

A calibration transform for families of spectroscopes

B. Baeumer¹R. Nicholson²S. W. Taylor³

(Received 22 August 2016; revised 26 July 2017)

Abstract

We seek a transform that will map between the outputs of pairs of similar spectroscopes. The spectroscopes are used to sort fruit and must regularly be calibrated to allow for seasonal fruit variation. The aim is to obtain transforms that would allow calibration of a number of such spectroscopes from the calibration of one of them.

Contents

1 Introduction

M269

2 Using principal component analysis to develop transforms

M272

DOI:10.21914/anziamj.v57i0.11238, © Austral. Mathematical Soc. 2017. Published August 1, 2017, as part of the Proceedings of the 2015 Mathematics and Statistics in Industry NZ Study Group. ISSN 1445-8810. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to the DOI for this article. Record comments on this article via

<http://journal.austms.org.au/ojs/index.php/ANZIAMJ/comment/add/11238/0>

1

Introduction

M269

3

Geometric mean transform

M278

3.1

Fitting data to the model

M281

3.2

Testing the model

M283

1

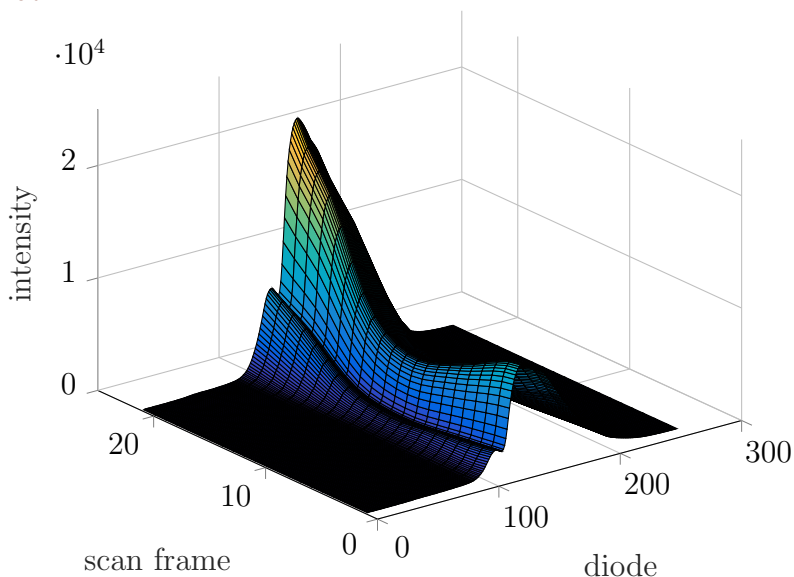
Introduction

Compac Sorting designs fruit sorting equipment for orchards. A key component of this equipment is a set of several spectroscopes, each of which is used to handle a line of fruit. The fruit varies from season to season and from year to year, and each spectroscope needs to be manually calibrated for each type of fruit to allow for these variations in fruit. This recalibration operation is time consuming because it requires human examination of large quantities of fruit. Further, even though the spectroscopes are built with identical design specifications, the small differences between spectroscopes make it necessary for each spectroscope to be calibrated separately.

Compac Sorting challenged the Study Group with the task of creating a transform that would map the output of one spectroscope to the output of another. In other words, we were tasked with the problem of using the knowledge of the spectral measurement of an item of fruit from some chosen machine to predict the spectral measurements of that item of fruit from all of the other machines. If we could achieve this, then only one of the several machines in each orchard would need to be calibrated manually; the others could be calibrated by using the recalibration of the one chosen machine and the transform.

The company provided data consisting of scans of various kinds of fruit. The data was sorted into groups of specific fruit types (for example, specific types of apple) that had been scanned in specific sorting houses. The problem was to find a transform that would map the data of one machine to that of all of the others in any particular sorting house. The data itself consisted of readings from each of the 256 diodes of the spectroscopes as the fruit went through

Figure 1: An example of a scan of an apple. This scan consisted of 21 frames for each of the 256 diodes and the intensity is represented as the z values in a surface plot.



the machine. Each item of fruit would be scanned up to approximately thirty times as it went through the machine. Each diode is designed to respond to a certain narrow band of light wavelengths so that all 256 diodes give a good spectral representation of the fruit. The spectrometers give a dimensionless numerical value for diode responses. As we need to compare these values for different spectrometers, it is useful to think of these values as approximations for the same physical quantity which we call the *intensity*. Our “intensity” is not the “intensity” that is physically related to the energy of a light wave, but it is likely to be related to it. We call each set of 256 intensity measurements taken at a particular time a *scan frame* or *frame*. A typical example showing approximately twenty frames is shown in [Figure 1](#).

Because of the way the data was arranged, it was easiest for us to to start by

trying to construct a transform that works for a particular fruit type. The group started the analysis by plotting the data and soon discovered that plots from different machines were indeed different, but they had many features in common. This inspired the group to see if the data from each machine could be represented as a finite linear combination of certain “shapes”. This could be done mathematically by calculating eigenvectors of the covariance matrices of the data. There is a well known procedure for doing this, called “principal component analysis”, or more commonly, “PCA”, which is closely related to singular value decomposition.

The group had some success with this PCA approach, measured by comparing the results of Compac’s quality testing algorithm acting on both raw data and data that had been represented in terms of these eigenvectors (the group called these *eigen-apples*).

The next step was to see if we could construct mappings from the PCA representation of one machine to the PCA representation of others. We made many attempts at doing this, as well as attempts at matching the raw data of one machine to another. The method of measuring success or failure in these approaches was to test our results with Compac’s quality testing algorithm. It seemed that we were getting close at times, but we were not able to obtain results that were useful. Details of this work are given in [Section 2](#).

Another approach that might work is to treat the problem as an “inverse problem”. This is a very successful area of applied mathematics which involves constructing a mathematical model for the physical system being measured and using measurements to estimate any unknown parameters in the model. It is conceivable that the difference in outputs of these machines, which have the same design specifications, could be explained with just a few differing parameters in a model. It is also possible that a suitable model could be discovered empirically. Our initial work on the problem indicates that it would be worth trying this or possibly other approaches to get a suitable solution. We detail some initial work in this direction in [Section 3](#).

2 Using principal component analysis to develop transforms

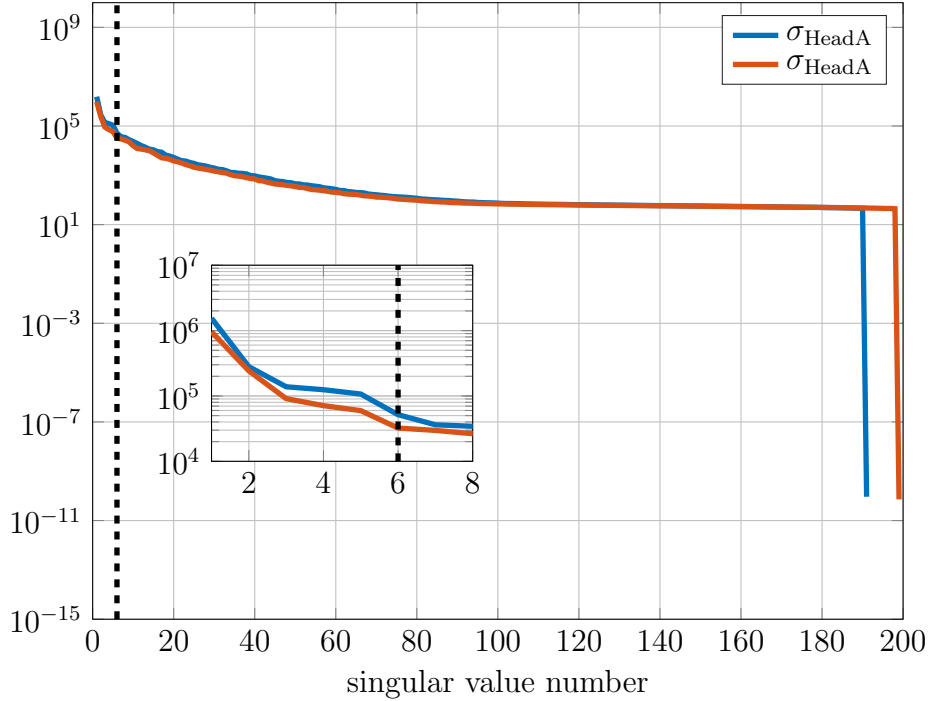
Here we outline the attempts made at finding a transform which maps the spectral measurement of a specific apple of a chosen machine to spectral measurements of the same apple taken by a different chosen machine. These attempts mostly relied on principal component analysis (PCA), for which we refer the reader to the tutorial by Shlens [3]. PCA is very closely related to singular value decomposition (SVD) (Golub and Van Loan [1]).

We briefly outline the PCA algorithm used. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_\ell]$ be our data matrix, where each column consists of the vectorised spectral measurements of a specific apple. Unfortunately the number of scans performed on each apple varies. To overcome this obstacle we only consider the intensity measurements from the diodes of the middle nine measurements of each apple. This allows us to eliminate the first and last couple of scans for each apple which are sometimes corrupted due to the measurement system. Recalling that each scan consists of 256 diode measurements of the intensity, the data matrix $\mathbf{X} \in \mathbb{R}^{2304 \times \ell}$, and for the data made available by Compac, we typically have $\ell \approx 200$ apples.

The first step in the PCA procedure is to centralise the data, that is, we subtract the mean of each row from all columns of \mathbf{X} . We denote by $\bar{\mathbf{X}}$ the centralised version of \mathbf{X} , that is $\bar{\mathbf{X}} = \mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}} \otimes \mathbf{1}_{2304}$, where $\boldsymbol{\mu}_{\mathbf{X}}, \mathbf{1} \in \mathbb{R}^{2304}$ with $\boldsymbol{\mu}_{\mathbf{X}} = \frac{1}{\ell} \sum_{k=1}^{\ell} \mathbf{x}_k$ being the column vector of means for each of the 256 diode measurements across each of the nine scans, and $\mathbf{1}_{2304}$ is the 1's vector of length 2304, and \otimes denotes the outer product.

Next we simply compute the SVD of $\bar{\mathbf{X}}$ to find the singular components. These are the eigenvectors of the covariance matrix $\Gamma_{\mathbf{X}} = \bar{\mathbf{X}}\bar{\mathbf{X}}^T/(\ell - 1)$. The governing idea behind PCA is that if our data is (highly) correlated in a linear sense, a (largely) reduced number of principal components accurately represents the data. It is for this reason that PCA is often used for data

Figure 2: Log plot of the singular values for each head

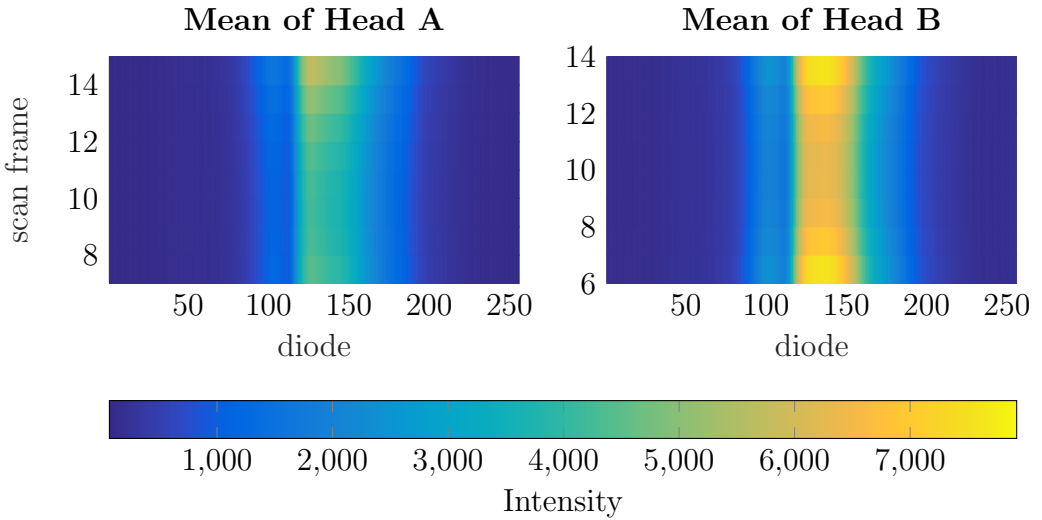


summarisation and compression.

The industry often refers to a particular sorting line as being a “head”, and we adopt this terminology to refer to the spectrometers. Thus “Head A”, “Head B” simply refer to two spectrometers labelled A and B.

Figure 2 presents in a log plot the singular values for the data from Head A and from Head B. The figure shows that the *effective rank* of the covariance matrix Γ_X is very small, with 87% of the variation in the data of Head A and 88% of the variation in the data of Head B being explained by the mean and first six respective principal components. The *mean apple* for both heads are shown in Figure 3, while the first six “eigen-apples” for Head A and Head B are shown in Figures 4 and 5 respectively.

Figure 3: The *mean apples*



Some effort went into finding a linear transform (say L_1) which mapped the first few principal components of Head A to those of Head B; that is, find L_1 such that

$$L_1 \bar{U}_{\text{HeadA}} = \bar{U}_{\text{HeadB}} ,$$

where $\bar{U}_{\text{HeadA}}, \bar{U}_{\text{HeadB}} \in \mathbb{R}^{2304 \times 6}$ are the matrices made up of the first six principal components as columns of Head A and Head B respectively. However, due to noise and other anomalies, it is more appropriate to search for L_1 such that,

$$\|L_1 \bar{U}_{\text{HeadA}} - \bar{U}_{\text{HeadB}}\|_F^2 \tag{1}$$

is minimised, with $\|\cdot\|_F$ denoting the Frobenius norm. One possible solution is $L_1^{\text{MP}} = \bar{U}_{\text{HeadB}} \bar{U}_{\text{HeadB}}^\dagger$, where $\bar{U}_{\text{HeadB}}^\dagger \in \mathbb{R}^{2304 \times 2304}$ denotes the right hand pseudoinverse. However, results showed that for intensity measurements for a specific apple from Head A and Head B (say $I_{\text{Head A}}$ and $I_{\text{Head B}}$), we had

Figure 4: The first six principal components (*eigen-apples*) for Head A

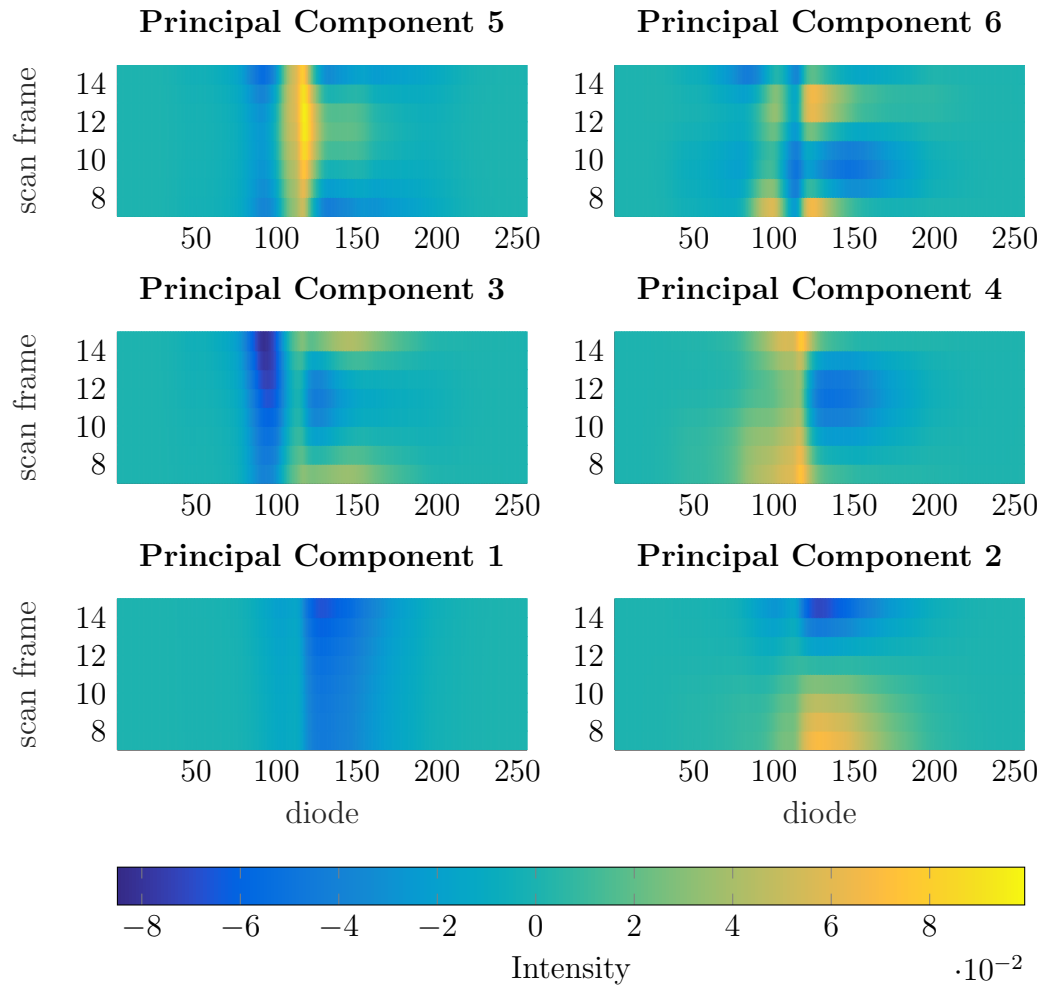
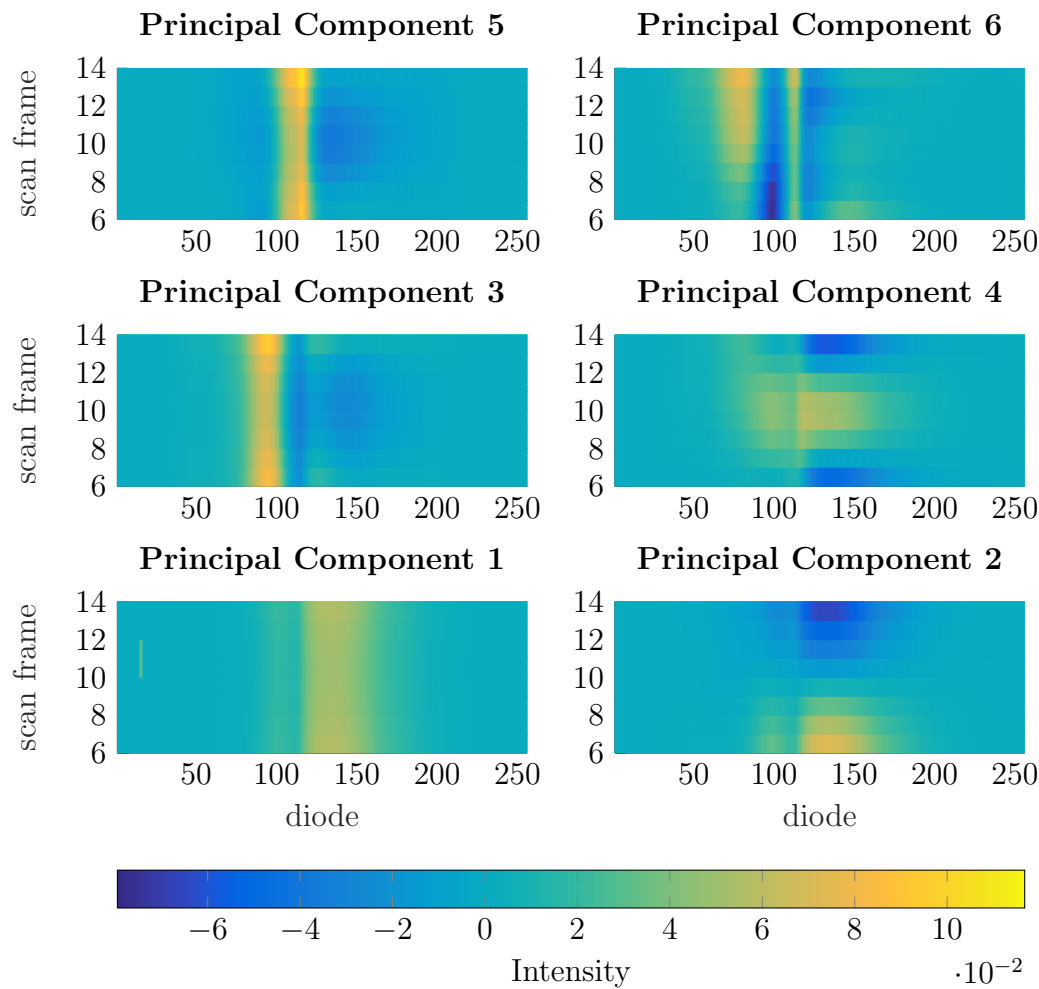


Figure 5: The first six principal components (*eigen-apples*) for Head B



$\|L_1^{\text{MP}} I_{\text{HeadA}} - I_{\text{HeadB}}\| > \|I_{\text{HeadA}} - I_{\text{HeadB}}\|$ in general, meaning L_1^{MP} was not a suitable mapping.

One reason for the poor results attained by using L_1^{MP} may be that although $L_1^{\text{MP}} \bar{U}_{\text{HeadA}} \approx \bar{U}_{\text{HeadB}}$, and is optimal in the sense of (1), it does not take into account the singular values. A remedy of this shortcoming is to introduce a family of mappings, L_1^α ($\alpha \geq 0$), such that L_1^α minimises the updated functional

$$\|L_1^\alpha \bar{U}_{\text{HeadA}} - \bar{U}_{\text{HeadB}}\|_{\text{F}}^2 + \alpha^2 \|L_1^\alpha \bar{\Sigma}_{\text{HeadA}} - \bar{\Sigma}_{\text{HeadB}}\|_{\text{F}}^2, \quad (2)$$

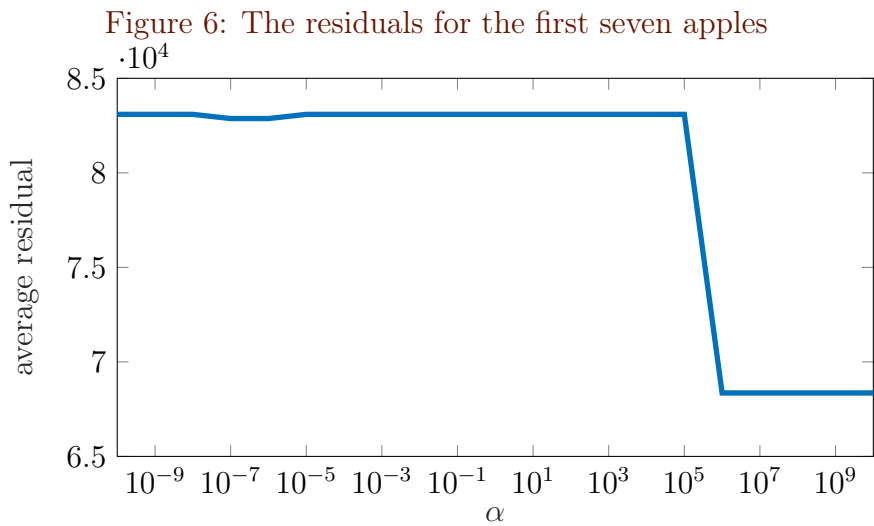
where $\bar{\Sigma}_{\text{HeadA}}, \bar{\Sigma}_{\text{HeadB}} \in \mathbb{R}^{2304 \times 6}$ have the singular values of Head A and Head B on the diagonal respectively, and are zero elsewhere. Different choices of the *regularisation parameter* α correspond to different weights put on either matching the singular values, or matching the principal components, with higher values of alpha lending more weight to minimising the residual $\|L_1^\alpha \bar{\Sigma}_{\text{HeadA}} - \bar{\Sigma}_{\text{HeadB}}\|_{\text{F}}^2$. The form of functional of (2) is a common one associated with regularisation, and is often encountered when fitting data or parameter estimation, as discussed by Kaipio and Somersalo [2]. By making the substitutions $V_{\text{HeadA}} = [\bar{U}_{\text{HeadA}} \quad \alpha \bar{\Sigma}_{\text{HeadA}}]$ and $V_{\text{HeadB}} = [\bar{U}_{\text{HeadB}} \quad \alpha \bar{\Sigma}_{\text{HeadB}}]$ we restate our problem as find L_1^α such that

$$\|L_1^\alpha V_{\text{HeadA}} - V_{\text{HeadB}}\|_{\text{F}}^2 \quad (3)$$

is minimised. We take the solution as $L_1^\alpha = V_{\text{HeadB}} V_{\text{HeadB}}^\dagger V_{\text{HeadA}}^\dagger$.

We neglected both head means, μ_{HeadA} and μ_{HeadB} , and thus a more natural test for the effectiveness of L_1^α is to apply it to centralised data. That is, we consider the residuals $r = \|L_1^\alpha (I_{\text{HeadA}} - \mu_{\text{HeadA}}) - (I_{\text{HeadB}} - \mu_{\text{HeadB}})\|$. To set a suitable value for the regularisation parameter we vary α and observe the behaviour of the residuals: Figure 6 shows the average residuals for the first 20 apples.

In general the residuals appear smallest for $\alpha \gtrsim 10^6$. However, this would imply basically all weight should be placed on the term $\|L_1^\alpha \bar{\Sigma}_{\text{HeadA}} - \bar{\Sigma}_{\text{HeadB}}\|_{\text{F}}^2$,



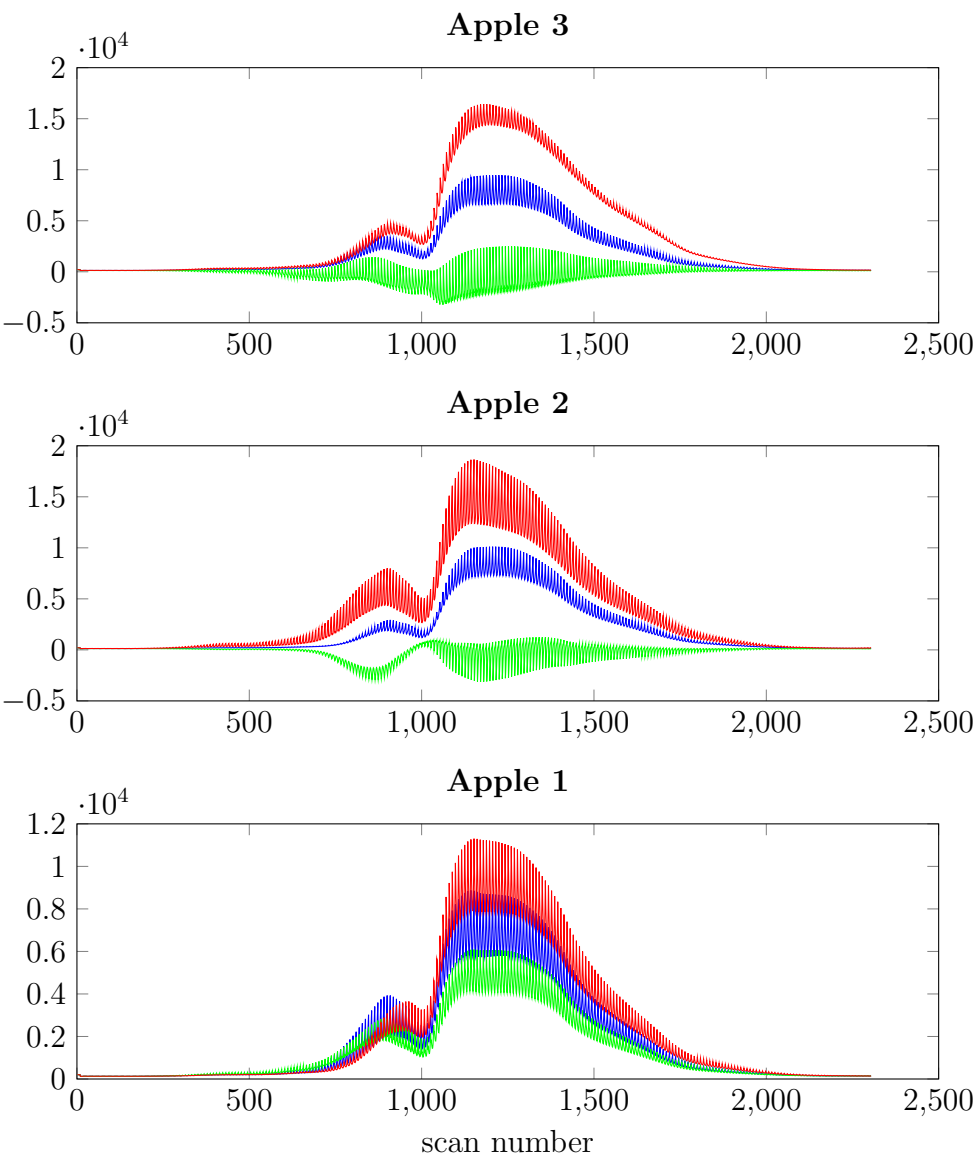
and our mapping would necessarily be diagonal, with only the first ten entries on the diagonal nonzero. A slight dip in the residual can be seen for values of the regularisation parameter at approximately 10^{-6} ; however, results showed that using this value for α gave no improvement to the results.

Figure 7 compares the results for three different apples.

3 Geometric mean transform

In this section we construct a mapping from the scan intensities measured at one machine (called Head A) to those measured at another (Head B). Let \mathbf{x} and \mathbf{y} denote data measured at A and B respectively. We can think of \mathbf{x} and \mathbf{y} as being functions of \mathbf{d} and \mathbf{s} , where $1 \leq \mathbf{d} \leq 256$, \mathbf{d} represents diode number, and \mathbf{s} is an appropriate index for the scan frames of a data set.

Figure 7: Values of I_{HeadA} (red), I_{HeadB} (blue), and $L_1^\alpha I_{\text{HeadA}}$ (green) for the 2304 scans of three apples.



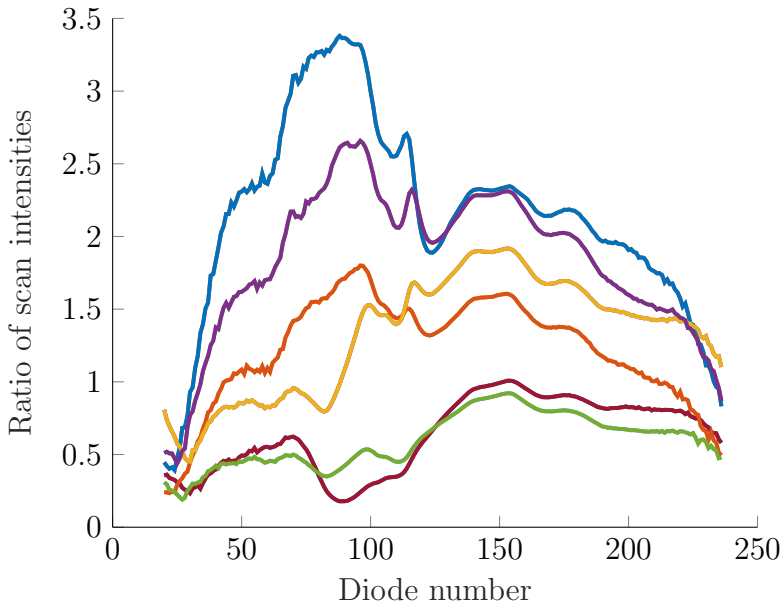


Figure 8: Ratio of scan intensity for two different heads. The “dark reference” measurements have been subtracted from the intensities. Each colour represents a particular item of fruit.

The simplest model that one might expect to work is

$$\mathbf{y}(\mathbf{d}, \mathbf{s}) = f(\mathbf{d})\mathbf{x}(\mathbf{d}, \mathbf{s}), \tag{4}$$

which says that the response at diode \mathbf{d} for machine B is proportional to that at diode \mathbf{d} for machine A, with the proportionality constant $f(\mathbf{d})$ depending on the diode number. If the machines were truly identical, then $f(\mathbf{d})$ would be equal to one for all \mathbf{d} .

If the model (4) were accurate, then the ratio $\mathbf{y}(\mathbf{d}, \mathbf{s})/\mathbf{x}(\mathbf{d}, \mathbf{s})$ would be equal to $f(\mathbf{d})$, and thus would be independent of \mathbf{s} . The plots in Figure 8 of this ratio for several scans indicate that this does depend on the scan in some random way.

The graphs have a similar shape, at least for diode numbers greater than approximately 100. As shown in [Figure 1](#), the intensities can be quite small for diode numbers less than 100, so the discrepancy in shape for such diode numbers may not be significant. The similarities in the shapes of the graphs suggests that we should modify Equation (4) to

$$\mathbf{y}(\mathbf{d}, \mathbf{s}) = \mathbf{f}(\mathbf{d})\mathbf{p}(\mathbf{s})\mathbf{x}(\mathbf{d}, \mathbf{s}). \quad (5)$$

The proportionality parameter $\mathbf{p}(\mathbf{s})$ depends on \mathbf{s} but not on the diode number. The idea is that $\mathbf{p}(\mathbf{s})$ accounts for the random effects that arise due to fruit placement in each machine. Thus, in this model, fruit placement in the machine scales scan intensity by the same random factor $\mathbf{p}(\mathbf{s})$ for each diode, and this factor can be different for each scan.

3.1 Fitting data to the model

It is easiest to fit a set of data to the model (5) by taking logarithms, and thus converting the product to a sum,

$$\log(\mathbf{y}(\mathbf{d}, \mathbf{s})) = \log(\mathbf{f}(\mathbf{d})) + \log(\mathbf{p}(\mathbf{s})) + \log(\mathbf{x}(\mathbf{d}, \mathbf{s})).$$

We write this as

$$\mathbf{Y}(\mathbf{d}, \mathbf{s}) = \mathbf{F}(\mathbf{d}) + \mathbf{P}(\mathbf{s}) + \mathbf{X}(\mathbf{d}, \mathbf{s}), \quad (6)$$

where $\mathbf{Y}(\mathbf{d}, \mathbf{s}) = \log(\mathbf{y}(\mathbf{d}, \mathbf{s}))$, $\mathbf{X}(\mathbf{d}, \mathbf{s}) = \log(\mathbf{x}(\mathbf{d}, \mathbf{s}))$, $\mathbf{F}(\mathbf{d}) = \log(\mathbf{f}(\mathbf{d}))$ and $\mathbf{P}(\mathbf{s}) = \log(\mathbf{p}(\mathbf{s}))$.

We have measured data $\mathbf{Y}_{\mathbf{d}\mathbf{s}}$ and $\mathbf{X}_{\mathbf{d}\mathbf{s}}$ for $\mathbf{Y}(\mathbf{d}, \mathbf{s})$ and $\mathbf{X}(\mathbf{d}, \mathbf{s})$ respectively and we wish to find parameters $\mathbf{F}_{\mathbf{d}} = \mathbf{F}(\mathbf{d})$ for each diode \mathbf{d} and $\mathbf{P}_{\mathbf{s}} = \mathbf{P}(\mathbf{s})$ for each scan frame \mathbf{s} such that (6) holds as closely as possible. We do this by minimising the sum of squares of $\mathbf{Y}_{\mathbf{d}\mathbf{s}} - \mathbf{X}_{\mathbf{d}\mathbf{s}} - \mathbf{F}_{\mathbf{d}} - \mathbf{P}_{\mathbf{s}}$. There is a small amount of redundancy in these parameters in that this quantity remains the same if we add some number to $\mathbf{F}_{\mathbf{d}}$ and subtract the same number from $\mathbf{P}_{\mathbf{s}}$. We remove this redundancy by including an extra condition in our model

to reflect the idea that the scale factor $p(s)$ should vary randomly but not too far from the value one. We do this by stipulating that the average value of $\log(p(s))$ is zero. Thus we minimise our sum of squares subject to the constraint that $\sum_s P_s = 0$. This constraint is incorporated into our least squares minimisation by use of a Lagrange multiplier term, with Lagrange multiplier λ . Thus we minimise

$$J = \frac{1}{2} \sum_{s,d} (Y_{ds} - X_{ds} - F_d - P_s)^2 + \lambda \sum_s P_s,$$

and then choose λ so that the constraint holds. Differentiating with respect to F_d gives

$$0 = \sum_s (Y_{ds} - X_{ds} - F_d - P_s) = \sum_s (Y_{ds} - X_{ds} - F_d),$$

because of the constraint. This can be written as

$$F_d = \frac{1}{N} \sum_s (Y_{ds} - X_{ds}), \quad (7)$$

where N is the number of scans in the data set. Differentiating with respect to P_s gives

$$0 = \sum_d (Y_{ds} - X_{ds} - F_d - P_s) + \lambda,$$

so

$$P_s = \frac{1}{M} \sum_d (Y_{ds} - X_{ds} - F_d) + \frac{\lambda}{M}, \quad (8)$$

where M is the number of diodes used in the data set. The Lagrange multiplier λ is determined so that the constraint is satisfied. For this, we need

$$\lambda N + \sum_{s,d} (Y_{ds} - X_{ds} - F_d) = 0,$$

so

$$\lambda = -\frac{1}{N} \sum_{s,d} (Y_{ds} - X_{ds} - F_d) = 0.$$

Thus Equation (8) gives

$$P_s = \frac{1}{M} \sum_d (Y_{ds} - X_{ds} - F_d). \quad (9)$$

Taking exponentials gives the desired estimate of the function

$$\begin{aligned} f(d) &= e^{F_d} = \exp \left\{ \frac{1}{N} \sum_s (Y_{ds} - X_{ds}) \right\} \\ &= \exp \left\{ \frac{1}{N} \sum_s [\log(y_{ds}) - \log(x_{ds})] \right\} \\ &= \left(\prod_s \frac{y_{ds}}{x_{ds}} \right)^{1/N}, \end{aligned} \quad (10)$$

where \prod_s denotes the product of a quantity indexed by the scan frame parameter s . Thus $f(d)$ is the *geometric mean* of the ratio of scan intensities at diode d .

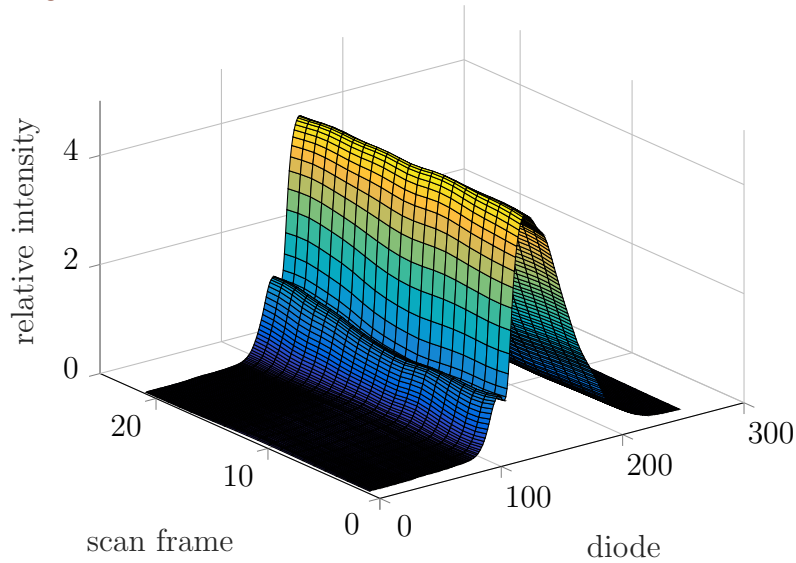
Similarly, $p_s = e^{P_s}$, which measures the random scale factor in scan frame s is given by a geometric mean over all diodes:

$$p_s = e^{P_s} = \left(\prod_d \frac{1}{f(d)} \frac{y_{ds}}{x_{ds}} \right)^{1/M}.$$

3.2 Testing the model

We chose to use measurements of scan intensities near the middle of the 20 or so frames for each fruit scan. Thus for 21 frames we would choose the 11th. However, $f(d)$ depends on the shape of the scan curve, not its scale. The frames of the fruit scan shown in [Figure 1](#), are re-rendered in [Figure 9](#) by dividing the scan intensity by its average value for the particular scan frame.

Figure 9: The scan shown in [Figure 1](#), re-rendered after dividing the intensity by its average for each frame.



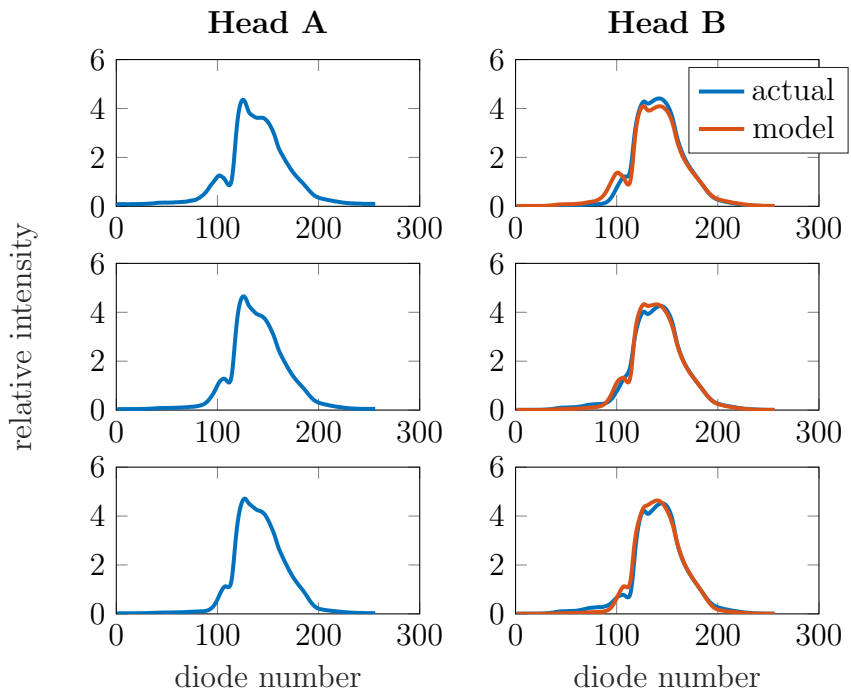
This figure indicates that the shape does not change much for each frame, so choosing the middle frames was probably unnecessary.

We used Equation [\(10\)](#) to estimate the scale function $f(d)$ for two machines on a certain date, and used the same function to predict the scan intensity at machine B from that at machine A on the same date, and at different dates. The graphs show the *relative intensities*, calculated by dividing the intensity by its mean value.

The foregoing plots of intensity are all relative to the average intensity and thus the random scale factor $p(s)$ has been scaled out of these plots. To give an idea of the effect of this scale factor, [Figure 13](#) presents a histogram of the values of $P(s) = \log(p(s))$ for the test data.

The standard deviation for this data is 0.90, which means that the random scale factor for intensity when comparing scans at Head A and Head B is

Figure 10: Actual and predicted relative intensities on the calibration data



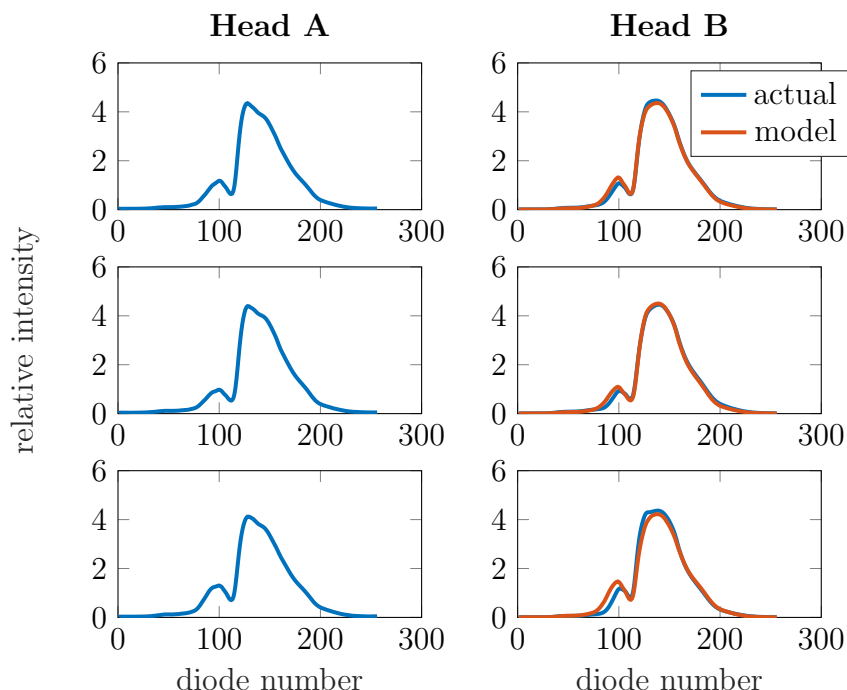
usually in the vicinity of $e^{-0.9} = 0.41$ to $e^{0.9} = 2.5$.

Acknowledgements We are grateful to Owen Dillon and Pascal Cheon (University of Auckland) and industry representatives Scott Rusby, Scott Walbran and Kate O’Byrne (Compac Sorting) for their work on this problem.

References

[1] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Fourth. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins

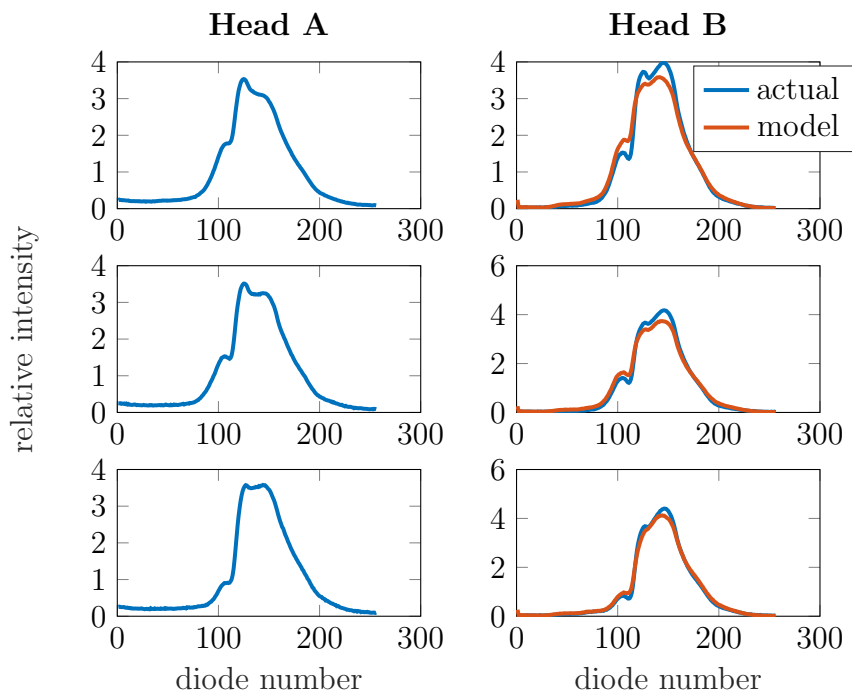
Figure 11: Actual and predicted relative intensities on the the same date but with calibration data on an earlier run



University Press, Baltimore, MD, 2013, pp. xiv+756. ISBN: 978-1-4214-0794-4; 1-4214-0794-9; 978-1-4214-0859-0 (cit. on p. M272).

- [2] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*. Vol. 160. Applied Mathematical Sciences. Springer-Verlag, New York, 2005, pp. xvi+339. ISBN: 0-387-22073-9 (cit. on p. M277).
- [3] Jonathon Shlens. *A Tutorial on Principal Component Analysis*. Tech. rep. <http://arxiv.org/abs/1404.1100>, 2014 (cit. on p. M272).

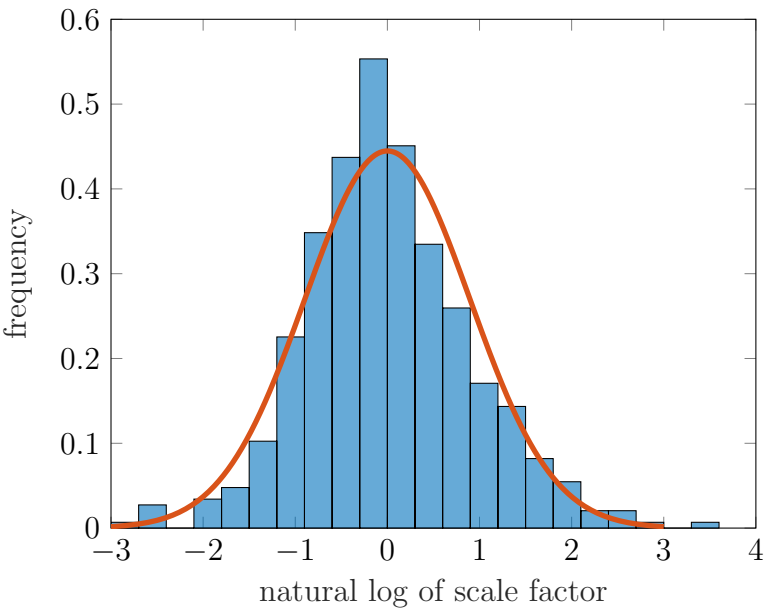
Figure 12: Actual and predicted relative intensities two years later than calibration



Author addresses

1. **B. Baeumer**, Department of Mathematics and Statistics, University of Otago, PO Box 56, Dunedin 9054, NEW ZEALAND.
<mailto:bbaeumer@maths.otago.ac.nz>
 orcid:0000-0003-3841-1094
2. **R. Nicholson**, School of Natural Sciences, University of California, Merced, CA 95443, USA.
<mailto:rnicholson@ucmerced.edu>
3. **S. W. Taylor**, Mathematics Department, The University of Auckland,

Figure 13: A histogram of $P(s) = \log(p(s))$ for the test data. A normal density function with mean zero and the same variance has been superimposed.



Private Bag 92019, Auckland, NEW ZEALAND.
<mailto:s.taylor@auckland.ac.nz>
orcid:[0000-0003-1633-3582](https://orcid.org/0000-0003-1633-3582)