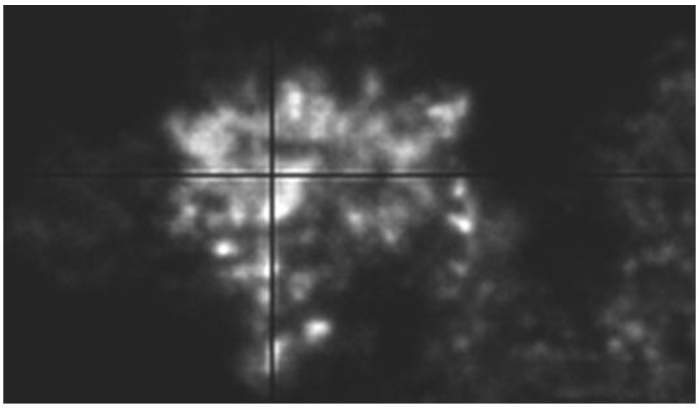
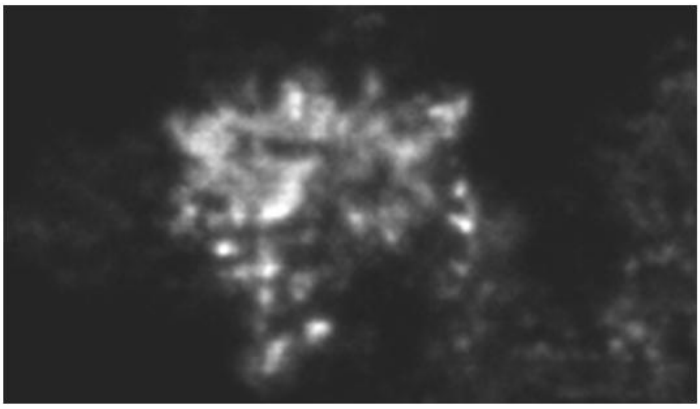
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Image restoration and analysis of biomolecular radioactivity images





Note no SAMBA/27/08

Authors Marit Holden
Ingrid K. Glad
Håvard H. Hauge
Knut Liestøl

Date 26th June 2008

The authors

Biomolex AS

Biomolex AS was established in May 2001 by Ideas and the Research Foundation (RF) of the Norwegian Radium Hospital. Biomolex AS supplies world leading systems for functional genomics and proteomics. The main product is the Biomolex MicroArray Reader which supporting real time digital imaging with high resolution and sensitivity for accurate quantification of protein expression arrays and tissue samples. The technology supports detection of biological samples labeled with radioisotopes. Compared to traditional detection system for radioactive emissions our technology offers advantages like; real time detection through single event counting, energy registration, signal linearity over 5 log, increased sensitivity and improved image resolution.
www.biomolex.no

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modeling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Image restoration and analysis of biomolecular radioactivity images**
Authors	**Marit Holden** <marit.holden@nr.no> **Ingrid K. Glad** <glad@math.uio.no> **Håvard H. Hauge** <hauge@biomolex.no> **Knut Liestøl** <knut@ifi.uio.no>
Date	26th June 2008
Publication number	SAMBA/27/08

Abstract

Biomolex has technology for real-time imaging of radioactive emissions. Images are acquired for measuring radioactivity in tissue slices for assessing transport and uptake of labeled biomolecules or in spots in protein kinase arrays for evaluating protein phosphorylations. The sensor used is a silicon strip detector, and a problem is that some strips produce no signal. Also, some other strips might have too high or low signals.

A method for restoring the images and imputing the missing data has been developed for images of tissue slices. For the protein kinase arrays a method for estimating the radioactivity in each spot in the image has been developed. The image after missing data have been imputed, is one of the outputs when applying this method.

The Biomolex technology opens up for the possibility of using multiple isotopes during a single image acquisition. A method for decomposing the measured signal into signals for the different isotopes is proposed and tested on simulated data.

Keywords	Biomolecular radioactivity images; Image analysis; Image restoration; Missing stripes; Estimating spot intensities; Separation of signals
Target group	
Availability	Open
Project	SFI08MisStri
Project number	220308
Research field	Health, Bioinformatics
Number of pages	34

Contents

1 Introduction

Biomolex has technology for real-time digital imaging of radioactive emissions. In Section 1.1 we will describe this technology, while its applications will be described in Section 1.2. Finally, challenges with the Biomolex technology, which is also the topic of this note, will be introduced in Section 1.3.

1.1 The Biomolex technology

The detector of the Biomolex technology is a double-sided silicon strip detector (DSSD). It consists of 640×1280 parallel strips running orthogonal on each side of a silicon core with a strip pitch of $50\mu m$ (see Figure 1). The DSSD will respond to energy depositions in the silicon core of the sensor. For each radioactive event that hits the detector the energy and the exact position, with a precision down to $50\mu m$ is registered. The high sensitivity

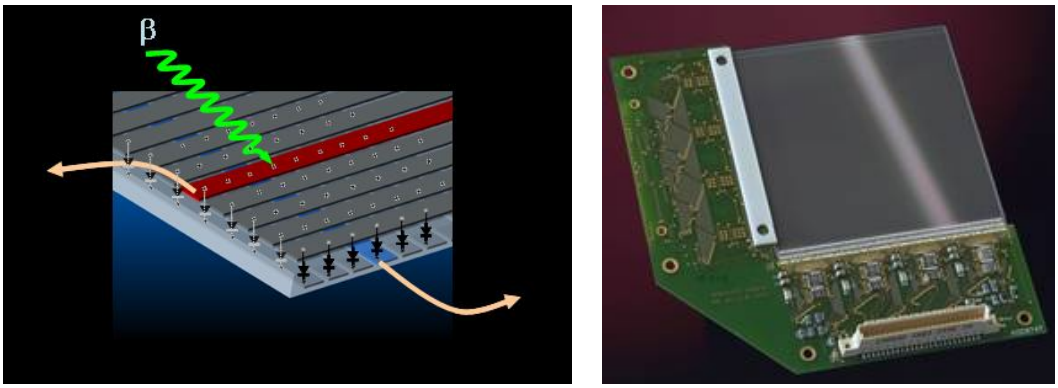


Figure 1. *Left panel: A schematic illustration of the DSSD principle. Right panel: The DSSD attached to its adjoining read-out electronics.*

of the patented read-out electronics allows detection of particles with energies as low as 20-25 keV, thus supporting the use of the most common β-emitters used for labeling of biomolecules e.g. ^{33}P, ^{35}S, ^{14}C, ^{131}I, ^{18}F, ^{75}Se, ^{89}Sr, ^{99}Tc, ^{153}Sa. These are all low energy emitters which are easily shielded.

During an array recording, the sensor is bombarded with a high number of β-particles. For each of the pixels/strip intersection ($50\times50\mu m$) the energy of all events are registered. Then an image is produced where the intensity value of each pixel is the number of registered events/hits.

1.2 Applications of the Biomolex technology

The two main applications of the Biomolex technology considered in this note are protein kinase analysis and tissue autoradiography.

Protein Kinase Analysis - Array images Kinase arrays are used to analyze protein and peptide phosphorylations, a reversible modification of a protein that serves to modify / regulate protein activity. The function of these post-translational modifications is to alter the substrate's activity, subcellular localization, binding properties, and/or associa-

tion with other proteins.

Phosphorylation is a ubiquitous regulatory mechanism found in both eukaryotes and prokaryotes. These mechanisms are regulated by protein kinases and phosphatases for adding and removing phosphate groups respectively. The reversible phosphorylation of proteins can result in the activation or termination of many important cellular events including cell signaling, growth, and differentiation. An example of an array image is shown in Figure 2.

Tissue autoradiography - Tissue images Microscopic imaging and dissymmetry of antibody distribution in tumor and tissues in time and space are essential in order to verify pharmacokinetic modeling and elucidate internalization rates. Detailed imaging with high sensitivity and high spatial resolution are crucial to get accurate knowledge of transport and the uptake of labeled biomolecules.

Tissue imaging opens for a wider use of isotopes, thus making it possible to exploit the underlying capabilities, provided by the technology, of using multiple isotopes during a single acquisition. Since the technology is able to detect the accurate amount of energy each particle deposits in the detector, an exact energy spectrum can be defined for any given radioactive isotope. An example of a tissue image is shown in Figure 3.

1.3 Challenges with the Biomolex technology

The production of DSSDs is a very complicated process and production of flawless sensors is practically impossible. One challenge is then that some strips are defect, i.e. they do not register any signal. This results in a corresponding missing stripe in the image, i.e. a stripe where all pixel values are missing. Besides, for some of the missing stripes the signals in the neighbor stripes will be too high, because they register signal that should have been registered by the missing stripe. Also, the signal strengths of the stripes may vary. A method for restoring images for reducing these problems will be proposed and tested in Section 2.

An important issue when analyzing array images is to estimate the total number of hits in each spot in the image. A method for estimating such spot sizes will be proposed in Section 3. Standard methods for estimating spot and background intensities used in microarray imaging cannot be used because the spots consist of fewer pixels and events/hits.

The Biomolex technology opens up for the possibility of using multiple isotopes during a single image acquisition. A challenge is then to decompose the measured signal into signals for the different isotopes, i.e. decide which fractions of the signals/hits that come from which isotope. In Section 4 a method for decomposing signals will be proposed and tested on simulated data. The method is based on using information about the energy spectra and decay times for each of the radioactive isotopes.

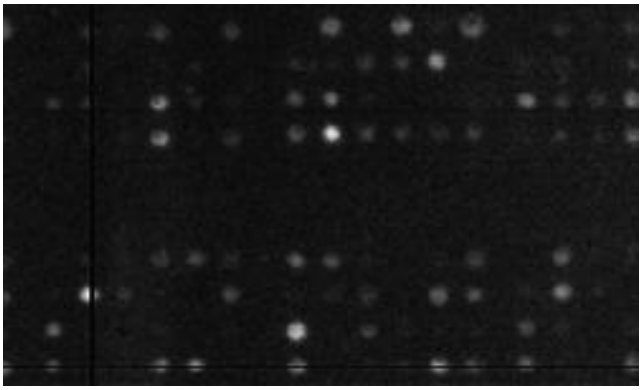


Figure 2. *Part of an array image.*

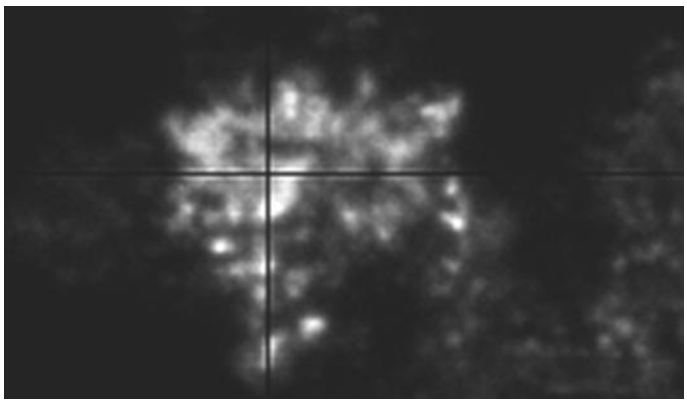


Figure 3. *Part of a tissue image.*

2 Missing stripes

For images acquired using the silicon strip detector there is a problem that some stripes are missing signal, i.e. no hits are registered in that stripe. This is caused by defect strips in the detector. There are two main types of such missing stripes. In both cases the pixel values of the stripe will be zero. First, if the strip in the detector is broken it will not absorb any electrons. The electrons will then be absorbed by one of the neighbor strips instead. This will result in too many registered hits in these strips, i.e. the corresponding stripes in the images will become too bright. Second, if the strip is not broken, but the read out channel of the ASIC is defect, the strip will absorb electrons, but the hits will not be registered. In this case the missing stripe will not influence the pixel values of the neighbor stripes.

Besides defect strips, there is a problem that some strips are overactive. Such strips are registering hits continuously even if there are no hits close to the strip. This results in high pixel values in the corresponding stripe in the image and the values in the stripe will be more or less proportional to the projection of the image onto the stripe. Such overactive stripes are often, but not always, neighbors of missing stripes.

One source of systematic noise in the images is the somewhat varying detection capability of the different strips. This leads to varying signal strengths of the stripes in the image.

2.1 Method for restoring the tissue images

In this section we describe the algorithm for restoring the image, including imputation of values in the missing and overactive stripes. First the missing and overactive stripes are identified (see Section 2.1.1). As the overactive stripes contain no information about the true signal of the stripes, the pixel values in these stripes are set to missing. Then missing values are imputed/estimated (see Section 2.1.2). Besides imputing values in a missing stripe, the pixel values in its neighbor stripes are adjusted to reduce the effect of potential extra signal that should have been registered in the missing stripe and not in its neighbors. Finally, a smoothing method is used to reduce the varying signal strengths of the strips.

The described method assumes that not more than two consecutive stripes are missing or overactive. The proposed algorithm for imputing missing values is based on the assumption that the image is quite smooth, i.e. that it contains quite large objects and that the true signal of two neighbor stripes are quite similar. This is also an important assumption when stripes are adjusted to reduce varying signal strength or to reduce the effect of extra signal from a neighbor stripe that is missing. Tissue images are normally quite smooth (see Figure 3), while this is not the case for array images. Figure 4 shows a part of an array image and two single spots, one with and one without a missing stripe. We observe that the objects consist of only a few pixels and that the pixel values have a large variation from stripe to stripe in the spots. The method for restoring images described in this section is therefore meant for restoring tissue images only.

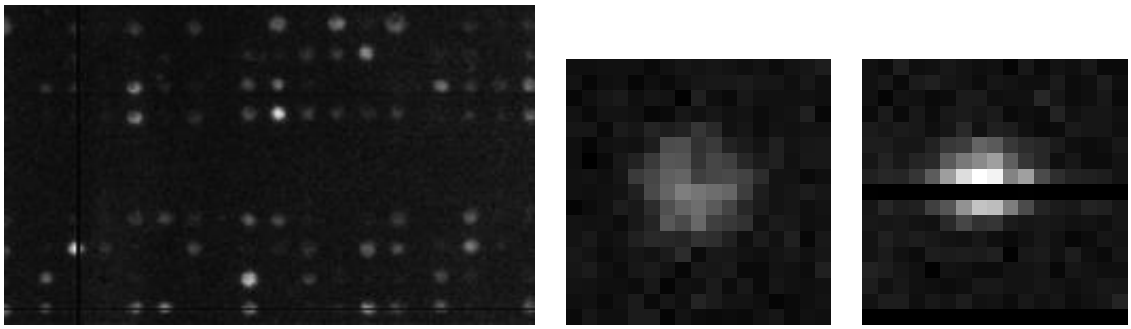


Figure 4. *Part of an array image and two single spots, one without and one with a missing stripe.*

2.1.1 Identifying missing and overactive stripes

The missing and overactive stripes are identified from the image of interest.

A missing stripe is a stripe in the image that corresponds to a defect strip in the detector. In such a strip, no hits will be registered, i.e. the corresponding pixel values will be zero. Because of inaccuracies in the detector, a few pixels might nevertheless be different from zero. A stripe s in the image will be identified as being a missing stripe if almost all pixel values are zero and it in addition is very different from at least one of its two nearest neighbors, stripes $s-1$ and $s+1$ (see Figure 5).

	Pixel values					
Stripe number						Projection value
$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
$s-2$	$\dots$	$a_{s-2,t-1}$	$a_{s-2,t}$	$a_{s-2,t+1}$	$\dots$	p_{s-2}
$s-1$	$\dots$	$a_{s-1,t-1}$	$a_{s-1,t}$	$a_{s-1,t+1}$	$\dots$	p_{s-1}
s	$\dots$	$a_{s,t-1}$	$a_{s,t}$	$a_{s,t+1}$	$\dots$	p_s
$s+1$	$\dots$	$a_{s+1,t-1}$	$a_{s+1,t}$	$a_{s+1,t+1}$	$\dots$	p_{s+1}
$s+2$	$\dots$	$a_{s+2,t-1}$	$a_{s+2,t}$	$a_{s+2,t+1}$	$\dots$	p_{s+2}
$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
Stripe number	$\dots$	$t-1$	t	$t+1$	$\dots$	

Figure 5. *An image indicating stripe s and its neighbors stripes $s-2$, $s-1$, $s+1$, $s+2$, etc. The a's are pixel values. If stripe s is a missing or overactive stripe, one or more of stripes $s-2$, $s-1$, $s+1$ and $s+2$ will be used for imputing data in stripe s. The projection value for a horizontal stripe s is the sum of the pixel values in that stripe, i.e. $p_s = \sum_{t=1}^{t=T} a_{s,t}$, where T is the number of vertical stripes. The projection of the vertical stripes are similarly defined. If $a_{s,t}$ occurs in the missing or overactive stripe s and missing values have not yet been imputed, we assume $a_{s,t} = 0$. S is the number of horizontal stripes.*

A property of an overactive stripe is that it will be more or less proportional to the

projection of the image onto the stripe. This means that if the vertical stripe t in Figure 5 is overactive we expect its pixel values $(a_{1,t}, a_{2,t}, \ldots, a_{S-1,t}, a_{S,t})$ to be highly correlated with the projection of the image $(p_1, p_2, \ldots, p_{S-1}, p_S)$. In practice a stripe t is identified as an overactive stripe if its correlation with the projected image is high (> 0.5) and either the correlation of the neighbor stripes $t-1$ and $t+1$ with the projected image is low (< 0.2), or the projection of stripe t is at least twice that of stripe $t-1$ and stripe $t+1$.

2.1.2 Imputing missing values

If stripe s is a missing stripe, the pixel values in stripe s are estimated from the pixel values in its neighbor stripes. Before imputing the missing values, the pixel values in its neighbor stripes are adjusted too reduce the effect of extra signal that should have been registered in the missing stripe. This is done by multiplying the pixels in stripe $s-1$ and $s+1$ by $\frac{p_{s-2}}{p_{s-1}}$ and $\frac{p_{s+2}}{p_{s+1}}$, respectively (see Figure 5). When pixel values in the neighbor stripes have been adjusted, the pixel values in stripe s are estimated from the pixel values in its neighbor stripes as explained in Figure 6.

	New value for $a_{s,t}, t = 1, \ldots, T$
Stripe $s+1$ not missing Stripe $s-1$ not missing	$\frac{1}{2}a_{s-1,t} + \frac{1}{2}a_{s+1,t}$
Stripe $s+1$ missing Stripe $s-1$ not missing Stripe $s+2$ not missing	$\frac{2}{3}a_{s-1,t} + \frac{1}{3}a_{s+2,t}$
Stripe $s+1$ missing Stripe $s-1$ not missing Stripe $s+2$ missing	$a_{s-1,t}$
Stripe $s-1$ missing Stripe $s+1$ not missing Stripe $s-2$ not missing	$\frac{2}{3}a_{s+1,t} + \frac{1}{3}a_{s-2,t}$
Stripe $s-1$ missing Stripe $s+1$ not missing Stripe $s-2$ missing	$a_{s+1,t}$
Stripe $s-1$ missing Stripe $s+1$ missing	No new value computed Pixel value set to 0

Figure 6. *Formulas for imputing values in a missing stripe s. See Figure 5 for definition of the a's etc. In the formulas information from a larger neighborhood might also have been used. Each $a_{j,t}$ in the formulas could for example have been substituted by the average of $a_{j,t-1}$, $a_{j,t}$ and $a_{j,t+1}$.*

2.1.3 Reducing varying signal strength

One source of systematic noise in the images is the somewhat varying detection capability of the different strips. This leads to varying signal strengths of the stripes in the image. To reduce this systematic noise, we should adjust pixel values in the image of interest such that the signal strengths of the stripes become more similar. It is difficult to obtain realistic estimates of the strips detection capability, i.e. of the corresponding stripes signal strength, using only information in the image itself. However, we might do some smoothing of the image such that the signal strengths become more similar. This is done under the assumption that neighbor stripes should have approximately the same projection values. This they will have because of the assumption of quite smooth images. The smoothing is done by multiplying the pixels in each stripe s with a constant c_s that is such that a smoothing of the projections of the stripes is obtained. More precisely $c_s = \frac{m_s}{p_s}$, where m_s is the mean of the projection values in a neighborhood around stripe s, i.e. the $2n + 1$ values p_{s-i}, p_{s+i},where $i = 0, 1, \ldots, n$ (see Figure 5). After multiplying the pixels in stripe s with c_s, m_s will be the new projection value of stripe s.

2.1.4 Using information from calibration experiments

In the method described above only information from the image itself is used when restoring an image. An alternative that might strengthen the results could be to include information from a calibration experiment that is such that the concentration of isotopes is constant over the image. An assumption for using such information is that the strips in the detector have the same characteristics during the acquisition of the calibration image and the image of interest when it comes to defect strips, overactive strips and detection capabilities.

In a calibration experiment we expect the projection values to be quite similar (see Figure 5). Missing and overactive stripes will be outliers. A stripe s might be identified as an overactive stripe if its projection p_s is much larger than the median projection value, while it might be identified as a missing stripe if its projection p_s is close to zero and is much smaller than the median projection value. Varying signal strengths might be reduced as described in Section 2.1.3, but where m_s and p_s are computed from the calibration image and $m = m_s$ is the mean of all projection values $(p_1, \ldots p_S)$ that are not outliers.

The indicated approaches have not been implemented and tested as calibration experiments were not available.

2.2 Results

The method described in Sections 2.1.1, 2.1.2 and 2.1.3 has been tested on several tissue images and has also been implemented in the Biomolex system. The results of the tests will be described and discussed in this section. The images in the test set consisted of the images shown in Figures 7-11 and were chosen to illustrate different challenges with restoring tissue images acquired with the Biomolex detector.

We conclude that for all images in the test set, missing and overactive stripes were identified. The images presented in Figures 7 and 8 are satisfactorily restored both when

it comes to imputing missing values and when it comes to removing the extra signal in the neighbors of the missing stripes. For the image in Figure 8, with zoomed subimages in Figure 9, we observe that use of smoothing reduces the problem of varying signal strength between stripes. The images in Figures 10 and 11 contain no missing stripes,

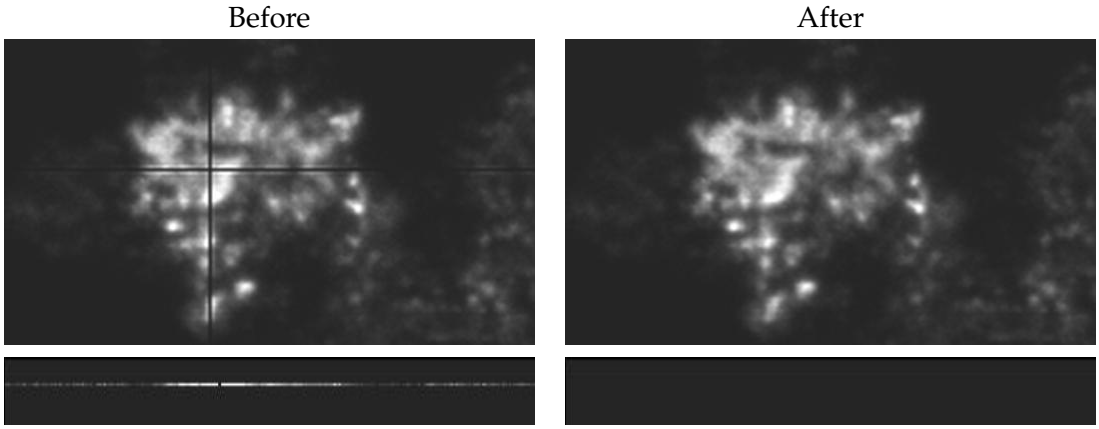


Figure 7. *Parts of a tissue image before and after image restoration. The upper images contain a large object in the image. There is one missing stripe in x-direction and one in y-direction. The neighbors of the missing stripes have too high signal. The lower images are a part of the image without an object, i.e. with very low signal, except that it contains an overactive stripe.*

but both images have several stripes that seems to have a too low signal strength. We observe that use of smoothing reduces this problem even though it has not disappeared completely. One possible way of reducing the problem with weak stripes even more is to set such stripes to missing and then impute values using information from the neighbour stripes. This has been tested for the image in Figure 11 and the results of using the missing stripes approach are shown in the right column in that figure. We observe that setting one of the weakest stripes to missing reduces the problem with varying signal strength even more. The problem has still not disappeared completely because there are several consecutive stripes that are missing, and the method for restoring the images assumes that not more than two consecutive stripes are missing or overactive.

To conclude, the proposed method gives satisfactory results for the test set. It is a promising method for restoring tissue images that are quite smooth and that do not contain more than two consecutive stripes that are missing or overactive or that have too low signal strength. Improved results might have been obtained by including a calibration experiment.

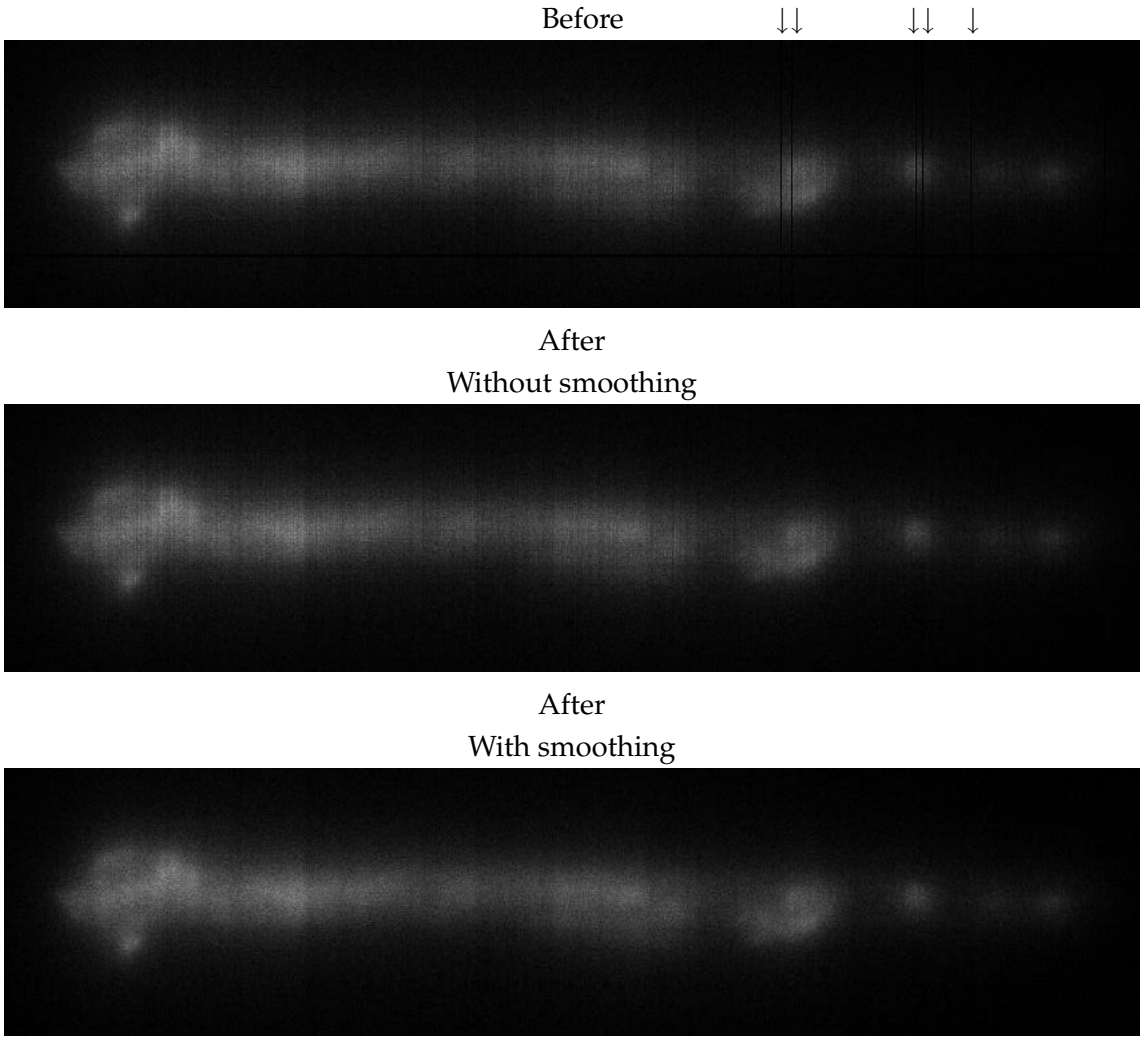


Figure 8. *Image of a rat arterie. There are five vertical missing stripes in the image (indicated by arrows), and one horizontal missing stripe below the arteria (not easy to see as it sourroundings are quite dark). Also, the signal strength of the stripes is varying.*

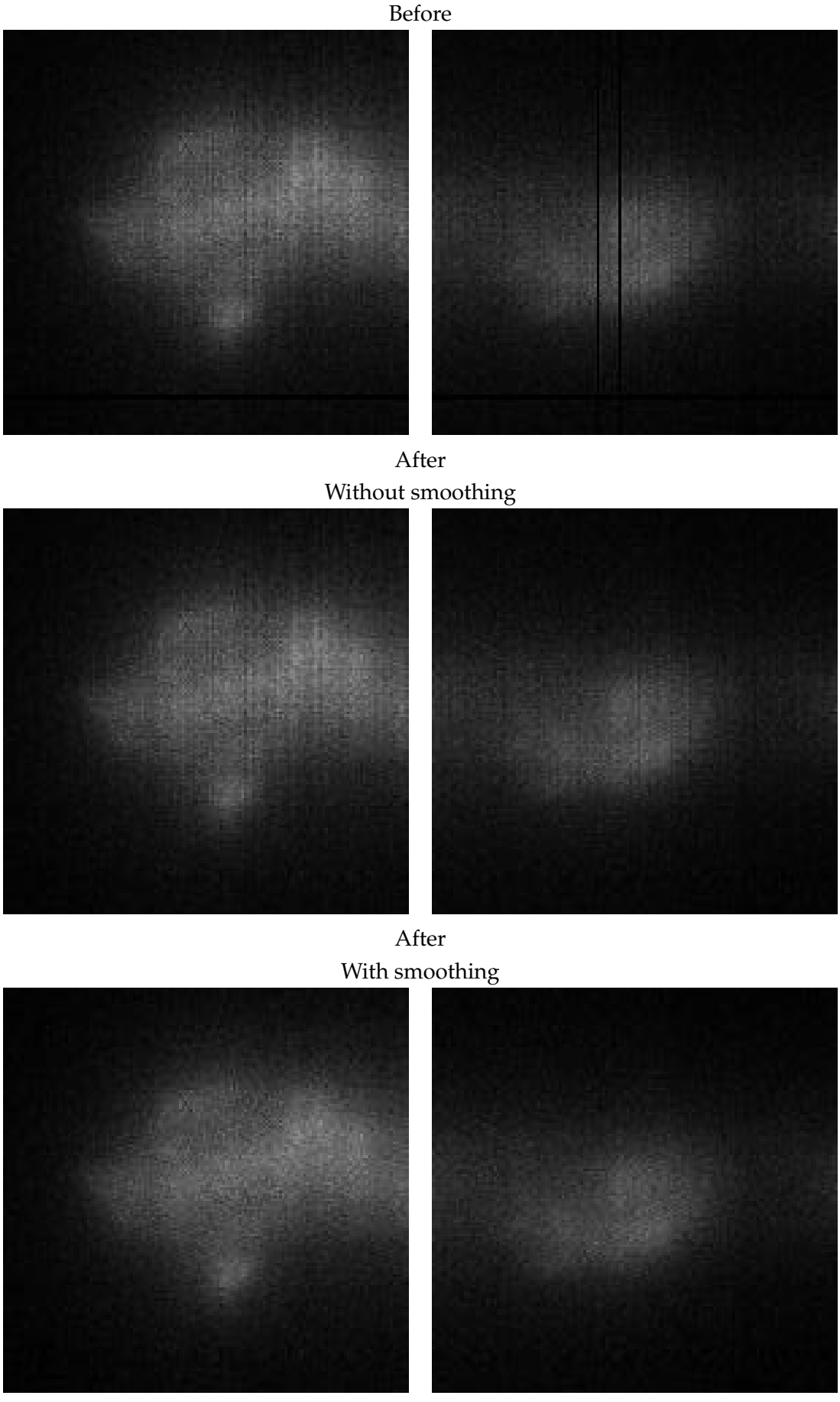


Figure 9. *Two zoomed parts of the images in Figure 8.*

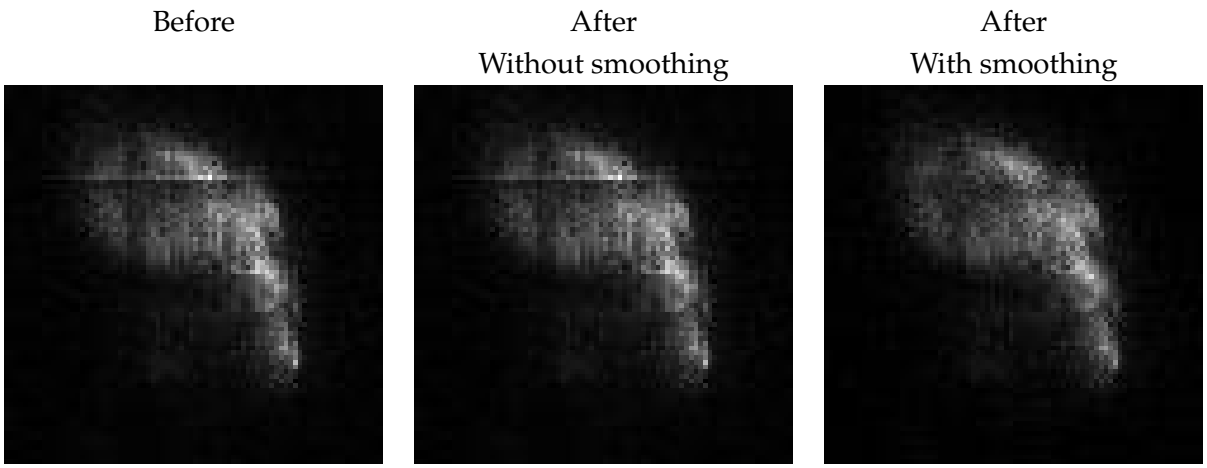


Figure 10. *Part of an image of tissue from a mouse. There is one overactive stripe in the image, and several stripes that seems to have a too low signal strength.*

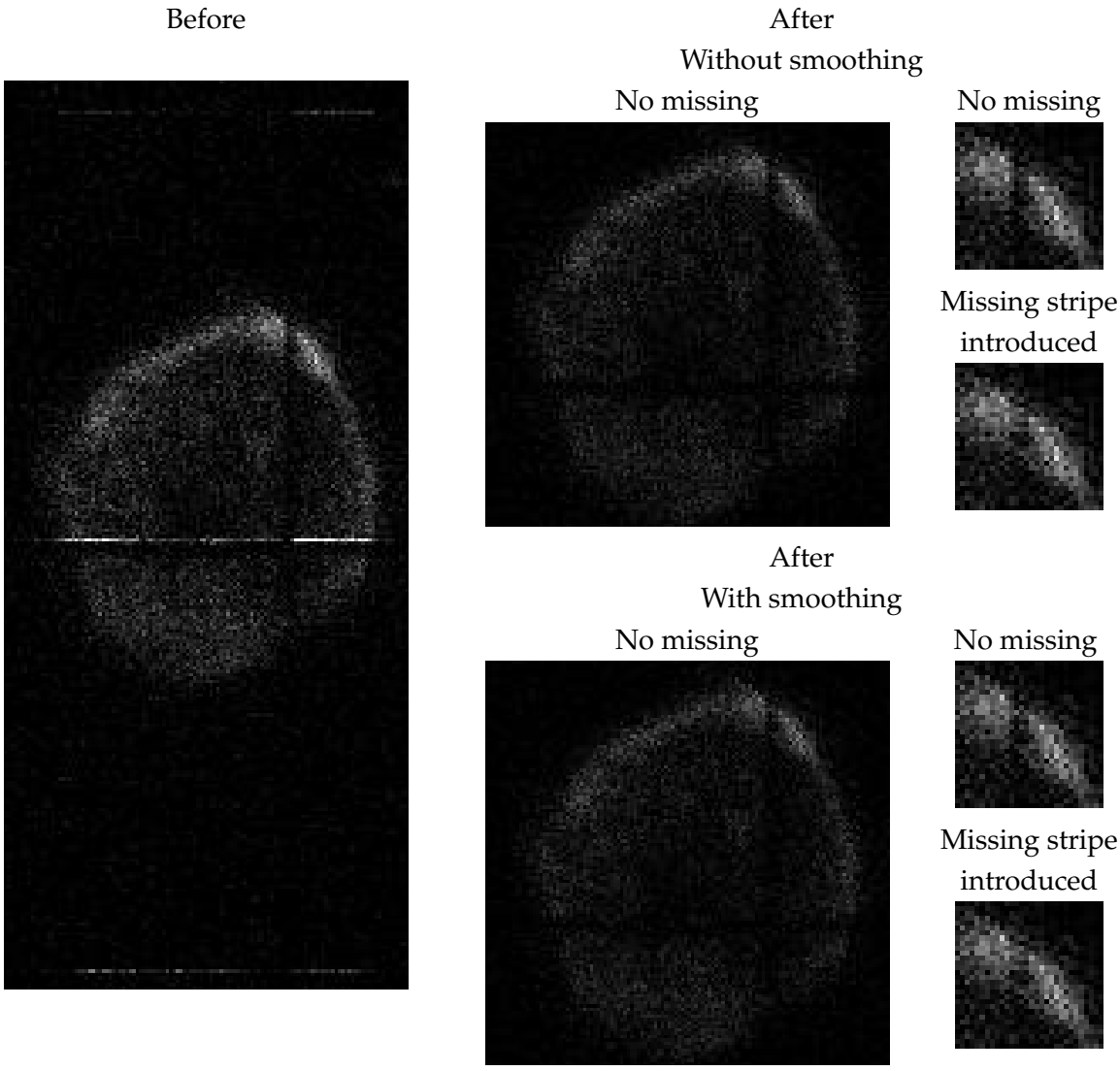


Figure 11. *Parts of a tissue image with three overactive stripes. Several stripes seem to have a too low signal strength. Also, there is a problem that the values in the object are very low. To reduce this problem the original image was multiplied by 7.5 before the image was restored. The left column show the largest part of the image, including all the three overactive stripes. The middle column show the main object after the image has been restored. The images in the right column are zooms of the upper right part of the main object. In the "Missing stripe introduced" images one of the weak stripes has artificially been set to missing before the image has been restored.*

3 Estimation of spot size in array images

An important issue when analyzing array images is to estimate the total number of hits in each spot, here denoted *spot size*. More precisely this is the assessment of the number of hits (β-particles) arising from each biological sample on an array.

A method for estimating spot size when the spots are well separated is described in Section 3.1, while the method is extended to overlapping spots in Section 3.2. In both cases it is assumed that the the grid and the spot centers have been identified. A method for doing this has already been implemented in the Biomolex system, while the method for estimating spot size will be implemented later.

3.1 Well separated spots

Two factors complicate the estimation:

- Hits arising from background sources - these should be subtracted.
- Missing or atypical stripes.

These problems imply that straightforward counting of the number of hits within the area corresponding to the biological sample may give misleading results.

One possible solution is to fit a two-dimensional surface to the pixel values. Suppose the hits from the specific spot under consideration are found within a rectangular area with pixel values $x_{ij}, i = 1, \ldots, N, j = 1, \ldots, M$, and that no other spot (significantly) affects this area. Over the area we fit a function (surface) of the form

$$B + Sf(i, j; \Theta) \tag{1}$$

where f is a two-dimensional distribution function, e.g., the two-dimensional Gaussian distribution. $f(i, j; \Theta)$ is then understood as the function evaluated at the pixel centers, B represent the background, while S will represent the size of the spot (since f integrates to 1).

If we use a Gaussian distribution for f, that is,

$$f(x, y; \Theta) = \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right)}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \tag{2}$$

the unknown parameters in Θ will be the location of the spot center (μ_x, μ_y), the standard deviations σ_x and σ_y and the correlation ρ. Together with B and S this adds to 7 unknown parameters. These parameters may be estimated by minimization of

$$\sum_{i,j \subset \Omega} (x_{ij} - f(i, j; \Theta))^2 \tag{3}$$

over Θ. Here, Ω represents the non-missing pixels (the missing stripes thus 'disappear' by simply not being included in the calculations).

In our tests, the minimizations were carried out using the SAS system procedure NLIN with the Marquart minimization algorithm (a modified Gauss-Newton method).

Explicit values for the derivatives for S and B were used, while the algorithm approximated the other derivatives. When using the algorithm, specific starting values must be given for each spot center, while for the other parameters default starting values seem to work well. Figure 12 shows examples of original spot images and model fits. The up-

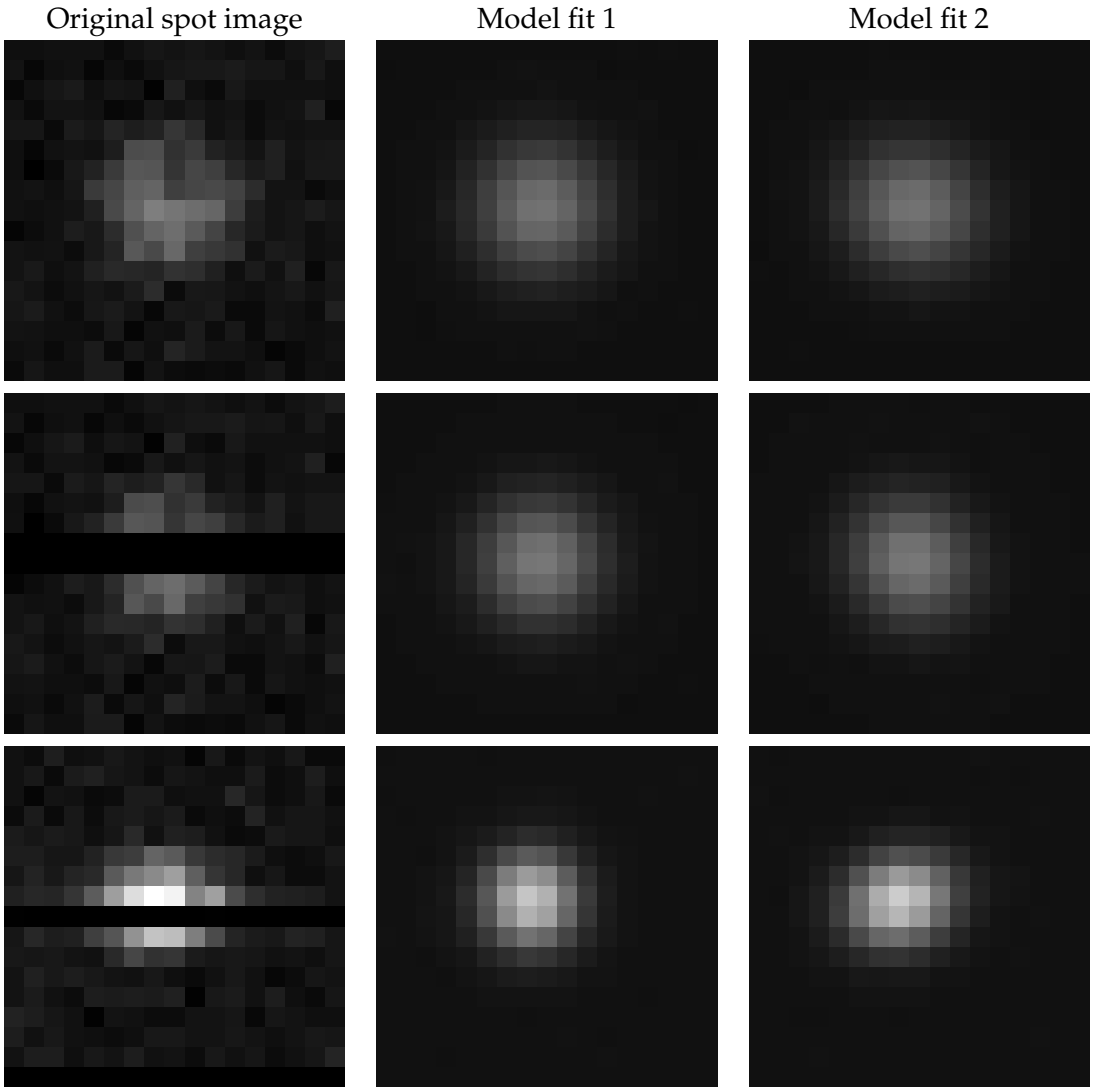


Figure 12. *Examples of original spot images and model fits. The spot in the upper row has no missing stripes. The second row is the spot in the upper row where two rows are set to missing. The lower row has a real missing stripe. Also its two nearest neighbors have been set to missing when fitting the model. In model fit 1 f is equal to a Gaussian with correlation=0 and equal variances. Model fit 2 is as model fit 1, but without restrictions on the correlation and variances.*

per row shows to the left an original spot image (this image has no missing stripes), the middle panel shows an approximation obtained by fitting (1) with f equal to a Gaussian with correlation=0 and equal variances, and the right panel the resulting surface using a Gaussian without restrictions on the correlation and variances. The second row shows the same spot, but two rows have artificially been set to missing. Note that the fit only changes marginally (the estimate of the spot size changes from 1272 to 1292). The lower

row of Figure 12 shows an example with a real missing stripe, the corresponding row and its nearest neighbors (with too high values) have been set to missing in the estimations shown in the middle and right panels.

Tests have been carried out on about 20 spots, partially with (artificially) missing rows or columns. The algorithm seems robust, provided reasonable starting values are provided for the center positions. However, before used in commercial systems, more extensive tests on images of varying quality should be carried out.

3.2 Overlapping spots

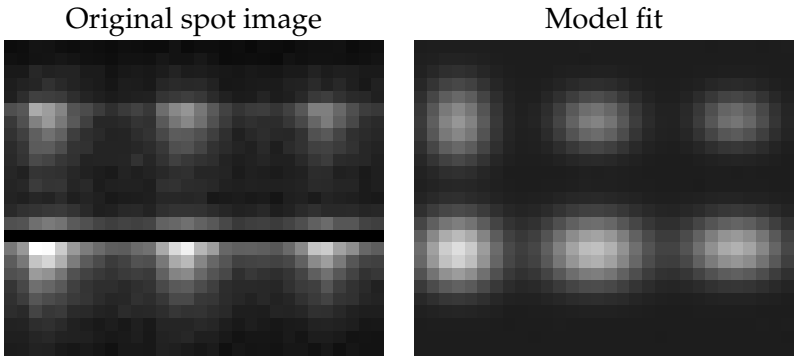


Figure 13. *An example of overlapping spots.*

Using isotopes that emit high-energy particles, like ^{32}P, the β-particles from one biological sample may be counted in areas corresponding to other samples. Thus a pattern of overlapping spots may emerge as illustrated in Figure 13. We have carried out preliminary tests to see if reliable estimation is still possible. The idea is to fit a set of distributions simultaneously to a spot and its nearest neighbors. Figure 13 shows an example of such a fit. Technically, the procedure seems to work well, and despite a fairly high number of parameters it only uses a few cpu-seconds. Use in real situations will, however, require solutions to other problems. Most importantly, the 'background' seems to be influenced by particles that move over long distances (several spot diameters), possibly in the supporting glass structures on the chip. It will then obviously be challenging to assess weak spots situated in an area with strong spots. While it is probably hard to solve this through a model based on physical principles, ad hoc solutions may be possible. One possible partial solution may be to utilize the energy of the hits, but we have not yet evaluated this options.

If the problems connected to overlapping spots can be overcome, an interesting perspective is the use of two radioactive emitters, one with high energies (contributing to a broad spot) and one with low energies (with focused hits in the central area). This may enable separation of the signals by fitting two distributions with the same center, and supporting the estimation by using the energies in the hits.

3.3 Remarks

In the method described in this section we assume that overactive and missing stripes already have been identified and that it already has been decided whether the neighbors of these stripes have too high signal. Besides we assume that varying signal strength is not a problem.

A slightly modified version of the algorithm described in Section 2.1.1 might have been used for identifying missing and overactive stripes. For identifying neighbors of missing stripes with too high signal we might either use the parts of the image that are outside the grid with spots or we might use information from a calibration experiment. This might also be the solution for reducing the problem with varying signal strength.

4 Separation of signals in tissues

It is of great interest to be able to generate images of the separate distributions of two (or more) different isotopes present in the same tissue sample. This allows to study different processes taking place in the same tissue simultaneously, for example comparing uptake patterns of different radiopharmaceuticals in the same tissue sample. In order to do this, it is crucial to be able to separate the contributions from the two isotopes. The underlying energy spectra (distribution of energies) for the two isotopes are considered known. In each pixel, we need to decompose the observed mixed energy spectrum into a linear combination of the two underlying spectra. Below, we summarize some previous work on this problem, and suggest a simpler and faster method. The new method has been tested on simulated data. Unfortunately, no real data were available.

4.1 Notation

Known probabilities. The energy spectra for isotope 1 and isotope 2 are known. The energy range is divided into m bins. Let $h_i^{(1)}$ be the probability that a hit from isotope 1 has energy in bin i. Similarly, let $h_i^{(2)}$ be the probability that a hit from isotope 2 has energy in bin i. We then have $\sum_{i=1}^m h_i^{(1)} = \sum_{i=1}^m h_i^{(2)} = 1$.

Data/observations. Hits are counted for one pixel. Let n be the total number of hits in the pixel and n_i the number of hits in energy bin i, $i = 1, \ldots, m$, i.e. $\sum_{i=1}^m n_i = n$.

From the data $n_1, n_2, \ldots, n_m$ and the probabilities (spectra) $h_i^{(1)}$ and $h_i^{(2)}$, $i = 1, \ldots, m$, we want to estimate the contribution from each isotope.

4.2 Previous work

The methods currently available for the Biomolex system are described in a paper by Kvinnsland and Skretting[1]. The aim of the two methods in that paper is to find estimates of $\lambda^{(1)}$ and $\lambda^{(2)}$, where $\lambda^{(1)}$ and $\lambda^{(2)}$ are the (expected) number of hits originating from isotope 1 and 2, respectively.

Method based on the least square method. The first method described in the paper by Kvinnsland and Skretting is based on the the least square (LS) method. Estimates of $\lambda^{(1)}$ and $\lambda^{(2)}$ are found by minimizing the expression

$$S = \sum_{i=1}^{m} (n_i - \lambda^{(1)} h_i^{(1)} - \lambda^{(2)} h_i^{(2)})^2 \tag{4}$$

This minimization might be done analytically.

Method based on a maximum likelihood approach. The second method described in the paper by Kvinnsland and Skretting is based on a maximum likelihood (ML) approach and models the probability of measuring b hits with energies $e_1, \cdots, e_b$, as follows:

$$p(e_1, \cdots, e_b | \lambda^{(1)}, \lambda^{(2)}) = \sum_{l=0}^{b} p_1(l, b-l | \lambda^{(1)}, \lambda^{(2)}) \cdot p_2(e_1, \cdots, e_b | l, b-l) \tag{5}$$

1. Y. Kvinnsland and A. Skretting. Methods for separation of contributions from two radionuclides in autoradiography with a silicon strip detector. Phys. Med. Biol., 2000, 45, 1183–1193

where

- $p_1(l, b-l|\lambda^{(1)}, \lambda^{(2)})$ is the probability that l hits originated from isotope 1 and $b-l$ from isotope 2 given the expectation values $\lambda^{(1)}$ and $\lambda^{(2)}$ and

- $p_2(e_1, \cdots, e_b|l, b-l)$ is the probability of measuring the energies $e_1, \cdots, e_b$ given l hits originating from isotope 1 and $b-l$ from isotope 2.

This model can incorporate temporal information. However, the likelihood and the maximization of the likelihood is computationally demanding even without this complication. The LS and ML methods above have been tested on simulated data and on a small real image with three drops, one with pure isotope 1 (here ^{35}S), one with pure isotope 2 (here ^{33}P) and one drop with a mixture of the two. The results are reasonably good, but the ML method is very computationally intensive, and does not show substantial improvements compared to the simple LS method.

With time information incorporated, the ML method has been tested with only two time windows, due to computational cost, and with no improvement in estimates.

4.3 Alternative method

Here we present an alternative to the two methods presented in the paper by Kvinnsland and Skretting. First, we reformulate the description of the mixing of the two isotopes. Instead of $\lambda^{(1)}$ and $\lambda^{(2)}$, our unknown will be the fraction a defined by:

Of the n observed hits, a fraction a is from isotope 1
and a fraction $(1-a)$ is from isotope 2.

Hence our aim is simplified to estimating one parameter a. Let h_i be the probability that a hit from a mixture of isotope 1 and 2 has energy in bin i. We can write

$$\begin{aligned} h_i &= P(\text{energy in bin } i) \\ &= P(\text{energy in bin } i|\text{from isotope 1}) \cdot P(\text{isotope 1}) + \\ &\quad P(\text{energy in bin } i|\text{from isotope 2}) \cdot P(\text{isotope 2}) \\ &= a \cdot h_i^{(1)} + (1-a) \cdot h_i^{(2)} \end{aligned} \tag{6}$$

where a is defined above. We wish to find an estimate of a.

Given n, the bin numbers for the energies $e_1, \ldots, e_n$ can be considered as a series of n independent trials. The outcome of each trial is one of m bin numbers. The random variables $N_1, \ldots N_m$, where N_i = the total number of n trials where the outcome is bin i, have a multinomial distribution, i.e.

$$P(N_1 = n_1, N_2 = n_2, \ldots, N_m = n_m) = \frac{n!}{n_1! n_2! \cdots n_m!} h_1^{n_1} h_2^{n_2} \cdots h_m^{n_m} \tag{7}$$

Assuming the model given in 7, an estimate of a could be found using the maximum likelihood approach. Maximizing the likelihood $P(a|N_1 = n_1, N_2 = n_2, \ldots, N_m = n_m)$ is

equivalent to maximizing the expression

$$\begin{aligned} L(a) &= log(\prod_{i=1}^{m} h_i^{n_i}) = \sum_{i=1}^{m} n_i \cdot log(h_i) \\ &= \sum_{i=1}^{m} n_i \cdot log(a \cdot h_i^{(1)} + (1-a) \cdot h_i^{(2)}). \end{aligned} \tag{8}$$

The value that maximizes the log-likelihood, $\hat{a}$, is found numerically using an iterative optimization method.

4.3.1 Maximization of the likelihood

An estimate of a that maximizes the likelihood is based on the derivative of L with respect to a,

$$\begin{aligned} L'(a) &= \sum_{i=1}^{m} \frac{n_i \cdot (h_i^{(1)} - h_i^{(2)})}{a \cdot h_i^{(1)} + (1-a) \cdot h_i^{(2)}} \\ &= \sum_{i=1}^{m} \frac{n_i \cdot (h_i^{(1)} - h_i^{(2)})}{a \cdot (h_i^{(1)} - h_i^{(2)}) + h_i^{(2)}}. \end{aligned} \tag{9}$$

We might assume that $n_i \neq 0$ and $h_i^{(1)} \neq h_i^{(2)}$ as such terms otherwise could have been removed from the expression to be maximized. We might also assume that the bin numbers are sorted such that $h_i^{(1)} > h_i^{(2)}$ for $i \leq m_1$ and $h_i^{(1)} < h_i^{(2)}$ for $i > m_1$ for some number m_1, $1 \leq m_1 \leq m$. We then rewrite

$$L'(a) = \sum_{i=1}^{m_1} \frac{n_i \cdot (h_i^{(1)} - h_i^{(2)})}{a \cdot (h_i^{(1)} - h_i^{(2)}) + h_i^{(2)}} - \sum_{i=m_1+1}^{m} \frac{n_i \cdot (h_i^{(2)} - h_i^{(1)})}{a \cdot (h_i^{(1)} - h_i^{(2)}) + h_i^{(2)}}. \tag{10}$$

We observe that both the numerator and the denominator in all terms in the two sums are positive. If a increases, the denominators in the first sum increases and the denominators in the second sum decreases. This means that if a increases, the first sum will decrease and the second sum will increase. Then $L'(a)$ will decrease. Similarly, if a decreases, $L'(a)$, will increase. From this we conclude that:

$$\begin{aligned} a_1 < a_2 &\Rightarrow L'(a_1) > L'(a_2) \\ &\Rightarrow L'(0) > L'(1) \\ L'(0) < 0 &\Rightarrow L'(a) < 0 \text{ for all } a \\ &\Rightarrow a_{max} = 0 \text{ maximizes the likelihood } L(a) \\ L'(1) > 0 &\Rightarrow L'(a) > 0 \text{ for all } a \\ &\Rightarrow a_{max} = 1 \text{ maximizes the likelihood } L(a) \\ L'(0) > 0 > L'(1) &\Rightarrow \text{ there exists an } a_{max}, 0 < a < 1, \text{ such that } L'(a) = 0 \\ &\Rightarrow a_{max} \text{ maximizes the likelihood } L(a). \end{aligned} \tag{11}$$

Algorithm The implications above will be used in the following algorithm for maximizing the likelihood. In the algorithm, aM is the current estimate of a_{max} and aS is the start estimate of a_{max}. Also, $[i_1, i_2]$ is the current interval in which we search for a_{max} and this interval is such that $L'(i_1) > 0 > L'(i_2)$.

```
< compute L'(0) and L'(1) >
if (< L'(0) < 0 >)
   aM = 0
else if (< L'(1) > 0 >)
   aM = 1
else {
  < compute aS >
  aM = aS
  i1 = 0
  i2 = 1
  while (< i1 is not approximately equal to i2 >) {
    if ( L'(aM) > 0 )
      i1 = aM
    else
      i2 = aM
    < choose a new value for aM in the interval [i1,i2] >
  }
}
```

The new value chosen in the interval $[i_1, i_2]$ could either be chosen in the middle of the interval, drawn uniformly from the interval, chosen closest to the interval limit with the smallest derivative or by some other algorithm. A new value closest to the interval limit with the smallest derivative could for example be chosen equal to

$$i_1 - \frac{i_2 - i_1}{L'(i_2) - L'(i_1)} \cdot L'(i_1) \tag{12}$$

This corresponds to assuming that the derivative is linear. The start value for the iterative algorithm could either be chosen in the interval $[0, 1]$ in the same way as new values are chosen, or it might be found using a least square approach and minimizing

$$S = \sum_{i=1}^{m} (n_i - a \cdot n \cdot h_i^{(1)} - (1 - a) \cdot n \cdot h_i^{(2)})^2 \tag{13}$$

This is equal to Equation 4 where $a \cdot n = \lambda^{(1)}$ and $(1 - a) \cdot n = \lambda^{(2)}$.

All above calculations are intended for one pixel only. If there are enough counts in each pixel, the estimation of a should be pixelwise, in order not to introduce unnecessary smoothing. Of course, it is possible to assume that neighboring pixels have the same fraction of the two isotopes, and hence aggregate counts in several pixels through a moving window f.ex. This assures enough data, but might introduce bias.

4.3.2 Extension to time

The data/observations for each pixel are now obtained within K time windows, $k = 1, \ldots, K$. Let n be the total number of hits in the pixel and $n_{i,k}$ the number of hits in energy bin i and time window k, $i = 1, \ldots, m$ and $k = 1, \ldots, K$, i.e. $\sum_{k=1}^{K} \sum_{i=1}^{m} n_{i,k} = n$. The known probabilities, $h_i^{(1)}$ and $h_i^{(2)}$ and the definition of the unknown fraction a are the same as before.

Let $T_h^{(1)}$ and $T_h^{(2)}$ be the half-lifes of isotope 1 and 2, respectively. Let $p_k^{(1)}$ be the probability that a hit from isotope 1 is observed in time window k. We derive

$$p_k^{(1)} = \frac{c^{(1)} \cdot ln(2)}{T_h^{(1)}} \cdot \int_{t_k}^{t_k+l_k} 2^{-\frac{t}{T_h^{(1)}}} dt = c^{(1)} \cdot 2^{-\frac{t_k}{T_h^{(1)}}} \cdot (1 - 2^{-\frac{l_k}{T_h^{(1)}}}), \tag{14}$$

where t_k is the start time for time window k, l_k is the length of time window k and $c^{(1)}$ is a constant that is such that $\sum_{k=1}^{K} p_k^{(1)} = 1$. Similarly, let $p_k^{(2)}$ be the probability that a hit from isotope 2 is observed in time window k, and derive

$$p_k^{(2)} = c^{(2)} \cdot 2^{-\frac{t_k}{T_h^{(2)}}} \cdot (1 - 2^{-\frac{l_k}{T_h^{(2)}}}). \tag{15}$$

Let $h_{i,k}^{(1)}$ be the probability that a hit from isotope 1 has energy in bin i and is observed in time window k, i=1, $\ldots, m$ and $k = 1, \ldots, K$. Since the energy spectrum does not change over time, $h_{i,k}^{(1)} = h_i^{(1)} \cdot p_k^{(1)}$. $h_{i,k}^{(2)}$ is similarly defined for isotope 2 and $h_{i,k}^{(2)} = h_i^{(2)} \cdot p_k^{(2)}$. We observe that $\sum_{k=1}^{K} \sum_{i=1}^{m} h_{i,k}^{(1)} = \sum_{k=1}^{K} \sum_{i=1}^{m} h_{i,k}^{(2)} = 1$.

Let $h_{i,k}$ be the probability that a hit from a mixture of isotope 1 and 2 has energy in bin i and is observed in time window k. We can write

$$\begin{aligned} h_{i,k} &= P(\text{energy in bin } i \text{ and time window } k) \qquad (16)\\ &= P(\text{energy in bin } i \text{ and time window } k | \text{from isotope 1}) \cdot P(\text{isotope 1}) + \\ &\quad P(\text{energy in bin } i \text{ and time window } k | \text{from isotope 2}) \cdot P(\text{isotope 2}) \\ &= a \cdot h_{i,k}^{(1)} + (1-a) \cdot h_{i,k}^{(2)} \end{aligned}$$

Given n, the bin numbers and time windows for the energies $e_1, \ldots, e_n$ can, as before, be considered as a series of n independent trials. The outcome of each trial is one of $m \cdot K$ combinations of m bin numbers and K time windows. The $m \cdot K$ random variables $N_{i,k}$, where $N_{i,k}$ = the total number of n trials where the outcome is bin i and time window k, have a multinomial distribution, i.e.

$$\begin{aligned} & P(N_{1,1} = n_{1,1}, \ldots, N_{m,1} = n_{m,1}, \ldots, N_{1,K} = n_{1,K}, \ldots, N_{m,K} = n_{m,K}) \qquad (17)\\ = \; & \frac{n!}{n_{1,1}! \cdots n_{m,1}! \cdots n_{1,K}! \cdots n_{m,K}!} h_{1,1}^{n_{1,1}} \cdots h_{m,1}^{n_{m,1}} \cdots h_{1,K}^{n_{1,K}} \cdots h_{m,K}^{n_{m,K}} \end{aligned}$$

Assuming the model given in 17, an estimate of a could be found using the maximum likelihood approach. Maximizing the likelihood $P(a|N_{1,1} = n_{1,1}, \ldots, N_{m,1} = n_{m,1}, \ldots, N_{1,K} = n_{1,K}, \ldots, N_{m,K} = n_{m,K})$ is equivalent to maximizing the expression

$$\begin{aligned} L(a) &= log(\prod_{k=1}^{K} \prod_{i=1}^{m} h_{i,k}^{n_{i,k}}) = \sum_{k=1}^{K} \sum_{i=1}^{m} n_{i,k} \cdot log(h_{i,k}) \qquad (18)\\ &= \sum_{k=1}^{K} \sum_{i=1}^{m} n_{i,k} \cdot log(a \cdot h_{i,k}^{(1)} + (1-a) \cdot h_{i,k}^{(2)}) \end{aligned}$$

We want to estimate a. As for the case without time, the a-value that maximizes the log-likelihood, $\hat{a}$, is found numerically using an iterative optimization method.

An estimate of a that maximizes the likelihood is based on the derivative of L with respect to a,

$$\begin{aligned} L'(a) &= \sum_{k\in\{1,\ldots,K\},m\in\{1,\ldots,M\}} \frac{n_{i,k}\cdot(h_{i,k}^{(1)}-h_{i,k}^{(2)})}{a\cdot h_{i,k}^{(1)}+(1-a)\cdot h_{i,k}^{(2)}} \\ &= \sum_{k\in\{1,\ldots,K\},m\in\{1,\ldots,M\}} \frac{n_{i,k}\cdot(h_{i,k}^{(1)}-h_{i,k}^{(2)})}{a\cdot(h_{i,k}^{(1)}-h_{i,k}^{(2)})+h_{i,k}^{(2)}}. \end{aligned} \tag{19}$$

We might assume that $n_{i,k}\neq 0$ and $h_{i,k}^{(1)}\neq h_{i,k}^{(2)}$ in all terms as such terms otherwise could have been removed from the expression to be maximized. As before, we then rewrite the expression above by splitting the sum into two sums depending on whether $h_{i,k}^{(1)} > h_{i,k}^{(2)}$ or $h_{i,k}^{(1)} < h_{i,k}^{(2)}$, i.e.:

$$\begin{aligned} L'(a) &= \sum_{k\in\{1,\ldots,K\},m\in\{1,\ldots,M\} \text{ where } h_{i,k}^{(1)}>h_{i,k}^{(2)}} \frac{n_i\cdot(h_{i,k}^{(1)}-h_{i,k}^{(2)})}{a\cdot(h_{i,k}^{(1)}-h_{i,k}^{(2)})+h_{i,k}^{(2)}} - \\ &\quad \sum_{k\in\{1,\ldots,K\},m\in\{1,\ldots,M\} \text{ where } h_{i,k}^{(1)}<h_{i,k}^{(2)}} \frac{n_i\cdot(h_{i,k}^{(2)}-h_{i,k}^{(1)})}{a\cdot(h_{i,k}^{(1)}-h_{i,k}^{(2)})+h_{i,k}^{(2)}} \end{aligned} \tag{20}$$

The arguments for the implications in (11) still hold and the algorithm in Section 4.3.1 might be used for estimating a.

The fraction a is defined by:

Of the n observed hits, a fraction a is from isotope 1 and a fraction $(1-a)$ is from isotope 2.

This is not the fraction of interest. What we want to estimate is the the fraction a_0 at the initial time before the decay of the radioactivity, i.e.

$$a_0 = \frac{R^{(1)}}{R^{(1)}+R^{(2)}} = \frac{1}{1+\frac{R^{(2)}}{R^{(1)}}} \tag{21}$$

where $R^{(1)}$ and $R^{(2)}$ are the initial rate of hits from isotope 1 and 2, respectively. An expression for a_0 as a function of a is found using the following formula for the expected number of hits from isotope 1 in time window k

$$\lambda_k^{(1)} = \int_{t_k}^{t_k+l_k} R^{(1)} 2^{-\frac{t}{T_h^{(1)}}}\,dt = R^{(1)}\frac{\mathrm{T}_h^{(1)}}{\ln 2} 2^{-\frac{t_k}{T_h^{(1)}}}(1-2^{-\frac{l_k}{T_h^{(1)}}}) \tag{22}$$

and the similar formula for isotope 2. We derive

$$\begin{aligned} a &= \frac{\sum_{k=1}^K \lambda_k^{(1)}}{\sum_{k=1}^K \lambda_k^{(1)}+\sum_{k=1}^K \lambda_k^{(2)}} \\ &= \frac{R^{(1)}e^{(1)}}{R^{(1)}e^{(1)}+R^{(2)}e^{(2)}} = \frac{e^{(1)}}{e^{(1)}+\frac{R^{(2)}}{R^{(1)}}e^{(2)}} = \frac{e^{(1)}}{e^{(1)}+(\frac{1}{a_0}-1)e^{(2)}}, \end{aligned} \tag{23}$$

where $e^{(1)} = T_h^{(1)} \sum_{k=1}^{K} 2^{-t_k/T_h^{(1)}}(1-2^{-l_k/T_h^{(1)}})$ and $e^{(2)} = T_h^{(2)} \sum_{k=1}^{K} 2^{-t_k/T_h^{(2)}}(1-2^{-l_k/T_h^{(2)}})$ are known constants. An estimate of a_0 is then found from a using the formula

$$a_0 = \frac{e^{(2)} \cdot a}{e^{(1)} - (e^{(1)} - e^{(2)}) \cdot a}. \tag{24}$$

Remark Assuming that there is no time gap between the time windows, and that $t_1 = 0$, a simpler expression for a as a function of a_0 might be derived as follows:

$$\begin{aligned} a &= \frac{\int_0^T R^{(1)} 2^{-\frac{t}{T_h^{(1)}}} dt}{\int_0^T R^{(1)} 2^{-\frac{t}{T_h^{(1)}}} dt + \int_0^T R^{(2)} 2^{-\frac{t}{T_h^{(2)}}} dt} \\ &= \frac{R^{(1)} T_h^{(1)} (1 - 2^{-\frac{T}{T_h^{(1)}}})}{R^{(1)} T_h^{(1)} (1 - 2^{-\frac{T}{T_h^{(1)}}}) + R^{(2)} T_h^{(2)} (1 - 2^{-\frac{T}{T_h^{(2)}}})} \\ &= \frac{T_h^{(1)} (1 - 2^{-\frac{T}{T_h^{(1)}}})}{T_h^{(1)} (1 - 2^{-\frac{T}{T_h^{(1)}}}) + \frac{R^{(2)}}{R^{(1)}} - T_h^{(2)} (1 - 2^{-\frac{T}{T_h^{(2)}}})} \\ &= \frac{T_h^{(1)} (1 - 2^{-\frac{T}{T_h^{(1)}}})}{T_h^{(1)} (1 - 2^{-\frac{T}{T_h^{(1)}}}) + (\frac{1}{a_0} - 1) T_h^{(2)} (1 - 2^{-\frac{T}{T_h^{(2)}}})}, \end{aligned} \tag{25}$$

where T is the sum of the lengths of the time windows. We observe that formula (25) is equivalent to formula (23) when there is no time gap between time windows, and $t_1 = 0$.

4.3.3 The method applied on array (spot) images

For array images, we might assume that the fraction a_S of isotope 1 is constant inside a spot when hits from the background are ignored. Similarly, we might assume that the fraction a_B of isotope 1 is constant inside a local background when hits from the spot are ignored. The fraction of interest is a_S.

Assuming a spot and a local background area have been identified in the image, estimates of a_S (and a_B) might be computed from estimates of

- a_{FG}, the fraction of hits from isotope 1 in the spot area (foreground, FG),
- a_{BG}, the fraction of hits from isotope 1 in a local background area (BG),
- b_{FG}, the fraction of hits from the spot in the spot area (foreground, FG) and
- b_{BG}, the fraction of hits from the spot in a local background area (BG)

by solving the following equations with two unknowns, a_S and a_B,

$$\begin{aligned} a_{FG} &= b_{FG} \cdot a_S + (1 - b_{FG}) \cdot a_B \\ a_{BG} &= b_{BG} \cdot a_S + (1 - b_{BG}) \cdot a_B. \end{aligned}$$

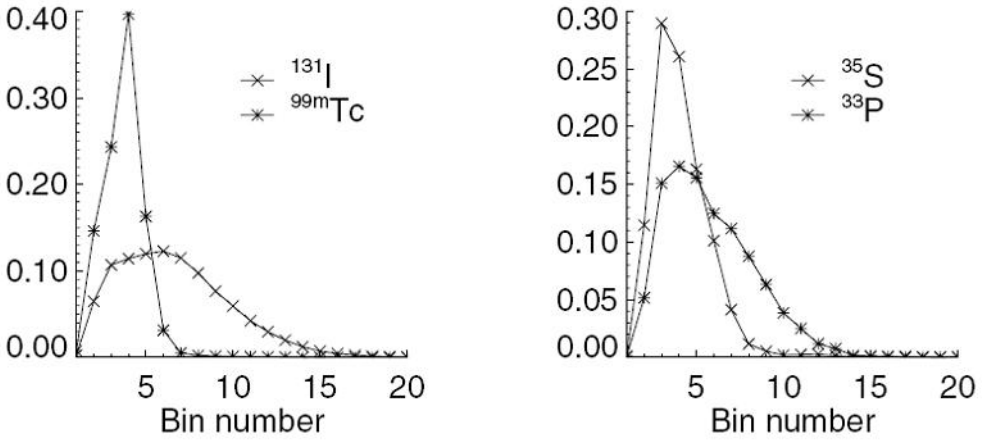


Figure 14. *Spectra for the isotopes used in the paper by Kvinnsland and Skretting. The figure has been copied from the paper: Y. Kvinnsland and A. Skretting. Methods for separation of contributions from two radionuclides in autoradiography with a silicon strip detector. Phys. Med. Biol., 2000, 45, 1183–1193*

An estimate of a_{FG} might be obtained by aggregating all hits inside the spot area, and using the above pixelwise estimation on the full spot area. Similarly, an estimate of a_{BG} might be obtained from all hits inside the local background area.

The spot and local background areas might be identified using the results of applying the method for estimating spot size described in Section 3. When the parameters in the model in Equation 1 have been estimated we could for example define a pixel to belong to the spot area if $2B \leq f(i,j;\Theta)$, and to the local background area otherwise. The fraction of hits from the spot inside the spot and local background area, b_{FG} and b_{BG}, respectively, may then be computed as

$$b_{FG} = \frac{\sum_{(i,j)\in FG} Sf(i,j;\Theta)}{\sum_{(i,j)\in FG}(B + Sf(i,j;\Theta))}$$
$$b_{BG} = \frac{\sum_{(i,j)\in BG} Sf(i,j;\Theta)}{\sum_{(i,j)\in BG}(B + Sf(i,j;\Theta))}$$

4.4 Results

The above described methods have to be tested on real images, unfortunately not available at the moment. However, the method has been tested on simulated data.

Data were generated for each of the 36 simulated experiments described in Tables 1 and 2. We have chosen the same experimental setup as the one described in the paper by Kvinnsland and Skretting to be able to compare their and our results. The spectra assumed for the isotopes are similar, but not equivalent, to those assumed in the paper by Kvinnsland and Skretting [2].

In all experiments there are either one or two time windows. In the tables, t_1 and t_2 are the start times for time window one and two, respectively. $\lambda^{(1)}$ and $\lambda^{(2)}$ are the expected number of hits from isotope 1 and 2 in a pixel for the first time window, and a_0

2. See Figure 14 for the spectra used in the paper by Kvinnsland and Skretting.

is the corresponding fraction of hits from isotope 1 at the initial time. The image in each experiment consists of 10000 pixels. Data for an experiment were obtained as follows: First, a, the fraction of hits from isotope 1, and the expected number of hits in a pixel is computed using information about $\lambda^{(1)}$, $\lambda^{(2)}$, t_1, t_2, the lengths of the time windows and the spectra and the halflives of the two isotopes. The total number of hits registered in the experiment is then set equal to 10000 times the expected number of hits in a pixel. For each hit the following information is drawn randomly: i) its pixel number; ii) which isotope it comes from; and iii) its energy bin and time window.

For each pixel, $n^{(1)}$ and $n^{(2)}$ are the number of hits from isotope 1 and 2, respectively, in the first time window. a_0, the fraction of hits from isotope 1 at the initial time, is computed from $n^{(1)}$ and $n^{(2)}$ using the formula in (24) with $K = 1$. In Tables 1 and 2, means and standard deviations computed from the simulated data are reported for each of $n^{(1)}$, $n^{(2)}$ and a_0.

Ignoring information about which isotope the different hits come from, estimates of a_0 are found from the simulated data using the method described in Section 4.3. $\hat{n}^{(1)}$ and $\hat{n}^{(2)}$ are then computed from $\hat{a}_0$. In Tables 1 and 2, means and standard deviations computed from these estimates are reported for each of $n^{(1)}$, $n^{(2)}$ and a_0.

The results in Tables 1 and 2 are very similar to those obtained for the maximum likelihood method in the paper by Kvinnsland and Skretting. As was concluded in their paper, we conclude that: i) The mean values of the estimates are very close to the true values, except when the fraction a is close to 0 or 1 ; and ii) The spectra of the (^{131}I,^{99m}Tc)-pair are better separated than the spectra of the (^{35}S,^{33}P)-pair, and then, as expected, the standard deviations for the (^{131}I,^{99m}Tc)-pair are lower than those for the (^{35}S,^{33}P)-pair. We observed that the new method is less computationally demanding than the maximum likelihood method proposed in the paper by Kvinnsland and Skretting, as can also be concluded by simply comparing the descriptions of the two algorithms.

4.5 Discussion

It is obvious that the separation of contributions from two different isotopes is easier the more separated the two underlying energy spectra are. For this reason, it is tempting to combine low energy with high energy isotopes such as ^{32}P. However, it is considered an additional problem that high energy isotopes travel along the media and hits are detected in other positions than where they were emitted. The extent of and implications of this mechanism must be studied further. It is hence an aim to use isotopes with low and medium energy, and still be able to separate well their contributions. The results obtained for simulated data indicate that this should be possible also for real data, but final conclusions cannot be drawn before the proposed method has been tested in practice on such data.

Experiment					Data			New method		
t_1	t_2	$\lambda^{(1)}$	$\lambda^{(2)}$	a_0	$n^{(1)}$	$n^{(2)}$	a_0	$\hat{n}^{(1)}$	$\hat{n}^{(2)}$	$\hat{a}_0$
0	-	0	5	0	0±0	5±2.23	0±0	0.14±0.47	4.89±2.24	0.03±0.11
0	-	5	0	1	5±2.25	0±0	1±0	4.29±2.36	0.75±1.17	0.82±0.27
0	-	5	5	0.46	5.01±2.21	4.99±2.24	0.47±0.17	4.97±3	5.03±3	0.47±0.26
0	$0.5T_h^{(2)}$	5	5	0.46	5±2.23	5.02±2.24	0.46±0.12	4.98±2.4	5.05±2.62	0.47±0.2
0	$1T_h^{(2)}$	5	5	0.46	5.03±2.28	5±2.23	0.47±0.13	5.02±2.39	5.01±2.74	0.48±0.21
0	-	10	0	1	10±3.16	0±0	1±0	8.97±3.4	1.03±1.59	0.88±0.18
0	-	10	5	0.63	10±3.21	5±2.23	0.63±0.13	9.94±4.11	5.06±3.4	0.63±0.22
0	$0.5T_h^{(2)}$	10	5	0.63	9.97±3.17	4.99±2.24	0.63±0.1	9.92±3.45	5.05±2.92	0.63±0.18
0	$1T_h^{(2)}$	10	5	0.63	10.03±3.14	4.98±2.25	0.64±0.1	10±3.35	5.02±3.08	0.64±0.18
0	-	0	10	0	0±0	10±3.13	0±0	0.17±0.56	9.83±3.17	0.02±0.06
0	-	5	10	0.3	4.99±2.24	10.01±3.16	0.3±0.12	4.97±3.13	10.03±3.85	0.3±0.19
0	0	5	10	0.3	5.01±2.23	9.97±3.14	0.3±0.08	5.02±2.41	9.96±3.28	0.3±0.13
0	$0.5T_h^{(2)}$	5	10	0.3	5.01±2.21	9.99±3.19	0.3±0.09	4.99±2.41	10.01±3.44	0.3±0.14
0	$1T_h^{(2)}$	5	10	0.3	5.02±2.24	9.97±3.14	0.3±0.09	5.03±2.36	9.96±3.57	0.31±0.14
0	-	10	10	0.46	9.97±3.14	10.03±3.21	0.46±0.11	9.94±4.26	10.06±4.33	0.46±0.19
0	-	20	10	0.63	20.05±4.49	9.95±3.13	0.63±0.09	20.04±5.9	9.96±4.94	0.64±0.16
0	-	10	20	0.3	9.98±3.2	20.02±4.45	0.3±0.08	9.97±4.4	20.03±5.39	0.3±0.13
0	-	50	50	0.46	50.01±7.05	49.99±7.13	0.46±0.05	49.93±9.55	50.07±9.66	0.46±0.08

Table 1. *Simulation results when isotope 1 is ^{131}I and isotope 2 is ^{99m}Tc. The length of all time windows is $l_k = 0.5T_h^{(2)}$.*

Experiment					Data			New method		
t_1	t_2	$\lambda^{(1)}$	$\lambda^{(2)}$	a_0	$n^{(1)}$	$n^{(2)}$	a_0	$\hat{n}^{(1)}$	$\hat{n}^{(2)}$	$\hat{a}_0$
0	-	0	5	0	0±0	5±2.25	0±0	1.29±1.86	3.74±2.62	0.28±0.38
0	-	5	0	1	5±2.24	0±0	1±0	4.46±2.4	0.57±1.25	0.87±0.27
0	-	5	5	0.47	5.01±2.24	4.99±2.24	0.47±0.17	5.24±3.85	4.76±3.9	0.51±0.36
0	$0.5T_h^{(2)}$	5	5	0.47	5±2.26	5.02±2.24	0.47±0.12	5.11±3.27	4.92±3.5	0.5±0.3
0	$1T_h^{(2)}$	5	5	0.47	5.05±2.24	4.95±2.25	0.48±0.13	5.03±3.19	4.97±3.7	0.5±0.3
0	-	10	0	1	10±3.15	0±0	1±0	9.27±3.52	0.73±1.65	0.92±0.19
0	-	10	5	0.64	10.02±3.16	4.98±2.27	0.64±0.13	9.93±5.04	5.07±4.53	0.65±0.3
0	$0.5T_h^{(2)}$	10	5	0.64	10.02±3.21	5.01±2.25	0.64±0.1	10.03±4.17	4.99±3.88	0.65±0.23
0	$1T_h^{(2)}$	10	5	0.64	9.98±3.15	5.03±2.22	0.64±0.1	10±4.1	5.01±4	0.66±0.24
0	-	0	10	0	0±0	10±3.14	0±0	1.87±2.74	8.13±3.93	0.19±0.28
0	-	5	10	0.31	4.98±2.22	10.02±3.13	0.31±0.12	5.61±4.79	9.39±5.21	0.36±0.31
0	0	5	10	0.31	5±2.26	9.99±3.17	0.31±0.08	5.18±3.81	9.81±4.42	0.33±0.23
0	$0.5T_h^{(2)}$	5	10	0.31	5±2.22	10.01±3.14	0.31±0.09	5.14±3.71	9.86±4.68	0.33±0.24
0	$1T_h^{(2)}$	5	10	0.31	5.03±2.26	9.97±3.14	0.31±0.09	5.19±3.68	9.81±4.76	0.34±0.24
0	-	10	10	0.47	10.01±3.14	9.99±3.19	0.47±0.11	10.16±5.98	9.84±6.05	0.49±0.28
0	-	20	10	0.64	20±4.49	10±3.16	0.64±0.09	20.01±7.46	9.99±6.67	0.65±0.22
0	-	10	20	0.31	9.95±3.17	20.05±4.47	0.31±0.08	10.33±7.23	19.67±7.88	0.32±0.23
0	-	50	50	0.47	49.99±7.11	50.01±7.15	0.47±0.05	50.09±14.01	49.91±13.95	0.47±0.13

Table 2. *Simulation results when isotope 1 is ^{35}S and isotope 2 is ^{33}P. The length of all time windows is $l_k = 0.5T_h^{(2)}$.*

5 Conclusion

Biomolex has technology for real-time imaging of radioactive emissions. Images are acquired for measuring radioactivity in tissue slices for assessing transport and uptake of labeled biomolecules or in spots in protein kinase arrays for evaluating protein phosphorylations. The sensor used is a silicon strip detector, and there is a problem that some strips are defect, i.e. they do not register any signal. For some of the missing stripes the signals in the neighbor stripes will be too high. Also, the signal strengths of the stripes may vary.

A method for restoring the images and imputing the missing data in tissue images has been developed and it has been implemented in the Biomolex software. The proposed method gives satisfactory results for the test set. It is a promising method for restoring tissue images that are quite smooth and that do not contain more than two consecutive stripes that are missing or overactive or that have too low signal strength. Improved results might have been obtained by including a calibration experiment.

An important issue when analyzing array images is to estimate the total number of hits in each spot. A method for estimating such spot sizes has been developed and will soon be implemented in the Biomolex software. The image after missing data have been imputed, is one of the outputs when applying this method. The algorithm for estimating spot sizes seems robust. However, both this method for array images and the method for restoring tissue images should be tested more extensively on images of varying quality before being used in commercial systems.

The Biomolex technology opens up for the possibility of using multiple isotopes during a single image acquisition. A challenge is then to decompose the measured signal into signals for the different isotopes, i.e. decide which fractions of the signal/hits that come from which isotope. A method for decomposing signals has been proposed and has been tested on simulated data. The results are promising. Unfortunately, no real data were available. The method is based on using information about the energy spectrum and decay times for each radioactive isotope.