Tina Memo No. 2016-009
Internal.

Estimating Noise Models for Arbitrary Images.

N. A. Thacker, S. Deepaisarn and A. McMahon.

Last updated
27 / 4 / 2016

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

Estimating Noise Models for Arbitrary Images.

Abstract

Whilst image processing routines are often treated as general purpose black boxes, it is clear that all images are not created equally, and some algorithms have difficulty with some forms of data. More formally, we can characterise images in terms of the statistical properties and we can assess the consequences of these properties for subsequent analysis using methods such as error propagation. A quantitative assessment of an image (random noise behaviour) can also allow image acquisition to be optimised.

In the simplest case, the noise on images can be assumed to be homogeneous and Gaussian. Relatively simple methods exist to allow noise to be estimated from an image in these circumstances. In this work we are specifically interested in images formed using MALDI, a spectroscopy technique capable of quantifying amounts of biological samples. In principle, such images should provide measurements consistent with a Poisson sampling process. MALDI imaging is a complex and relatively new technology. We would like to be able to optimise the imaging process, to both ensure that statistical behaviour is as expected and also to minimise the absolute level of measurement error. In order to support this we have developed a general technique, based upon the construction of Bland-Altman plots, which allows the statistical behaviour of an image to be quantified.

Introduction

Bland-Altman plots are constructed by plotting the observed error (residual) on data against measured signal. The distribution seen in these plots provides information regarding the functional behaviour of the error process and general quality of data. In **some cases** the dominant source of noise is a simple function of the signal, allowing for simple modifications to algorithms which provide better statistical efficiency (e.g. homoscedastic transforms). More generally, an understanding of these errors then supports an analysis of the expected performance of data processing algorithms, via analytic means (error propagation) or sample based evaluation (Monte-Carlo testing). A range of possible error models (such as proportional or Poisson), result as a consequence of real world measurement, the most common being independent Gaussian noise.

Estimating the residual on a measurement can be done in several ways, if we have ground truth then variations from this value can be computed directly. However, ground truth is often unavailable, particularly in legacy datasets. A more practical approach is to make repeat measurements and look at the difference. With a small amount of error propagation, this difference can be related directly to the measurement error, allowing the difference in repeat measurements to be used as a substitute to a residual. Quality assessment of images is often done using repeated measurements, but in some circumstances obtaining a second image may be either time consuming, costly or impractical. In these cases it would be convenient to find some way of constructing a Bland-Altman plot which does not require either ground truth or a repeat measurement.

In this work we consider several candidate approaches for the estimation of a pixel grey-level using the local neighbourhood of values. The ideal approach would make an estimate with the best accuracy and be statistically independent of the original (central) value. Also, the estimate would have understandable (predictable) statistical characteristics. Such an estimate can then be used as a substitute for a repeat measurement, and Bland-Altman plots constructed. As a final step the Bland-Altman distribution can be fitted to a parametrised error model in order to provide a quantitative summary of measurement error, suitable for evaluation and optimisation.

Key to this work is the need to make a statistically independent estimate of a grey-level value from a neighbourhood of values. As images contain significant structure at a local level, such an estimate needs to be able to appropriately deal with genuine signal. We test four purpose written methods here. First we estimate a central value using two pixels on either side, these are chosen from the direction which minimises the local image gradient (Tangential Filtering, TF). Two more methods are used based upon the same idea, but with alternative ways of selecting the sample direction. Second-order Filtering (SF) selects the sample direction which minimises the quadratic contribution in the observed spatial structure. A third technique, which we will call Fast Smoothing (FS), chooses the direction which is perpendicular to that which maximises the difference between the central pixel and a linearly interpolated value. Our final method, Polynomial Fitting (PF), is based on a generalised method for estimating the central value using a polynomial fit to the surrounding values.

Our four methods are tested on simulated images to find the statistical efficiency of central pixel prediction and the level of consistency and independence. The filtering methods are also tested on real data, by performing experiments which assess the ability to remove noise in real data. In each case, statistical predictions are made

regarding the expected performance for comparison with the measure data. We go on to use the best performing algorithm and show the quantitative estimation of an error model. As error propagation may be unfamiliar to many readers, Key mathematical relationships are derived for the simplest cases examples in Appendices.

Methods

Tangential Filtering (TF)

The smallest possible symmetrical an-isotropic filter is one which takes a simple average of two values, from either side of the pixel of interest. In tangential smoothing these two values are selected in the direction which aligns with the local image structure. The tangential vector t is computed as that perpendicular to the direction of maximum gradient, using the conventional first derivatives (though other methods could be used), and two grey-level values are quadratically interpolated at a radial distance of one pixel away from the centre (x) in this direction (figure 2). These two values are then averaged.

$$I'(x) \;=\; \frac{I(x+t) \;+\; I(x-t)}{2}$$

This new image has some useful statistical properties. As the output is simply a form of simple averaging, for an input image with independent random noise (σ), the output image has locally similar noise level which is approximately $\sqrt{2}$ smaller.

Second-order Filtering (SF)

Second-order filtering is similar to TF, except for the method used to obtain the averaging pixels. In this case the second derivative of local structure is estimated, and the direction chosen which makes the observed quadratic behaviour as small as possible. As a consequence, the pixels chosen for averaging are those that minimise any error in this process (approximately $\sigma/\sqrt{2}$).

Fast Smoothing (FS)

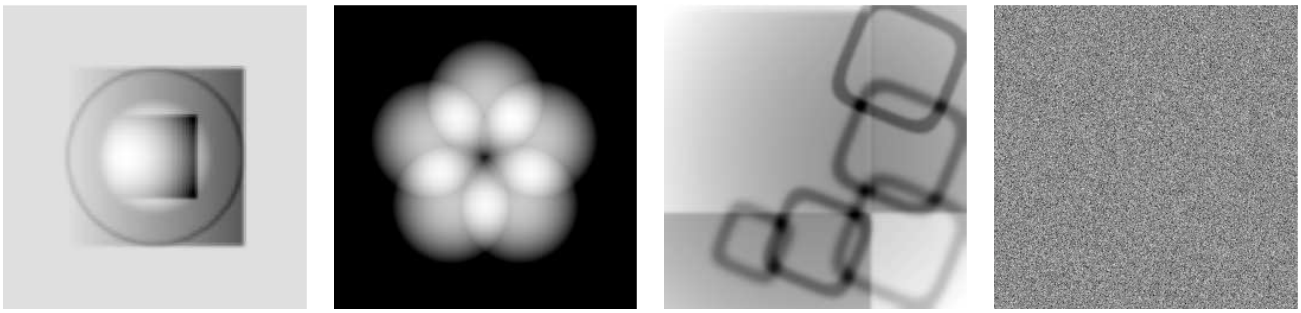
The opposing adjacent pixels are used to estimate multiple candidate averages, and the interpolated value chosen using a quadratic interpolation of the squared difference between these values and the original pixel. Unlike the other methods, the use of the central value in this way introduces the possibility of a correlation between the random noise on the original image and the filtered output. Otherwise, we might expect the error on the filtered output to be typically $\sigma/\sqrt{2}$.

Polynomial Fitting (PF)

General methods for interpolation are often based upon polynomial models. There are a multitude of possible polynomial functions (polynomial order and weighted combinations of terms) and choices for input pixel patterns. For each we can construct sets of linear equations and solve them to find the best interpolating polynomial factors. However, with a little insight we can avoid the need to derive the kernels relating to all these possibilities.

The least squares solution to a general polynomial fit of a patch of pixels always generates the same characteristic solution. The estimate of the central pixel value takes the form of a normalised convolution kernel with spatially symmetrical coefficients. We can avoid correlation between the input and output values by excluding the central pixel from this process. This allows us to find the best convolution kernel in real images by direct evaluation of prediction of the original image data, without specifying the order (or computational structure) of the assumed polynomial function. Prior to the work presented below, the real MR images were used to optimise the weighting co-efficients by minimising the squared difference between the original image and filtered output. We settled on the following choice of convolution mask $h(x)$.

$$\frac{1}{48}\begin{bmatrix} 0 & 0 & -3 & 0 & 0 \\ 0 & -5 & 20 & -5 & 0 \\ -3 & 20 & 0 & 20 & -3 \\ 0 & -5 & 20 & -5 & 0 \\ 0 & 0 & -3 & 0 & 0 \end{bmatrix}$$



(a) (b) (c) (d)

Figure 1: Ground truth and noise images

Error propagation can be used to estimate the variance on the filtered output as $\sum_x h(x)^2\sigma^2 = \sigma^2/1.33$.

Evaluation of Filtering Methods

As each filter makes minimal use of the central value, the result has noise which might be expected to be independent of the original. Therefore, it is legitimate to average the output image with the input to obtain a noise filter.

$$I' = \frac{\alpha f(I) + I}{1+\alpha} \tag{1}$$

Where I ,$f(I)$ and I' represent an original image, a filtered image and a weighted-filtered image, respectively, and α is a weighting factor of the filtered image (varying here from 1.0 to 3.0).

Appendices A-C explain how the linear combination of estimates affects the noise in an output image composed as a weighted average of the filtered and unfiltered images. Based upon this theory, for any image filter there should exist an optimal α which gives the best linear combination of input and output images with minimum noise. The position and value of this minimum tell us two things; if the minimum is consistent with combination of two independent measurements then it should lie along a predictable curve ($\sigma/\sqrt{1+\alpha}$, see appendix C). If the minimum is not consistent with the theory then the filtered image ($f(I)$) can not be independent of the input.

The optimal value of α can be interpreted in several ways, not only is it the best value for a linear combination (for independent image data), it is also the ratio of the noise variances between $f(I)$ and I, or equivalently the ratio of the information. The larger the value of α the more efficient the image filters use of input data and the better its noise properties. For example a value of $\alpha = 2$ implies that the output image has noise which is $\sigma/\sqrt{2}$, with each pixel in $f(I)$ being equivalent to the information contained in two pixels in I.

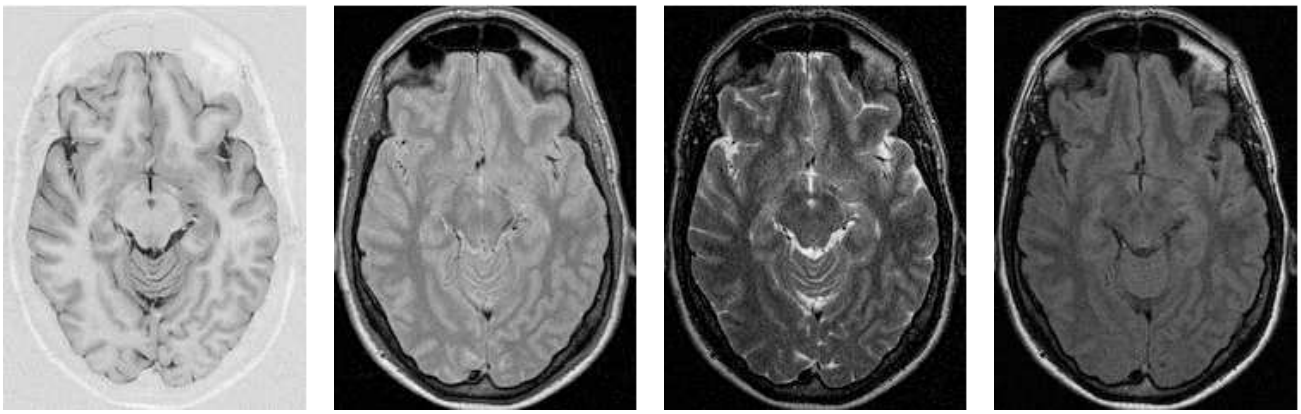
We seek a method which consistently filters images with the same (minimal) level of output noise i.e. the largest value of α which is consistent with being independent of the input data I. Ideally, the observed values should also be consistent with predictions made based upon the design of the method. Such a filter will give the most reliable starting point for constructing a Bland-Altman plot.

For each of our four filters, an image of varied linear combinations of the filtered and the original images were constructed. The resulting graph of image noise against weight factor was used to evaluate filtering efficiency and the level of correlation (in noise) between the original (I) and filtered values ($f(I)$).

The initial experiments used simulated images and the real-world (MR) images to test algorithm performance. Simulated images were constructed which contained a variety of different image structures (Figure 1(a), (b) and (c)), with various levels of local spatial curvature and image smoothness, with the intention of testing the degree of consistency of any results. A known random noise distribution ($\sigma = 5$) was added to each image shown in Figure 1(d). In addition, the image of random noise only was also used in a fourth test.

Knowledge of ground truth pixel values, before adding random noise, was used to determine the amount of noise in I', (computed using a root mean squared difference).

Our real image data comprised 4 real MR images of a human brain. These were chosen simply as four instances of real noise distributions. As we do not know the ground truth for these images, the noise in original and filtered images was obtained by measuring the width of the peak at zero in a histogram of spatial second derivatives. This technique is one we find to be a reliable method for estimating homogenous image noise in arbitrary images. However, the technique does assume that the input images have spatially independent noise, and some of our filters



(a) (b) (c) (d)

Figure 2: Four MR images of brain sections

can introduce mild spatial correlation. Therefore, although the best weighting factor can be estimated the absolute value of estimated noise may have a slight bias.

Construction of a Parametric Noise Model

Characteristics of the noise in an image can be observed using a Bland-Altman plot. Here we will use the estimate obtained from our filter as a surrogate ground truth. The intention then is to fit the data density distribution seen in the BA plot to obtain a polynomial description of the dependency of the noise on the signal $\sigma = g(I)$. However, this approach is complicated by the fact that the estimated residual $I - f(I)$ is correlated with both I and $f(I)$. This modifies the observed density distribution, introducing a rotation of the point noise processes.

We have no choice but to use the residual estimate $I - f(I)$, but we are free to modify our estimate of signal. What is required is the definition of the variable for the signal axis which is uncorrelated with the residual. Appendix D shows that the same linear combination of the input and filtered image (which minimises the error on a combined estimate as described above), will also decorrelate the noise. Allowing us to fit a symmetrical density model to the data.

In order to test the model fitting routine to look at Poisson distributed noise, Poisson-like signal-to-noise has to be carefully simulated in a specific way. That is a simulated image was generated by multiplying a known uniform random noise with the square root of a simulated ground truth (ideally pure signal) image, G, and subsequently added to the original image. This means, for each data point i on a simulated image of N data points, the signal value G_i determines the error on the data point by the following relationship.

$$\epsilon_i = \sigma\sqrt{G_i}$$

See Figure7, where the uniform random noise $\sigma = 0.5$ was fed in as a test.

A power law model is selected as a convenient way of characterising the Poisson distribution and similar behaviours. We attempted to find the best parameters that would fit an error model of the form[1]

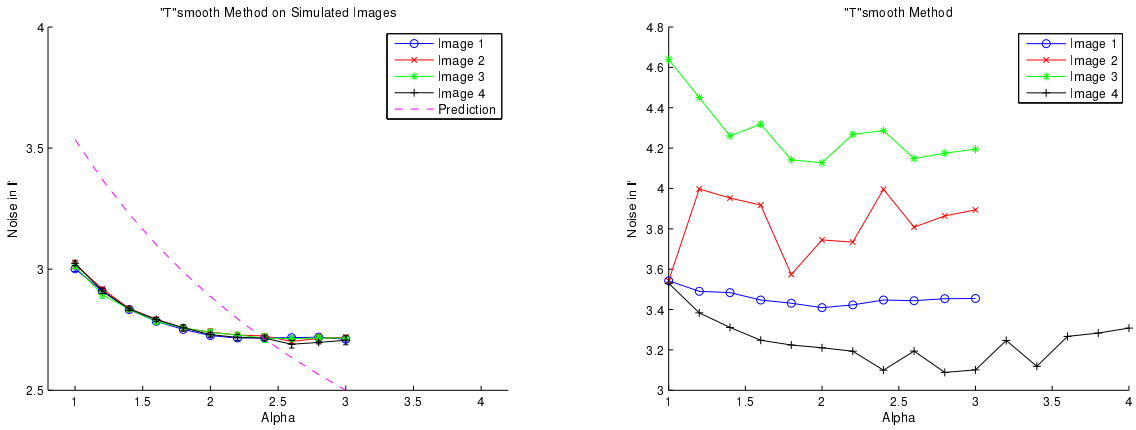
$$\sigma_i = a_1 \left(\frac{I_i'}{a_1}\right)^{\frac{0.5}{a_2}}$$

Given that I_i' represents a weighted-filtered image as discussed in the previous section. a_1 and a_2 are variable parameters that were adjusted such that the calculated Likelihood function,

$$\mathscr{L} = \Pi_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} exp\left(-\frac{1}{2}\left(\frac{\Delta_i}{\sigma_i(a_1, a_2)}\right)^2\right)$$

was maximised. Where $\Delta_i = \frac{I_i - f(I_i)}{1.18}$, with 1.18 being the coefficient of the expected standard deviation on the residual image, $\sqrt{var(I_i - f(I_i))}$. Appendix D explains the way we estimate the error on a filtered image as a function of the error on the original image. Since the curve of error for the tangential filtering image varying α

[1]This particular form has approximately homoscedastic variables, making it easier to compute meaningful parameter estimation errors.



(a) (b)

Figure 3: *Plot of estimated noise in the TF weighted images against weighting factors. Where the original data sets are (a) simulated images 1-4 (b) MR images 1-4*

intersects with the predicted line at $\alpha = 2.5$, $\sqrt{var(I_i - f(I_i))}$ takes a value of $1.18\sigma_i$. Therefore, Δ_i represents the expected Pull distribution of the noise (the residual written in term of unit standard deviations). The Downhill Simplex method was used to optimise the parameters in 2 dimensions with respect to a cost function determined by $-log\mathscr{L}$ (to be minimised).

$$-log\mathscr{L} \;=\; \sum_{i=1}^{N}\left[\frac{1}{2}\left(\frac{\Delta_i}{\sigma_i(a_1,a_2)}\right)^2 + log(\sigma_i)\right] + constant$$

Results

Evaluation of Image Filters

Results for the simulated data are shown in figures 3-6. Illustrating a range of effectiveness and consistency.

Results for real MR data, (although less accurate) show minima which are consistent with the simulated results.

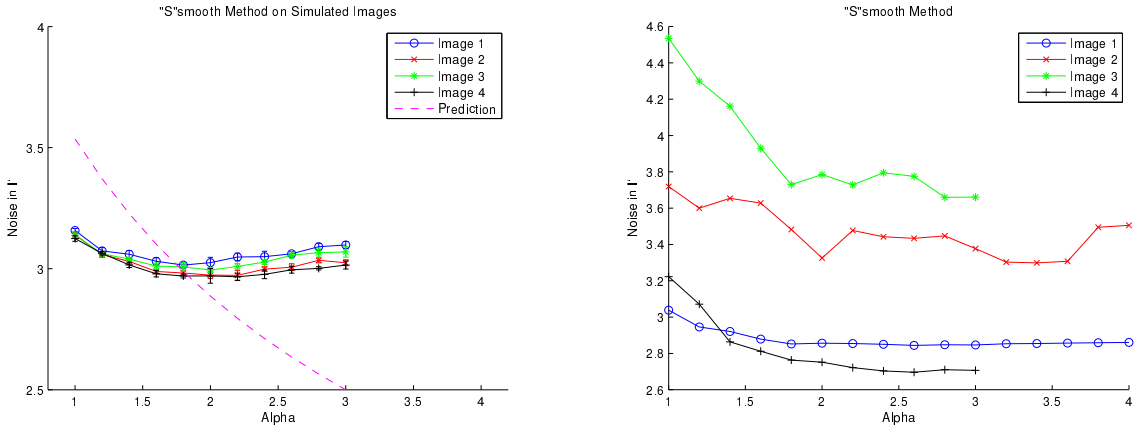
The level of noise reduction for Tangential Filtering approaches is slightly better than that predicted. We believe this to be due to the use of quadratic interpolation for the estimation of the two end samples, which has a slight smoothing effect. Otherwise the result is what we might expect to be close to the limit of any method which does not assume a specific structural model of image content. The results obtained using polynomial fitting (PF) are less consistent than we might have hoped, but the statistical efficiency for three of the tests ($\alpha = 1.33$) is as we would predict from error propagation for this convolution kernel. The FS algorithm has results which are inconsistent with an assumption of grey-level independence, and otherwise too poor to be of practical value.

Computing a Parametric Error Model

The first (vertical) dimension of our Bland-Altman (BA) plot estimates the noise in an image by subtracting the tangential filtered image from the original image. A variance on this axis is about 1.18 times the variance in the original image. The second (horizontal) dimension of the plot is a selected linear combination of the raw image data and the smoothed data as a result of tangential filtering (where $\alpha = 2.5$) which was experimentally agreed with the theoretical prediction to introduce least correlation to the estimated noise values. The model for characterising Poisson distribution was modified slightly by inputing the values of residual image and use the factor of 1.18 to compensate for the variance in $I - f(I)$ as follows.

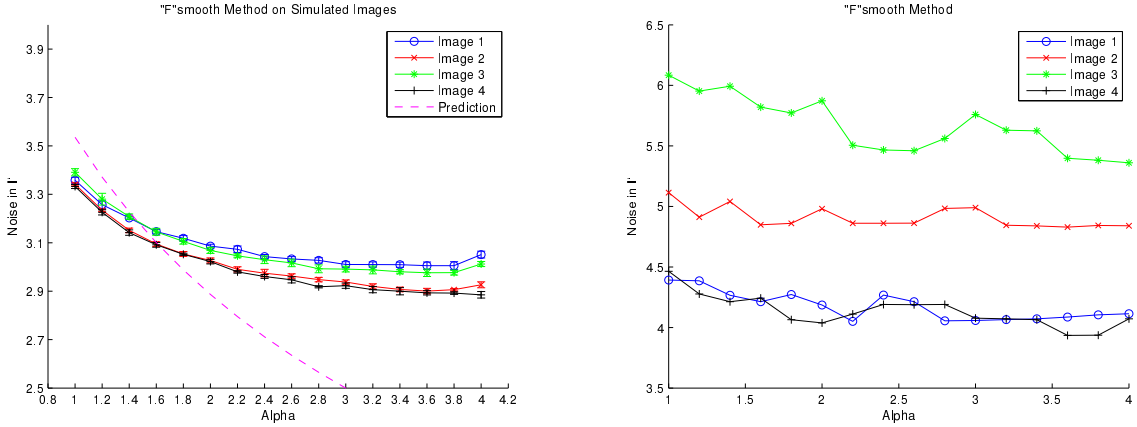
$$\epsilon \;=\; a_1\left(\frac{(I - f(I))}{(1.18)a_1}\right)^{\frac{0.5}{a_2}}$$

Where the power law applies as expected, the parameter a_2 is close to the value of 1. The parameter a_1 takes values of the variance, σ^2, on the original image I within 2 standard deviations.



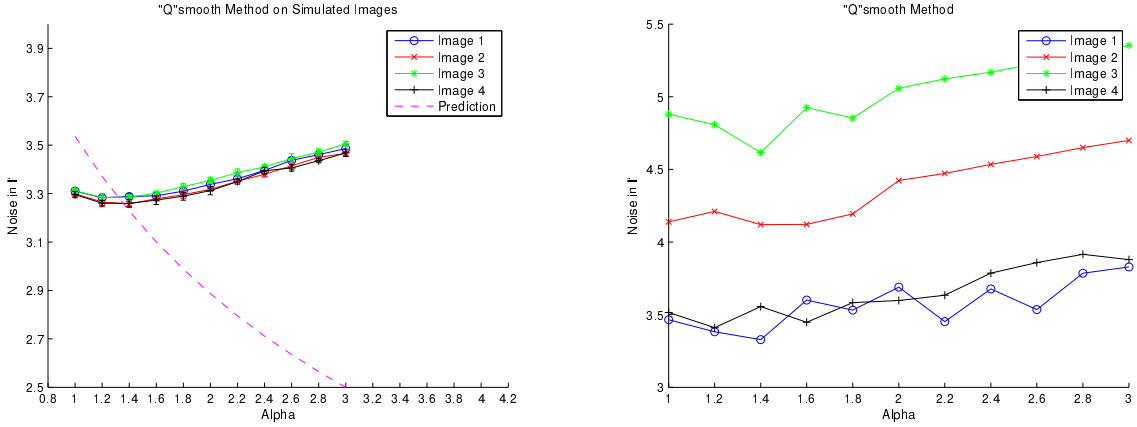
(a) (b)

Figure 4: *Plot of estimated noise in the SF weighted images against weighting factors. Where the original data sets are (a) simulated images 1-4 (b) MR images 1-4*



(a) (b)

Figure 5: *Plot of estimated noise in the FS weighted images against weighting factors. Where the original data sets are (a) simulated images 1-4 (b) MR images 1-4*



(a) (b)

Figure 6: *Plot of estimated noise in the PF weighted images against weighting factors. Where the original data sets are (a) simulated images 1-4 (b) MR images 1-4*

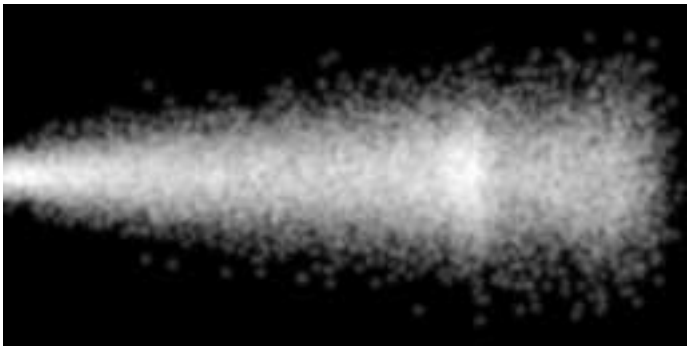


Figure 7: *Bland-Altman plot of a simulated image carrying Poisson noise characteristics*

The BA plot of a simulated image with Poisson noise demonstrates the property that the estimated noise (y-axis) is proportional to square root of the signal parameter (x-axis) (see Figure 7). Fits of this distribution generate the expected parameter values within errors.

Conclusions

The filter which gives an independent estimate of the input image data with the least noise is the Tangential Filter (TF) (with $\alpha = 2.5$). Second-order filtering (SF) was also independent and consistent, but with less statistical efficiency ($\alpha = 1.8$). Both of these results were broadly consistent with theoretical predictions based upon the design of the filters. The other two filters were either not independent or inconsistent, and neither were very efficient estimators. Our Bland-Altman (BA) plots were designed to illustrate the characteristics of noise associated to the signal on arbitrary images. With the use of TF at the appropriate α, the data set on a BA plot can be fit onto a model of Poisson-like distribution that would return a value for the variance on the image data. And hence a method of estimating noise models for an image. It is now our intention to use this quantitative assessment of Bland-Altman plots as a way to evaluate, and then optimise, MALDI image generation.

Appendix

Appendix A : Error propagation of a linear function

For any $x \pm \sigma_x$ and $y \pm \sigma_y$, the error propagates onto $z = x + y$ is

$$\sigma_z(x,y) \;=\; \sqrt{\left(\frac{\partial z(x,y)}{\partial x}\sigma_x\right)^2 \;+\; \left(\frac{\partial z(x,y)}{\partial y}\sigma_y\right)^2} \;=\; \sqrt{\sigma_x^2 \;+\; \sigma_y^2} \tag{2}$$

Appendix B : Optimisation of Linear Combination between 2 Parameters

If z is a linear combination of some parameters $x \pm \sigma_x$ and $y \pm \sigma_y$, it can be written as

$$z \;=\; \frac{\alpha x \;+\; \beta y}{\alpha \;+\; \beta}$$

Using equation(2) , the error on z in Equation (1) is therefore

$$\sigma_z \;=\; \sqrt{\left(\frac{\alpha}{\alpha \;+\; \beta}\sigma_x\right)^2 \;+\; \left(\frac{\beta}{\alpha \;+\; \beta}\sigma_y\right)^2} \tag{3}$$

For simplicity, re-write the weighted factors on x and y in Equation (3) as α' ($= \frac{\alpha}{\alpha + \beta}$) and β' ($= \frac{\beta}{\alpha + \beta}$), respectively. These two terms are summed up to 1. Substitute these into Equation (4) in term of α' ($\beta' = 1 - \alpha'$). Then,

$$\sigma_z \;=\; \sqrt{(\alpha'\sigma_x)^2 \;+\; ((1-\alpha')\sigma_y)^2} \tag{4}$$

Differentiate σ_z with respect to α' to work out a value for α' which gives minimum σ_z corresponding to levels of σ_x and σ_y.

$$\alpha' \ = \ \frac{\sigma_y^2}{\sigma_x^2 \ + \ \sigma_y^2} \tag{5}$$

and

$$\sigma_z^2 \ = \ \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 \ + \ \sigma_y^2} \tag{6}$$

Appendix C : Predicted Values for Noise in a Weighted-filtered Image

We combine the original and filtered images as in Equation (1). Where σ_x, σ_y and σ_z represent errors (noise levels) in the filtered image, original image and the final (weighted-filtered) image, respectively. Recall $\alpha' = \frac{\alpha}{\alpha \ + \ \beta}$) where $\beta = 1$, and in order to achieve least σ_z, it must follow the expression from Appendix B (Equation (6)) that is

$$\alpha' \ = \ \frac{\alpha}{\alpha + 1} \ = \ \frac{\sigma_y^2}{\sigma_x^2 \ + \ \sigma_y^2} \tag{7}$$

Rearranging,

$$\alpha \ = \ \frac{\sigma_y^2}{\sigma_x^2} \tag{8}$$

Substitute Equation (6) and (7) into the solution for σ_z

$$\sigma_z^2 \ = \ \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 \ + \ \sigma_y^2} \ = \ \alpha' \frac{\sigma_y^2}{\alpha} \ = \ \frac{\sigma_y^2}{\alpha + 1}$$

Finally,

$$\sigma_z \ = \ \frac{1}{\sqrt{\alpha + 1}} \sigma_y \tag{9}$$

i.e. the error (noise) on the processed image should reduce by a factor of $\sqrt{\alpha + 1}$ from the error (noise) contained in the original image that gives the corresponding minimal σ_z for all α.

Appendix D : Minimisation of Correlation in a Bland-Altman Plot

In a Bland-Altman plot, the vertical axis will be our estimated noise evaluated from the difference between the original and the filtered images $(I - f(I))$, carrying $(\sigma_y, -\sigma_x)$ errors.

This is plotted against the horizontal axis which is in the form of linear scale of the original and the filtered images $(\eta I + (1 - \eta) f(I))$, carrying $(\eta \sigma_y, (1 - \eta)\sigma_x)$ errors.

Orthogonality of the errors on both axes imples the zero dot product,

$$(\sigma_y, -\sigma_x) \cdot (\eta \sigma_y, (1 - \eta)\sigma_x) \ = \ 0,$$

$$\eta \sigma_y^2 \ - \ (1 - \eta)\sigma_x^2 \ = \ 0,$$

$$\frac{\eta}{1 - \eta} \ = \ \frac{\sigma_y^2}{\sigma_x^2} \tag{10}$$

Recall $\alpha' \ = \ \frac{\alpha}{\alpha + 1}$ and rearrange,

$$\alpha \ = \ \frac{\alpha'}{1 - \alpha'} \tag{11}$$

Then, Equation (8) = (10) = (11)

$$\alpha \ = \ \frac{\eta}{1 - \eta} \ = \ \frac{\alpha'}{1 - \alpha'} \tag{12}$$

i.e.

$$\eta \ = \ \alpha',$$

Therefore, this ideal linear scale which should minimise the correlation in the Bland-Altman plot agrees with the predicted values from Appendix C where its errors are minimised. Further, given a value of α using equation (7), we can estimate the error on f(I) as $\sigma_f(I)^2 = \frac{\sigma_I^2}{\alpha}$.