Tina Memo No. 2015-005
Internal.

Background Subtraction and Noise Modelling for Spectroscopy Data

Ashley Seepujak, Neil A Thacker, Paul D Tar

Last updated
25/03/2015

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Background Subtraction and Noise Modelling for Spectroscopy Data.

Ashley Seepujak, Neil A Thacker, Paul D Tar. 25/03/2015

Abstract

In order to apply analysis processes to spectroscopy data, it is generally necessary to first pre-process it to make the data more consistent with the assumed signal generation model. In this work, we are specifically interested in applying a linear Poisson model (LPM) [4], and need to confirm that background subtraction delivers data with specific (Poisson) noise characteristics.

A suitable background subtraction strategy for utilisation with noisy datasets is presented. The strategy is based upon an iterative process designed to reduce the biasing effects of control parameters, and is found to produce a numerically-stable output. It is demonstrated that a power-law error model fitted to RELAX mass spectrometry data [10], are approximately (within measurement error) consistent with Poisson-like behaviour. This represents a necessary condition for application of the LPM. Future work will investigate the sensitivity of LPM analyses to the specific level of approximation using Monte Carlo simulations [5].

1 Mathematical Background

1.1 The Linear Poisson Model

Accurate and unbiased quantitative separation of background and signal counts represents a common problem with experimentally-measured datasets such as mass spectra [4] and electron energy-loss spectra [6]. In general, such separation necessitates some form of background subtraction so that measured data can be directly fitted to assumed theoretical models of signal. The behaviour of highly-sensitive analogue amplifiers combined with digital circuitry and operation in a noisy environment generates a smoothly-varying background across any range of masses. However, any attempt to subtract background must result in data which is suitable for subsequent use, this requires some appreciation of algorithmic assumptions.

A popular approach to modelling complex data is to decompose using linear basis functions, as found in methods including principal component analysis, factor analysis and independent component analysis (ICA). These techniques are primarily designed for continuous data. They typically make the assumption of a Gaussian residual distribution (noise) which is either uniform, isotropic or negligible. Some formulations of these problems, especially ICA, explicitly exclude consideration of data noise for simplicity and tractability.

Histogram data has the properties of discrete, non-negative frequencies and non-uniform, non-isotropic distributed residuals. Whilst it is computationally possible to apply continuous techniques to histograms, it would be difficult to interpret results in a physically meaningful way. Results may include non-physical negative components and unknown destabilising effects due to propagated noise. Although the statistical definitions of histogram data are well known, there are fewer linear decomposition techniques designed for them specifically. Gaussian mixture models (GMMs) can describe histograms with Poisson bin noise (i.e. Y-axis direction), but only if the signal contains Gaussian distributions (i.e. multiple bell-shapes in the X-axis direction).

Linear Poisson models (LPMs) [4] are intended to describe histograms as linear combinations of more general probability mass functions (PMFs). These can be viewed as non-parametric versions of GMMs, where any PMFs can be used, rather than just Gaussians, thus,

$$\mathbf{H}_X \approx \mathbf{M}_X = \sum_k P(X|k)\mathbf{Q}_k \tag{1}$$

$\mathbf{H}_X$ is the observed frequency in bin X of histogram $\mathbf{H}$; $\mathbf{M}_X$ is the modelled frequency of bin X; $P(X|k)$ is the probability of an event occurring in bin X, originating from independent component k; and $\mathbf{Q}_k$ is the quantity of component k within the histogram. The estimation of PMFs requires multiple independent histogram examples. Solutions to the linear model are based upon maximising the following extended maximum likelihood function:

$$\ln \mathcal{L} = \sum_r \sum_X \ln \left[\sum_k P(X|k)\mathbf{Q}_{(r)k} \right] \mathbf{H}_{(r)X} - \sum_k \mathbf{Q}_{(r)k} \tag{2}$$

r is an independent example histogram; $\mathbf{H}_{(r)X}$ is the X bin for histogram r; and $\mathbf{Q}_{(r)k}$ is a linear model weight for component k in histogram r. Hence, all histograms are described using a common set of PMFs, $P(X|k)$, with histogram specific weights $\mathbf{Q}_{(r)k}$. This likelihood, which assumes independent Poisson distributed bins, is maximised using expectation maximisation (EM). Given a successful solution, statistical and systematic error covariances are estimated for model weighting quantities, $\mathbf{Q}$, via error propagation:

$$\mathbf{C} = \mathbf{C}_{stat} + \mathbf{C}_{sys} \tag{3}$$

$$\mathbf{C}_{ij(stat)} = \sum_X \left[\left(\frac{\partial \mathbf{Q}_i}{\partial \mathbf{H}_X} \right) \left(\frac{\partial \mathbf{Q}_j}{\partial \mathbf{H}_X} \right) \sigma^2_{\mathbf{H}_X} \right] \tag{4}$$

$$\mathbf{C}_{ij(sys)} = \sum_X \left[\sum_k \left(\frac{\partial \mathbf{Q}_i}{\partial P(X|k)} \right) \left(\frac{\partial \mathbf{Q}_j}{\partial P(X|k)} \right) \sigma^2_{P(X|k)} \right] \tag{5}$$

$\mathbf{C}_{stat}$ is a statistical error originating from incoming histogram data; $\mathbf{C}_{sys}$ is a systematic error originating from training data from which PMFs are sampled; $\sigma^2_{\mathbf{H}_X}$ is the variance of independent histogram bin X; and $\sigma^2_{P(X|k)}$ is the variance of the probability estimate $P(X|k)$. This error theory assumes independent noise on each bin (i.e. the sum over X with no cross-terms). It also assumes Poisson bin noise, as $\sigma^2_{\mathbf{H}_X}$ is taken to be equal to $\sqrt{< \mathbf{H}_X >}$. $\sigma^2_{P(X|k)}$ is computed using a similar but slightly more complex method, based upon the same Poisson assumption.

1.2 Indirect Poisson Data

LPMs assume that each entry within a histogram is a discrete count of an individual event. However, there are sources of histogram data where these discrete counts are only indirectly measured and/or contaminated by the measurement process. Raw mass spectra are an example of such histograms. A cascade multiplier might be used to amplify the detection of a single ion (one Poisson event), resulting in a peak in output current (a continuous change). A detector might also signal counts by using a dip in current rather than a peak, giving an inverted signal. Whilst underlying ion detections correspond to independent Poisson events, their recording is achieved using continuous electronic means and as such they are subject to electronic and thermal noise. Charging and discharging effects can also change the sensitivity of detectors if many events occur in close succession. The final result can be a Poisson-like histogram, but with significant amounts of background noise which can be non-uniform along the range of the spectrum.

Given that background noise is non-Poisson this must be removed if LPMs are to be applied. As mass spectrometry signals can also be scaled, this scaling factor must be estimated. Ultimately, a quantitative test is required to check the behaviour of indirect Poisson data to assess how closely it accords with conventional Poisson behaviour. The preprocessing of indirect Poisson data should aim to achieve the most Poisson-like behaviour possible in filtered outputs.

2 A Relatively Simple Approach to Background Subtraction

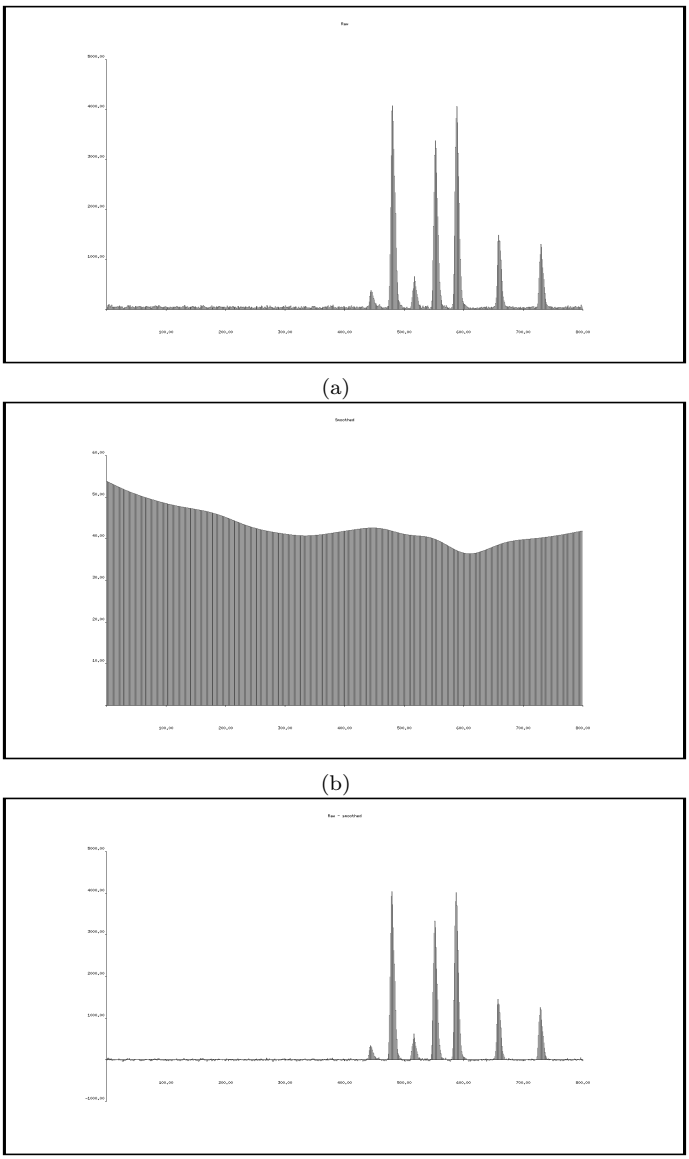
2.1 Background Estimation in the Presence of Missing Data

As a starting point for background subtraction (and assuming that signal is only a small fraction of the total measured data), continuous data $s(x)$ can be smoothed using a large scale filter to estimate the mean signal metric or a varying background component. This is performed by computing some form of locally-weighted mean. Numerical considerations generally favour the use of well-behaved and smooth-windowed functions, such as a Gaussian function. The functional form of the Gaussian peak can be described as,

$$G(b) \;=\; a \; exp\left[-\frac{b^2}{2w} \right] \tag{6}$$

In Eqn. (1), w represents a length scale, with a representing a normalisation constant chosen s.t. $\sum_b \; G(b) \;=\; 1$. Then,

$$m(x) \;=\; \sum_b \; s(x-b)G(b) \tag{7}$$



(a)

(b)

(c)

Figure 1: Stages of the background subtraction process. (a) A raw mass spectrum, in which the background is clearly evident. (b) A raw spectrum from which peaks have been extracted, and subsequently smoothed until convergence. (c) The residual of the raw spectrum detailed in Fig. 1(a) minus the smoothed spectrum detailed in Fig. 1(b).

Gaussian smoothing can be interpreted as a low-pass filter. The present approach is therefore justified, if the frequency components of the assumed signal process model are likely to be of lower spatial frequency than the implied length scale w. This condition is guaranteed for measurement systems which have a limiting resolution, in which case w must be smaller than, or close to, the standard deviation of the equivalent point-spread function. It may also be more generally true for larger values of w for the background generation process.

Computing a background, using all data (including signal) will provide over-estimation of background signal. It is therefore desirable to remove signal prior to smoothing. Thresholding, detailed in the following section, involves extraction of peaks above a pre-determined intensity from raw spectra. As a consequence, gaps in data, arise when background counts and signal counts are separated. In order to correctly account for this data we use the following background method. Defining

$$\delta(x - b \ \in z) = 0 \tag{8}$$

$$\delta(x - b \ \notin z) = 1 \tag{9}$$

The delta function assumes a binary value of either unity when the data $s(x-b)$ is known, or zero otherwise. Our background estimation is then given by

$$m(x) \ = \frac{\sum_b \ s(x-b)G(b)\delta(x-b \notin z)}{\sum_b \ G(b)\delta(x-b \notin z)} \tag{10}$$

The justification for this aproach is as follows. Missing data $x \in z$ can be replaced with the (unknown) local mean metric $m(x)$. Indeed, the local mean metric represents the unique value which will not bias the computed estimate. Assigning a delta function for two discrete regions, It can be observed that Eqn. (3), is invariant to the addition further datapoints equal to the mean, to the data vector. Eqn. (3) is therefore equivalent to,

$$m(x) \ = \ \frac{\sum_b \ s(x-b)G(b)\delta(x-b \ \notin z) \ + \ m(x)G(b)\delta(x-b \ \in z)}{\sum_b \ G(b)\delta(x-b \ \notin z) \ + \ G(b)\delta(x-b \in z)} \tag{11}$$

Rewriting Eqn. (6),

$$m(x) \ = \ \frac{\sum_b \ s(x-b)G(b)\delta(x-b \ \notin z) \ + \ m(x)G(b)\delta(x-b \ \in z)}{\sum_b \ G(b)} \tag{12}$$

$$= \ \sum_b \ [s(x-b)\delta(x-b \ \notin z) \ + \ m(x)\delta(x-b \ \in z)]G(b) \tag{13}$$

Eqn. (8) thus replaces unknown values with the resulting mean.

This approach can therefore be utilised for the estimation of a smoothed background, on the condition that there exists a reliable means to identify background (or equivalently, that which is not signal), in data.

2.2 Hysteresis Thresholding

Conventional thresholding involves the separation of signal data from the background (data $s(x)$ above the threshold t is assigned as signal). However, use of a fixed user-defined input for the threshold results in application of thresholding being a black art, causing significant variability in data processing. Moreover, the intrinsic nature of noisy data often makes a reliable and statistically-efficient detection of signal difficult, as too high a threshold will miss quite a lot of genuine signal. Lowering this threshold increases the identification of signal, but consequently increases the possibility of labelling noise as signal. Furthermore, thresholding only works appropriately when the background distribution has a uniform mean (ideally zero). The process can be made more statistically meaningful by relating thresholds to observed noise levels, typically the standard deviation of known background (eg, $t > 3\sigma$).

For the previous local-average calculation, imperfect thresholding of signal can also result in biased estimates of background, owing to signal contamination. When the signal has some form of structure, such as extended regions of nearby data being generated by the same underlying process (i.e. a peak), the reliability of signal detection can be increased using a process termed hysteresis thresholding. This involves the application of two thresholds, one upper threshold t_u, and one lower threshold t_l. All data above the upper threshold is immediately labelled as clear signal. The second threshold is used to identify additional data which is also likely to be signal, on the basis that it is near to (or locally connected to), known clear signal. Assuming that all initial data values are strictly greater than zero, data to be removed are set to zero. In a second pass of the entire data vector, if the value of datapoint x_i is 0, and the value of datapoint x_{i+1} is greater than the lower threshold, then the value of datapoint x_{i+1} is set to zero. In a third pass, if the value of x_i is already 0, and value of datapoint x_{i-1} is greater than the lower threshold, then the value of datapoint x_{i-1} is also set to zero. These thresholds can be established on the basis of assumed noise; typically $0 > t_l > 2\sigma$ and $2\sigma > t_u > 3\sigma$ represent reasonable choices.

Hysteresis thresholding presents two major advantages over conventional thresholding. Firstly, it increases the amount of signal identified within data, but also reduces the sensitivity to threshold selection. This is because any missed signal can generally be expected to be below the lower threshold, and can therefore be made insignificant. Furthermore, by being more able to identify signal (and therefore also background), running-average estimates of background such as those detailed in Section 1, are also expected to be more accurate.

2.3 Iterated Background Subtraction

Hysteresis thresholding necessitates the condition of a background with a zero mean. This precludes its immediate use for background subtraction, as it is a slowly-varying non-zero value of the background which requires estimation.

However, the process of hysteresis thresholding can be applied iteratively, so that successive iterations gradually converge on a self-consistent separation of the background and signal.

Following selection of initial thresholds t_{l1} and t_{u1}, the initial smoothing iteration is performed. For the following and subsequent smoothing iterations, the upper and lower thresholds t_{ln} and t_{un} (n=2, 3, ...) are adjusted, based upon the observed background variance, such that $t_{ln} = 2\sigma$ and $t_{un} = 4\sigma$. Convergence of the iterative process is attained when no difference exists between the zero regions of contiguous iterations of thresholded spectra.

As the estimate of smoothed background are completely deterministic given labelled signal and background, convergence is reached, typically after a small number (typically three) iterations. The combined approach of hysteresis thresholding along with smooth local-averaging is detailed in Fig. 1.

Local-Gaussian smoothing is trivial, in comparison to supposedly more sophisticated interpolation techniques such as Kriging [1-3]. Kriging assumes a zero-mean Gaussian process and associated correlation model, but with apparently no better justification than the present approach (Appendix A). Ultimately, we require only a sufficiently accurate estimate, such that any resulting errors in background subtraction are negligible, in comparison to other sources of measurement error. This requirement is easily tested in real data.

3 Experimental Details

The current document is intended as a simple demonstration of the approach used for confirming Poisson-like behaviour of background subtracted spectroscopy data. The specific data set data used here was acquired utilising air shot calibration (containing typically 1000 atoms of xenon), using the refrigerator enhanced laser analyser for xenon (RELAX) instrumentation [10]. Measurements are acquired using a micro-channelplate detector, which were then written to computer via a transient recording amplifier. Typical spectra are detailed in Fig. 1. These data were acquired sequentially, allowing investigation measurement repeatability using Bland-Altman plots (see below).

3.1 Bland-Altman Plots

A Bland-Altman plot represents a scatterplot of individual values of the data vector versus the fitted residual of that datapoint, and describes the reproducability of a parameter, over a large range of values for that parameter.

Bland-Altman plots can also be equivalently constructed in repeatability experiments, by plotting the difference between two repeat measurements R_{12} against the mean, (or any other statistically-equivalent values) for a set of data J. Its purpose is to establish any parametric dependency of the measurement error on the measurement itself (i.e. $x \pm f(x)$), for the design of algorithms using specific likelihood functions, or for confirming of the appropriateness of data for use in an existing algorithm. When taking differences between repeat values, the distribution of R_{12} (denoted henceforth as $\epsilon_{R_{12}}$) is a scaled function of $f(x)$.

A common error-dependency model, assumes a power-law function, i.e. $f(x_i) = ax_i^N$. A data vector x_i exhibiting a Poisson-like characteristic will possess an error function with a power-law function s.t. $N \approx \frac{1}{2}$. That is, the statistical error is proportional to the square root of the value of that datapoint. The systematic component assumes a linear function. If N $>$ 0.7, then residuals are greater than those predicted by Poisson noise. Overtraining is indicated by residuals being too small, typically N $<$ 0.3. In order to calculate the residual of a datapoint, a series of three consecutively-acquired spectra were selected, which we may denote as spectra n-1, n and n+1. The mean value of the intensity at each mass channel in spectra n-1 and n+1 was deduced, which was then subtracted from the intensity of each mass channel in spectrum n. The final values of parameters of the curve fitted to the datapoints, or their respective uncertainties, were invariant with changing the values of t_{l1} and t_{u1} by a factor of 2.

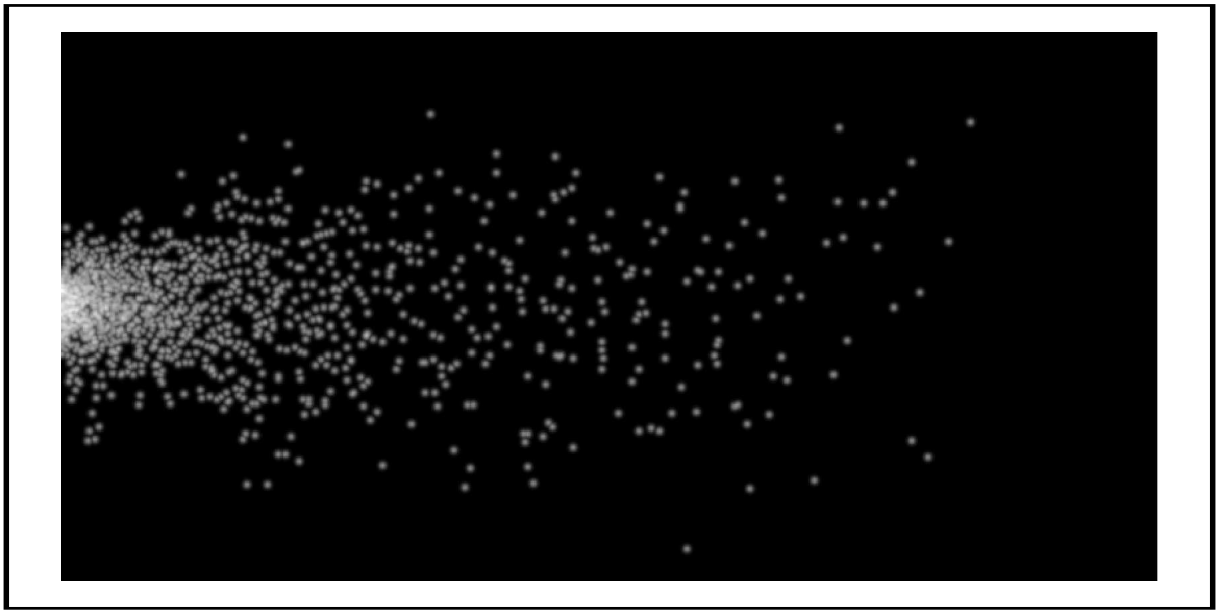
Datapoints from mass spectra (n=29) are plotted in a Bland-Altman plot, detailed in Fig. 2. Datapoints with value less than the observed level of background noise (in this case 50), are purposely excluded from the plot sample space, since these do not possess the same noise characteristics as the signal which we are interested in modelling.

3.2 Noise Model Fitting

We minimise the following log-likelihood cost function [7], in order to calculate fitting parameters,

$$-logL \ = \sum_{i=1}^{I} \left[\frac{1}{2} \left(\frac{R_{12}\,[i]}{\epsilon_{R_{12[i]}}} \right)^2 + log\left(\sqrt{2\pi} \epsilon_{R_{12[i]}} \right) \right] \tag{14}$$

The term $\frac{R_{12}[i]}{\epsilon_{R_{12[i]}}}$ is a Pull distribution with unit standard deviation. i represents the sample number for the whole population I. The error model for $R_{12[i]}$ is given by $\epsilon_{R_{12[i]}}$. The fit was perfomed with iterative optimisation using the Simplex algorithm of Nealder and Meade. Final values of the fitting parameters were a=9.482800 ± 1.678087, and N=0.465157 ± 0.027671. Estimated errors were determined numerically.



(a)

Figure 2: A Bland-Altman plot of residuals versus the value of the datapoint. The value of the index N=0.465157 ± 0.027671, equal to that of a square-root function (to within 3 s.d.), represents a test for Poisson-like data.

4 Conclusion

In order to confirm the validity of using the LPM approach with a mass spectrum data vector, it is critical to deduce whether data are consistent with Poisson statistics. If the Poisson statistics condition is met, a power-law function fitted to the Bland-Altman plot has an index of $\frac{1}{2}$ to within two standard deviations of the error, i.e. the function is square-root in form. We confirm that in the case of RELAX mass spectra data, Poisson statistics are indeed obtained, within the measured limits.

In the present study, we are intending to apply LPM analyses to a variety of other mass spectra data, including matrix-assisted laser desorption ionisation (MALDI) data. This will be accompanied by Monte Carlo simulation work. The next stage of the study will be to confirm the Poisson-like nature of MALDI data, and to assess the effects of any departures from Poisson statistics using Monte Carlo simulations [5].

Appendix A: Gaussian Process Interpolation

Gaussian smoothing is based upon making assumptions regarding the nature of correlations between nearby data samples. Precisely what these assumptions are, are not trivial to recognise. Ideally, any principled analysis should be based upon probability theory, in a way which relates the parameters we wish to estimate to the data available, in this case $p(s_2|s_1)$. A conventional likelihood fit using a parametric form for the underlying signal generator would be sufficient to achieve this. Gaussian smoothing could even be extended to linear prediction, however it is not sensible to make such model assumptions when dealing only with noise. One body of theory which is consistent with the requirement is the Gaussian process (GP) model.

Consider a sample of data $s(x)$ extracted from a multivariate Gaussian distribution with zero mean and covariance C. Decomposing s into two sub-vectors s_1 and s_2 so that $s \approx N(0, C)$ gives,

$$\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \approx N\left(0, \begin{bmatrix} D & E^T \\ E & F \end{bmatrix}\right) \tag{15}$$

The conditional distribution of s_2 is itself a Gaussian distribution. If matrix C were diagonal, then the knowledge of s_1 would tell us nothing regarding s_2, since $s_1 and s_2$ are orthogonal. If however, the off-diagonal terms E are non-zero then,

$$p(s_2|s_1) = N(ED^{-1}s_1, F - ED^{-1}E^T) \tag{16}$$

Then, our best estimate of the vector s_2 is,

$$\bar{s_2} = ED^{-1}s_1 \tag{17}$$

Eqn. (17) elucidates the link between conventional Gaussian smoothing and GP regression (Kriging) [1-3]. The parameters E and D^{-1} represent the linear weighting coefficients previously assumed to be the Gaussian kernal, with appropriate modification for missing data.

For many applications, this interpolation scheme is applied by assuming a particular parametric form for the coefficients of C, for example,

$$C_{ij} = \theta_1 \, exp\left[-\frac{(x_i - x_j)^2}{2\theta_2^2}\right] \tag{18}$$

More generally, for a radial function $C_{ij} = f(|x_i - x_j|, \theta)$, for which parameters θ require estimation.

In comparison, with conventional smooth interpolation, this method changes the assumption, from that of a parametric smoothing function (e.g. Gaussian weight factors), to the more fundamental GP assumption of a particular parametric form of correlation. At this point, it is debatable which of these two forms of assumption is more appropriate.

Consider the special case of data being on a regular lattice ($x_i = i\Delta_x$), and applying the GP model. The parametric form can be replaced with direct estimates of correlation factors, based on known samples of data with a specific separation denoted by $b = \frac{(x_i - x_j)}{\Delta_x}$, i.e.,

$$C_b = \frac{1}{N}\sum_x (s_1(x_i) - s_0))^T \otimes (s_1(x_j) - s_0) \tag{19}$$

In Eqn. (19), s_0 represents the mean of the data vector $s_i(x)$. If σ represents the estimated noise of the data, then the covariance matrix C is given by,

$$C = \begin{bmatrix} \sigma & C_1 & C_2 & C_3 & .. & C_N \\ C_1 & \sigma & C_1 & C_2 & .. & C_{N-1} \\ C_2 & C_1 & \sigma & C_1 & .. & C_{N-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ C_N & C_{N-1} & C_{N-2} & C_{N-3} & .. & \sigma \end{bmatrix} \tag{20}$$

The rows and columns of C require adjustment, in order to group together known and unknown data vectors (s_1 and s_2). In this way, our estimates of C, and therefore s_2, become nonparametric. The obvious similarity between the elements of this matrix and auto-regression should also be noted.

It would be naive to assume that owing to GPs being more fundamental, that they represent the inherently correct assumption to apply in all situations. Although we can always compute a correlation for any set of data, the existence of a fixed dependency of C_{ij} on $|x_i - x_j|$ cannot be considered to be generally applicable. For example, if we split the data vector into two groups (such as high x_i and low x_i), there is no guarantee that the corresponding correlation coefficients would be consistent.

Kriging must be rather ineffective in comparison to conventional interpolation methods, if there exists a genuine parametric form of signal, which generates varying effective correlations across data subsets. Then, data cannot really be assumed to be a Gaussian random variable with zero mean without introducing significant estimation bias on extreme values. The GP assumptions are more justifiable when applied to background estimation, except

we have already seen there may be some problems with independent noise (which is a real possibility when performing background subtraction). Although we would probably not be able to predict in advance which of the two alternatives (Kriging or Gaussian smoothing) are more likely to most efficiently interpolate real data, we could easily evaluate either approach, by using cross-validation.

Acknowledgements

Funding from Leverhulme is gratefully acknowledged. We thank Jamie Gilmour and Adam McMahon for data and useful discussions.

References

[1] D. G. Krige. A Statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand. MSc thesis, University of Witwatersrand (1951).

[2] G. Matheron. Principles of Geostatistics. Economic Geology, 58 1246 (1963).

[3] G. Matheron. The Intrinsic Random Functions, and their Applications. Adv. Appl. Prob., 5 439 (1973).

[4] P. D. Tar and N. A. Thacker. Linear Poisson Models: A Pattern Recognition Solution to the Histogram Composition Problem. Tina Memo 2013-006.

[5] W. R. Gilks, S. Richardson, D. J. Spiegelhalter, (Eds.). Markov Chain Monte Carlo in Practice. Boca Raton, FL: Chapman and Hall (1996).

[6] A. Seepujak et al. Red-Shift and Optical Anisotropy in Multi-Walled Nanotubes. Physical Review B.

[7] H. Ragheb, N. A. Thacker, R. Pathak, D. M. Morris and A. Jackson. Multi-site Liver Tumour ADC Reproducibility at 1.5 T. Tina Memo 2014-007.

[8] M. Guilhaus. Principles and Instrumentation in Time-of-Flight Mass Spectrometry. Journal of Mass Spectrometry 30 1519 (1995).

[9] W. H. Press et al. Numerical Recipies in C. Cambridge University Press (1991).

[10] S. Crowther, R. Mohapatra, G. Turner, D. Blagburn, K. Kehm, J. Gilmour. Characteristics and Applications of RELAX, an Ultrasensitive Resonance Ionization Mass Spectrometer for Xenon. Journal of Analytical Atomic Spectrometry 23 938 (2008).