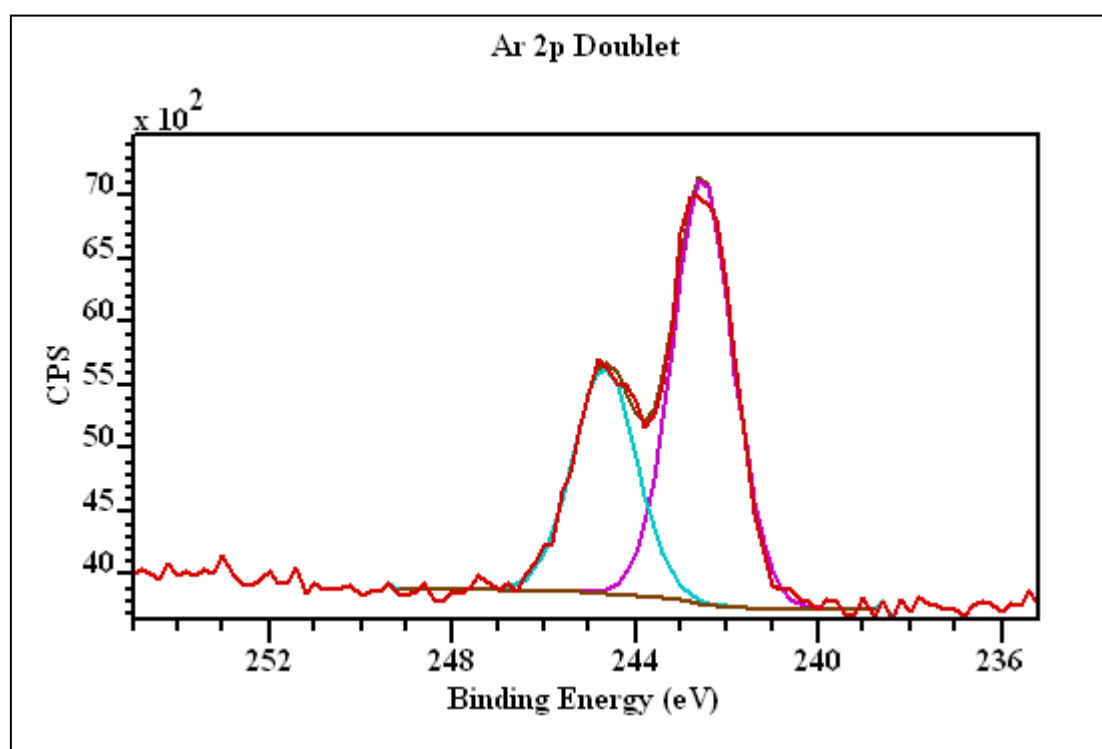


## Principal Axes and Peak Fitting Uncertainties

When peaks are fitted to data the answer returned by the fitting procedure for the peak areas is just one of many possible answers. The source for these alternative outcomes to the fitting algorithm is random variations in the data due to noise which differs with each measurement. The influence noise on the outcome is dependent on the peak model used to approximate the data envelope and the nature of the envelope in terms of underlying peak proximity.

The following data from a sample measured after bombardment with argon ions by XPS pulse counted signal will be used to facilitate a discussion of the issues associated with estimating errors in the parameters adjusted to fit the synthetic peaks to the data. An energy range appropriate for electrons emitted from the Ar 2p core level provides an example of a doublet electron state yielding two overlapping peaks. The discussion will focus on the determination of the peak areas for each of the Ar 2p<sub>1/2</sub> and Ar 2p<sub>3/2</sub> component peaks.



A Monte Carlo procedure is used to simulate repeating the measurement many times and with each simulated data set the peak area is recalculated. The result of the Monte Carlo simulation is a table of variations from the initial peak parameters calculated from the Ar 2p spectrum. In this current example of fitting two synthetic peaks the number of adjustable parameters is six, namely, two sets of area, position and FWHM parameters, one set per synthetic peak. While these six parameters yield six error distributions, the discussion will proceed by focusing only on the two area distributions in isolation. This is only to highlight the nature of the analysis involved in understanding error distributions. The determination of the uncertainties for these area parameters in CasaXPS includes all fitting parameters and not just the two distributions now discussed.

## The Poisson distribution and Pulse Counted Data

For pulse counted data it can be assumed there exists a count rate  $\nu$  such that

- 1) The probability of a single counting event occurring in a small time interval of length  $\delta t$  is approximately equal to  $\nu \delta t$ .
- 2) The probability of more than one counting event occurring in a small time interval  $\delta t$  is negligible when compared to a single counting event occurring in the same time interval.
- 3) The numbers of counting events in non-overlapping time intervals are independent.

Given these assumptions it can be shown that the number of counting events occurring in a period of time  $t$  has a Poisson distribution with parameter  $\lambda = \nu t$ . If the random variable  $X(t)$  denotes the number of counting events in the time interval  $t$  then  $P[X(t) = r] = \frac{e^{-\nu t} (\nu t)^r}{r!}$  for  $r = 0, 1, 2, \dots$

Given that  $X$  is a Poisson distributed random variable; the expected value and variance for  $\lambda = \nu t$  are as follows

$$\mathcal{E}[X] = \lambda$$

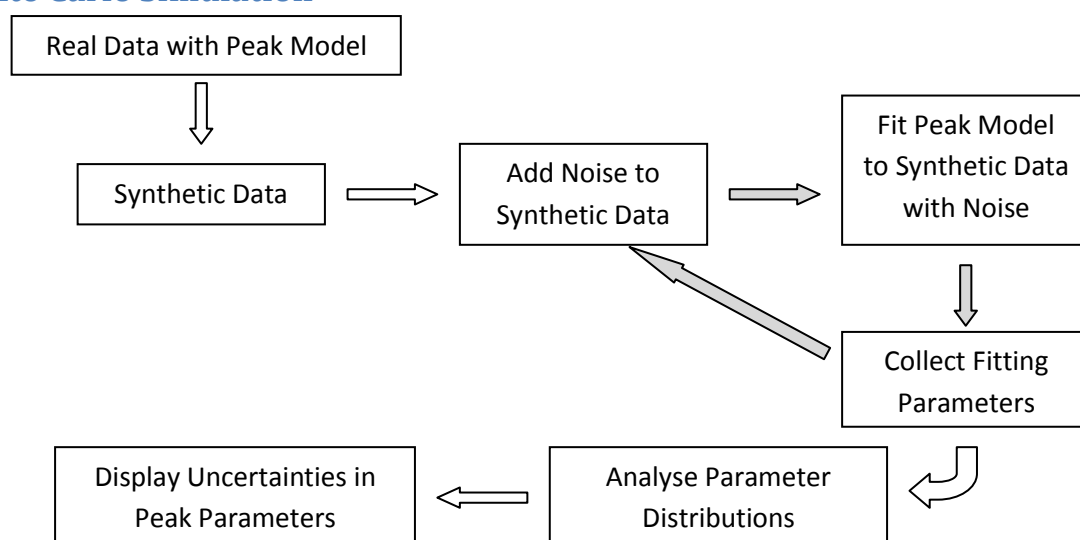
and

$$\text{var}[X] = \lambda$$

Since for pulse counted XPS data  $\lambda$  corresponds to the counts per bin, the standard deviation in the counts per bin is  $\sigma = \sqrt{\text{var}[X]} = \sqrt{\lambda}$ .

Provided an XPS spectrum can be expressed as counts per bin, assuming Poisson behavior for the noise in the data allows error estimates for peak fitting parameters to be calculated using a Monte Carlo approach.

## Monte Carlo Simulation



A Monte Carlo procedure involves a simple sequence of steps aimed at estimating the precision error in output quantities from a calculation. By synthesising the problem before adding noise back to the synthetic problem, followed by repeating the calculation an understanding of how noise

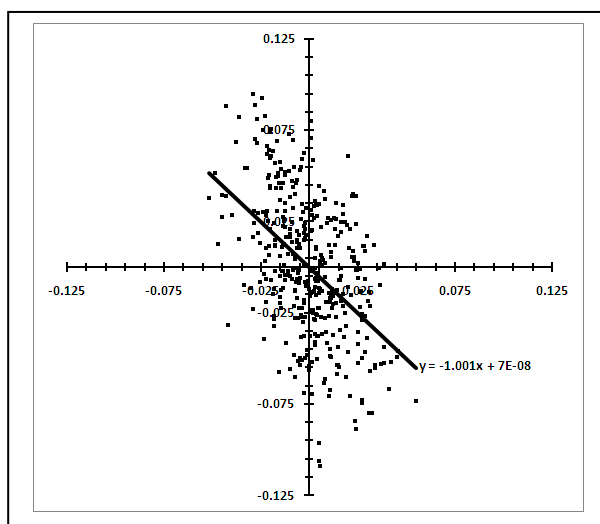
perturbs the current set of peak parameters is established. After iterating through these steps collecting the output parameters at each iteration, distributions for the output parameters are gathered where the variation in the output values are due to the influence of noise on the calculation in question. For the problem of fitting peaks to data, the calculation is that of optimising a set of peak parameters so as to reproduce the data envelope in a least squares sense.

### Methods for Characterising Trends in Scatter Plots

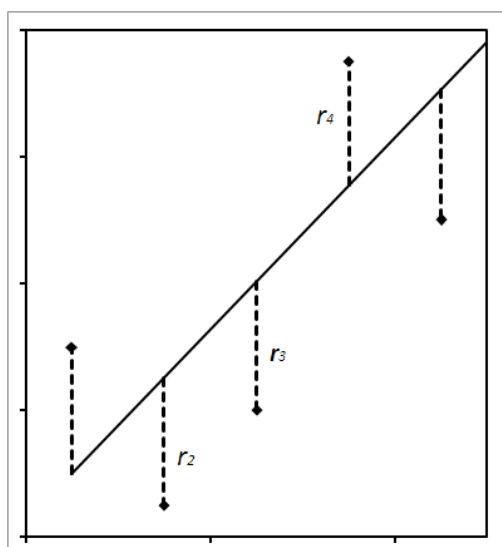
The peak areas, measured relative to the area parameter calculated from the data, are plotted for the two peaks fitted to the Ar 2p spectrum as a set of points on Cartesian axes. Since the two peaks overlap it is logical that if one peak increases in area, to fit the same data envelope, the second peak must reduce in area. It is therefore reasonable to believe the scatter plot for these area parameters is anti-correlated. A regression line calculated for the set of coordinates

$$\left( \frac{area_{simulated\ peak_1} - area_{peak_1}}{area_{peak_1}}, \frac{area_{simulated\ peak_2} - area_{peak_2}}{area_{peak_2}} \right)$$

supports this theory.



A Scatter Diagram in which the y-axis is the area calculated for the Ar 2p<sub>1/2</sub> peak and the x-axis is the area for the Ar 2p<sub>3/2</sub> peak. The peak areas are plotted centred with respect to the initial values for the areas determined from fitting the model to the data and normalised to these initial areas.

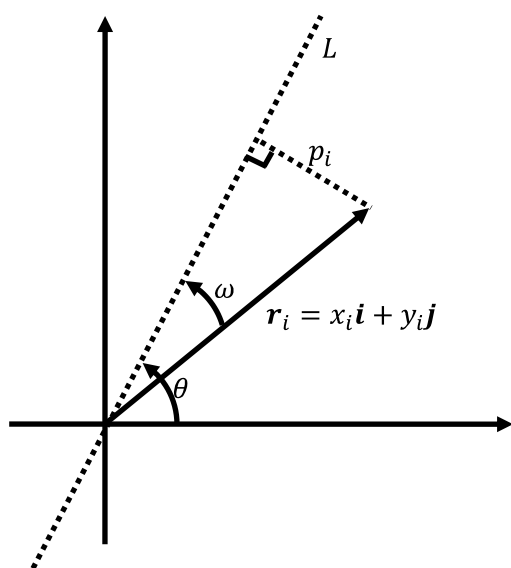
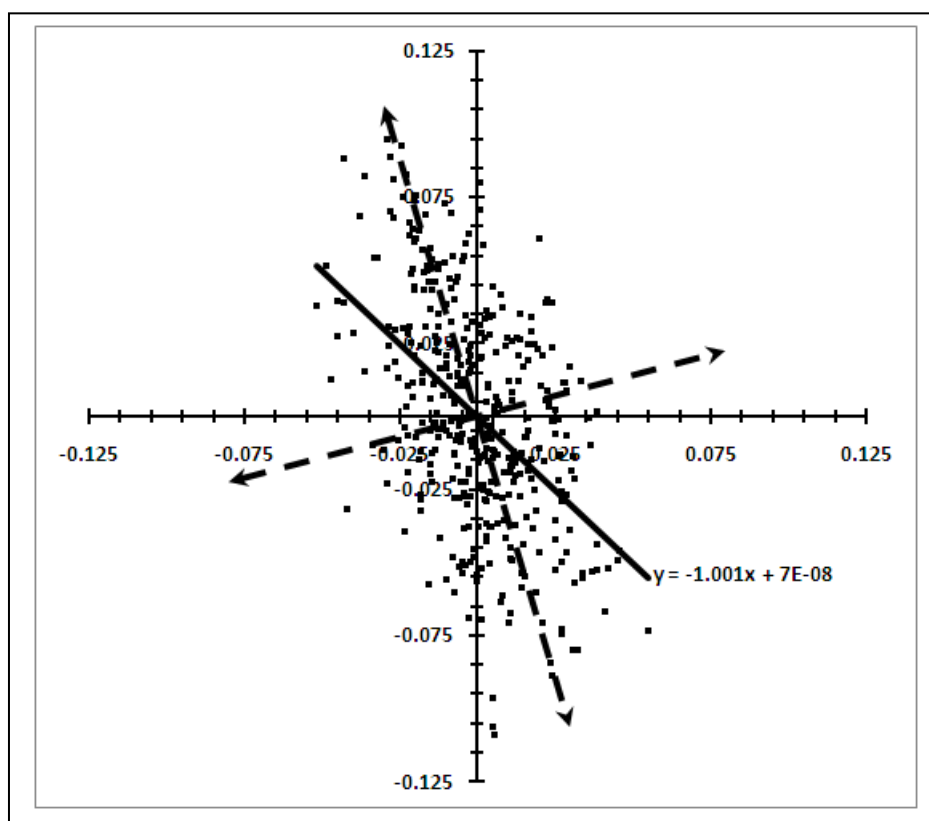


The trend in the above scatter diagram is illustrated using a regression line. The values  $r_i = y_i - (a + bx_i)$  are called the residuals and are depicted graphically as vertical lines between the data points and the line of best fit.

The line of best fit is therefore considered to be the line which minimises the sum of the squares of the residuals.

$$\chi^2 = \sum_{i=1}^n r_i^2$$

While a regression line is often used to determine the line of best fit, the error distribution is visually different in nature from the regression line. The regression line is limited to minimising the residuals with respect to one distribution only. An alternative approach for lines passing through the origin is to consider the shortest distance from each point on the scatter plot to a line of best fit. If the set of points are considered to be the positions for a set of unit mass particles relative to the centre of mass located at the origin, then the problem of calculating the lines of best fit for these two distributions is equivalent to finding the principal axes for the moment of inertia for a collection of particles.



Consider a set of particles of unit mass with position vectors  $\mathbf{r}_i = x_i \mathbf{i} + y_i \mathbf{j}$  and a line  $L$  making an angle  $\theta$  with the positive  $x$  direction. Let the shortest distance between the point with position vector  $\mathbf{r}_i$  and the line  $L$  be  $p_i$ .

A line of best fit through the origin may be obtained by calculating the minimum for

$$I = \sum_{i=1}^n p_i^2$$

Let the unit vector in the direction of the line  $L$  be  $\hat{l} = \alpha\mathbf{i} + \beta\mathbf{j}$  the  $p_i = |\mathbf{r}_i| \sin \omega = |\mathbf{r}_i \times \hat{l}|$

$$\mathbf{r}_i \times \hat{l} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_i & y_i & 0 \\ \alpha & \beta & 0 \end{vmatrix} = \mathbf{k}(x_i\beta - y_i\alpha)$$

Therefore  $p_i^2 = (x_i\beta - y_i\alpha)^2$  and so

$$\begin{aligned} I &= \sum_{i=1}^n p_i^2 = \sum_{i=1}^n (x_i\beta - y_i\alpha)^2 \\ \Rightarrow I &= \sum_{i=1}^n (x_i^2\beta^2 - 2x_iy_i\alpha\beta + y_i^2\alpha^2) \\ \Rightarrow I &= \beta^2 \sum_{i=1}^n x_i^2 + \alpha^2 \sum_{i=1}^n y_i^2 - 2\alpha\beta \sum_{i=1}^n x_iy_i \end{aligned}$$

Let

$$\begin{aligned} A &= \sum_{i=1}^n x_i^2, B = \sum_{i=1}^n y_i^2 \text{ and } C = \sum_{i=1}^n x_iy_i \\ \Rightarrow I &= A\beta^2 + B\alpha^2 - 2C\alpha\beta \end{aligned}$$

Since  $\hat{l} = \alpha\mathbf{i} + \beta\mathbf{j}$  can be expressed in terms of the angle  $\theta$  between the line  $L$  and the  $x$ -axis, the expression for  $I$  can also be expressed in terms of the angle  $\theta$ ; the values  $A$ ,  $B$  and  $C$  are all calculated from the data and are therefore known.

$$\alpha = \cos \theta \text{ and } \beta = \sin \theta$$

Therefore

$$I = A \sin^2 \theta + B \cos^2 \theta - 2C \sin \theta \cos \theta$$

The line of best fit can be obtained by minimising  $I$  with respect to  $\theta$ .

$$\frac{dI}{d\theta} = 2A \sin \theta \cos \theta - 2B \sin \theta \cos \theta - 2C(\cos^2 \theta - \sin^2 \theta)$$

Since  $\cos 2\theta = \cos^2 \theta - \sin^2 \theta$  and  $\sin 2\theta = 2 \sin \theta \cos \theta$

$$\frac{dI}{d\theta} = (A - B) \sin 2\theta - 2C \cos 2\theta$$

The minimum in  $I$  occurs when  $\frac{dI}{d\theta} = 0$  therefore

$$(A - B) \sin 2\theta - 2C \cos 2\theta = 0$$

$$\tan 2\theta = \frac{2C}{(A - B)}$$

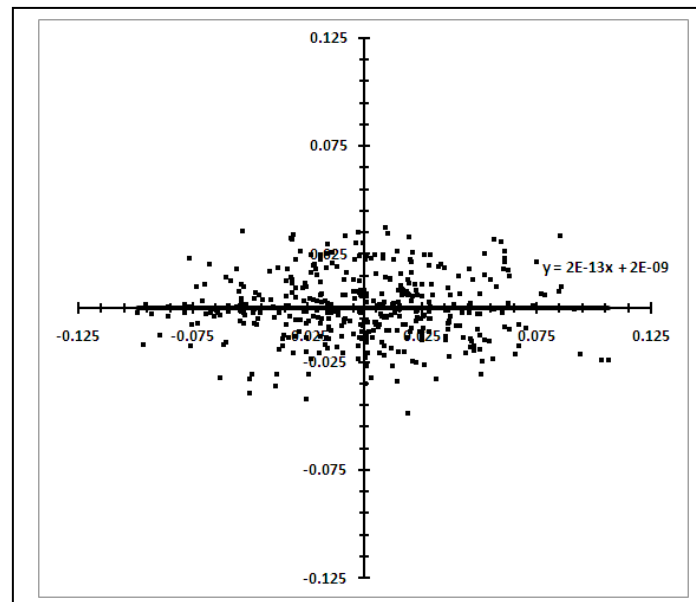
Applying the result to the two distributions for the area parameters from the Ar 2p spectrum the minimum for  $I$  occurs for ( $A = \sum_{i=1}^{400} x_i^2 = 0.132165691$ ,  $B = \sum_{i=1}^{400} y_i^2 = 0.563587216$  and  $C = \sum_{i=1}^{400} x_i y_i = -0.132312195$

$$\tan 2\theta = \frac{2(-0.132312195)}{(0.132165691 - 0.563587216)} \Rightarrow 2\theta = \tan^{-1} 0.613377809$$

$$\Rightarrow \theta = \frac{0.550198098}{2} + k \frac{\pi}{2} \text{ radians}$$

Therefore two extrema occur for lines at  $\theta = 15.76^\circ$  and  $\theta = 15.76^\circ + 90^\circ = 105.76^\circ$  with respect to the  $x$ -axis.

If the set of data points are transformed by rotation by  $-105.76^\circ$  the image for the distributions appears as follows.



An alternative means of finding the principal axes which generalises to multidimensional distributions is to minimise the function  $I = A\beta^2 + B\alpha^2 - 2C\alpha\beta$  subject to the constraints  $\alpha^2 + \beta^2 = 1$ .

Applying the method of Lagrange multipliers to this constrained optimisation problem involves minimising

$$\Psi(\alpha, \beta) = I(\alpha, \beta) - \lambda(\alpha^2 + \beta^2 - 1)$$

The extrema are obtained by the condition  $\frac{\partial \Psi}{\partial \alpha} = 0$  and  $\frac{\partial \Psi}{\partial \beta} = 0$

$$\Rightarrow \frac{\partial \Psi}{\partial \alpha} = 2B\alpha - 2\beta C - 2\lambda\alpha = 0 \text{ and } \frac{\partial \Psi}{\partial \beta} = 2A\beta - 2\alpha C - 2\lambda\beta = 0$$

Resulting in the simultaneous equations

$$\begin{aligned}
 2B\alpha - 2\beta C - 2\lambda\alpha &= 0 \\
 2A\beta - 2\alpha C - 2\lambda\beta &= 0 \\
 \Rightarrow (B - \lambda)\alpha - \beta C &= 0 \\
 -\alpha C + (A - \lambda)\beta &= 0
 \end{aligned}$$

or in matrix notation

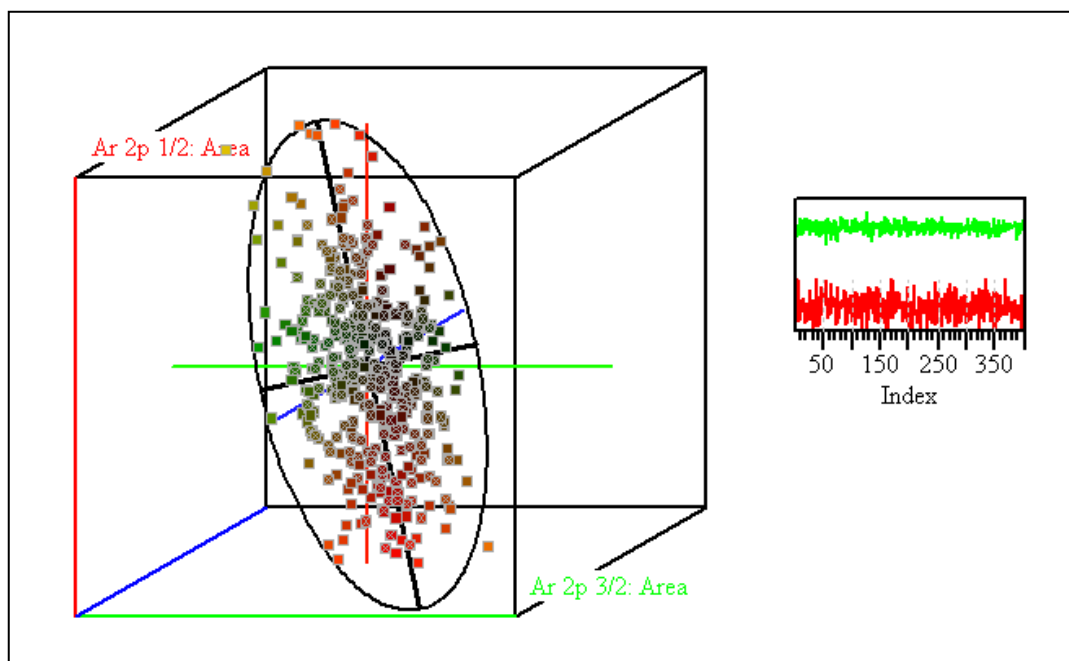
$$\begin{bmatrix} (B - \lambda) & -C \\ -C & (A - \lambda) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \mathbf{0}$$

A non-trivial solution is only possible if these two lines are parallel which means mathematically  $\lambda$  is an eigenvalue of the matrix

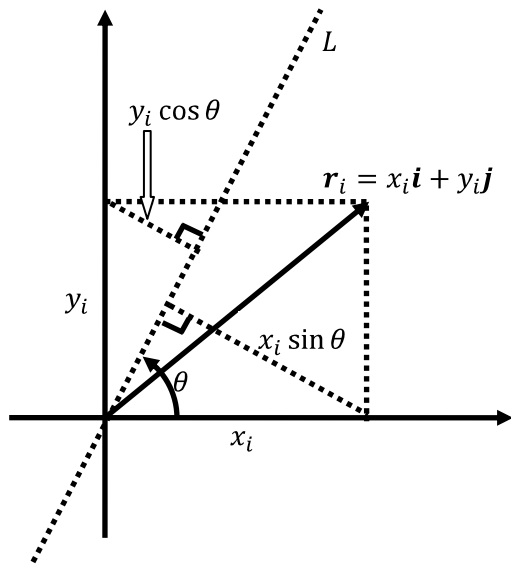
$$\mathbf{Z} = \begin{bmatrix} B & -C \\ -C & A \end{bmatrix}$$

The eigenvalues and eigenvectors for  $\mathbf{Z}$  are the principal axes shown on the scatter diagram.

A geometric interpretation, at least for the 2D problem, is the principal axes are the directions about which the variation in the data points plotted in the plane are a minimum in one direction and a maximum in the orthogonal direction. Principal axes are calculated for multi-dimensional distributions by performing an eigenanalysis. The residual option on the toolbar of CasaXPS turns these principal axes on and off when data are displayed as a scatter plot.



## Principal Component Analysis and Principal Axes



Consider a set of particles of unit mass with position vectors  $\mathbf{r}_i = x_i \mathbf{i} + y_i \mathbf{j}$  and a line  $L$  making an angle  $\theta$  with the positive  $x$  direction.

The direction cosines for the line  $L$  are  $\alpha \mathbf{i} + \beta \mathbf{j} = \cos \theta \mathbf{i} + \sin \theta \mathbf{j}$ .

A line of best fit through the origin may be obtained by calculating the minimum for

$$J = \sum_{i=1}^n (x_i \beta + y_i \alpha)^2$$

subject to the constraint  $\alpha^2 + \beta^2 = 1$

The analysis for  $I$  can be applied for  $J$  resulting in the following optimisation problem.

$$\text{Minimise } J = A\beta^2 + B\alpha^2 + 2C\alpha\beta \text{ subject to the constraint } \alpha^2 + \beta^2 = 1$$

Where again

$$A = \sum_{i=1}^n x_i^2, B = \sum_{i=1}^n y_i^2 \text{ and } C = \sum_{i=1}^n x_i y_i$$

So applying the method of Lagrange Multipliers leads to finding the eigenvectors and eigenvalues of

$$\mathbf{W} = \begin{bmatrix} B & C \\ C & A \end{bmatrix} = \begin{bmatrix} \mathbf{y} \cdot \mathbf{y} & \mathbf{x} \cdot \mathbf{y} \\ \mathbf{x} \cdot \mathbf{y} & \mathbf{x} \cdot \mathbf{x} \end{bmatrix}$$

where  $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$  and  $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ . The matrix  $\mathbf{W}$  is the covariance matrix associated with a Principal

Component determination. The eigenvalues for  $\mathbf{W}$  and  $\mathbf{Z}$  are identical since both are calculated from the same quadratic

$$(B - \lambda)(A - \lambda) - C^2 = 0$$

The eigenvectors for  $\mathbf{W}$  and  $\mathbf{Z}$  are related by a reflection in the  $x$  axis. The principal axes can therefore be obtained using either  $\mathbf{W}$  or  $\mathbf{Z}$ .

## Line of Best Fit using $\mathbf{Z}$

As an alternative to linear regression, in general, a line of best fit in the moment of inertia sense where the distance between a point and the line is measured using the shortest distance from the point to the line is obtained by calculating the gradient for the line of best fit for a line passing



through the mean coordinate for the two data distributions with gradient obtained from the principal axis vector determined from the matrix  $\mathbf{Z}$ .

Applying the notation used for linear regression, the mean centred distributions are summarised as follows:

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

where

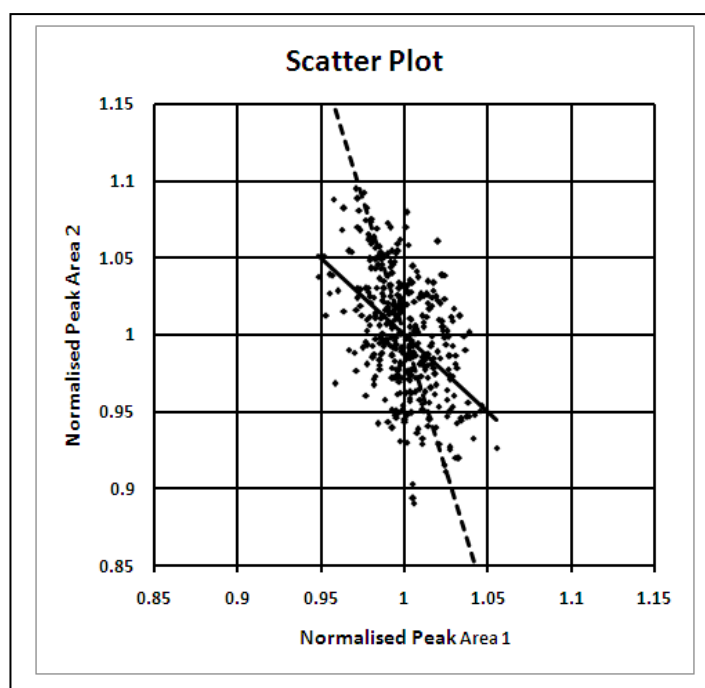
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

The equation of best fit  $y = a + bx$  is given by

$$b = \frac{(S_{yy} - S_{xx}) + \sqrt{(S_{yy} - S_{xx})^2 + 4(S_{xy})^2}}{2S_{xy}}$$

and

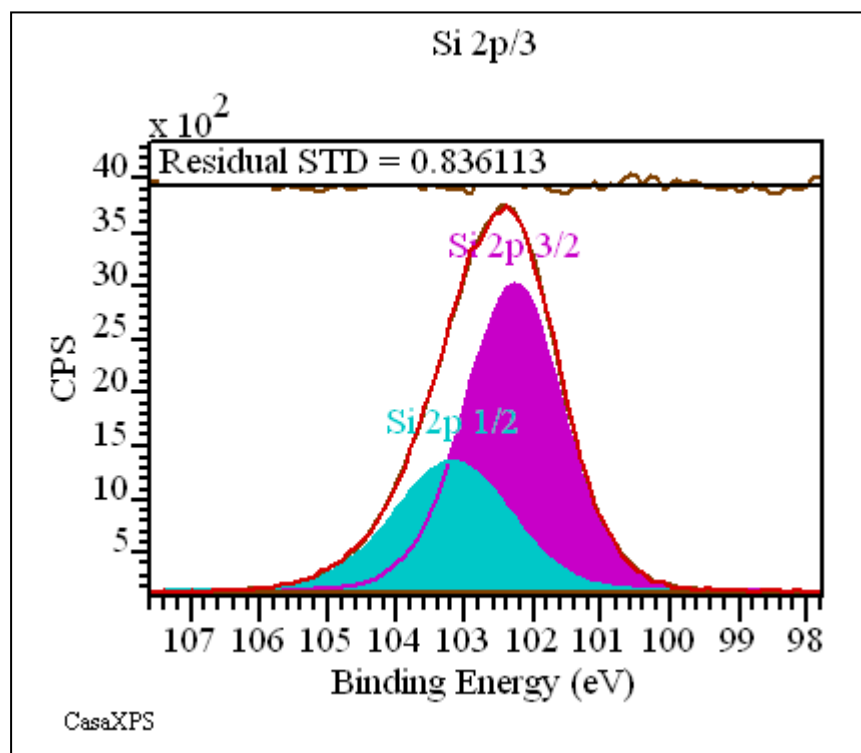
$$a = \bar{y} - b\bar{x}$$



For the scatter plot of the peak area for each of the two Ar 2p peaks relative to the initial values for the peak areas in the Monte Carlo simulation, the two possible lines of best fit have gradient  $-1$  in the case of linear regression, while the line of best fit based on principal axis has gradient  $-3.54$ . Both lines pass through the mean coordinate for the data  $(\bar{x}, \bar{y}) = (1,1)$ .

## Peak Fitting and Error Estimates

Quantification of a sample using XPS is typically presented as a set of atomic concentrations for the elements evident in the data. Evidence of an element in the sample consists of a set of peaks in the spectra and the ability to measure the contribution from each peak to an atomic concentration calculation is dependent on separating peaks arising from different elements. The separation of overlapping peak intensities is achieved by constructing a peak model from known lineshapes and fitting these component peaks to the data envelope. The following data envelope is from a silicon dioxide sample. The measured data envelope is a simple example of two component peaks.



The usual objective for modelling a data envelope is to estimate the relative intensity of elements using peak area. XPS often presents situations where peaks can be identified from the same element with differing position and FWHM as well as from different elements with coincidentally overlapping peaks. When peaks overlap there are two problems involved in the calculation: the first, and instrumental in the atomic concentration calculation, involves determining the peak area for each peak underlying a peak structure; the second problem is to estimate the precision associated with these area measurements. The subject addressed in this article is the latter, namely, assuming a peak model is correctly defined, estimate the uncertainty in the peak area values determined from the model.

## Peak Parameters

Consider the case in which a component peak is defined by a Gaussian lineshape:

$$G(E) = Ae^{-\left[\frac{(E-P)}{F}\right]^2} \quad \dots \quad (1)$$

The functional form in Equation (1) contains three parameters  $P$ ,  $A$  and  $F$  which alter the mean position for the functional form, the area between the abscissa and the function, and the spread of peak area over the energy axis, respectively. Given a data envelope which can be well approximated by a single Gaussian as defined by Equation (1), the problem is to choose values for these parameters  $P$ ,  $A$  and  $F$  which minimise the chi square

$$\chi^2 = \sum_{i=1}^n \left( \frac{d_i - Ae^{-\left[\frac{(E_i-P)}{F}\right]^2}}{\sigma_i} \right)^2$$

where  $\{d_1, d_2, \dots, d_n\}$  are the measured data intensities corresponding to energies  $\{E_1, E_2, \dots, E_n\}$  with individual standard deviations  $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ .

There are many methods for minimising the  $\chi^2$ . Essentially these methods iteratively adjusting the current parameter set  $P$ ,  $A$  and  $F$  until the  $\chi^2$  function appears to be at a minimum, and the method chosen to minimise the  $\chi^2$  is of no importance to the problem provided the method yields a reliable minimum in a timely fashion.

## Monte Carlo Simulation

The area for the Gaussian in Equation (1) can be calculated from the three fitting parameters once a minimum is achieved. The uncertainty for the measured area may be estimated by taking a set of identical samples, repeating the measurement several times and fitting the same model Gaussian to these independent measures. Since the variable element in each determination of the peak area is the noise contribution to the measured signal, the fitting parameters will only vary from the first set determined due to the instrumental noise contribution in the data.

The principle behind Monte Carlo error estimation is analogous to the experimental method just described. The only difference between repeating the experimental measurement and the Monte Carlo approach is noise is introduced into the results from a single measurement using a random number generator rather than allowing the noise inherent in the measurement process for a sequence of measurements to alter the initial conditions to the fitting procedure. The output from both approaches is a list of fitting parameters differing from the original values only due to the influence of noise.

The uncertainties for the adopted fitting parameters and therefore the peak area are calculated from these variations in the parameter sets resulting from noise. One difference between an empirical approach and Monte Carlo is the measurement process will necessarily introduce noise characteristic of the instrument and sample while the Monte Carlo method requires a theoretical specification of the noise.

For pulse counted data measured from a large sample of random events, in this case induced by photoionisation, the noise distribution is modelled using a Poisson distribution where the standard deviation in the recorded intensity in a data channel is the square root of the counts per bin. The validity of assuming a Poisson noise distribution for the counts per bin is perhaps the weak link in estimating the errors using Monte Carlo or any other theoretical methods reliant on knowledge of the noise distribution. Problems with such an assumption exist for instruments for which data are collected using multiple detectors or for detecting systems which are not strictly reporting raw

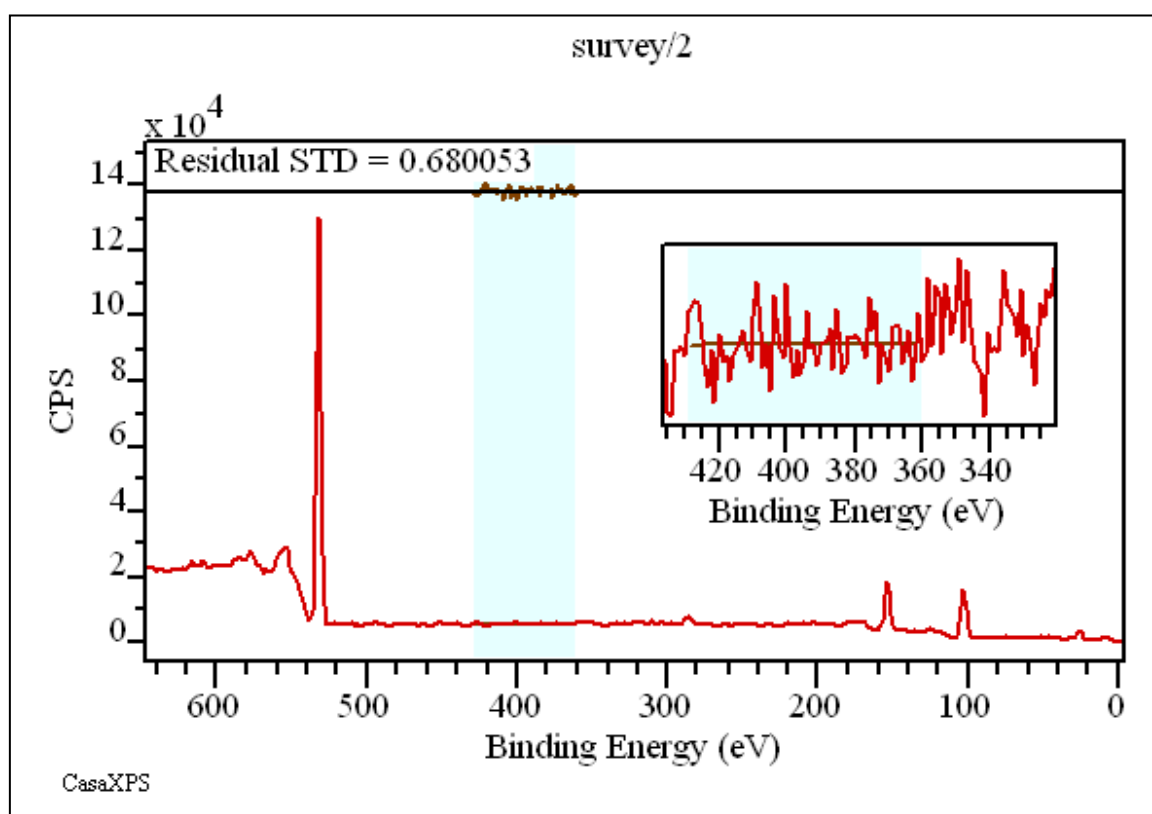
intensities. The consequence of multiple detector systems is the *raw* data appears smoothed by the averaging procedures typically adopted when combining spectral information from multiple data streams.

A simple procedure for testing the validity of assuming Poisson statistics for the spectral data bins is to measure an energy range without any peaks in the data. Using the Regions property page, add a region to the spectrum and select the regression background type for the region. A linear background is added to the data chosen to minimise

$$\chi^2 = \sum_{i=1}^n \left( \frac{d_i - [a(E_i - P_{region}) + b]}{\sigma_i} \right)^2$$

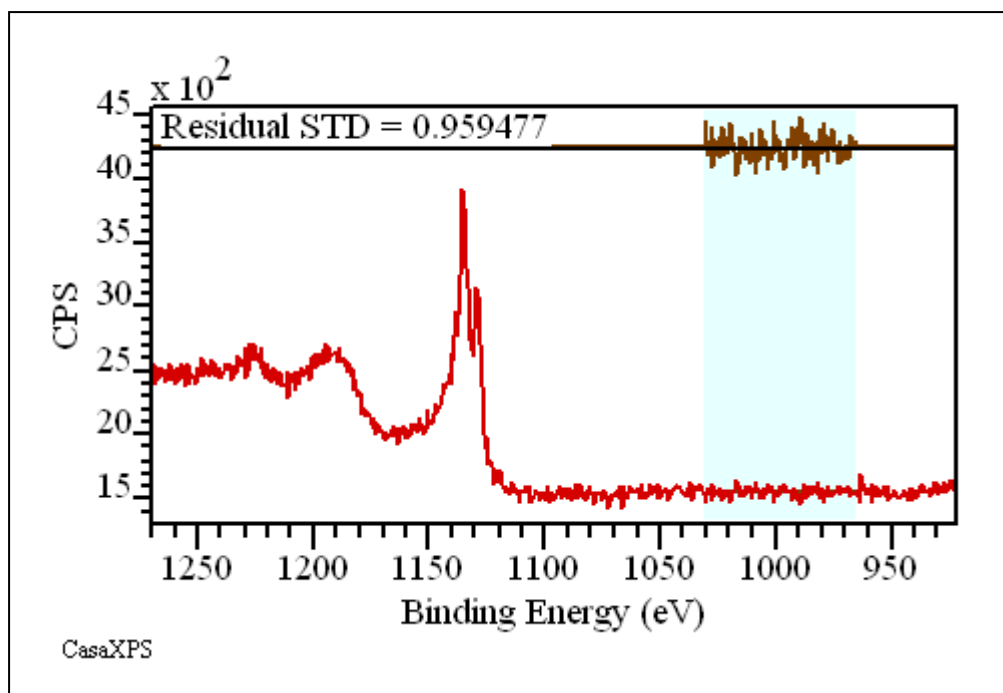
by calculating the linear parameters  $a$  and  $b$  over the interval defined by the region. The standard deviation reported for the residual will be close to unity suggests the noise in the data channels obey Poisson statistics.

The following data are from a multiple detector XPS instrument. The residual standard deviation is too good and is the result of merging multiple data streams to produce the spectrum.



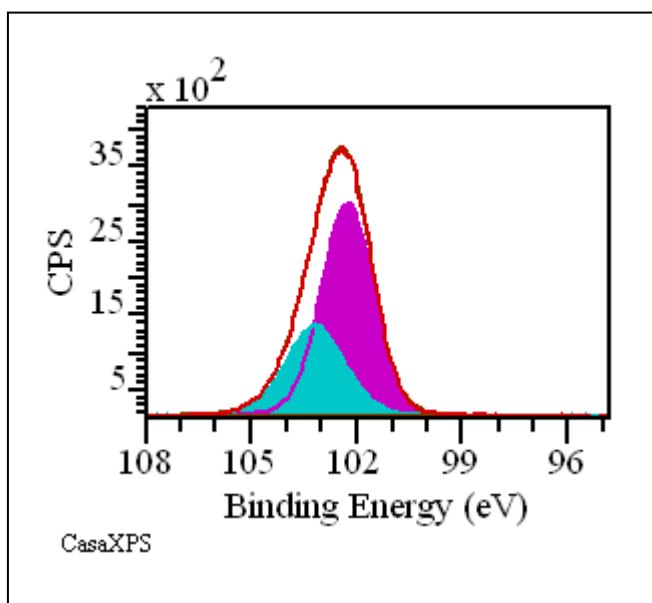
Error estimates assuming Poisson distributed noise will be conservative for cases where the data have a better than expected residual standard deviation.

Data collected using a single channeltron electron detector typically results in the expected Poisson behaviour. The following data are collected with such a detection system which is typical of older instruments. However, it is also possible that an instrument can introduce more than the expected noise, in which case the uncertainties will be under estimated by the Monte Carlo approach.

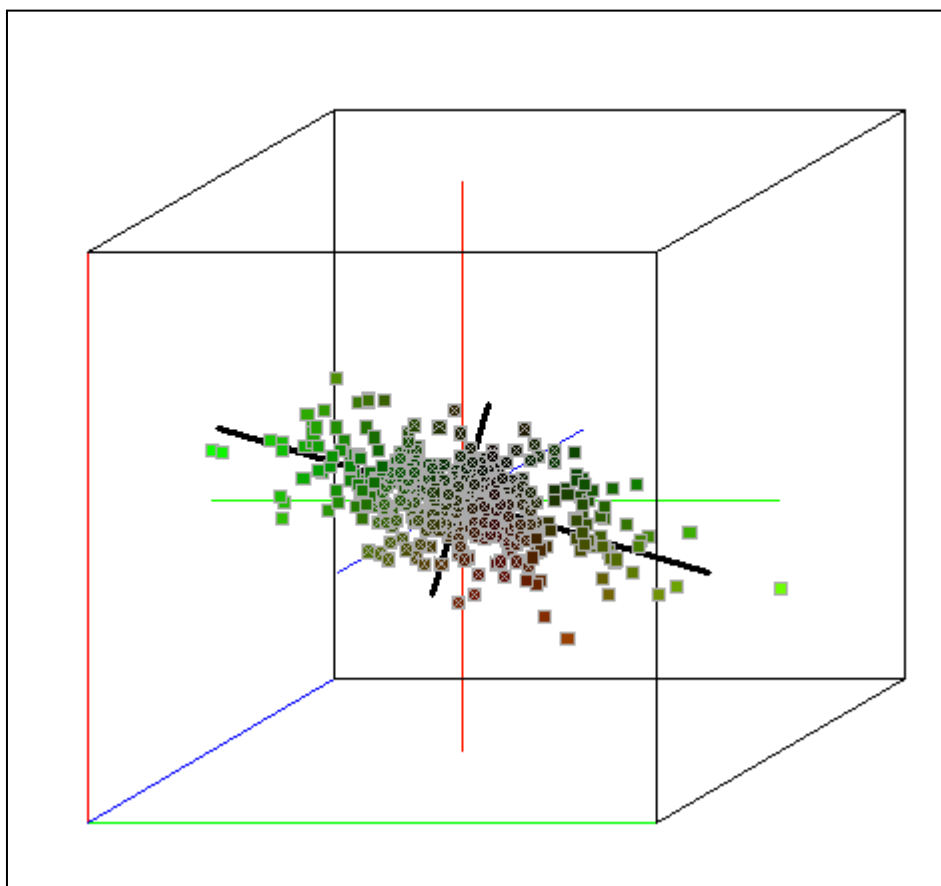


### Estimating the Errors in Peak Parameters

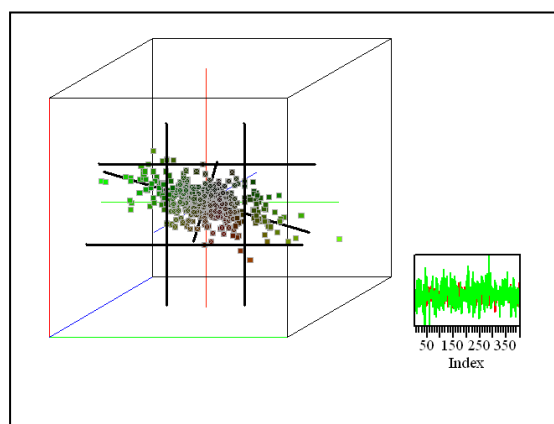
For uncorrelated fitting parameters, one standard deviation in the distribution for the individual fitting parameters offers the uncertainty interval with a 68.3% confidence. Unfortunately the only time peak fitting is performed is precisely when the peaks are correlated and therefore the fitting parameters such as those in Equation (1) are correlated too. Instead of considering the range of variation for each individual parameter it becomes necessary to consider a multi-dimensional distribution from which a region containing 68.3% of the parameter sets must be determined and by projecting the extent of the region within which 68.3% of the possible parameter sets lie from the mean, the uncertainty for each of the fitting parameters can be established.



To illustrate the procedure consider the case where two peaks are defined which model a data envelope and for which the peaks are assumed to have fixed position and FWHM but the area for each of the peaks is allowed to vary. After performing a Monte Carlo simulation for such a peak model, the two free fitting parameters produce two distributions from which the uncertainties for the areas can be estimated.



A scatter plot of these two parameter distributions illustrates anti-correlation between these two fitting parameters and also highlights the set of points belonging to the region which contains 68.3% of the simulations outcomes. The scatter plot represents the outcome for the peak areas determined for each optimisation step in the Monte Carlo procedure by positioning a marker at the coordinate  $(area_{peak1}, area_{peak2})$ . These markers are filled with a colour determined by the size of the coordinates relative to the mean for the individual distributions, where each coordinate axis is assigned to a colour intensity red, green, blue (RGB) as a right-handed coordinate system. Markers within the set of peak areas lying within an elliptical (in the case of a 2D plot) region containing 68.3% of these coordinates are additionally marked with a cross. An estimate for the uncertainty in these two peak area parameters is obtained by projecting the extreme values from this 2D confidence region onto the coordinate axes as illustrated below.



The principal axes determined from the parameter distribution are also shown. The length of these axes are three standard deviations in the distributions obtained from the coordinates of the area parameters after transformation to coordinates with respect to the principal coordinate axes.

While discussed in terms of two fitting parameters, the principles illustrated with 2D scatter diagrams can be extended to  $m$ -dimensional problems involving  $m$  fitting parameters. The visualisation of these  $m$  dimensional distributions is limited to a projection of these distributions onto at most 3D scatter plots, but the mathematics for determining the confidence region in  $m$ -dimensional space remains the same and can be used to obtain error estimates for fitting problems requiring multiple component peaks with parameters equivalent to those illustrated by Equation (1).

## Calculating Parameter Uncertainties in CasaXPS 2.3.16

The procedure for generating estimates for the uncertainty in peak areas and ultimately atomic concentrations derived from peak models requires a Monte Carlo simulation to be performed on each piece of data in the atomic concentration calculation.

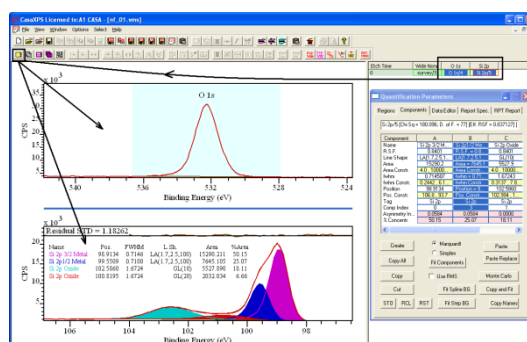
There are two ways of performing a Monte Carlo simulation on a spectrum. Either an individual spectrum is identified by displaying the VAMAS block in the active tile of the scrolled list in the left-hand pane, before pressing the Monte Carlo button on the Components property page; or a set of VAMAS blocks are prepared with peak fits, selected in the right-hand pane before using the Browser Operations dialog window to propagate the calculation of error bars to the selected blocks. Since performing a Monte Carlo operation is a lengthy process, calculating error bars using the propagation option on the Browser Operations dialog window may take some time, but not so long as if each spectrum is individually processed using the Components property page.

It is important to note that the results for a Monte Carlo simulation used to calculate the atomic concentration uncertainties are transient. If any parameter associated with the peak modelling processed is altered, the Monte Carlo results are deleted and the process must be repeated a further time. It is therefore important to proceed directly from a Monte Carlo error analysis to generating a report using the Report Spec property page of CasaXPS.

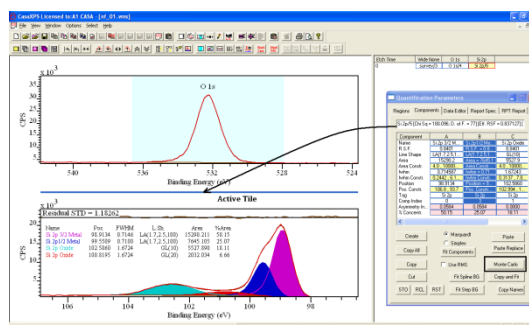
## Calculating Error Estimates for a Single VAMAS Block

The following steps cause the error estimates for the peak area to be entered into a quantification report based on two VAMAS blocks containing an O 1s spectrum and a Si 2p region. The Si2p data uses a peak model to differentiate the elemental silicon peaks from silicon oxide peaks.

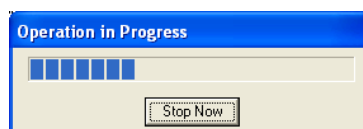
1. Select the VAMAS blocks in the right-hand pane and display the data in the left-hand pane, one spectrum per tile.



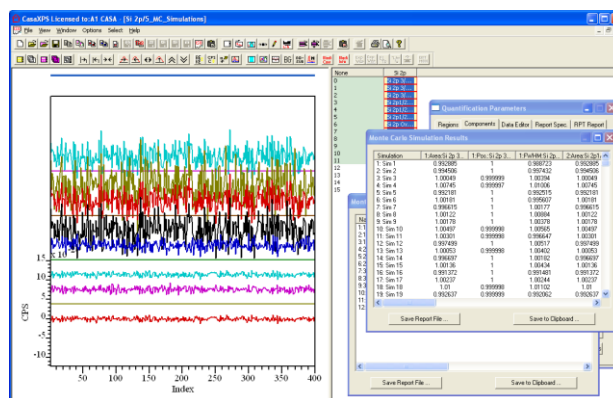
- Left-click in the tile displaying the peak model and data for silicon to make the Si 2p VAMAS block the active data, ensure the peak model is fitting to the data before pressing the Monte Carlo button on the Components property page.



A progress metre appears showing the progress for the individual simulation steps used to perform the Monte Carlo procedure on the data in the active tile.



On completion, the progress dialog is replaced by two dialog windows offering the results from the Monte Carlo simulation steps and a new experiment frame is created containing a set of VAMAS blocks in a state suitable for interrogating the information for each of the parameter distributions as scatter plots.

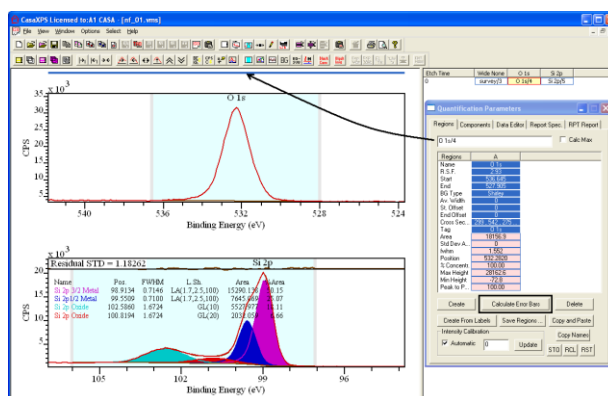


The error estimates for the peak areas are at this point included in the component quantification information for each of the components listed on the Components property page. These error estimates will remain valid provided the component peaks and the quantification regions parameters are left unaltered for the given silicon data.

Before a quantification report can be generated the error estimate for the O 1s region must be computed.

- Select the O 1s spectrum using the left-hand pane. The tile containing the O 1s must be the active tile to ensure the information displayed in the Regions property page is for the O 1s region.



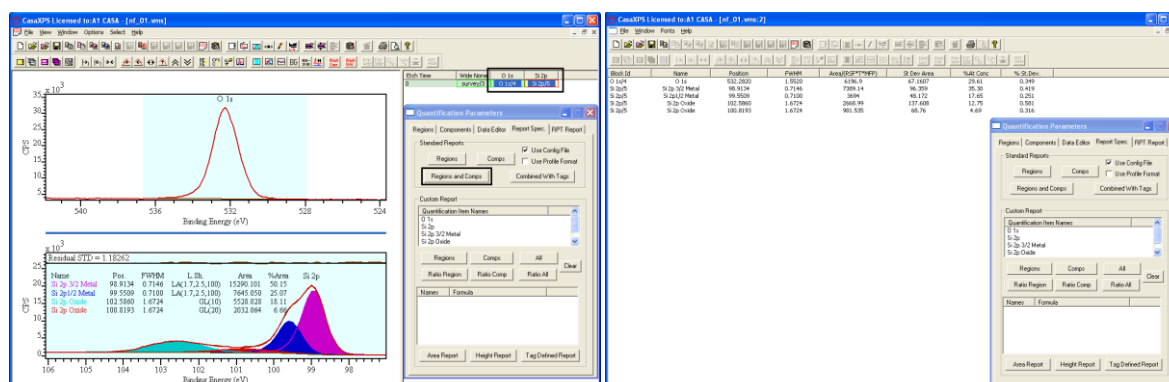


4. Press the Calculate Error Bars button on the Regions property page.

Again a Progress metre indicates the progress of a Monte Carlo calculation based on the region information. Once completed the error information is added to the region as the standard deviation for the region area.

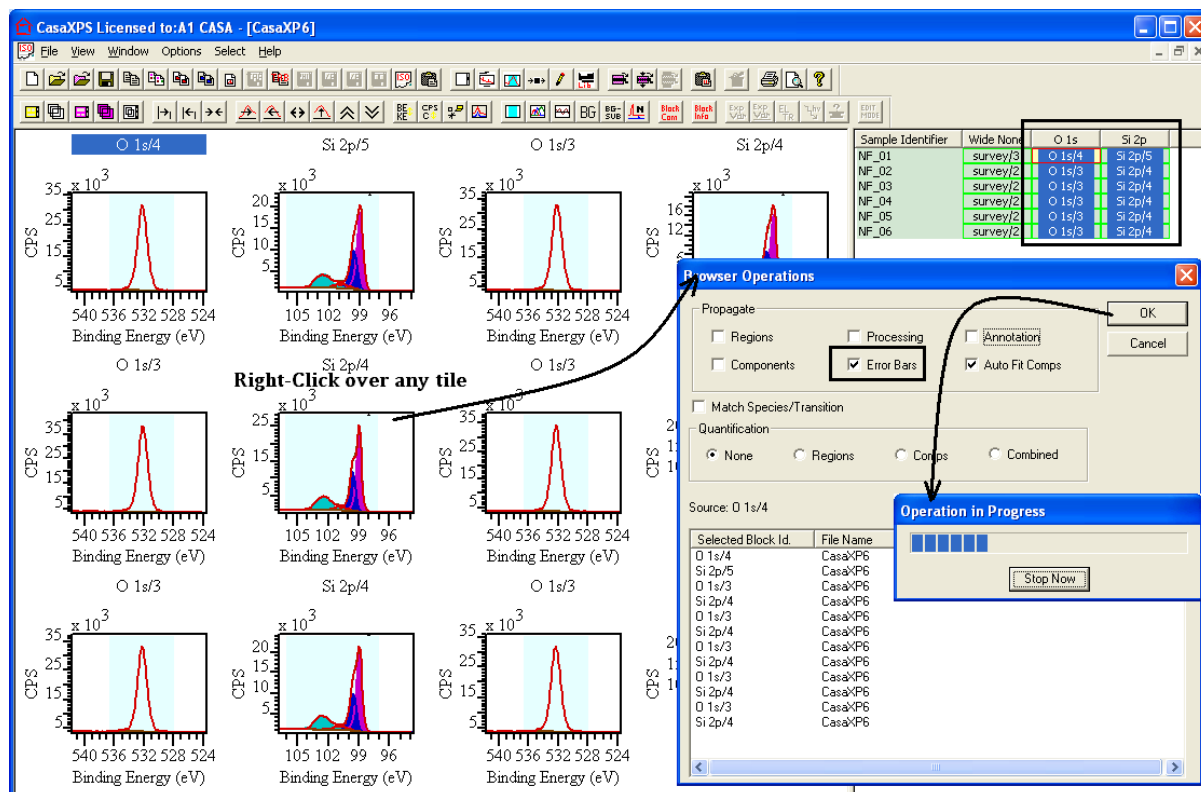
To view the results of the Monte Carlo error analysis for both the O 1s region and the Si 2p peak model, a report specification configuration file must be prepared as shown in the Notepad editor below.

5. Select both the O 1s and the Si 2p VAMAS blocks in the right hand pane.
6. On the Report Spec property page, press the button labelled Regions and Comps.



## Calculating Error Estimates for a Multiple VAMAS Blocks

As an alternative, the uncertainties for fitting parameters associated with components in a set of VAMAS blocks can be populated using the Browser Operations dialog window.



The Browser Operations dialog window includes an option for propagating Error Bars. To perform the propagation of error bars it is sufficient to only tick the Error Bars radio button in the Propagate section of the dialog window, however it is assumed each spectrum has appropriate regions and components already defined and fitted. Right-clicking the mouse over any tile in the left-hand pane will invoke the Browser Operations dialog window and since the propagation of error bars will act on whatever quantification items already defined the choice of tile to initiate the Browser Actions is irrelevant to the error bars operation.

Monte Carlo simulation used to calculate the error bars for components is time-consuming; therefore it is usually best to prepare the data without using the error bar propagation option at the same time as propagating the regions and components. The Region and Comps button is applied to the selection in the right hand pane to produce the quantification table below.

