DOI: 10.58895/ksp/1000128686-13 erschienen in:
OCM 2021 - 5th International Conference on Optical Characterization of Materials, March 17th – 18th, 2021, Karlsruhe, Germany : Conference Proceedings
DOI: 10.58895/ksp/1000128686 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000128686/

Line Spectra Analysis: A Cumulative Approach

Achim Kehrein[1] and Oliver Lischtschenko[2]

[1] Rhine-Waal University of Applied Sciences,
Faculty of Technology and Bionics,
Marie-Curie Str. 1, 47533 Kleve, Germany
[2] Ocean Insight - A Brand of Ocean Optics B.V.,
Maybachstr. 11, 73760 Ostfildern, Germany

Abstract An optical spectrometer uses detector pixels that measure the integrated intensity over a certain interval of wavelengths. These integrated pixel values are divided by the interval width and then interpreted as estimates of function values of the wanted spectral irradiance. Hence each pixel measurement constitutes an averaging process. But, averaging biases at maxima: pixel data feature lower maxima. This paper proposes the conceptual use of a cumulated spectrum to estimate spectral data. The integrated quantities are placed in their natural habitat. The motivation originates from the fact that pixel data as integrated quantities are exact values of the cumulated spectrum. Averaging becomes obsolete. There is no information loss. We start with a single spectral line. This "true" spectrum is blurred mimicking the instrument function of the spectrometer optics. For simplicity we consider the instrument function to have Gaussian shape. We integrate the blurred spectrum over subintervals to simulate the pixel measurements. We introduce a cumulated spectrum approach. We compare the cumulated approach with the approach that interpolates the averaged function value estimates of the non-cumulated spectrum. The cumulated approach requires only basic mathematical concepts and allows fast computations.

Keywords Spectrum, spectral fitting, cumulative spectrum, data processing, plasma, line emission, spectroscopy

1 Introduction

The analysis of emitted line spectra provides insight into the qualitative and quantitative composition of materials. Virtually all analysis is carried out by using non-linear fitting of (multi) Gaussians to the recorded data. This paper replaces the all-purpose non-linear fitting by direct computation of the peak parameters. This reduces dramatically the computational effort and eliminates the dependence on intelligent initial parameter guesses.

Probability theory or statistics knows two perspectives on Gaussians, i.e. on normal distributions. On the one hand there is famous bell-shaped curve, the probability density function (PDF). This is how we usually visualize the normal distribution. On the other hand there is the tabulated cumulative distribution function (CDF) with which we perform almost all calculations. The PDF possesses non-negative values and finite area. The CDF is a non-decreasing function.

A spectrum is also a function with non-negative spectral irradiance values and finite area which corresponds to the finite irradiance. This is the usual visualization of a spectrum. In this paper we add the equivalent perspective of a cumulated spectrum to perform our calculations. After all the irradiance of the spectrum is nothing but the cumulated spectral irradiance (integrated with respect to wavelength). This paper introduces how the cumulated approach can be used to fit Gaussian peaks to discrete spectral data in a simple fashion.

In 2020 the global market volume of applications relying on spectral line emission analyses is estimated at a trillion US-Dollar. Examples of these applications are semi-conductor production, quality control of physical coating processes [1], minimally invasive cancer surgery with live tissue diagnostics [2], natural resource exploration [3], and magnetic plasma fusion research [4].

2 Mathematical Model

We use a simple model. An object sampled by a spectrometer possesses a “true spectrum” which is a function that assigns to each wavelength a certain non-negative value, the spectral irradiance. The wavelength can have any positive value, so mathematically, the true spectrum is a

function S (signal) defined on the interval $(0, \infty)$ with non-negative values. We restrict our attention to a true spectrum with a single spectral line which we approximate by a Dirac delta function,

$$S(\lambda) = S_\mu \cdot \delta(\lambda - \mu). \tag{2.1}$$

2.1 Blurring - Spectral Lines Become Gaussians

A spectrometer blurs the true spectrum S by convoluting it with the so-called instrument function I. We call the result of this convolution $I * S$ the blurred spectrum,

$$B(\lambda) = \int_0^\infty I(\lambda - \ell) \cdot S(\ell)\, d\ell\ . \tag{2.2}$$

This blurring depends on the used spectrometer. In this paper we assume that the instrument function is sufficiently well approximated by the Gaussian shape

$$I(\lambda) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{\lambda^2}{2\sigma^2}} \tag{2.3}$$

with the width parameter[3] $\sigma > 0$. Selecting a Gaussian is purely illustrative. Any shape can be cumulated and treated analogously.

The blurring convolutes the line spectrum, equation (2.1), into the Gaussian peak

$$B(\lambda) = S_\mu \cdot I(\lambda - \mu)\ . \tag{2.4}$$

2.2 Pixel Detectors - Spectral Irradiances Are Locally Integrated

A spectrometer measures the blurred spectrum on a finite interval of wavelengths $[\lambda_0, \lambda_R]$. Its detector consists of finitely many pixels that detect adjacent parts of the blurred spectrum. Each pixel measures the cumulated spectral irradiance on a subinterval of wavelengths. Let the rth pixel cover the subinterval $[\lambda_{r-1}, \lambda_r]$, $1 \le r \le R$. Then the spectral irradiance of the blurred spectrum measured by the rth pixel is

$$B_r = \int_{\lambda_{r-1}}^{\lambda_r} B(\lambda)\, d\lambda\ . \tag{2.5}$$

[3] We do not want to call σ standard deviation, since this is a non-statistical use.

We divide these cumulated values by their subinterval widths to convert B_r into the same unit as the blurred spectrum and obtain the averaged values

$$B_r^* = \frac{B_r}{\lambda_r - \lambda_{r-1}} = \frac{1}{\lambda_r - \lambda_{r-1}} \cdot \int_{\lambda_{r-1}}^{\lambda_r} B(\lambda)\, d\lambda \ . \tag{2.6}$$

A spectrometer with R pixels provides the finitely many values B_1^*, B_2^*,...,B_R^* that approximate the blurred spectrum over the interval $[\lambda_0, \lambda_1] \cup \cdots \cup [\lambda_{R-1}, \lambda_R] = [\lambda_0, \lambda_R]$.

Each pixel value B_r^* – though obtained by averaging – can be interpreted as a function value estimate at the pixel's mean wavelength

$$\lambda_r^* = \frac{\lambda_{r-1} + \lambda_r}{2} \ . \tag{2.7}$$

Then the measured spectrum is a function whose graph consists of the finitely many points (λ_1^*, B_1^*), (λ_2^*, B_2^*), ..., (λ_R^*, B_R^*).

3 Proof of Concept - Matching a Gaussian Peak

As an example we consider a single blurred spectral line shown in Figure 3.1. This is a Gaussian peak centered at wavelength $\mu = 500$ nm with width parameter $\sigma = 1$ nm and irradiance 100 μW/cm^2 (integrated spectral irradiance). We assume that the wavelength interval from $\lambda_0 = 497$ nm till $\lambda_{11} = 503$ nm is detected by a row of eleven equidistant pixels. Their averaged values B_1^*, ..., B_{11}^* are shown as function value estimates. Figure 3.1 illustrates in particular the effect of averaging the concave function segment close to the maximum: averaging underestimates a maximum value.

3.1 Non-linear All-purpose Fitting of the Function Value Estimates

As control case for our example we use the all-purpose Matlab command `fit` to fit a Gaussian peak to the pixel values (λ_1^*, B_1^*),..., $(\lambda_{11}^*, B_{11}^*)$. The model function to be fitted is

$$\hat{B}(\lambda) = a \cdot e^{-\frac{(\lambda - b)^2}{c}} \tag{3.1}$$

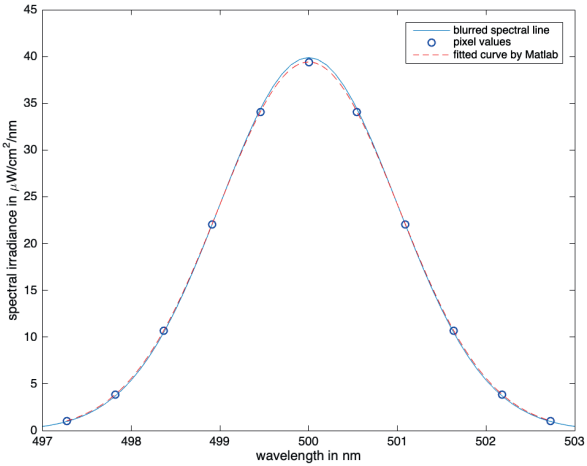

Figure 3.1: A Gaussian peak models a blurred spectral line. The wavelength interval $[497\,\text{nm}, 503\,\text{nm}]$ is divided into 11 equidistant subintervals (pixels). The pixel values are placed at the midpoints of the subintervals. The dashed curve is the fitted result by Matlab. The maximum of the fit is lower than the maximum of the blurred spectrum.

with model parameters a, b, and c. The underlying algorithm for nonlinear least squares requires intelligent initial parameters to produce convergence. We provided the initial parameters $a_0 = 39.4$ (maximal pixel value), $b_0 = 500$, and $c_0 = 2$.

Matlab finds $\hat{a} = 39.41$, $\hat{b} = 500$, and $\hat{c} = 1.432$ with 95% confidence intervals $(39.40, 39.41)$, $(500, 500)$, and $(1.432, 1.432)$ respectively. The estimated model parameters translate into: the line is located at wavelength 500 nm with width parameter $\sigma = \hat{c}/\sqrt{2} = 1.0126$ nm and irradiance $\hat{S}_\mu = \hat{a} \cdot \sigma\sqrt{2\pi} = 100.0287\ \mu\text{W/cm}^2$. Figure 3.1 also shows the fitted curve. It fits the blurred spectrum well, but features the biased smaller maximum. The estimated irradiance corresponds to the area under the curve. The fit overestimates the irradiance.

3.2 Cumulative Spectrum - Natural Use of Pixel Values

This section presents the alternative way to match a Gaussian peak to the pixel data. Instead of interpreting the pixel data as function value estimates, we use them exactly as the integrated quantities that they are. To do so, we consider the blurred spectrum cumulated over the measured interval

$$C(\lambda) = \int_{\lambda_0}^{\lambda} B(\ell)\, d\ell \quad , \lambda_0 \leq \lambda \leq \lambda_R \,, \tag{3.2}$$

and the cumulated pixel values

$$C_r = \sum_{i=1}^{r} B_i \quad , \qquad r = 0, 1 \ldots, R. \tag{3.3}$$

The cumulated pixel values are exact samples of the cumulated blurred spectrum at the subinterval endpoints, $C_r = C(\lambda_r)$, $r = 0, 1 \ldots, R$. For our example, these quantities are shown in Figure 3.2 and in Table 13.1.[4]

Table 13.1: Cumulated pixel values of example.

k	λ_k	C_k	k	λ_k	C_k	k	λ_k	C_k	k	λ_k	C_k
0	497.00	0.00	3	498.64	8.50	6	500.27	60.61	9	501.91	97.05
1	497.55	0.57	4	499.18	20.53	7	500.82	79.20	10	502.45	99.16
2	498.09	2.68	5	499.73	39.12	8	501.36	91.23	11	503.00	99.73

3.3 Estimating Parameters from Cumulative Values

The cumulative normal distribution function

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(\zeta-\mu)^2}{2\sigma^2}}\, d\zeta \tag{3.4}$$

[4]The blurred spectrum B (spectral irradiance) is the derivative of the cumulative spectrum C (irradiance). Consistently, the earlier introduced averaged values $B_r^* = (C_r - C_{r-1})/(\lambda_r - \lambda_{r-1})$ located at the midpoint λ_r^* turn out to be the centered differences known from numerical differentiation.

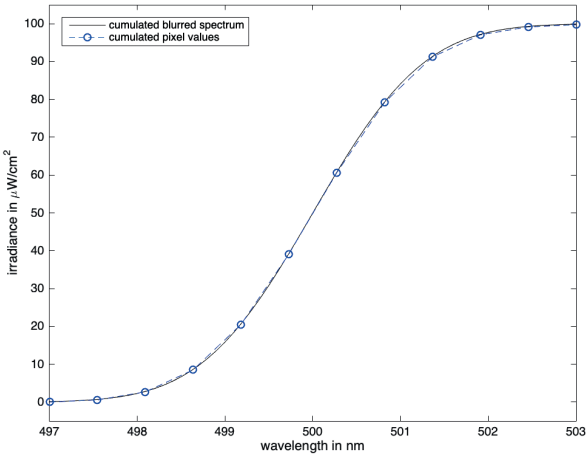

Figure 3.2: The cumulative Gaussian peak function (solid curve) and the cumulated pixel values at the respective endpoints of the subintervals. The dashed curve is the piecewise linear interpolation. The cumulated pixel values are exact values of the cumulative Gaussian peak function, not just estimates.

satisfies $F(\mu) = 0.5$ and can be normalized, $z = (x - \mu)/\sigma$, into the standard normal distribution $\Phi(z)$. In particular, we obtain the tabulated value

$$F(\mu - \sigma) = \Phi(-1) = 0.1587 \,. \tag{3.5}$$

We use these properties to estimate the parameters of the blurred spectrum.

The irradiance of the spectral line is estimated as the area under the measured blurred spectrum, which equals a difference in the cumulated spectrum,

$$\hat{S}_\mu = C_R - C_0 = \int_{\lambda_0}^{\lambda_R} B(\lambda)\, d\lambda \approx \int_0^\infty B(\lambda)\, d\lambda \approx S_\mu \,. \tag{3.6}$$

In Figure 3.2 this is $C_{11} - C_0 = 99.73 - 0 = 99.73$. After we will have obtained a first estimate for all parameters S_μ, μ, and σ, we will check for consistency and update them.

Now we estimate the wavelength μ of the spectral line from the cumulated pixel values. First, we find the subinterval $[\lambda_{j-1}, \lambda_j]$, in which the spectral line is located,

$$C_0 < \cdots < C_{j-1} < 0.5 \cdot \hat{S}_\mu \leq C_j < \cdots < C_R \ . \tag{3.7}$$

In Figure 3.2 this is $[\lambda_5, \lambda_6]$. We estimate $\hat{\mu}$ by linear interpolation,

$$\hat{\mu} = \lambda_{j-1} + \frac{0.5 \cdot \hat{S}_\mu - C_{j-1}}{C_j - C_{j-1}} (\lambda_j - \lambda_{j-1}) \ . \tag{3.8}$$

Since the cumulative normal distribution function is almost linear at the half-height value, linear interpolation is sufficient. However, the interpolation accuracy depends on the pixel width $[\lambda_{j-1}, \lambda_j]$. A narrower pixel width produces a better approximation. Figure 3.2 shows a good match between the cumulative blurred spectrum and the linear interpolation on $[\lambda_5, \lambda_6]$. We compute $\hat{\mu} = 500$ nm.

Now we estimate σ from the cumulated pixel values. We find the subinterval $[\lambda_{k-1}, \lambda_k]$, in which $\mu - \sigma$ is located (see equation 3.5),

$$C_0 < \cdots < C_{k-1} < 0.1587 \cdot \hat{S}_\mu \leq C_k < \cdots < C_R \ . \tag{3.9}$$

For the following interpolation we need $k \geq 2$ which is a reasonable assumption on the number and widths of the pixels covering the spectral line. In Figure 3.2 this is the interval $[\lambda_3, \lambda_4]$. The curvature over this interval is not negligible. Linear interpolation would systematically overestimate σ due to the convexity. Thus we compute $\hat{\mu}$ by quadratic interpolation through (λ_{k-2}, C_{k-2}), (λ_{k-1}, C_{k-1}), and (λ_k, C_k). [5] We use Newton's form of the interpolating parabola. The required divided differences are

$$[C_k, C_{k-1}] = \frac{C_k - C_{k-1}}{\lambda_k - \lambda_{k-1}} \quad , \quad [C_{k-1}, C_{k-2}] = \frac{C_{k-1} - C_{k-2}}{\lambda_{k-1} - \lambda_{k-2}} \ , \tag{3.10}$$

and

$$[C_k, C_{k-1}, C_{k-2}] = \frac{[C_k, C_{k-1}] - [C_{k-1}, C_{k-2}]}{\lambda_k - \lambda_{k-2}} \quad . \tag{3.11}$$

[5] We use the indices $k-2$, $k-1$, k instead of $k-1$, k, $k+1$ for numerical reasons.

The interpolating parabola is

$$P(\lambda) = C_k + [C_k, C_{k-1}](\lambda - \lambda_k) + [C_k, C_{k-1}, C_{k-2}](\lambda - \lambda_k)(\lambda - \lambda_{k-1}) \tag{3.12}$$

and we compute $\hat{\mu} - \hat{\sigma}$ from $P(\hat{\mu} - \hat{\sigma}) = 0.1587 \cdot \hat{S}_\mu$. We plug in

$$\begin{aligned} C_k &+ [C_k, C_{k-1}](\hat{\mu} - \hat{\sigma} - \lambda_k) \\ &+ [C_k, C_{k-1}, C_{k-2}](\hat{\mu} - \hat{\sigma} - \lambda_k)(\hat{\mu} - \hat{\sigma} - \lambda_{k-1}) = 0.1587 \cdot \hat{S}_\mu \end{aligned} \tag{3.13}$$

and sort for the powers of the unknown $\hat{\sigma}$,

$$\begin{aligned} 0 = {} & C_k + [C_k, C_{k-1}](\hat{\mu} - \lambda_k) - 0.1587 \cdot \hat{S}_\mu \\ & + [C_k, C_{k-1}, C_{k-2}](\hat{\mu} - \lambda_k)(\hat{\mu} - \lambda_{k-1}) \\ & + \big([C_k, C_{k-1}] + [C_k, C_{k-1}, C_{k-2}](\hat{\mu} - \lambda_k + \hat{\mu} - \lambda_{k-1})\big) \cdot \hat{\sigma} \\ & + [C_k, C_{k-1}, C_{k-2}] \cdot \hat{\sigma}^2 \end{aligned} \tag{3.14}$$

In the template $0 = a\hat{\sigma}^2 + b\hat{\sigma} + c$ we obtain the following coefficients for our example.

$$\begin{aligned} a &= [C_k, C_{k-1}, C_{k-2}] = 10.730 \\ b &= 22.278 + 10.730(0.82 + 1.36) = 45.669 \\ c &= 20.53 + 22.278 \cdot 0.82 - 15.827 + 10.730 \cdot 0.82 \cdot 1.36 = 34.937 \end{aligned} \tag{3.15}$$

From the geometry, the solution of the quadratic formula with the difference is the relevant one. The result rounded to four significant digits is

$$\hat{\sigma} = \frac{b - \sqrt{b^2 - 4a\,c}}{2a} = 1.000\ . \tag{3.16}$$

3.4 Updating the Estimates

Now that we have a complete set of estimates, $\hat{S}_\mu$, $\hat{\mu}$, and $\hat{\sigma}$ we use them to update the estimate of the irradiance S_μ. By design our first estimate of S_μ is an underestimate. We do not know how much of its area we actually considered. Now $\hat{\mu}$ and $\hat{\sigma}$ allow us to assess this underestimate. We can compute the area covered by the interval $[\lambda_0, \lambda_R]$ using

the tabulated normal distribution. This becomes particularly important when we have covered a wavelength range significantly smaller than $[\mu - 3\sigma, \mu + 3\sigma]$ (which corresponds to 99.7% of the area). This is also useful when we have an asymmetric interval about the location of the spectral line.

Such an update becomes important when spectral lines are located close to one another and their areas partially overlap. A similar problem arises when a spectral line is close to an endpoint λ_0 or λ_R so that a significant part of the blurred spectrum lies outside our observed interval.

For our simple example the numbers will not change much, but let us describe the updating process. Our interval was $[\lambda_0, \lambda_{11}] = [497\text{ nm}, 503\text{ nm}]$. With respect to our estimates $\hat{\mu} = 500$ nm and $\hat{\sigma} = 1$ nm the interval endpoints correspond to the standardized z-values -3 and 3 respectively (computed from $\lambda_i = \hat{\mu} + z\hat{\sigma}$). From the tabulated normal distribution we read off $\Phi(-3) = 0.0013$ and $\Phi(3) = 0.9987$. Therefore, the area covers $0.9987 - 0.0013 = 0.9974 = 99.74\%$ of the irradiance. Hence we update the irradiance estimate

$$\hat{S}_\mu = \frac{\text{original estimate}}{\text{covered area}} = \frac{99.73}{0.9974} = 99.99\ . \tag{3.17}$$

With this updated estimate we recalculate $\hat{\mu}$ and $\hat{\sigma}$. If the updated values deviate significantly from the original ones, we iterate this updating process until the estimates become stationary.

4 Outlook

We have only considered the simple example of an isolated spectral line in a sufficiently large wavelength interval. The results of the cumulated view point are promising. Only linear and quadratic equations were required in combination with only two tabulated values of the normal distribution. However, many aspects are still in need of further investigation:

- The whole process starts with the determination of an interval $[\lambda_0, \lambda_R]$ which covers most of the irradiance of a blurred spectral line. We have to automate the partitioning of a spectrum with several spectral lines into such intervals.

- We need to study examples of spectral lines whose blurred spectra overlap. Which separation of lines is required for the method to work properly? How will the method react when the lines are too close? Similarly, how do we handle spectral lines that are close to the endpoints of the measured interval?
- We want to transfer this method to other templates like a Lorentzian or a Voigt profile instead of a Gaussian peak.

The presented approach significantly reduces the calculation complexity of spectral matching by avoiding all-purpose non-linear fit routines. Due to the simplicity of the computation steps, the presented cumulative approach can be implemented in FPGAs. This allows immediate on-camera real-time data processing.

This affects many applications. Control circuits for online plasma analysis will benefit from shorter latencies. Processes like laser induced breakdown spectroscopy (LIBS) that so far require a complex post mortem analysis have the potential to be treated live. Adapting the approach to complex fusion spectra will significantly reduce the amount of currently needed a-priori knowledge. The whole analysis becomes more deterministic and less dependent on intelligent initial estimates. The only information to be determined in advance is the cumulative distribution of the instrument function.

References

1. A. Mackus, S. Heil, E. Langereis, H. Knoops, M. Van de Sanden, and W. Kessels, "Optical emission spectroscopy as a tool for studying, optimizing, and monitoring plasma-assisted atomic layer deposition processes," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol. 28, no. 1, pp. 77–87, 2010.
2. R. Kanawade, F. Mehari, C. Knipfer, M. Rohde, K. Tangermann-Gerk, M. Schmidt, and F. Stelzle, "Pilot study of laser induced breakdown spectroscopy for tissue differentiation by monitoring the plume created during laser surgery?an approach on a feedback laser control mechanism," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 87, pp. 175–181, 2013.
3. A. J. R. Bauer and S. G. Buckley, "Novel applications of laser-induced breakdown spectroscopy," *Applied spectroscopy*, vol. 71, no. 4, pp. 553–566, 2017.

4. M. G. von Hellermann, G. Bertschinger, W. Biel, C. Giroud, R. Jaspers, C. Jupén, O. Marchuk, M. O'Mullane, H. Summers, A. Whiteford *et al.*, "Complex spectra in fusion plasmas," *Physica scripta*, vol. 2005, no. T120, p. 19, 2005.