



Evaluation of uncertainties in atomic data on spectral lines and transition probabilities

Alexander Kramida^a

National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Received 20 December 2023 / Accepted 23 February 2024

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024

Abstract. The present article discusses some recent advances in methods of critical evaluation of experimental data on wavelengths of spectral lines and theoretical data on transition probabilities and oscillator strengths for atoms and atomic ions. In particular, recently developed new statistical approaches to estimation of uncertainties of weighted means of multiple measurements are described, and a numerical toolbox implementing these new approaches is presented. There are also some new developments in estimation of uncertainties of theoretical transition probabilities. A short review of literature implementing these new procedures is provided, including a description of the methodology.

1 Introduction

There is a distinction between the concepts of “numbers” and “data.” This is seemingly trivial, but reading current literature on atomic spectroscopy indicates that not everybody understands it. There are many properties of physical and mathematical objects that can be expressed as exact numbers. A typical example is the ratio of circumference to the diameter of a circle. It is the number π , which can be calculated with any required precision. So, we can say that it is known exactly. However, if one tries to measure circumference or diameter of any circle with available measuring tools, the result is never exact. Repeated measurements with the same instrument or with different instruments will give a set of slightly differing values. The result of such measurement can be calculated as a weighted mean of individual results with weights equal to inverse squares of measurement uncertainties.¹ Uncertainty of this result can be reliably estimated by studying the distribution of individual measurements around the weighted mean. This is a typical example of what we call “data.” Uncertainty is its intrinsic property that cannot be omitted. Without uncertainties, the results of a measurement or a calculation become unusable collections of senseless numbers, as they lack any

indication of their dependability. For example, if we omit the uncertainty when giving our result of the measured diameter of some cylinder, users of this result can never be sure that this cylinder will fit in a hole of given dimensions.

As indicated by the above example, uncertainty is a statistical property, and its determination largely relies on comparisons of repeated measurements or calculations. For measurements, there exists a very detailed guideline to determination and expression of uncertainties [2]. This guideline is not exact, as there are many difficult circumstances preventing an exact knowledge of all sources of measurement errors.² Nevertheless, there are good recipes for most circumstances.

For measurements of wavelengths of atomic spectral lines and estimation of uncertainties associated with them and propagating to other indirectly determined atomic properties, such as energy levels, there exists a methodology described in my old review [3]. Since that time, there appeared some new developments in statistical theory, which now allow a more robust estimation of uncertainties. In particular, Dr. Andrew L. Rukhin of the National Institute of Standards and Technology, USA (NIST) has developed a procedure to estimate statistical uncertainties in heterogeneous measurements [4–6]. This procedure was successfully utilized in recent atomic spectroscopy works [7, 8]. One of the purposes of this article is to explain this new statistical approach and promote its further use in atomic spectroscopy.

Estimation of uncertainties in theoretical calculations is a more difficult problem. Here, if one repeats the same calculations using the same computer code and input

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1140/epjd/s10053-024-00820-y>.

^a e-mail: alexander.kramida@nist.gov (corresponding author)

¹ In this paper, the term “uncertainty” means “standard uncertainty” as defined in Taylor et al. [1] (i.e., “square root of variance”), unless stated otherwise.

² In this paper, the term “error” is synonymous to “difference from the true value.”

parameters, the results are normally exactly the same, which by no means implies that they are infinitely accurate. Calculating the same atomic properties, e.g., transition probabilities, with a different code or with different input parameters usually gives different results, often by orders of magnitude. Comparisons of results obtained with different computational approaches can provide means for a robust estimation of uncertainties [3, 9]. The methods suggested in the above papers have been extended by several authors (see, e.g., [10, 11]).

Back in 2016, in a talk on this subject at the 12th International Colloquium on Atomic Spectra and Oscillator Strengths (ASOS12) [12], a report on the current status of the literature on atomic transition probabilities was given. In particular, it was mentioned that less than 2 % of published theoretical papers contained estimates of uncertainties of the reported theoretical values. There has been a significant progress in this regard: during the last two years, the percentage of such theoretical works containing uncertainty estimates grew to about 10 %. There is still a long way to go, and many researchers still make inadequate estimates. The second purpose of the present article is to provide an update on the procedures of estimation of uncertainties of theoretical transition probabilities in order to disseminate the newly developed methodologies.

2 Uncertainties in wavelength measurements

A common problem in all wavelength measurements is estimation of measurement uncertainties. As with all measurements, researchers must rely on the current recommendations of the International Bureau of Weights and Measures (BIPM) [2]. The NIST Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results [1], despite narrowly targeting NIST researchers in the title, are of general use as well. From these recommendations, one can see that uncertainty should be considered as a purely statistical concept.³ One makes a series of measurements, looks at the statistical distribution of the results, adds estimated contributions of possible systematic effects, calculates the weighted mean of the results, and somehow derives an uncertainty of this weighted mean. There is

³ Some readers object that only the “type A” uncertainties, as defined by Taylor et al. [1], are evaluated by statistical methods, while the “type B” uncertainties are evaluated by other means. However, from the explanations and examples given in Section 4 of Taylor et al. [1], one can see that evaluation of the type B uncertainties is designed to approximate standard statistical uncertainties as closely as possible, so that they can be treated by statistical methods on equal footing with type A uncertainties. In particular, the total standard uncertainty is calculated as a combination in quadrature of both type A and type B uncertainties (see Section 5 of Ref. [1]).

a long-standing problem with this procedure: there is no strict statistical recipe that fits all scenarios.

Let us look at one simple example. The Newtonian gravitational constant, G , has been precisely measured by several independent teams, and a weighted mean of these measurements is adopted by the Committee on Data of the International Science Council (CODATA2018) [13]. The currently recommended value is $6.67430(15) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$. Let us take it as one measurement and consider it together with a result of a hypothetical very imprecise undergraduate physics experiment. In this experiment, a student measured the acceleration of the free fall at the Earth’s surface, say, with a reversible pendulum. Then the Earth’s radius and mass data obtained from a web search were plugged into the Newtonian gravitational law to derive a value of G equal to $6.690(3) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$. Now, let us treat these two results as a series of repeated measurements and find the standard statistical weighted mean value with instrumental weights (inverse squares of measurement uncertainties) from the formulae below:

$$v_{\text{wm}} = \frac{\sum v_i w_i}{\sum w_i}, \quad (1)$$

$$u_{\text{wm}} = \frac{1}{(\sum w_i^2)^{1/2}}, \quad (2)$$

where v_i is the i -th measured value and $w_i = 1/u_i^2$ is its weight in the averaging (u_i being the standard uncertainty of the i -th measurement).

The weighted mean of our two measurements turns out to be $6.67434(15) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$. We seem to have slightly changed G compared to the CODATA value. However, we want to be careful and account for the fact that the standard statistical formulae (1, 2) give biased results. A seemingly good formula for unbiased uncertainty of weighted mean of a series of measurements can be found in Wikipedia [14] (referring to [15]):

$$u_{\text{biased}}^2 = \frac{\sum w_i (v_i - v_{\text{wm}})^2}{V_1}, \quad (3)$$

$$u_{\text{unbiased}} = \frac{u_{\text{biased}}}{(1 - V_2/V_1^2)^{1/2}}, \quad (4)$$

where $V_1 = \sum w_i$ and $V_2 = \sum w_i^2$.

From Eq. (3), we get $u_{\text{biased}} = 0.0008$, and from Eq. (4), $u_{\text{unbiased}} = 0.011$ (in the same units as G), 5 and 73 times greater than the CODATA uncertainty, respectively. This implies that the undergraduate experiment has hopelessly spoiled the joint effort of several leading laboratories!

The reason for this apparently wrong result is that Eqs. (3) and (4) are applicable only to the case where all data belong to a sample from a series of measurements having the same statistical distribution. This is obviously not true in this case, as our set of two measured values is *inhomogeneous*: each of them has its

own statistical distribution and its own set of systematic uncertainties, about which there is no information at this stage of data processing. In this situation, the concept of *dark uncertainties* comes in very naturally. To rectify the result, we need to add in quadrature some unknown dark uncertainties to each of the values participating in the averaging. This is where the new statistical theory of Rukhin [4–6] is indispensable. Both his *clustered maximum likelihood* and *clustered reduced maximum likelihood* estimators (CMLE and CRLME) unequivocally and correctly single out our undergraduate experiment as the culprit of disagreement between the two values. They both assign a large value of dark uncertainty of 0.016 (in the same units as G) to our undergraduate experiment, which results in the weighted mean value and its uncertainty coinciding with CODATA.

However, the reader must be warned right away that the averaging procedure in this example is methodologically wrong. The result we got from the refined statistics with dark uncertainties should only be used to make a conclusion that one of our experiments was faulty (i.e., the measured value and/or its uncertainty were determined incorrectly) and must be investigated and corrected. It is appropriate to quote here the reference Guide to the Expression of Uncertainty in Measurement (GUM) [2] (section 3.4.7): “Blunders in recording or analyzing data can introduce a significant unknown in the result of a measurement... Measures of uncertainty are not intended to account for such mistakes”. The point I want to make here is that statistics helped us to identify the erroneous measurement, but it is wrong to average correct measurements together with incorrect ones; the latter should rather be corrected (from physical considerations) or excluded from the averaging.

Turning to wavelength measurements, I note that many of them are intrinsically heterogeneous even in the case when a large number of spectral lines are measured in the same experiment (i.e., using the same equipment and methodology). This is because the lines have different spectral profiles (e.g., some are broadened by self-absorption or saturation effects), and many lines are usually overlapping in a quasi-random fashion with other lines or are affected by hyperfine structure and/or isotope shifts.

Rukhin’s method can be used not only in direct comparisons of several measured wavelengths of the same spectral line but also in indirect comparisons, such as analysis of deviations between observed and Ritz wavelengths. Such analysis was suggested in [3] as an easy means to estimate the wavelength uncertainties where they are not available from measurement reports. However, at that time Rukhin’s theory was yet unpublished. Now it is clear that it can nicely supplement this analysis by allowing one to easily detect abnormally deviating wavelengths. One has to keep in mind that no statistical theory can replace a thorough analysis of measurement errors. It can only serve as an indicator of problems in some measurements. Finding the causes of these problems and eliminating them is a task of the physicist.

As mentioned above, in the course of our joint work on Li-like ions [7] I have developed a statistical toolbox implementing Rukhin’s CMLE and CRMLE methods, as well as Mandel–Paule and DerSimonian–Laird dark uncertainty estimators (see [4]) and the functions necessary to make the so-called normal probability plots [16]. This toolbox is implemented in a Visual Basic module embedded in a Microsoft Excel⁴ file included as the Supplementary Information with this article. Its use is illustrated in the following subsection.

2.1 Statistical toolbox

To understand the general ideas used in this statistical toolbox, the reader is advised to study Section 4 of Ref. [7], as well as Rukhin’s articles [4–6]. The content and usage of the Toolbox are explained in Appendix A.

The most useful of the many functions included in the Toolbox are `dark_unc_MP` (the Mandel–Paule estimator), which provides for an exact reduction of reduced χ^2 (defined further below) to unity in the cases where the measurements participating in the averaging disagree beyond the stated uncertainties, and the two functions based on Rukhin’s clustered estimators, `dark_unc_CMLE` and `dark_unc_CRLME`.

The example included in the toolbox is quoted from our work on Li-like ions [7]. It consists of a set of 15 measurements of energy of the $1s(2S)2s2p(^3P^o)^2P_{1/2}^o$ level of Li-like oxygen, O VI (for explanation of values and references to their sources, see [7]). To see the dark uncertainties and corresponding weighted mean values, click on the Run button. As explained in [7], there is only one value that significantly disagrees with the rest of the data. This is the energy of $4539260(470) \text{ cm}^{-1}$ deduced from the measurement of an absorption feature reported by Liao et al. [17]. Abnormality of this value is unanimously detected by the CMLE and CRMLE estimators, which assign to it a dark uncertainty of about 1600 cm^{-1} , five times greater than the stated uncertainty (470 cm^{-1}). One can see that the Mandel–Paule estimator assigns equal dark uncertainties of about 200 cm^{-1} to all measurements, which results in a weighted mean of $4540570(220) \text{ cm}^{-1}$. With either CMLE or CRMLE, the weighted mean is $4540880(230)$, while the standard statistical formulae without dark uncertainties give $4540590(210) \text{ cm}^{-1}$. One can see that the Mandel–Paule method gives a mean value close to the standard statistical formula, while the result of CMLE and CRMLE is significantly different (by 310 cm^{-1} , or 1.4σ). As follows from the analysis of Azarov et al. [7], the result of the MP estimator disagrees with the recommended value, $4541169(17) \text{ cm}^{-1}$, by 3σ , while for the result of CMLE and CRMLE the discrepancy is much smaller, only 1.3σ . Here, however,

⁴ The identification of commercial products in this paper does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the items identified are necessarily the best available for the purpose.

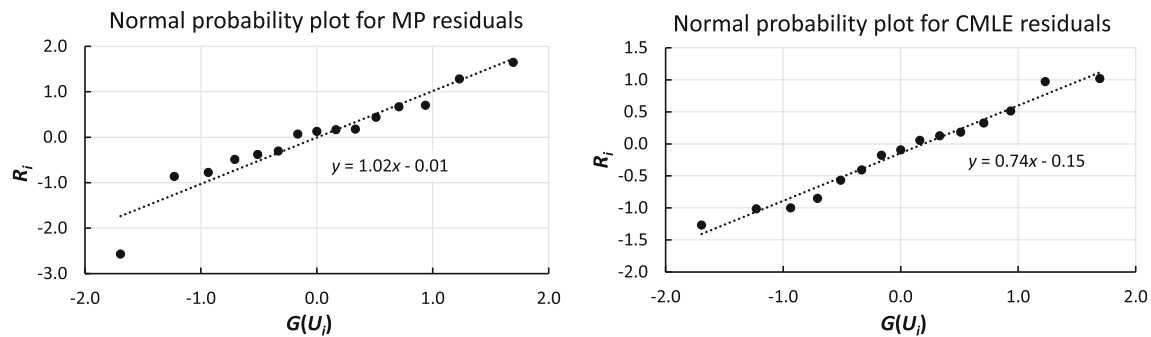


Fig. 1 Normal probability plots of normalized residuals (R_i ; see text) for calculations made with the Mandel–Paule (MP) and clustered maximum likelihood (CMLE) estimators of dark uncertainties. The quantity $G(U_i)$ on the horizontal axes of these plots is the percent point function of the uniform order statistic median of the normal distribution [16]. The dotted lines are linear fits to the data points

the main point is that different ways of estimating dark uncertainties result in different values of both the weighted mean and its uncertainty. Application of statistical methods should always be taken with a grain of salt: their underlying assumptions can lead you to a wrong result.

Note that the values of weighted means and their uncertainties can easily be calculated using the functions `WMDU()` and `unc_WMDU()` included in the toolbox. These functions take as input three vertical arrays of cells: values, stated uncertainties, and dark uncertainties. The latter can be specified as an array of empty cells for calculation without dark uncertainties.

The procedure called by the Run button of the toolbox also displays the normal probability (NP) plots (see [16]) corresponding to the three averaging procedures (MP, CMLE, and CRMLE). The MP and CMLE plots are shown in Fig. 1 (the plot obtained with CRMLE is omitted, as it is almost exactly the same as the CMLE plot). In these plots, R_i is the normalized residual of the i -th measurement (difference of the measured value from the weighted mean divided by the stated uncertainty combined in quadrature with the dark uncertainty). One can see that the uniformly distributed dark uncertainty resulting from the MP estimator does not rectify the irregularity of the NP plot: there is one data point that strongly deviates from the straight line near which the rest of the data points are clustered. This data point corresponds to the faulty measurement discussed above. On the other hand, the NP plot drawn from the CMLE calculation is much closer to a straight line, but it has a slope of about 0.74, while one would expect this slope to be close to 1.0 for normally distributed measurements with correctly estimated uncertainties. The corresponding value of χ^2 is 0.72, while for the MP calculation it is exactly 1.0, by design of this estimator. One has to come at peace with it: yes, the χ^2 and the slope of the NP plot are allowed to be smaller than unity. This can happen due to presence of significant systematic uncertainties in the measurements.

One can see that the MP estimator is conceptually simpler and easier to use than the cumbersome CMLE and CRMLE estimators. However, it has an intrinsic

deficiency: it treats all measurements on equal footing. If the measurements are statistically inconsistent, it assigns the same dark uncertainty to all of them. If the measurements are, in fact, unequal (in the sense that they use different methods or different experimental conditions, or there are other factors leading to different statistical distributions in the measurements), this leads to both the weighted mean value and its uncertainty being wrong. The CMLE and CRMLE estimators, when they agree with each other (this is not always the case), give much better estimates for the weighted mean and its uncertainty, even if one chooses not to investigate the causes of the large dark uncertainty they assign to outlying measurements. However, one should understand their limitations. They are designed to be valid under certain assumptions, the most important of which is that there are only two classes of measurements, “good” (receiving no dark uncertainty) and “bad” (all of the latter being equally bad and receiving the same value of dark uncertainty). If one wants to be fast and efficient, the easiest way of getting a statistically consistent set of data is to simply discard those to which CMLE and CRMLE unanimously assign large dark uncertainties. However, there is danger of accidentally throwing away a piece of gold together with muck.

As noted above, the detection of a faulty measurement by a statistical procedure is in itself insufficient for a proper interpretation of the measurements. One has to look for physical reasons of the detected problem. In the case of the O VI line measurement by Liao et al. [17], the most probable cause is the intrinsic deficiency of the measurement method: it was designed to measure the positions of absorption peaks, while absorption to autoionizing levels, such as the one selected for this example, is described by a Fano profile, which is asymmetric and includes both an absorption and an emission feature. The true position of the level should be found by fitting a Fano profile to the observed feature in a significantly wider spectral area than just the width of the absorption peak. In this case, indeed, Figure 3 of Liao et al. [17] shows the presence of a weaker emission feature at a wavelength somewhat shorter than that of the absorption peak. A proper modeling of the spectral

profile would give a higher energy, consistent with the observed deviation from the mean experimental value, but probably with a significantly larger uncertainty.

As mentioned above, Rukhin’s estimators of dark uncertainties can be used in the analysis of differences between observed and Ritz wavelengths (or wave numbers). Let us take as an example the measurements of Zr I wave numbers by Lawler et al. [18]. Although the results of their level optimization look generally good, there are many lines for which the observed wave numbers deviate from the Ritz values by more than twice the standard uncertainty. Among the total list of about 370 lines, the fraction of such strongly deviating lines is about 8 %, while for a normal statistical distribution one would expect 5 %. The simplest way to detect problematic lines is by calculating the Mandel–Paule dark uncertainty taking the differences of observed wave numbers from the Ritz values ($\Delta E_{\text{obs-Ritz}}$) as the measured values with uncertainties equal to those of the observed wave numbers (u_{obs}). These dark uncertainties should be calculated for small subsets of lines originating from the same upper level. In this way, it is easy to see that the largest dark uncertainty gets assigned to lines from the $4d^2 5s 5p z^5 P_1^o$ level (25489.87 cm^{-1} in the NIST Atomic Spectra Database (ASD) [19]). The data associated with the four transitions from this level included in Lawler et al. [18] are given in Table 1.

As noted above, the dark uncertainties given in the last three columns of Table 1 are estimated by treating the wave number difference $\Delta\sigma_{\text{obs-Ritz}}$ as measured quantity with uncertainty equal to that of the observed wave number, u_{obs} . Some people argue that, in the treatment of $\Delta\sigma_{\text{obs-Ritz}}$, the uncertainty of the Ritz wave number should be added in quadrature to u_{obs} . However, in the least-squares level optimization procedure (in all of its flavors known to me), the minimized quantity is the sum of squares of the ratio $\Delta\sigma_{\text{obs-Ritz}}/u_{\text{obs}}$ (denoted as $\Delta\sigma/u_{\text{obs}}$ in Table 1 for brevity). This sum is called residual sum of squares (RSS). Ideally, if the measurements are a sample from a normal statistical distribution, and all uncertainties u_{obs} are properly estimated, the optimization procedure should result in the ratio of RSS to the number of degrees of freedom (number of observed transitions minus the number of excited energy levels) equal to one. This ratio is otherwise known as reduced χ^2 . Conversely, if $\chi^2 > 1$, it means that one or both of the assumptions made above do not hold: some uncertainties are underestimated and/or the measurements are not normally distributed. Note that uncertainties of the Ritz wave numbers do not participate in weighting of the observed wave numbers. These properties of least-squares fitting follow from rigorous statistical considerations (see details in the LOPT article [20]).

The particular transitions that are most likely causing the problem in determination of the upper level energy can be pinpointed by using Rukhin’s CMLE and CRMLE methods implemented in the present statistical toolbox. Both these estimators unanimously suggest (by assigning the large dark uncertainties of about 0.006 cm^{-1}) that the transitions listed in the first and

Table 1 Selected spectral lines of Zr I

σ_{obs}^a	u_{obs}^1	N_{spectra}^b	$E_{1,\text{ASD}}^c$	$E_{2,\text{ASD}}^c$	$E_{1,\text{opt}}^d$	$E_{2,\text{opt}}^d$	$\Delta\sigma_{\text{obs-Ritz}}^e$	$\Delta\sigma/u_{\text{obs}}$	DU_{MP}^f	$\text{DU}_{\text{CMLE}}^g$	$\text{DU}_{\text{CRMLE}}^h$
14473.2603	0.0015	4	11016.65	25489.87	11016.6440	25489.8995	0.0048	3.21	0.0051	0.0061	0.0062
14604.5628	0.0015	4	10885.36	25489.87	10885.3362	25489.8995	-0.0005	-0.35	0.0051	0.0000	0.0000
21303.8870	0.0026	3	4186.11	25489.87	4186.0080	25489.8995	-0.0045	-1.73	0.0051	0.0000	0.0000
25489.8915	0.0025	2	0.00	25489.87	0.0000	25489.8995	-0.0080	-3.20	0.0051	0.0061	0.0062

^aMean observed wave number and its standard uncertainty as given in Table 6 of Lawler et al. [18] (cm^{-1})

^bNumber of spectra used in the wave number averaging

^cLower and upper levels of the transition as given in NIST ASD [19] (cm^{-1})

^dOptimized values for the lower and upper levels of the transition (cm^{-1} ; see text)

^eDifference between the observed and Ritz wave numbers (cm^{-1})

^fDark uncertainty estimated by the Mandel–Paule method (cm^{-1})

^gDark uncertainty estimated by Rukhin’s CMLE (cm^{-1})

^hDark uncertainty estimated by Rukhin’s CRMLE (cm^{-1})

last rows of Table 1 are the most likely causes of the problem. However, it is not a good practice to blindly use these estimated dark uncertainties to degrade the originally estimated uncertainties. This can lead to a physically unjustified shift of the resulting energy levels. As noted above, no statistical considerations can replace a careful analysis of physical reasons of observed problems. They can be used to pinpoint the likely culprits, but the reasons for large deviations should be investigated. In the case of the measurements of Lawler et al. [18], each measured wave number results from averaging of measurements in several spectra. The number of spectra used for this averaging is given in the column N_{spectra} in Table 1. Analyzing the statistics of individual measurements could give more insights about the problem; looking at the line profiles in each spectrum is even more informative. Typical causes of large disagreements between different measurements are blending with neighboring lines, asymmetries caused by partially resolved hyperfine structure or self-absorption, noise, and calibration errors. In typical atomic spectroscopy experiments, the number of observed lines can be very large, measured by thousands or even tens of thousands. It is very difficult to analyze profiles of every measured line. However, isolating the problems to a limited number of lines can make the analysis of outstanding errors manageable. The statistical toolbox given here can be a good help in this matter.

It should be noted that the optimized level values given in Table 1 result from my level optimization made with the LOPT code of Kramida [20]. They slightly differ from the values given by Lawler et al. [18] (by one or two units of the last decimal place). These small differences are due to rounding errors that are treated differently in LOPT and in the code used by Lawler et al. [18].

Regarding the use of the MP, CMLE, and CRMLE estimators of dark uncertainties, my current recommendation is to follow the procedure described above: first, use the fast MP estimator to find potentially problematic measurements, then use both CMLE and CRLME to further narrow down the problem. The reason for using both these estimators is that they do not always agree with each other. If they do agree, it gives a strong indication that the measurements they endow with a large dark uncertainty are the culprit of the problem. If they do not agree, the problem widens. In any case, only a careful examination of all suspicious measurements can give a good solution to the problem of making all measurements statistically consistent with each other.

3 Uncertainties in calculated transition probabilities

There are two radically different approaches to estimation of uncertainties of theoretical calculations. One approach, which is used, e.g., by CODATA [13], is to

explicitly include all neglected terms in the equations (such as the omitted high-order effects in a theory accounting only for a limited number of terms), estimate possible sizes of these omitted terms, and combine them in quadrature to arrive at an estimate of the total possible error. In calculation of transition matrix elements (determining transition probabilities), it is a very difficult task. However, it was done in several works. See, e.g., Safronova and Safronova [21]. It is fair to say that this approach is limited to a restricted number of atomic systems and requires a high level of expertise in atomic theory.

The second approach, which is conceptually simpler and technically easier to implement, is based on comparisons utilizing statistics in a way that is similar to evaluation of experimental results. The practical methodology of such evaluation was described in my review article [3], and its possible extensions were outlined in a subsequent article [9].

The main requirement for applicability of statistical evaluation of theoretical uncertainties is availability of a benchmark data set to compare with. In selection of benchmark data, the first priority should always be experimental data, if they are available. However, not all experimental methods are reliable, and it requires considerable expertise in both theory and experiment to detect unreliable experimental data. Some good guidelines for critical evaluation of transition probability data can be found in the review made by Wiese [22]. As one can see from that review, critical evaluation of published data (both experimental and theoretical) is not an easy task. Not many theorists possess the skills and knowledge required for it. Thus, one should use the published products of available critical evaluations. One common source of such critically evaluated data is the NIST ASD [19]. However, it is far from being complete; there are many spectra for which critically evaluated data on transition probabilities do not exist or are too sparse. (Note that the data compiled at NIST are not always the most accurate available. To extract the most reliable and accurate data, one should always consult the literature. Lists of relevant references are conveniently rendered in the links provided in the NIST ASD. These lists are extracted from the NIST bibliographic databases (see Sect. 4), which are fairly complete and up to date.) In the absence of reliable external benchmarks, one can resort to internal comparisons, in which different results of the same computer code can be compared with each other. The most commonly used method is to compare the results for transition rates computed in the length and velocity forms (in a non-relativistic approximation; these two forms roughly correspond to results obtained in the Babushkin and Coulomb gauges in a relativistic calculation). Some of the widely used modern atomic codes do produce transition rates in these two forms, e.g., the General Relativistic Atomic Structure Package (GRASP; see Froese Fischer et al. [23]) and the Flexible Atomic Code (FAC). The latter code was originally developed by Gu [24] and subsequently parallelized and extended to include many-body perturbation theory (MBPT) calculations [25].

To facilitate uncertainty estimation, the GRASP code outputs the so-called uncertainty indicator dT (see the GRASP2018 manual included in the Supplementary Materials for Ref. [23]):

$$dT = \frac{|A_l - A_v|}{\max(A_l, A_v)}, \tag{5}$$

where A_l and A_v are the length and velocity forms of transition rate.

This was apparently inspired by the articles of Froese Fischer [26] and Ekman et al. [27]. However, this definition of the uncertainty indicator differs from the one originally proposed in the above papers:

$$\frac{\delta A'}{A'} = (\delta E + \delta S), \tag{6}$$

where A' is the energy-adjusted transition rate (see Eq. (13) below) computed from the observed transition energy (E_{obs}), $\delta E = |E_{\text{calc}} - E_{\text{obs}}|/E_{\text{obs}}$ is the relative error in the transition energy, and $\delta S = |S_l - S_v|/\max(S_l, S_v)$ is the relative discrepancy between the length and velocity forms of the line strength (see Eqs. (4–7) of Ekman et al. [27]).

The quantity δE in the above equation practically never goes to zero, while its absence in Eq. (5) results in dT equal to zero in too many cases just due to chance coincidences of A_l and A_v . However, in most cases on which theorists work nowadays, experimental energies are not known. This prompted Ekman et al. [27] to give a truncated recipe in their Eq. (8): $\delta A = (\delta S)A$, which ultimately reduces to Eq. (5).

It is important to understand that the above equations (5, 6) are meant to be only *uncertainty indicators* and not estimates of uncertainty. Many theorists forget about it and give dT in their tables describing it as “the uncertainties in the computed transition rates” (see, e.g., Atalay et al. [28] and Manai et al. [29]). As follows from the examples of application of dT to evaluation of uncertainties in Ekman et al. [27], the initially proposed use of dT is strictly statistical: the estimate of uncertainty of calculated rates for a certain group of transitions expected to be calculated with a similar accuracy is given as some average of the magnitude of dT for this group of transitions. How exactly to divide all transitions into groups with similar accuracy is still an open question (see below).

It must be noted that the original uncertainty estimator proposed by Froese Fischer [26], given by Eq. (6), has no statistical justification.

Furthermore, there is a big problem in the definition of dT : due to the use of the max function in the denominator of Eq. (5), it *always* underestimates the uncertainties. This is because, for any statistically significant data set, there are similar numbers of cases when $A_v < A_l$ and when $A_v > A_l$. Always using the maximum value in the denominator results in roughly half the cases in which the value of dT is smaller than the actual relative discrepancy. When the differences

between A_l and A_v are small, this underestimation is insignificant, but unfortunately, in most calculations of complex spectra, the vast majority of calculated transition rates have very poor accuracy, with discrepancies between A_l and A_v reaching orders of magnitude for very weak transitions. To some extent, this drawback of Eq. (5) can be rectified by using min instead of max in the denominator. This, however, would lead to overestimation of uncertainties. Arguably, it would give a more prudent estimate of uncertainty, since it neglects the errors in the transition energies.

A somewhat better statistical indicator of uncertainty could be given as

$$dL = \ln(S_1/S_2), \tag{7}$$

where S_1 and S_2 are any two forms of line strength of the same transition. For example, they can be the length and velocity forms (S_l and S_v) from the same calculation, or the results from two adjacent layers of a series of active-space calculation (for explanation of this terminology, see, e.g., Jönsson et al. [30], Section 3.2), or the results of two different calculations. To give a reasonable estimate of uncertainty, S_1 and S_2 in Eq. (7) should be calculated in sufficiently developed physical models. For example, if both S_1 and S_2 are calculated with a limited number of interacting configurations, they can be close to each other; then Eq. (7) would give an unreasonably small uncertainty estimate. An indication of a sufficiently good quality of the models can be obtained, e.g., by examining convergence trends in a series of calculations with increasing complexity.

For a group of transitions with similar accuracy, the relative uncertainty of the line strength can be estimated from the root mean square (rms) of dL , $\langle dL \rangle$:

$$u_S \approx e^{\langle dL \rangle} - 1. \tag{8}$$

In the limit of S_l being very close to S_v , when they are used as S_1 and S_2 in Eq. (7), u_S tends to dT , which is the main reason for using the natural logarithm and natural exponent in these equations.

To my knowledge, the first implementation of this method with the use of results of two successive layers of an active-space multiconfiguration Dirac–Hartree–Fock (MCDHF) calculation was made by El-Sayed in her work on Pd XLII [31]. In that work, as well as in several others, including Ref. [10] mentioned in the Introduction, she uses decimal logarithms instead of natural ones, but the difference from Eqs. (7, 8) is only technical. Her works illustrate an important fact: the dL estimator can be used for evaluating uncertainties not only of electric-dipole (E1) and electric-quadrupole (E2) transitions, but also for magnetic-dipole (M1) and magnetic quadrupole (M2) transitions, where there is no velocity form to compare with.

In another large series of works, a similar method was employed to estimate the uncertainties of transition rates based on comparison of results of two different calculations—MCDHF and MBPT—implemented

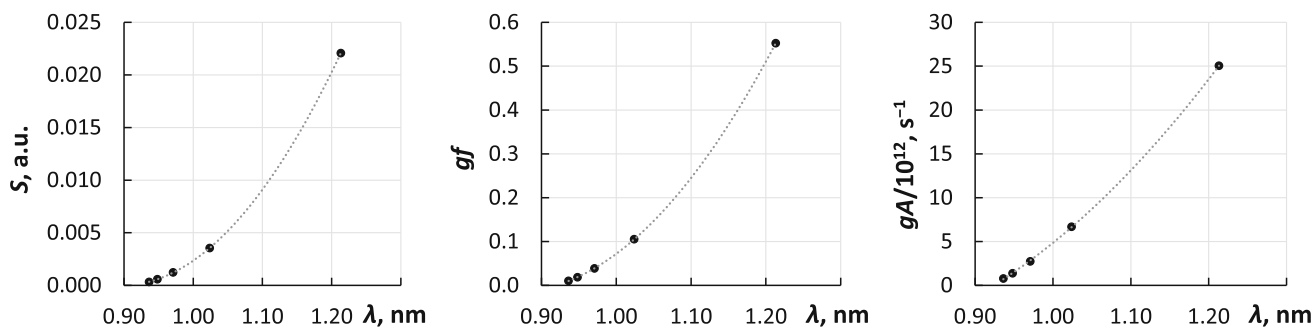


Fig. 2 Line strength S (left), weighted oscillator strength gf (middle), and weighted transition probability gA (right) as a function of transition wavelength λ for the $1s - np_{3/2}$ ($n = 2$ to 6) resonance transitions of H-like Ne (Ne X). The data are from the NIST ASD [19] (originating from Jitrik and Bunge [37]). Smooth curves are quadratic interpolations

in GRASP and FAC, respectively; see, e.g., Zhao et al. [32].

One should keep in mind that Eqs. (7) and (8) are also statistically unjustified in the sense that there is no evidence for statistical distribution of dL to be close to normal. In fact, my numerical experiments on M1 and E2 transitions in Fe V [9] showed that it is not the best choice of function from this point of view. Different groups of transitions have different shapes of statistical distributions even when the input parameters of the computing code [33,34] are varied randomly with a normal statistical distribution. It was found that for most M1 and E2 transitions of Fe V the best function (i.e., having statistical distribution closest to normal) is not $\ln(S_1/S_2)$ but $[(S_1/S_2)^{1/3} - 1]/(1/3)$. Nevertheless, its asymptotic for $S_1 \rightarrow S_2$ is the same, while for large discrepancies, there is not much sense in exactly answering the question: is the uncertainty one or two orders of magnitude?

As mentioned above, the big outstanding question is: how to divide all transitions into groups with similar expected uncertainties? In my article of 2013 [3], it was argued that the line strength S (in length form) is a natural choice on which such grouping should be made, because it does not explicitly depend on transition energies. However, this is true to some extent only for non-relativistic calculations, and probably not for all of them. For example, in several textbooks (see, e.g., Cowan [33] and Corney [35]) it was shown that for resonance lines of H-like ions, the oscillator strength f does not depend on nuclear charge Z , while the line strength is proportional to Z^{-2} , which is close to the dependence of wavelength on Z (see also Wiese and Fuhr [36]). However, if one plots the dependencies of either S or f on transition energy along any chosen series of transitions with increasing principal quantum number for the same Z or for several values of Z , neither shows a constant behavior in practically any spectrum. This is illustrated in Fig. 2 showing the line strength S , weighted oscillator strength gf , and weighted transition probability gA ⁵ as

a function of transition wavelength λ for the $1s - np_{3/2}$ resonance transition of H-like Ne (Ne X) for $n = (2-6)$.

The dynamic range of variation of the three quantities plotted in Fig. 2 can be taken as the ratio of the $n = 2$ and $n = 6$ values (the rightmost and leftmost points on the plots, respectively). It turns out that it is the smallest for gA (32), while for gf and S it is 53 and 69, respectively. Thus, in this case, the line strength S displays the strongest dependence on transition energy.

Nevertheless, it is an empirical fact that for the vast majority of calculated sets of transition rates, uncertainties in the calculated values are strongly correlated with line strength, S , while correlation with other quantities, such as gf , gA , and also the cancellation factor is much weaker. There are exceptions that can be found in the literature. Most notably, in the recent MCDHF calculations of transition rates for neutral nitrogen and oxygen [38,39], the differences between the Babushkin and Coulomb gauges were found to be most strongly correlated with the A values rather than S . Thus, the general methodology outlined in [3] should be amended by an additional step: one should compare the distribution of computational errors when plotted against different quantities, such as S , gf , gA , or perhaps the branching fractions for radiative decay and choose the quantity that displays the strongest correlation with computational errors. This is illustrated in Fig. 3.

As one can see in the top right panel of Fig. 3, discrepancies between the Babushkin- and Coulomb-gauge results (S_B and S_C) do not correlate with S_B . Correlation with S_C is even worse and is not shown here. On the other hand, there is a good correlation with gA_C (top left), which justifies the use of this quantity as a guide to uncertainties. The bottom panel is discussed further below.

Most textbooks, as well as a majority of research articles, recommend using the length form (or Babushkin gauge) as producing more reliable values of transition rates. However, recent research indicates that this is not universally correct. E.g., the MCDHF study of Papoulia et al. [40] indicated that in some cases, par-

⁵ One should keep in mind that the statistical weight g entering the commonly used designations gf and gA pertain

to different levels of a transition: the lower level in the case of gf and the upper level in the case of gA .

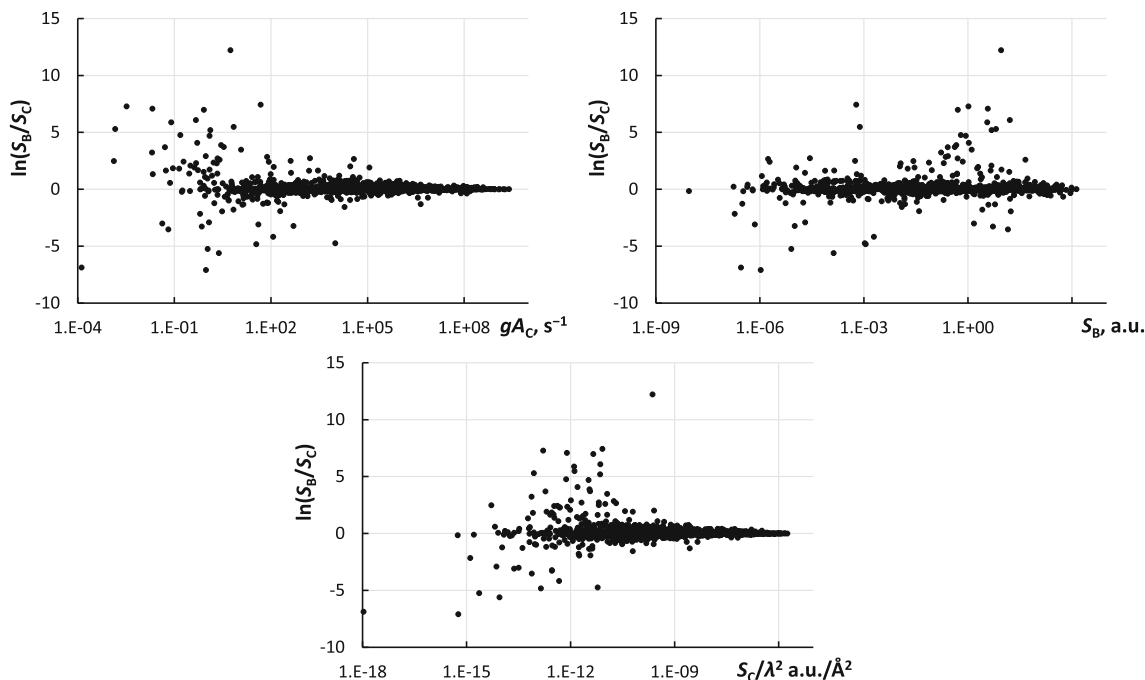


Fig. 3 Discrepancy between the line strengths computed in the Babushkin and Coulomb gauges (S_B and S_C) plotted against weighted transition probability computed in the Coulomb gauge (gA_C ; upper left), against S_B (upper right), and against S_C/λ^2 (bottom; λ is transition wavelength). The data are from the MCDHF calculation of atomic nitrogen [38]

ticularly for transitions involving high Rydberg states, the results in the Coulomb gauge may converge faster and be more reliable than the results in the Babushkin gauge, depending on the chosen computational scheme. One should also keep in mind that the line strength computed in the length and velocity forms have different sensitivity to errors in the computed transition energy. From Eqs. (8) and (9) of Bilal et al. [41], one can see that the leading term of S in the length form (Babushkin gauge) does not depend on transition wavelength, while the velocity form (Coulomb gauge) is proportional to its square. This can be derived from relativistic formulae given by Grant [42]. Relativistic atomic codes, such as GRASP [23] and FAC [24,25], do not directly compute line strengths. Instead, they compute reduced transition matrix elements, and then the transition rates (A) are derived from them. Formulae for them can be found, e.g., in the paper of Grant quoted above [42]. From these formulae, by using a standard relation between A and S [36], one can derive approximate expressions for line strength as a function of transition frequency (ω). For magnetic transitions,

$$S_{\alpha\beta}^m \propto \left[\int_0^\infty (P_\alpha Q_\beta - Q_\alpha P_\beta) r^L dr \right]^2, \quad (9)$$

where L is the multipolarity (1 for dipole transitions, 2 for quadrupole, etc.), while for electric transitions the formulae depend on gauge. In the Babushkin gauge

(equivalent to length form),

$$S_{\alpha\beta}^e(B) \propto \left[\int_0^\infty R_\alpha R_\beta r^L dr \right]^2, \quad (10)$$

and in the Coulomb gauge (equivalent to velocity form),

$$S_{\alpha\beta}^e(C) \propto \frac{1}{\omega^2} \left[\int_0^\infty R_\beta r^{(L-1)/2} \left\{ \frac{d}{dr} + \frac{(l_\alpha - l_\beta)(l_\alpha + l_\beta + 1)}{2r} \right\} r^{(L-1)/2} R_\alpha dr \right]^2. \quad (11)$$

In Eqs. (9), (10), and (11), α and β denote the lower and upper states of a transition, P and Q are the real and imaginary components of the wave function, and definitions of other quantities can be found either in Grant’s paper [42] or in his book [43]. As noted above, these equations are approximate. They were obtained by using the leading terms in Taylor series expansion of Bessel functions of $\frac{\omega r}{c}$ entering radial integrals in relativistic formulae of Grant [42].

Equations (9) and (10) seem to confirm the idea of Ref. [3]: the line strength (of any magnetic transition or of the length form of electric transitions) does not explicitly depend on transition energy and thus should be a good discriminating quantity in evaluation of uncertainties of line strengths. However, as follows from Eq. (11), the velocity form of line strength is not a good quantity in this regard, as it is inversely proportional to transition frequency. It seems reasonable to

assume that, if we divide it by squared wavelength, this would cancel the energy dependence, so S_C/λ^2 should be a good discriminating quantity. This is confirmed to some extent by the lower panel of Fig. 3: it looks quite regular, with only one strongly deviating point having $\ln(S_B/S_C) \approx 12$. This point corresponds to the transition having the longest wavelength in the tables of Ref. [38].

There is an important consequence of Eq.(11): in high-precision calculations, the results in the Coulomb gauge should be adjusted to experimental wavelengths (if known):

$$S_{\text{adj}}^v = S_{\text{calc}}^v \frac{\lambda_{\text{exp}}^2}{\lambda_{\text{calc}}^2}. \tag{12}$$

To my knowledge, this adjustment was first introduced by Bilal et al. [41]. Note that this adjustment is additional to the well-known adjustment factors for A - and f -values (of E1 transitions, in the length form):

$$\begin{aligned} A_{\text{adj}}^{\text{len}} &= A_{\text{calc}}^{\text{len}} \frac{\lambda_{\text{calc}}^3}{\lambda_{\text{exp}}^3}, \\ f_{\text{adj}}^{\text{len}} &= f_{\text{calc}}^{\text{len}} \frac{\lambda_{\text{calc}}}{\lambda_{\text{exp}}}, \end{aligned} \tag{13}$$

meaning that, for the velocity form, the right parts of Eq. (13) must be additionally multiplied by the same factor as used in Eq. (12), yielding

$$\begin{aligned} A_{\text{adj}}^{\text{vel}} &= A_{\text{calc}}^{\text{vel}} \frac{\lambda_{\text{calc}}}{\lambda_{\text{exp}}}, \\ f_{\text{adj}}^{\text{vel}} &= f_{\text{calc}}^{\text{vel}} \frac{\lambda_{\text{exp}}}{\lambda_{\text{calc}}}. \end{aligned} \tag{14}$$

Another interesting development in comparisons of length and velocity forms was recently introduced by Zhang et al. [11]. These authors have found that occasional closeness of results obtained in different gauges is not always random. Sometimes it is due to insensitivity of the calculation to the choice of gauge; then the closeness of results in different gauges indicates real high accuracy of these results. The idea stems from the work of Grant [42] whose treatment allows one to consider gauge not as a strictly discreet choice between a few options but as a continuously varying parameter G . If the wavefunctions are exact, the results have no dependence on G . However, in imprecise numerical calculations, there will be a difference depending on G . The dependence of line strength on gauge is parabolic, with a positive quadratic term [44, 45]:

$$S(G) = aG^2 + bG + c. \tag{15}$$

Its coefficients can be determined from three data points. Two are provided by the values of S in the Coulomb ($G = 0$) and Babushkin ($G = \sqrt{(L+1)/L}$)

gauges, where L is the multipolarity (1 for E1 transitions, 2 for E2, etc.). The third point is provided by the non-trivial fact that S equals zero at exactly one point given by Eq. (9) of Rynkun et al. [45]:

$$G_{S=0} = \frac{\sqrt{2}}{1 - (M_B/M_C)}, \tag{16}$$

where M_C and M_B are reduced matrix elements computed in the Coulomb and Babushkin gauges. In the special case of $M_B = M_C$, $G_{S=0}$ goes to infinity, and S becomes independent of G .

Radžiūtė and Gaigalas [46] have investigated the dependence of dL (with $S_1 = S_B$ and $S_2 = S_C$) on the parameter $G_{S=0}$ in several As-like spectra from Br III to Sr VI. As seen from their Figures 8 and 10, in all these spectra, accuracy better than 50 % is achieved when $|G_{S=0}| \gtrsim 5$ for both the E1 and E2 transitions, and the dependence of dL on $G_{S=0}$ looks very regular, which prompted a discussion of the use of this dependence in evaluation of uncertainties: “Function ($S(G)$) minimum position correlate[s] well with accuracy class: transitions with [greater] $|G_{S=0}|$ values are in better accuracy class” [46]. However, this behavior is a simple consequence of definition of $G_{S=0}$ (see Eq. 16). Inverting this equation and noting that $S_B/S_C = (M_B/M_C)^2$, one obtains

$$\frac{S_B}{S_C} = \left(1 - \frac{\sqrt{2}}{G_{S=0}}\right)^2, \tag{17}$$

from which it follows that $S_B \rightarrow S_C$ when $|G_{S=0}|$ increases beyond $\sqrt{2}$, and the precise boundaries of the 50 % accuracy ($0.5 < S_B/S_C < 1.5$) are at $G_{S=0} = \frac{2}{\sqrt{2}-\sqrt{3}} \approx -6.29$ and $G_{S=0} = \frac{2}{\sqrt{2}+1} \approx 4.83$. In my view, the methodology described by Radžiūtė and Gaigalas [46] reflects a simple fact: these authors are assuming that there is no randomness in cases when length and velocity forms give close results, and this closeness always implies a high accuracy of those results. This is not justified, as discussed below.

Various ideas described in the articles mentioned above [11, 45, 46] seem interesting, as they give new insights into possible ways to analyze the discrepancies between the length and velocity forms of calculated transition rates. However, at present I do not see any clear-cut recipe to decide whether occasional closeness of these two forms indicates real accuracy of the calculation or it is a quasi-random computational artifact. In the literature, there are many examples of calculations yielding very close results in the length and velocity forms, both being very wrong. This was spelled out by Hibbert [47]: “However, even though exact agreement between the two forms is achieved in a local potential approximation, the common value is not necessarily correct. It is sometimes possible to achieve good length and velocity agreement even in the HF approximation (a non-local potential method), but again the common value can be incorrect” (here, ‘HF’ means Hartree–

Fock). He further gives some examples of such computational artifacts.

One of the ideas tried by Rynkun et al. [45] and Gaigalas et al. [48] is investigation of the dependence of cancellation effects on gauge. The cancellation factor (CF) originally used in Cowan's atomic structure codes [33] is a numerical measure of the extent to which contributions of different sign cancel each other in the computation of transition matrix elements (see Eq. (6) of Gaigalas et al. [48]). Its smallness (generally, being smaller than about 0.1) indicates a high degree of cancellation, resulting in unreliable computed values of S and all other related parameters. It was found in both these works [45, 48] that calculations in Babushkin gauge are less sensitive to cancellations than in Coulomb gauge. Rynkun et al. [45] also reported a few transitions of Ce IV for which the largest values of CF (hence, more reliable results for S) were achieved with $G = 1$, which is in between the Coulomb ($G = 0$) and Babushkin ($G = \sqrt{2}$) gauges for E1 transitions. The extension of the GRASP code package, with which these computations were made, is not yet publicly available at the distribution site of GRASP2018 (<https://www.github.com/compas/grasp2018>), but it would be very beneficial for atomic physics research to have it there.

It should be kept in mind that the smallness of the cancellation factor does not always mean that the calculation is hopelessly unreliable. As demonstrated in my work on forbidden transitions of Fe V [9], there are many cases when transitions with very small CF values still have very accurate A values. For the spectrum studied in that work, uncertainties in A correlated much stronger with S than with CF.

Figure 24 of Rynkun et al. [45] illustrates an important subtlety: distributions of errors in the computed line strengths may be different for certain subgroups of transitions. One of the most common reasons for some transitions to be less accurate than others is a sharp difference in the amount of correlation effects included in the calculation for energy levels with different principle quantum numbers n . For example, in the work of Rathi and Sharma [49] on Na-like Ar, Kr, and Xe, the active layers included in the calculations were restricted by $n \leq 11$, while the results include transitions from levels with n up to 9. It is clear that the wavefunctions of states with $n = 8$ and 9 must be less accurate than those with smaller n , which was confirmed numerically by comparison of results of MCDHF and MBPT calculations performed by these authors. Thus, transitions from the $n = 8, 9$ levels should be separated from the rest, which was done in that work in estimation of their uncertainties.

3.1 Uncertainties in computed lifetimes: error propagation

The best method to estimate the uncertainties of theoretical radiative lifetimes is to compare them with accurately measured experimental values. However, such

benchmark data are available for only a limited number of energy levels in a limited number of spectra. Moreover, a database providing critically evaluated reference data for lifetimes does not exist. So, in estimating their uncertainties, theorists resort to scouring the literature and extracting the data they think are the most reliable. The most convenient way to do it is by using the NIST Atomic Transition Probability Bibliographic Database [50]. Even if the literature search provides some reference data, it is usually incomplete. So, theorists resort to alternative methods of uncertainty estimation.

The most commonly encountered incorrect practice is to compare the results in the length and velocity forms – similar to the method used for line strengths, but applied directly to lifetimes. One of the drawbacks of this method is that it cannot be complete: there exist metastable levels that decay only via magnetic transitions, for which the velocity form does not exist and cannot be used for comparisons. Second, comparison of length and velocity forms is controversial even for line strengths (see the previous section). Third, the lifetimes are computed from transition probability values, which have contributions not only from errors in line strengths, but also from wavelengths. These contributions can be eliminated when all transition probabilities are rescaled to experimental wavelengths. However, those are often unavailable for some levels, which requires careful estimation of uncertainties of calculated wavelengths. Many theorists do it incorrectly, by assuming a fixed percentage error assigned to all excited levels and applying the same percentage error to transition energies. Even for excitation energies, it is often incorrect, because excitation energies are small differences between two large quantities, the total energies of the two levels: the ground level and an excited level. The error in the total energy of the ground level is usually rather large and produces a systematic error in calculation of all excitation energies. An example of correct estimation of errors in calculated energy levels can be found in the work of Li et al. on N I [38]; see their Figure 1 and its discussion in the text. The work of that group on O I [39] provides a similar example of good practice.

Even when uncertainties of calculated excitation energies, E_{calc} , can be expressed as percentage errors p , wave numbers of transitions between excited levels have uncertainties that are not proportional to wave number (σ). If the errors in E_{calc} can be treated as uncorrelated, uncertainties of the calculated wave numbers are combinations in quadrature of the uncertainties of the two levels. If the levels are close (i.e., there is a long-wavelength transition between two close levels), the wave number uncertainty, u_{σ} , can easily be hundreds of times greater than p :

$$u_{\sigma} \approx \sqrt{2}pE_{\text{calc}} \quad (18)$$

(when σ is much smaller than both calculated levels).

The work of Rathi and Sharma [49] gives an example of a correct account of contributions from uncertainties

in line strengths S and in transition wavelengths in the total uncertainties of calculated transition probabilities; see their Eqs. (1) and (2).

If the computational errors in individual A_{ki} values (for transition from upper level k to lower level i) are statistically independent, relative uncertainty in the computed lifetime of level k , $u(\tau_k)$, can easily be determined using the standard formula for error propagation:

$$\frac{u(\tau_k)}{\tau_k} = \tau_k \sqrt{\sum_i u(A_{ki})^2}, \quad (19)$$

where the summation goes over all transitions to lower levels. This formula was applied, e.g., in the work of Ruczkowski and Elantkowska [51], see their Eq. (14); the paper of Rathi and Sharma [49] is another good example, see their Eq. (3).

An alternate method for estimation of lifetime uncertainties was used, e.g., by Singh et al. [52]. These authors have calculated the lifetimes of $n \leq 3$ levels of He-like W and Au for each successive layer of their active state calculations, which included virtual states up to $n = 7$, and compared the lifetime values from the last two layers. Although there is no evidence that their calculation has converged with respect to line strengths and lifetimes, and the contribution of errors in the computed wavelengths was not accounted for, such comparison can provide a reasonable estimate of possible errors.

4 Conclusions and outlook

This work has reviewed several recent developments in evaluation of uncertainties of transition wavelengths and calculated transition rates, which appeared in the literature after the publication of my last review [3]. The fact that many new interesting ideas have been published in the last several years is very encouraging, as it testifies to a certain success of my attempts to promote good practices of uncertainty estimation in atomic physics. The greatly increased percentage of theoretical papers containing uncertainty estimates (from $< 2\%$ seven years ago to about 10% in the last few years) is also very gratifying. However, the remaining 90% are still published without uncertainty evaluation. Considering the rate at which these papers are published (for the topic of transition probabilities, steady at about 150 papers per year, see Fig. 4⁶), it is clear that no single

⁶ The data for most recent years are not shown in Fig. 4 because of their incompleteness. They are always available by directly querying the databases referred to in the Figure caption. One should keep in mind that about 70% of the relevant references are usually found with only a short time delay of about two months. To find the remaining 30% takes a few years. Completeness of coverage varies depending on the subject. The most complete coverage of literature is for

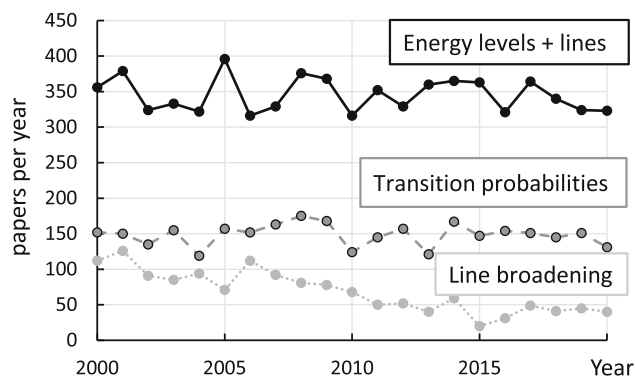


Fig. 4 Number of new papers added to the three NIST bibliographic databases per publication year, see Refs. [50, 53, 54]

person or agency is able to critically evaluate the massive amount of theoretical results continuing to appear in literature. I intentionally call them “results” rather than “data,” because without uncertainties they are not data. In my opinion, publication of any results, whether experimental or theoretical, without carefully analyzed uncertainties should be banned altogether, as their only effect is confusion.

With experimental papers, there is a similar problem. Although they usually do contain estimates of measurement uncertainties for new reported measurements, most of them are limited to a small number of observed spectral lines in a narrow range of wavelengths. To be incorporated in updated sets of reference atomic data, these newly measured wavelengths must be considered together with all other previously published data on the studied spectrum to derive a consistent set of energy levels that fit all observations. This means that, if the authors are revising the previous data on energy levels by using their new line measurements, they should evaluate not only their own wavelength uncertainties, but also those of previous works, which may concern different sets of spectral lines not directly related to the authors’ measurements but affecting the energy levels involved. In such cases, which are very common, ensuring consistency requires careful assessment of old measurements, which often do not provide uncertainties for every measured line.

I hope that the methods of uncertainty evaluation described in the present work will help researchers to make progress in these matters, and deficiencies in the existing methodology will receive further treatment.

Acknowledgements The Author is grateful to Dr. Andrew L. Rukhin and Gregory W. Haber of the NIST Statistical Engineering Division for their help in implementing computer codes for clustered maximum likelihood uncertainty estimation.

optical data on atomic structure, which is the primary focus of the NIST Atomic Spectroscopy Group. For line broadening and shifts, coverage is less complete.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: This article has no associated data beyond those included in the main text and the Supplementary Information.]

Declarations

Conflict of interest and Funding There are no Conflict of interest associated with this article. This research received no external funding.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Statistical toolbox user guide

The Statistical Toolbox included as a supplementary online material for this article contains the following groups of functions:

- Functions for estimation of dark uncertainties;
- Functions for calculation of weighted mean and its uncertainty;
- A function for making normal probability plots;
- A function for generating normally distributed random numbers;
- A number of functions internally used by the above.

The functions for estimation of dark uncertainties are `dark_unc_MP` (Mandel–Paule estimator), `dark_unc_DL` (DerSimonian–Laird estimator), `dark_unc_CMLE` (clustered maximum likelihood estimator), and `dark_unc_CRMLE` (clustered reduced maximum likelihood estimator). All of them follow definitions given by Rukhin [4–6] and take the same input parameters: a set of spreadsheet cells containing values of the analyzed quantity and a set of cells containing the corresponding standard uncertainties.

The function `dark_unc_MP` is most often used in my practice, so its interface is somewhat better developed. Its input arrays of cells can either be both vertical (columns) or both horizontal (rows) of cells, and these arrays can contain empty cells, which is convenient when analyzing large arrays of data. The only requirement is that every non-blank value of the analyzed quantity must have a corresponding non-zero uncer-

tainty. This function can be called directly from an Excel spreadsheet by writing “=dark_unc_MP(...)” in a cell; the returned value of the dark uncertainty will be in the same cell.

All other dark uncertainty functions accept only vertical ranges of cells as input, and they do not accept blank values between non-blank ones.

The function `dark_unc_DL` is included only for completeness, as some users may be curious about its results as compared to the MP estimator. So far, I did not find any cases where it could be useful for atomic spectroscopy measurements. However, it has some statistical foundation and has been used in other applied sciences.

The functions `dark_unc_CMLE` and `dark_unc_CRMLE` are “safe” (i.e., do not result in data corruption) only when called from a subroutine such as the one included in the toolbox. It is possible to call them from a spreadsheet by specifying the formulae in certain cells, e.g., “=dark_unc_CMLE(...)”. However, after each use of such a formula, the user must delete the equation, e.g., by copying the cell with it and pasting its value back in the same cell. This is because, unlike normal Excel functions, these two modify contents not only of the cell with the formula, but also of several cells below it.

The functions `WMDU` and `unc_WMDU` are used for calculation of the weighted mean and its uncertainty, respectively, in presence of non-zero dark uncertainties. Both take the same three input parameters, which are columnar ranges of cells containing the quantity values, their nominal uncertainties, and the corresponding dark uncertainties. It can also be used for calculating the weighted mean without dark uncertainties. For that, the cells specified for the dark uncertainties must be empty or zero.

For making a normal probability (NP) plot, it is necessary to sort the normalized residual values, R_i in increasing order, number them sequentially, and calculate the value of the percent point function of the uniform order statistic median of the normal distribution [16] for each value of the sequential index i . For the latter calculation, the function included in the Toolbox is called “G”. The calculation of the uniform order statistic median, U_i , is not used anywhere outside of this function, so it is simply hard-coded within the G function, making its use very simple. It takes two parameters: the sequential index, i , and the total number of residuals, N . The sorted R_i values are plotted against the values of $G(i, N)$ to display the NP plot.

Function `RandNormal()` is used to generate normally distributed random numbers for numerical simulations. It takes two parameters: the mean and standard deviation of the normal distribution.

References

1. B.N. Taylor, C.E. Kuyatt, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. National Institute of Standards and Tech-

- nology, Gaithersburg (2001). <http://physics.nist.gov/TN1297>. Accessed 27 July 2023
2. Joint Committee for Guides in Metrology, Working Group 1 (JCGM/WG 1): JCGM 100:2008: Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. International Bureau of Weights and Measures (BIPM), Sèvres (2008). <https://www.bipm.org/en/committees/jc/jcgm/publications>. Accessed 6 June 2023
 3. A. Kramida, Critical evaluation of data on atomic energy levels, wavelengths, and transition probabilities. *Fusion Sci. Technol.* **63**, 313–323 (2013). <https://doi.org/10.13182/FST13-A16437>
 4. A.L. Rukhin, Weighted means statistics in interlaboratory studies. *Metrologia* **46**, 323–331 (2009). <https://doi.org/10.1088/0026-1394/46/3/021>
 5. A.L. Rukhin, Homogeneous data clusters in interlaboratory studies. *Metrologia* **56**, 035002 (2019). <https://doi.org/10.1088/1681-7575/ab1559>
 6. A.L. Rukhin, Estimating heterogeneity variances to select a random effects model. *J. Statist. Plann. Inference* **202**, 1–13 (2019). <https://doi.org/10.1016/j.jspi.2018.12.003>
 7. V.I. Azarov, A. Kramida, Y. Ralchenko, A critical compilation of experimental data on the $1s2l2l'$ core-excited states of Li-like ions from carbon to uranium. *At. Data Nucl. Data Tables* **149**, 101548 (2023). <https://doi.org/10.1016/j.adt.2022.101548>
 8. M. Allegrini, E. Arimondo, L.A. Orozco, Survey of hyperfine structure measurements in alkali atoms. *J. Phys. Chem. Ref. Data* **149**, 043102 (2022). <https://doi.org/10.1063/5.0098061>
 9. A. Kramida, Assessing uncertainties of theoretical atomic transition probabilities with Monte Carlo random trials. *Atoms* **2**, 86–122 (2022). <https://doi.org/10.3390/atoms2020086>
 10. S.M. Attia, F. El-Sayed, Relativistic excitation energies and transition data for in XL and Sn XLI. *J. Quant. Spectrosc. Radiat. Transf.* **283**, 108139 (2022). <https://doi.org/10.1016/j.jqsrt.2022.108139>
 11. X.H. Zhang, G. Del Zanna, K. Wang et al., Benchmarking multiconfiguration Dirac–Hartree–Fock calculations for astrophysics: Si-like ions from Cr XI to Zn XVII. *Astrophys. J. Suppl. Ser.* **257**, 56 (2021). <https://doi.org/10.3847/1538-4365/ac2a3f>
 12. A. Kramida, Estimation of uncertainties of calculated oscillator strengths, in Invited Talk at the 12th International Colloquium on Atomic Spectra and Oscillator Strengths (ASOS12), São Paulo, Brazil, 4–7 July 2016 (unpublished) (2016)
 13. E. Tiesinga, P.J. Mohr, D.B. Newell, B.N. Taylor, CODATA recommended values of the fundamental physical constants: 2018. *J. Phys. Chem. Ref. Data* **50**, 033105 (2021). <https://doi.org/10.1063/5.0064853>
 14. Wikipedia Article on Weighted Arithmetic Mean. https://en.wikipedia.org/wiki/Weighted_arithmetic_mean. Accessed 8 Dec 2023 (see section “Weighted sample variance”)
 15. Gnu.org: GNU Scientific Library—Reference Manual: Weighted Samples. <https://www.gnu.org/software/gsl/doc/html/statistics.html#weighted-samples>. Accessed 16 June 2023
 16. J.J. Filliben, A. Heckert, Normal probability plot, in NIST/SEMATECH E-Handbook of Statistical Methods, Chap. 1.3.3.21. National Institute of Standards and Technology, Gaithersburg (2013). <https://www.itl.nist.gov/div898/handbook/>. Accessed 16 June 2023
 17. J.-Y. Liao, S.-N. Zhang, Y. Yao, Wavelength measurements of K transitions of oxygen, neon, and magnesium with X-ray absorption lines. *Astrophys. J.* **774**, 116 (2013). <https://doi.org/10.1088/0004-637X/774/2/116>
 18. J.E. Lawler, J.R. Schmidt, E.A. Den Hartog, Energy levels of singly ionized and neutral zirconium. *J. Quant. Spectrosc. Radiat. Transf.* **289**, 108283 (2022). <https://doi.org/10.1016/j.jqsrt.2022.108283>
 19. A. Kramida, Y. Ralchenko, J. Reader, NIST ASD Team: NIST Atomic Spectra Database (Version 5.10). National Institute of Standards and Technology, Gaithersburg (2022). <https://physics.nist.gov/asd>. Accessed 19 July 2023. <https://doi.org/10.18434/T4W30F>
 20. A.E. Kramida, The program LOPT for least-squares optimization of energy levels. *Comput. Phys. Commun.* **182**, 419–434 (2011). <https://doi.org/10.1016/j.cpc.2010.09.019>
 21. M.S. Safronova, U.I. Safronova, Relativistic many-body calculation of energies, oscillator strengths, transition rates, lifetimes, polarizabilities, and quadrupole moment of a Fr-like Th IV ion. *Phys. Rev. A* **87**, 062509 (2013). <https://doi.org/10.1103/PhysRevA.87.062509>
 22. W.L. Wiese, The critical assessment of atomic oscillator strengths. *Physica Scripta* **65**, 188–191 (1996). <https://doi.org/10.1088/0031-8949/1996/T65/028>
 23. C. Froese Fischer, G. Gaigalas, P. Jönsson, J. Bieroń, GRASP2018-A Fortran 95 version of the general relativistic atomic structure package. *Comput. Phys. Commun.* **237**, 184–187 (2019). <https://doi.org/10.1016/j.cpc.2018.10.032>
 24. M.F. Gu, The flexible atomic code. *Can. J. Phys.* **86**, 675–689 (2008). <https://doi.org/10.1139/P07-197>
 25. M.F. Gu, Flexible Atomic Code, v. 1.1.5 (2017). <https://www-amdis.iaea.org/FAC/>. Accessed 20 June 2023
 26. C. Froese Fischer, Evaluating the accuracy of theoretical transition data. *Phys. Scr.* **T134**, 014019 (2009). <https://doi.org/10.1088/0031-8949/2009/T134/014019>
 27. J. Ekman, M.R. Godefroid, H. Hartman, Validation and implementation of uncertainty estimates of calculated transition rates. *Atoms* **2**, 215–224 (2014). <https://doi.org/10.3390/atoms2020215>
 28. B. Atalay, P. Jönsson, T. Brage, Extended relativistic multiconfiguration calculations of energy levels and transition properties in singly ionized tin. *J. Quant. Spectrosc. Radiat. Transf.* **294**, 108392 (2023). <https://doi.org/10.1016/j.jqsrt.2022.108392>
 29. S. Manai, S. Ben Nasr, D. Elhak Salhi, H. Jelassi, Multi-configuration Dirac–Hartree–Fock and relativistic many body perturbation theory calculations of energy levels, wavelengths, weighted oscillator strengths, transitions rates, lifetimes, isotope shifts, HF constants and Landé g -factors of Zr XXXIX. *At. Data Nucl. Data Tables* **148**, 101536 (2022). <https://doi.org/10.1016/j.adt.2022.101536>
 30. P. Jönsson, G. Gaigalas, P. Rynkun, L. Radziūtė, J. Ekman, S. Gustafsson, H. Hartman, K. Wang, M. Godefroid, C. Froese Fischer, I. Grant, T. Brage, G.

- Del Zanna, Multiconfiguration Dirac–Hartree–Fock calculations with spectroscopic accuracy: applications to astrophysics. *Atoms* **5**(2), 16 (2017). <https://doi.org/10.3390/atoms5020016>
31. F. El-Sayed, Energy levels, wavelengths, and transition probabilities for transitions in Pd XLII. *J. Quant. Spectrosc. Radiat. Transf.* **254**, 107204 (2020). <https://doi.org/10.1016/j.jqsrt.2020.107204>
 32. Z.L. Zhao, K. Wang, S. Li et al., Multi-configuration Dirac–Hartree–Fock calculations of forbidden transitions within the $3d^k$ ground configurations of highly charged ions ($Z = 72\text{--}83$). *At. Data Nucl. Data Tables* **119**, 314–353 (2018). <https://doi.org/10.1016/j.adt.2017.01.002>
 33. R.D. Cowan, *The Theory of Atomic Structure and Spectra* (University of California Press, Berkeley, 1981)
 34. A. Kramida, *A Suite of Atomic Structure Codes Originally Developed by RD Cowan Adapted for Windows-Based Personal Computers* (National Institute of Standards and Technology, Gaithersburg, 2021). <https://doi.org/10.18434/T4/1502500>
 35. A. Corney, *Atomic and Laser Spectroscopy* (Oxford University Press Inc, New York, 1977)
 36. W.L. Wiese, J.R. Fuhr, Accurate atomic transition probabilities for hydrogen, helium, and lithium. *J. Phys. Chem. Ref. Data* **38**, 565–726 (2009). <https://doi.org/10.1063/1.3077727>. (Erratum: *J. Phys. Chem. Ref. Data* **38**, 1129 (2009))
 37. O. Jitrik, C.F. Bunge, Transition probabilities for hydrogen-like atoms. *J. Phys. Chem. Ref. Data* **33**, 1059–1070 (2004). <https://doi.org/10.1063/1.1796671>
 38. M.C. Li, W. Li, P. Jönsson, A.M. Amarsi, J. Grumer, Extended MCDHF calculations of energy levels and transition data for N I. *Astrophys. J. Suppl. Ser.* **265**, 26 (2023). <https://doi.org/10.3847/1538-4365/acb705>
 39. W. Li, P. Jönsson, A.M. Amarsi, M.C. Li, J. Grumer, Extended atomic data for oxygen abundance analyses. *Astron. Astrophys.* **674**, A54 (2023). <https://doi.org/10.1051/0004-6361/202245645>
 40. A. Papoulia, J. Ekman, G. Gaigalas et al., Coulomb (velocity) gauge recommended in multiconfiguration calculations of transition data involving Rydberg series. *Atoms* **7**, 106 (2019). <https://doi.org/10.3390/atoms7040106>
 41. M. Bilal, A.V. Volotka, R. Beerwerth et al., High-precision calculations of the $1s^2 2s 2p \ ^1P_1 \rightarrow 1s^2 2s^2 \ ^1S_0$ spin-allowed E1 transition in C III. *Phys. Rev. A* **99**, 062511 (2019). <https://doi.org/10.1103/PhysRevA.99.062511>
 42. I.P. Grant, Gauge invariance and relativistic radiative transitions. *J. Phys. B* **7**, 1458–1475 (1974). <https://doi.org/10.1088/0022-3700/7/12/007>
 43. I.P. Grant, *Relativistic Quantum Theory of Atoms and Molecules: Theory and Computation* (Springer, New York, 2007)
 44. Z. Rudzikas, *Theoretical Atomic Spectroscopy* (Cambridge Univ. Press, Cambridge, 2007)
 45. P. Rynkun, S. Banerjee, G. Gaigalas et al., Theoretical investigation of energy levels and transition for Ce IV. *Astron. Astrophys.* **658**, A82 (2022). <https://doi.org/10.1051/0004-6361/202141513>
 46. L. Radziūtė, G. Gaigalas, Energy levels and transition properties for As-like ions Se II, Br III, Kr IV, Rb V, and Sr VI. *At. Data Nucl. Data Tables* **147**, 101515 (2022). <https://doi.org/10.1016/j.adt.2022.101515>
 47. A. Hibbert, Successes and difficulties in calculating atomic oscillator strengths and transition rates. *Galaxies* **6**, 77 (2018). <https://doi.org/10.3390/galaxies6030077>
 48. G. Gaigalas, P. Rynkun, L. Radziūtė et al., Energy level structure and transition data of Er^{2+} . *Astrophys. J. Suppl. Ser.* **248**, 13 (2020). <https://doi.org/10.3847/1538-4365/ab881a>
 49. S. Rathi, L. Sharma, Extended calculations of atomic structure parameters for Na-like Ar, Kr and Xe ions using relativistic MCDHF and MBPT methods. *Atoms* **10**, 131 (2022). <https://doi.org/10.3390/atoms10040131>
 50. A. Kramida, J.R. Fuhr, NIST Atomic Transition Probability Bibliographic Database (Version 9.0). National Institute of Standards and Technology, Gaithersburg (2010). <https://physics.nist.gov/fvalbib>. Accessed 22 June 2023. <https://doi.org/10.18434/T46C7N>
 51. J. Ruczkowski, M. Elantkowska, Semi-empirical determination of radiative parameters for metastable Sc II levels. *J. Quant. Spectrosc. Radiat. Transf.* **277**, 107996 (2022). <https://doi.org/10.1016/j.jqsrt.2021.107996>
 52. N. Singh, A. Goyal, S. Aggarwal, Study of excitation energies of doubly excited states and identification of EUV, SXR and HXR spectral lines in M1 transitions of W LXXIII and Au LXXVIII. *J. Electron Spectrosc. Relat. Phenom.* **257**, 147205 (2022). <https://doi.org/10.1016/j.elspec.2022.147205>
 53. A. Kramida, NIST Atomic Energy Levels and Spectra Bibliographic Database (Version 2.0). National Institute of Standards and Technology, Gaithersburg (2010). <https://physics.nist.gov/elevbib>. Accessed 22 June 2023. <https://doi.org/10.18434/T40K53>
 54. A. Kramida, J.R. Fuhr, NIST Atomic Spectral Line Broadening Bibliographic Database (Version 3.0). National Institute of Standards and Technology, Gaithersburg (2010). <https://physics.nist.gov/linebrbib>. Accessed 22 June 2023. <https://doi.org/10.18434/T4B59K>