

# Estimation of expression levels in spotted microarrays with saturated pixels

C.A. Glasbey

Biomathematics and Statistics Scotland  
King's Buildings, Edinburgh EH9 3JZ, Scotland

T. Forster & P. Ghazal

Division of Pathway Medicine  
University of Edinburgh Medical School  
Chancellor's Building, 49 Little France Crescent  
Edinburgh EH16 4SB, Scotland

## Abstract

Digital images obtained by the laser scanning of spotted microarrays often include saturated pixel values. These arise when the scan settings are sufficiently high that some pixels exceed the limit  $L = 65535$  and are instead set to  $L$ . Failure to adjust for this censoring leads to biased estimates of gene expression levels. To impute censored values, we propose a linear model based on the principal components of uncensored spots on the same array. This is computationally fast, flexible to adapt to distinctive spot shapes and profiles on different arrays, and is shown to be more effective than the polynomial-hyperbolic model in correcting for the bias. The application to biological data demonstrates the potential for enhancing the dynamic range of detection. Fortran90 subroutines implementing these methods are available at <http://www.bioss.ac.uk/~chris>

**Key words:** censored pixel values, eigenshapes, principal components, regularisation, shape model.

## 1 Introduction

A commonly-used microarray technology involves printing genetic material using either contact- or non-contact transfer methodology. Typically, DNA corresponding to specific genes are placed as spots on a glass slide, which is then hybridised with either one or two samples, labelled with fluorescent dyes, and the microarray is laser scanned at the corresponding wavelengths of the fluors, to produce a digital image. The first stage in the analysis of the data is the use of image analysis to summarise the information in each spot, to estimate the gene expression level in each

sample and the ratio of expressions between pairs of samples (see, for example, Glasbey and Ghazal, 2003). Computer packages such as *QuantArray* (GSI Luminomics, 1999), *ScanAlyze* (Eisen, 1999), *GenePix* (Axon Instruments, Inc, 1999) and *UCSF Spot* (Jain et al., 2002), have implemented a range of methods. However, none of these packages satisfactorily address the problem of saturated pixel values, which arise when the laser scan settings are sufficiently high that some pixels should exceed the two-byte limit  $L = 2^{16} - 1 = 65535$ , but are instead assigned this value. Ignoring this problem of censoring leads to biased estimates of expression levels.

Wit and McClure (2003) developed a method for correcting this bias, using only the summary statistics of mean, variance and median for each spot. However, they acknowledge that adjustments made at a pixel level would be far more effective. If interest is solely in estimating the ratio of expressions between pairs of samples in dual-dye experiments, then Dodd et al. (2004) have developed an elegant approach that models the correlation between channels. To estimate expression levels for individual samples, Ekstrom et al. (2004) proposed a regression approach for imputing censored values, by fitting a parametric model to the uncensored pixels associated with a spot. Several models have been suggested for what they term ‘spot shape’ but which may better be described as ‘spot profile’ because others have used ‘spot shape’ to refer to the contour of the spot, i.e. the set of pixels, identified by image segmentation, associated with the spot foreground. Models for spot profile include a scaled bivariate Gaussian density function (Steinfath et al., 2001), a difference of two Gaussian densities and a cylinder (Wierling et al., 2002), though Ekstrom et al. (2004) found a polynomial-hyperbolic model to perform best for one array. We note that in all these models, the set of pixels in the spot foreground have the shape of a circular disc, which may not always be appropriate. Further, the best choice of model is likely to vary from array to array, and to depend on many aspects of microarray manufacture. Each array with censored spots will also have many uncensored spots, which can then be used as the basis for a model. So, an alternative approach, developed in other domains of image analysis, is the use of principal components or eigenvectors computed from these uncensored spots (Turk and Pentland, 1991; Cootes et al., 1995). These are sufficiently flexible to model both non-circular spot shapes and profiles that do not conform to parametric models.

In §2 we present the methodology for using principal components to impute censored pixel values, using data from a dye-swap experiment, designed to examine the effects of ingestion of apoptotic cells on murine macrophage gene expression for illustration. In §3 we evaluate the method using a real biological experiment. Finally, in §4 we discuss the results, biological significance and extensions.

## 2 Principal components model

Let  $y^{(j)}$  denote the vector of pixel values associated with spot  $j$  on a digital image of one channel on a single- and double-channel microarray. In this section we will treat each channel separately, so the distinction between single- or double-channels is irrelevant. For simplicity, we consider an  $N \times N$  square centred on each spot, where  $N$  is slightly larger than the diameter of any of the spots, so some background pixels are also included, and  $y^{(j)}$  is of length  $K$  ( $= N^2$ ) with pixels in raster-scan order. The locations of the squares can be identified using the semi-

automatic gridding algorithms available in almost all microarray computer packages. Also for notational simplicity, we measure the intensity of a spot, corrected for background intensity, as the mean of the pixels in a square minus the mean of the  $M$  smallest pixel values:

$$\mu^{(j)} = \frac{1}{K} \sum_{k=1}^K y_k^{(j)} - \frac{1}{M} \sum_{k=1}^M y_{(k)}^{(j)}, \quad (1)$$

where  $y_k^{(j)}$  denotes the  $k$ th element in  $y^{(j)}$  and  $y_{(k)}^{(j)}$  denotes the  $k$ th smallest element. We would expect very similar results in what follows even if more sophisticated methods had been used to estimate  $\mu^{(j)}$ .

If there are  $J$  spots on an array which are free of censored pixels, then we estimate the mean spot shape/profile,  $u$ , by

$$u = \frac{1}{J} \sum_{j=1}^J y^{(j)},$$

where  $u$  is also a vector of length  $K$ . In order to model the variability in spot shape/profile, we first compute the variance matrix of  $y$ :

$$V = \frac{1}{J} \sum_{j=1}^J (y^{(j)} - u) (y^{(j)} - u)^T,$$

where  $V$  is of size  $K \times K$ . We then obtain the eigenvectors and eigenvalues of  $V$ , using a standard algorithm (see, for example, Jolliffe, 2002). Let  $e_i$  denote the  $i$ th eigenvector and  $\lambda_i$  the corresponding eigenvalue, for  $i = 1, \dots, K$ . The  $i$ th principal component (PC), denoted by the  $K$ -vector  $x_i$ , is given by

$$x_i = \sqrt{\lambda_i} e_i.$$

To illustrate, Fig 1 shows the first 4 PCs for channel 1, array 1, of the murine macrophage experiment. The array consisted of 9248 spots, and we extracted a  $17 \times 17$  square of pixels for each spot using QuantArray software (GSI Luminomics, 1999). PC1 is very similar to  $u$  (not shown), and models the overall intensity level of a spot. PC2 when added to  $u$  adjusts the diameter of a spot, whereas PC3 has the effect of shifting the spot down the array and PC4 across the array. Some shifting is necessary, because gridding does not align spots perfectly, and even if alignment was to the nearest pixel, there would still be some need for subpixel adjustments.

The  $J$  uncensored spots can each be re-expressed as a linear combination of all the  $K$  PCs:

$$y^{(j)} = u + \sum_{i=1}^K x_i \beta_i^{(j)} = u + X \beta^{(j)}, \quad (2)$$

where  $X$  denotes the  $K \times K$  matrix with  $x$ 's as columns, and the  $K$ -vector of parameters,  $\beta^{(j)}$ , can be obtained either directly from (2) by premultiplying by  $X^{-1}$ , because  $X$  is of full rank, or by least squares, for consistency with the censored case:

$$\begin{aligned} \beta^{(j)} &= \arg \min_{\beta} [(y^{(j)} - u - X\beta)^T (y^{(j)} - u - X\beta)] \\ &= (X^T X)^{-1} X^T (y^{(j)} - u). \end{aligned} \quad (3)$$

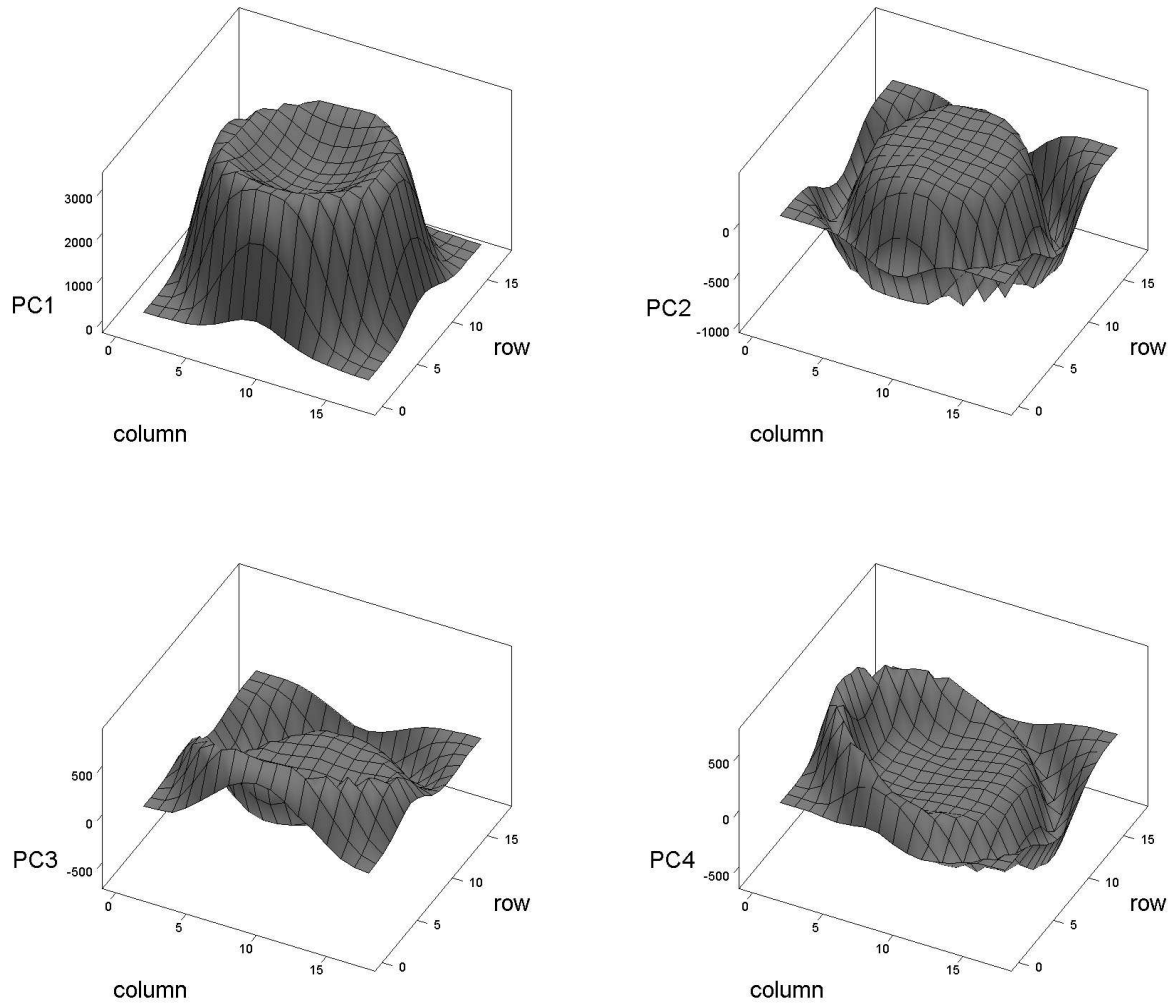


Figure 1: First 4 principal components for channel 1 in array 1 of murine macrophage experiment, which explain 90.4%, 3.2%, 1.7% and 1.0% of variability.

Averaged over the  $J$  spots used to compute the PCs,

$$\sum_{j=1}^J \beta^{(j)} (\beta^{(j)})^T = I, \quad (4)$$

where  $I$  denotes the  $K \times K$  identity matrix. This follows because the  $\beta$ 's are the right singular vectors of the data matrix (see, for example, Mardia et al., 1979).

Spots with censored values can be similarly modelled, but pixels with value  $L = 65535$  are excluded from the estimation of  $\beta$ . Let  $z^{(j)}$  denote the vector of length  $K' < K$  obtained from  $y^{(j)}$  by omitting all censored pixels. Least squares can no longer be used, because (3) does not have a unique solution, so we regularise by penalising large values of  $\beta$ , motivated by (4), by adding a penalty term  $\lambda \beta^T \beta$  to the least squares criterion, as follows:

$$\hat{\beta}^{(j)} = \arg \min_{\beta} \left[ (z^{(j)} - u^{(j)} - X^{(j)} \beta)^T (z^{(j)} - u^{(j)} - X^{(j)} \beta) + \lambda \beta^T \beta \right] \quad (5)$$

$$= ((X^{(j)})^T X^{(j)} + \lambda I)^{-1} (X^{(j)})^T (z^{(j)} - u^{(j)}), \quad (6)$$

where  $u^{(j)}$  denotes a  $K'$ -vector, obtained from  $u$  by omitting elements where the corresponding pixel in  $y^{(j)}$  is censored, and similarly,  $X^{(j)}$  is obtained from  $X$  by omitting rows, and is a  $K' \times K$  matrix. We then impute the correct values of those pixels affected by censoring, using:

$$\hat{y}_k^{(j)} = \max(L, (u + X \hat{\beta}^{(j)})_k) \quad \text{if } y_k^{(j)} = L. \quad (7)$$

Finally, we estimate  $\mu$ , denoted  $\hat{\mu}$ , by applying (1), but with censored pixels replaced by estimated values.

The penalty term in (5) was first proposed by Hoerl and Kennard (1970) for ridge regression. Alternatively, the term can be viewed as a Bayesian log-prior probability (Lindley and Smith, 1972). Parameter  $\lambda \geq 0$  controls the extent to which values of  $\beta$  are shrunk towards zero. As  $\lambda \rightarrow \infty$ ,  $\beta \rightarrow 0$  and  $y \rightarrow u$ , the mean spot shape/profile. We note that, because the model is linear in its parameters,  $\hat{\beta}^{(j)}$  is expressible analytically in (6), unlike for the nonlinear models considered by Ekstrom et al. (2004), which require a more computationally-intensive iterative optimisation. If errors are both independently distributed and homoscedastic, then least squares is equivalent to maximum likelihood estimation, and is fully efficient. In many imaging applications, these two assumptions are likely to be violated. However, least squares usually remains a relatively efficient method of estimation, and in the linear case is very fast to compute.

To choose the best value for  $\lambda$  we use a form of cross-validation. If  $C$  spots on an array are affected by censoring, then we artificially censor a further  $C$  spots (or 100 spots if  $C < 100$ ) by reducing the saturation threshold ( $L$ ) by a suitable amount, say to  $L'$ . For these additional spots now considered to be affected by censoring, we apply (6) and (7), but with  $L$  replaced by  $L'$ , and obtain  $\hat{\mu}$ . However, for these  $C$  spots, we also know the true values of  $\mu$ , i.e. the values computed in the absence of censoring at  $L'$ . So we can assess how well the imputation of censored pixels has performed, for example by computing the proportional root-mean-square error between estimated and true values of  $\mu$ , scaled by the true values:

$$\text{rmse}\% = \sqrt{\frac{1}{C} \sum_{j=1}^C \left( \frac{\hat{\mu}^{(j)} - \mu^{(j)}}{\mu^{(j)}} \right)^2} \times 100\%. \quad (8)$$

censored % =	0.1	0.3	1	3	10
bias %					
PC estimator	-0.3	-0.6	-0.7	-2.1	-5.7
parametric estimator	-1.0	-2.9	-2.9	-7.4	-11.2
uncorrected	-1.5	-6.6	-13.5	-17.2	-21.2
rmse %					
PC estimator	0.8	3.3	7.9	10.3	15.5
parametric estimator	2.3	5.9	8.9	14.0	20.3
uncorrected	3.3	9.7	20.5	26.4	31.5

Table 1: Proportional biases and root-mean-square errors, averaged over 2 channels on each of 2 arrays of murine macrophage experiment, from using the PCs model to estimate  $\mu$ , and also from using the polynomial-hyperbolic model and saturated pixel values without adjustment, for a range of percentages of spots including censored pixels.

By repeating the computation of (8) for a range of values of  $\lambda$ , we see which value minimises  $\text{rmse}\%$ , and then use this value to obtain  $\hat{\mu}$  for the  $C$  spots that are truly affected by censoring.

We applied the method to the murine macrophage experiment, using the  $M = 50$  smallest pixel values in each  $17 \times 17$  square to estimate background intensity. We created censoring artificially, by rescaling the pixel values so that a target number of spots included censored pixels. We could alternatively have based our ‘true’ spot means on output from a laser scan at a lower scan setting that avoided censored pixels, but this would have introduced between-scan variability into the assessment. To speed-up the computations we used only the first 100 of the 289 PCs in  $X$ , but this has a negligible effect on the results because these account for 99.90% of the variability, and so later columns in  $X$  are very small. We also tried using fewer PCs and omitting the penalty term, but the results were always poorer. For illustration, Fig 2 shows the results from channel 1 in array 1, with 1% of spots including some censored pixels. As we would expect,  $\mu$  is underestimated if no account is taken of saturation effects. The PCs model is successful in correcting for this bias, but the larger values show substantial variability.

Table 1 summarises results for 2 channels on 2 arrays, for several degrees of censoring. On average, 24% of the pixels in the  $17 \times 17$  squares associated with spots were censored. The cross-validation choices of  $\lambda$ , which ranged from  $10^2$  to  $10^4$ , were all close to the optimal values, which we know in this application because censoring was artificial. Summaries given are the proportional bias:

$$\text{bias}\% = \frac{1}{C} \sum_{j=1}^C \left( \frac{\hat{\mu}^{(j)} - \mu^{(j)}}{\mu^{(j)}} \right) \times 100\%. \quad (9)$$

and the  $\text{rmse}\%$  (8). For comparison we also fitted the parametric polynomial-hyperbolic model proposed by Ekstrom et al. (2004). For compatibility with the PC model, we did not consider power transformations, as they had done, and for computational speed we excluded saturated pixels from the fitting procedure. We standardised on values of  $a_s = 2.5$  and  $\gamma_s = 2$ , as these gave the best fit to the mean spot shapes/profiles ( $u$ ) for each array. We see that the principal components model corrects for more of the  $\text{bias}\%$  than the parametric model, and also  $\text{rmse}\%$ ’s are least.

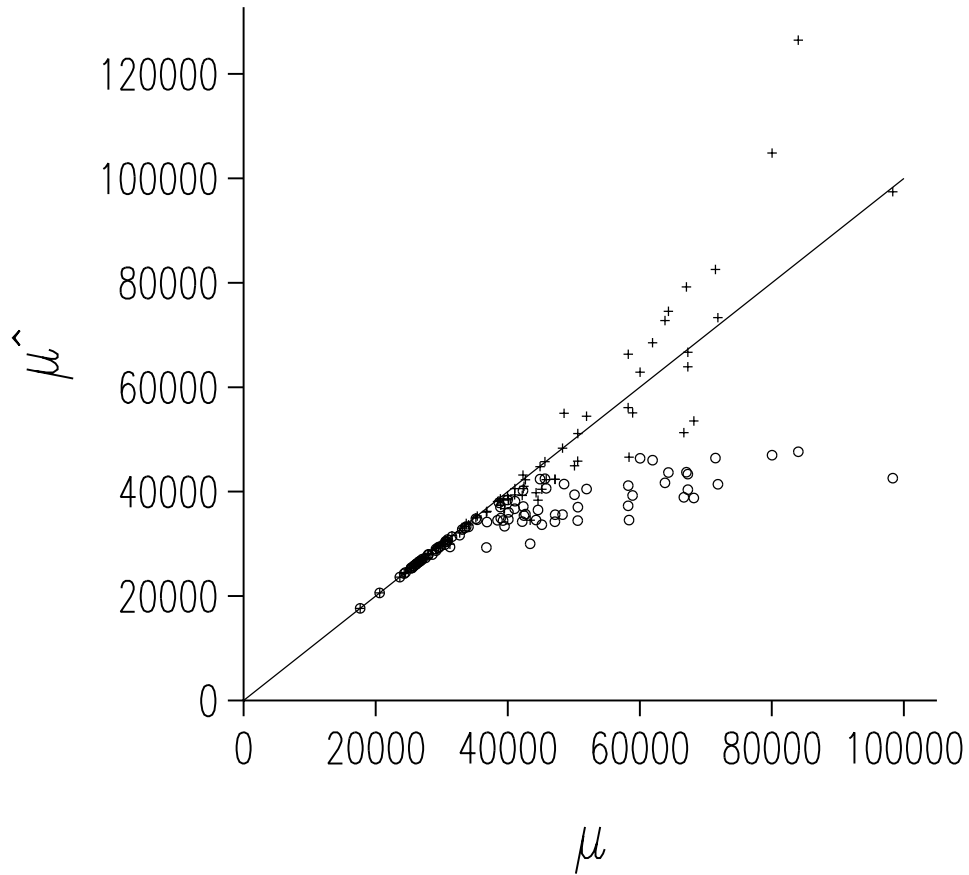


Figure 2: Estimated means ( $\hat{\mu}$ ) of spots with censored pixels for channel 1 in array 1 of murine macrophage experiment, plotted against true means ( $\mu$ ), when 1% of spots included some censored pixels:  $\circ$  uses saturated pixel values without adjustment,  $+$  uses model predictions with  $\lambda = 10^4$ , and  $—$  shows the 1:1 line of perfect agreement.

% pixels censored	bias %	rmse %
0–10	−0.7	2.7
10–20	−1.1	2.4
20–30	−2.3	5.0
30–40	−3.6	7.2
40–50	−6.8	14.2
50–60	−9.8	21.4
60–70	−12.7	29.6
70–80	−8.1	31.6

Table 2: Proportional biases and root-mean-square errors, from using the PCs model to estimate  $\mu$ , for different percentages of censored pixels in  $17 \times 17$  squares.

Table 2 shows the results of the PCs model again, but now presented in terms of percentages of censored pixels in the  $17 \times 17$  squares associated with spots. We see that the quality of the imputation deteriorates as censoring increases, rising to a root-mean-square error of 30% when more than 60% of pixels are censored.

### 3 Evaluation on a real biological experiment

A laboratory experiment was conducted to assess the viability of PC correction in a real biological experiment. Data for such a validation is not readily available in public microarray databases, in that studies are often insufficiently annotated, stored only with numerical data or only one scan image is provided. Instead, we selected a suitable in-house array platform for which we were able to prospectively perform the necessary scans.

The selected platform is a microarray in form of a silicon nitride ‘wafer’ (so named because it is not based on a glass microscopy slide) coated with amino-silane, printed and quality-controlled at DPM (Terry et al., 2006). The wafer consists of 15 unique probes representing murine cytomegalovirus genes and murine cellular genes. This particular wafer was chosen on the basis of availability and replication level, i.e. the physical wafer was ‘fresh’ and available for the necessary laser scanning steps, and each gene on the wafer was replicated a total of 75 times at various locations on the wafer, allowing the use of more precise estimates for fluorescence measurements. There are a total of 1800 spots on the array, of which  $15 \times 75 = 1125$  are genes, the remainder are spots consisting of only buffer solution. The wafer was hybridised with mouse tissue total RNA, which was known to have high, medium and low expression for 5 genes each on the wafer.

The wafer was scanned using a GSI Lumonics ScanArray 5000 confocal laser scanner, with photo-multiplier-tube (PMT) settings covering 100%, 90%, 80%, 70% and 60%, with the laser power fixed at 100%. The resulting TIFF images then underwent image processing using QuantArray software, resulting in pixel fluorescence summarised at spot level. From this range of data sets, we selected the highest scan for which there was no apparent spot saturation, i.e. no

printed spot had a median pixel fluorescence of 65535. This condition is met by a PMT setting of 70%. In addition to this, the scanned TIFF images of the 3 higher scans were corrected using the PC algorithm as described above, and subsequently re-processed with identical image processing parameters to those used in the uncorrected images.

The wafer was first checked for spatial variation in signal and background fluorescence measurements, such as misprinted regions on the wafer or bleeding of signal from one gene probe across others. Basic descriptive statistics were then computed for each of the 15 unique gene probes across 75 replicate instances on the wafer, separately for each laser scan setting. Of the 15 probes, only one was found to match the requirements for the PCA model validation in being consistently saturated in the highest scan. The gene referred to as ‘cMU012’ is saturated in 72 out of 75 instances in the highest (PMT=100%) scan, 67 out of 75 instances in the 90% scan and 25 out of 75 instances in the 80% scan. It is not saturated in any instances on the PMT=70% scan, and has a coefficient of variation in that scan of 0.13. Four other gene probes with expected strong response to the hybridised target fulfilled this biological expectation but were not measured close to saturation. For selection of a reference gene with lower expression, ‘vMC149’ was chosen amongst the remainder of medium and low expressed genes on the basis of having the best reproducibility ( $CV = 0.17$ ) and best correlation (albeit low at  $r=0.18$ ) with cMU012, as measured across 75 replicates on the wafer scanned at 70% PMT. The reference gene is merely used to have a robust control that is not affected by pixel censoring issues even at the highest scan setting.

Fig 3 shows results for the 75 replicates of each the two genes. Scans at laser settings of 80%, 90% and 100%, which were effected by saturation, are plotted against results for laser setting 70%, which was unaffected. All values should lie on the diagonal line, not only those for the low expressed gene. We see that uncorrected values for the high expressed gene lie below the line. PC correction improves the agreement with the diagonal, but progressively less effectively at increasing laser settings, possibly because spot shape/profile changes with intensity.

The ratio  $R$  between measured gene expression for cMU012 and vMC149 was calculated for all corrected and uncorrected scan data sets. For the scan without saturated spots (PMT=70%)  $R = 17.3$ , which can reasonably be assumed to be the biologically true measured expression ratio between these two genes. Calculating  $R$  for saturated but uncorrected scan data yields  $R_{80} = 15.2$ ,  $R_{90} = 6.9$  and  $R_{100} = 3.3$ , which clearly and expectedly show that the 16-bit measurement barrier deflates this ratio as the fluorescence for the lowly expressed reference gene rises with the high expressed gene fixed at this barrier. Following PC correction of the high scan images,  $R$  is re-calculated and yields  $R_{80,pc} = 17.2$ ,  $R_{90,pc} = 9.2$  and  $R_{100,pc} = 3.8$ . Consistent with Fig 3, it appears that PC correction is effective for low to moderate levels of censoring, but not if it is severe.

## 4 Discussion

If no account is taken of censoring, where scanned microarrays have saturated pixel values, then estimates of gene expression levels are biased. We have shown that a linear model, based on

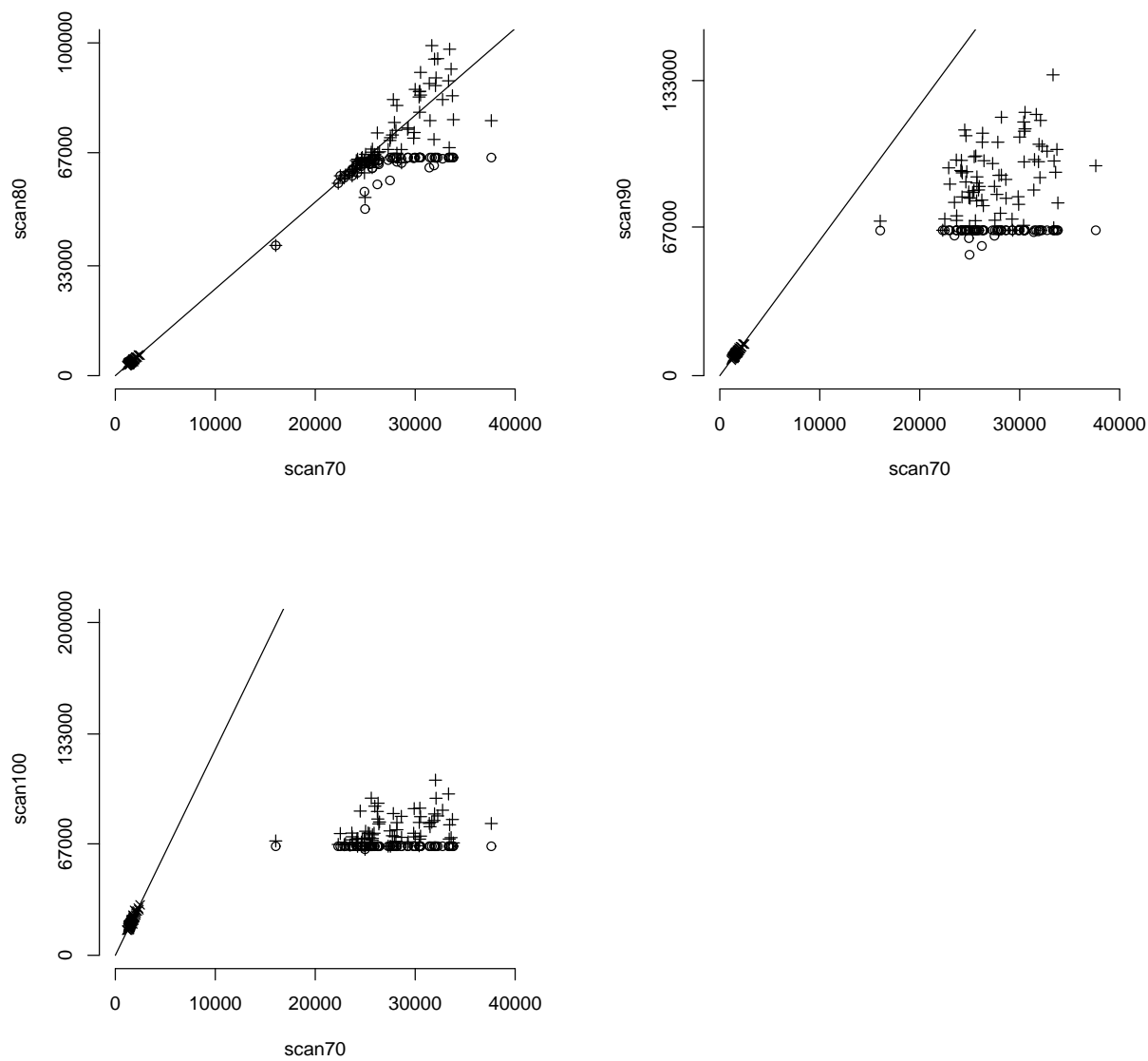


Figure 3: Estimated signals from 75 replicates of two reference genes from three scans affected by saturation plotted against one that was not:  $\times$  - low expressed gene,  $\circ$  - high expressed gene,  $+$  - high expressed gene after PC correction. The diagonal line indicates where points should lie.

the PCs of uncensored spots on the same array, is effective in correcting for the bias. It is also computationally fast, and flexible to adapt to distinctive spot shapes and profiles on different arrays.

PC correction appears effective for low to moderate levels of saturation. This suits experimental practice, as arrays are usually scanned only once and the laser power and PMT settings are chosen to optimise fluorescent signal for lower expressed genes while minimising saturation for highly expressed genes. Data from array images which show large amounts of saturation (a subjective assessment) are normally discarded, and would also not be suitable for correction by the proposed PC method. However, where the trade-off between detection of low/high expressed genes is not optimal, a small proportion of genes can be saturated and may prove to be recoverable by PC correction.

The PC approach could be modified in several ways. For example, we could use log-transformed pixel values, or omit the mean term in computing the PCs. If the number of uncensored spots is comparatively small, we could augment the set from which the PCs are derived by including shifted, rotated and reflected versions. Instead of using least squares estimation, we could take account of censored values in estimating  $\beta$  by replacing least squares by maximum likelihood estimation, as Ekstrom et al. (2004) did. Similarly, in imputing censored values, we could use conditional means instead of ordinary means. The method is also applicable to digital images arising in other genomic technologies, such as electrophoresis gels (Rogers et al., 2003), and in the case of dual-dye experiments it may be possible to combine our model with that of Dodd et al. (2004) to exploit the correlation between channels.

## Acknowledgements

The work was supported by funds from the Scottish Government. Design and production of the silicon nitride wafer was DTI funded. We thank Colin Campbell and Alan Ross for providing the data in §3.

## References

- Axon Instruments, Inc (1999). *GenePix 400A User's Guide*.
- Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models – their training and application. *Computer Vision and Image Understanding*, 61:38–59.
- Dodd, L. E., Korn, E. L., McShane, L. M., Chandramouli, G. V. R., and Chuang, E. Y. (2004). Correcting log ratios for signal saturation in cDNA microarrays. *Bioinformatics*, 20:2685–2693.
- Eisen, M. B. (1999). *ScanAlyze*. (Available at <http://rana.Stanford.EDU/software/>).
- Ekstrom, C. T., Bak, S., Kristensen, C., and Rudemo, M. (2004). Spot shape modelling and data transformation for microarrays. *Bioinformatics*, 20:2270–2278.

- Glasbey, C. A. and Ghazal, P. (2003). Combinatorial image analysis of DNA microarray features. *Bioinformatics*, 19:194–203.
- GSI Luminomics (1999). *QuantArray Analysis Software, Operator's Manual*.
- Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67.
- Jain, A. N., Tokuyasu, T. A., Snijders, A. M., Segraves, R., Albertson, D. G., and Pinkel, D. (2002). Fully automatic quantification of microarray image data. *Genome Research*, 12:325–332.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer-Verlag, New York, 2nd edition.
- Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B*, 34:1–41.
- Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press, London.
- Rogers, M., Graham, J., and Tonge, R. P. (2003). Statistical models of shape for the analysis of protein spots in two-dimensional electrophoresis gel images. *Proteomics*, 3:887–896.
- Steinfath, M., Wruck, W., Seidel, H., Lehrach, H., Radelof, U., and O'Brien, J. (2001). Automated image analysis for array hybridization experiments. *Bioinformatics*, 17:634–641.
- Terry, J. G., Campbell, C. J., Ross, A. J., Livingston, A. D., Buck, A. H., Dickinson, P., Mountford, C. P., Evans, S. A. G., Mount, A. R., Beattie, J. S., Crain, J., Ghazal, P., and Walton, A. J. (2006). Improved silicon nitride surfaces for next-generation microarrays. *Langmuir*, 22:11400–11404.
- Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of the Optical Society of America, Series A*, 4:519–524.
- Wierling, C. K., Steinfath, M., Elge, T., Schulze-Kremer, S., Aanstad, P., Clark, M., Lehrach, H., and Herwig, R. (2002). Simulation of DNA array hybridization experiments and evaluation of critical parameters during subsequent image and data analysis. *BMC Bioinformatics*, 3:29.
- Wit, E. and McClure, J. (2003). Statistical adjustment of signal censoring in gene expression experiments. *Bioinformatics*, 19:1055–1060.