions should be done.



After clicking the 'run TB CAT' button, the frame for the chemical informations appears. The required data should be correspond to the values previously identified in the CAT runs. The TBCAT process is time-consuming. A careful analysis with 2000-5000 TBCAT cycles easily requires several hours even on a GHz CPU.

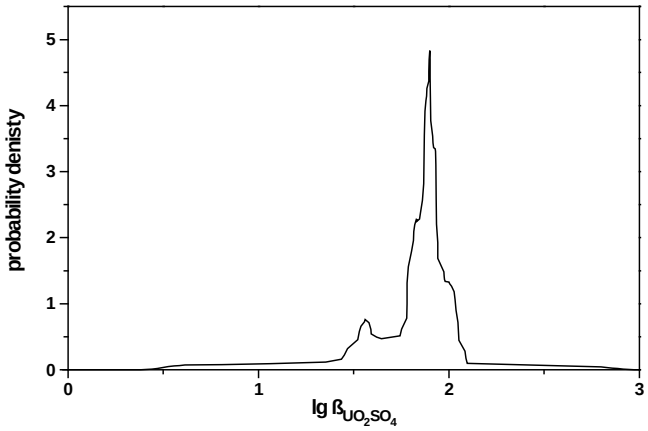
The directory 'Spectra_Evaluation' holds the output of a test run with just 75 repetitions for illustration purposes. TBCAT builds up a large number of output files holding the spectral and the numerical information from the individual runs. This bunch of information is transformed into empirical distributions and probability densities under the 'Evaluate' menu item. 'Spectral uncertainty' evaluates the empirical cumulative probability distribution for each wavelength, reads the confidence limits, and writes this information into an output file with defined file name.

The process requires that the confidence limits are checked and the single component spectrum of interest is chosen. The 'Spectral Uncertainty' frame lists the first spectrum of all

available components from which one can be selected. The process may take some time depending on the total number of TBCAT cycles and the number of wavelengths in the files. The result is save under a name derived from the user input for the respective species. The defaults are 'species(1)', 'species(2)' and 'metal ion'. If more characteristic species names have been selected, e.g. 'UO2SO4' the result will appear as ' CDF_UVUO2SO4001_3.dat'. As additional information, the filename holds the key and the position of this species.

From each TBCAT cycle, formation constants of the respective species are obtained. These are automatically saved after finishing the TBCAT process. The filename convention is 'cdf_UO2SO4001.dat' for a species 'UO2SO4'. Selecting the 'Differentiate' item from the 'Evaluation' menu, the respective frame will appear.

The probability density will be saved with the prefix 'dif'. On basis of the 75 test cycles, the following distribution for the formation constant lg ß$_{11}$ of the UO_2SO_4 species has been obtained:



Reference:

Meinrath G, Lis S, Piskula Z, Glatty Z (2006) An application of the total measurement uncertainty budget concept to the thermodynamic data of uranyl(VI) complexation by sulfate. J Chem Thermodynamics 38: 1274 - 1284.