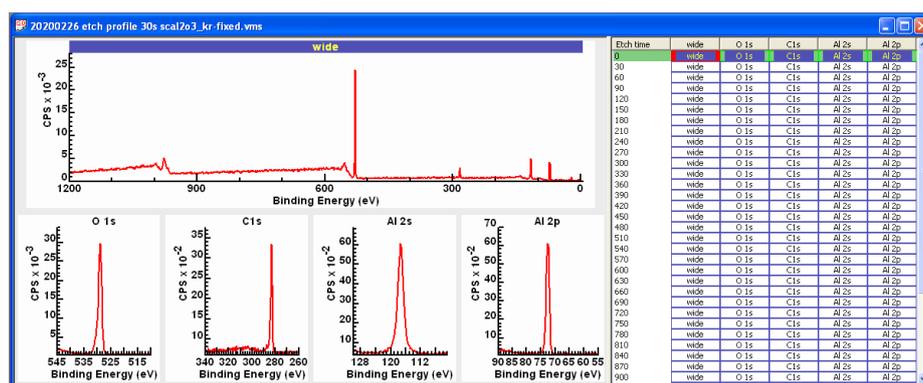


Quantification from Narrow Scan Spectra

A VAMAS file contains VAMAS blocks, where each VAMAS block holds data. When a VAMAS file is opened in CasaXPS, data are read and displayed in an experiment frame which is a window divided into two panes. The left-hand-pane displays data such as spectra or images, while the right-hand-pane displays a logical arrangement of VAMAS blocks within the VAMAS file. Before quantification is performed, a VAMAS file when displayed in CasaXPS must appear with a specific format for VAMAS blocks when visualised via the right-hand-pane of an experiment frame. The arrangement of VAMAS blocks allows quantification information to be gathered from different spectra all acquired as part of an experimental unit aimed at characterising a sample surface.



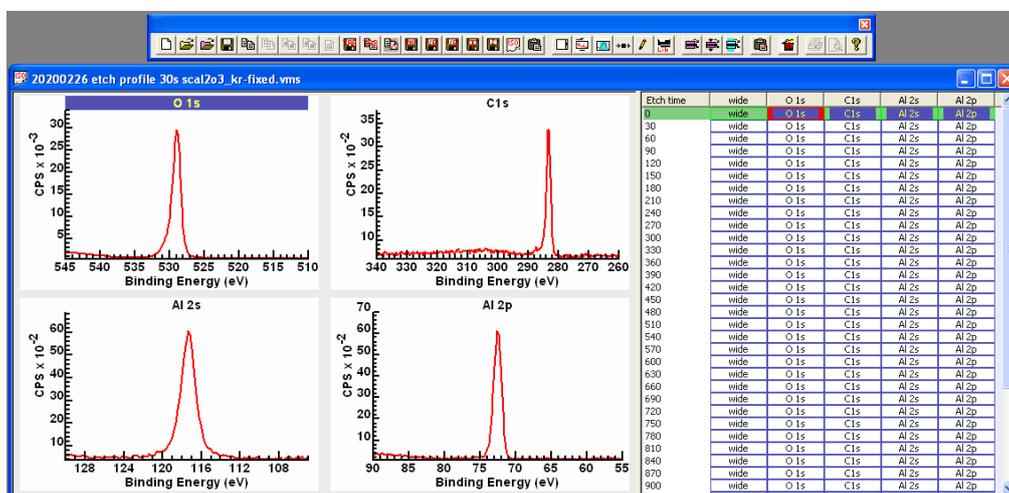
When a VAMAS file is loaded into CasaXPS, the right-hand-pane arranges VAMAS blocks making use of the experimental variable VAMAS block field to align rows of VAMAS blocks. The experimental variable is a value assigned to spectra acquired from a sample in a given state during the course of an experiment. For example, a sample that is repeatedly etched using an ion beam followed by a sequence of XPS measurements, the state of the sample surface following an etch cycle can be defined by specifying the elapse time for ion beam sputtering. Columns of VAMAS blocks are formed based on a string constructed from two VAMAS block fields, namely, the element string and the transition string which together specify the origin of spectra held within the VAMAS block.

For each VAMAS block within a VAMAS file that is intended for use in computing percentage atomic concentration, quantification items must be defined, meaning regions defining an energy interval and background and/or component peaks fitted to data are available to compute signal intensity for photoemission peaks. VAMAS blocks organised into rows in the right-hand-pane link together data acquired at the same time from the same state of the sample surface. Selecting VAMAS blocks in the right-hand-pane in the same row specifies a subset of spectra that are appropriate for computing percentage atomic concentration for a specific measurement.

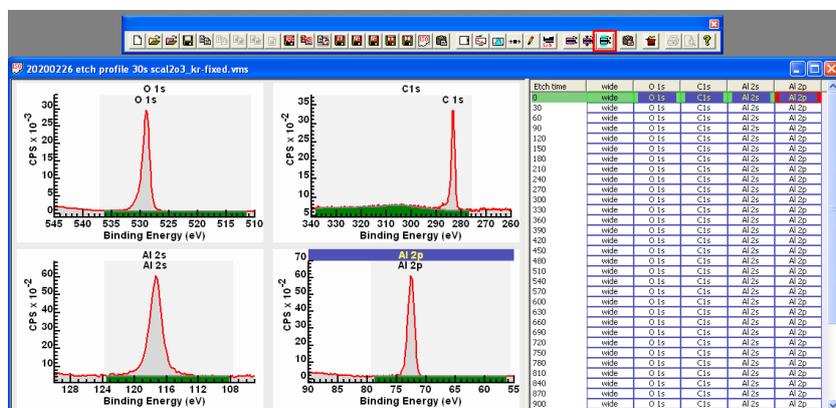
The example of a VAMAS file used here contains a sequence of XPS measurements interleaved by sputter cycles designed to remove adventitious carbon. Each XPS measurement includes a survey spectrum and four narrow scan spectra. The narrow scan spectra include data collected over energy intervals corresponding to O 1s, C 1s, Al 2s and Al 2p. Two of these narrow scan spectra are acquired using energy intervals recording signal from Al atoms. When quantification in terms of elemental composition is performed only one of these Al spectra per row should be selected together with O 1s and C 1s spectra.

Survey spectra also include photoemission signal measured from O1s, C 1s, Al 2s and Al 2p. However, these photoemission peaks are measured as part of survey spectra using different acquisition conditions from the equivalent energy intervals measured using narrow scan spectra saved as separate VAMAS blocks. Therefore these survey spectra should not be used as part of an atomic concentration calculation involving narrow scan spectra of different energy resolution. While it is possible mixed-operating mode data can be used to quantify a sample, an instrument used in this way would need careful characterisation to ensure the same analysis area and angular acceptance at the sample are measured for all operating modes used to collect spectra assumed to be equivalent.

In this example, quantification regions will be used to define photoemission for each narrow scan spectrum. Peak models defined in terms of components representing chemical state signal could equally well be used to measure photoemission intensity, but for simplicity in terms of discussing quantification of narrow scan spectra, regions alone will be used in this discussion.



Regions can be added to spectra for which the correct element/transition fields are used within the VAMAS blocks representing narrow scan data. Each of these four narrow scan data are correctly assigned therefore selecting VAMAS blocks before pressing the indicated toolbar button creates regions for each spectrum so indicated. The element/transition VAMAS block fields allow the creation mechanism to extract appropriate RSFs from the current element library.

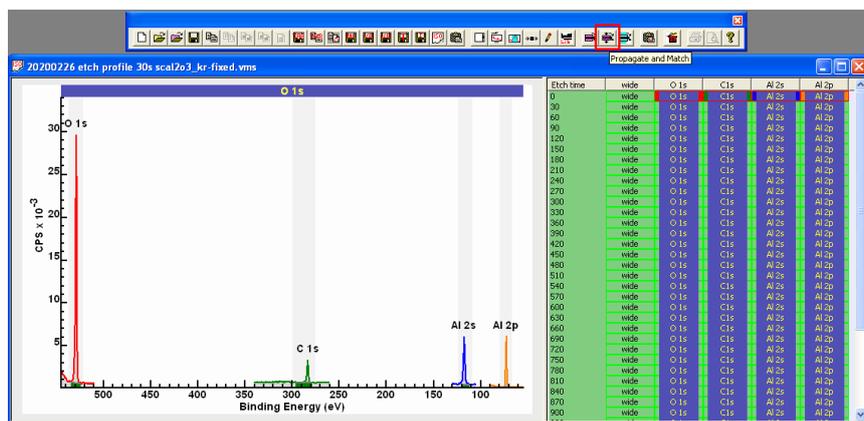


The same toolbar button would not work for survey data. Survey data, by definition, is not specific to an element or transition therefore it is not possible to use element/transition VAMAS block fields

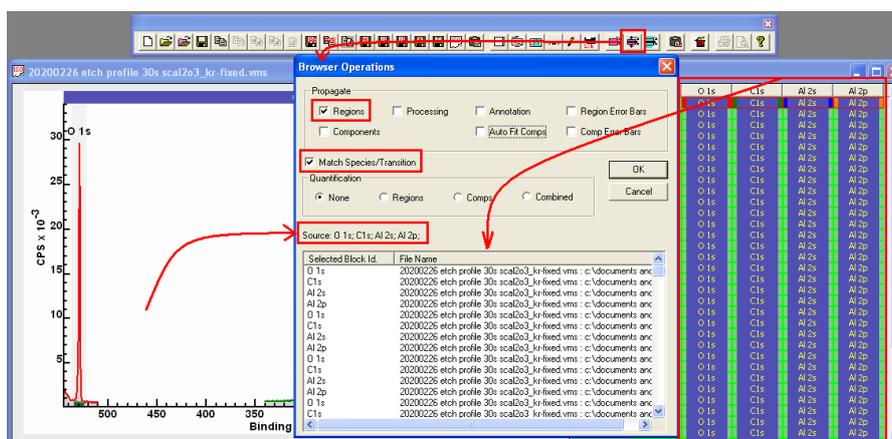
information to link RSFs from the element library with newly created regions via the toolbar button. Options on the Element Library dialog window are used to manage the creation of regions on survey data.

The background type used to create these regions is the last-used background type. In this example a Tougaard background was the default background type specified in the ParameterFile.txt configuration file found in the directory CasaXPS.DEF. The specific Tougaard background makes use of cross-sections parameters defined by Tougaard for polymer data. Since these data selected include adventitious carbon, there is at least some consistency in adopting a hydrocarbon type background approximation for these photoemission peaks. The Tougaard background is also used for the Al and O photoemission peaks on the basis adventitious carbon is an overlayer and so O and Al photoemission are attenuated by the hydrocarbon, coupled with assuming Al₂O₃ has a sizeable band gap.

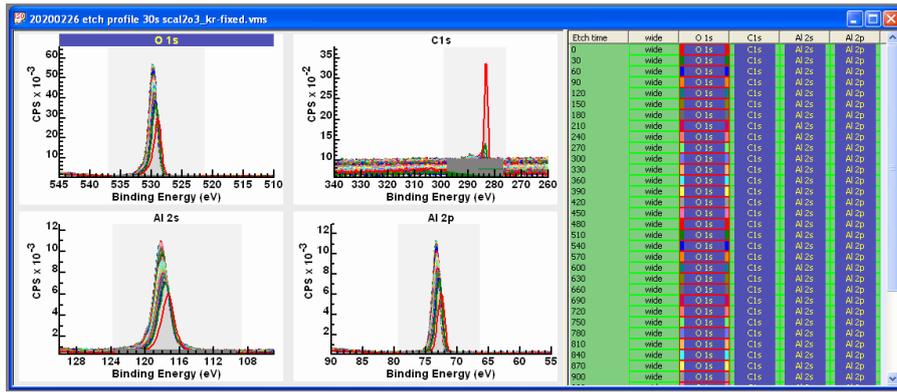
These regions defined for the first row of VAMAS blocks need further adjustments to the energy interval limits, after which these can be propagated via Browser Operations to VAMAS blocks with identical element/transition fields.



The four VAMAS blocks with regions are overlaid in the active tile. A selection of VAMAS blocks in the right-hand pane provides the target VAMAS blocks for the Propagate and Match toolbar button.



Following propagation each narrow scan VAMAS block includes a region appropriately defined to allow the computation of atomic concentration for oxygen, carbon and aluminium.



A text report generated from these regions is defined in terms of the region names fields, peak areas calculated from integrated signal after background subtraction from raw data and adjustments to integrated peak areas to allow amount of substance to be computed. These corrections involve relative sensitivity for different photoemission using RSFs, instrumental response via transmission functions included in VAMAS blocks and escape depth correction to account for sampling depth variation with energy. The instrument in question is designed using the magic angle for angular distribution correction.

The Standard Report option Regions makes use of a configuration file to create a report from regions defined on selected VAMAS blocks.

The image shows the 'Quantification Parameters' dialog box with the 'Regions' tab selected. To the right is a Notepad window titled 'RegionQuantTable.txt' containing the following configuration:

```

VARIABLE
NAME
CONCENTRATION
UNCORRECTED_RSF
RSF
TOTAL_SENSITIVITY_FACTOR
TRANSMISSION
MEAN_FREE_PATH
POSITION
FWHM
RAW_AREA
    
```

Variable	Name	%At Conc	Uncorrected RSF	Library RSF	Total RSF	Transmission	MFP	Position
0	O 1s	51.11	2.93	2.93	2.43527	1	0.831149	529.00
	C 1s	14.03	1	1.0	1.00124	1	1.00124	283.20
	Al 2p	34.86	0.5371	0.5371	0.614986	1	1.14501	72.60
30	O 1s	56.90	2.93	2.93	2.43486	1	0.83101	529.20
	C 1s	4.11	1	1.0	1.00096	1	1.00096	283.60
	Al 2p	30.98	0.5371	0.5371	0.614913	1	1.14488	72.60
60	O 1s	58.55	2.93	2.93	2.43445	1	0.83087	529.40
	C 1s	2.51	1	1.0	1.00096	1	1.00096	283.60
	Al 2p	38.95	0.5371	0.5371	0.61484	1	1.14474	73.00
90	O 1s	59.20	2.93	2.93	2.43445	1	0.83087	529.40
	C 1s	1.16	1	1.0	1.00069	1	1.00069	284.00
	Al 2p	39.64	0.5371	0.5371	0.61484	1	1.14474	73.00
120	O 1s	50.93	2.93	2.93	2.43445	1	0.83087	529.40
	C 1s	1.13	1	1.0	1.00069	1	1.00069	284.00
	Al 2p	39.95	0.5371	0.5371	0.61484	1	1.14474	73.00
150	O 1s	58.15	2.93	2.93	2.43445	1	0.83087	529.40
	C 1s	1.78	1	1.0	1.00055	1	1.00055	284.20
	Al 2p	40.06	0.5371	0.5371	0.61484	1	1.14474	73.00
180	O 1s	58.22	2.93	2.93	2.43445	1	0.83087	529.40
	C 1s	1.11	1	1.0	1.00041	1	1.00041	284.40
	Al 2p	40.67	0.5371	0.5371	0.61484	1	1.14474	73.00
210	O 1s	59.30	2.93	2.93	2.43445	1	0.83087	529.40
	C 1s	1.11	1	1.0	1.00041	1	1.00041	284.40
	Al 2p	39.58	0.5371	0.5371	0.61484	1	1.14474	73.00

After a text report is created, the Copy toolbar button is used to place text reports onto the clipboard so that results can be pasted via Ctrl+V into a spreadsheet or other programs.

Clipboard Selection

Report Copied on: Monday, April 13, 2020, 15:22:27
 c:\documents and settings\neal\my documents\workingdir\customerdata\swansea\klaudia\20200226 etch profile 30s scal

Etch time	Name	%At	Conc	Uncorrected	RSF	Library	RSF	Total	RSF	Transmission	MFP	Pr
0	O 1s	51.11	2.93	2.93	2.43527	1	0.831149	529.00	1.40	49798.5		
	C 1s	14.03	1	1	1.00124	1	1.00124	283.20	1.53	5618.99		
	Al 2p	34.86	0.5371	0.5371	0.614986	1	1.14501	72.60	1.30	8576.66		
30	O 1s	56.90	2.93	2.93	2.43486	1	0.83101	529.20	1.38	59953.5		
	C 1s	4.11	1	1	1.00096	1	1.00096	283.60	1.92	1781.27		
	Al 2p	38.98	0.5371	0.5371	0.614913	1	1.14488	72.80	1.24	10373.2		
60	O 1s	58.55	2.93	2.93	2.43445	1	0.83087	529.40	1.37	63318.5		
	C 1s	2.51	1	1	1.00096	1	1.00096	283.60	2.19	1114.43		
	Al 2p	38.95	0.5371	0.5371	0.61484	1	1.14474	73.00	1.28	10637.8		
90	O 1s	59.20	2.93	2.93	2.43445	1	0.83087	529.40	1.36	64048.4		
	C 1s	1.16	1	1	1.00069	1	1.00069	284.00	1.85	515.763		
	Al 2p	39.64	0.5371	0.5371	0.61484	1	1.14474	73.00	1.26	10832.7		
120	O 1s	58.93	2.93	2.93	2.43445	1	0.83087	529.40	1.37	65400.9		
	C 1s	1.13	1	1	1.00069	1	1.00069	284.00	1.62	513.579		

Buttons: OK, Save..., Copy Selection to Clipboard, Copy All to Clipboard

Book1 - Microsoft Excel

Report Copied on: Monday, April 13, 2020, 15:22:27

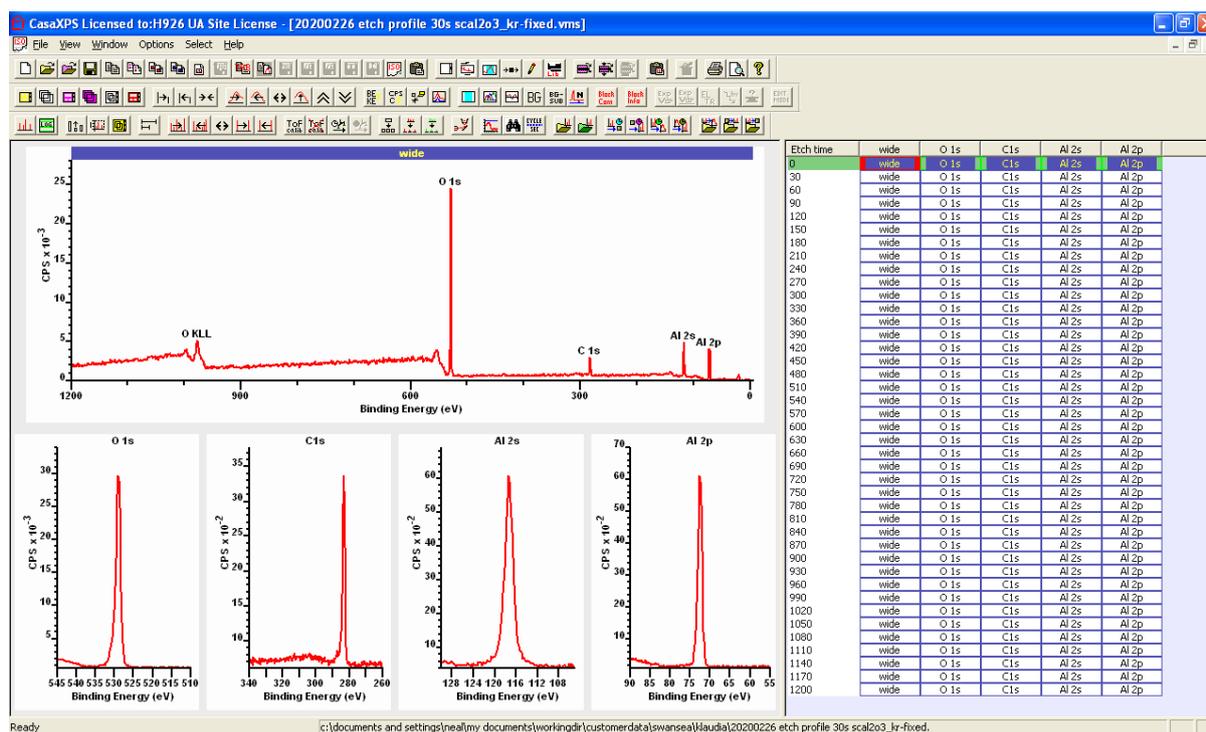
Etch time	Name	%At	Conc	Uncorrect	Library	RS	Total	RSF	Transmiss	MFP	Position	FWHM	Raw Area
0	O 1s	51.11	2.93	2.93	2.43527	1	0.831149	529	1.4	49798.5			
	C 1s	14.03	1	1	1.00124	1	1.00124	283.2	1.53	5618.99			
	Al 2p	34.86	0.5371	0.5371	0.614986	1	1.14501	72.6	1.3	8576.66			
30	O 1s	56.9	2.93	2.93	2.43486	1	0.83101	529.2	1.38	59953.5			
	C 1s	4.11	1	1	1.00096	1	1.00096	283.6	1.92	1781.27			
	Al 2p	38.98	0.5371	0.5371	0.614913	1	1.14488	72.8	1.24	10373.2			
60	O 1s	58.55	2.93	2.93	2.43445	1	0.83087	529.4	1.37	63318.5			
	C 1s	2.51	1	1	1.00096	1	1.00096	283.6	2.19	1114.43			
	Al 2p	38.95	0.5371	0.5371	0.61484	1	1.14474	73	1.28	10637.8			
90	O 1s	59.2	2.93	2.93	2.43445	1	0.83087	529.4	1.36	64048.4			
	C 1s	1.16	1	1	1.00069	1	1.00069	284	1.85	515.763			
	Al 2p	39.64	0.5371	0.5371	0.61484	1	1.14474	73	1.26	10832.7			

Excel status bar: Average: 4422.700271 Count: 5656 Sum: 14143795.47

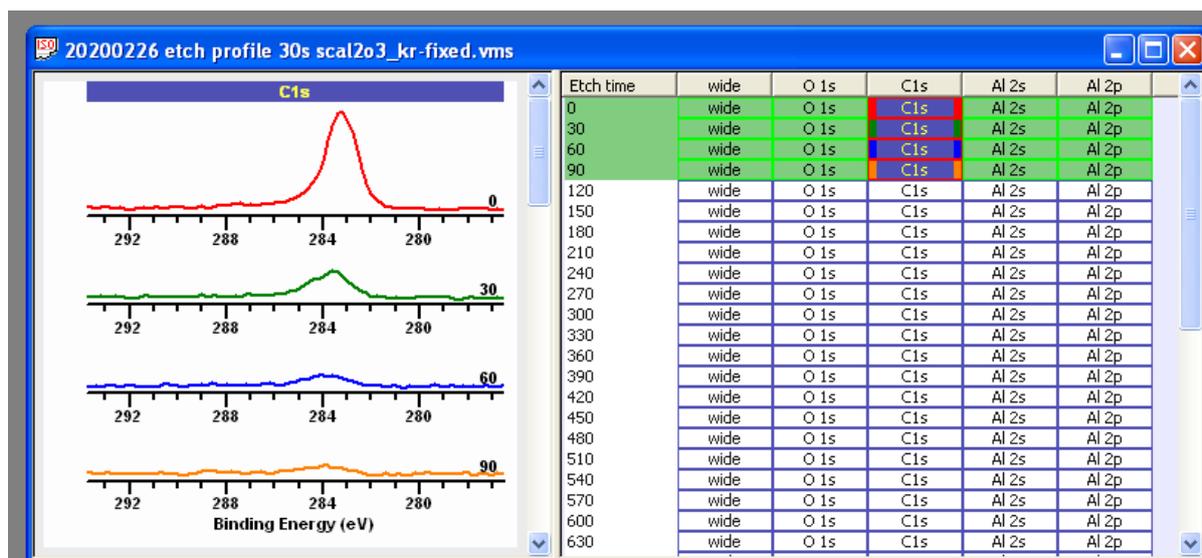
Data Analysis by Example

The first steps in understanding samples of scientific interest is to develop strategies for the measurement and analysis of these samples using samples with known composition similar in character to the samples of interest. The objective when measuring samples of lesser interest but of known substance is to evaluate different measurement protocols and data treatment options that will be of use when considering samples of greater scientific interest.

The data set considered here is part of a preliminary study considering chemistry related to aluminium oxide. A sequence of measurements was performed using a sample prepared as Al_2O_3 . Each measurement consists of a survey spectrum (meaning data are acquired over an energy interval equivalent to photoemission peaks accessible by mono Al $K\alpha$ X-ray source) and a set of spectra measured over narrow energy intervals focused on photoelectron peaks of significance to Al_2O_3 . The function of a survey spectrum is to monitor photoemission not included in narrow scan spectra and just as importantly provides information about trends in background signal. Shapes within the background can indicate changes in in-depth distribution of material and in this case supports the assumption the material remains, despite repeated sputtering with an ion beam, essentially homogeneous in depth. Narrow scan spectra form the basis for chemical state analysis and are collected using smaller energy steps and increased dwell-time compared to the survey measurement to improve quality of curve shapes in data and signal-to-noise.

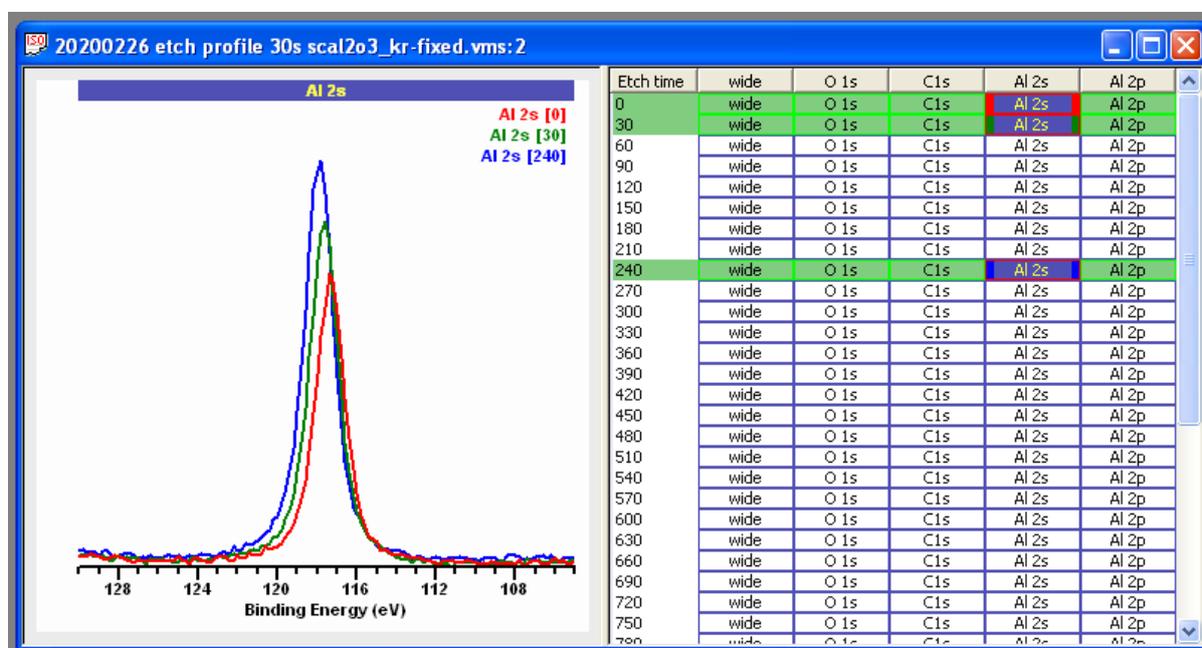


The data set measured from the same location on the same sample of Al_2O_3 is acquired by measuring spectra followed by sputtering the sample with relatively low energy argon ions. The purpose of sputtering the surface with an ion beam is to gently remove contamination from the Al_2O_3 surface. Evidence of ion beam sputtering includes changes to C 1s narrow scan spectra. Also, charge state alters with ion beam etch time which further indicates changes have occurred to the surface measured by XPS as a consequence of sputtering with the ion gun. These changes in charge state with etch time requires the use of charge correction to align spectra throughout the data file.



Charge Correction

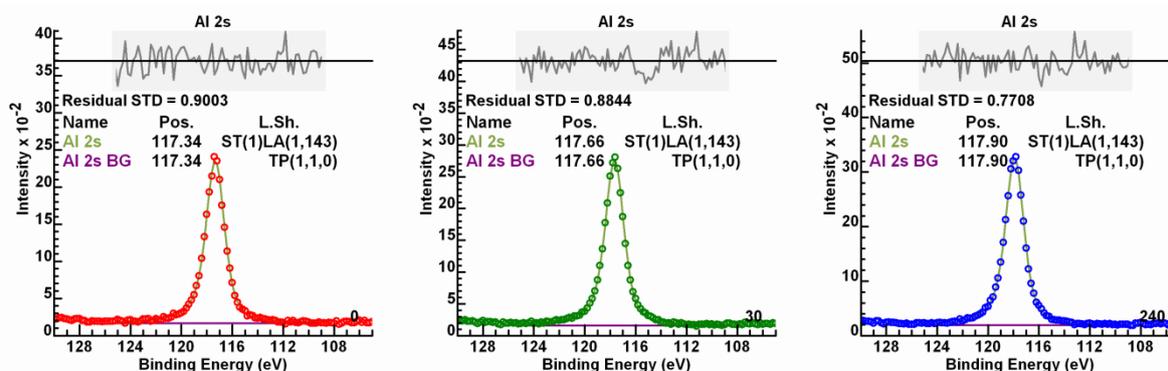
Data recorded from an insulating material such as Al_2O_3 requires the use of charge compensation, the mechanism by which a steady state charge distribution is established by returning electrons to the sample to compensate for the emission of electrons from the sample due to irradiation by X-rays. For an experiment involving sputtering with an ion beam or even simply small changes in analysis conditions and/or sample chemistry with exposure to X-rays may result in changes to the steady state charge on the sample for each XPS measurement. As a consequence each measurement in the Al_2O_3 data set represents a different charge compensation state and requires calibration in terms of binding energy scale to align signal with respect to energy.



Binding energy is calibrated by applying a shift in energy to each spectrum considered to be acquired under identical charge compensation conditions. An appropriate shift is calculated using a peak of known binding energy and shifting the measured spectra to position the peak maximum at the known energy. The common practice of calibrating spectra using C 1s as the reference peak has in

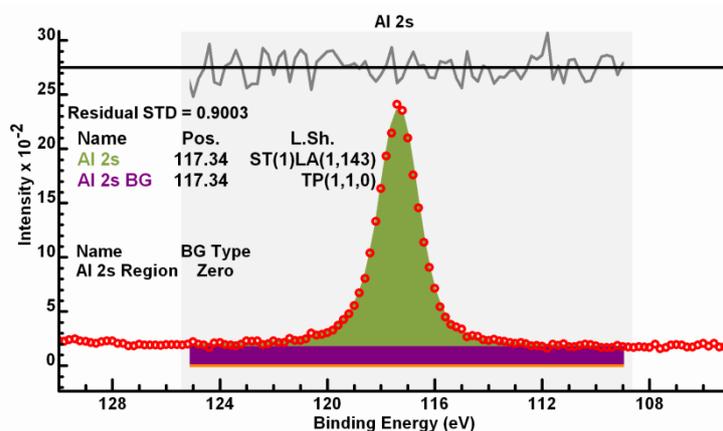
recent times been called into question. Traditionally spectra are aligned to place the lowest binding energy C 1s component peak, established by fitting a model to adventitious C 1s data, at a binding energy of 285 eV. Such an approach has merit in providing improved precision for aligning photoemission peaks with binding energy but lacks accuracy due to carbon potentially being part of the sample chemistry and the assumption that adventitious carbon is a uniform layer over the sample surface. The latter point is relevant for samples with islands of contamination, in which case charge compensation state for such islands may not be the same as for fully exposed substrate. Charge compensation in the case of sputtered Al₂O₃ is a particular problem for binding energy calibration based on adventitious carbon since the cleaning action of the ion beam is intended to remove adventitious carbon. For this reason charge correction for each row in the Al₂O₃ data set will be calibrated using the Al 2s to align spectra in binding energy rather than making use of C 1s.

Al 2s is a broad peak with Lorentzian character. Both these characteristics can be exploited to provide a precise systematic tool for identifying the maximum intensity for photoemission from Al 2s. While aluminium chemistry is the reason for making these XPS measurements, the chemical differences within these spectra result in relatively minor changes to Al 2s and Al 2p compared to changes in C 1s and O 1s spectra. Hence the use of Al 2s to calibrate binding energy for each measurement is seen as preferable to the use of C 1s data for charge correction. The approach adopted here is to use a synthetic component representative of Al 2s fitted to the Al 2s spectrum in each row of VAMAS blocks within the VAMAS file and shift each row of VAMAS blocks representing spectra acquired under identical charge compensation conditions by aligning these Al 2s synthetic components throughout the sputter sequence.

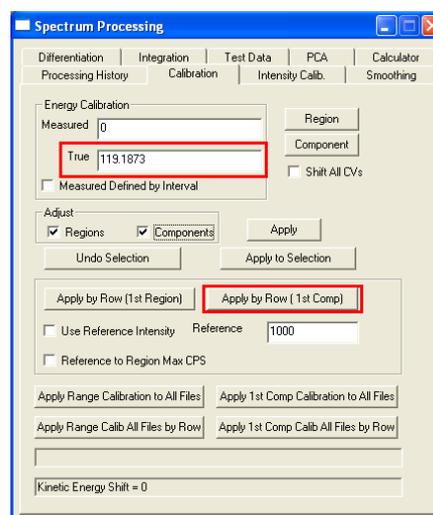
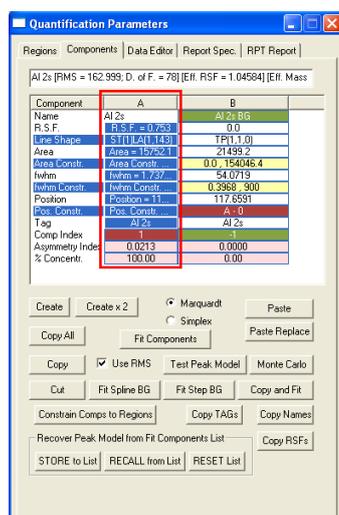


These Al 2s peaks include extended tails that can be modelled using the line shape LA(1,143) modified by asymmetry using the prefix ST(1) to the LA line shape. The objective is to fit each Al 2s spectrum with a precision consistent with a residual standard deviation of close to 0.9. A value for the residual standard deviation of 0.9 is expected for pulse counted data combined from multiple detector streams to form data fitted by the line shape ST(1)LA(1,143). The Test Peak Model option on the Quantification Parameters, Components property page is used to obtain the line shape selected for use with the Al 2s spectra. The model fitted to these Al 2s data includes the use of TP(1,1,0) line shape representing a trapezium (TP) of equal height to both sides, that is, in this case a rectangle. The TP(1,1,0) provides a constant offset to the ST(1)LA(1,143) bell-shaped curve with Lorentzian character that is determined as part of the optimisation steps for each spectrum. Owing to a significant band gap for Al₂O₃ the background trend can be modelled using a constant

background set to zero. Thus TP(1,1,0) is the means by which background signal is accounted for when fitting Al 2s photoemission.

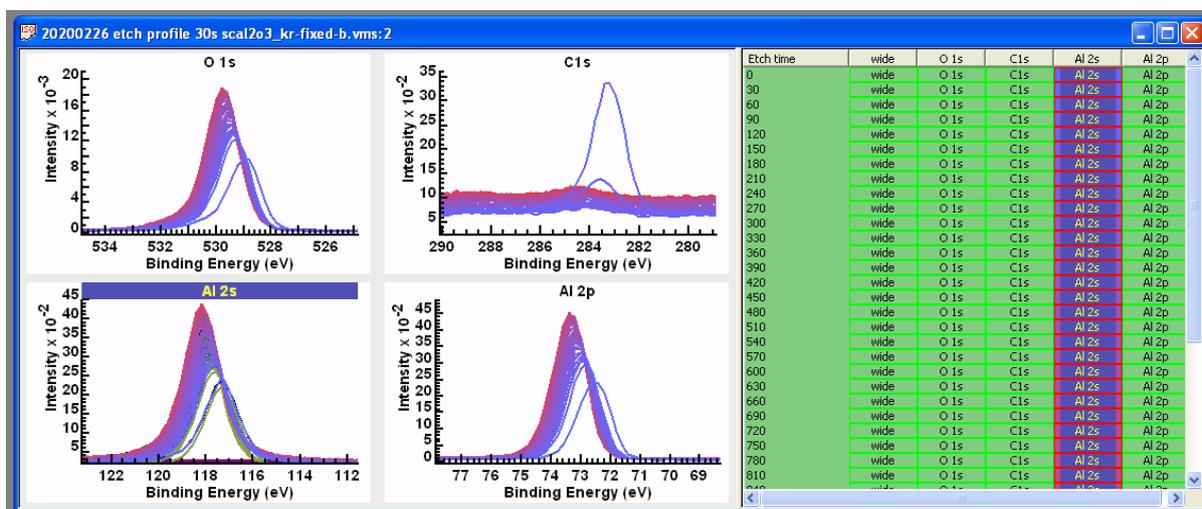


Charge correction for a data set arranged so that each measurement cycle is organised as a row of VAMAS blocks is performed by creating a peak model for a specific column of VAMAS blocks corresponding to particular spectra. In the current example the spectra used for calibrating the data file are the Al 2s spectra. The peak model for Al 2s is arranged such that the component defined in Column A in the list of components on the Component property page corresponds to the energy entered on the Spectrum Processing, Calibration property page in the text-field labelled True.

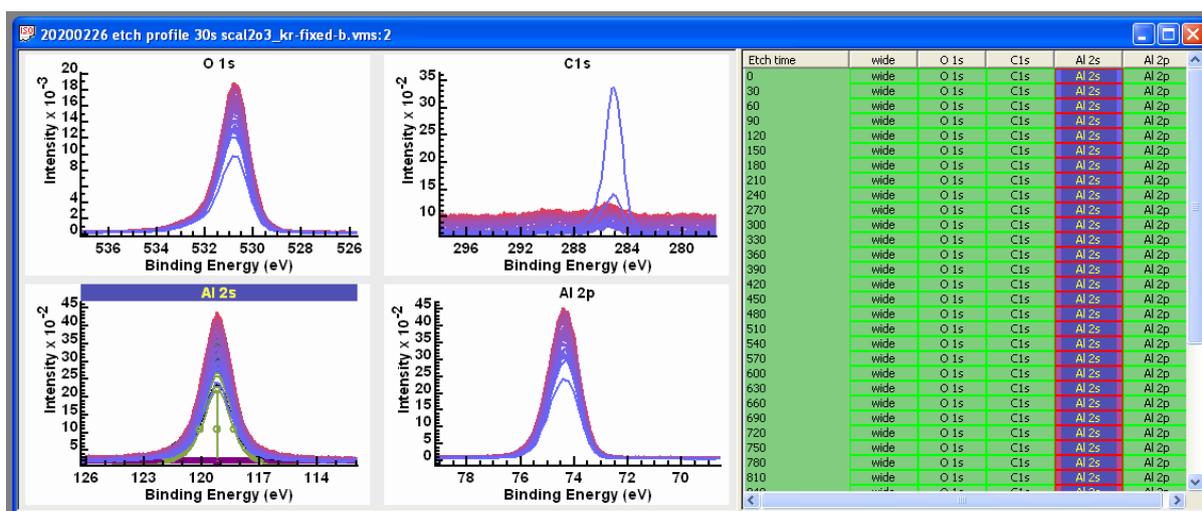


Calibration of all VAMAS blocks in the VAMAS file is performed by selecting the column of VAMAS blocks for which the peak model is defined and only that column of VAMAS blocks, then pressing the button on the Spectrum Processing, Calibration property page labelled **Apply by Row (1st Comp)**. The action of the **Apply by Row (1st Comp)** button causes an offset to be calculated using the Al 2s peak model and the value entered in the True text-field of the Calibration property page. The offset so computed is designed to align the component in Column A of the Components property page so that the energy for components in Column A matches the value in the True text-field in each row of VAMAS blocks. For each row of VAMAS blocks the offset calculated using the Al 2s spectrum is applied to these spectra in the same row of VAMAS blocks.

Before calibration is performed, an overlay of each column of VAMAS blocks in separate display tiles, one for each column of spectra with the same element/transition labels clearly demonstrates offset peaks for O 1s, C 1s, Al 2s and Al 2p.



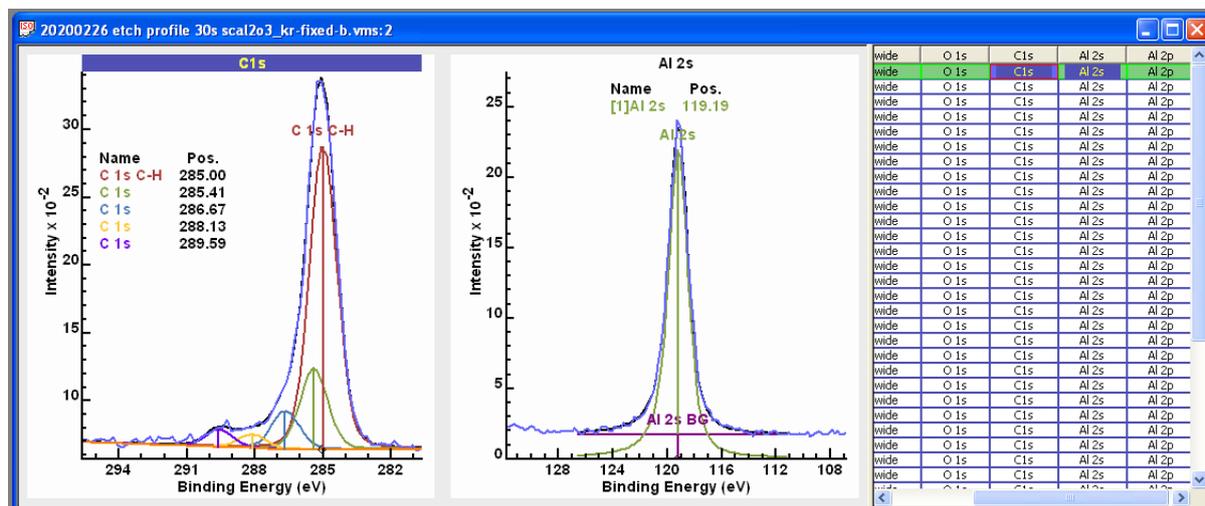
After calibration has been performed, these same display tiles illustrate the action of the **Apply by Row (1st Comp)** button. Each set of overlaid spectra are aligned and any misalignment in spectra other than Al 2s is assumed to be chemical shifts of significance to sample chemistry.



Motivation for Charge Correction

The normal motivation for charge correcting XPS data is a desire to identify chemical state by comparing binding energy assignments for peaks in data from a sample with unknown chemistry to binding energy from samples of known chemistry. The problem encountered in calibrating the binding energy scale for an insulating material is there is no absolute reference other than assuming a peak can be identified corresponding to photoemission of a known binding energy. For many samples the best we can do is assume a peak in C 1s spectra can be identified corresponding to C-H type carbon chemistry and assign this peak binding energy to 285 eV. The source for C 1s signal is assumed to be adventitious carbon. When presenting energy calibrated data, the importance of stating the calibration assumption should not be underestimated since all assumptions about assignments of other peaks reported in a paper will be dependent on this initial assumption. Indeed

the energy calibration performed here makes use of a binding energy for Al 2s, the value for which was determined from the initial measurement before sputtering with an ion-gun making use of adventitious carbon, available at this point in the experiment, to calculate the binding energy for Al 2s that would align C-H component within the corresponding C 1s spectrum with binding energy 285.0 eV. It is not certain that the precision of identifying the component in C 1s assigned to 285 eV or the accuracy in terms of the energy assigned to the component is sufficient to support Al 2s assigned to 119.19 eV, but in choosing to make such a calibration the normal motivation is achieved.



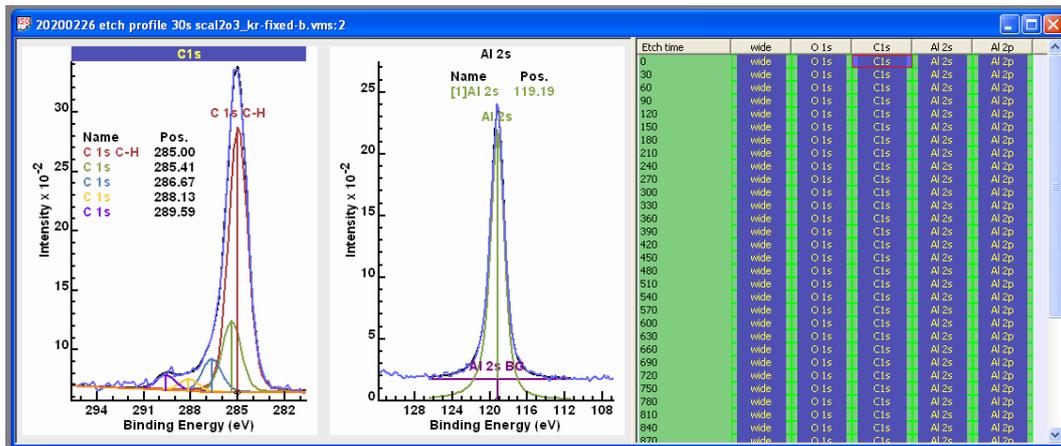
A secondary motivation is to align spectra in binding energy throughout the experiment so that data can be manipulated in the sense of vector analysis. This vector analysis could be as simple as summation of spectra to improve signal to noise or analyse columns of data into principal components through an application of singular valued decomposition (SVD) followed by noise reduction based on principal component analysis (PCA). The objective in performing these types of vector analysis is to check for consistency in these measurements and by creating signal enhanced data identify subtle changes within these data.

Verification by Vector Analysis

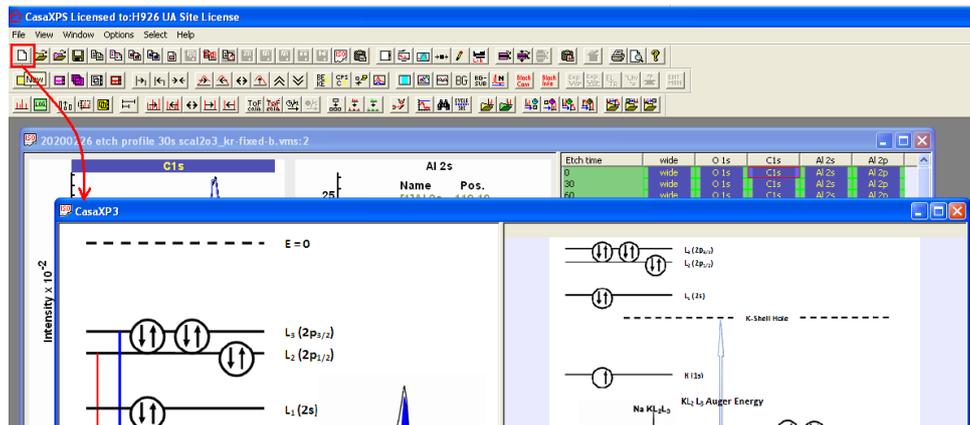
When presented with a data set in which multiple measurements from the same sample have been performed the option is available to make use of the total acquisition time to improve signal to noise in each measurement via PCA. Before PCA can be applied to data the energy scale must be calibrated to remove variation due to charge compensation state. Once energy calibration is performed, the next step is to copy the set of VAMAS blocks to a new VAMAS file extracting processed data. The reason data needs to be copied is energy calibration is performed by shifting the energy scale rather than moving intensity between data bins. PCA or any vector operation requires the movement of intensities between data bins so that data bins correlate with binding energy. Correlating data bins with binding energy is performed by first copying the processed data then an option on the Spectrum Processing, Calculator property page is used to re-bin intensities so that data bins correspond to binding energy directly.

The following steps create a new VAMAS file containing spectra from the original data set with intensities moved between data bins so that data bins correlate with binding energy.

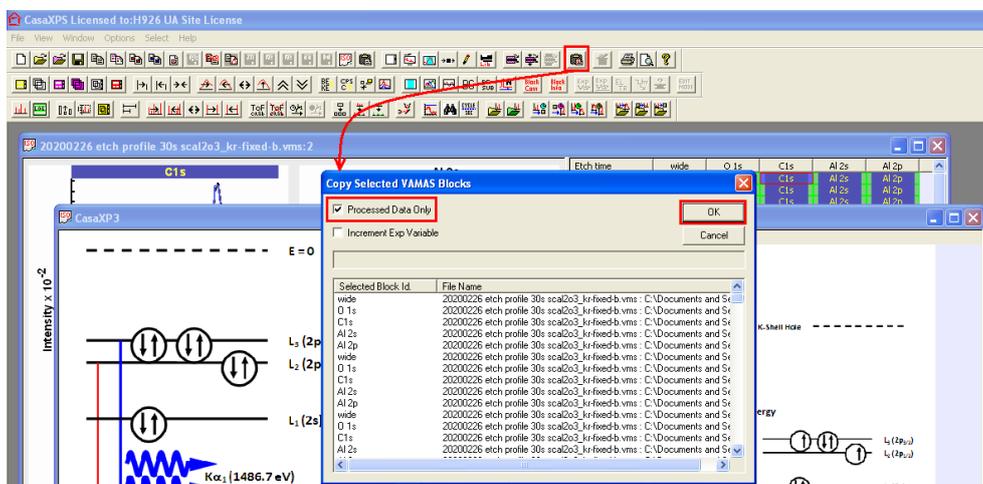
1. Select all VAMAS blocks in the original VAMAS file that have previously been calibrated in terms of binding energy.



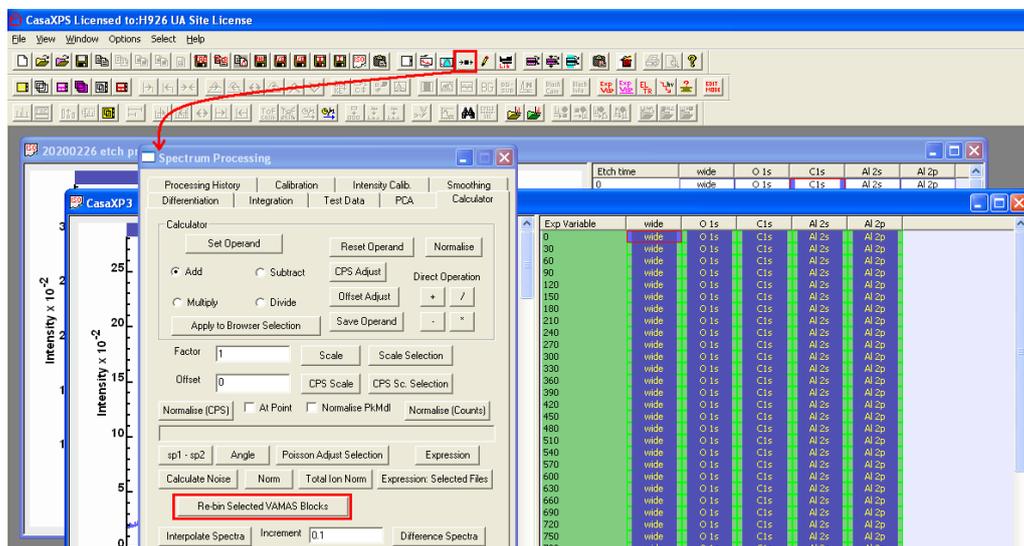
2. Using the top toolbar button or the New File menu option, create an empty experiment frame window into which processed data can be copied.



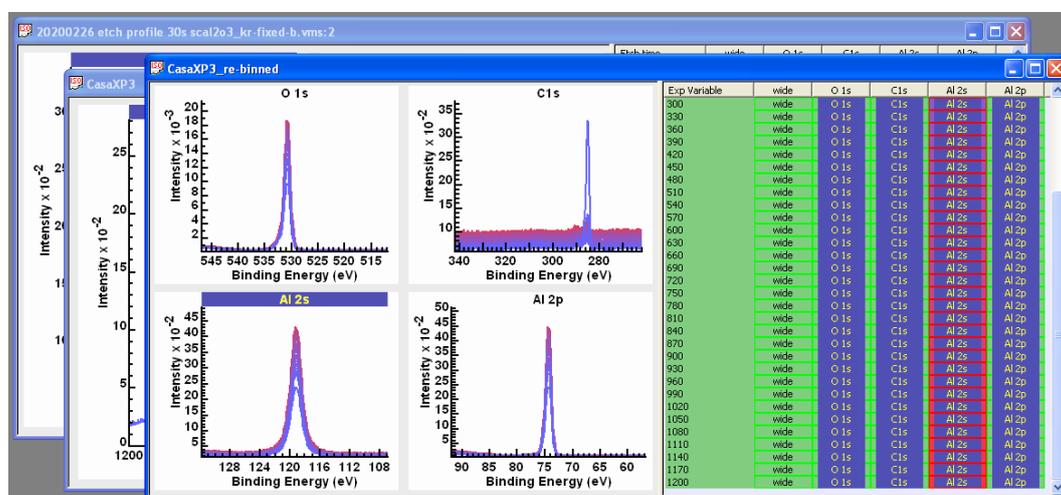
3. The empty frame window should be given focus before pressing the Copy VAMAS blocks toolbar button. Select the tick-box labelled **Process Data Only** on the Copy VAMAS blocks dialog window the press the OK button.



- Select all the VAMAS blocks in the newly populated experiment frame before invoking the Spectrum Processing dialog window and selecting the Calculator property page. The set of selected VAMAS blocks are acted upon by the **Re-bin Selected VAMAS Blocks** button to create a new experiment frame containing data re-binned so that data bins align with binding energy.



The resulting experiment frame contains data in a format suitable for vector analysis steps including PCA calculations aimed at suppressing noise in each spectrum.

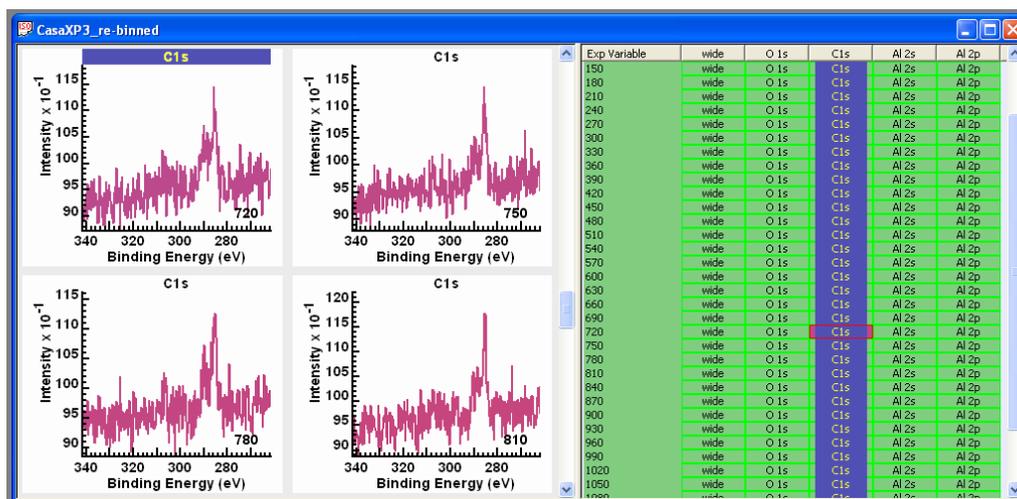


The spectra of most interest are the O 1s and C 1s. Both of these photoemission lines potentially respond to chemical state and despite low signal to noise in C 1s spectra, these C 1s data are of interest and so enhancing these C 1s spectra is an objective for the following PCA calculations.

Principal Component Analysis for C 1s Spectra

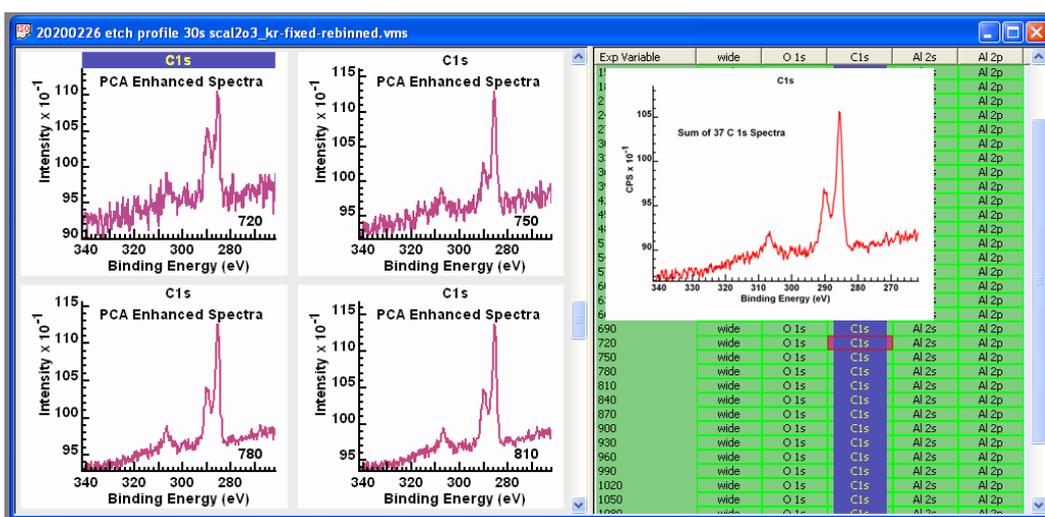
While the as-received surface provides a C 1s spectrum with adventitious carbon, following cleaning cycles of sputtering with an ion beam the nature for C 1s signal changes, reducing in intensity leading to spectra with low signal-to-noise. The shapes within C 1s signal from these sputtered surfaces are potentially of interest with respect to the chemistry of the sample, so there is a need to improve

signal to noise for these C 1s data to allow an analysis in terms of a model involving peaks representative of chemical state for carbon.



One way to investigate these C 1s spectra is to perform a PCA using all C 1s spectra.

The underlying concept of PCA is that a set of spectra can be partitioned into spectral shapes that are important to the analysis and a set of shapes that are unimportant and can be attributed to random noise. If we can determine a small number of shapes of significance and a large number of shapes representing noise, then if we fit each spectrum in a data using shapes with significance only, in a least squares sense, then the result is a set of spectra with reduced noise. The problem is therefore to first identify the number of shapes that can be considered of significance and then replace the spectra by linear least squares approximations to each spectrum. The success of such an approach depends on making use of the total acquisition time in the same sense that signal to noise in a spectrum formed by summing all spectra results in data with much improved signal to noise, but by virtue of the PCA calculation retaining individual spectral information for each measurement, at the cost of including a proportion of noise.

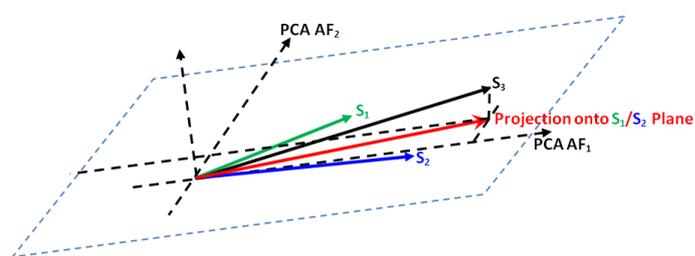


Background to SVD and PCA

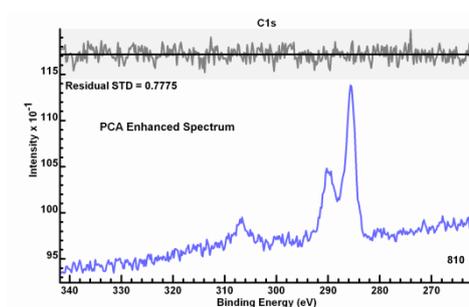
Singular Valued Decomposition creates a set of orthogonal eigenvectors that can be ranked according to corresponding eigenvalue magnitude. For this reason the results of SVD are often

referred to as Principal Component Analysis. The terminology PCA encapsulates two concepts, namely, the analysis involves component shapes in an analogous way to a peak model is constructed from bell shaped components and secondly these components can be identified as principal meaning we are able to order these components so identified as some being more important than others. SVD is a mathematical procedure for computing eigenvectors and eigenvalues of real symmetric matrices. PCA is the interpretation of these mathematical results in terms of data description.

In essence, from the perspective of SVD a set of spectra are treated as vectors in a multidimensional vector space. The act of performing SVD generates a set of orthogonal vectors that form a natural coordinate system for the data set as a whole. The adjective natural applied to an n-dimensional coordinate system conveys the meaning that one can order the basis vectors by magnitude of associated eigenvalues, and the basis vector with largest eigenvalue points in the direction of the greatest variance within the original data set.



Given two linearly independent vectors (spectra) S_1 and S_2 , a plane is defined by these two vectors. SVD computes from these two vectors S_1 and S_2 two new vectors PCA AF₁ and PCA AF₂ that are orthogonal and also lie in the same plane as S_1 and S_2 . A third vector S_3 is approximated by the projection of S_3 on to the plane P . The projection can be defined in terms of $P = \alpha S_1 + \beta S_2$ or $P = \gamma \text{PCA AF}_1 + \eta \text{PCA AF}_2$ with the only difference between these two ways of specify P is the PCA vectors can be ordered in terms of significance.



The ordering of vectors by eigenvalues is very useful because the eigenvector with smallest eigenvalue contains the smallest variation in signal and this is noteworthy because such a characteristic would be a typical goal for fitting of data with curves. That is, a necessary condition for a peak model constructed from bell shaped curves fitted to data in an acceptable form is the difference between the sum of all curves and the spectrum should exhibit a uniform shape without structure. For a peak model the removal of intensity deemed to be of least significance is precisely the outcome of an optimisation procedure that fits components to data. This same logic applied to curve fitting using bell shaped curves can also be applied to SVD results, namely, if a data set is projected onto the n-1 eigenvectors with largest eigenvalues the result is a data set with less variation than the original data set. If the vector with smallest eigenvalue contains no significant variation or put another way, simply has the appearance of noise, then SVD provides a means of reducing the influence of noise within each of the original data vectors through the act of projecting onto the subspace spanned by the eigenvectors with the largest n-1 eigenvalues. The same reduction in dimension can be repeated for each eigenvector starting with the smallest eigenvalue and continuing until it is deemed an eigenvector contains shape information of significance to the interpretation of spectra. The process of removing eigenvectors from the original data in order of least significance transforms the dimension for the original data set into a subspace of dimension defined by the number of eigenvectors of significance onto which the original data set is projected.

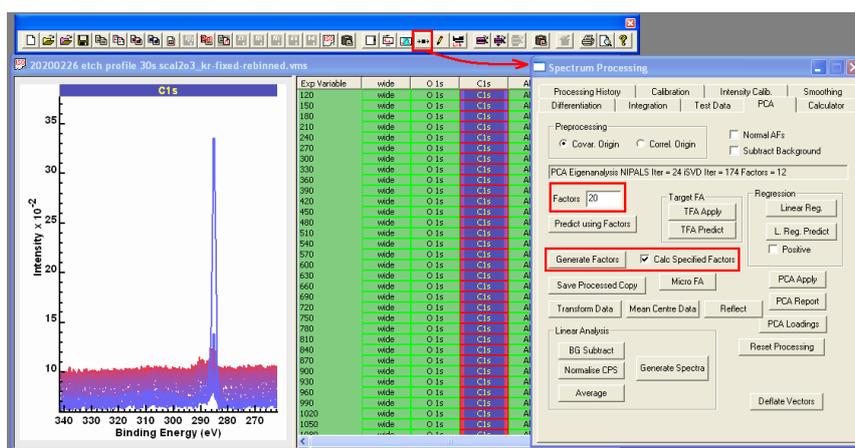
The outcome to such a procedure is spectra with the influence of noise minimised and hence spectra width enhanced spectral shapes. A measure of success for SVD/PCA noise reduction is the residual standard deviation should be close to but below unity, and the appearance of the residual is uniform for the entire energy interval.

The first step is to inspect the abstract factors (AF) for a PCA calculation. Abstract factor are computed as part of SVD and are a set of abstract spectral forms with the property that each abstract factor is orthogonal to all abstract factors computed for a set of spectra. The most significant result of the SVD algorithm is these abstract factors are not simply mutually orthogonal but are also determined by SVD such that the first abstract factor points in the direction of the greatest variance for the spectra involve by virtue of a least squares criterion. The iterative SVD (iSVD) algorithm used in CasaXPS computes for n initial spectra the first abstract factor by dividing the processed data into the first abstract factor and a set of processed vectors with the property that all vectors in the n-1 set of processed vectors are orthogonal to the first abstract factor but not necessarily mutually orthogonal. The procedure is then applied to the set of n-1 processed vectors to create the second abstract factor for PCA leaving n-2 processed vectors in a state suitable for computing subsequent abstract factors. The user interface for PCA in CasaXPS reflects the sequential determination of abstract factors.

PCA Options and the PCA Property Page

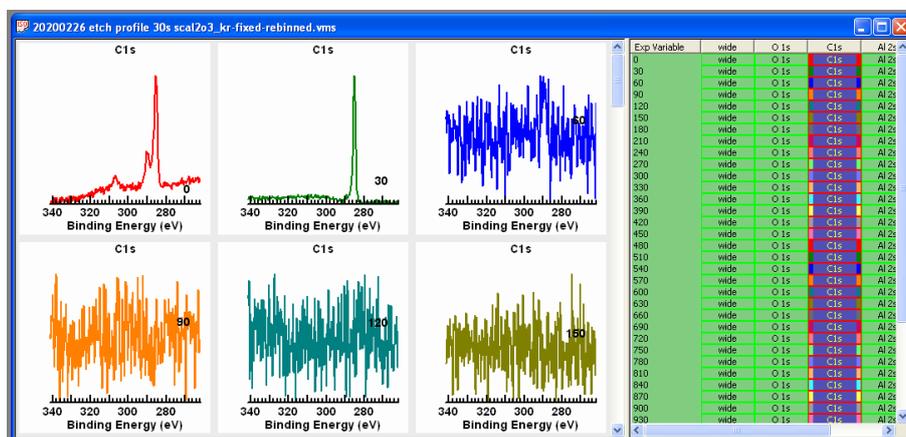
The Spectrum Processing dialog window, PCA property page includes options that perform two steps leading to signal enhanced spectra.

The first step is to compute a number of abstract factors based on a set of VAMAS blocks. Select the column of C 1s VAMAS blocks and display these C 1s spectra overlaid in the active tile. The PCA property page, **Generate Factors** button performs an iSVD calculation using data displayed in the active display tile computing the number of abstract factors specified in the **Factors** text-field. The tick-box labelled **Calc Specified Factors** forces the calculation of the specified number of factors indicated in the **Factors** text-field. If **Calc Specified Factors** is not ticked, iSVD will terminate once abstract factors sequentially determined resemble Poisson distributed noise.



Abstract factors and partially computed abstract factors are saved in the processed form for data in a VAMAS block. Pressing the **Reset Processing** button on the PCA property page returns each VAMAS block to the data before iSVD was applied to data displayed overlaid in the active tile.

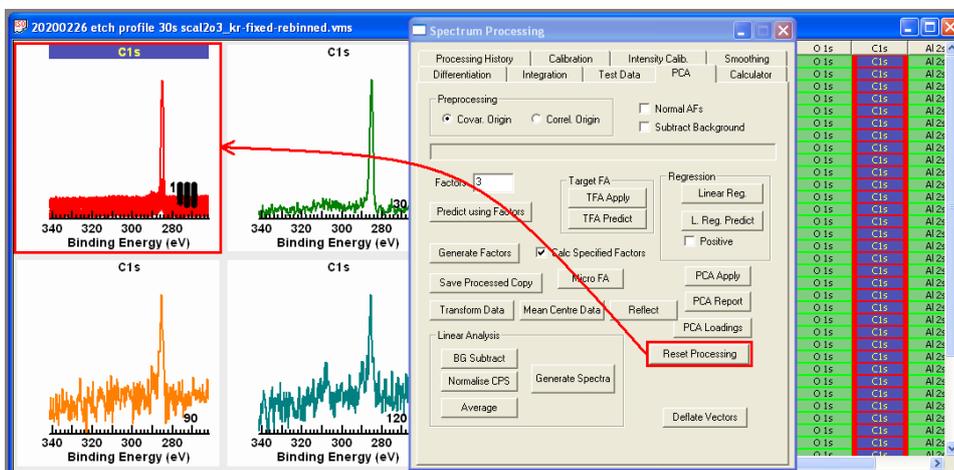
Displaying the selected C 1s VAMAS blocks one per tile allows the inspection of abstract factors. For the case of C 1s spectra processed to form 20 abstract factors, the results show two very significant spectral shapes followed by potentially a third with some spectral shape accompanied by a sizable noise contribution but following the first three most significant abstract factors the remaining factors have the appearance of noise. As a consequence of these observations these raw data may be reconstructed using the three most significant abstract factors.



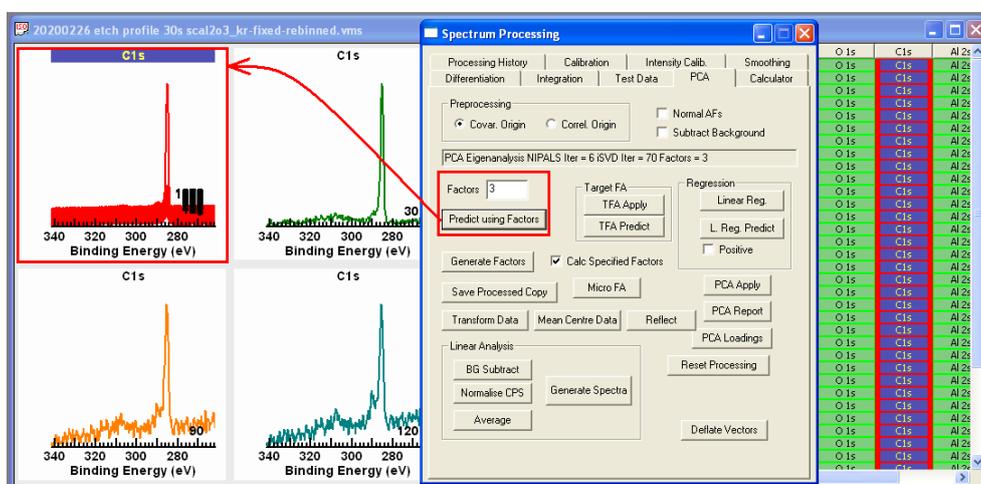
The term reconstruction means these first three most significant abstract factors form a three dimensional subspace of the vector space containing all C 1s spectra used in the iSVD calculation. Reconstruction is the act of projecting each spectrum onto the 3D subspace defined by these three abstract factors. The vectors formed by this projection operation contain linear combinations of only these three abstract factors and since we assume all other abstract factors formed by a full SVD applied to C 1s spectra represent noise the reconstructed also referred to as predicted spectra suppress noise.

The second step in creating noise reduced spectra makes use of the first step involving computation of abstract factors by selecting to use three abstract factors in the prediction step. Data as seen in the display tiles must be returned to the original C 1s spectra before iSVD was applied. The steps are therefore as follows.

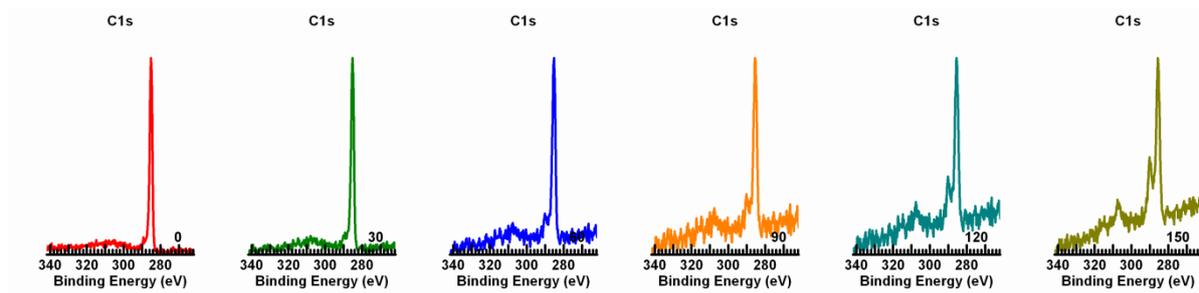
1. Overlay the set of C 1s VAMAS blocks in the active tile before pressing the **Reset Processing** button on the PCA property page.



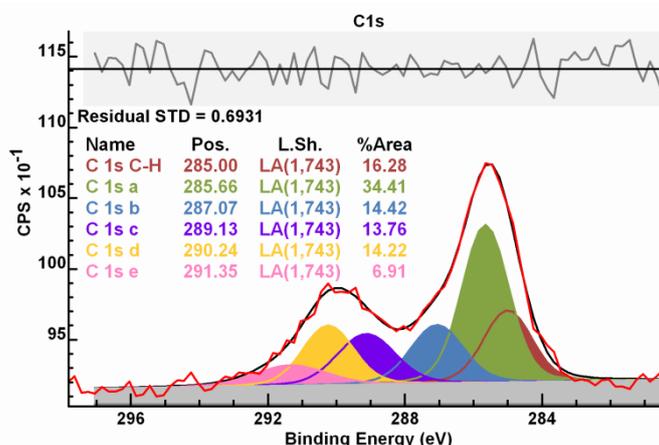
2. Enter the number of abstract factors intended for use in the reconstruction of C 1s spectra using abstract factors in the **Factors** text-field.
3. Press the **Predict using Factors** button on the PCA property page.



Once signal is enhanced relative to noise the character of C 1s spectra can be seen to alter as a consequence of cleaning the surface with an ion beam. A spectrum typical of adventitious carbon is replaced progressively by spectral forms containing information relating to carbon chemistry no longer dominated by C-H bonded C 1s. Further, after sputtering C 1s peaks are superimposed on a declining background, which is an important consideration when constructing a peak model to analyse carbon chemistry for these types of samples.



The challenge is then to establish component peaks within these C 1s spectra and correlate component peaks with carbon chemistry.



Fitting a set of component peaks to a data envelope with good reproduction as indicated by the residual standard deviation and uniformity of residual plot is no guarantee of physically meaningful component binding energies or relative intensities. At the very least, a peak model should be consistent with results from other sources. In the current case study, the existence of C 1s component peaks assigned to bonds with oxygen or aluminium must correlate with signal within O 1s and Al 2p spectra. However it is not always possible to uniquely determine such correlations, for example Al 2p in this case study lacks any features that would allow meaningful separation of chemical state by means of Al 2p. In the absence of such supporting evidence conclusions drawn from fitting C 1s in isolation should be treated as circumstantial evidence rather than proof of hypothesis.

