JAAS

Accepted Manuscript

Hg

Pb

Fe

ROYAL SOCIETY
OF CHEMISTRY

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

ROYAL SOCIETY
OF CHEMISTRY

www.rsc.org/jaas

Detection in LA-ICPMS: construction and performance evaluation of decision rules

Alex Ulianov,*[a] Othmar Müntener,[a] Urs Schaltegger[b] and François Bussy[a]

[a]Institute of Earth Sciences, University of Lausanne, Géopolis, 1015 Lausanne, Switzerland
[b]Section of Earth and Environmental Sciences, University of Geneva, Rue des Maraîchers 13, 1205 Geneva, Switzerland

Abstract

Laser ablation inductively coupled plasma mass spectrometry (LA-ICPMS) is frequently employed for the analysis of minute isotope contents in the presence of a background noise. Distinguishing between the sample signal and the background noise at a given confidence level thus represents a routine challenge. For count numbers N_b and N_s collected during (equally long) background and sample measurements, respectively, the statistical significance of their net value, N_s-N_b, can be evaluated: how probable is it to obtain such value by subtracting two count number estimates coming from a common statistical distribution (i.e., when N_s and N_b represent measured estimates of the same mean activity)? If, based on the analysis of a model distribution of the net count numbers, we decide that this is probable, the signal is statistically indistinguishable from the background: the analysed isotope is not detected. If the corresponding (one-sided) probability is below some threshold, it is detected. The net signal value on the divide between the above alternatives, given in net counts or mass (content) units, is called critical level; optionally, it can be complemented by the computation of the detection limit; such values are often reported in the literature. Less discussed is the appropriateness of computational methods used to estimate these values. Troubles arise from attempts to apply Gaussian confidence intervals to small, discretely distributed count numbers contained in real LA-ICPMS acquisitions, and from a non-optimal estimation of the net count number standard deviation in some of the methods used for the computation of critical levels for paired measurements. Combined, these factors may result in uncontrolled, excessively high rates of false detections (background reported as detection of analyte in the sample). Here, we provide a review of methods, otherwise called decision rules, available for the critical level estimation and discuss how to evaluate the performances of these rules to enable an educated computation of LA-ICPMS detection capabilities, including the case of small count numbers.

Introduction

The International Union of Pure and Applied Chemistry (IUPAC) defines the chemical measurement process as 'a fully specified analytical method that has achieved a state of statistical control'[1]. The latter part of this definition implies that routines to evaluate accuracy, precision, and detection capabilities of the method are available, and that the measurement results are reported accordingly. At this time, in the field of trace element LA-ICPMS, a 'state of statistical control' does not seem to be achieved. This review covers questions related to the estimation of LA-ICPMS detection capabilities. The appropriate solutions have been available for a while, most of them remaining beyond the ICPMS literature - in texts of mathematical statistics, radioactivity monitoring / protection, and biometrics.

The problem of quantifying the detection capability of an analytical method has several aspects. Firstly, one may wish knowing if a sample signal is statistically distinguishable from the background noise. This question can be formalised somewhat differently, depending on the amount of data available to characterise the background noise and the way chosen to correct the gross signal for the background value in order to estimate the net activity (net count number, or count number per second / intensity, cps). In the general case, the hypothesis is discussed that the measured sample signal value and the measured background value(s), expressed in count number or intensity units, belong to statistical distributions having the same mean. As a special case of this hypothesis, when the background is well characterised, it is often discussed if the measured sample signal value belongs to the distribution of the background values. A net value, or level, characterising the difference between the sample signal and the background values is usually provided, above which the one-sided probability of accepting this hypothesis, when it is true, is below some threshold set by the analyst. Mathematical statistics and metrology offer a number of methods (decision rules, tests), to quantify this level; above it, the sample signal is considered detected. In modern terminology, this level is usually called *critical level*, or *critical value*[1-3]. As alternatives, the terms *decision level, decision threshold* and *minimum significant measured activity* are used, e.g., in studies on the detection of Poisson-distributed signals in the field of radioactivity monitoring[4-6]. Secondly, for a sample showing some specified, precisely known net

activity, or containing a precisely known amount of analyte, calculations can be made to obtain the probability that the signal will be detected using a decision rule above. This probability characterises the *power* of the decision rule selected, which is, in turn, closely related to the concept of *detection limit* in its actual understanding[1-3]. It follows that critical level and detection limit values are always associated with some probability of *false detection* and *false non-detection*, respectively. For some decision rules, declared and actually observed values of this probability are poorly matched in the range of small count numbers; studying the difference between them is an important part in the evaluation of a decision rule[5-8]. Finally, in addition to the computational aspects above, discussing detection capabilities has a potential for terminological controversy. A record of such controversy is contained in the annals of chemical metrology[1-3,9].

Still, quantifying detection capabilities is worth discussing. Otherwise, (i) too frequently reporting the background noise as the detected presence of analyte in a sample becomes possible; (ii) performances of analytical instruments and methods cannot be adequately compared; it becomes difficult to numerically describe the improvement of an analytical technique operating with trace analyte contents.

In this review, we mainly focus on the definition and computation of the critical level: this parameter is sufficient for taking the decision 'detected' or 'not detected' and for the corresponding screening of the data collected[1-6]. Such screening is part of all trace element LA-ICPMS data reduction programmes and is applied to most trace element LA-ICPMS analyses obtained in practice. In particular, we consider decision rules suitable for the computation of critical levels for weak LA-ICPMS backgrounds containing very few counts: this practically important topic is not covered in the available ICPMS literature, albeit receiving much attention in the literature of biometrics and radioactivity monitoring. In compliance with the past and most of the recent LA-ICPMS practices, the concept of detection limit is given less attention. The latter rather represents a tool to characterise the inherent performance of a method (instrument) than to decide if the analyte is detected or not in an individual sample signal; besides, it is derived from and defined relative to the critical level[1-4,6]. Still, we provide an introduction to the field of detection limits also, to help the reader to avoid mixing the different detection concepts and terms. The current IUPAC recommendations and terminology[1,2] are followed when possible; in the domain of small counts, the IUPAC recommendations are insufficient and substituted by an introduction to the original literature on tests for the equality of two Poisson means.

This review includes two parts. Part 1 covers general questions of chemical metrology and mathematical statistics, including the formal definitions of critical level and detection limit and associated concepts from the theory of hypothesis testing, as well as the statistical interpretation of count number distributions and calculation of standard deviations. Whenever possible, the concepts are introduced in a distribution-independent form, but illustrated by examples from the Poisson distribution, given the importance of the latter for the description of ICPMS data. This part can be omitted by an experienced reader, or by a reader that looks for a ready solution for the practical calculation of critical levels. Part 2 covers questions related to the construction, performance evaluation, comparison and application of selected decision rules and their different realisations, including one or two recommendable, well performing and computation-friendly solutions suitable for the practical treatment of LA-ICPMS data.

It worth noting that, although selected to fit with the analytical protocol and data statistics typical of trace element LA-ICPMS, decision rules described in this review can be used for the quantification of detection capabilities in other areas of analytical chemistry where weak Poisson-distributed noise is encountered, e.g., in secondary ion mass spectrometry. Some relevant details are contained in Part 2.

Part 1. Mathematical tools of detection

1.1. Critical level: basic notions

Traditionally, two concepts related to the definition and calculation of critical levels in metrology are distinguished – *well-known background* and *paired measurements*[1-4,6]. The basic difference between them consists in the precision, with which the net activity is defined. There are, however, practically important situations when these concepts converge, which leads to the construction of identically formulated decision rules. Below, both of these concepts are introduced from the standpoint of the theory of hypothesis testing, followed by a more detailed analysis of the 'paired' concept in Part 2.

Well-known background

Let us consider a sample signal measurement that yields N_s counts. Besides, some number (k) of replicate background measurements, each of them yielding N_{bi} counts and performed during an acquisition time and at operating conditions identical to the sample measurement, are available to estimate the mean value of the background activity, which is further used to obtain the net sample activity value. If k is very large, such setup is called *well-known background*, or *well-known blank*. The requirement for k to be large can be alleviated, and the above setup - extended to any number of background replicates ($k \geq 1$), ultimately allowing to consider the well-known background as a generalisation of the concept of paired measurements. Although at smaller k the background value becomes less precise, the core idea of the background being well known suffers little alteration: to compute the net activity in the most precise way possible, all available background replicates are averaged.

In the context of a well-known background, one may wish to test the hypothesis that the signal and the background values come from the same statistical distribution, representing measured estimates of the same 'true' mean count number value equal to the mathematical expectation for this distribution. In statistics, this hypothesis is usually called *null hypothesis* (H_0). The competing hypothesis states that the signal and the background values come from two different distributions, with the mean signal value being larger than the mean background value. It is called *alternative hypothesis* (H_1). The null hypothesis above can be given in an equivalent formulation stating that the mean net signal count number is equal to zero, which immediately means that the mean signal count number is equal to the mean background count number. It is this formulation that is used (and is most convenient to use) in mathematical realisations of most statistical tests, including metrological applications.

If the null hypothesis holds true at a specified level of confidence, then the signal is considered as not detected. If it is false (meaning that the alternative hypothesis is true), then it is considered as detected. Let us discuss how to test if the null hypothesis, as formulated in the previous paragraph, is true.

From k individual background count number acquisitions, each of them yielding N_{b_i} counts, a mean count number value $\overline{N}_b$ is computed:

$$\overline{N}_b = (N_{b_1} + ... + N_{b_k})/k \qquad (1)$$

This mean value can be obtained with a high precision that improves with increasing k. At large k (typically, >15-20 in practical calculations), it can be considered precisely, or 'well', known.

Besides, a net count number value is defined for each of the individual background acquisitions:

$$N_{b\,net_i} = N_{b_i} - \overline{N}_b \qquad (2)$$

and also for the sample signal:

$$N_{s\,net} = N_s - \overline{N}_b \qquad (3)$$

Under the null hypothesis, all net values for the individual background acquisitions and the net sample value (eqn 2 and 3, respectively) form a common distribution with a mean value of zero. The net sample value is then compared to the common distribution above, to find out if it is probable for the net sample value to be part of this distribution, i.e. to test the null hypothesis for contradiction.

The standard deviation of the common distribution constructed under the null hypothesis [denoted as $s(N_s - \overline{N}_b)$] is not explicitly specified and needs to be estimated. By error propagation, we obtain[3,6]:

$$s(N_s - \overline{N}_b) = \sqrt{s^2(N) + \frac{s^2(N)}{k}} = s(N)\sqrt{1 + \frac{1}{k}} \qquad (4)$$

Under the null hypothesis, there is no difference between sample signal values and background values. Thus, N in the above equation denotes the number of counts in any measurement, including both background and sample measurements. Pooling these measurements to estimate $s(N)$, and computing the corresponding confidence intervals can be done in several ways, depending on the number of measurements available, type of their statistical distribution and the way used to estimate the standard deviation. For example, in the widely known two-sample Student's *t*-test, variances (squares of stand-

ard deviation) for the signal and the background are estimated from the sums of squared residuals, after which $s(N)$ is pooled as their weighted mean[10,11]. This way to estimate standard deviations is relatively imprecise; using t-values can be necessary to account for the increased uncertainty of the pooled standard deviation in the absence of a large number of replicates[10,11]. In the case of ordinary Poisson statistics often applicable for the description of ICPMS data, the corresponding standard deviation can be pooled as a square root of the mean count number computed from all available measurements, including both background measurements and sample measurement[7,8]:

$$s(N) = \sqrt{\overline{N}} = \sqrt{\frac{N_{b_1} + ... + N_{b_k} + N_s}{k+1}} \qquad (5)$$

The value under the radical of eqn (5) represents the most probable estimate for the mean, or mathematical expectation, of the count number distribution under the null hypothesis; it is also called *constrained maximum likelihood estimate, CMLE*[7,8]. This estimate is very precise [$s(\sqrt{\overline{N}}) = 1/2\sqrt{k+1}$; see the section 'Statistical distributions, standard deviations, and ICPMS data'].

Eqn (4,5) are applicable at any positive integer k. At $k=1$, for ordinary Poisson distributed data, they yield the famous relationship

$$s(N_s - \overline{N}_b) = \sqrt{s^2(N) + \frac{s^2(N)}{1}} = s(N)\sqrt{2} = \sqrt{\frac{N_b + N_s}{2}}\sqrt{2} = \sqrt{N_b + N_s} \qquad (6)$$

which constitutes the basis of some of the earliest decision rules for Poisson distributed data known in mathematical statistics and metrology[4]; it will be discussed in detail in Part 2 of this text.

At large k, eqn (4,5) can be simplified. In the expression under the radical of the right-hand part of eqn (5), the background values have a much larger weight than the signal value, provided k is indeed large. Consequently, it is a common practice to pool the standard deviation of the common count number distribution arising under the null hypothesis from the background values only, e.g.:

$$s(N) = s(N_b) = \sqrt{\frac{N_{b_1} + ... + N_{b_k}}{k}} = \sqrt{\overline{N}_b} \Bigg|_{large\ k} \qquad (7)$$

and simultaneously simplify the formula for the net activity standard deviation:

$$s(N_b - \overline{N}_b) = s(N_b) = \sqrt{\overline{N}_b} \Bigg|_{large\ k} \qquad (8)$$

The null hypothesis is then restated as follows: the net signal count number value belongs to the statistical distribution of the net background count numbers, with the mean and standard deviation of this distribution being estimated from the background acquisitions only (Fig. 1a).

If the net sample value falls in the range where encountering a value from the common distribution of the net activities constructed under the null hypothesis is probable, we conclude that the null hypothesis holds true and the sample activity is not detected. At large k, this condition is usually reformulated according to the modification of the null hypothesis from the previous paragraph: If the net sample value falls in the range where encountering a net background value is probable, the null hypothesis holds true and the sample activity is not detected, as there are too many chances to obtain such activity as part of the background activity distribution (Fig. 1a). If the net sample value falls in the range of higher values (*critical range, critical region*) where the probability to encounter a net value from the common distribution constructed under the null hypothesis is below some (reasonably small) threshold set by the analyst, then the null hypothesis is rejected as contradictory and the sample activity is detected: obtaining the corresponding net sample value from the common distribution is unlikely.

The decision 'detected' vs. 'not detected' is thus probabilistic rather than absolute: it is made at a confidence level and allows for the presence of a fraction of *false detections*, when the net values from the common distribution constructed under the null hypothesis fall, albeit rarely, in the critical region (Fig. 1a). Such false detections are called *false positive errors*, or just *false positives*; syn.: *α-type errors, type I errors*. The rate of false positives is called *size*, or *significance level*, of decision rule. The actual (observed, simulated) and the declared (nominal) sizes are distinguished; they do not always coincide.

Based on the probabilistic approach above and assuming that the actual distribution of the net values of Poisson distributed count numbers is well approximated by the Gauss distribution, the critical level for detection decision can be defined as follows (Fig. 1a)[1-6]:

$$L_c = z_\alpha s(N_s - \overline{N}_b) = z_\alpha s(N) \sqrt{1 + \frac{1}{k}} = z_\alpha \sqrt{N} \sqrt{1 + \frac{1}{k}} \approx z_\alpha \sqrt{\overline{N}_b} \Big|_{large\ k} \quad (9)$$

where quantile z_a characterises the confidence level, at which the detection decision is made, and is defined through the cumulative function of the Gauss distribution:

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{z_\alpha}^{+\infty} e^{-x^2/2} dx \quad (10)$$

α denotes probability of false positive error (declared rate of false detections, declared size). Some frequently used values of z_a and the corresponding probabilities / confidence intervals are shown in Fig. 1a,b. At $N_{s\ net} > L_c$, the signal is considered as detected, at $N_{s\ net} \le L_c$, - as not detected.

Paired measurements

Replicating background acquisitions is not always practical or possible. Acquiring a large number of such acquisitions takes time and cannot be done anew before the execution of each particular sample analysis. Thus, the same 'pack' of background acquisitions has to be used to compute the critical value for several consecutive analyses. Contamination / drift issues arise. For example, fast ablation of standards or samples with moderately high trace element contents, such as the NIST 612 and especially 610 glasses, leaves elevated gas blanks slowly decreasing during at least several minutes, provided a sensitive spectrometer is used (perhaps, combined with a non-optimal design of the ablation cell). Thus, background acquisitions acquired prior to sample analyses do not always represent a good proxy to the blank level at the time of analysis of a given sample. Estimating the background prior to each particular analysis becomes advisable; in LA-ICPMS, this is done by measuring the gas blank for some time before each sample ablation. The corresponding metrological concept is called *paired measurements* to emphasize that each sample acquisition receives its own background estimate that is supposed to better characterise the actual background activity during the sample analysis.

Mathematical statistics, biometrics and metrology offer a plethora of different, partly related decision rules for paired measurements. To simplify their study, distinguishing *replicate* and *true paired measurements* is helpful. Replicate paired measurements represent an educational concept rarely used (but readily misused) in practice; learning it helps better understanding relationships between the different detection rules. In replicate paired measurements, the background (mean value and standard deviation) is well-known, defined by replication, but the net activity is calculated by subtracting a background activity estimate obtained from a single measurement (and not its 'well-known' value) from an other background activity estimate, or signal estimate, obtained during another single measurement. In true paired measurements, both background activity (its mean and standard deviation) and net activity values are estimated from one analysis made of two parts: background measurement and signal measurement. True paired measurements are not only an educational but also a practical concept with many applications in biometrics, radioactivity monitoring, chemical metrology, and in LA-ICPMS.

Let us consider a paired analysis containing a measurement of background activity (N_b) followed by the ablation of a sample to record its gross signal (N_s); for simplicity, we assume the same counting times for both background and signal. The net sample activity is calculated as follows:

$$N_{s\ net} = N_s - N_b \quad (11)$$

Besides, some number ($k \ge 1$) of paired background acquisitions are collected, of which the structure is shown in Fig. 1b. For each of them, a net value can be computed:

$$N_{b\ net_i} = N_{b_{i2}} - N_{b_{i1}} \quad (12)$$

The null hypothesis is formulated as usual: both signal and background activity values come from a common distribution (i.e., they represent measured estimates of the same mean activity). An equivalent formulation of it states that the net signal and net background activity values come from a common distribution with a mean value of zero. The alternative hypothesis states that the signal and the

background values come from two different distributions, the mean signal value being larger than the mean background value. Under the null hypothesis, the signal is not detected. If this hypothesis is rejected at a pre-set confidence level, the alternative hypothesis is accepted; the signal is detected. To construct the common statistical distribution above and to estimate its standard deviation, both signal and background values need to be pooled: they are indistinguishable under the null hypothesis (cf. eqn 4-6). For the simplest case of ordinary Poisson distributed data, this can be done as follows:

$$s(N_s - N_b) = \sqrt{s^2(N) + s^2(N)} = \sqrt{2}s(N) = \sqrt{2\overline{\overline{N}}} = \sqrt{\frac{1}{k+1}\left[\sum_{i=1}^{k}(N_{b_1} + N_{b_2})_i + N_s + N_b\right]} \qquad (13)$$

Here, $s(N_s-N_b)$ denotes the standard deviation of the common distribution of net count number under the null hypothesis, and N - the number of counts in any measurement, including both background measurements and sample measurement. Eqn (13) is applicable at any non-negative integer k. At $k=0$ (true paired measurements: only two values, N_b and N_s, are available), it degenerates as follows:

$$s(N_s - N_b) = \sqrt{\frac{N_s + N_b}{2} + \frac{N_s + N_b}{2}} = \sqrt{N_s + N_b} \qquad (14)$$

Eqn (14) is identical to eqn (6) from the previous section; this is one important case when the concepts of well-known background and paired measurements converge. The second case of convergence implies that the background is measured longer than the sample; it is considered in detail in Part II.

At large k, the weight of the net sample value in the common distribution is insignificant, and the null hypothesis is usually restated as follows: the net signal value belongs to the statistical distribution of the net background values, with the standard deviation of this distribution being estimated as follows:

$$s(N_{b_2} - N_{b_1}) = \sqrt{s^2(N_{b_2}) + s^2(N_{b_1})} = \sqrt{2}s(N_b)\Big|_{large\ k} \qquad (15)$$

For ordinary Poisson distributed count numbers, we obtain:

$$s(N_{b_2} - N_{b_1}) = \sqrt{2\overline{N}_b} = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(N_{b_1} + N_{b_2})_i}\Bigg|_{large\ k} \qquad (16)$$

The net background values form a distribution with larger 'wings' (larger standard deviation) compared to the net background activity, as it is defined in the previous section (cf. eqn 8 and 15,16).

We compare the net sample activity value to the common distribution constructed under the null hypothesis (of which the standard deviation is pooled using all available measurements; eqn 13). When the null hypothesis is modified for large k, this condition is reformulated: we compare the net sample activity obtained during a paired measurement to a model distribution of the net background activities with a standard deviation defined by eqn (15,16) (Fig. 1b). If the net sample activity value is large enough to fall in the range where the probability to encounter a net value from the common distribution is below some threshold, the null hypothesis is rejected and the signal - detected (Fig. 1b).

Assuming the Gaussian approximation of genuinely Poisson data, the critical level can be given as follows (Fig. 1b):

$$L_c = z_\alpha s(N_s - N_b) = z_\alpha \sqrt{2}s(N) = z_\alpha \sqrt{2\overline{N}} \qquad (17)$$

At large k, based on eqn (15), we obtain: $L_c = z_\alpha s(N_{b_2} - N_{b_1}) = z_\alpha \sqrt{2}s(N_b) = z_\alpha \sqrt{2\overline{N}_b}\Big|_{large\ k} \qquad (18)$

Comparing eqn (9) and (18) shows that, at large k, the detection capability deteriorates by a factor of $\sqrt{2}$, if replicate paired measurements are employed (for background and signal of equal duration). This explains why the concept of replicate paired measurements and the 'imprecise' net activity definition used in this concept have little use in practice: the amount of work to characterise the background is similar to the realisation of a 'well-known' background, whereas the detection capability is worse.

It must be noted that eqn (13,14), frequently used in scientific disciplines beyond chemical metrology[4,6-8], are somewhat at odds with the current practice of analytical chemistry. In fact, rule (18), otherwise known as '$\sqrt{2}s(N_b)$' or '$\sqrt{2N_b}$' rule, is often employed in analytical chemistry. It appears for the first time in a canonical article by H. Kaiser[12], and can be found in a number of later texts, including an influential review by L. Currie[3] and the IUPAC recommendations currently in vigour[1,2]. In the prac-

tice of LA-ICPMS, this rule has been used for twenty years now[13,14]. It, and its generalised forms, are included in data reduction packages and are traceable in almost all trace element LA-ICPMS data available at this time. The explanation is simple: it is unconsciously employed for the treatment of true paired measurements without background replication [using time-resolved background count number acquisitions or Poissonian statistic to estimate $s(N_b)$ from a single background measurement]. Then, eqn (15,16) return a negatively biased standard deviation estimate of the net activity, and eqn (18) - the correspondingly biased estimate of the critical level. An unbiased (Poissonian) estimate is given by eqn (13,14,17). At large count numbers, the difference between the above estimates is insignificant and has no harmful consequences in practice. At small count numbers, the situation is detailed in Part 2 of this text, along with a systematic discussion of decision rules[4-8,15-19] indeed capable of handling true paired measurements.

1.2. Detection limit: basic notions

Critical values are estimated relative to an arbitrarily selected rate of false positives errors that appears acceptable to the analyst, who thus consciously allows some fraction of the background acquisitions to be reported as detection of the sample signal. A complementary problem may also be considered: how often will the application of a given decision rule (e.g., eqn 9 or 17,18) result in the detection of analyte in a sample with some mean, precisely known, activity above the mean background value? The result is clearly below 100%. For example, for a mean net sample activity equal to L_c, and for a symmetrical distribution of the net activity estimates, only 50% of all replicate sample measurements fall in the critical range (Fig. 2a,b). The power of the decision rule equals 0.50 in this case. Confidently detecting the presence of analyte in a sample, provided its mean activity equals L_c, is clearly impossible. In the past, there was an animated discussion in the metrological literature regarding how to describe this situation in appropriate terms. It was concluded (and later reflected in the IUPAC recommendations)[1-3] that, to minimise controversy, *detection limit* must be introduced as a separate concept of chemical metrology not synonymous with critical value, although related to it. It is defined as a mean, precisely known sample activity that ensures some arbitrarily set power (probability, e.g., 95%) of analyte detection in the sample using a given decision rule. Still, a fraction of replicate sample measurements will fall below the critical level (Fig. 2a,b). Such false non-detections are called *false negative errors*, or just *false negatives*; syn.: *β-type errors*, *type II errors*.

Distinguishing when each of the complementary concepts, critical level and detection limit, applies is quite important. If we perform a sample measurement, we compare the net sample signal to the critical level. If the net sample signal is above the critical level, it is detected. Otherwise, it can be labelled as not detected (*n.d.*) or below detection (*b.d.*). Detection limit does not even need to be calculated to take the detection decision. Its area of application is different. It is most frequently used to characterise the performance of an instrument or method against requirements set by a particular analytical task or customer: a specific power of detection, usually set to be high (e.g., 95%), is ensured for a sample, of which the mean, precisely known signal is equal to the detection limit. If detection of such signal with a specified probability, and detection of stronger signals with even larger probability, meets the customer's requirement, the method is approved[1,2]. Besides, detection limits are used to compare two analytical methods[1,2], although critical levels can also be applied for this purpose.

Applications of the detection limit concept can be illustrated by the following example. Let us consider a set of samples. We need to identify those of these samples that contain more than *x* activity units (*x* being a threshold for the maximum admissible level of contamination, or of radiation exposure, or a minimum commercially interesting analyte content, etc.). If, for the analyte in question, our analytical method ensures a detection limit equal to *x*, then the probability to miss a potentially important sample because it is labelled 'below detection' - and its activity is likely to come from the distribution of the background fluctuations - corresponds to the rate of false negative errors associated with the detection limit formulation. This probability is usually set to be small (e.g., 5%). Interestingly, if the presence of analyte is detected, then we still cannot tell for sure if its 'true', precisely known activity or abundance in the sample is above *x*; the only result ensured at this stage is that this sample will not be lost for further study aimed to quantify *x* more precisely (e.g., by replication).

Undoubtedly, there are applications that require samples to be checked against a threshold, as in the example above. Environmental pollution control and quality monitoring in industry are among them. In LA-ICPMS, such applications are still rare, as are studies illustrating a practical need for the quantification of detection limits. For years, detection limits in the field of LA-ICPMS were defined, calculated and applied, as they would be critical levels, to decide if an individual sample signal is detected[13]. This is an odd testament to the limited practical use of the detection limit concept in the range of established LA-ICPMS applications. Besides, estimating the critical level is a pre-requisite for computing the detection limit[1-3]. In the following sections of this review, we mainly focus on the different rules of the critical level estimation; computing the corresponding detection limits for the most basic decision rules is discussed in Fig. 2, further explanations can be found in the literature[1-4,6,9].

A note of caution is warranted. In no case can detection limit be compared to the outcome of a single sample measurement to infer if the sample signal is detected: this is a misuse of the detection limit concept that is inconsistent with the definition of detection limit. It may only be added in the case of non-detection that the precise value of analyte activity in the sample is below the detection limit at a given confidence level, if the measured estimate of this value is below the critical level (Fig. 2a,b).

An other note of caution addresses the common practice of reporting measurement results below the critical level as 'below detection' without providing a measured value of the net sample activity (even if it happened to be below zero). In this case, the precise net sample activity value is not necessarily zero. It lies between zero and the detection limit at a confidence level associated with the latter: obtaining an activity estimate below the critical level for a precise sample activity value equal or superior to the detection limit is less than $\beta\%$ likely. The corresponding estimate can be reported[1,2,9]; especially, it must always be reported if averaging several replicate net sample activity values is planned; otherwise, the resulting mean is positively biased[1,2,9] (Fig. 3). Reporting it can be done, for example, as follows: $x^{b.d.}$, where x is an activity (content) estimate that lies below the critical level. Alternatively, colour marking can be used, especially for the LA-ICPMS data reduction software output.

1.3. Statistical distributions, standard deviations, and ICPMS data

The critical level, as well as the detection limit, is defined at a confidence level. Hence, knowing the probability function(s) describing the count number distribution in LA-ICPMS is essential for the construction and performance evaluation of decision (detection) rules. As many of such rules rely on the link between the net activity standard deviation and the associated confidence interval, estimating the standard deviation for the relevant probability distributions should also be considered. The genetic and possibly the most general approach to ICPMS data from the probabilistic standpoint is to interpret them as a doubly stochastic Poisson distribution[20]. In applications, the Poisson distribution usually arises as a special form of the binomial distribution, provided a large number of 'candidates' are exposed to a low probability to realise a particular event; the number of 'realisations' is then Poisson distributed[10,11,20-22]. For example, a number of ions are formed in the ICP per observation interval, with a low probability for each of them to be detected because of low ion extraction and transmission efficiency of the ICP mass spectrometer. The numbers of ions actually counted per observation are then Poisson distributed[20]. For acquisitions containing very few counts, sources of emission other than the ICP, including the skimmer cone, extraction lens, and the first dynode(s) of the secondary electron multiplier, can be added. In all cases, the Poisson argument appears to apply: e.g., a large number of ions and electrons are amenable to thermionic and field emission from the surface of the first dynode(s) of a secondary electron multiplier, but the probability for a particular ion (electron) to be emitted is low, leading to the appearance of a Poisson distributed detector noise. If the supply of ions to the spectrometer and the ion transmission efficiency is stable, the resulting distribution can be called ordinary Poisson distribution. If ions are extracted from a fluctuating (flickering) source doubly stochastic Poisson distribution applies[20]. For the doubly stochastic Poisson distribution, at a constant ion transmission, the count number standard deviation can be given as follows[20,21]:

$$s(N) = \sqrt{\overline{N} + p^2 s^2(M)} \qquad (19)$$

where $\overline{N}$ denotes the mean count number, p - ion transmission efficiency and M - number of ions exposed to the spectrometer interface per time interval. The term $p^2 s^2(M)$ is called *excess variance*.

The doubly stochastic model is quite useful, as far as the fundamental origin of ICPMS uncertainties is discussed[20]. However, it is less useful for the practical estimation of the standard deviation of a particular measurement, as parameters p and M are rarely known. Fortunately, LA-ICPMS background acquisitions can often be treated using the much simpler apparatus of the ordinary Poisson distribution. For an ordinary Poisson distribution, the term $p^2 s^2(M)$ in eqn (19) is negligible (e.g., when both ion source and ion transmission efficiency fluctuate little compared to the Poisson scattering related to the 'binomial selection' of ions[20,21]). Omitting this term allows computing the standard deviation as the square root of the mean number of counts, the latter often being estimated from a single measurement:

$$s(N) = \sqrt{\overline{N}} \; ; \; s(N) = \sqrt{N} \qquad (20)$$

Such estimates are very precise for all but very small mean count numbers:

$$s(s(N)) = s(\sqrt{N}) = \frac{\partial(\sqrt{N})}{\partial N} s(N) = \frac{1}{2\sqrt{N}} \sqrt{N} \bigg|_{N=\overline{N}} = 1/2 \qquad (21)$$

In ICPMS, the ordinary Poisson distribution appears in weak backgrounds and signals: according to experimental data and modelling, the term $\overline{N}$ in eqn (19) decreases linearly with signal strength, while the term $p^2 s^2(M)$ decreases in quadrature, vanishing at small counts[20,23-27]. The latter case is practically important, as weak backgrounds ranging from zero to a few counts are typical of many isotopes (e.g., lanthanides, Th, U) analysed during routine multi-elemental LA-ICPMS work. For an ordinary Poisson distribution, it is easy to give an analytical expression for the probability function[10,11,21,22]:

$$p(N) = \overline{N}^N e^{-\overline{N}} / N! \qquad (22)$$

where N is the outcome of a particular acquisition in a series of replicates with a mean of $\overline{N}$ counts.

The ordinary Poisson distribution has two important properties[10,11,21]: (i) in a replicate series of acquisitions, the individual outcomes are independent (uncorrelated, show no covariance); (ii) it is possible to reduce the duration of acquisitions so that each of them will contain 0 or 1, but no more than 1 count (simultaneous counts are impossible). Using these properties, it can be instructive to derive eqn (20) from the general, distribution-independent definition of standard deviation. At a level of rigour required for the present text, this could be done as follows. We assume that each of the acquisitions above represents a sweep in a series of sweeps constituting time resolved ICPMS background noise. According to the general definition of standard deviation, we have:

$$s(N_{sw}) = \sqrt{E(N_{sw}^2) - [E(N_{sw})]^2} \qquad (23)$$

where $E(N_{sw}^2)$ and $E(N_{sw})$ are mathematical expectations defined as follows:

$$E(N_{sw}^2) = 1^2 \times p(1) + 2^2 \times p(2) + ... \; ; \; E(N_{sw}) = 1 \times p(1) + 2 \times p(2) + ... \qquad (24)$$

In other words, for discrete data, mathematical expectation is defined as a sum of products of the value of each outcome, e.g., 1,2,…, and probability p to obtain such outcome. For a particular statistical sample (measurement), the definition of mathematical expectation evolves into the definition of arithmetic mean. For a measurement made of k sweeps, we have:

$$s(N_{sw}) = \sqrt{\frac{1}{k-1}\left[\sum_{i=1}^{k} N_{sw\,i}^2 - k\left(\overline{N}_{sw}\right)^2\right]} \text{ or, equivalently, } s(N_{sw}) = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}\left(N_{sw\,i} - \overline{N}_{sw}\right)^2} \qquad (25)$$

where $N_{sw\,i}$ denotes the number of counts in an individual sweep from this measurement, $\overline{N}_{sw}$ - the mean count number per sweep, and $s(N_{sw})$ - standard deviation for the individual sweep count numbers. Once again, eqn (23,25) are valid for any statistical distribution.

In the absence of correlation, the total count number variance can be represented as a sum of the individual sweep variances, without any specific requirements regarding the sweep duration; reducing the latter and simultaneously increasing the number of sweeps does not change the total variance:

$$N_b = \sum_{i=1}^{k} N_{sw\,i};\ Var(N_b) = Var\left(\sum_{i=1}^{k} N_{sw\,i}\right) = Var(N_{sw\,1}) + ... + Var(N_{sw\,k}) = kVar(N_{sw})\Big|_{at\ any\ k} \tag{26}$$

$$s(N_b) = \sqrt{k}\,s(N_{sw})$$

where N_b denotes the total number of counts in a background measurement, and $s(N_b)$ denotes its standard deviation. The remaining parameters are defined as in eqn (25).

Combining eqn (23,26), we obtain: $s(N_b) = \sqrt{k\left(E(N_{sw}^2) - \left[E(N_{sw})\right]^2\right)}$ (27)

Combining eqn (25,26) leads to a similar result:

$$s(N_b) = \sqrt{\frac{k}{(k-1)}\left[\sum_{i=1}^{k} N_{sw\,i}^2 - k\left(\overline{N}_{sw}\right)^2\right]}\ \text{or, equivalently,}\ s(N_b) = \sqrt{\frac{k}{(k-1)}\sum_{i=1}^{k}\left(N_{sw\,i} - \overline{N}_{sw}\right)^2} \tag{28}$$

Eqn (27,28) still show a large degree of distribution independence and are valid for any statistics of *uncorrelated* count numbers (confidence intervals assigned to the $\pm s(N_b)$ range may, however, depend on the distribution, see the next sections). Based on property (ii) of the ordinary Poisson distribution, eqn (27,28) can be very much simplified. Let us cut the whole acquisition in very short segments ('sweeps'). With increasing k, it is possible to make the segment duration sufficiently short for each of these segments to contain either 0 or 1, but no more than 1 count:

$$s(N_b) = \sqrt{k\left(E(N_{sw}^2) - \left[E(N_{sw})\right]^2\right)} = \sqrt{k\left[p(0)\times 0^2 + p(1)\times 1^2\right] - k\left[p(0)\times 0 + p(1)\times 1\right]^2}$$

$$= \sqrt{k\left[\frac{\overline{n}(1)}{k}\right] - k\left[\frac{\overline{n}(1)}{k}\right]^2} = \sqrt{\overline{n}(1) - \frac{\left[\overline{n}(1)\right]^2}{k}} \approx \sqrt{\overline{n}(1)} = \sqrt{\overline{N}_b}\Big|_{k\ very\ large} \tag{29}$$

where $n(1)$ is the number of unity outcomes.

For a statistical sample (such as an individual measurement), we obtain the following estimate:

$$s(N_b) = \sqrt{\frac{k}{(k-1)}\left[\sum_{i=1}^{k} N_{sw\,i}^2 - k\left(\overline{N}_{sw}\right)^2\right]} \approx \sqrt{\sum_{i=1}^{k} N_{sw\,i}^2 - k\left(\overline{N}_{sw}\right)^2}$$

$$\approx \sqrt{\left[0^2 + 1^2 + ... + 0^2 + 1^2 + 0^2 + ...\right] - N_b^2/k} = \sqrt{N_b - N_b^2/k} \approx \sqrt{N_b}\Big|_{k\ very\ large} \tag{30}$$

The above derivations are unidirectional: they follow from eqn (23,25) as a special case, at the condition that the actually observed count numbers follow an ordinary Poisson distribution. Either of eqn (20,28-30) can be used to estimate the standard deviation, provided this condition is filled.

The above equations can be readily presented in the intensity-based notation:

$$s(I_b) = s\left(\frac{N_b}{t_b}\right) = \sqrt{\frac{1}{k(k-1)}\sum_{i=1}^{k}\left(\frac{N_{sw\,i}}{t_b/k} - \frac{\overline{N}_{sw}}{t_b/k}\right)^2} = \sqrt{\frac{1}{k(k-1)}\sum_{i=1}^{k}\left(I_{sw\,i} - I_b\right)^2} \tag{31}$$

$$s(I_b) = \frac{s(N_b)}{t_b} = \frac{\sqrt{\overline{N}_b}}{t_b} \approx \frac{\sqrt{N_b}}{t_b} = \sqrt{\frac{I_b}{t_b}} \tag{32}$$

where I_b denotes mean intensity pooled from all sweeps constituting a background acquisition, $I_{sw\,i}$ - intensity of an individual sweep number i, and t_b - total duration of the background measurement.

In strong signals and backgrounds with a large contribution of the flicker noise, the term $p^2 s^2(M)$ in eqn (19) cannot be neglected. The $s(N)=\sqrt{N}$ dependence, applicable for ordinary Poisson data, does not hold anymore [short, closely spaced in time, consecutively acquired sweep count numbers begin to correlate[20,28-32]: if the previous sweep count number is high, then the actual sweep count number is also likely to be high, reflecting the fluctuation pattern of the ion source and resulting in a violation of eqn (26)]. Strong LA-ICPMS backgrounds rarely occur in practice, [39]K and [28-30]Si isotopes being the most

common examples. To estimate the standard deviation for the individual sweep count number for such isotopes, eqn (25) is used. To estimate the standard deviation of the total number of counts per measurement, eqn (28) is a good proxy, provided a multi-elemental acquisition protocol is employed. In such protocols, in the single-collector regime, the sweep-to-sweep acquisitions of the same isotope are well spaced in time; their outcomes show little correlation that is usually neglected[13,33-35]. Respectively, to estimate the standard deviation in the intensity-based notation, eqn (31) can be employed.

It follows that using eqn (28,31) is the most general solution for the computation of the standard deviation of LA-ICPMS data. Moreover, for years eqn (31) has been the only mathematical solution for the estimation of standard deviations in LA-ICPMS detection calculations[13]. At the same time, many decision rules are easier to formulate based on the ordinary Poisson distribution (eqn 20,29,30,32).

1.4. Weak and strong backgrounds

The terms 'weak / strong' and 'small / large' are subjective. The simplest, albeit *qualitative*, way to verify if a background is weak enough to be described by the ordinary Poisson distribution is to check if the standard deviation values computed using eqn (20) and (28) (or eqn 31 and 32) are approximately equal to each other. If eqn (20) clearly underestimates the standard deviation compared with eqn (28), then the ordinary Poisson model is not applicable; the doubly stochastic Poisson / Gauss model applies. To *quantitatively* study any particular time-resolved LA-ICPMS measurement for a suitable model distribution, a goodness-of-fit test can be used (for examples, see Electronic Appendix 1). Such tests confirm that LA-ICPMS backgrounds <500-1000 cps are well described by the ordinary Poisson distribution. The same often applies to higher backgrounds (and signals) containing thousands and even tens of thousands cps[33] (Electronic Appendix 1); as the flickering strength can be instrument dependent, such data are still better to subject to a goodness-of-fit test before eqn (20) is employed for the estimation of standard deviations; alternatively, eqn (28, 31) can be employed as a more universal, distribution-independent solution (in the absence of correlation between the different sweeps).

1.5. About the Poisson-Gauss approximation of confidence intervals

A standard deviation is not a self-contained value; it is rather a pre-requisite for the estimation of confidence intervals, on which decision and detection rules are built. In practice, well-known Gaussian confidence intervals are used (e.g., eqn 9,18). However, the count number distribution observed in ICPMS is not the continuous symmetrical Gauss distribution but the discrete Poisson distribution, doubly stochastic at large and ~ ordinary - at small count numbers (Fig. 4a). In the latter case, it is not only discrete but also skewed. A standard deviation value can be obtained following equations from the previous section. Is a Gauss confidence interval assigned to this value a good proxy for the Poisson statistics? Fortunately, at large counts, both doubly stochastic and ordinary Poisson distributions are well approximated by the Gauss distribution and its confidence intervals. This can be considered as a consequence of the central limit theorem; for the ordinary Poisson distribution, a direct proof (by deriving the Gauss probability density function from the binomial probability function) is available[10,11,36].

The Gaussian description of confidence intervals is simple to use (e.g., eqn 9,18). Thus, at small count numbers, it is still common to see Poisson distributed data approximated by the Gauss distribution with a mean equal to the mean count number in the Poisson probability function (eqn 22) and standard deviation defined as its square root (eqn 20,29,30) or obtained from the statistics of the individual sweeps (eqn 28). It is assumed that the Gauss confidence intervals for the standard deviation are applicable to the actually observed Poison count number distribution. At small count numbers, this assumption is questionable: the corresponding Poisson statistics may contain 3-4 non-negligible probability classes, while the Gauss distribution remains to be continuous (Fig. 4a). Besides, at mean count numbers ≤1, when the probability to obtain zero counts from a particular measurement is high, symmetry disappears from the Poisson distribution pattern (Fig. 4a; admittedly, the skewness mostly disturbs the left-hand part of the distribution that is less essential for the computation of critical levels). The approximation quality can be slightly improved if the confidence interval is corrected for discontinuity (Fig. 5). For example, the critical level in eqn (9) (well-known background) is corrected as follows[10,11]:

$$L_c = 1/2 + z_\alpha s(N_b) = 1/2 + z_\alpha \sqrt{N_b} \qquad (33)$$

At a declared rate of false positives α=5% ($z_{5\%}$ =1.645), this equation can be used at $\overline{N}_b$ as small as ~2-5 counts (Table 1, Fig. 4a; further illustrations can be found in Ref. 36). Still, neither discreteness nor skewedness of the original distribution are not removed, and the actual rate of false positive errors in the range of small counts will somewhat differ from the declared rate - a problem that will be extensively studied in Part II of this text. To quantify this difference and to simultaneously estimate the critical level without using the Gaussian approximation, an exact analytical solution based on using the probability function of the ordinary Poisson distribution can be employed (Table 1):

$$L_c = N - \overline{N}_b \Big| \ p(N+1) + p(N+2) + ... \leq \alpha \ and \ p(N) + p(N+1) + p(N+2) + ... > \alpha \quad (34)$$

where the probabilities $p(N)$, $p(N+1)$, $p(N+2)$, *etc*. are obtained from formula (22). This solution can be extended to large count numbers, provided the computation is computer-aided.

The Poisson-Gauss approximation is also applicable to paired measurements. Moreover, estimating the critical level in all LA-ICPMS data obtained since the introduction of this technique and until now has been based solely on this approach[13,14]. The net count number distribution obtained by subtracting two random Poisson values (estimates) represents a derivative of the Poisson distribution that is sometimes called Skellam distribution (Fig. 4b). It is symmetrical around zero, provided the count number estimates are derived from the replicate measurement of the same Poisson process and correspond to the same mean, as it is the case under the null hypothesis. The problem of skewness, inherent to the Poisson distribution of individual small counts, disappears. However, the problem of discreteness does not disappear. Besides, in the case of true paired measurements, the net distribution under the null hypothesis is not constrained precisely, leaving some liberty regarding how to estimate its standard deviation (see Part 1, section 1.1 'Paired measurements', and Part 2). Construction and performance evaluation of the Poisson-Gauss (Skellam-Gauss) approximations and related decision rules for true paired measurements is a well-developed area of statistical, as discussed in Part 2 of this text. Some of the exact analytical solutions not involving the Poisson-Gauss approximation are also presented there.

Part II. Critical levels for paired measurements - the different decision rules, their performance

2.1. The $\sqrt{2N_b}$ rule and its size

Principles of size evaluation; the $\sqrt{2N_b}$ rule as an example

The most basic of all decision rules for paired measurements is the $\sqrt{2N_b}$ rule (eqn 18) and a generalisation of it suitable for unequal sample / background counting times:

$$L_c = z_\alpha \sqrt{2} s(N_b) = z_\alpha \sqrt{2\overline{N}_b} \quad (35)$$

In practice (section 1.1 'Paired measurements'), the count number standard deviation under the null hypothesis is usually estimated from one single background measurement constituting the first part of a paired LA-ICPMS analysis:

$$L_c = z_\alpha \sqrt{2N_b} \quad (36)$$

Let us test the actual size of rules (35,36). Such tests are based on either of two independent approaches: Monte-Carlo simulation or exact combinatorial calculation[5-8,19]. The first approach requires generating a large set of Poisson distributed estimates for a given mean background count number value (for a given mean value characterising the common distribution constructed under the null hypothesis, see section 1.1). These estimates model the background count number distribution that could be obtained if we would replicate the background measurement many times. From this set, two estimates are randomly selected. The first of them is stored as a background estimate to compute L_c using the decision rule tested. The second is stored as a sample signal, though it is also a background estimate. The net value of the two estimates is computed and compared to L_c. Detecting such a 'sample signal' means committing a false positive error (false detection). The process is replicated many times; in the end, it returns two numbers - number of false detections and number of non-detections. The fraction of false detections, as obtained during the simulation, is compared to the rate of false positives declared for the decision rule tested (e.g., α in eqn 35,36). The simulation result represents an unbiased estimate of the real rate of false positive detections for this decision rule. It is somewhat noisy, however, depending on the number of the net values simulated: the larger it is, the weaker is the noise. Practical measure-

ments are not required, nor a knowledge of probability theory. A modification of this approach exists (but has little practical use): instead of a random number generator, a spectrometer is employed to collect a large set of background estimates; the remaining part of the procedure suffers no alteration.

The second, combinatorial approach requires computing the probability of obtaining each particular background estimate for a given mean background activity. For each of such estimates, the critical level (L_c) is computed. For each of the L_c values, the probability is calculated to obtain a paired background estimate that is sufficiently large for detection according to the decision rule and for the mean background activity value that are tested. Upon summation of such probabilities for all L_c values, the total actual rate of false positives is obtained. The outcome is consistent with Monte-Carlo simulation results, but devoid of noise[5,6]. It is compared to the rate (α) declared based on the Poisson-Gauss approximation. Henceforth, we will use the combinatorial approach only, reserving the Monte-Carlo simulation to the area of its excellence - problems where no analytical solution can easily be found.

An application example of the combinatorial approach is given in Table 2. Let us first discuss the case when the mean background count number and thus the count number standard deviation under the null hypothesis are estimated with a good precision [e.g., using replicate paired measurements, as shown in Fig. 1b; eqn (35) applies]. Only the quality of the Gauss approximation of discrete Poisson data remains to be tested. We will set the mean background count value to 1.5 counts - a rather common level in LA-ICPMS background measurements. For a mean of 1.5 counts, the probabilities to obtain $N_b=0$, 1, 2, etc. counts in a particular background measurement are calculated using the probability function of the Poisson distribution (eqn 22). For example, for $N_b=1$, $p(1)=0.335$ (Table 2). At a commonly used confidence level of 95% corresponding to a rate of false positive errors of 5%, $z_a=1.645$ (assuming the Poisson-Gauss approximation). L_c amounts to $1.645 \times \sqrt{2} \times 1.5 = 2.849$ counts, irrespective of the outcome of a particular background measurement (Table 2). At $N_b=1$ and $L_c=2.849$, the probability of detection equals the cumulative probability to obtain 4 and more counts for a mean of 1.5 (to obtain a net value > 2.849). Hence, the fraction of false detections for $N_b=1$ is as follows: $p(1) \times [p(4)+p(5)+p(6)+...]=0.022$. For a mean of 1.5, the probability to obtain an N_b value >8 is very low; for simplicity, the corresponding probabilities are omitted in our summation. The exercise is repeated for all N_b values with non-negligible probabilities, yielding a total actual false detection rate of 6.98% (Table 2). This differs from 5% declared based on the Poisson-Gauss approximation.

In practice, the background count number is almost never known precisely, but estimated from one single paired measurement (background and sample signal). An intelligent estimation of the standard deviation under the null hypothesis is required. In the $\sqrt{2N_b}$ decision rule, it is estimated using the background count number only [eqn (36) applies]. The combined effect of the quality of the Poisson-Gauss approximation and of the choice of the standard deviation estimate is now to be tested. The mean background activity value will be set at 1.5 counts, as before. For this mean, the probabilities to obtain $N_b=0$, 1, 2, etc. counts in a particular background measurement are calculated (Table 2); this part of the exercise is identical to the case of replicate paired measurements discussed in the previous paragraph. However, the corresponding L_c values need to be computed for each of the N_b values specifically. For example, for $N_b=2$, L_c equals $1.645 \times \sqrt{2} \times 2 = 3.290$. At $N_b=2$ and $L_c=3.290$, the probability of detection equals the cumulative probability to obtain 6 and more counts for a mean of 1.5 (to obtain a net value > 3.290). Hence, the fraction of false detections for $N_b=2$ is as follows: $p(2) \times [p(6)+p(7)+p(8)+...]=0.0011$. Repeating the exercise for other N_b values yields a total actual false detection rate of 19.64% (Table 2). This is much higher compared to the declared level of 5%.

It can be argued that most false detections according to the $\sqrt{2N_b}$ decision rule happen when the first background acquisition yields zero counts[5,39,40]. Then, any positive count number contained in the second acquisition results in detection. ICPMS analysts often attempt to fix this issue by adding one count to an 'empty' background, an arbitrary practice that seems to lack theoretical grounds (cf. Ref. 5, decision rule 9 and its explanation). Such practice also allows for a combinatorial test; it results in a decrease of the rate of false detections, but still fails to bring it down to the declared level (Table 2).

Calculations presented in Table 2 can be repeated to tabulate the actual rate of false detections for a range of mean background count numbers and confidence levels (Fig. 6); such tabulations for several decision rules can be found in the literature[5-8] and in the subsequent sections of this text (representative examples of the source code are available in Electronic Appendix 2 in the form of *excel vba* mac-

ros). At a 5% declared level of false detections, their actual rate for eqn (36) reaches a maximum of 25.2% at a mean of 0.72 counts per background acquisition (Fig. 6)! This decision rule returns elevated rates of false detections, approaching the declared level only asymptotically, at large count numbers (Fig. 6). The actual size of this rule is thus larger than declared (except for extremely small mean count numbers limiting to zero, where the actual size limits to 1%; for a proof, see Ref. 5). Decision rules, of which the actual size is larger than declared, are called *liberal*[7]. Those with an actual size smaller than declared are called *conservative* (they yield less false detection than declared). Rules, of which the actual size falls in the range of ~ ±20% of the declared value (in the studied range of mean count numbers), are called *robust*[7]. For the practical estimation of critical levels, conservative or (if possible) robust rules are preferred[5-8,37,38]. If a liberal decision rule is still used for whatever reason, it is quite essential to know the extent of its liberality (i.e., its actual size).

Randomisation of the $\sqrt{2N_b}$ rule to eliminate skewness and to adjust its actual size

Testing a Poisson-Gauss decision rule, such as the $\sqrt{2N_b}$ rule, for size is equivalent to considering how well the corresponding statistics, e.g.,

$$(N_s - N_b)\Big/\sqrt{2N_b} \qquad (37)$$

approximate standard normal variates, i.e. how close they are to a Gauss distribution centered around zero with a standard deviation of one[4-8,10,15-19,40]. Under the null hypothesis, it is easy to show that statistics (37) indeed have a standard deviation of one (see Electronic Appendix 3). At the same time, ratios of two approximately Gauss distributed variables, such as the numerator and the denominator in the above statistics, are not necessarily Gauss distributed. Statistics (37) are asymmetrically distributed (skewed) relative to zero, resulting in a deviation from the Gauss distribution that is more marked at smaller N_b values[40]. The skewness is due to a negative correlation between the numerator N_s-N_b and denominator $\sqrt{2N_b}$; the corresponding distribution pattern is schematically shown in Fig. 7. The positive 'tail' of this pattern corresponds to the area where false positive errors occur, partly explaining an increase in their rate compared to the prediction from the Gauss model. Other statistics can be suggested that are devoid of this shortcoming; they are described in the next section of this text.

The recognition of the inadequate performance of the $\sqrt{2N_b}$ rule for the treatment of true paired measurements during the last fifteen years[5,6,40] motivated attempts to 'randomise' statistic (37) in order to break the correlation pattern in it[40]:

$$(N_s - N_b')\Big/\sqrt{2N_b''} \qquad (38)$$

It was suggested to compute the numerator and the denominator using two independent background count number estimates (N_b' and N_b'', respectively)[40]. This improves the match between the declared and actually observed rates of false positives, although the liberality of the resulting 'randomised' $\sqrt{N_b}$ decision rule at small count numbers is still significant (Fig. 6). At the same time, an inconsistency was created. Instead of carrying out a paired measurement, a triple measurement needs to be performed to acquire N_b', N_b'' and N_s; two different, equally eligible net sample signal count numbers and critical values can then be obtained:

$$N_{s\,net} = N_s - N_b'; \ \ L_c = z_\alpha\sqrt{2N_b''} \ \ \text{and} \ \ N_{s\,net} = N_s - N_b''; \ \ L_c = z_\alpha\sqrt{2N_b'} \qquad (39)$$

In both cases, the standard deviation of the background activity under the null hypothesis is estimated with a bias; the maximum likelihood value is $(N_b' + N_b'' + N_s)/3$ (see sections 1.1 and 2.2). We conclude that randomising the $\sqrt{2N_b}$ decision rule does not improve its performance sufficiently, but creates undesirable side effects. The skewness of statistics (37) represents a reason to avoid using the $\sqrt{2N_b}$ rule[5,6] instead of attempting to repair this rule[40].

2.2. Unbiased rules for true paired measurements based on the Poisson-Gauss approximation

The $\sqrt{N_s + N_b}$ rule

Decision rules based on the Poisson-Gauss approximation require (i) a good convergence of the net count number distribution to the Gauss distribution in the observed range of counts and (ii) an unbiased estimation of the standard deviation of net count numbers. At small count numbers, the $\sqrt{2N_b}$ rule

does not meet these requirements. Even if the background activity is estimated with a good precision (e.g., by replicate paired measurements), the approximation quality is fair only for backgrounds containing ~20 counts and more (Fig. 6). Mathematical transforms exist to improve the approximation quality; they will be discussed below. The $\sqrt{2N_b}$ rule does not invoke such transforms. The performance of the $\sqrt{2N_b}$ rule deteriorates further, if the background activity is estimated from one single paired analysis (Fig. 6): using the background count number only, as in the $\sqrt{2N_b}$ rule, returns a biased estimate of the net count number standard deviation. This question was already considered in section 1.1; here, we provide a recapitulation. As a null hypothesis, we state that both background and sample signal count numbers come from the same Poisson distribution (with the same mean $\overline{N_s} = \overline{N_b}$, i.e., $\overline{N_s - N_b} = 0$). The alternative hypothesis is $\overline{N_s} > \overline{N_b}$. We first assume that the null hypothesis is true. The distribution of the net values $N_s - N_b$ is then centred around zero, with a standard deviation

$$s(N_s - N_b) = \sqrt{s^2(N_s) + s^2(N_b)} = \sqrt{\overline{N_s} + \overline{N_b}} = \left[= \sqrt{2\overline{N_s}} = \sqrt{2\overline{N_b}} \right] \quad (40)$$

Under the null hypothesis, the most probable (best estimated[4], maximum likelihood[7,8]) value for the variances (mean count numbers) under the radicals of this equation is $(N_s+N_b)/2$; it is pooled using both parts of a paired measurement, as neither the background nor the signal part has a 'priority', and both of them are used to estimate the standard deviation of their common distribution (see also section 1.1). In other words, under the null hypothesis, each of the values $s^2(N_s)$ and $s^2(N_b)$ is best estimated as $(N_s+N_b)/2$. We have:

$$s(N_s - N_b) = \sqrt{N_s + N_b} \quad (41)$$

Using the Poisson-Gauss approximation, we compare the measured difference $N_s - N_b$ to the estimated standard deviation of the net distribution (eqn 40,41). If the net difference is too large (Fig. 8), i.e.

$$N_s - N_b > z_\alpha \sqrt{N_s + N_b} \quad (42)$$

then there is little chance for this difference to come from an $N_s - N_b$ distribution centred around zero. This means that $\overline{N_s - N_b} > 0$, or $\overline{N_s} > \overline{N_b}$. The null hypothesis is rejected; the sample signal is detected at a nominal rate of false positives equal to α. Otherwise, the null hypothesis is accepted: the sample signal is not detected. The divide between these alternatives corresponds to the critical level:

$$L_c = N_s - N_b = z_\alpha \sqrt{N_s + N_b} \quad (43)$$

Eqn (41) underlying the above derivations can also be readily obtained from the general formulations of both 'well-known background' and 'replicate paired measurements' detection concepts [see section 1.1, eqn (5,6) and (13,14)], which highlights the close relationships existing between the different approaches to detection that are based on the equivalent formulations of the null hypothesis.

The difference compared to the $\sqrt{2N_b}$ decision rule lies in the estimation of the standard deviation of the net count number: in the case of true paired measurements, without background replication, its unbiased value is $\sqrt{N_s+N_b}$ and not $\sqrt{2N_b}$. Noteworthy, N_s-N_b and N_s+N_b values do not correlate under the null hypothesis, ensuring the symmetry of the normalised net statistics. The $\sqrt{2N_b}$ decision rule is not the only biased rule possible. Indeed, the unbiased estimator of the mean activity of the common distribution of the background and signal count numbers under the null hypothesis is unique under the condition of equal background and signal counting times and amounts to $(N_s+N_b)/2$. In contrast, the number of biased estimators is large - from N_b to N_s. From this standpoint, rule $N_s-N_b > z_a\sqrt{2N_b}$ is equally eligible as rule $N_s-N_b > z_a\sqrt{2N_s}$. The latter rule has to be conservative. A curious reader can model its size - an exercise beyond the scope of this text.

By solving a quadratic equation, it is possible to exclude N_s from the right-hand part of eqn (43) and to obtain a maximum N_s-N_b value at which the signal is not yet detected as a function of N_b and z_a only. This value corresponds to L_c:

$$N_s - N_b = z_\alpha \sqrt{N_s + N_b}$$

$$\Rightarrow N_s = N_b + \frac{z_\alpha^2}{2} + z_\alpha \sqrt{\frac{z_\alpha^2}{4} + 2N_b}; \; L_c = \frac{z_\alpha^2}{2} + z_\alpha \sqrt{\frac{z_\alpha^2}{4} + 2N_b} \quad (44)$$

Decision rule (42-44) and similarly constructed decision rules with an improved convergence to normality are often discussed in texts of mathematical statistics[7,8,10,11,15,19,36], while the $\sqrt{2N_b}$ rule does not appear in such texts because of the biased estimation of the mean background count number. The $\sqrt{2N_b}$ rule, by its derivation from replicate paired measurements and by its application field, is restricted to those very rare situations when the background is replicated and its mean value and uncertainty are well known, but the detection decision is taken based on the concept of paired measurements [Fig. 1b; an other, more positive way of thinking about this rule is to consider it as a starting point to obtain eqn (44); the relevant derivation is not devoid of mathematical elegance, see Electronic Appendix 4].

Decision rule (42-44) has been known for at least 75 years now[10,11,15]. To the analytical community, it was introduced by B. Altshuler and B. Pasternack[4] and W. Nicholson[16,17], their texts representing classical works on the mathematical principles of detection. Beyond the field of radioactivity monitoring[4-6,16,17], in chemical metrology, decision rule (42-44) still has a rather limited use. Size calculations show that using this rule at small count numbers quite significantly improves the match between the actual and the declared rate of false detections, especially in the framework of equal counting times (Fig. 9a).

Decision rule (42-44) can be corrected for discontinuity[10,11,15]: $L_c = 1 + z_\alpha \sqrt{N_s + N_b}$ (45)

Compared to the case of well-known background (eqn 33), the correction term increases from ½ to 1[10,11]; this is because the net count number for a given value of L_c has the same parity as the $N_s + N_b$ value (both odd or both even); thus, the discreteness in probability has a step of 2 (not 1, as in Fig. 5). Eqn (45) can be further modified by excluding the signal count number (cf. eqn 44):

$$N_s - N_b = 1 + z_\alpha \sqrt{N_s + N_b} \Rightarrow N_s = N_b + 1 + \frac{z_\alpha^2}{2} + z_\alpha \sqrt{\frac{z_\alpha^2}{4} + 2N_b + 1}$$

(46)

$$L_c = N_s - N_b = 1 + \frac{z_\alpha^2}{2} + z_\alpha \sqrt{\frac{z_\alpha^2}{4} + 2N_b + 1}$$

The discontinuity-corrected formulation above is conservative (Fig. 9b). It is mostly discussed in classical texts of statistics[10,11] and has little practical use in metrology at this time. It can be used as a simpler substitute to the canonical version of the binomial decision rule (cf. Fig. 9b and f, see section 2.3).

Decision rules using square root variance stabilising count number transformations

Decision rule (42-44) uses the count numbers values N_s and N_b as they were acquired during a paired measurement; accordingly, the null hypothesis is $\overline{N_s} = \overline{N_b}$, or $\overline{N_s - N_b} = 0$. A faster convergence to normality can be obtained, if the same approach as in eqn (42-44) is applied to transformed count numbers. For a series of Poisson distributed counts, let us consider a complementary series of their square roots, of which the mean value is $\overline{\sqrt{N}}$. The null hypothesis to be tested will be as follows:

$$\overline{\sqrt{N_s}} = \overline{\sqrt{N_b}}, \text{ or } \overline{\sqrt{N_s} - \sqrt{N_b}} = 0 \quad (47)$$

This hypothesis is equivalent to the hypothesis $\overline{N_s} = \overline{N_b}$ discussed in the previous section:

$$s^2\left(\sqrt{N}\right) = \overline{N} - \left(\overline{\sqrt{N}}\right)^2; \ s^2\left(\sqrt{N}\right) = 1/4 \Rightarrow \left(\overline{\sqrt{N}}\right)^2 = \overline{N} - 1/4 \quad (48)$$

assuming the standard deviation of the 'Poisson roots' to be equal to ½ (see eqn 21). Consequently, the decision rule under which the null hypothesis is rejected becomes as follows[41]:

$$\sqrt{N_s} - \sqrt{N_b} > z_\alpha s(\sqrt{N_s} - \sqrt{N_b}) \Rightarrow \sqrt{N_s} - \sqrt{N_b} > z_\alpha \sqrt{s^2(\sqrt{N_s}) + s^2(\sqrt{N_b})}$$

$$\sqrt{N_s} - \sqrt{N_b} > z_\alpha \sqrt{1/4 + 1/4}; \ \sqrt{N_s} - \sqrt{N_b} > z_\alpha / \sqrt{2}$$

(49)

Solving the latter inequality relative to N_s allows presenting rule (49) in a more convenient form:

$$N_s > N_b + z_\alpha \sqrt{2N_b} + \frac{z_\alpha^2}{2} \Rightarrow L_c = \frac{z_\alpha^2}{2} + z_\alpha \sqrt{2N_b} \quad (50)$$

At large background count numbers, rule (50) converges to rule (44) that is based on the original, non-transformed data. However, at very small mean count numbers, rule (50) is - strictly - not applicable. Rule (50) is quite easy to obtain, owing to the simple variance stabilising transform [$s(\sqrt{N})=1/2$] used in derivation (49). This transform is not applicable at near-zero mean count numbers, as the square

root function cannot be satisfactorily linearised in this range, for which calculating the standard deviation of the square root statistics can be done analytically [using the Poisson probability function (eqn 22); see Electronic Appendix 5]. Such calculations show that, at very small mean count numbers, the 'Poissonian' square roots are over-dispersed compared to the ½ prediction from eqn (21). The size modelling confirms the inadequate performance of rule (50) for very small count numbers (Fig. 9c).

Fortunately, to extend the applicability of the square root transformation to very small counts, a simple fix is available. It was noted long ago that adding a small constant value to a Poisson distributed variable improves the convergence of the standard deviation of the resulting roots to ½: $s(\sqrt{N+d}) \to ½$. Initially, a value of $d=0.5$ was suggested[42], and already ensured a good convergence quality. Modern studies often use a value of $d=3/8$[18,19] shown to be optimal in a paper of F. Anscombe[43]; the $\sqrt{N+3/8}$ transform is usually called *Anscombe transform*. The corresponding decision rule is as follows:

$$\sqrt{N_s+d} - \sqrt{N_b+d} > z_\alpha s(\sqrt{N_s+d} - \sqrt{N_b+d}) \;\Rightarrow\; \sqrt{N_s+d} - \sqrt{N_b+d} > z_\alpha/\sqrt{2}$$

$$L_c = N_s - N_b = \frac{z_\alpha^2}{2} + z_\alpha\sqrt{2}\sqrt{N_b+d} \tag{51}$$

In radioactivity monitoring, this rule is also known as *Stapleton formula*[5,6]; the recommended d value is fairly close to 3/8 and amounts to 0.4. For example, at 5% of false positives ($z_{5\%}=1.645$), we obtain:

$$L_c = 1.645^2/2 + 1.645\sqrt{2}\sqrt{N_b+0.4} = 1.353 + 2.326\sqrt{N_b+0.4} \tag{52}$$

This rule has been suggested for practical use in the field of radioactivity monitoring for 15 years now[5,6]. Size calculations show a robust behaviour of this rule in a range of count numbers (Fig. 9d,e).

2.3. Beyond the Poisson-Gauss approximation: unbiased rules based on exact methods

The binomial decision rule

In the previous sections, we demonstrated that the Gauss distribution with its quantiles (z_a) is a good proxy for Poisson distributed count numbers and their differences in a wide range of mean count number values, provided principles used for the construction of the corresponding approximations are not abused. However, as a discrete distribution, and at small count numbers especially, the Poisson distribution and its derivatives are inherently prone to exact modelling using combinatorial methods. The problem of detection is no exception. In 1940, two agricultural scientists, Józef Przyborowski and Henry Wilenski, presented a combinatorial decision rule that is often called *binomial test*, or *C-test*[15]. The rate of false positives for this test is exactly known from the test derivation and guaranteed to be no larger than declared; a liberal behaviour is thus excluded without a need for size modelling; a conservative behaviour is possible, however. This test does not invoke a Poisson-Gauss approximation of any kind, but coincides with such approximations at larger count numbers (moreover, it coincides with the discontinuity corrected $\sqrt{N_s+N_b}$ rule at almost any count number). Nowadays, this test has reached a canonical status; it often serves as a reference decision rule, to which other rules are compared[5-8,44].

The binomial decision rule can be introduced as follows. Let us consider an example when, during a paired measurement, we collect a sum $N_b+N_s=5$ counts. We will assume, as always before, that the background and the signal counting times are equal to each other. The null hypothesis will be stated as usual: on average, $N_b=N_s$. In other words, under the null hypothesis, the probability for a specific count from the sum N_b+N_s to belong to the background is equal to the probability for this count to belong to the signal, both of these probabilities being equal to ½. These probabilities can be estimated by counting N_b and N_s values over a large series of paired replicate measurements. Of course, this does not mean that, in each particular paired measurement, $N_b=N_s$; moreover, for an (odd) sum of $N_b+N_s=5$, such an outcome is excluded. The possible (N_b,N_s) outcomes are as follows: (0,5), (1,4), (2,3), (3,2), (4,1), and (5,0). Each of these outcomes is characterised by some rate, or probability, of appearance. This rate is defined according to the probability function of the binomial distribution:

$$p(N_b=x, \; N_s=N_b+N_s-x) = \binom{N_b+N_s}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{N_b+N_s-x} = \binom{N_b+N_s}{x}\left(\frac{1}{2}\right)^{N_b+N_s} \tag{53}$$

where $\begin{pmatrix} N_b + N_s \\ x \end{pmatrix}$ denotes the binomial coefficient : $\begin{pmatrix} N_b + N_s \\ x \end{pmatrix} = \dfrac{(N_s + N_b)!}{x!(N_b + N_s - x)!}$ (54)

In our example, $p(N_b = 0, \ N_s = 5) = 1\left(\dfrac{1}{2}\right)^5 = 0.0313, \ or \ 3.13\%$

$$p(N_b = 1, \ N_s = 4) = 5\left(\dfrac{1}{2}\right)^5 = 0.1563, \ or \ 15.63\%, \text{ etc.}$$

Under the null hypothesis, the probability to obtain combination (0,5) is below 5%, which is our usual detection threshold (declared rate of false positives). More exactly, this probability equals 3.13%. It is too unlikely to obtain combination (0,5) under the null hypothesis. We conclude that, at the confidence level selected, the null hypothesis is rejected, and the signal - detected. The probability to obtain a combination of (1,4) is 15.625%. It is not too unlikely to obtain such an outcome under the null hypothesis: the signal is not detected. In a one-sided test, as ours, the same applies to all other outcomes. Interestingly, for a declared rate of false positives of 5%, detecting the signal in the presence of an 'empty' background acquisition of the same duration requires at least 5 counts to be registered during the signal acquisition - a rather conservative result, also from an intuitive standpoint.

An example of size calculation for the binomial decision rule is presented in Table 3. Similar calculations, extended over a range of mean count numbers, are summarised in Fig. 9f (an example of the source code can be found in Electronic Appendix 2). These calculations indeed confirm that the canonical version of the binomial rule described above is quite conservative in the range of small to intermediate counts (for the mid-p adjusted version of this rule, see the next section).

Apart from its use in biometrical applications, the binomial decision rule has been known in the field of radioactivity monitoring for decades, where it was usually called *Nicholson test*[5,6,17]. B. Altshuler and B. Pasternack also mention this rule as a solution for near-zero backgrounds in their classical paper on detection[4] (referring to 'Statistical theory' by K. Brownlee[10]). In 2008, L. Currie[40] reintroduced it to the metrology community based on the original text of J. Przyborowski and H. Wilenski.

Rules using the numerical approximation of p-values; bootstraps

For any statistic, adjusting the size of the corresponding decision rule to a desired value can be done numerically. This approach is conceptually similar to using inequalities (34) in the case of well-known background, but requires a computationally intensive double summation to account for the bivariate structure of paired measurements. It is based on the following formula:

$$p\left[T > t(N_b, N_s) \ under \ H_0\right] = \sum_{x=0}^{+\infty}\sum_{y=0}^{+\infty} \dfrac{\overline{N}^x e^{-\overline{N}}}{x!} \dfrac{\overline{N}^y e^{-\overline{N}}}{y!} I\left[T(x,y) > t(N_b, N_s)\right]$$ (55)

where T is the statistic to be tested, $t(N_b,N_s)$ is a particular value of this statistic that was obtained from a paired measurement, and $I[A]$ is the indicator function: $I[A] = \begin{cases} 1 \ if \ A \ is \ true \\ 0 \ otherwise \end{cases}$

The mean count number in eqn (55) is not known precisely, but can be estimated from a paired measurement: $\overline{N} = (N_s + N_b)/2$. The simplest statistic that can be tested represents the net count number: $T(x,y)=y-x$. The tested condition is $y-x>N_s-N_b$. By the summation of probabilities of all outcomes (x,y) that satisfy this condition, the actual size p of the critical region represented by the net values larger than N_s-N_b is evaluated. We reject the null hypothesis and report 'detection', if the p-value in eqn (55) is smaller than the required level of false positives (declared size, α). This rule is equivalent to a rule that would be based on the summation of probabilities of the individual net count number values in the right-hand parts of Skellam distributions shown in Fig. 4b. Moreover, it can be shown to be equivalent to the canonical formulation of the binomial decision rule[15] discussed in the previous section.

Statistics used in Poisson-Gauss approximations can also be tested. Especially, the statistic

$$T = \frac{y - x}{\sqrt{y + x}} \qquad (56)$$

from decision rule (42-44) has recently been studied in detail[8,19,44] and shown to be promising, along with some logarithmic statistics[8,19] not described in this text.

At this time, the method of numerical approximation of p-values is mostly used in biometrics[8,19,44]. We are unaware of its use in chemical metrology and closely related sciences. We believe that, at the level of rigour presently required in chemical metrology, using the numerical approximation of p-values for paired measurements is not compulsory because of (i) computational complexity of this method and (ii) availability of well performing Poisson-Gauss approximations (e.g., using the Anscombe transform, eqn 51); another good alternative is the mid-p adjusted version of the binomial rule described in the next section. For details, the reader is addressed to the literature[8,19,44] where examples of size simulation for rules based on the numerical approximation, as well as generalised formulae applicable under the condition of unequal counting times can be found. It seems likely that the method of numerical approximation will be more widely used in the future, given that the calculations are computer-aided.

The probability of $T>t(N_b, N_s)$ can also be estimated from a Monte-Carlo simulation using x and y values randomly selected from an 'artificial' Poisson distribution with a mean set to $(N_s + N_b)/2$. The resulting decision rule is usually called *bootstrap test*. The method of numerical approximation of p-values can be considered as an exact realisation of such test[44].

2.4. Unequal counting times

The $\sqrt{N_s + N_b}$ rule and its discontinuity corrected version

Though using equal time intervals to collect the background and the signal in a paired measurement is 'often the situation'[13], extending the background duration compared to the signal is also fairly common in the practice of LA-ICPMS. The decision rules derived in the previous sections need to be adapted accordingly, and their sizes - evaluated for a range of the background / signal time settings. Most decision rules are easy to adapt for an arbitrary setting of background and signal acquisition times (t_b and t_s, respectively) by substituting count numbers by their respective intensities:

$$I_b = \frac{N_b}{t_b}; \; I_s = \frac{N_s}{t_s} \qquad (57)$$

For example, the classical decision rule (42-44) can be re-formulated as follows:

$$I_s - I_b > z_\alpha s(I_s - I_b) \Rightarrow I_s - I_b > z_\alpha \sqrt{s^2(I_s) + s^2(I_b)} \Rightarrow I_s - I_b > z_\alpha \sqrt{\frac{\overline{N_s}}{t_s^2} + \frac{\overline{N_b}}{t_b^2}} \qquad (58)$$

Under the null hypothesis, $\overline{I_s} = \overline{I_b}$. Hence, the most probable estimate for the mean intensity (constrained maximum likelihood estimate[7,8]) is pooled as follows:

$$\frac{\overline{N_s}}{t_s} = \frac{\overline{N_b}}{t_b} = \frac{t_s}{t_s + t_b} \frac{N_s}{t_s} + \frac{t_b}{t_s + t_b} \frac{N_b}{t_b} = \frac{N_s + N_b}{t_s + t_b} \qquad (59)$$

From formulae (58,59), we obtain:

$$I_s - I_b > z_\alpha \sqrt{\frac{\overline{N_s}}{t_s^2} + \frac{\overline{N_b}}{t_b^2}} \Rightarrow I_s - I_b > z_\alpha \sqrt{\frac{1}{t_s} \frac{N_s + N_b}{t_s + t_b} + \frac{1}{t_b} \frac{N_s + N_b}{t_s + t_b}} \Rightarrow I_s - I_b > z_\alpha \sqrt{\frac{N_s + N_b}{t_s t_b}}$$

$$(60)$$

$$I_s - I_b > z_\alpha \sqrt{\frac{I_s}{t_b} + \frac{I_b}{t_s}}$$

The same result can be presented in the count number / time notation:

$$\frac{N_s}{t_s} - \frac{N_b}{t_b} > z_\alpha \sqrt{\frac{N_s + N_b}{t_s t_b}} \Rightarrow N_s - \frac{t_s}{t_b} N_b > z_\alpha \sqrt{(N_s + N_b)\frac{t_s}{t_b}} \qquad (61)$$

A curious reader will easily show that eqn (60,61) are equivalent to eqn (4,5,19) from part I.

Excluding the signal intensity (signal count number) from the right-hand part of the above expressions yields the following equations for the critical level:

19

$$L_c = (I_s - I_b)_c = \frac{z_\alpha^2}{2t_b} + z_\alpha \sqrt{\frac{z_\alpha^2}{4t_b^2} + I_b\left(\frac{1}{t_s} + \frac{1}{t_b}\right)}; \quad L_c = (N_s - \frac{t_s}{t_b}N_b)_c = \frac{z_\alpha^2 t_s}{2t_b} + z_\alpha \sqrt{\frac{z_\alpha^2 t_s^2}{4t_b^2} + \frac{N_b t_s}{t_b}\left(1 + \frac{t_s}{t_b}\right)}$$

$$(62)$$

In the practice of radioactivity monitoring[6,16,17] and biometrics[7], this decision rule has been recommended for a while, though criticised for a relatively liberal behaviour at $t_b > t_s$[5,6] (Fig. 9a). More exactly, its actual size 'explodes' at $t_b / t_s > z_\alpha^2$; the idea of the corresponding proof can be found in Ref. 5. Eqn (58-62) can be corrected for discontinuity [with a correction term equal to $(t_s + t_b)/2t_s t_b$][10,11]:

$$I_s - I_b > \frac{t_s + t_b}{2t_s t_b} + z_\alpha \sqrt{\frac{I_s}{t_b} + \frac{I_b}{t_s}}$$

$$L_c = (I_s - I_b)_c = \frac{t_s + t_b}{2t_s t_b} + \frac{z_\alpha^2}{2t_b} + z_\alpha \sqrt{\frac{z_\alpha^2}{4t_b^2} + I_b\left(\frac{1}{t_s} + \frac{1}{t_b}\right) + \frac{1}{2t_b}\left(\frac{1}{t_s} + \frac{1}{t_b}\right)} \quad (63)$$

$$L_c = (N_s - \frac{t_s}{t_b}N_b)_c = \frac{t_s + t_b}{2t_b} + \frac{z_\alpha^2 t_s}{2t_b} + z_\alpha \sqrt{\frac{z_\alpha^2 t_s^2}{4t_b^2} + (N_b + 1/2)\frac{t_s}{t_b}\left(1 + \frac{t_s}{t_b}\right)}$$

The discontinuity corrected version above is conservative; its robustness improves at $t_b/t_s \geq 2$ (Fig. 9b).

The $\sqrt{2N_b}$ rule

Based on eqn (58,59), it is, perhaps, worth to re-consider the $\sqrt{2N_b}$ decision rule (eqn 36) and extend it to any combination of counting times. The $\sqrt{2N_b}$ rule is biased regarding the estimation of the mean count number: under the null hypothesis, only the background value is used. This is equivalent to assigning a weight of 0 to the signal intensity and a weight of 1 - to the background intensity in eqn (59):

$$\frac{\overline{N_s}}{t_s} = \frac{\overline{N_b}}{t_b} = \frac{N_b}{t_b}; \quad I_s - I_b > z_\alpha \sqrt{\frac{\overline{N_s}}{t_s^2} + \frac{\overline{N_b}}{t_b^2}} \Rightarrow I_s - I_b > z_\alpha \sqrt{\frac{N_b}{t_s t_b} + \frac{N_b}{t_b^2}} \Rightarrow L_c = z_\alpha \sqrt{I_b\left(\frac{1}{t_s} + \frac{1}{t_b}\right)}$$

$$(64)$$

The robustness of rule (64) naturally improves, if the background is measured longer than the signal: the biased weighing is less important in this case. Still, for practically important t_b/t_s ratios, at small count numbers, the robustness of rule (64) is insufficient (Fig. 10a,b). Except for large t_b/t_s ratios and (or) high background intensities, this rule cannot be approved for practical use (though, from a purely empirical standpoint, its use can perhaps be tolerated, provided a conservative z_a value is employed and the rule is empirically corrected for an 'empty' background, see section 2.10 and Fig. 10a,b).

Rules based on the square root variance stabilising transformations

Detection rules based on variance-stabilising transformations, such as the Anscombe transform, are also easily adaptable for arbitrary background / signal time settings:

$$\sqrt{\frac{N_s + d}{t_s}} - \sqrt{\frac{N_b + d}{t_b}} > z_\alpha s\left(\sqrt{\frac{N_s + d}{t_s}} - \sqrt{\frac{N_b + d}{t_b}}\right) \Rightarrow \sqrt{\frac{N_s + d}{t_s}} - \sqrt{\frac{N_b + d}{t_b}} > \frac{z_\alpha}{2}\sqrt{\frac{1}{t_s} + \frac{1}{t_b}}$$

$$L_c = (I_s - I_b)_c = d\left(\frac{1}{t_b} - \frac{1}{t_s}\right) + \frac{z_\alpha^2}{4}\left(\frac{1}{t_s} + \frac{1}{t_b}\right) + z_\alpha \sqrt{(I_b t_b + d)\frac{1}{t_b}\left(\frac{1}{t_s} + \frac{1}{t_b}\right)} \quad (65)$$

$$L_c = (N_s - \frac{t_s}{t_b}N_b)_c = d\left(\frac{t_s}{t_b} - 1\right) + \frac{z_\alpha^2}{4}\left(1 + \frac{t_s}{t_b}\right) + z_\alpha \sqrt{(N_b + d)\frac{t_s}{t_b}\left(1 + \frac{t_s}{t_b}\right)}$$

This decision rule is very recommended for its robustness, especially at $t_b > t_s$[5,6,19] (Fig. 9d,e).

The binomial rule and its mid-p adjusted version

Modifying the binomial decision rule is equally simple. Under the null hypothesis, instead of using probabilities of ½ for both signal and background, the following definitions naturally arise[5,6,15]:

$$p_b = \frac{t_b}{t_b + t_s}; \quad p_s = \frac{t_s}{t_b + t_s} \qquad (66)$$

For example, for a paired measurement, of which the background acquisition is three times longer than the signal acquisition, $p_b=3/4$ and $p_s=1/4$. Accordingly, we obtain:

$$p(N_b = x, \ N_s = N_b + N_s - x) = \binom{N_b + N_s}{x}\left(\frac{3}{4}\right)^x\left(\frac{1}{4}\right)^{N_b+N_s-x} \qquad (67)$$

Developing this example further, for a sum $N_b+N_s=7$ counts, under the null hypothesis, the test returns the following probabilities: $p(N_b = 0, \ N_s = 7) = 1\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^7 = 0.000061, \ or \ 0.0061\%$

$$p(N_b = 1, \ N_s = 6) = 7\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^6 = 0.001282, \ or \ 0.1282\%$$

$$p(N_b = 2, \ N_s = 5) = 21\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^5 = 0.011536, \ or \ 1.1536\%$$

$$p(N_b = 3, \ N_s = 4) = 35\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^4 = 0.057678, \ or \ 5.7678\%, \text{etc.}$$

At a rate of false positives set to be no higher than 5%, for a paired measurement with an N_b+N_s value of 7 counts, combinations (0,7), (1,6) and (2,5) imply that that signal is detected: their cumulative probability amounts to 1.29%, which is below the 5% threshold. The remaining combinations are too likely to obtain under the null hypothesis: the signal is not detected. Such calculations can be generalised: for all non-negative integers x from 0 to n, combinations $(N_b = x, \ N_s = N_b + N_s - x)$

result in detection, provided $p_{cumulative} = \sum_{x=0}^{n} \binom{N_b + N_s}{x}\left(\frac{t_b}{t_b + t_s}\right)^x\left(\frac{t_s}{t_b + t_s}\right)^{N_b+N_s-x} \leq \alpha \qquad (68)$

The size calculation shows a conservative behaviour of the binomial rule (Fig. 9f) that is quite similar to the behaviour of eqn (63) from the previous section. Again, the robustness improves at $t_b/t_s\geq2$, although the rule still remains rather conservative (Fig. 9f). The conservatism of the binomial rule motivated attempts to adjust its formulation in order to improve the robustness[17,45,46]. The underlying idea of such adjustments is to provide a recipe allowing in some cases to consider as detected the (N_b,N_s) combination next to the critical region [as estimated using the canonical version of the rule, e.g., combination (3,4) in the example above]. Here, we introduce the *mid-p adjusted* version of this rule after H. Lancaster[45,46]. We modify formula (68) for the cumulative probability as follows:

$$p_{cumulative}^{mid-adjusted} = \sum_{x=0}^{n} \binom{N_b + N_s}{x}\left(\frac{t_b}{t_b + t_s}\right)^x\left(\frac{t_s}{t_b + t_s}\right)^{N_b+N_s-x} + 0.5\binom{N_b + N_s}{x=n+1}\left(\frac{t_b}{t_b + t_s}\right)^x\left(\frac{t_s}{t_b + t_s}\right)^{N_b+N_s-x} \leq \alpha$$

$$(69)$$

For all non-negative integers x from 0 to $n+1$, for which the above inequality holds true, combinations $(N_b = x, \ N_s = N_b + N_s - x)$ result in detection. In the example above,

$$p_{cumulative}^{mid-adjusted} = 1.29 + 0.5 \times 5.77 = 4.17\% \leq 5\%.$$

Thus, combination (3,4) is also detected when using the mid-p adjusted version of the binomial rule, contrary to the outcome of its canonical version. The actual size becomes less conservative and more closely approaches the declared value, provided the mid-p adjustment is employed (Fig. 9g).

It is also worth briefly mentioning an alternative way to adjust the size of the binomial rule. It is based on the method of the auxiliary random experiment, and combines the possibility to reproduce the declared size exactly with more complexity and ambiguous results for the (N_b,N_s) combination next to the critical region (the same measurement returns either detection or non-detection, depending on the outcome of the auxiliary experiment, not of the measurement itself)[38,46]. For decades, this method has been suggested for use in the field of radioactivity monitoring, although its practical application seems to be restricted, for the above reasons. We refer the interested reader to the literature[5,6,17].

2.5. Conservatism of decision rules for mean background count numbers limiting to zero

In the range of mean background count numbers limiting to zero ($\overline{N}_b t_s / t_b < 0.5$-1 count), it does not seem possible with most decision rules, including those characterised by robust and even liberal behaviour at somewhat larger count numbers, to reach the declared rate of false detections: the actual size is smaller than declared (i.e., the rule is conservative, see Fig. 9a-g). This issue is somewhat alleviated by the difficulty to decrease the background strength of practical measurements indefinitely: the low level of dark noise observed using some ICPMS (e.g., 0.2 cps for mass 220 for Element 2 and XR mass spectrometers) is not easily reproducible during the analysis of naturally available isotopes, partly because of the inevitable, though weak, contamination of the instrument during the routine work. Still, at small t_s values, practical background measurements may approach the above range.

At this time, no real solution exists to overcome the conservatism of detection at mean count numbers limiting to zero. The only 'universally' robust decision rule available is the binomial rule used in combination with the method of the auxiliary random experiment (see section 2.4). However, it is the auxiliary experiment and not the measurement outcome that decides if the sample activity is detected, provided the mean background count number limits to zero[46]! In this case, it just needs to be accepted that the actual rate of false positive errors can be smaller than declared: if the analyte is detected, it is done with more confidence than declared. There is no big fault with this circumstance[37,38,46].

2.6. Extending the applicability of decision rules beyond the ordinary Poisson approximation

For strong backgrounds (usually - at least several thousands cps, see Table 4 in Electronic Appendix 1 for an example), the presence of non-negligible excess variance is possible. Such backgrounds cannot be completely described by the ordinary Poisson process; the square root dependence $s(N) = \sqrt{N}$ yields only a minimum estimate of the standard deviation (see section 1.3). Then, decision rules based on the ordinary Poisson distribution cannot be employed in their original form. Returning to the basic definition of a decision rule based on the Poisson-Gauss approximation is necessary, e.g.:

$$N_s - N_b > z_\alpha \sqrt{s^2(N_s) + s^2(N_b)} \qquad (70)$$

where standard deviations $s(N_b)$ and $s(N_s)$ have to be estimated under the null hypothesis (consult texts on two-sample z- and t-tests regarding how to pool the standard deviation under the null hypothesis, provided the signal and the background are considered as time-resolved measurements containing k_s and k_b sweeps, respectively, resulting in a common distribution with a degree of freedom $k_s + k_b - 2$[10,11]). Eqn (70) and the associated calculus can be substantially simplified, if a biased estimate of the net activity standard deviation is accepted. Assuming that both estimates, N_b and N_s, come from the same distribution with a standard deviation equal to $s(N_b)$ [instead of pooling the standard deviation as a sort of weighted average of $s(N_b)$ and $s(N_s)$ values[10,11]], eqn (70) returns the familiar formula $L_c = z_\alpha \sqrt{2} s(N_b)$. Combining this formula with eqn (28), we obtain:

$$L_c = z_\alpha \sqrt{\frac{2k}{(k-1)} \sum_{i=1}^{k} \left(N_{sw\,i} - \overline{N}_{sw} \right)^2} \qquad (71)$$

where k is the number of background sweeps, $N_{sw\,i}$ - count number in an individual sweep, $\overline{N}_{sw}$ - mean count number per sweep. For an ordinary Poisson distributed background as a special case, we obtain the equation $L_c = z_\alpha \sqrt{2} s(N_b) = z_\alpha \sqrt{2 N_b}$ analysed in the previous sections of this text.

Using the intensity-based notation combined with a biased estimation of the net count number standard deviation, eqn (71) can be extended to any combination of the background / signal counting times:

$$I_s - I_b > z_\alpha \sqrt{s^2(I_s) + s^2(I_b)}$$

$$s(I_b) = \frac{s(I_{sw\,b})}{\sqrt{k_b}}; \quad s(I_s) = \frac{s(I_{sw\,s})}{\sqrt{k_s}} = \frac{s(I_{sw\,b})}{\sqrt{k_s}}, \quad where \; s(I_{sw\,b}) = \sqrt{\frac{1}{k_b - 1} \sum_{i=1}^{k_b} \left(I_{sw\,i} - \overline{I}_{sw} \right)^2} \qquad (72)$$

Accordingly, the critical level can be given as follows[13]:

$$L_c = z_\alpha s(I_{sw\,b})\sqrt{\frac{1}{k_b}+\frac{1}{k_s}}; \ \ L_c = z_\alpha s(I_b)\sqrt{1+\frac{t_b}{t_s}} \qquad (73)$$

where $s(I_{sw\,b})$, $s(I_{sw\,b})$ and $s(I_b)$, $s(I_s)$ denote standard deviations for the individual sweep intensities and mean intensities of background and sample signal, respectively, $I_{sw\,i}$ - intensity of an individual sweep, $\overline{I}_{sw}$ - mean intensity per sweep, k_b, k_s and t_b, t_s - number of sweeps and acquisition times, respectively.

For ordinary Poisson distributed data, eqn (73) coincides with eqn (64), of which the size is shown in Fig. 10a,b. The liberal behaviour of eqn (64) becomes less marked with increasing the count number (value of the square root function changes little upon changing its argument, if the latter is large: replacing $2N_b$ by N_s+N_b returns similar result at large count numbers). For the sake of its simplicity, we can recommend eqn (73) as a solution applicable (*only*) for large counts. For a rigorous user, more complex unbiased solutions can be proposed. One of them is outlined below, using the method of square root transform (eqn 49-51) as an example. At $\overline{N}_b t_s / t_b > 5$, the term d improving the quality of variance stabilisation at small counts can be set to zero (see Fig. 9c). By error propagation, we obtain:

$$\sqrt{N_s}-\sqrt{N_b} > z_\alpha s(\sqrt{N_s}-\sqrt{N_b}) \ \Rightarrow \ \sqrt{N_s}-\sqrt{N_b} > z_\alpha\sqrt{s^2(\sqrt{N_s})+s^2(\sqrt{N_b})}$$

$$\Rightarrow \ \sqrt{N_s}-\sqrt{N_b} > z_\alpha\sqrt{\frac{s^2(N_s)}{4\overline{N}_s}+\frac{s^2(N_b)}{4\overline{N}_b}} \qquad (74)$$

Under the null hypothesis, $\overline{N}_b = \overline{N}_s$; the most probable estimate of this value is the mean $(N_b+N_s)/2$ of the measured count numbers. Using eqn (19) and the property $s(M)/M=Const^{20,23\text{-}27}$ of the excess variance, the standard deviation for a distribution centered around this mean is estimated as follows:

$$s^2(N) = \overline{N}+p^2s^2(M), \ \ \overline{N}=\frac{N_s+N_b}{2}, \ \ s(M)=s(M_b)\frac{N_s+N_b}{2N_b}$$

$$\Rightarrow s^2(N)=\frac{N_s+N_b}{2}+p^2s^2(M_b)\left(\frac{N_s+N_b}{2N_b}\right)^2 = \frac{N_s+N_b}{2}+\left(s^2(N_b)-N_b\right)\left(\frac{N_s+N_b}{2N_b}\right)^2 \qquad (75)$$

Inequality (74) is reorganised accordingly:

$$\sqrt{N_s}-\sqrt{N_b} > z_\alpha\sqrt{\frac{2\left[\frac{N_s+N_b}{2}+\left(s^2(N_b)-N_b\right)\left(\frac{N_s+N_b}{2N_b}\right)^2\right]}{4\frac{N_s+N_b}{2}}} \qquad (76)$$

$$\Rightarrow \ \sqrt{N_s}-\sqrt{N_b} > \frac{z_\alpha}{\sqrt{2}}\sqrt{1+\left(s^2(N_b)-N_b\right)\frac{N_s+N_b}{2N_b^2}}$$

The signal uncertainty is excluded from the right-hand part of inequality (76) and does not need to be estimated, which simplifies calculations. By solving a quadratic equation, it is further possible to derive an analytical expression for the critical level that is a function of N_b, $s(N_b)$ and z_a only:

$$L_c = (N_s-N_b)_c = \left[\frac{\sqrt{N_b}+\sqrt{N_b-\left[1-\left(s^2(N_b)-N_b\right)\frac{z_\alpha^2}{4N_b^2}\right]\left[N_b-\frac{z_\alpha^2}{4}\left(1+\frac{s^2(N_b)}{N_b}\right)\right]}}{\left[1-\left(s^2(N_b)-N_b\right)\frac{z_\alpha^2}{4N_b^2}\right]}\right]^2 - N_b \qquad (77)$$

This equation can be given in the intensity-based notation and extended to any combination of the background and signal counting times (for derivation, see Electronic Appendix 6):

$$L_c = (I_s - I_b)_c = \left[\frac{\sqrt{I_b} + \sqrt{I_b - \left[1 - \frac{(t_b s^2(I_b) - I_b)z_\alpha^2}{4 t_b I_b^2}\right]\left[I_b - \frac{z_\alpha^2}{4}\left(\frac{1}{t_b} + \frac{t_b s^2(I_b)}{t_s I_b}\right)\right]}}{\left[1 - \frac{(t_b s^2(I_b) - I_b)z_\alpha^2}{4 t_b I_b^2}\right]} \right]^2 - I_b \quad (78)$$

For ordinary Poisson data, eqn (77, 78) return decision rule (50) that is the simplest among the rules using the square root transform. At the same time, eqn (78) is also applicable to strong backgrounds showing an excess variance. Compared to the much simpler eqn (73), for large count numbers, eqn (78) returns slightly higher critical values (with a maximal difference of a few percentage points). We recommend using eqn (78), especially as an extension of decision rules based on the square root trans-form. Strictly, it is applicable in the range of robustness of rule (78), i.e., at $\overline{N}_b t_b / t_s > 5$ counts. Its genuine application area represents strong backgrounds beyond the ordinary Poisson approximation.

2.7. Power of detection

Actual size of a decision rule characterises the probability of detection using this rule, provided the null hypothesis is true (mean net sample activity equals zero; no analyte in the sample). Power of a decision rule characterises the probability of detection using the above rule, provided the null hypothesis is false (positive mean net sample activity; presence of analyte in the sample). The positive net sample activity can be set to any value, including values infinitely close to zero. Hence, the concept of power can be considered as a generalisation of the concept of size. Liberal decision rules tend to be more powerful in the range of small counts, as they have larger size in this range. Conservative decision rules tend to be less powerful (smaller size). For example, the canonical formulation of the bino-mial decision rule is, unfortunately, well known for its relatively small power[15,19,44]. With increasing the mean net sample activity (i.e., with increasing the difference between the mean background value and the mean sample value), the power of detection limits to unity. With decreasing the mean net sample activity, it limits to the actual size of the decision rule used. As already discussed, detection limit represents a sample activity value ensuring a specified power of detection (usually, a 95% proba-bility of detection; see sections 1.1-2, Fig. 2a,b and Fig. 11a,b). Power studies often complement the study of size of decision rules[7,8,15,19,41,44]. An 'ideal' decision rule is robust (its declared and actual sizes are closely matched) and has a large power. For two equally robust decision rules (~ same size), the rule having more power is preferred for the purposes of practical data treatment.

Studying the power of detection of a decision rule has much in common with studying its actual size. Two basic approaches coexist: Monte-Carlo simulation and exact combinatorial calculation. The Mon-te-Carlo approach implies creating two large sets of random Poisson distributed values for pre-defined mean background and mean sample signal count number values, respectively. From each of these sets, a value is randomly selected. The first value is stored as a background estimate (N_b). The second is stored as a sample signal estimate (N_s). The net value of these estimates is then tested using the studied decision rule: is it large enough for detection? The process is replicated many times; in the end, it re-turns two numbers - number of detections and number of non-detections. The so-estimated probability of detection characterises the power of the studied rule. The combinatorial approach implies compu-ting the probability of appearance of all (N_b, N_s) pairs that return 'detection', using the probability func-tions of the Poisson distribution parameterised at pre-defined mean background and mean sample sig-nal count number values, respectively. Here, we employ the combinatorial approach only (the interest-ed reader can refer to the source code in Electronic Appendix 7). Based on the size calculations pre-sented in this text and on the published size studies[5-8,19], two decision rules appear to be of particular interest because of their robustness - the Poisson-Gauss approximation using the square root transform and the mid-p adjusted version of the binomial rule. Their power values are shown in Fig. 11a,b. The square root transform has somewhat more power in the range of small counts; representative literature data show that, in general, this decision rule represents one of the most powerful and robust rules for testing the equality of two Poisson means[19]. It can definitely be recommended for practical use.

2.8. Detection as a function of acquisition parameters

Based on the analysis of a decision rule selected for the data treatment, it is possible to optimise the measurement process in order to facilitate detection. Let us consider decision rule (65) based on the square root transform. This rule can be reformulated as follows:

$$N_s > N_b \frac{t_s}{t_b} + d\left(\frac{t_s}{t_b} - 1\right) + \frac{z_\alpha^2}{4}\left(1 + \frac{t_s}{t_b}\right) + z_\alpha \sqrt{(N_b + d)\frac{t_s}{t_b}\left(1 + \frac{t_s}{t_b}\right)} \qquad (79)$$

Detection is ensured, if N_s is larger than the right-hand term of formula (79). Three special cases arise:

(i) N_s is adjustable, other parameters are constant. Under this condition, increasing N_s facilitates detection, because the right-hand term of formula (79) remains constant. Increasing N_s can be achieved, for example, by increasing the ablation pit size, or by using raster ablation;

(ii) t_b is adjustable, N_b/t_b (in average), as well as other parameters are constant. Under this condition, increasing t_b facilitates detection, since the right-hand term of formula (79) decreases, while N_s remains constant. Obviously, increasing t_b means increasing the background acquisition time;

(iii) t_s is adjustable, other parameters are constant. It is, perhaps, the most interesting case. Let us consider an ablation pit of some diameter and depth; its volume is constant. Some amount of analyte is extracted from the pit during the analysis. In a first approximation, N_s can thus be considered constant. The same pit size can be obtained at different ablation speeds (e.g., by varying the laser repetition rate). Accordingly, t_s changes. Reducing t_s by increasing the ablation speed (for the same amount of material ablated) results in a decrease of the right-hand term of formula (79), facilitating detection.

To summarise, using larger pit sizes and ablation speeds and increasing the background duration are factors facilitating detection, which agrees with the routine practice of many trace element LA-ICPMS laboratories and the literature data[47]. Extending the background duration, especially in the range of t_b/t_s ratios from 1 to 3, also helps improving the robustness and power of detection (Fig. 9-11).

2.9. Presenting critical levels in mass content units

Although decisions 'detected' and 'not detected' in particle counting techniques are usually taken at a level of counts (per second)[1-6], it is also possible to consider a distribution of analyte contents resulting purely from the net activity fluctuations under the null hypothesis, when the sample signal and the background are statistically indistinguishable, and to describe the critical level in mass content units. In LA-ICPMS, the analyte content is quantified in two steps, combining external and internal standardisation[13,14,25]. For isotopes x and y representing analyte and internal standard, respectively, we have:

$$\frac{I_{s\,net}^x}{I_{s\,net}^y} = \beta \frac{c^x}{c^y} \qquad (80)$$

where $I_{s\,net}^x$ and $I_{s\,net}^y$ are net intensities of the above isotopes: $I_{s\,net}^x = I_s^x - I_b^x, \;\; I_{s\,net}^y = I_s^y - I_b^y$ (81)

and c^x and c^y - their mass contents.

In the first step, eqn (80) is applied to a standard reference material with reasonably well known c^x, c^y values. The mean net intensity ratio $I_{s\,net}^x / I_{s\,net}^y$ for isotopes x and y is measured by ICPMS. Coefficient β, representing a *relative sensitivity factor* in the notation of eqn (80), is the only unknown in this equation and can easily be calculated. In the second step, eqn (80) is applied to a sample. The sample intensity ratio is measured by ICPMS. The relative sensitivity factor specific to the sample is generally not known, but assumed to be the same as for the standard. The only remaining unknown is the isotope content ratio in the sample. From this ratio, the analyte content is easily derivable, provided the internal standard content is known *a priori*. The whole procedure is reduced to the following equation:

$$c_{sample}^x = \frac{I_{s\,net}^x}{\beta(I_{s\,net}^y / c_{sample}^y)}, \text{ where } \beta = \frac{I_{s\,net\,std}^x / c_{std}^x}{I_{s\,net\,std}^y / c_{std}^y} \qquad (82)$$

the term $\beta(I_{s\,net}^y / c_{sample}^y)$ of this equation being often called *sensitivity*: $S = \beta(I_{s\,net}^y / c_{sample}^y)$:

$$c_{sample}^x = \frac{I_{s\,net}^x}{S} \qquad (83)$$

The factor β in eqn (80) can also be determined using more than one standard reference material. In this case, a regression of the concentration vs. intensity ratio data can be carried out to determine β as the slope of the regression line. Since element contents in standards currently available for trace element LA-ICPMS work often show non-negligible uncertainties, such regressions represent measurement error models, i.e. regressions with uncertainties in both variables, concentration ratio and intensity ratio. The LA-ICPMS practice of using several standards to determine β is still □□□□□□□□□□□□; the interested reader is referred to the literature on measurement error models[48].

Using eqn (83), a net intensity value can be directly converted to mass units. Converting (propagating) a confidence interval, as in the case of critical levels, is also rather straightforward. Let us first assume that sensitivity is known precisely. Substituting $I_{s\,net}^x$ in eqn (83) by net intensity values $I_{net\,H_0}$ from a common distribution constructed under the null hypothesis (from a large series of paired background acquisitions) yields a series of estimated analyte contents c_{H_o} resulting from background fluctuations and not from the presence of analyte (Fig. 12). The probability of committing a false positive error at a specified critical level remains the same irrespective of whether the content or the intensity notation is used, the corresponding critical levels being connected as follows:

$$L_c^x(content) = \frac{L_c^x(intensity)}{S} \qquad (84)$$

In practice, sensitivity has an uncertainty, though its scatter never extends to zero. Hence, the bivariate function $c(I_{s\,net}, S) = I_{s\,net} / S$ is satisfactorily approximated by a plane in the vicinity of point $(0, \bar{S})$:

$$Var(c_{H_o}) = Var\left(\frac{I_{net\,H_0}}{S}\right) = \frac{1}{S^2}Var(I_{net\,H_0}) + \frac{I_{net\,H_0}^2}{S^4}Var(S) \Bigg|_{I_{net\,H_o} = \bar{I}_{net\,H_0} = 0,\ S = \bar{S}}$$

$$(85)$$

$$Var(c_{H_0}) = \frac{1}{\bar{S}^2}Var(I_{net\,H_0}) \Rightarrow s(c_{H_0}) = \frac{s(I_{net\,H_0})}{\bar{S}}$$

Therefore, eqn (84) still holds. This agrees with the previous experience of critical level computation in LA-ICPMS[13,14] that is solely based on using eqn (84), not including the sensitivity uncertainty.

2.10. An assessment of the existing practices of critical level estimation in LA-ICPMS

Analysing (ultra-) trace element contents always was an important field of application for ICPMS. The need to distinguish weak signals from the background noise has resulted in the appearance of a few articles providing recommendations for the LA-ICPMS community regarding the calculation of critical levels and, more recently, detection limits[13,14,47]. In particular, the following formula for the critical level ('detection limit', using the authors' terminology) for paired measurements was suggested[13]:

$$L_c(intensity) = 3s(I_{sw\,b})\sqrt{\frac{1}{k_b} + \frac{1}{k_s}} \qquad (86)$$

Eqn (86) is obtained as a special case of eqn (73) by setting z_a to 3; it uses the same notation. It is used for any distribution of uncorrelated sweep intensities that can be approximated by the Gauss distribution (see section 1.3). For ordinary Poisson distributed data, it is equivalent to eqn (64) for $z_a=3$:

$$L_c(intensity) = 3s(I_{sw\,b})\sqrt{\frac{1}{k_b} + \frac{1}{k_s}} = 3\sqrt{\frac{\bar{I}_{sw\,b}}{t_{sw}}}\sqrt{\frac{1}{k_b} + \frac{1}{k_s}} = 3\sqrt{I_b\left(\frac{1}{t_{sw}k_b} + \frac{1}{t_{sw}k_s}\right)} = 3\sqrt{I_b\left(\frac{1}{t_b} + \frac{1}{t_s}\right)} \qquad (87)$$

where t_{sw} denotes per-sweep dwell time for the isotope of interest. For $k_b=k_s$ ($t_b=t_s$, background and signal of equal duration), eqn (86,87) return the $\sqrt{2N_b}$ decision rule.

The actual size of rule (87) is shown in Fig. 10b. For weak backgrounds, this rule is liberal even in the frequently practiced case of adding one count to an 'empty' background: the actual rate of false positive errors much exceeds the declared rate of 0.135% defined by the quantile (z_a=3) of the Gauss distribution. This is not surprising: the liberality of the 'biased' decision rules of the $\sqrt{2N_b}$ family has been much discussed in the previous sections of this text. Still, from a purely practical standpoint, using decision rule (86,87) empirically corrected for an 'empty' background is, perhaps, not entirely unreasonable, because its maximum size does not exceed 2.4% ($t_b \geq t_s$), owing to the very conservative selection of the quantile. It slowly decreases to the declared level with increasing the mean background count number, varying by a factor of ~18 (!) in the practical range of LA-ICPMS backgrounds.

From 1996 till 2008, rule (86) has been the only decision rule used for LA-ICPMS data treatment; it still remains in use with some laboratories. Noteworthy, all critical values presented in the LA-ICPMS literature before 2008 were called 'detection limits', following the practice that prevailed among analysts during several decades after the appearance of a well-known paper by H. Kaiser in 1947[12]. A return of the ICPMS community to the modern meaning of the above terms, which is part of the current IUPAC recommendations[1,2], began in 2008[47]. It does not seem to be achieved in practice: the communication level between statistics, metrology, and ICPMS is insufficient at times.

The above article[47] from 2008 considers only the relatively simple concept of well-known background [$L_c = z_a \sqrt{\overline{N}_b}$], no performance analysis of decision rules for paired measurements was given. At the same time, practical LA-ICPMS data usually represent true, non-replicated paired measurements.

In 2012, the problem of detection in LA-ICPMS was approached again[14]. The decision rule suggested coincides with rule (64) from this text for z_a=1.645 (nominal rate of false positive errors set to 5%):

$$L_c = 1.645 \sqrt{I_b \left(\frac{1}{t_s} + \frac{1}{t_b} \right)} \qquad (88)$$

Compared to decision rules (86,87), this rule uses a less conservative z_a value; its declared size is larger. Its actual size is shown in Fig. 6 (t_b=t_s) and 11a ($t_b \geq t_s$); increasing the declared size makes this rule too liberal for practical use at small background count numbers and at commonly employed t_b/t_s ratios.

In both cases, difficulties related to the approximation of Poisson data by the continuous Gauss distribution at small counts were mentioned[14,47]. Important difficulties concerned with the biased estimation of the net count number standard deviation and the liberal behaviour of the resulting decision rule(s) were not discussed, nor hidden from a reader willing to follow the literature cited in these articles. We fully admit the significant motivational value of these articles that, for the first time, introduced some modern texts on detection to the ICPMS community.

Unfortunately, this effort has remained isolated. Eqn (86-88) appear to exhaust the list of decision rules suggested to the users of LA-ICPMS for thirty years. Sharing the same construction and deficiencies, they are a poor choice at small count numbers.

In this context, one technical aspect is worth mentioning. It follows from the literature reviewed in this text that the $\sqrt{2N_b}$ decision rule and its derivatives can be substituted by rules that perform much better in the range of small counts, while most background measurements acquired in the current practice of LA-ICPMS fall in this range. Indeed, 'modern ICPMS instruments with their curved ion channels supress the photon noise; setting a correct discriminator threshold minimises the internal noise of the secondary electron multiplier [operated in the counting regime]; the gas blank is clean for most mass-to-charge ratios.'[20]. However, early ICPMS instruments seem to have been less refined regarding the background suppression. For example, according to the specification of Elan 6000 ICPMS introduced in 1994 by Perkin Elmer, the mean background intensity at mass 220 must be <30 cps, while the corresponding value for Elan 6100 DRC ICPMS, appeared in 1999, is <2 cps. The modern instruments generally match the latter value, some of them having exceptionally low dark (detection system related) noise. Thus, at the time of introduction of decision rule (86) to the ICPMS community[13], the problem of correct handling of weak backgrounds seems to have been somewhat less acute than now.

Still, the lack of interest from the ICPMS community to the large body of literature beyond ICPMS concerned with tests for the equality of two Poisson means represents a somewhat regrettable aspect of

27

the history of quantification of LA-ICPMS detection capabilities. In fact, an LA-ICPMS user is confronted with an excessive choice of reasonably well performing decision rules rather than with a lack of such rules (Fig. 13). In reconciliation, it could be said that the IUPAC texts[1,2] are also (partly) deficient regarding the choice and performance analysis of decision rules for Poisson distributed data.

2.11. Tests for the equality of Poisson means and the different techniques of analytical chemistry

The core part of this text is dedicated to the mathematical theory of tests for comparing two Poisson means, especially in the range of small count numbers. Consequently, this text can, first of all, be used to serve analytical techniques (i) dealing with very weak Poisson distributed backgrounds and signals and (ii) employing the concept of paired measurements. The LA-ICPMS is one of the most obvious representatives of such techniques. For example, in optical emission ICP and laser induced breakdown spectrometry (ICP-OES and LIBS, respectively), there is a continuous background emission resulting in an increased background noise level, as well as a long-standing practice to replicate background measurements to compute critical values and detection limits (which requires simpler decision rules that are generally available in the literature of chemical metrology[1-3,40]). A continuous background (*Bremsstrahlung*) is also typical of analytical techniques based on X-ray fluorescence.

Apart from LA-ICPMS, one technique that could benefit from approaches reviewed in this text is secondary ion mass spectrometry (SIMS). As a proxy, the ordinary Poisson distribution is frequently used to describe count number statistics of SIMS data[49,50]. Besides, SIMS is known as a technique often dealing with weak background noise. If the background is analysed using the concept of paired measurements, then unbiased decision rules from Part II of this text can be recommended as a replacement for the $\sqrt{2}N_b$ rule previously proposed to the SIMS community[49] (see also the section 'Conclusions and recommendations'). However, in the practice of SIMS, in contrast to LA-ICPMS, replicate background measurements are also encountered. At an arbitrary number (k) of background replicates, in the absence of instrument drift, unbiased decision rules from Part II of this text can still be used, assuming that $t_b/t_s=k$ and N_b is equal to the total number of background counts from all available replicate background measurements. For example, for decision rule (65), we obtain:

$$L_c = (N_s - \frac{N_b}{k})_c = d\left(\frac{1}{k}-1\right) + \frac{z_\alpha^2}{4}\left(1+\frac{1}{k}\right) + z_\alpha\sqrt{(N_b+d)\frac{1}{k}\left(1+\frac{1}{k}\right)}$$

$$L_c = (N_s - \overline{N}_b)_c = d\left(\frac{1}{k}-1\right) + \frac{z_\alpha^2}{4}\left(1+\frac{1}{k}\right) + z_\alpha\sqrt{\left(\overline{N}_b+\frac{d}{k}\right)\left(1+\frac{1}{k}\right)}$$

(89)

With increasing k (e.g., >15-20; well-known background), it becomes possible to use less complex rules (9) or (33) from Part I of this review (see also Refs. 1-3 and 40) or rule based the numerical approximation of p-values in simple one-dimensional realisation (eqn 34), possibly in combination with Lancaster's mid-p correction to improve the robustness. Analysing the size, power, and overall practical applicability of such solutions, especially at small count numbers, could be a topic of future metrological research in the area of SIMS.

An other technique where weak Poisson-distributed background noise is sometimes encountered is noble gas mass spectrometry (NG-MS). In particular, remarkably low helium blanks have been reported over the last few years[51,52]. Signal quantification in NG-MS usually invokes extrapolating a time-resolved signal back to inlet to account for consumption of the gas by ionisation during measurement[53]. However, for weak signals and backgrounds, the average of peak intensities or total count number are also employed: the individual values in a time-resolved acquisition 'show so wide a distribution that it is impossible to determine a trend'[54]. Then, unbiased decision rules from Part II of this text can be recommended, if a decision regarding detection needs to be made.

Besides, it could be helpful to use tests for the equality of two Poisson means for the treatment of some data obtained by traditional solution nebulisation (SN-) ICPMS. Background count numbers can be fairly small again; by far the commonest approach to the estimation of critical values in SN-ICPMS is, however, to replicate background measurements. Since collecting many background measurements before each sample measurement is practically unreasonable / unrealisable, the same 'pack' of background acquisitions is usually applied to a series of consecutively measured sample solutions. Thus, the system can suffer from contamination from the samples already measured (as well as from the

standards). Performing paired measurements in the solution nebulisation regime can thus be appealing, especially if samples with very different analyte contents are analysed in the same batch.

2.12. Metrological concepts associated with or complementary to the concept of detection

If the signal activity is detected using a well-constructed decision rule, then two further questions can be posed: (i) how much the gross signal activity is higher than the background activity and (ii) what is the precision of the resulting net activity? The first of these questions is part of the theory of statistical hypotheses testing as it is described in the previous sections of this text. The condition to be tested is, however, modified. Instead of testing (in the intensity-based notation) if, on the average, $I_s - I_b > 0$, the condition $I_s - I_b > \delta$ is tested. Here, δ denotes a positive value set by the analyst. The corresponding decision rules can be found in the literature[8,19,44]; they are somewhat more complex compared to the rules formulated for $\delta = 0$ and reviewed in this text, but constructed using exactly the same principles. The second question is part the theory of error propagation. It is practically important: net activity estimates, especially close to the critical level, are imprecise, and estimating the corresponding confidence intervals is essential in order to avoid data misinterpretation. In the majority of cases, the confidence interval needs to be given not only for net counts (per second) but also for the analysed element content. In the case of ordinary Poisson distributed data, this usually implies calculating the net intensity uncertainties of the analyte and the internal standard using an error propagation scheme frequently employed in the previous sections of this text $[s(I_{net}) = \sqrt{(s^2(I_s) + s^2(I_b)} = \sqrt{(I_s/t_s + I_b/t_b)}]$. Then, the net uncertainties are propagated according to eqn (82) along with the estimated uncertainty for the internal standard content in the sample and uncertainty of the relative sensitivity factor. This algorithm, as well as alternative solutions valid also for doubly stochastic Poisson distributed data are analysed in several general texts[20,33] and articles dedicated to specific questions[34,35] of the LA-ICPMS data treatment.

Conclusions and recommendations

All decision rules ever used in the LA-ICPMS practice are related to the $\sqrt{2N_b}$ rule, which is inadequate for the treatment of true paired measurements by the basic principles of its construction (biased estimation of the net activity standard deviation in the absence of replication) and by its performance (large liberality). Unfortunately, the $\sqrt{2N_b}$ rule has been popularised in the IUPAC texts[1,2] and in an earlier review by L. Currie[3], on which some of these texts are based. Admittedly, the IUPAC provides recommendations that could be most readily understood by a practically minded analyst (e.g., based on the concept of replicate paired measurements). On the negative side, the use the $\sqrt{2N_b}$ rule in the absence of replication, and at small counts, where it is very liberal, has essentially been left at the analyst's discretion. It should also be noted that more performant decision rules, such as the $\sqrt{N_s + N_b}$ rule and the mid-p adjusted version of the binomial rule, were already available at the beginning of the 1960th, but did not find their way to the IUPAC texts. The different scientific disciplines treated this situation differently. In the field of radioactivity monitoring / protection, the use of the $\sqrt{2N_b}$ rule for weak backgrounds has been disapproved[5,6]. In disciplines more distant from analytical chemistry, such as mathematical statistics and biometrics, this rule does not seem to have ever been used. We also do not recommend using this rule in the range of small count numbers, though we admit that its deficiencies become less acute at larger count numbers and when the background is measured longer than the signal. Based on the evaluation of decision rules included in this review, among decision rules based on the Poisson-Gauss approximation, we especially recommend using rules based on the square root transform (eqn 49-52,65). The corresponding formula, in the intensity based notation, for any t_b/t_s ratio, at a frequently used z_a value of 1.645 (5% of false positives) is as follows:

$$L_c = (I_s - I_b)_c = d\left(\frac{1}{t_b} - \frac{1}{t_s}\right) + 0.677\left(\frac{1}{t_s} + \frac{1}{t_b}\right) + 1.645\sqrt{(I_b t_b + d)\frac{1}{t_b}\left(\frac{1}{t_s} + \frac{1}{t_b}\right)} \quad (90)$$

This recommendation agrees with earlier recommendations made in texts of radioactivity monitoring[5,6], where this formula is appreciated for its robust behaviour in a wide range of mean background count numbers and t_b/t_s ratios. It also agrees with recommendations given in texts of mathematical statistics and biometrics[18,19], where, in addition to the robust behaviour, it is approved for its relatively high power[19]. The value of d can be set either to 3/8 or to 0.4, it does not seem that either of these values is clearly preferable. In our practice, we usually set d to 0.4 (Stapleton's version), as the actual size of formula (90) is somewhat better controlled in the practically important range of $t_b/t_s \leq 3$, provided

this value is employed. A simpler version using $d=0$ can only be approved at $\overline{N}_b t_b / t_s > 5$. The reader interested in the estimation of the detection limit for decision rules based on the square root transform will find the relevant formulae in Ref. 6 (p. 58-61). Some special cases are described in Fig. 11a.

A viable alternative to decision rules based on the square root transform is the mid-p adjusted version of the binomial decision rule[19,45,46]. It is slightly more conservative and less powerful, however, and less computation friendly. Several other potent solutions exist among decision rules based on the numerical approximation of p-values[19,44]. The volume and the merely educational orientation of this text forces us to leave these relatively complex and computationally intensive solutions beyond its scope.

For strong backgrounds that cannot be described by the ordinary Poisson process, using eqn (73) or its unbiased counterpart (78) can be recommended. The decision to use one of these equations instead of eqn (90) can be taken based on the results of a goodness-of-fit test (see Electronic Appendix 1), or when the background intensity reaches some (safely set) threshold, e.g., 500-1000 cps. In all cases, the Gaussian quantile (z_a) needs to be set to the same value (the value of 1.645 being largely preferred in the literature[1-6] and most extensively studied regarding the size and power of detection).

A note of caution is warranted. A good understanding of the computational aspects of chemical analysis is important. Although it does dot fully protect against using, interpreting, and publishing analyte abundances that correspond to fluctuations of the background noise in the absence of analyte in the sample, it helps to reduce the rate of such errors (false positive errors) to a desired level. However, reporting a positive analyte activity in a sample, when there is none, is also possible without the aid of a decision rule, especially when samples with very small analyte abundances are analysed.

Then, any contamination mobilised during the analysis influences its result. One example is the mineral olivine (Mg_2SiO_4, orth.) and the analysis of the light rare earth elements (LREE, e.g., La, Ce, Pr) in it by LA-ICPMS. The natural abundances of these elements in olivine can vary depending on the geological conditions; in most cases, they are very low, as the crystal structure of olivine is incompatible with their presence. At the same time, olivine often contains alteration products and small inclusions that can be relatively rich in the LREE. Sampling the bulk of olivine, its alteration products and inclusions during the ablation can lead to detection of a light rare earth element that is mostly concentrated in the alteration products and inclusions. To circumvent this obstacle, two approaches can be employed. The first approach requires that the background activity is acquired from a gas blank (as it is usually done in LA-ICPMS), but the sample is thoroughly studied before analysis to avoid the contaminated material and, of course, the sample surface is cleaned. The second approach requires that the background activity is acquired by analysing a 'no-analyte' sample; in the present case, it represents olivine grains, with all their inclusions and alterations, but devoid of LREE in the crystal structure of the olivine itself. Neither of these approaches can be easily employed in practice (some LA-ICPMS techniques, such as the analysis of lithium borate glass beads[55], still allow for an uncomplicated preparation of a surrogate 'no-analyte sample' made of 100% lithium borate).

The adequacy of the gas blank for the background characterisation in real LA-ICPMS measurements can also be questioned in the context of spectral (and possibly non-spectral) interferences related to the composition of the sample matrix: background count numbers obtained from the gas blank can thus be underestimated, resulting in an underestimation of the corresponding critical values and detection limits. This issue is of general significance and applies to the detectability and accuracy of many trace-level analytical results obtained by LA- and SN-ICPMS, as well as by ICP-OES, LIBS, and SIMS.

We refrain from further generalisations. In LA-ICPMS, there are isotopes affected by spectral interferences, but there are many isotopes and practically important matrices showing no such interferences even in the most common mode of instrument operation (low resolution; collision/reaction cells and MS/MS mode not available or not used). For example, in geological applications, where LA-ICPMS excels, there are minerals (e.g., garnet, zircon, orthopyroxene, olivine) depleted in LREE. The REE menu analysed in these minerals does not show any practically significant spectral overlaps, while the analysis of their REE contents is of crucial importance for geology. Obviously, spectral interferences are specific to the isotope and matrix studied, and the decision of whether the gas blank is a good

proxy for a given isotope needs to be made on an individual basis. There are computer programmes (e.g., ICPMS interference simulator from Thermo Scientific[56]) helping to take this decision rationally.

It is worth mentioning that an additional yet complementary part of the 'gas blank' issue exists: under-estimating the blank count number results in a systematic positive error in the analyte content. Mathematical corrections of spectral interferences are rarely used in the modern practice of multi-elemental LA-ICPMS and are usually unavailable in trace element data reduction programmes. So, we would not only need to refrain from estimating critical values (detection limits), but also to stop producing trace element data or analysing ever-smaller targets by LA-ICPMS because the results could be positively biased. We are convinced that this decision must be taken individually for each of the studied isotopes, and where there are issues - instrumental solutions need to be found to supress the interference(s) as much as possible. It also seems likely that we will see a wider use of mathematical corrections of spectral interferences in the practice of trace element LA-ICPMS in the future (for a brief discussion of modified versions of decision rules for paired measurements that include such corrections, see Ref. 6).

As an ultimate solution, using a matrix-matched material devoid of the analyte ('no-analyte' sample) for the analysis of real samples seems feasible in some cases. It is possible that we will see more development in this direction in the future, in which case the mathematical methods reviewed in this text will suffer no alteration. The only alteration will be to compare count numbers acquired from a no-analyte material, and not from a gas blank, to those acquired from a sample.

Thus, we believe that the existence of interferences and possible sample contamination are no excuse not to practice (ultra-) trace element spectrometry. We admit and emphasise that care should be exercised regarding the signal source in samples analysed by LA-ICPMS, even if the signal is detected by using a well-constructed decision rule: 'The fascinating impressiveness of rigorous mathematical analysis, with its atmosphere of precision and elegance, should not blind us to the defects of the premises that condition the whole process.'[57] This does not mean that using poorly performing decision rules can be encouraged: such rules do not allow deciding if a signal is detected at a well-defined confidence level irrespective of the source of this signal.

It is obvious that the continuous attempt of the LA-ICPMS community to analyse ever-smaller isotope abundances includes several (partly) independent aspects and requires: (i) using well-constructed decision rules, (ii) investigating the material to be analysed for contamination and interferences and (iii) increasing the sensitivity of LA-ICPMS instruments and, perhaps, further supressing their background noise. A lack of progress in any of these areas hampers the user of LA-ICPMS to fully benefit from the (ultra-) trace element capabilities of this technique. The area covering the quantification of LA-ICPMS detection capabilities has received relatively little attention for two decades, hence this review.

Examples

Electronic Appendix 8 contains four examples of LA-ICPMS background measurements covering a range of intensities. For each of these examples, critical levels are estimated and conclusions provided.

Acknowledgements

References

1   L.A. Currie, Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC recommendations 1995), *Pure Appl. Chem.*, 1995, **67**, 1699-1723.

2   IUPAC publications, Chapter 18. Quality assurance of analytical processes, Section 18.4.3.7. Detection and Quantification capabilities, http://iupac.org/publications/analytical_compendium, consulted on the 21st of May 2015.

3   L.A. Currie, Limits of qualitative detection and quantitative determination, Application to Radio-chemistry, *Anal. Chem.*, 1968, **40**, 586-593.

4   B. Altshuler and B. Pasternack, Statistical measures of the lower limit of detection of a radioactivity counter. *Health Phys.*, 1963, **9**, 293-298.

5   D.J. Strom and J.A. MacLellan, Evaluation of eight decision rules for low-level radioactivity counting, *Health Phys.*, 2001, **81**, 27-34.

6   MARLAP Multi-Agency Radiological Laboratory Analytical Protocol Manual, vol. 19-20, version from July 2004, http://www.epa.gov/radiation/marlap/manual.html, consulted on the 21st of May 2015.

7   H.K.T. Ng and M.L. Tang, Testing the equality of two Poisson means using the rate ratio, *Stat. Med.*, 2005, **24**, 955-965.

8   H.K.T. Ng, K. Gu and M.L. Tang, A comparative study of tests for the difference of two Poisson means, *Comp. Stat. Data Anal.*, 2007, **51**, 3085-3099.

9   L.A. Currie, Detection and quantification limits: origins and historical overview, *Anal. Chim. Acta*, 1999, **391**, 127-134.

10 K.A. Brownlee, *Statistical theory and methodology in science and engineering*, John Wiley & Sons, New York, 1965, p. 590.

11 A. Hald, Statistical theory with engineering applications, John Wiley & Sons, New York, 1967, p. 783.

12 H. Kaiser, Die Berechnung der Nachweisempfindlichkeit, *Spectrochimica Acta*, 1947, **3**, 40-67.

13 H.P. Longerich, S.E. Jackson and D. Günther, Laser ablation inductively coupled plasma mass spectrometric transient signal data acquisition and analyte concentration calculation, *J. Anal. At. Spectrom.*, 1996, **11**, 899-904.

14 T. Pettke, F. Oberli, A. Audétat, M. Guillong, A.C. Simon, J.J. Hanley and L.M. Klemm, Recent developments in element concentration and isotope ratio analysis of individual fluid inclusions by laser ablation single and multiple collector ICP-MS, *Ore Geol. Rev.*, 2012, **44**, 10-38.

15 J. Przyborowski and H. Wilenski, Homogeneity of results in testing samples from Poisson series, *Biometrika*, 1940, **31**, 313-323.

16 W.L. Nicholson, Statistics of net-counting-rate estimation with dominant background corrections, *Nucleonics*, 1966, **24**, 118-121.

17 W.L. Nicholson, Fixed time estimation of counting rates with background corrections. AEC R&D Report HW-76279, Hanford Works, Richland, Washington, 1963.

18 M.D. Huffman, An improved approximate two-sample Poisson test, *Appl. Stat.*, 1984, **33**, 224-226.

19 K. Gu, H.K.T. Ng, M.L. Tang and W.R. Schucany, Testing the ratio of two Poisson rates, *Biom. J.*, 2008, **50**, 283-298.

20 A. Ulianov, O. Müntener and U. Schaltegger, The ICPMS signal as a doubly stochastic Poisson process, *J. Anal. At. Spectrom.*, 2015, **30**, 1297-1321.

21 H.H. Barrett and K.J. Myers, *Foundations of Image Science*, John Wiley & Sons, Hoboken, New Jersey, 2004, p. 1540.

22 L.J. Rainwater and C.S. Wu, Applications of probability theory to nuclear particle detection, *Nucleonics*, 1947, **1**, 60-69.

23 N. Furuta, C.A. Monnig, P. Yang and G.M. Hieftje, Noise characteristics of an inductively coupled plasma-mass spectrometer, *Spectrochim. Acta, Part B*, 1989, **44**, 649-656.

24 P.K. Appelblad, I. Rodushkin and D.C. Baxter, Sources of uncertainty in isotope ratio measurements by inductively coupled plasma mass spectrometry, *Anal. Chem.*, 2001, **73**, 2911-2919.

25 D.C. Baxter, I. Rodushkin and E. Engström, Isotope abundance ratio measurements by inductively coupled plasma-sector field mass spectrometry, *J. Anal. At. Spectrom.*, 2012, **27**, 1355–1381.

26 E.D. Prudnikov and R.M. Barnes, Theoretical calculation of the standard deviation in inductively coupled plasma mass spectrometry, *J. Anal. At. Spectrom.*, 1999, **14**, 27-31.

27 M.P. Goudzwaard and M.T.C. de Loos-Vollebregt, Characterisation of noise in inductively coupled plasma - atomic emission spectrometry, *Spectrochim. Acta, Part B*, 1990, **45**, 887-901.

28 R. Antanavičius, P. Serapinas and P. Šimkus, Observation of the frequency dependence of ICP radiation noise, *J. Phys. D: Appl. Phys.*, 1989, **22**, 254-257.

29   J.W. Olesik, L.J. Smith and E.J. Williamsen, Signal fluctuations due to individual droplets in inductively coupled plasma atomic emission spectrometry. *Anal. Chem.*, 1989, **61**, 2002-2008.

30   S.E. Hobbs and J.W. Olesik, Inductively coupled plasma mass spectrometry signal fluctuations due to individual aerosol droplets and vaporizing particles, *Anal. Chem.*, 1992, **64**, 274-283.

31   J.W. Olesik, Investigating the fate of individual sample droplets in inductively coupled plasmas, *Appl. Spectrosc.*, 1997, **51**, 158A-175A.

32   D.R. Cox and V. Isham, *Point processes*, London-New York-Washington, Chapman & Hall, 1980, p. 188.

33   Y. Luo, S. Gao, H.P. Longerich, D. Günther, S. Wunderli, H.L. Yuan and X.-M. Liu, The uncertainty budget of the multi-element analysis of glasses using LA-ICP-MS, *J. Anal. At. Spectrom.*, 2007, **22**, 122-130.

34   C.M. Fisher, H.P. Longerich, S.E. Jackson and J.M. Hanchar, Data acquisition and calculation of U–Pb isotopic analyses using laser ablation (single collector) inductively couple plasma mass spectrometry, *J. Anal. At. Spectrom.*, 2010, **25**, 1905-1920.

35   A. Ulianov and O. Müntener, Differencing as a method to estimate the uncertainty of a transient LA-ICPMS signal, *J. Anal. At. Spectrom.*, 2014, **29**, 934-940.

36   D.R. Cox and P.A.W. Lewis, *The statistical analysis of series of events*, Methuen&Co., London, 1968, p. 285.

37   R.A. Fisher, The logical inversion of the notion of the random variable, *Sankhyā*, 1945, **7**, 129-132.

38   E. S. Pearson, On questions raised by the combinations of tests based on discontinuous distributions, *Biometrika*, 1950, **37**, 383-398.

39   M. Tanner and D. Günther, Short transient signals, a challenge for inductively coupled plasma mass spectrometry, a review, *Anal. Chim. Acta*, 2009, **633**, 19-28.

40   L.A. Currie, On the detection of rare, and moderately rare, nuclear events, *J. Radioanal. Nucl. Chem.*, 2008, **276**, 285-297.

41   C. Han, Comparing two independent incidence rates using conditional and unconditional exact tests, *Pharm. Stat.*, 2008, **7**, 195-201.

42   M.S. Bartlett, The square root transformation in the analysis of variance, *J. R. Statist. Soc. Suppl.*, 1936, **3**, 68-78.

43   F.J. Anscombe, The transformation of Poisson, binomial and negative-binomial data, *Biometrika*, 1948, **35**, 246-254.

44   K. Krishnamoorthy and J. Thomson, A more powerful test for comparing two Poisson means, *J. Stat. Plan. Inf.*, 2004, **119**, 23-35.

45   H.O. Lancaster, Statistical control of counting experiments, *Biometrika*, 1952, **39**, 419-422.

46   H.O. Lancaster, Significance tests in discrete distributions, *J. Am. Stat. Assoc.*, 1961, **56**, 223-234.

47   M. Tanner, Shorter signals for improved signal to noise ratio, the influence of Poisson distribution, *J. Anal. At. Spectrom.*, 2010, **25**, 405–407.

48   W.A. Fuller, *Measurement Error Models.* John Wiley & Sons, New York, 1987, p. 440.

49   H.W.Werner, New developments in secondary ion mass spectrometry, in *Applied surface analysis,* ed. T.L. Barr and L.E. Davis, Am. Soc. Test. Mater., Spec. Tech. Publ. 699, 1980, pp. 81-110.

50   C.W. Fitzsimons, B. Harte and R.M. Clark, SIMS stable isotope measurement: counting statistics and analytical precision, *Mineral. Mag.*, 2000, **64**, 59-83.

51   L. Zimmermann, P.H. Blard, P. Burnard, S. Medynski, R. Pik and N. Puchol, A new single vacuum furnace design for cosmogenic ^{3}He dating, *Geostand. Geoanal. Res.*, 2012, **36**, 121-129.

52   S. Assonov and T. Prohaska, Gas source isotope ratio mass spectrometry for the analysis of noble gases, in *Sector Field Mass Spectrometry for Elemental and Isotopic Analysis,* ed. T. Prohaska, J. Irrgeher, A. Zitek and N. Jakubowski, The Royal Society of Chemistry, London, 2014, pp. 550-581.

53   P. Burnard, L. Zimmermann and Y. Sano, The noble gases as geochemical tracers: History and background, in *The Noble Gases as Geochemical Tracers*, ed. P. Burnard, Springer Verlag, Berlin-Heidelberg, 2013, pp. 1-15.

54   H. Tamura, K. Sato, T. Hanyu and H. Kumaga, New noble gas isotope analysis system: design and implementation, IFREE Report for 2003–2004, 2005, **2**, 1–5.

55   S.M. Eggins, Laser ablation ICP-MS analysis of geological materials prepared as lithium borate glasses, *Geostand. Geoanal. Res.*, 2003, **27**, 147-162.

56   R. Pesch, Thermo Finnigan Interference Workshop, software version 2.5, 2001, Finnigan MAT GmbH, Bremen, Germany.

57 T.C. Chamberlin, Lord Kelvin's address on the age of the Earth as an abode fitted for life. *Science*, 1899, **9**, 889-901.

Table 1. Critical levels and actual probabilities of false positive errors for the case of well-known background (a number of replicates is available to characterise the mean and standard deviation of the background noise). Both canonical and discontinuity-corrected versions are described. k_a is set to 1.645 (5% nominal rate of false positives).

N_b	$L_c=1.645\sqrt{N_b}$			$L_c=1/2+1.645\sqrt{N_b}$		
mean	L_c	$N_{s\ gross}$*	$p(N) > N_{s\ gross}$&	L_c	$N_{s\ gross}$*	$p(N) > N_{s\ gross}$&
1	1.645	3	0.0803	2.145	3	0.0190
2	2.326	5	0.0527	2.826	5	0.0527
3	2.849	6	0.0839	3.349	6	0.0335
5	3.678	9	0.0681	4.178	9	0.0318
7	4.352	12	0.0533	4.852	12	0.0533
10	5.202	16	0.0487	5.702	16	0.0487
20	7.357	28	0.0525	7.857	28	0.0525

*min gross signal count number that is considered as detected; &probability to commit a false positive error (i.e., to obtain a count number value > min detected gross signal).

Table 2. Critical levels and actual probabilities of false positive errors for the $\sqrt{2N_b}$ rule (background and signal of equal duration). For the sake of illustration, the mean background count number value is set to 1.5 counts. Three cases are considered: replicate paired measurements, true paired measurements, and true paired measurements modified by substituting a zero background estimate by one count. k_a is set to 1.645 (5% nominal rate of false positives).

N_{b1}	$p(N_{b1})$	L_c	$p(N_{b2})$										probability of false detection
			$p(0)$	$p(1)$	$p(2)$	$p(3)$	$p(4)$	$p(5)$	$p(6)$	$p(7)$	$p(8)$	$p(9)$, etc.	
replicate paired measurements: computing L_c from a well-known value of the standard deviation of the net background count number ($L_c=1.645\sqrt{3}$)													
0	0.2231	2.849	0.2231	0.2231	0.2231	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**	… …	$p(0)$ x $(p(3)+p(4)+\dots)$=**0.0426**
1	0.3347	2.849	0.2231	0.2231	0.2231	0.2231	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**		$p(1)$ x $(p(4)+p(5)+\dots)$=**0.0220**
2	0.2510	2.849	0.2231	0.2231	0.2231	0.2231	0.2231	**0.2231**	**0.2231**	**0.2231**	**0.2231**		$p(2)$ x $(p(5)+p(6)+\dots)$=**0.0046**
3	0.1255	2.849	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	**0.2231**	**0.2231**	**0.2231**		$p(3)$ x $(p(6)+p(7)+\dots)$=**0.0006**
4	0.0471	2.849	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	**0.2231**	**0.2231**		$p(4)$ x $(p(7)+p(8)+\dots)$=**0.0000**
5	0.0141	2.849	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	**0.2231**		$p(5)$ x $(p(8)+p(9)+\dots)$=**0.0000**
6	0.0035	2.849	…									… … …	…
…		…										… … …	**0.0698**
true paired measurements: computing L_c from one single background count number estimate ($L_c=1.645\sqrt{2N_{b1}}$)													
0	0.2231	0.000	0.2231	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**	… …	$p(0)$ x $(p(1)+p(2)+\dots)$=**0.1733**
1	0.3347	2.326	0.2231	0.2231	0.2231	0.2231	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**		$p(1)$ x $(p(4)+p(5)+\dots)$=**0.0220**
2	0.2510	3.290	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	**0.2231**	**0.2231**	**0.2231**		$p(2)$ x $(p(6)+p(7)+\dots)$=**0.0011**
3	0.1255	4.029	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	**0.2231**		$p(3)$ x $(p(8)+p(9)+\dots)$=**0.0000**
4	0.0471	4.653	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231		…
5	0.0141	5.202	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231		…
6	0.0035	5.698	…									… … …	…
…	…	…										… … …	**0.1964**
true paired measurements: computing L_c from one single background count number estimate ($L_c=1.645\sqrt{2N_{b1}}$), with replacing a zero background estimate by unity													
1→0	0.2231	2.326	0.2231	0.2231	0.2231	0.2231	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**	… …	$p(0)$ x $(p(4)+p(5)+\dots)$=**0.0146**
1	0.3347	2.326	0.2231	0.2231	0.2231	0.2231	**0.2231**	**0.2231**	**0.2231**	**0.2231**	**0.2231**		$p(1)$ x $(p(4)+p(5)+\dots)$=**0.0220**
2	0.2510	3.290	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	**0.2231**	**0.2231**	**0.2231**		$p(2)$ x $(p(6)+p(7)+\dots)$=**0.0011**
3	0.1255	4.029	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	**0.2231**		$p(3)$ x $(p(8)+p(9)+\dots)$=**0.0000**
4	0.0471	4.653	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231		...
5	0.0141	5.202	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231	0.2231		…
6	0.0035	5.698	…									… … …	…
…	…	…										… … …	**0.0625**

Table 3. Actual probabilities of false positive errors for the canonical version of the binomial rule (background and signal of equal duration). For the sake of illustration, the mean background count number value is set to 2 counts (4 counts for a sum of two independent background estimates). The nominal rate of false positives is set to 5%.

$N_{b1}+N_{b2}$	$p(N_{b1}+N_{b2})$	$1/2^n$	$p(0,n)$	$p(1,n-1)$	$p(2,n-2)$	$p(N_{b1},N_{b2})$ ...	...	...	...	...	probability of false detection
0	0.0183	1.00000	1.0000 (1)								
1	0.0733	0.50000	0.5000 (1)	0.5000 (1)							
2	0.1465	0.25000	0.2500 (1)	0.5000 (2)	0.2500 (1)						
3	0.1954	0.12500	0.1250 (1)	0.3750 (3)	0.3750 (3)	0.1250 (1)					
4	0.1954	0.06250	0.0625 (1)	0.2500 (4)	0.3750 (6)	0.2500 (4)	0.0625 (1)				
5	0.1563	0.03125	**0.0313 (1)**	0.1563 (5)	0.3125 (10)	0.3125 (10)	0.1563 (5)	0.0313 (1)			$p(5) \times p(0,5)=0.0049$
6	0.1042	0.01563	**0.0156 (1)**	0.0938 (6)	0.2344 (15)	0.3125 (20)	0.2344 (15)	0.0938 (6)	0.0156 (1)		$p(6) \times p(0,6)=0.0016$
7	0.0595	0.00781	**0.0078 (1)**	0.0547 (7)	0.1641 (21)	0.2734 (35)	0.2734 (35)	0.1641 (21)	0.0547 (7)	...	$p(7) \times p(0,7)=0.0005$
8	0.0298	0.00391	**0.0039 (1)**	**0.0313 (8)**	0.1094 (28)	0.2188 (56)	0.2734 (70)	0.2188 (56)	0.1094 (28)		$p(8) \times (p(0,8)+p(1,7))=0.0011$
9	0.0132	0.00195	**0.0020 (1)**	**0.0176 (9)**	0.0703 (36)	0.1641 (84)	0.2461 (126)	0.2461 (126)	0.1641 (84)		$p(9) \times (p(0,9)+p(1,8))=0.0003$
10	0.0053	0.00098	**0.0010 (1)**	**0.0098 (10)**	0.0439 (45)	0.1172 (120)	0.2051 (210)	0.2461 (252)	0.2051 (210)		$p(10) \times (p(0,10)+p(1,9))=0.0001$
11	0.0019	0.00049	**0.0005 (1)**	**0.0054 (11)**	**0.0269 (55)**	0.0806 (165)	0.1611 (330)	0.2256 (462)	0.2256 (462)		$p(11) \times (p(0,11)+p(1,10)+p(2,9))=0.0001$
12	0.0006	0.00024	**0.0002 (1)**	**0.0029 (12)**	**0.0161 (66)**	0.0537 (220)	0.1208 (495)	0.1934 (792)	0.2256 (924)	...	$p(12) \times (p(0,12)+p(1,11)+p(2,10))=0.0000$
...	...	...	...							...	**0.0086**

Figure captions

Fig. 1. Detection in the framework of the Poisson-Gauss approximation: (a) well-known background; (b) replicate paired measurements. For the sake of illustration, the number of replicates is assumed to be large; the mean background activity is arbitrarily set at 13 counts; the model distribution is Gaussian, set at a mean of zero and at standard deviations of $\sim\sqrt{13}$ and $\sqrt{26}$ for the cases of well-known background and paired measurements, respectively. Confidence intervals corresponding to some frequently used quintiles (k_a=1; 1.645; 2; 3) are delimited. Detection is achieved, provided the net sample activity value falls in the critical region of the net values larger than L_c. Within the critical region, the nominal probability of appearance of a net background value is below some level α set by the analyst (e.g., to 5%); it is constrained by the quintile used. Thus, if a net sample activity value falls within the critical region, it is unlikely that it comes from fluctuations of the background activity: analyte is detected. If the net sample activity value falls outside the critical region (i.e., it does not surpass L_c), then the presence of analyte in the sample is still not excluded, but there is a relatively large probability that the above activity comes from fluctuations of the background activity: analyte is not detected.

Fig. 2. Critical level (L_c) and Detection limit (L_d) as two different parameters characterising detection: (a) well-known background; (b) replicate paired measurements. The number of replicates is assumed to be large; the mean background activity is arbitrarily set at 13 counts; the model distributions represent a Gauss approximation of the original Poisson statistics that are markedly heteroscedastic at small mean count number values (their standard deviation increases with increasing the mean count number)[1-3]. As a simpler model applicable at large count numbers only, the homoscedastic solution is also shown. Deciding whether the analyte activity is detected requires comparing the net sample activity value to L_c. However, for a sample, of which the mean (precisely known, 'true') activity is equal to L_c, only 50% of all measurements, or estimates, of the net activity will result in detection. The detection limit (L_d) represents a value of the mean ('true', precisely known) sample activity that ensures a reasonably high probability of detecting net sample activity values obtained from individual sample measurements. This probability is usually set at 95%. L_d is defined relative to a decision rule (that allows to decide if the net sample activity above is detected). Estimating L_c is a pre-requisite for computing L_d. Using L_d values is necessary, provided the analytical task requires a specific value of 'true' analyte activity to be detected at some level of confidence (e.g., 95%; for details, see section 1.2).

Fig. 3. The mean sample activity value estimated from replicate measurements is positively biased, provided the sample activity values below detection are excluded from its computation. An unbiased value is obtained by averaging all net sample activity values including those that are below detection (including those that are below zero). The unbiased value is approximately equal to the 'true' sample activity, the quality of this approximation improving with the increase of the number of replicates.

Fig. 4. The Poisson-Gauss approximation. (a) Well-known background. Net values obtained by subtracting the mean background activity from the individual Poisson outcomes form a skewed, discrete distribution that improves in symmetry and converges to the Gauss distribution with the increase of the mean count number value. (b) Replicate paired measurements. Net values obtained by subtracting the individual outcomes sampled from the same Poisson distribution are symmetrically centered around zero at any mean count number; the resulting distribution is sometimes called Skellam distribution. It is still discrete, but also converges to the Gauss distribution with increasing the mean count number value. This allows using the Gaussian quintiles and the associated confidence intervals for the estimation of critical levels (delimited for the case of a large number of replicates).

Fig. 5. Correcting discrete data for discontinuity. In a discrete distribution, the probability is concentrated in the individual 'probability bars', whereas in a continuous distribution, the probability density is continuously distributed. Here, for the sake of illustration, we consider an arbitrarily selected case of a well-known background based on a Poisson distribution with a mean of 12 counts that is approximated by the Gauss distribution. Accordingly, L_c=1.645x$\sqrt{12}$≈5.70 counts. This would mean that a net sample value of 6 counts is detected. However, the corresponding probability bar covers the range from 5.5 to 6.5 counts. Including this probability bar in the critical region causes the size of the critical region to exceed 5%, meaning that the decision rule behaves liberally. In the present case, the distribution step is equal to unity. Adding one-half of it to L_c allows excluding the probability bar of 6 counts

from the critical region, though the decision rule employing such a correction becomes somewhat conservative. Strictly, for an L_c value obtained using the Gauss approximation of Poisson distributed data at a declared level of false positive errors set to α, the range of Poisson outcomes with a cumulative probability not exceeding α begins from $L_c+1/2$.

Fig. 6. Actual rate of false positive errors for the $\sqrt{2N_b}$ decision rule (background and signal of equal duration). The performance of four versions of this rule is illustrated: (1) replicate paired measurements: the standard deviation of the net background activity is known precisely; (2) true paired measurements: the standard deviation of the net background activity is not known, bit assumed to be equal to $\sqrt{2N_b}$, (3) same as (2), modified by adding one count to an 'empty' background estimate without changing other (positive) background estimates; (4) randomised formulation (for details, see section 2.1). In all cases, k_a was set to 1.645 (5% declared level of false positive errors).

Fig. 7. A schematic illustration of correlation-induced skewness in the distribution of the normalised statistics associated with the $\sqrt{2N_b}$ decision rule (non-randomised formulation under the null hypothesis; see section 2.1).

Fig. 8. Testing the hypothesis of equality of two Poisson means using the $\sqrt{N_b+N_s}$ decision rule (background and signal of equal duration). The net count number estimates obtained from the individual paired measurements are distributed around zero, provided their mean is zero. The latter condition represents the null hypothesis (H_0) to be tested. H_0 is first assumed to be true. For a paired measurement, the most probable estimate of the net count number standard deviation is $\sqrt{N_b+N_s}$. We compute L_c based on the Poisson-Gauss approximation, and compare the net sample activity obtained from a paired analysis to L_c. If the net sample activity falls in the critical region of the net distribution that is centered around zero, then there is little chance to obtain it from such distribution (i.e., when H_0 holds true). We conclude that H_0 is false, which implies a positive mean net sample activity: the corresponding distribution is shifted relative to zero; obtaining the net sample activity above from such shifted distribution is more probable. Hence, the analyte activity in the sample is detected.

Fig. 9. Actual rate of false positive errors for several decision rules based on an unbiased estimation of the net count number standard deviation, in the range of t_b/t_s ratios from 1 to 5. In all cases, k_a was set to 1.645 (5% declared level of false positive errors). The most robust rules are the Poisson-Gauss approximation using the square root transform [Huffman's ($d=3/8$) and Stapleton's ($d=0.4$) versions] and the mid-p adjusted version of the binomial decision rule.

Fig. 10. Actual rate of false positive errors for the $\sqrt{2N_b}$ decision rule adapted for arbitrary t_b/t_s ratios: (a) $k_a=1.645$, or 5% declared level of false positive errors; (b) $k_a=3$, or 0.135% declared level of false positive errors. The original formulation, as well as the formulation empirically corrected for the 'empty' background are described. A liberal behaviour is observed in all cases, although the actual size becomes closer to the declared value with increasing the t_b/t_s ratio.

Fig. 11. Power of detection for two decision rules in the range of t_b/t_s ratios from 1 to 5: (a) Poisson-Gauss approximation using the square root transform [Stapleton's version ($d=0.4$)]; (b) mid-p adjusted version of the binomial rule. $\overline{N}_b(t_s/t_b)$ is fixed at 1, 5, and 10 counts. At $\overline{N}_b(t_s/t_b)=\overline{N}_s$, the diagrams return the actual rate of false positive errors (actual size) of the corresponding decision rule. At a power of 95%, they return the gross signal count numbers corresponding to the detection limit.

Fig. 12. Scaling confidence intervals for the net activity distribution constructed under the null hypothesis by the sensitivity factor (see section 2.9).

Fig. 13. A systematics of decision rules used for testing the equality of two Poisson means. For the sake of completeness, several rules (e.g., the likelihood ratio test test and the Poisson-Gauss approximation using the logarithmic transform) are included, which were omitted in the main text because of their complex construction, arbitrary solutions used in their construction, or liberal performance.
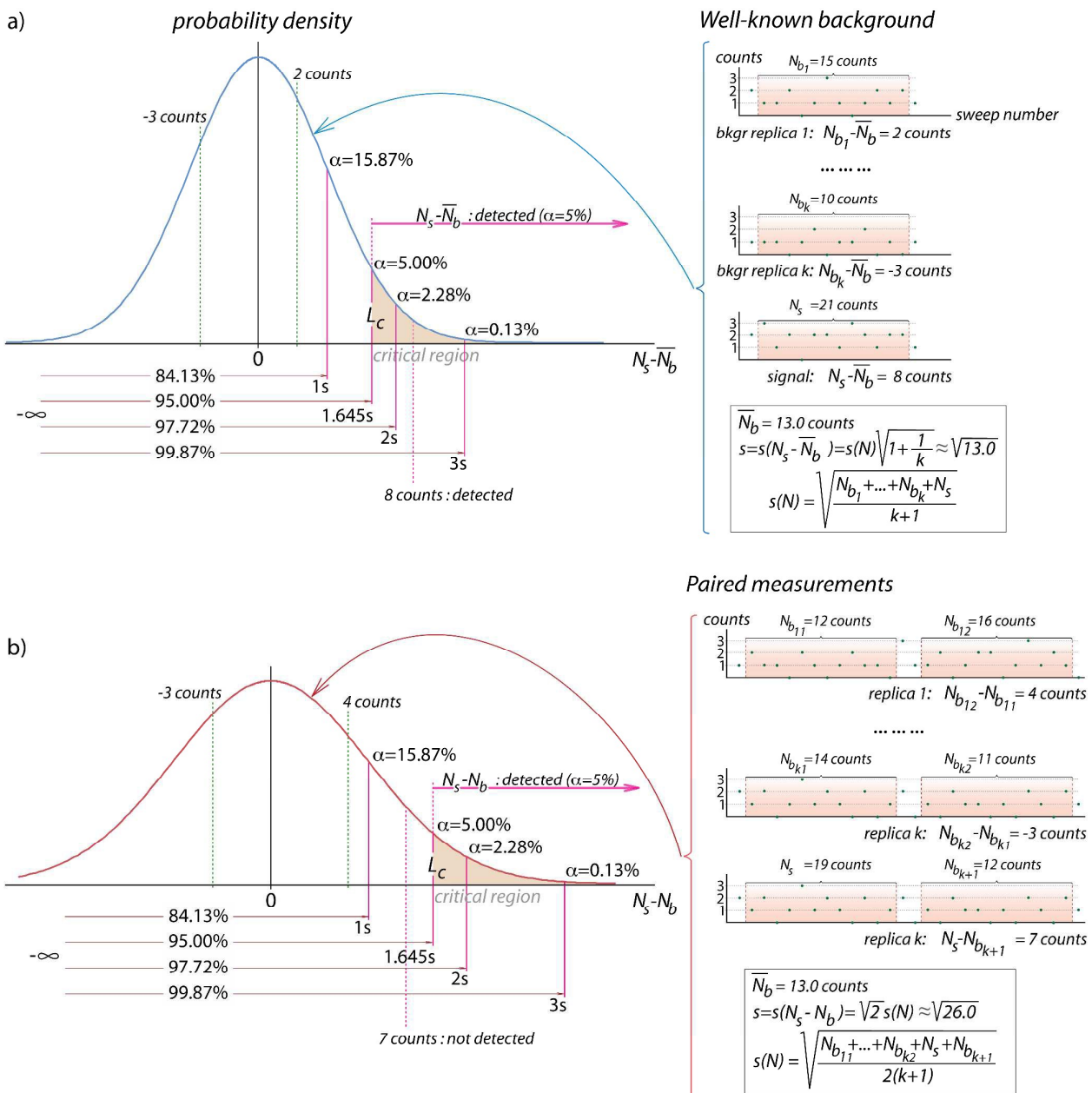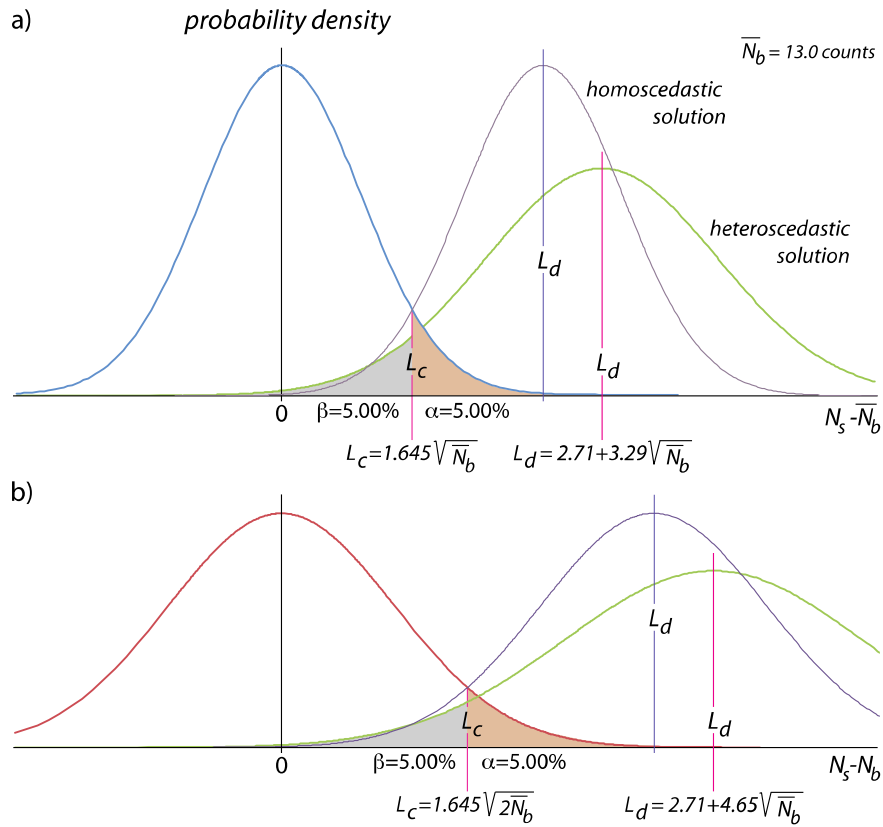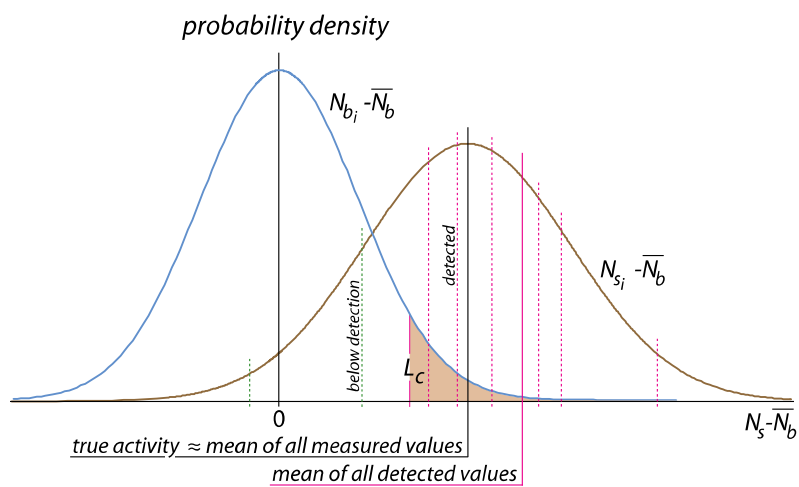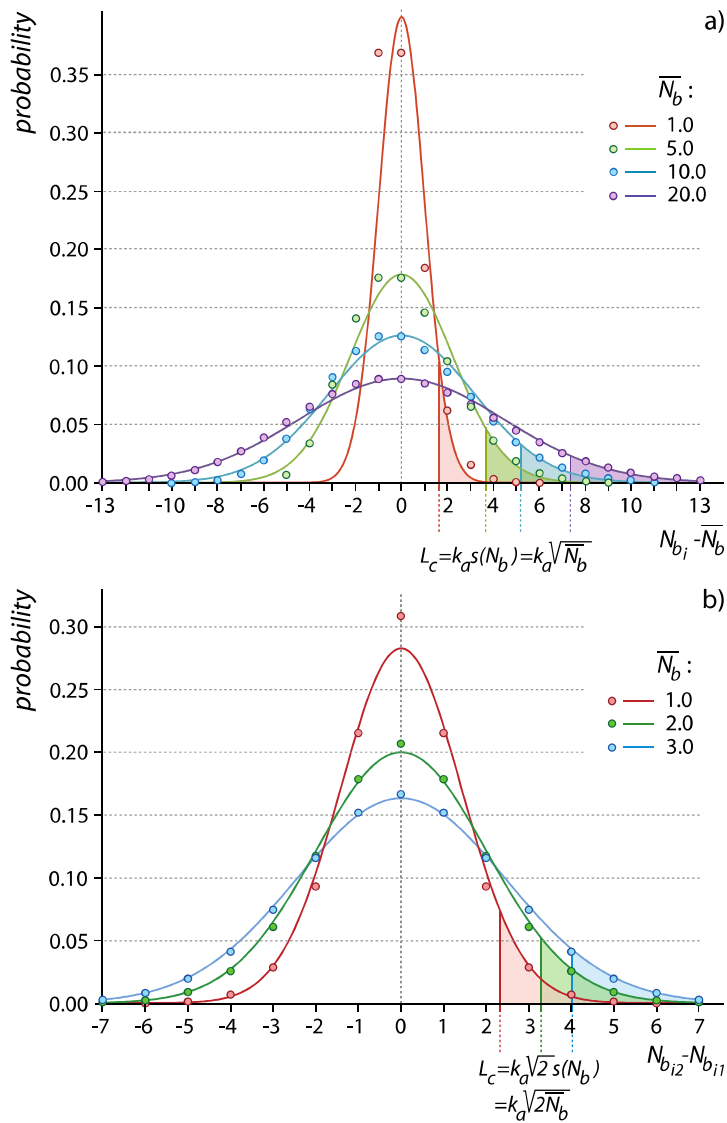
Fig. 1

a)   *probability density*

$\overline{N_b} = 13.0\ counts$

homoscedastic solution

heteroscedastic solution

L_d

L_d

L_c

0   β=5.00%   α=5.00%

$N_s - \overline{N_b}$

$L_c = 1.645\sqrt{\overline{N_b}}$ $L_d = 2.71 + 3.29\sqrt{\overline{N_b}}$

b)

L_d

L_d

L_c

0   β=5.00%   α=5.00%

$N_s - N_b$

$L_c = 1.645\sqrt{2\overline{N_b}}$ $L_d = 2.71 + 4.65\sqrt{\overline{N_b}}$

Fig. 2

probability density

$N_{b_i} - \overline{N_b}$

below detection

detected

L_c

$N_{s_i} - \overline{N_b}$

$N_s - \overline{N_b}$

0

true activity ≈ *mean of all measured values*

mean of all detected values

Fig. 3

Fig. 4

Fig. 5

Fig. 6

Fig. 7

probability density

$\overline{N_s} - \overline{N_b} = 0$

$\overline{N_s} - \overline{N_b} > 0$

$L_{C\,5\%}$

$N_s - N_b$ *in a sample*

$N_s - N_b$

Fig. 8

Fig. 9

Fig. 10

Fig. 11

Fig. 12

Decision rules used to test the equality of two Poisson means

Poisson-Gauss approximations   *numerical approximations of p-values*   *bootstraps*   *binomials*   *likelihood ratio*

$N_s\text{-}N_b / \sqrt{2N_b}$ *statistic*

$N_s\text{-}N_b / \sqrt{N_s+N_b}$ *statistic*

statistics based on transforms, e.g., on the square root transform

canonical vers.

mid-p adjusted vers.

auxiliary experiment
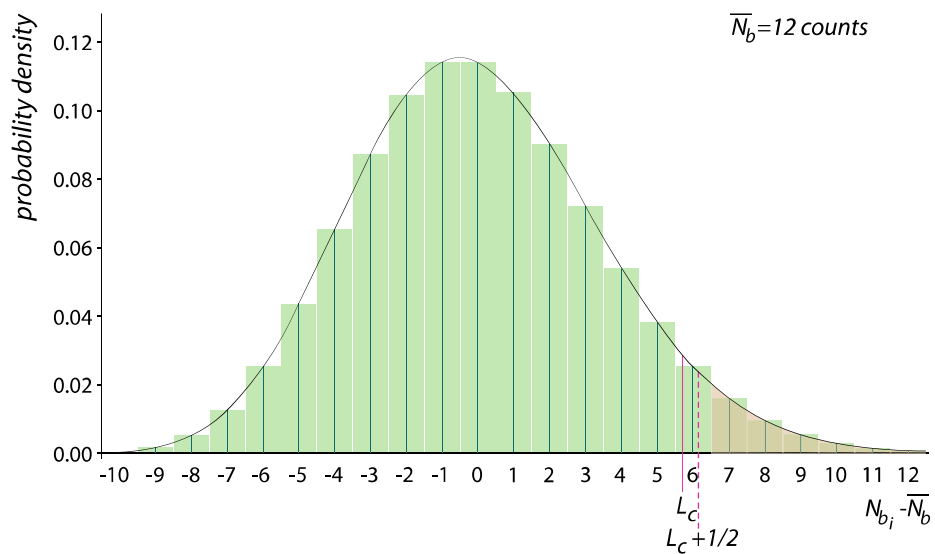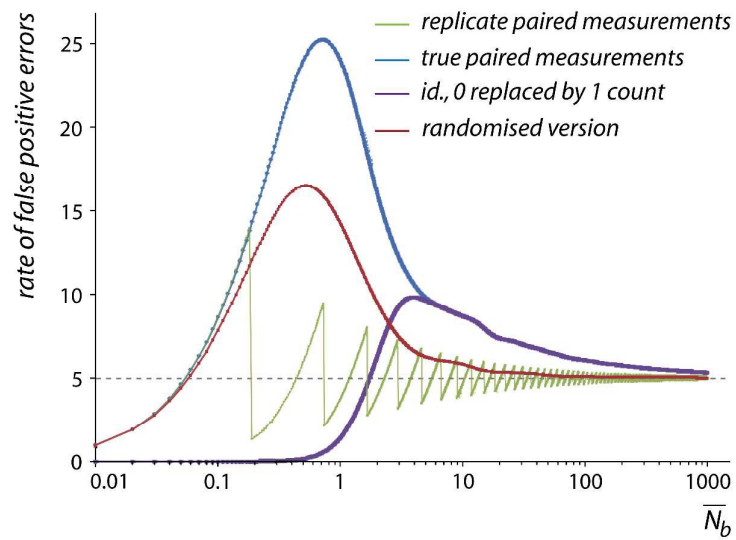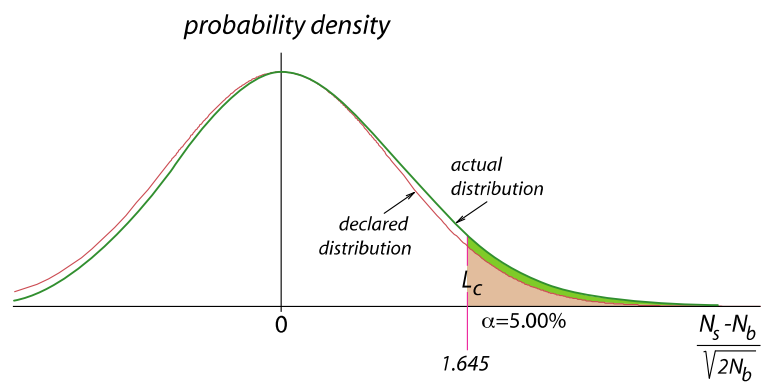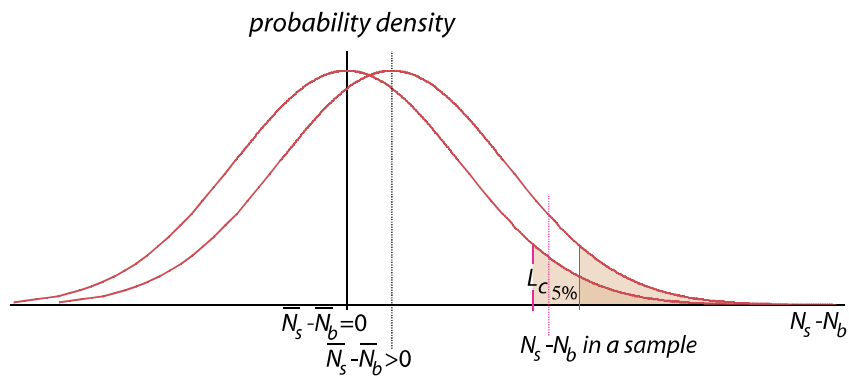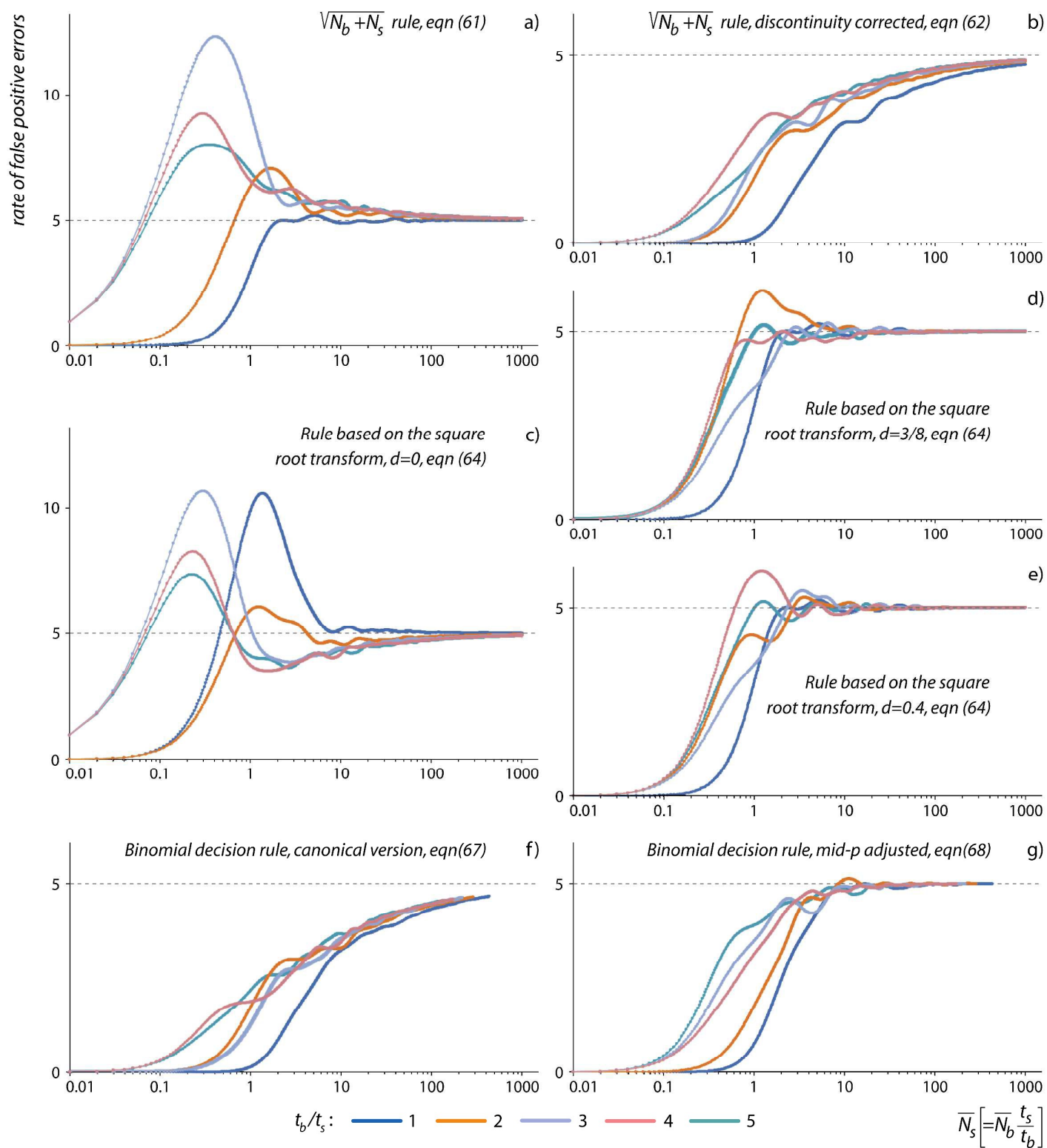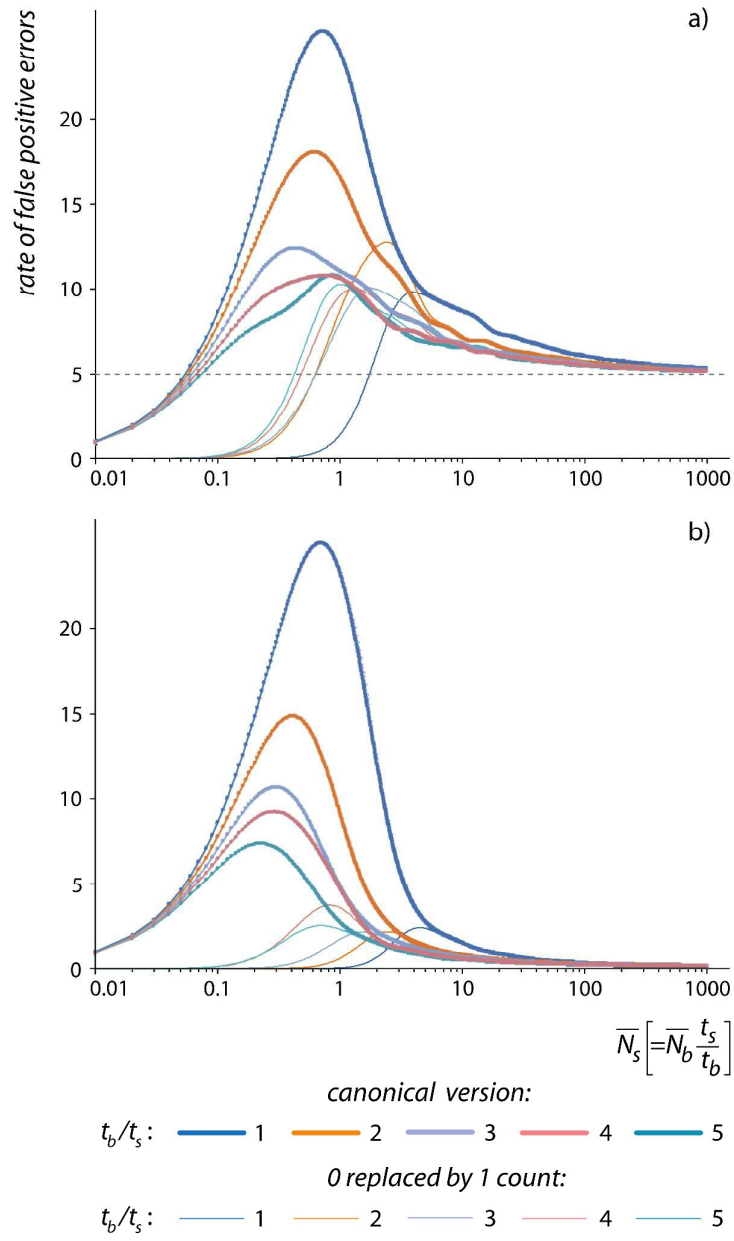
Fig. 13

The theory of tests for the equality of two Poisson means is reviewed, with a special emphasis on the detection of LA-ICPMS signals in the presence of a very weak, Poisson distributed background noise.