Microscopy and Microanalysis (2019), **25**, 378–388
doi:10.1017/S1431927618016276

Microscopy AND Microanalysis

Original Article

Quantifying Uncertainty from Mass-Peak Overlaps in Atom Probe Microscopy

Andrew J. London

United Kingdom Atomic Energy Authority, Culham Science Centre, Abingdon, Oxon, OX14 3DB, UK

Abstract

There are many sources of random and systematic error in composition quantification by atom probe microscopy, often, however, only statistical error is reported. Significantly larger errors can occur from the misidentification of ions and overlaps or interferences of peaks in the mass spectrum. These overlaps can be solved using maximum likelihood estimation (MLE), improving the accuracy of the result, but with an unknown effect on the precision. An analytical expression for the uncertainty of the MLE solution is presented and it is demonstrated to be much more accurate than the existing methods. In one example, the commonly used error estimate was five times too small.

Literature results containing overlaps most likely underestimate composition uncertainty because of the complexity of correctly dealing with stochastic effects and error propagation. The uncertainty depends on the amount of overlapped intensity, for example being ten times worse for the CO/Fe overlap than the Cr/Fe overlap. Using the methods described here, accurate estimation of error, and the minimization of this could be achieved, providing a key milestone in quantitative atom probe. Accurate estimation of the composition uncertainty in the presence of overlaps is crucial for planning experiments and scientific interpretation of the measurements.

Key words: APT, atom probe tomography, confidence interval, decomposition, error bars, mass spectrum analysis, maximum likelihood, overlap solving, peak deconvolution

(Received 10 July 2018; revised 7 December 2018; accepted 24 December 2018)

Introduction

Atom probe microscopy (APM) is a technique capable of resolving nano-scale composition measurements with a broad application in many areas (Gault et al., 2012). The composition depends on the elemental and molecular identity of ions inferred from time-of-flight (TOF) mass spectroscopy measurements (Müller et al., 1968; Kelly & Larson, 2012). The TOF measurements are converted to mass-to-charge ratios (m/z) and peaks are assigned to ionic species based on the presence and amplitude of resultant mass-peaks. The composition measurement hinges on this assignment, therefore accurate peak identification is key. There are various sources of errors and uncertainty inherent to APM, which have not been consistently quantified.

Sources of error may be random, mainly affecting the precision/scatter, or they are systematic/biased, reducing the accuracy of the measured composition. The true sample composition is clouded by the uncertainties of both the evaporation physics and the limits of detection. After optimizing analysis conditions, the principal composition errors may occur in the following order of significance: identification of peaks as the leading cause of composition measurement uncertainty (Haley et al., 2015), followed species-specific loss mechanisms [preferential evaporation (Hyde et al., 2011), detector pile-up (Meisenkothen et al., 2015), ion dissociation (Gault et al., 2016)], global background subtraction, peak-tail interference, peak area/ranging (Hudson et al., 2011), and counting error (Danoix et al., 2007*a*). However, the impact these factors have on the measurement uncertainty is dependent on the material system. Additionally, spatial uncertainty will also strongly affect the composition measurement of multi-phase materials (Marquis, 2008).

A general discussion of the errors and factors involved in APM has been presented (Gault et al., 2012; Larson et al., 2013). However, analysts have lacked suitable methods and tools for quantifying uncertainty. Therefore, an empirical approach is often taken to quantify the random error which involves taking the standard deviation of multiple atom probe experiments. Resampling within the observed data is a common statistical approach (Efron, 1979) and has been applied to lateral-composition profiles (Thuvander, 2016) and cluster measurements (Philippe et al., 2010), but is not widely reported in APM. Because an accurate theoretical understanding is lacking, the most commonly reported error is the counting error.

Statistical counting error comes from working with limited counts from a stochastic process, in the case of atom probe, the detection of individual ions in specific m/z bins or ranges (Miller et al., 1996). Danoix et al. (2007*a*) constructed a statistical model of this detection by assuming the atom probe measurement is representative of the probed material and that it is a random solid solution. It was shown that the number of detected atoms of a particular type is given by a binomial distribution (Danoix

Author for correspondence: Andrew J. London, E-mail: andy.london@ukaea.uk

Cite this article: London AJ (2019) Quantifying Uncertainty from Mass-Peak Overlaps in Atom Probe Microscopy. *Microsc Microanal* **25**, 378–388. doi:10.1017/S1431927618016276

et al., 2007*b*), if and, only if, detected events are statistically independent. Danoix et al. (2007*a*) showed the counting error (one standard deviation) for a specific species σ_i with composition C_i, and the total number of counts N is:

$$\sigma_i = \sqrt{\frac{C_i(1-C_i)}{N}} \quad (1)$$

This is the fundamental limit and cannot be improved without taking a larger N. Counting error is the minimum level of uncertainty on any measurement. In most atom probe experiments N is of the order 10^5–10^7 rendering the counting error insignificant, but looking at a sub-section of the mass spectrum, a minority species, or sub-volumes, the count of ions involved is reduced and the counting error can become significant. The counting error and resampling techniques only address the random error present and not systematic bias or loss of accuracy.

Haley et al. (2015) investigated the inter-operator systematic bias by comparing composition results produced by multiple operators. They found the identification of ions in m/z peaks could cause a large systematic error in measured composition. This identification is made more difficult when an overlap or interference of two or more species occurs within the same m/z range.

The typical mass resolving power (MRP) of a modern local electrode atom probe (800–1200) is more than capable of identifying individual isotopes of the same element (Kelly & Larson, 2000). However, peaks separated by <0.1 Da are typically considered "directly" overlapped, although with suitably fast electronics or a long enough flight-path the MRP can be improved enough to separate $^{56}Fe^{2+}$ and $^{14}N_2^+$ (0.04 Da peak separation), for example (Liu et al., 1991).

Where the identity is uncertain, multiple analyses can be made with different ions identified in the same peak (Li et al., 2011). However, the accuracy can often be improved by using information about the isotopic ratios to solve the peak overlap. This analysis is usually called decomposition or deconvolution (Larson et al., 2013); these analyses not to be confused with methods used to sharpen the peak shape or otherwise extract peak counts (Ferrige et al., 1991; Johnson et al., 2013).

Solving overlaps improves the accuracy but the influence on the precision or scatter of the result should also be considered. Miller et al. (1996) introduced maximum likelihood estimation (MLE) to solve direct peak overlaps and London et al. (2017) extended it to solve any general overlap. Both showed the advantage of MLE over using only a single non-overlapped peak (SP) or least squares (LS) approaches, but neither addressed the impacts on uncertainty in measured composition. This paper aims to address this and discuss factors affecting the systematic bias.

Materials and Methods

An example data treatment was performed on an oxide dispersion strengthened (ODS) steel [Fe–0.3Ti–0.3Y_2O_3 wt% (London et al., 2015*b*)], which contains Y–Ti-oxide clusters. The following ionic species were considered for solving the overlap problem: 24 Da Ti^{2+}/C_2^+, 32 Da TiO^{2+}/O_2^+, and 64 Da TiO^+/Fe_2O^{2+}. Data were acquired with a LEAP 3000X HR (Cameca Instruments, Madison, WI, USA) in laser pulsing mode (0.4 nJ, 200 kHz rate) with the sample held at 50 K. The following sections define the details of the overlap solving method.

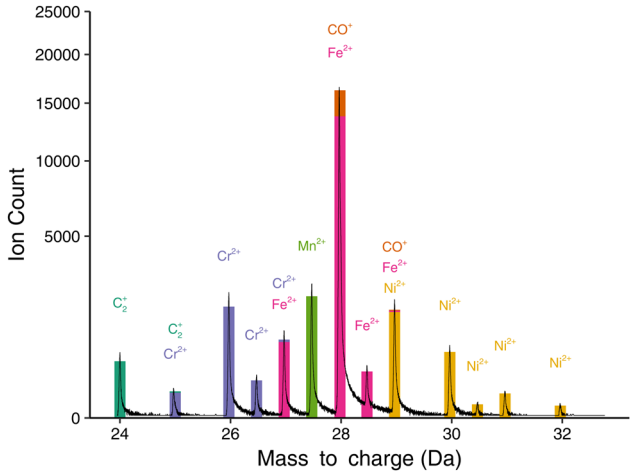

Fig. 1. Labeled mass spectrum. Colored bars show the windows or "ranges" of each peak, and the stacked heights represent the contribution of different ions to the total peak count.

Defining Overlaps

All the ionic species which may be overlapped must be specified. For each species an isotopic fingerprint is calculated from tabulated data (Berglund & Wieser, 2011). The peak width is defined by a parameter related to the mass resolution of the data (London et al., 2017), and peaks are assigned to specific m/z windows or ranges. When an m/z range contains two or more theoretical mass-peak positions, then those ions are considered overlapped. Because theoretical mass positions are used, the m/z spectrum must be properly calibrated, including voltage, bowl and m/z registration corrections. Figure 1 shows the section of an example m/z spectrum, with various ions contributing to different m/z peaks.

Table 1 shows the various ionic species from Figure 1 at specific m/z positions and their natural abundance forming an abundance matrix [**A**]. Pairs of overlapped ions can be constructed by finding rows in [**A**] with multiple entries. Each overlap pair can also overlap with additional species. For example, C_2^+ at 25 Da is connected to Cr^{2+}, but Cr^{2+} is also connected to Fe^{2+} at 27 Da, hence C_2^+ is connected to Fe^{2+} through Cr^{2+}. Therefore a graph of overlapped ions can be constructed. All the ions connected within one graph are called an overlap group. The overlap graph for the ions from Figure 1 is shown in Figure 2. Note the Mn^{2+} peak has some contribution from the preceding peak's tail. It is assumed this kind of overlap can be resolved by suitable choice of background removal and range width; therefore it does not appear in the overlap graph. There may be more than one group of overlaps in the same mass spectrum, but these can be considered as independent problems to solve. This equates to separating the abundance matrix into smaller parts, which simplifies the numerical solving of the overlaps.

If the abundance matrix of a specific overlap problem contains more ions than peaks then there is no unique solution; the matrix is rank deficient as noted by London et al. (2017).

Separating Overlaps

Solving the overlaps can be considered as an optimization problem where the composition of overlapped ions **p** is given by:

$$N\mathbf{A}_{ij}\mathbf{p}_j = \mathbf{r}_i \quad (2)$$

https://doi.org/10.1017/S1431927618016276

Table 1. Abundance Matrix with Ranges as Rows and Ions as Columns.

m/z	C2	Cr	Fe	Mn	CO	Ni
24.0	0.978					
25.0	0.022	0.043				
26.0		0.838				
26.5		0.095				
27.0		0.024	0.058			
27.5				1.000		
28.0			0.917		0.989	
28.5			0.022			
29.0			0.003		0.011	0.683
30.0						0.261
30.5						0.011
31.0						0.036
32.0						0.009

The entries are the natural abundances for each ion in each range. Rows of multiple entries are overlapped peaks.

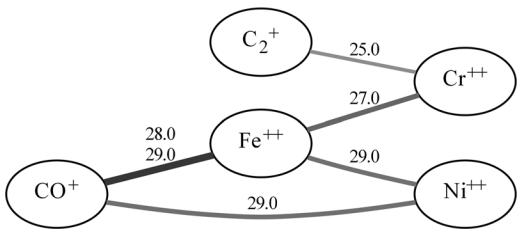

Fig. 2. Overlap graph: overlapping ionic species (nodes) are connected by lines, peak positions are shown in the text. The thickness and darkness of the line indicates the uncertainty of solving that overlap assuming an equal quantity of all ions present. Thicker and darker means more uncertainty as detailed in the Section "Overlap Uncertainty".

where N (scalar) is the total number of counts of ions, $\mathbf{A}$ is the abundance matrix defined above, and $\mathbf{r}$ are the background subtracted counts in each peak. The background subtraction should remove counts occurring from the thermal tails of preceding peaks and the global background. Equation (2) can be optimized in a number of ways: minimizing the sum of squared residuals, LS, or maximizing the likelihood, MLE. Likelihood expresses the probability of observing a particular set of data as a function of a parameterized statistical model (Pawitan, 2001). MLE describes the likelihood of a specific overlap composition (parameters) given the observed integrated and background subtracted peak area counts (data). Miller et al. (1996) modeled the probability of an individual ion being detected in one particular mass range as a multinomial distribution. Following the work of London et al. (2017), the following log-likelihood function based on the multinomial distribution is proposed (derived in Appendix A):

$$\mathcal{L}(\mathbf{r}, \mathbf{p}) = \mathbf{r}^{\mathrm{T}} \log([\mathbf{A}]\mathbf{p}) \tag{3}$$

Note, this corrects an oversight made by London et al. (2017) where the MLE was formulated in terms of the number of counts of each ion. The estimated composition $\hat{\mathbf{p}}$ is found by minimizing the negative log-likelihood function numerically by varying $\mathbf{p}$. The total number of fitted counts N is not optimized by the likelihood estimate. While N does not affect the overlap composition, it does alter the overall composition when the overlap group is summed with other overlap groups and non-overlapped peaks. Therefore, a LS estimate of the fitted counts $\tilde{N}$ is made using equation (2) with $\mathbf{p} = \hat{\mathbf{p}}$.

Overlap Uncertainty

With an aim of defining the uncertainty arising from peak overlaps two examples are considered, Fe^{2+} overlapped with either Cr^{2+} or CO^{+}. Figures 3a and 3c show the natural abundance and peak positions of Cr^{2+} with Fe^{2+} and CO^{+} with Fe^{2+}, respectively. Intuitively, comparing these plots, there is a different degree of difficulty in estimating the composition. For Cr^{2+}/Fe^{2+} the composition could be estimated by simply ignoring the contribution of the 27 Da peak with little loss in accuracy and precision, whereas with CO^{+} and Fe^{2+} the overlap cannot be separated so easily.

Using equation (3), the likelihood of a particular set of peak counts can be calculated as a function of composition. Simulated peak counts of 90% Fe and 1,000 total counts were made. The resulting log-likelihood is plotted in Figures 3b and 3d. Both plots have a maximum at the expected composition of 10%, but with Cr^{2+} the likelihood is sharp and with CO^{+} the likelihood is much broader. This is a valuable property of the likelihood function, the maximum is useful for obtaining point estimates, but whole shape shows the contribution of the data to the parameter of interest. In particular, the curvature at the maximum can, with certain conditions, be inversely related to the standard deviation of the estimated parameter (see Appendix B). The curvature of equation (3) is given by:

$$\frac{\partial^2 \mathcal{L}}{\partial \mathbf{p}_m \partial \mathbf{p}_k} = \sum_i \frac{(\mathbf{r}_i \mathbf{A}_{ik} \mathbf{A}_{im})}{\left(\sum_j \mathbf{A}_{ij} \mathbf{p}_j\right)^2} \tag{4}$$

As evident from the numerator of equation (4), the curvature is related to the product of the abundance of different species and the counts in that mass range. Therefore, overlaps with a small product of abundance in a particular range will have a more curved likelihood function, be more sensitive to small changes in composition and have high precision. Conversely, species with a large shared abundance in the same peaks will give a slowly varying likelihood function with more uncertainty about the maximum, and therefore lower precision. For example, in Figure 3c CO^{+} and Fe^{2+} share significant intensity in the 28 Da peak, and only the much smaller peaks of Fe^{2+} 27 and 28.5 Da are available to constrain the amount of Fe^{2+} in the 28 Da peak. This results in a much broader log-likelihood in Figure 3d and a correspondingly larger range of uncertainty in the composition of CO^{+} 1.7–18.1 compared to 9.2–10.9 at%. Equation (4) can be used to mathematically estimate the composition uncertainty, but what is its origin?

When peaks overlap there is a loss of information. The ions detected have a specific identity, either CO^{+} or $^{56}Fe^{2+}$, but our detection of them is simplified to only the m/z ratio and the total count in a peak. Without other information, such as kinetic energy sensing (Kelly, 2011), we are blind to the true identities. MLE is making an estimate of the most likely proportion of

https://doi.org/10.1017/S1431927618016276

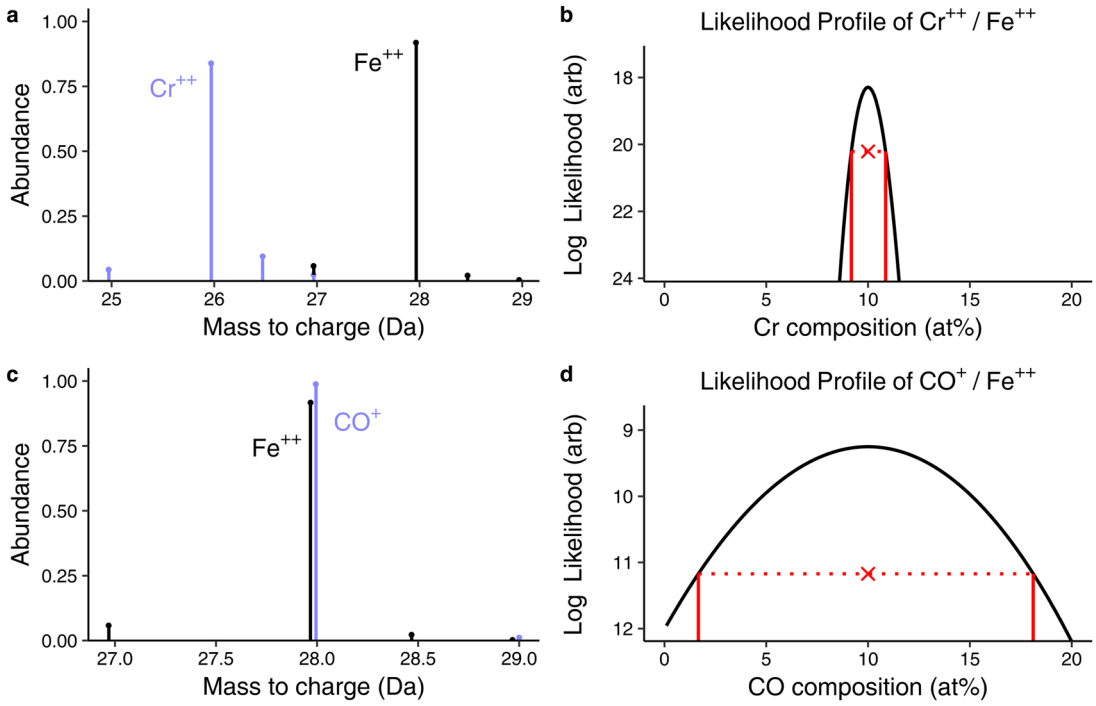

Fig. 3. Examples of simulated Fe/Cr and Fe/CO overlaps, 90% Fe in each with 1,000 counts total. **a, c:** Natural abundances and mass positions of Cr^{2+} with Fe^{2+} and CO^{+} with Fe^{2+} respectively (equal amounts). **b, d:** The log-likelihood function as a function of Cr and CO composition respectively. The cross indicates the maximum likelihood solution. The red vertical lines mark the 95% CI. Lower curvature of the likelihood profile indicates less information and lower certainty in the answer (wider CI_{95}).

ions in the overlapped peaks given the measured data and naturally takes account of stochastic uncertainty of information lost in the overlap. Appendix B shows how equation (4) is related to the standard deviation or standard error σ_i of composition $\mathbf{p}_i$. Because σ is proportional to the square root inverse of equation (4), then σ is reduced proportionally to $N^{-1/2}$ as in equation (1). Returning to Figure 3c, with sufficient counts, the composition becomes statistically more certain, because there is a lower probability of observing $^{56}Fe^{2+}$ without also observing some intensity in the other isotopes. However, the CO/Fe overlap will always have a larger σ than the Cr/Fe overlap for the same number of counts. Thus, MLE mathematically captures the "difficulty" of solving certain overlaps.

Uncertainty Measures

One common way of reporting uncertainty is to use a confidence interval (CI) and this interval gives a range of certainty for a measured value (Neyman, 1937). A CI is given for a specific confidence limit, such as 95%, written as CI_{95}. If an experiment is repeated, the predicted CI_{95} should contain the true value being measured 95% of the time. The vertical red lines in Figure 3 show the range of plausible values of composition estimated by the profile likelihood method (Pawitan, 2001). The CI_{95} is essentially the values of $\mathbf{p}$ that satisfy:

$$\mathcal{L}(\mathbf{r}, \mathbf{p}) > \mathcal{L}(\mathbf{r}, \hat{\mathbf{p}}) - \frac{1}{2}\chi_1^2(0.95) \tag{5}$$

where χ_1^2 is the chi-squared distribution with one degree of freedom and 0.95 is the level of confidence. When the MLE depends on more than two parameters, a Wald test can be used to give a CI (Pawitan, 2001). Again, as shown in Appendix B, the standard deviation (composition error) can be estimated via approximation to the normal distribution, and this can be used to define CI_{95} as:

$$\hat{\mathbf{p}} \pm 1.96\,\sigma \tag{6}$$

where 1.96 comes from a two-tailed test of the cumulative normal distribution for a 95% CI. These mathematical estimates of uncertainty are simple to calculate and can be applied to experimental measurements or to define the minimum number of ions required for sufficient precision in a planned experiment. Returning to Figure 2, the thickness of the connecting lines indicates the uncertainty calculated starting with equation (4). Thicker darker lines indicate more overlapped intensity and expected higher uncertainties. For the purpose of preliminary investigations, these uncertainties are made assuming equal amounts of all the ions in the overlap group—requiring no prior knowledge except to speculate at the ions present. In the next section, the new measurement of uncertainty defined above is compared to the existing methods.

Results

Comparison of Methods

The following example is taken from Sha et al. (1992) who investigated the overlaps of a Fe–l.85 wt.% C alloy. Sha observed an overlap of C_2^+/C_4^{2+} at 24 Da and calculated the ionic composition by a simple SP method using only the counts of the 24.5 Da peak and the isotopic abundance of $(^{12}C_3{}^{13}C)^{2+}$ species, values given in Table 2. For comparison with the SP method, the methods of LS and MLE were also applied to the same data. Each of the

https://doi.org/10.1017/S1431927618016276

Table 2. Details of the Peak Overlap at 24 Da Including the Peak Positions, Ions, and Natural Abundance of C_2 and C_4, Using Data from Sha et al. (1992).

m/z	Count	Species	C_2	C_4
24.0–24.2	732	$^{12}C_2^+$ or $^{12}C_4^{2+}$	97.79	95.63
24.5–24.7	17	$(^{12}C_3^{13}C)^{2+}$	0.00	4.29
25.0–25.1	4	$(^{12}C^{13}C)^+$ or $(^{12}C_2^{13}C_2)^{2+}$	2.20	0.07

methods gives a different C_4 concentration and CI_{95}: SP (52 ± 4)%, LS (58 ± 20)%, and MLE (64 ± 26)%. The fit achieved by SP is not perfect, as noted by Sha, 16 ions are expected at 25 Da, but only four are detected. This may indicate some isotopic detection bias. The predicted CIs also differ between the methods, with SP the narrowest and MLE having the widest. However, because the true composition is not known then it is difficult say which method provides an accurate estimate and CI.

To investigate the reliability of the predicted CIs, 10,000 simulated datasets were created from a multinomial probability [as per Miller et al. (1996)] with an ionic composition of exactly 40/60% C_2^+/C_4^{2+} at 24 Da, 50% detector efficiency, and 1,406 total counts giving 753 detected counts on average. Each simulated sample was solved by the three methods above and a 95% CI predicted. Equation (1) is used for SP, the residual standard error is used for LS, and a Wald test is used to predict the CI for the MLE solution. Fifty sorted random samples and their CI_{95} are shown in Figure 4. The mean and standard deviation of all the simulated data are shown on the left of Figure 4. The tally of how many times the predicted CI actually contained the true value is the CI coverage, and is shown on the right of Figure 4.

The three methods do not agree exactly on the average value. The SP method relies on a peak of very few counts (~8) and experiences some systematic bias (loss of accuracy) because of this. Additionally, this method has the highest standard deviation, and so has low precision. LS and MLE perform very similarly, with MLE having the lowest standard deviation (highest precision) for solving the overlap.

While the average and standard deviations of the results are similar between the different methods, the coverage of the predicted CIs is very different. The SP solution, with error given by equation (1), is much too narrow as it makes no account of the additional difficulty of solving the peak overlap. The predicted CI_{95} was only correct one in five times, rather than the expected 19 in 20 times. The CI for LS is given by the residual standard error with the assumptions of normally distributed residuals and equal variance for each peak (Hansen et al., 2013). Neither of these is good assumption in this situation due to the small peak counts, leading to a narrower CI which only captured the true value in 60% of trials. The CI predicted from MLE (Wald test) captures the true value 94% of the time, and therefore is an appropriate 95% CI. Sha et al. (1992) noted a 10% increase in C-concentration after solving the overlap, therefore the identity of this peak is significant in determining the overall carbon composition. The error bounds calculated by MLE here, although much larger than the counting error [Equation (1)], are much more representative of the uncertainty in the experiment. Next some other practical examples are considered.

Monte Carlo Uncertainty

The error calculated from equation (6) is approximate because it assumes a Gaussian distribution at the MLE composition. Even with a good error estimate, practically it can be difficult to propagate the error estimate because composition is a ratio, the difficulty of combining multiple overlap groups and due to highly correlated data treatments e.g., converting from atomic to weight%. However, numerical simulations, like that demonstrated in Figure 4, which sample from appropriate distributions can be used to produce error estimates (Anderson, 1976; Cox & Siebert, 2006). With a model for the statistical distribution of counts in each mass-peak, multiple random simulations, or Monte Carlo (MC) trials can be made and compositions estimated for each trial. In this manner, many simulated mass-peaks are generated, one such simulation is shown in Figure 5a. The input data are similar to that shown in Figure 1, with Cr20–Fe60–Ni20 at% and a multinomial distribution is assumed for the peak counts. The gray bars are the average input data and the black error bars indicate the spread of the distribution of MC trials used. One single set of peak counts is shown by the red crosses. Each set of peaks is solved and a distribution of compositions is accumulated. The average composition is shown in Figure 5b, again the red crosses show a single simulated data point and the error bar indicates twice the standard deviation of the estimated compositions. This numerical methodology is extended beyond a single overlap group to real data shown next.

Case Study: Line Profile

A section of atom probe data from a Fe-ODS steel is shown in Figure 6a. This contains two large Y–Ti oxide clusters and several smaller clusters. The composition of overlaps between Ti^{2+}/C_2^+, TiO^{2+}/O_2^+, and TiO^+/Fe_2O^{2+} ions is estimated, with suitable error bars estimated from the MC method in the previous section. All ions shown in Figure 6a were split into 12 sections and the composition calculated at each point to produce an overlap-free composition line profile, Figure 6b. The overlap-solved counts for every position and each *m/z* range are given in the Supplementary information. The TiO^{2+}/O_2^+ overlap can greatly alter the measured Ti:O ratio in oxide clusters in these materials, with approximately 30% of the total counts from the clusters in this overlapped peak (London et al., 2015*a*). The two large clusters in the line profile have very different amounts of TiO^{2+} and O_2^+ but this difference has been successfully accounted for by the MLE. Before solving the peak overlaps, the second cluster apparently contained an increased C-content, but this was due to the overlap at 24 Da Ti^{2+}/C_2^+. The Ti-composition profile from Figure 6b is shown in isolation in Figure 6c. As well as the overlap solved composition, the composition calculated by considering each of the overlaps as only having a single identity, either "R1" (Ti^{2+}, TiO^{2+}, and TiO^+) or "R2" (C_2^+, O_2^+, and Fe_2O^{2+}), are also shown. These single-identity composition profiles demonstrate systematic bias present from the peak identification, which is much greater than the error bars displayed [calculated from equation (1)]. The error bar at 26 nm, is about 40% larger than the counting error alone, this shows a modest increase the total composition uncertainty from the overlap uncertainty plus the non-overlapped species.

Discussion

If a peak is identified entirely as a single ion, a commonly applied approach, a reduction in accuracy occurs (as shown in Fig. 6c). Solving the overlap by one of the stated methods brings some improvement in accuracy but there is a variable amount of

https://doi.org/10.1017/S1431927618016276

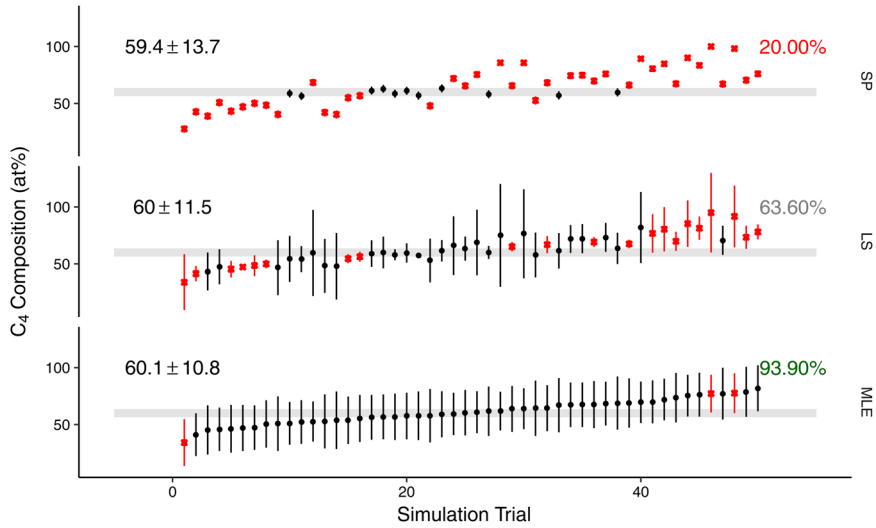

Fig. 4. Simulation results of the C_2^+/C_4^{2+} overlap (in the ratio exactly 40:60) solved using a SP, LS, or MLE, mean ± SD shown on the left. Arranged by increasing MLE estimate, each trial is independent. Points show the composition estimate and lines show the predicted CI. The gray band is the true spread of compositions in the simulation. CIs are colored black for containing the true value and red for missing it. How often the CI captures the true value is shown on the right.

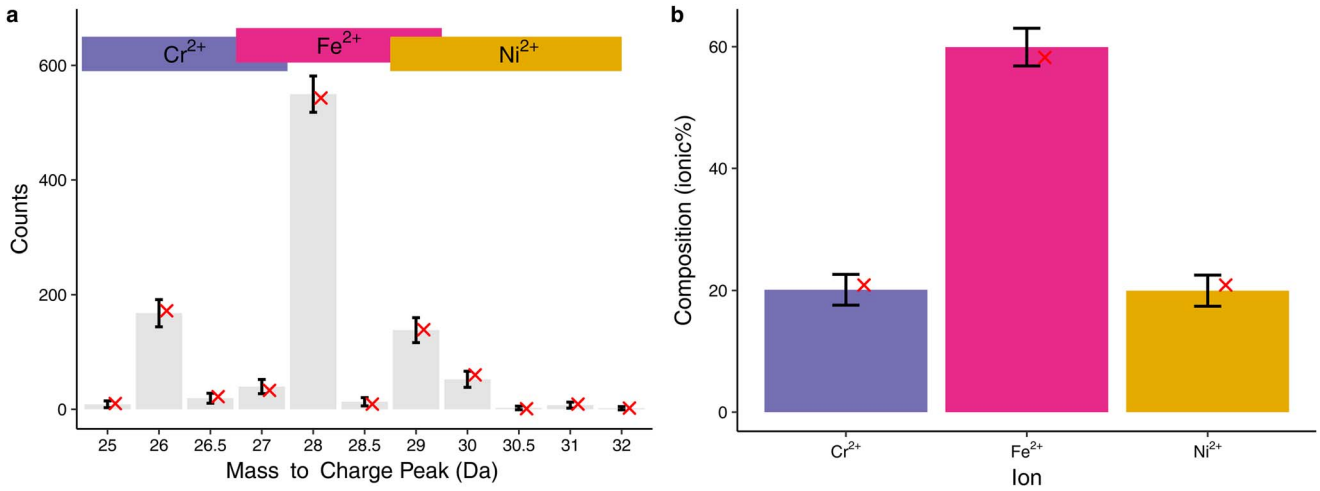

Fig. 5. a: Gray bars: background subtracted counts for each *m/z* peak, error bars: two standard deviations (2σ) of the multinomial distribution used in the MC method, red crosses: one individual MC trial. **b:** Colored bars: mean result of MLE of the MC simulated peak counts in (**a**), black error bars: 2σ of the resulting composition distribution, red crosses: the solution to the single MC measurement shown in (**a**).

uncertainty relating to the presence of overlaps. The overlapping ions and overlapped peaks can be concisely shown on an overlap diagram and the uncertainty of different overlaps can be represented visually by the line thicknesses, as shown in Figure 2.

The commonly used counting error equation [equation (1)] is a suitable estimate of the minimum composition error of non-overlapped peaks. Indeed, it does describe the distribution of counts of ions into overlapped peaks. However, when peaks are overlapped, information about the ion identity is lost and additional uncertainty is introduced. This additional source of uncertainty is well described by the error estimated from the MLE. The MLE-predicted CI, though wider than counting error alone, are more representative of the precision of the result (as shown in Fig. 4).

An error estimate can be made from the MLE in several ways, either using the analytical formula assuming normality (see Appendix B) or numerically using MC. The former is quick and less accurate, especially with very small compositions or low counts, and the MC method is slower but more accurate. For solving overlaps, MLE has also been shown to be the method with the smallest standard deviation and closest to the true value in simulations (Miller et al., 1996; London et al., 2017), showing it is an advantageous method to use. However, as with all techniques there are some disadvantages and limitations.

The primary disadvantage is the increased computation required in the MLE. While the calculation of a single MLE is very quick, several thousand calculations can take several minutes for the MC method above. Another, mostly practical disadvantage is the potential for numerical instability: the MLE may not converge to a maximum. This is a particular problem with overlap groups of very many ions (>20) or when compositions are very close to zero. This limitation can be overcome with suitably robust numerical methods but requires more input than a simple LS fit. Finally, the MLE is only as good as the assumption of

https://doi.org/10.1017/S1431927618016276

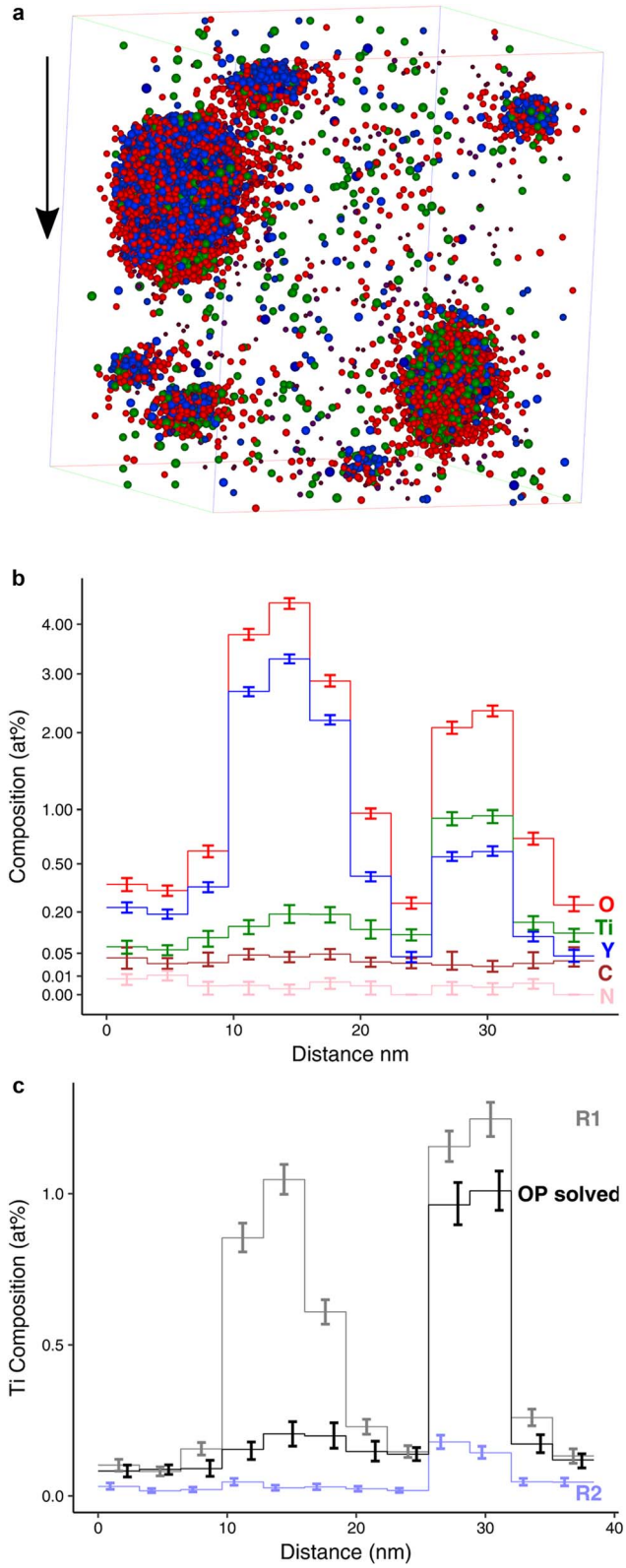

Fig. 6. a: Section of 3D reconstructed data from a Fe–Y–Ti ODS steel, volume 35 × 25 × 35 nm^3. Color key: Y-containing (blue), Ti-containing (green), O-containing (red), and C (brown). Fe ions are not rendered for clarity. **b:** Composition line profile following the arrow shown in (**a**), with the peak overlaps solved for each distance point. 95% confidence limits are shown by the error bars, calculated by the MC method described in the text. **c:** The Ti profile from (**b**), as calculated using single identities R1 (Ti^{2+}, TiO^{2+}, and TiO^{+}), R2 (C_2^{+}, O_2^{+}, and Fe_2O^{2+}) or by solving the overlaps (labeled OP solved).

multinomial distribution that underpins it. If, for whatever reason, this assumption is not valid, then the MLE will not contain valid information either. For example in the case of fewer than ten total counts, significant bias could be present and an alternative formulation may be necessary, similar to that of Danoix et al. (2007*b*).

In the case where ions have a significant percentage of intensity in a single peak, a monoisotopic limit may apply. Certain overlap problems may incur bias when there are insufficient counts of particular ions and they appear to only contain a single isotope. This is described in detail and quantified in Appendix C. A check should be made to ensure the counts of ions are enough to avoid this limit; greater than 415 counts for ^{12}C, 1,930 for ^{16}O, and 30,700 for ^{1}H for example. This limit affects all overlap-solving methods, however MLE is able to capture useful uncertainty information when this occurs.

One of the limitations of all overlap-solving methods is the reliance on natural abundance data. For some light elements, such as boron and carbon, there is an uncertainty in the terrestrial abundance; for example the ^{12}C abundance range is 0.9885–0.9904 (Berglund & Wieser, 2011). While not as important for smaller samples, in larger sample sizes this will place a lower limit on the minimum achievable error. Another key issue is the assumption that all ions are detected with equal probability. For certain species, significant loss of certain isotopes is observed (Thuvander et al., 2011; Meisenkothen et al., 2015). In principle, these can be accounted for in [**A**], but additional information would be required, also bringing its own uncertainty. However, the current implementation allows full customization of the elements and their natural abundance as required.

Most of the preceding discussion considers primarily random errors, whereas the misidentification of specific ions is a systematic bias, and will affect accuracy. To understand these errors it is necessary to consider how well the overlap-solved result fits with experiment (the goodness of fit) as well as the choice of ions used in the fit. Both of these are addressed in the following sections.

Goodness of Fit

If the overlap-solved fit is poor, not all the ions have been specified or the natural abundance does not match the experiment. Therefore it is important to take note of when the fit does not agree well with the measured data. Two methods are proposed to inspect the fit: (1) peak residuals and (2) χ^2 goodness of fit. Table 3 shows the counts of each ion in each peak as found by the three different methods for the example considered in the Section "Comparison of Methods". The peak residuals (ε) are the difference between the observed and fitted peak counts. The residuals are positive for all the methods for the peak at 24 Da, meaning this peak has too many counts compared to the expected value, and conversely, the residuals for the other peaks are negative. To reduce the residual, the 24 Da peak should be smaller or the other peaks larger, or some combination of the two; the residual offers no preference for which is correct.

The residual squared sum (RSS) describes the absolute magnitude of all of the residuals and indicates the total difference of the fit to the observed data. Inspecting the individual peak residuals may also help to identify peaks which are poorly fitted, however even when only a single peak is poorly fitted, the residual is balanced among all of the peaks. The LS method has the smallest RSS, as seen in Table 3, because LS minimizes this metric. The

https://doi.org/10.1017/S1431927618016276

Table 3. From the Fits Performed in the Section "Comparison of Methods": Fitted Counts for Both Ions, Residuals (ε), RSS, and χ^2 *P*-value with one Degree of Freedom χ^2 are Shown for Each Fitting Method.

	Exp. Counts	SP			LS		MLE			
		C_2	C_4	ε	C_2	C_4	ε	C_2	C_4	ε
24 Da	732	353.0	379.0	0.0	316.0	415.9	0.1	267.5	464.4	0.1
24.5 Da	17	0.0	17.0	0.0	0.0	18.7	−1.7	0.0	20.8	−3.8
25 Da	4	7.9	0.3	−4.2	7.1	0.3	−3.4	6.0	0.3	−2.4
RSS				17.8			14.4			20.3
χ^2 *P*-value				0.04			0.19			0.21

MLE method has the largest RSS and the SP method gives a smaller RSS but not as small as the LS fit. MLE maximizes the likelihood of the data, and this is not the same as minimizing the residual.

The χ^2 test statistic is another commonly applied metric of fit. The χ^2 test statistic is given in terms of the observed and expected (fitted) data:

$$\chi^2_c = \sum_i \frac{(\mathrm{Obs}_i - \mathrm{Exp}_i)^2}{\mathrm{Exp}_i} \tag{7}$$

This test statistic is compared to cumulative χ^2 distribution to derive a *P*-value with a number of degrees of freedom equal to the number of ranges minus the number of fitted ions. In the case of overlap solving, the solution should have as large a *P*-value as possible, meaning that the variation of the fit error is as random as possible. In Table 3 the χ^2 *P*-value indicates the order SP < LS < MLE, with MLE having the largest *P*-value. However, the fit was not optimized to maximize χ^2, therefore it is primarily useful to indicate significant errors, that is, when very small *P*-values are observed (<0.01), rather than to justify the fit is correct.

These two metrics of fit can help identify mistakes in the overlap solving, but how can mistakes identified be corrected? Usually a bad fit occurs because the wrong ions have been specified or an ion of significant quantity has not been included. The following section introduces a third metric which can be used to guide the identification of ions used and therefore reduce bias.

Identifying Ions

The identification of ions is required for all forms of APM and is a major source of systematic error (Haley et al., 2015). Defining what ions are present or expected is difficult due to experimental factors and is often determined by operator experience. Thus a significant inter-operator bias may be introduced. To investigate this effect, a simulated overlap of Fe^{2+}/Ni^{2+} ions was made by drawing from a multinomial distribution with a composition of 90% Fe and 10% Ni of 1,000 total counts. The simulation was repeated 10,000 times; additional information is included in the Supplementary information. The overlap was solved by MLE and LS with Si^+ included. Even though no Si should be detected, due to stochastic fluctuations in the data, MLE or LS will sometimes find it beneficial to include some Si in the fit to improve the likelihood or reduce the residual respectively. In this test, LS found 4.5% Si and MLE 4.2% Si on average. Such a large quantity of Si is found because of the intensity or difficulty of this overlap. However, MLE predicts a smaller quantity of the falsely included species than LS, suggesting MLE may be more robust to the identification of ions than LS.

The ions used to fit the data can be thought of as a model. Different combinations of ions constitute different models. Including Si in the overlapped ions (model) above is an example of over-fitting the data. However, there are methods of model selection which can reduce over fitting. One such method was introduced by Akaike (1974)—now known as the Akaike information criterion (AIC). The AIC can be used to compare the relative merit of models based on the number of parameters k and the maximum log-likelihood with those parameters $\mathcal{L}$:

$$\mathrm{AIC} = 2k - 2\mathcal{L} \tag{8}$$

The AIC is reduced with a larger likelihood (better fit) but is penalized for models with more parameters. The benefit of increased fit must be significant enough to justify adding another parameter. Therefore, with a set of different ions to be considered for an overlap, the AIC may be used to decide which ions to use. In the Fe/Ni/Si example, the AIC value of models with and without Si was compared and AIC selected the correct model 92% of the time. If the AIC is used to select between adding Si or not, then the average Si detected drops to 1.7 at%.

Figure 7 shows another example: the counts of ions near the 28 Da peak in a zirconium dataset from Hudson et al. (2011). These peaks can be fitted using Al/Fe with the possibility of CO^+ present at the 28 Da peak. When CO^+ is included, the measured composition is 9.16 ± 9.84% CO^+ and the residual equals zero because the choice of ions can perfectly fit the data. In this case, it is difficult to say whether CO^+ should be included or not, as the true answer is not known. Hudson et al. (2011) stated the relative isotopic abundances agreed "within the margin of error" and therefore the contribution of CO^+ was ruled out.

Can AIC guide the identification of the ions in this case? The Al/Fe/CO fit has a smaller AIC than the Al/Fe model, but only by 1.1. Burnham & Anderson (2003) suggest a difference in AIC of least two for it to be considered significantly different. Adding only ten additional counts to the 28.5 Da peak swaps the AIC result. Therefore AIC only offers weak evidence to include CO^+. Hudson et al. (2011) were probably correct to exclude CO^+ in this case, but this highlights difficulty in identifying ions, even when using AIC. The 28 Da peak contained 11,200 counts which would often be classified as "enough counts", ~1% counting error, but again the overlap difficulty plays a significant role. The large uncertainty in CO^+ content (0–20%) is symptom of the difficulty of this overlap problem.

https://doi.org/10.1017/S1431927618016276

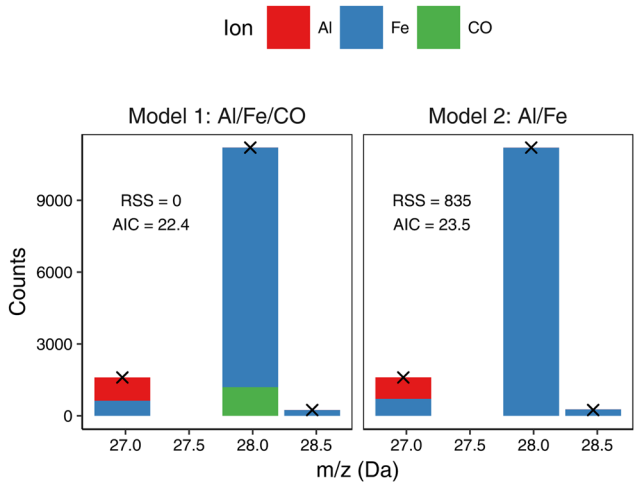

Fig. 7. Choice of ions in an Al/Fe/CO overlap. Two models are fitted, Al/Fe/CO and Al/Fe, the bars show the fitted results and the crosses mark the observed peak counts. The RSS and AIC are shown in the text to show the acceptability of each model's fit.

Returning to the Fe/Ni/Si example, the sensitivity of AIC to the Si content can be investigated by varying the content of Si in simulated data and comparing the AIC of Fe/Ni or Fe/Ni/Si models. With 1,000 counts, AIC prefers the simpler Fe/Ni model until the Si content reaches 1.5%, however, with more than 1.5% Si in the data, AIC prefers the more complex Fe/Ni/Si model. This threshold value is reduced proportionally with the square root of total counts. AIC is also more sensitive to simpler overlaps, if Si is replaced with Al, then the threshold value is reduced to 0.17% because Al has less overlapped intensity.

Therefore the overlap uncertainty affects not only the precision of the estimated composition, but also the difficulty of which ions to include in the fitting procedure. However, AIC can be used as a quantitative measure to show the relative merit of different ion identifications and therefore help reduce ion selection bias.

Improvements

In all the analysis presented above, the peak counts are quantified from the m/z spectrum in a very simple manner, using the background corrected counts in a specific m/z range. Using methods which take account of the whole peak shape will undoubtedly bring improvement, especially in robustly removing the influence of peak tails. For example, Johnson et al. (2013) used full-peak shape fitting and a LS method to constrain the isotopic abundances; this could be improved by applying MLE to constrain the solution instead.

A minimum abundance specification is necessary due to limited counts and the influence of the background which makes small peaks unreliable for quantification. The current use of a single value for the minimum detectable abundance does not take account of individual ionic compositions and a varying local background. These effects could be incorporated to the method, using a criterion: signal count $> 14.1\sqrt{\text{background}}$ (Larson et al., 2013), after estimating the signal count from an initial calculation of composition.

The assumption that detection efficiency is independent of composition and isotope, does not always hold. Additional information could be supplied to the MLE to take account of these discrepancies. However, this would require more detailed knowledge of both the evaporation physics and electronic detection system.

Conclusion

APM is ideal for detecting subtle changes in composition due to microstructural evolution, but the significance of these changes should be seen in the light of the measurement uncertainty. The uncertainty in composition can be accurately and reliably estimated using maximum likelihood methods, and in the presence of severe overlaps, the uncertainty can be much greater than that predicted from counting error alone. The uncertainty can be estimated quickly using an analytical formula or using MC methods for more complex analysis combining multiple overlap groups and non-overlapped mass peaks. An analyst needs to be careful not to misrepresent the data with overly optimistic/narrow error bounds. The difference in the size of an error bar may lead to noise being misinterpreted as being actual information.

The primary disadvantage of using MLE is the added complexity of the analysis and the associated computation cost. The implementation written in MATLAB is efficient enough to compute even the thousands of estimates required for the MC estimate of uncertainty in a few minutes, and was successfully demonstrated on real atom probe data (see the section "Case Study: Line Profile").

Ultimately, the MLE method presented primarily deals with the random counting error and not systematic errors arising from instrumental and physical effects, which may be much larger. The appropriateness or goodness of the resulting fit should be inspected and reported to show no ions have been misidentified; either RSS or χ^2 statistic could be used. The goodness of the fit and the total number of counts can be used to convey confidence in the measured composition, larger peaks carry more statistical significance and more greatly influence the total composition than smaller ones.

The overlap difficulty reduces the composition precision but also makes the identification and selection of ions more difficult, this in turn reduces the accuracy of the measurement. To address the identification of ions used, an objective function (such as AIC) is anticipated to reduce inter-operator bias and to give confidence to the identification. However, a greater effort is required by the community to address the biases and uncertainty arising from the instrument, sample, evaporation, detection, reconstruction, and peak identification processes.

Implementations of the methods described, written in MATLAB, are available at http://AtomProbeLab.sourceforge.net.

Supplementary Material. The supplementary material for this article can be found at https://doi.org/10.1017/S1431927618016276.

Author ORCIDs. Andrew J. London 0000-0001-6959-9849.

Acknowledgments. The author thanks the colleagues who provided discussion and feedback, especially Dan Haley and Hazel Gardner of the University of Oxford and Alfred Cerezo, as well as the members of CCFE, including Daniel Mason and Max Boleininger for many mathematical conversations. This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training program 2014–2018 under grant agreement no. 633053 and from the RCUK Energy Programme [grant number EP/P012450/1]. The views and opinions expressed herein do not necessarily reflect those of the European Commission. To obtain further information on the data and models underlying this paper please contact PublicationsManager@ccfe.ac.uk.

References

Akaike H (1974). A new look at the statistical model identification. *IEEE Trans Autom Control* **19**(6), 716–723.

Anderson G (1976). Error propagation by the Monte Carlo method in geochemical calculations. *Geochim Cosmochim Acta* **40**(12), 1533–1538.

Berglund M & Wieser ME (2011). Isotopic compositions of the elements 2009 (IUPAC technical report). *Pure Appl Chem* **83**(2), 397–410.

Burnham KP & Anderson DR (2003). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media. New York: Springer-Verlag.

Cox MG & Siebert BRL (2006). The use of a Monte Carlo method for evaluating uncertainty and expanded uncertainty. *Metrologia* **43**(4), S178–S188.

Danoix F, Grancher G, Bostel A & Blavette D (2007*a*). Standard deviations of composition measurements in atom probe analyses. Part I: Conventional 1D atom probe. *Ultramicroscopy* **107**(9), 734–738.

Danoix F, Grancher G, Bostel A & Blavette D (2007*b*). Standard deviations of composition measurements in atom probe analyses. Part II: 3D atom probe. *Ultramicroscopy* **107**(9), 739–743.

Efron B (1979). Bootstrap methods: Another look at the jackknife. *Ann Statist* **7**(1), 1–26.

Ferrige AG, Seddon MJ, Jarvis S, Skilling J & Aplin R (1991). Maximum entropy deconvolution in electrospray mass spectrometry. *Rapid Commun Mass Spectrom* **5**(8), 374–377.

Gault B, Moody MP, Cairney J & Ringer S (2012). *Atom Probe Microscopy*, vol. **160**. New York: Springer-Verlag, ISBN: 978-1-4614-3435-1..

Gault B, Saxey DW, Ashton MW, Sinnott SB, Chiaramonti AN, Moody MP & Schreiber DK (2016). Behavior of molecules and molecular ions near a field emitter. *New J Phys* **18**(3), 033031.

Geyer CJ (2013). Asymptotics of maximum likelihood without the LLN or CLT or sample size going to infinity. In *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton*, Jones G. and Shen X. (Eds.), pp. 1–24. Collections, Volume 10, Beachwood, OH: Institute of Mathematical Statistics.

Haley D, Choi P & Raabe D (2015). Guided mass spectrum labeling in atom probe tomography. *Ultramicroscopy* **159**, Part 2, 338–345.

Hansen PC, Pereyra V & Scherer G (2013). *Least Squares Data Fitting with Applications*. Baltimore, MD: JHU Press.

Hudson D, Smith G & Gault B (2011). Optimisation of mass ranging for atom probe microanalysis and application to the corrosion processes in Zr alloys. *Ultramicroscopy* **111**(6), 480–486.

Hyde J, Burke M, Gault B, Saxey D, Styman P, Wilford K & Williams T (2011). Atom probe tomography of reactor pressure vessel steels: An analysis of data integrity. *Ultramicroscopy* **111**(6), 676–682.

Johnson L, Thuvander M, Stiller K, Odén M & Hultman L (2013). Blind deconvolution of time-of-flight mass spectra from atom probe tomography. *Ultramicroscopy* **132**, 60–64.

Kelly TF (2011). Kinetic-energy discrimination for atom probe tomography: Review article. *Microsc Microanal* **17**(1), 1–14.

Kelly TF & Larson DJ (2000). Local electrode atom probes. *Mater Charact* **44** (1–2), 59–85.

Kelly TF & Larson DJ (2012). The second revolution in atom probe tomography. *MRS bulletin* **37**(2), 150–158.

Larson DJ, Prosa TJ, Ulfig RM, Geiser BP & Kelly TF (2013). *Local Electrode Atom Probe Tomography: A User's Guide*. New York: Springer-Verlag, ISBN: 978-1-4614-8720-3.

Lehmann E & Casella G (1998). *Theory of point estimation (Springer texts in statistics)*, by E. L. Lehmann, George Casella. New York: Springer-Verlag. ISBN 0-387-985026.

Li Y, Choi P, Borchers C, Westerkamp S, Goto S, Raabe D & Kirchheim R (2011). Atomic-scale mechanisms of deformation-induced cementite decomposition in pearlite. *Acta Mater* **59**(10), 3965–3977.

Liu J, Wu C & Tsong TT (1991). Measurement of the atomic site specific binding energy of surface atoms of metals and alloys. *Surf Sci* **246**(1–3), 157–162.

London AJ, Haley D & Moody MP (2017). Single-ion deconvolution of mass peak overlaps for atom probe microscopy. *Microsc Microanal* **23**(2), 300–306.

London AJ, Lozano-Perez S, Moody MP, Amirthapandian S, Panigrahi BK, Sundar CS & Grovenor CRM (2015*a*). Quantification of oxide particle composition in model oxide dispersion strengthened steel alloys. *Ultramicroscopy* **159**(2), 360–367.

London AJ, Santra S, Amirthapandian S, Panigrahi BK, Sarguna RM, Balaji S, Vijay R, Sundar CS, Lozano-Perez S & Grovenor CRM (2015*b*). Effect of Ti and Cr on dispersion, structure and composition of oxide nano-particles in model ODS alloys. *Acta Mater* **97**, 223–233.

Marquis EA (2008). Core/shell structures of oxygen-rich nanofeatures in oxidedispersion strengthened Fe–Cr alloys. *Appl Phys Lett* **93**(18), 181904.

Meisenkothen F, Steel EB, Prosa TJ, Henry KT & Kolli RP (2015). Effects of detector dead-time on quantitative analyses involving boron and multi-hit detection events in atom probe tomography. *Ultramicroscopy* **159**, 101–111.

Miller MK, Cerezo A, Hetherington MG & Smith GDW (1996). *Atom Probe Field ion Microscopy*. Oxford: Oxford Science Publications, ISBN: 978-0-19-851387-2.

Müller EW, Panitz JA & McLane SB (1968). The atom-probe field ion microscope. *Rev Sci Instrum* **39**(1), 83–86.

Neyman J (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **236**(767), 333–380.

Pawitan Y (2001). *In all Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford, UK: Oxford University Press.

Philippe T, Duguay S & Blavette D (2010). Clustering and pair correlation function in atom probe tomography. *Ultramicroscopy* **110**(7), 862–865.

Sha W, Chang L, Smith G, Cheng L & Mittemeijer E (1992). Some aspects of atom-probe analysis of Fe–C and Fe–N systems. *Surf Sci* **266**(1), 416–423.

Thuvander M (2016). On the accuracy of compositional quantification for atom probe tomography. *Microsc Microanal* **22**(S3), 642–643.

Thuvander M, Weidow J, Angseryd J, Falk L, Liu F, Sonestedt M, Stiller K & Andren H-O (2011). Quantitative atom probe analysis of carbides. *Ultramicroscopy* **111**(6), 604–608. Special Issue: 52nd International Field Emission Symposium.

Appendix A

The log-likelihood of the multinomial distribution can be derived from the multinomial probability distribution of K bins, for a random variable $\mathbf{R}$:

$$P(\mathbf{R}=\mathbf{r};N,\mathbf{q})=N!\prod_{i=1}^{K}\frac{\mathbf{q}_i^{\mathbf{r}_i}}{\mathbf{r}_i!}$$

where $\mathbf{r}_i$ is the count in the ith bin and $\mathbf{q}_i$ is the probability of any single trial belonging to that bin (success). In this distribution $\mathbf{r}_i$ sums to N and $\mathbf{q}_i$ sums to 1. The log-likelihood $\mathcal{L}$ is given by:

$$\begin{aligned}\mathcal{L}(\mathbf{q},\mathbf{r})&=\log P=\log\left(N!\prod_{i=1}^{m}\frac{\mathbf{q}_i^{\mathbf{r}_i}}{\mathbf{r}_i!}\right)\\&=\log(N!)+\log\left(\prod_{i=1}^{m}\frac{\mathbf{q}_i^{\mathbf{r}_i}}{\mathbf{r}_i!}\right)\\&=\log(N!)+\sum_{i=1}^{m}\log\left(\frac{\mathbf{q}_i^{\mathbf{r}_i}}{\mathbf{r}_i!}\right)\\&=\log(N!)+\sum_{i=1}^{m}\mathbf{r}_i\log\mathbf{q}_i-\sum_{i=1}^{m}\log(\mathbf{r}_i!)\end{aligned}$$

However, the probability for success in each bin is actually given by the matrix multiplication of the abundance matrix $[\mathbf{A}]$ and the overlap composition $\mathbf{p}$:

$$\mathbf{q}=[\mathbf{A}]\mathbf{p}$$

Therefore the summation over $\mathbf{q}_i$ above can be expressed as a matrix multiplication with $\mathbf{r}^{\mathrm{T}}$ to give:

$$\mathcal{L}(\mathbf{p},\mathbf{r})=\mathbf{r}^{\mathrm{T}}\log([\mathbf{A}]\mathbf{p})+\log\left(\sum_{j=1}^{m}\mathbf{r}_j\right)-\sum_{j=1}^{m}\log(\mathbf{r}_j!)\qquad(A.1)$$

https://doi.org/10.1017/S1431927618016276

Only the first term of $\mathcal{L}(\mathbf{p}, \mathbf{r})$ depends on $\mathbf{p}$, therefore the maximum likelihood estimate may be made by considering only the following:

$$\mathcal{L}(\mathbf{p}, \mathbf{r}) \propto \mathbf{r}^{\mathrm{T}} \log([\mathbf{A}]\mathbf{p})$$

Appendix B

The observed Fisher information is given by the expectation value of the second derivative of likelihood function with respect to θ (Lehmann & Casella, 1998):

$$\mathcal{I}(\theta) = -\text{Expectation}\left[\frac{\partial^2}{\partial\theta^2}\mathcal{L}(X;\theta)|\theta\right]$$

where $\mathcal{L}(X;\theta)$ is the likelihood function, the probability for the data X given the model parameter θ. $\mathcal{L}(X;\theta)$ is maximized to find the parameter estimate $\hat{\theta}$. If the MLE is a consistent estimator and the log-likelihood can be adequately approximated by a quadratic function (Geyer, 2013), the lower limit of the variance can be estimated from the Cramer–Rao bound:

$$\text{Var}(\hat{\theta}) \geq (\mathcal{I}(\theta))^{-1} \tag{B.1}$$

This means that the uncertainty of the solution to equation (3) is inversely related to the curvature at the solution. Using the symbols defined in the Section "Separating Overlaps", the second derivative of equation (3) with respect to $\mathbf{p}$ is:

$$\frac{\partial^2\mathcal{L}}{\partial\mathbf{p}_m\partial\mathbf{p}_k} = \sum_i \frac{(\mathbf{r}_i\mathbf{A}_{ik}\mathbf{A}_{im})}{\left(\sum_j \mathbf{A}_{ij}\mathbf{p}_j\right)^2} = \mathbf{H} \tag{B.2}$$

This is known as the negative Hessian, because it is derived from the negative log-likelihood and describes the local curvature.

Equation (B.1) also allows CIs to be computed. A Wald interval can be constructed, assuming $\mathcal{L}$ is regular (Lehmann & Casella, 1998). Using equation (B.1), the asymptotic normal covariance (fitting the $\mathcal{L}$ to a quadratic) is estimated by the inverse of the negative Hessian [equation (B.2)]:

$$\text{Covar}(\mathbf{p}) = (\mathcal{I}(\theta))^{-1} = \mathbf{C} \tag{B.3}$$

The standard deviation (SE) is given by the square root of the diagonal elements of the covariance matrix:

$$\sigma_j = \sqrt{\mathbf{C}_{jj}}$$

One limitation of this approach is that it ignores the requirement that $\mathbf{p}$ should sum to 1. This could be achieved by a transformation of variables, using a trigonometric or exponential function.

Appendix C

Monoisotopic Limit

When two monoisotopic (containing a single isotope) species overlap, there are more unknowns than knowns in equation (2) and it has no unique MLE estimate. However, a species may appear practically monoisotopic depending on the counts and abundance of side peaks. The problem may appear solvable, but if a species is practically monoisotopic, bias will be introduced. If the abundance of isotopes other than the main peak is small or the counts of the ion are low then a multi-isotope ion may appear as a single peak. When the monoisotopic limit occurs can be estimated.

In the case of an ionic isotopic distribution, of a single ion type, where there is one prominent peak, the probability that no other peaks are detected is given by:

$$P(p, n) = 1 - p^n$$

where p is the natural abundance of the prominent peak and n is the number of ions of this type. This can be rearranged to give the number of ions required so there is a $P = 99\%$ chance at least one count will be observed outside the prominent peak:

$$n = \frac{\log(1-P)}{\log(p)} - 1 \tag{C.1}$$

So when solving an overlap of TiO^{2+} with O_2^+, $p(^{48}Ti^{16}O^{2+}) = 0.736$, no side peaks of TiO^{2+} will be detected more than 1% of the time if there are fewer than 14 counts. This means, when there are only 14 counts of TiO^{2+}, with a probability of more than 1% no side peaks are detected and the overlap will *appear* to only contain O_2^+, even if it is purely TiO^{2+}. This introduces bias to the measured composition. The monoisotopic limit ($P = 99\%$) for other elements are: ^{12}C: 415 counts, ^{16}O: 1,930, and ^{1}H: 30,700. Therefore, depending on the overlapped ions and the number of counts, bias may be introduced from this practical limitation. When practically monoisotopic species overlap, there is not enough information present in the peak counts alone to determine the composition. The MLE will return a (biased) estimate, but it will not be possible to compute the uncertainty from equation (4) because the resulting matrix is singular. The uncertainty can be estimated using a method like the profile likelihood (see the "Uncertainty Measures" section), which captures information in the log-likelihood profile directly.

For example using the profile likelihood method, for 100 counts in the 32 Da peak, for the TiO^{2+}/O_2^+ overlap, a composition CI_{95} of 0.93–1.00 O_2 is estimated, but for the S^+/O_2^+ overlap a the CI_{95} is 0.62–1.00 (both excluding ^{18}O). This is because more intensity of the TiO peaks should fall outside of the overlapped peak. TiO/O_2 is therefore an "easier" overlap than S/O_2 with a correspondingly lower uncertainty.

https://doi.org/10.1017/S1431927618016276