Tina Memo No. 2010-004
Internal Memo

# Automatic Morphological Landmark Location using Local Image Patch Registration

P.A. Bromiley, H. Ragheb and N.A. Thacker

Last updated
25 / 10 / 2010


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

# Automatic Morphological Landmark Location using Local Image Patch Registration

P.A. Bromiley, H. Ragheb and N.A. Thacker
Imaging Science and Biomedical Engineering Division
Medical School, University of Manchester
Manchester, M13 9PT, UK
paul.bromiley@manchester.ac.uk

**Abstract**

We propose a system for automatic identification of morphological landmarks in 3D medical image volumes. The system will be implemented as a software package comprising three components: a manual landmark identification tool, allowing the generation of training data; a global registration tool, allowing alignment of new image volumes with volumes in the training data, in order to provide approximate, initial landmark locations; and a local registration tool, which will refine the approximate landmark locations using correlation between image patches around each landmark. This report deals with initial development of the third component.

The aim of this work was to develop an algorithm capable of providing optimised transformation model parameters for correlation-based local image patch registration, an estimate of the goodness-of-fit of the patches, and an estimate of the covariance matrix of the optimised parameters. The algorithm was tested using micro CT images of a *Mus musculus* skull. Monte-Carlo simulations were performed, demonstrating the ability of the algorithm accurately to locate landmark points and provide realistic estimates of the transformation parameter covariance matrix.

## 1 Introduction

Quantitative morphological analysis of biological specimens is an increasingly active research area, with the aim of investigating the links between morphology, phylogeny and ecology and, once the extent of these links is understood, of using morphological analysis as a measurement tool to infer phylogenic and ecological data. One currently popular approach is Procrustes analysis of manually identified landmarks. Manual landmarking is time consuming, particularly when performed on 3D data, and must typically be performed for at least several tens of specimens. Therefore, morphological analysis would benefit considerably from an automatic landmark identification algorithm.

In previous work [12, 14, 13, 16, 17, 15] we developed a system for automatic location of morphometric landmarks in 2D microscope images of *Drosophila* wings. In this work, we aim to extend the method to deal with 3D data, with particular focus on micro-CT images of rodent skulls, and to implement the method as a software package comprising three components. The first, a manual landmark identification tool, has already been developed [4]. This will be used to generate training data for the automatic landmark location tools. The second component will be a global registration tool, capable of aligning previously unseen image volumes with those stored in a database of training data. This will provide approximate, initial landmark locations in previously unseen data. However, shape differences between the training data and the previously unseen data will prevent accurate landmark location using global registration. Therefore, a third component will refine the approximate landmark locations using a local registration of image patches around each landmark in the training data. Shape variation between the training data and the previously unseen data will be minimal for sufficiently small patches, allowing accurate landmark location. The availability of multiple training data sets, and thus multiple hypothesised locations for each landmark, will then allow further refinement; for example, a subset of the hypothesised location with the best goodness-of-fit could be chosen, and then averaging performed to provide the final, automatic landmark location.

This report focuses on initial development of the third component; the local image patch registration. The proposed algorithm extracts three orthogonal image patches from around each landmark in the training data, aligned with the three major axes of the original image volume. Orthogonal patches, rather than a 3D data block, are used due to the high resolution of the image volumes used in morphological analysis; the patches must incorporate local but not global structure in order that landmarks are located on the basis of local shape, and visual inspection of example data sets indicated that a patch width of several tens of pixels will be sufficient to meet this. Orthogonal patches of this size are sufficient to ensure that the errors on the optimised transformation model parameters will

be limited by factors other than the amount of data used in the likelihood calculation i.e. using a 3D block would increase the processor time required by the algorithm without having any measurable effect on the final accuracy.

Rather than use the image patches directly, the algorithm used the horizontal and vertical derivatives of the patches. This provides independence to any constant offset between the intensity in the training and previously unseen data sets, such as might be introduced by differences in scanner parameters. However, a linear intensity scaling might also be expected, particularly when matching data from species of different sizes, due to differences in bone density/mineralisation. Therefore, a parameter is incorporated into the likelihood allowing for a linear scaling between patch derivatives. This likelihood is then optimised in order to register the derivative patches from the training data (i.e. the target) to those from previously unseen data (i.e. the source). In this report we use a transformation model consisting of three translation parameters; extension of the system to include other modes of variation would be trivial.

In order to make efficient use of the optimised transformation model parameters, the system is also required to estimate the goodness-of-fit of the source and target volumes, in order to allow detection of mis-registrations and to provide an indication of which hypothesised landmark locations, from multiple sets of training data, to use in the final, automatic landmark location. Furthermore, the system is required to estimate the covariance matrix on the optimised transformation model parameters, both as a test of measurement accuracy and in order that error propagation can be used to calculate the covariance matrix on the automatic landmark location, such that the location errors can be accounted for in any subsequent analysis of the results. A goodness-of-fit measure is provided by the $\chi^2$ per degree of freedom at the optimum, although this requires an accurate estimate of the noise on the data incorporated into the $\chi^2$ calculation. An estimate of the covariance matrix on the optimised transformation model parameters can be provided by the Minimum Variance Bound (MVB) [2]. In previous work [20, 7, 8, 5, 6], we investigated the use of this technique to estimate the errors on the transformation model parameters for mutual information based registration of MR images of the brain. Bansal et al. [1] have investigated a similar method in the same application. Here, we adopt a modified approach compared to our earlier work. Rather than attempting to calculate the MVB as a sum over individual data terms, we calculate it using finite stepping on the $\chi^2$ function, in order to avoid problems with numerical stability.

The remainder of this report is organised as follows. Section 2 presents some observations on the process of manual landmarking, and on the global registration stage of the automatic landmark location procedure. Section 3 describes the theoretical development of the local registration algorithm and the corresponding goodness-of-fit and parameter covariance estimation procedures. Evaluation of the algorithm was performed using Monte-Carlo simulations of local patch registration on micro-CT images of a *M. musculus* skull; the experimental procedure is described in Section 4, and the results of the evaluation are presented in Section 5. The aim of this initial phase of testing was to evaluate the numerical stability and implementation of the algorithm, and the experimental design reflected this; all experiments involved registration of patches from identical images, such that the correct solution was known exactly. The results therefore do not test the use of the algorithm in automatic landmarking, where landmark positions must be transferred between images of different rodent skulls and differences in shape may be present. More realistic experiments will be performed at a later date and presented in a companion document. Conclusions are presented in Section 6, together with a discussion of the limitation of the evaluation methodology and its relation to the application more realistic data sets incorporating variation in shape.

# 2 Real-world Issues

Here we discuss the main issues we should consider in order to deal with real-world data.

### 2.0.1 Scale

In morphological analysis, apparent scale change is introduced due to developmental processes as well as the inherent differences in mature samples. However, unlike more general computer vision problems, the overall scale of a given structure in an image is largely fixed by the nature of the imaging process and the overall structures are very similar. We therefore assume, in the development of the current algorithm, that one overall approximate scaling of the data is sufficient to allow us to locate small sub-regions (e.g. a scale factor error of 5% will produce a negligible one-pixel shift in data at the edges of a region 40 pixels across). This scaling will be provided by the global registration stage of the automatic landmarking procedure; the accuracy of this assumption will be evaluated during testing of the completed system, when all components are available.

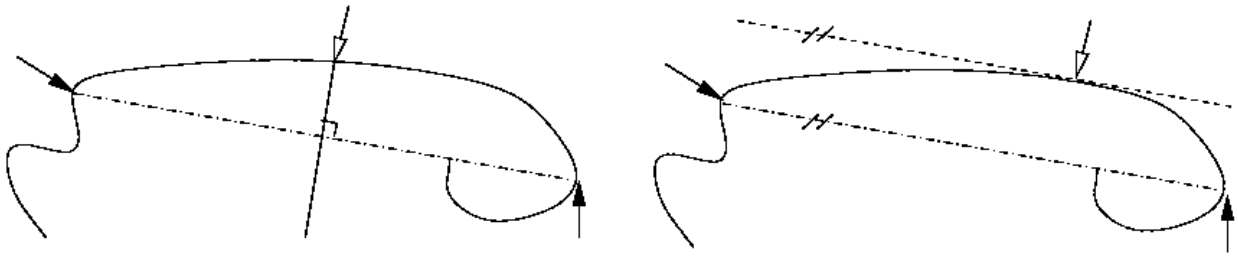

Figure 1: Manual mark-up strategy for locations with low curvature using two distinct landmarks; left: linear combination of existing points; right: parallel lines as guides to mark-up the tangential point. Full pointers show existing points; empty pointer show new points.

#### 2.0.2 Rotation

A mark-up strategy is needed mainly to make sure all example landmarks are correctly selected manually. For instance, there are places on the skull surface where the principal curvature in one direction (or even both directions) is very small (Fig. 1). These places are obviously not the best candidates for selecting landmarks, because it is difficult to maintain consistency between datasets. When such "semi-landmarks" are needed it is usually the case that two or more existing landmarks are used as a constraint. However, using a linear combination model to define a line of intersection results in a positional estimate that is correlated with the other two landmarks, indeed, the measurement between the constraint points is fixed entirely by their positions. The new point does not represent a fully independent measurement, and this invalidates the independence assumptions needed in later shape analyses (e.g. PCA). However, if the two reference points are used instead to define a tangent and a point of extremum identified, manual identification accuracy is then determined by the curvature of the bone boundary, which gives a better reflection of the location precision. As this approach is equivalent to specifying a rotation, it is consistent with localising the mark-up point using a reference image with fixed orientation. This approach therefore also simplifies the numerical implementation of the localisation process.

The accuracy of the automated estimate of initial orientation will need to be better than a degree in order to limit the effects of error on localisation (e.g. an error of half a degree will produce subpixel effects at the edge of a 40 pixel wide block) and its covariance. By matching to multiple candidate references we can however mitigate against large rotation errors using statistical match criteria. Again, we assume that this scaling will be provided by the global registration stage of the automatic landmarking procedure; the accuracy of this assumption will be evaluated during testing of the completed system, when all components are available. Due to the observations in this and the previous section, for the purpose of the evaluation presented here the transformation model used in the local patch registration is limited to translation components only; the suitability of this model will be evaluated when both the local and global registration components are incorporated into the automatic landmark location software, at which point any requirement for additional modes of variation will be reassessed.

#### 2.0.3 Irregularities

For specific local bone structures there are a variety of shapes for which consistent landmark points need to be estimated. Hence, a clear strategy of defining landmarks on such structures must be specified prior to manual landmark identification. Automation is therefore a challenging problem, because in some cases there is obviously no systematic way to define robust features using local information. One approach is to learn an appearance model [9] corresponding to the three orthogonal slices of each block. However, we expect that we will have to work with a minimum number of reference examples (insufficient to construct an accurate model), while at the same time topological differences between corresponding locations for identified structures are likely to be quite large. It is therefore unlikely, in general, that the examples provided will be efficiently described by a single linear model. If we can identify markup locations without using appearance models it is therefore sensible to do so [12, 14, 13, 16, 17, 15].

These facts suggest the alternative approach of matching individual reference data separately, subject to nominal deformations (which in the simplest case might be scaled or rigid matching). Note that, in general, we may wish to make simple linear models out of the individual example blocks. In doing this, we should shortlist a number of best "candidate" points matching against the landmark from different example structures found in our reference blocks. This strategy has the advantage of avoiding a highly dimensioned optimisation, which is time-consuming, replacing it with several low-dimensioned optimisations. If we assume that the candidate matches have been matched to the

data perfectly, then the corresponding reference landmark locations can be combined using a weighted mean and errors on this estimate propagated accordingly. However, this is beyond the scope of the current paper.

#### 2.0.4 Missing Data

In addition, it is a matter of concern in morphological studies incorporating species with small, delicate bone structures (which may be damaged during sample preparation) or museum specimens (which may have become damaged during long-term storage and handling) that parts of the data are missing in some cases. This results in inconsistency in the landmark points among samples. In our opinion this is an extreme case of irregularities where features corresponding to missing data do not exist at all. Hence, our method must be able to recognise such cases so that landmarks corresponding to these locations could be labelled as impossible/virtual landmarks. In order to quantitatively indicate how well sample data conforms to a predefined model, we plan to define a $\chi^2$ based hypothesis. Using this technique we also expect to be able to inform the user of those occasions when more reference examples are needed.

# 3 Theoretical Development of the Algorithm

## 3.1 Template Matching

The aim of this work is to take an image patch $I$ centred on a landmark location in an image volume for which a set of landmarks have already been manually identified, and to optimise a set of transformation parameters that bring this patch into alignment with a second image volume, by comparing $I$ to an image patch $J$ at the equivalent, transformed position in the second volume. Once this template matching task has been performed, the landmark location from the first image, when transformed using the optimised transformation parameters, gives an estimate of the landmark location in the second volume. This can be presented for visual inspection and refinement; however, a goodness-of-fit test and an estimation of the covariance matrix on the transformation parameters can also be obtained to provide some level of automatic accuracy checking, allowing the software to alert the user to landmarks where the fitting process is known to have failed.

The log-likelihood function for matching image patches $I$ and $J$, where an overall scaling difference is expected between the two images, is

$$lnL = \sum_{i=1}^{N} -\frac{(\gamma I_i - J_i)^2}{2\sigma_E^2}$$

where $I_i$ is a pixel from image patch $I$, $J_i$ is a pixel from the equivalent position in image patch $J$ (i.e. after having transformed one of the images using the transformation parameters being optimised; for the moment, we do not specify which image is being transformed), $i$ is an iterator over the pixels in the patches, $N$ is the number of terms in the sum i.e. the number of pixels in each patch, $\gamma$ accounts for the overall scaling difference between the images, and $\sigma_E$ is the effective noise i.e. the expected standard deviation of the difference between the patches at perfect alignment, assuming that there is no difference in shape between the structures in the two image patches. Therefore, using error propagation,

$$\sigma_e^2 = \gamma^2\sigma_I^2 + \sigma_J^2$$

where $\sigma_I$ is the standard deviation of the noise on image patch $I$ and $\sigma_J$ is the standard deviation of the noise on image patch $J$ i.e. the expected error on the difference between the images is the sum in quadrature of the noise on the individual patches, with $\gamma$ taken into account. If it is assumed that the noise is the result of an additive process occurring during image acquisition i.e. is independent of image contents, then it follows that $\sigma_I \approx \sigma_J$, and

$$\sigma_e^2 = (\gamma^2 + 1)\sigma_I^2$$

so

$$lnL = \sum_{i=1}^{N} -\frac{(\gamma I_i - J_i)^2}{2(\gamma^2 + 1)\sigma_I^2}$$

The constant factor of $-1/2$ is usually ignored when writing the log-likelihood function, using the transformation-invariance characteristic of maximum-likelihood techniques (i.e. optimising any monotonic function of the likelihood will give the same result as optimising the likelihood itself). However, this factor is retained here in order that the likelihood can also be used for covariance estimation, which is only possible when all scaling constants are retained.

### 3.2 Goodness-of-fit Estimation

Goodness-of-fit estimation was performed by calculating the $\chi^2$ for the matching process [2]

$$\chi^2 = \sum_{i=1}^{N} \frac{(\gamma I_i - J_i)^2}{\sigma_E^2}$$

The concept behind $\chi^2$ optimisation is that, if the model is a good fit to the data, then the only differences in the $(\gamma I_i - J_i)^2$ term at the optimum of the fitting process are due to random noise. The numerator of the sum is therefore expected to be equal to $N\sigma_E$, where $N$ is the total number of terms in the sum, giving a $\chi^2$ of $N$, or equivalently a $\chi^2$ of one per degree-of-freedom. In theory, a small adjustment to $N$ should be made by subtracting the number of parameters optimised in the process; however, since we are dealing with three parameters and $\approx 10,000$ data terms this is insignificant and can be ignored and the number of degrees-of-freedom assumed to be equal to the number of terms in the sum. Any departure from the expected $\chi^2/D.O.F$ of 1 indicates poor goodness-of-fit, which could either be due to an inability of the model to fit the data or a failure of the optimiser to converge on a global minimum. Since the $\chi^2$ distribution function is known, a probabilistic interpretation can be placed on the result.

It will be noted that there is a close similarity between this form of the $\chi^2$ and the log-likelihood. One major difference is that log-likelihoods are frequently constructed with arbitrary scaling, taking advantage of the transformation-invariance of maximum likelihood estimators. However, in cases where the log-likelihood is intended for uses beyond simply finding its (transformation-independent) optimum, e.g. for covariance estimation, the scaling factors must be retained. Due to the similarities between the log-likelihood and the $\chi^2$, and the fact that the $\chi^2$ must be properly scaled in all cases such that it can be compared to the known $\chi^2$ distribution function [2], we can therefore state that the log-likelihood is properly scaled (in the sense that it can be used in the calculation of the Minimum Variance Bound; see below) when it is distributed according to $-2logL = \chi^2$. This has lead some researchers (e.g. [3]) to assert that the $\chi^2$ is the more fundamental quantity than likelihood. We refrain from further comment here; additional material is available in [7].

$\chi^2$-based goodness-of-fit testing depends on the absolute value of the $\chi^2$ and it is therefore essential that the $\sigma_E$ term is a genuine estimate of noise on the data terms i.e. it is not simply the original image noise, but must include the effects of all steps of the algorithm. Since the algorithm operates on image derivatives, and these are calculated using finite differencing i.e. an equation of the form

$$\frac{\delta I_i}{\delta i} = \frac{I_{i+1} - I_{i-1}}{2}$$

error propagation implies an additional factor of 2 i.e.

$$\sigma_I^2 = \frac{\sigma_{I_o}^2}{2}$$

where $I$ is the derivative patch and $I_o$ is the original patch from which this derivative was obtained. However, it must also be remembered that the image patch $I_o$ is not simply a patch of raw image data but, in the case of the patch from the source volume, has been interpolated from the original data (during registration, the source image must be interpolated on the voxel grid of the target image using the current transformation model parameters, so that the voxel values used in the cost function originate from the same spatial positions). Since interpolation involves averaging over neighbouring voxels, it may have an effect on noise level that is a function of the transformation model parameters. This effect has been studied extensively in the area of medical image registration [18, 8].

In the experimental method used here this complexity was avoided by adding noise to the image patches after interpolation, so that subsequent error propagation involved only the $\gamma$ scaling and differential terms. Furthermore, the same image was used as both the source and target volume, primarily to ensure that there was no intrinsic shape variation between the two and that the correct registration solution was known, allowing the accuracy of the results to be evaluated. However, this also had the effect that the intrinsic image noise i.e. the noise added to the underlying data during image acquisition, was identical between the two volumes, so had no random effect and instead acted like structure or texture. Therefore, for these experiments

$$\sigma_E^2 = \frac{(\gamma^2 + 1)\sigma_A^2}{2}$$

where $\sigma_A$ was the standard deviation of the noise added to the patches in the course of the Monte-Carlo experiment. In the final version of the software, used for automatic landmark placement, it is intended that the noise will be measured directly from the interpolated patches using the techniques described by [11], such that the same equation can be used. Further comments on this issue are made in Section 6.1.

### 3.3 Estimating the Scaling Factor

$\gamma$ can be estimated through the usual maximum likelihood procedure i.e. setting the first derivative of the log-likelihood function, with respect to $\gamma$, equal to zero and solving for $\gamma$. So

$$\frac{\delta lnL}{\delta\gamma} = \frac{-1}{2\sigma_I^2}\sum_{i=1}^{N}[\frac{2(\gamma I_i - J_i)I_i}{1+\gamma^2} + \frac{(\gamma I_i - J_i)^2 2\gamma}{(1+\gamma^2)^2}] = 0$$

and

$$\sum_{i=1}^{N}\frac{(\gamma I_i - J_i)^2\gamma}{1+\gamma^2} = \sum_{i=1}^{N}(\gamma I_i - J_i)I_i$$

Multiplying both sides by $(1+\gamma^2)$, expanding the square and simplifying gives

$$\sum_{i=1}^{N}[(1-\gamma^2)I_iJ_i + \gamma(J_i^2 - I_i^2)] = 0$$

Put

$$\sum_{i=1}^{N} I_i^2 = |I|^2 \quad and \quad \sum_{i=1}^{N} J_i^2 = |J|^2$$

and use the familiar expression for the $I_iJ_i$ dot-product i.e.

$$\sum_{i=1}^{N} I_iJ_i = |I||J|cos\phi$$

where $\phi$ is a measure of the dissimilarity between the patches; if we assume that the transformation contains sufficient modes of variation to bring the image patches into perfect alignment, such that the only remaining difference between them is random noise, then we can assume that $\phi \approx 0$ i.e. $cos\phi \approx 1$, so

$$(1-\gamma^2)|I||J| + \gamma(|J|^2 - |I|^2) = 0 = -\gamma^2|I||J| + \gamma(|J|^2 - |I|^2) + |I||J|$$

This is quadratic in $\gamma$, and so can be solved using the usual expression

$$\gamma = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

where

$$a = -|I||J| \qquad b = |J|^2 - |I|^2 \qquad c = |I||J|$$

giving

$$\gamma = \frac{|J|}{|I|} \quad or \quad \gamma = \frac{-|I|}{|J|}$$

The moduli of the image $|I|$ and $|J|$ must be positive, as must $\gamma$, and so the first solution is the physical one, and provides the method for estimating $\gamma$ (the second represents cases where the scaling includes an inversion, which is not expected with the images used here).

### 3.4 Covariance Estimation

The maximum achievable accuracy, i.e. minimum achievable covariances, on the parameter estimates provided by a maximum likelihood technique is given by the minimum variance bound (otherwise known as the Cramer-Rao bound or Frechet inequality; it was first discovered by Aitken and Silverstone)

$$C_\theta^{-1} \leq < \frac{\partial \log L}{\partial\theta_r}\frac{\partial \log L}{\partial\theta_s} >$$

or

$$C_\theta^{-1} \leq - < \frac{\partial^2 \log L}{\partial\theta_r\partial\theta_s} >= -\left.\frac{\partial^2 \log L}{\partial\theta_r\partial\theta_s}\right|_{\theta=\theta_{max}}$$

where $\theta$ is the vector of parameters being optimised and $\theta_{max}$ is the vector of parameter values at the optimum. (a derivation is provides in [7]). This states that the variance of a given parameter, or the covariance of a pair

of parameters, is bounded by the second differential of the likelihood function with respect to the parameter(s). A simple way to view this result is that the accuracy with which the parameters can be determined is dictated by the width of the likelihood function around the optimum; the narrower the function (i.e the greater its second differential) the more accurately the position of the optimum can be located.

The condition for achieving the minimum variance bound is that the log-likelihood function is quadratic i.e. that the likelihood function is Gaussian. Note that this is a statement about the shape of the likelihood function, not about the error distribution on individual data terms contributing to it. Therefore, the Central Limit Theorem will ensure that the shape of the likelihood function converges to a Gaussian as the number of individual data contributing to the likelihood grows, even if the error distribution on the individual data is not Gaussian, as long as the errors on the data are not highly correlated. This condition is met in most computer vision and image analysis tasks, and so the minimum variance bound can be used directly as an estimate of the errors in such tasks i.e. we can assume that we will achieve the bound due to the large amount of data available. In previous work [20, 7, 8, 5, 6] we demonstrated the application of this approach to the estimation of the covariances for global registration of medical images using mutual information. Here we apply a similar approach to the estimation of covariances for the image patch registration.

Due to the relationship between log-likelihood and $\chi^2$, the expression for the minimum variance bound can be restated in terms of the individual data terms in the $\chi^2$ sum; let these be called the $\chi_i$, so [7]

$$C_\theta^{-1} = \sum_i (\bigtriangledown_\theta \chi_i)^T \otimes (\bigtriangledown_\theta \chi_i)|_{\theta=\theta_{max}}$$

However, in previous work [8, 5, 6] summing over the individual data terms in this way was found to be quite unstable and susceptible to image noise, requiring additional steps such as estimating covariances from a range of $\bigtriangledown_\theta$ and then taking the median value of the results. Therefore, in this work we adopt an alternative method in which the covariances are calculated from finite stepping on the $\chi^2$ function itself, using [19]

$$\chi^2 = \chi_0^2 + \Delta\theta^T C_\theta^{-1} \Delta\theta$$

where $\chi_0^2$ is the value of the $\chi^2$ function at its optimum. We can find linear equations between the unknown elements of the inverse covariance matrix $C_\theta^{-1}$ and the corresponding value obtained from the $\chi^2$ function. In our case there are transformation parameters only, which are the coordinates of the centre of the target block. We can, for instance, change the optimum values of these parameters by adding one pixel position in different combinations. With 3 parameters there are 8 combinations where (0,0,0) gives $\chi_0^2$. As the number of unknown elements in the $C_\theta^{-1}$ matrix is 6, and by leaving the combination with non-zero values for all parameter changes $\Delta x_i$, $\Delta y_i$ and $\Delta z_i$, there will be 6 linear equations with closed from analytic solution as follows.

$$\xi(\Delta\theta_i) = \xi(\Delta x_i, \Delta y_i, \Delta z_i) = \frac{\chi^2(\theta_{opt} + \Delta\theta_i) - \chi^2(\theta_{opt})}{\Delta\theta_i}$$

For the 6 combinations used, we can write down 6 linear equations

$$\xi(0,0,\Delta z_1) = \xi_1 = C_{zz}^{-1} \qquad \xi(\Delta x_6, \Delta y_6, 0) = \xi_6 = C_{xx}^{-1} + 2C_{yx}^{-1} + C_{yy}^{-1}$$

$$\xi(0,\Delta y_2,0) = \xi_2 = C_{yy}^{-1} \qquad \xi(0,\Delta y_3,\Delta z_3) = \xi_3 = C_{yy}^{-1} + 2C_{zy}^{-1} + C_{zz}^{-1}$$

$$\xi(\Delta x_4,0,0) = \xi_4 = C_{xx}^{-1} \qquad \xi(\Delta x_5,0,\Delta z_5) = \xi_5 = C_{xx}^{-1} + 2C_{zx}^{-1} + C_{zz}^{-1}$$

It follows that

$$C_{zz}^{-1} = \xi_1 \qquad C_{yx}^{-1} = 0.5(\xi_6 - \xi_2 - \xi_4)$$

$$C_{yy}^{-1} = \xi_2 \qquad C_{zy}^{-1} = 0.5(\xi_3 - \xi_1 - \xi_2)$$

$$C_{xx}^{-1} = \xi_4 \qquad C_{zx}^{-1} = 0.5(\xi_5 - \xi_1 - \xi_4)$$

The covariance matrix $C_\theta$ is then found by inverting the $C_\theta^{-1}$ matrix. Note that $\Delta x_i$, $\Delta y_i$ and $\Delta z_i$ should be chosen so that the difference between $\chi^2(\theta_{opt} + \Delta\theta_i)$ and $\chi^2(\theta_{opt})$ becomes about unity. However, this needs to be balanced with numerical accuracy and when dealing with pixel positions in image data, in practice one pixel change in x, y and z directions is found to be a good choice.

### 3.5 Coding the Cost Function

The previous steps in the theoretical development of the algorithm have been independent of any consideration of what the image patches $I$ and $J$ actually represent. As stated in Section 1, the algorithm proposed here operates on the derivatives of image patches, rather than on image patches directly. Therefore, six image patches are involved from the target volume for each landmark; horizontal and vertical derivatives (here called $x$ and $y$ derivatives, where the lower-case notation indicates the coordinate system within the patch) of each of three orthogonal patches taken from the volume, on planes normal to the major axes $X$, $Y$ and $Z$ (where the upper-case notation refers to the coordinate system of the target volume), centred at the landmark location. Corresponding patches are obtained from the source volume, generated by interpolation using the current estimate of the transformation model. Therefore, the modulus of the image patches is given by e.g.

$$|I|^2 = \sum_n^{N_X} \left[ \left( \frac{\partial I_n}{\partial x} \right)^2 + \left( \frac{\partial I_n}{\partial y} \right)^2 \right]_X + \sum_n^{N_Y} \left[ \left( \frac{\partial I_n}{\partial x} \right)^2 + \left( \frac{\partial I_n}{\partial y} \right)^2 \right]_Y + \sum_n^{N_Z} \left[ \left( \frac{\partial I_n}{\partial x} \right)^2 + \left( \frac{\partial I_n}{\partial y} \right)^2 \right]_Z$$

The registration cost function involves two loops over the data. The first calculates $\gamma$ using

$$\gamma = \frac{|J|}{|I|}$$

The modulus of the target volume patches is constant, as it is the source volume that is affected by the transformation model, and so one of these two moduli can be pre-calculated. The second loop over the data then calculates the registration cost function

$$\chi^2 = \sum_{i=1}^{N} \frac{(\gamma I_i - J_i)^2}{\sigma_E^2}$$

where $\sigma_E$ is the standard deviation of the noise on the voxels used in the numerator of this equation. This function is optimised using the simplex algorithm [10], a local optimiser.

Since three different processes are involved here (optimisation, goodness-of-fit estimation, and covariance estimation), some flexibility is available to deal with numerical stability issues and processor time requirements. Since a local optimisation algorithm is used, the optimisation process is susceptible to the effects of local minima, and so the cost function is required to be reasonably smooth. Therefore during optimisation the image patches are smoothed prior to differentiation with one iteration of Gaussian smoothing with a kernel size of 1 voxel. The denominator of the cost function is also omitted from the calculation i.e. the noise on the patches is assumed to be constant, in order to reduce processor time requirements, utilising the transformation-invariance property of maximum likelihood estimation. The noise term is obviously required in the calculation of the goodness-of-fit, as it provides the scaling for the numerator of the $\chi^2$ function. However, in order to avoid the requirement of performing error propagation on the smoothing algorithm, the $\chi^2$ for goodness-of-fit measurement is calculated from unsmoothed patches. The consequence of this is that the calculation gives the goodness-of-fit between unsmoothed patches, rather than the goodness-of-fit between smoothed patches; this is therefore a more accurate interpretation of the data. The covariance estimation measures the expected errors on the parameters from the width of the cost function optimised during registration, and so smoothing must be taken into account as it reduces the information content of the data. Furthermore, all scaling parameters must be included as the derivative of the cost function is being measured, so the covariance estimate uses the full $\chi^2$ (including the noise term) of the smoothed patches.

## 4 Evaluation

A micro-CT image of a *M. musculus* skull, consisting of 1003 slices of 658 by 658 voxels, was used in the evaluation of the algorithm. Manual landmarking was performed using the TINA Manual Landmark Tool [4]; the image volume was down-sampled by a factor of 2 along all axes during loading. Five landmark points were identified, as shown in Fig. 2; points 1 and 2 on the right and left coronoid processes, point 3 in an arbitrary location on the top of the skull along the line of bilateral symmetry, and points 4 and 5 in arbitrary locations on the sides of the skull. This selection provided points that were well-constrained along different axes. In order to ensure that there was no shape variation between the landmarked and non-landmarked volumes, the same image volume was used as both the source and target volume, allowing testing of the goodness-of-fit against a known result.

Monte-Carlo experiments were then performed, in which independent Gaussian random noise fields with a mean of zero were added to both the $I$ and $J$ image patches. The standard deviation of the noise was varied between 0.5 and 3.5 times the noise on the original image volume, measured using TINA noise measurement function, which

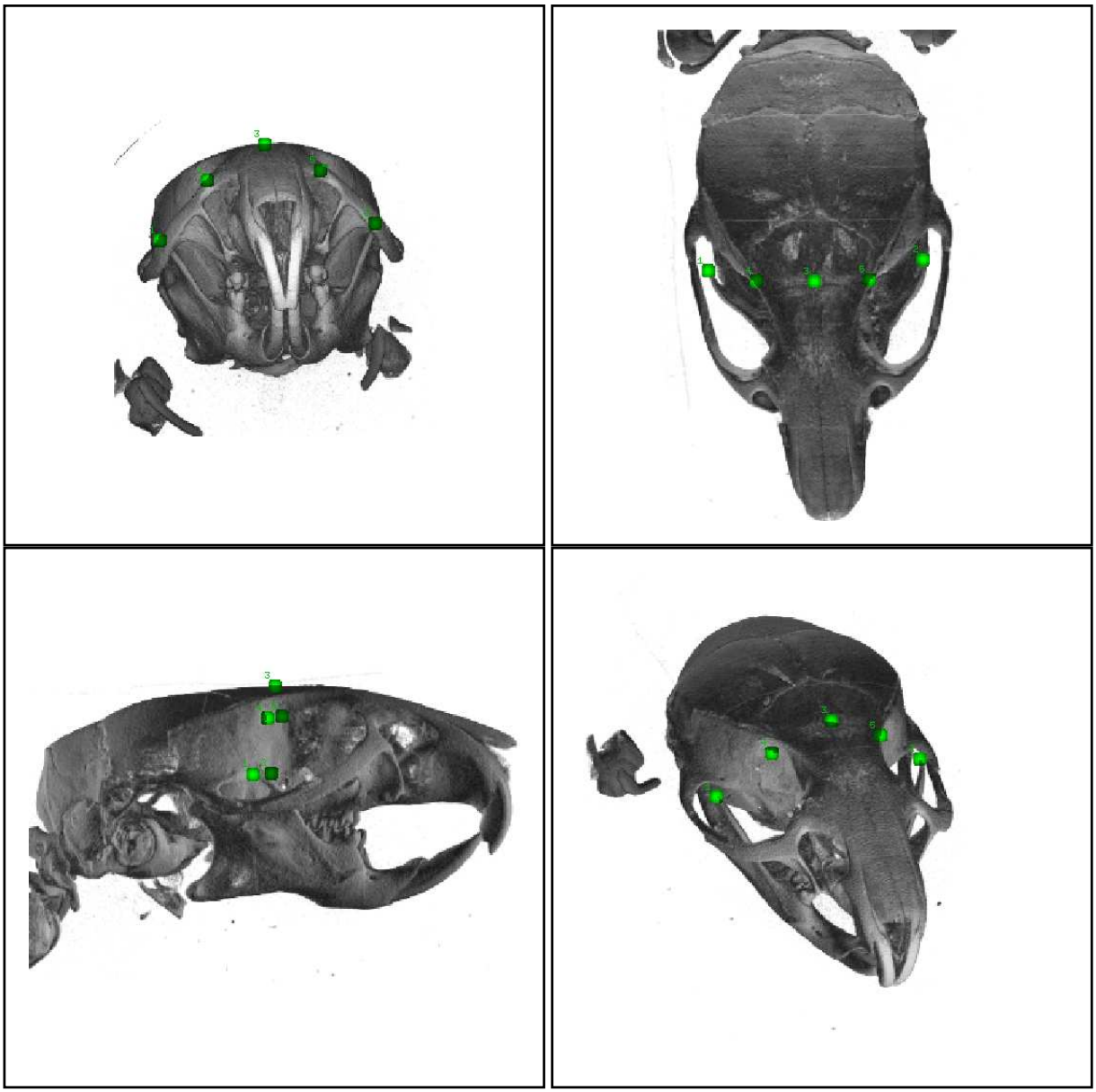

Figure 2: Locations of the landmark points used in the evaluation, shown on the *M. musculus* skull

is based on the width of zero-crossings in horizontal and vertical gradient histograms [11]. One hundred iterations of the Monte-Carlo experiment were performed at each level of added noise. A random offset was introduced into the alignment parameters prior to each registration, using samples drawn from a Gaussian distribution with a mean of zero and a standard deviation of one voxel, added to each of the transformation model parameters independently. The initial set of transformation model parameters, final set (after optimisation), initial and final $\chi^2$ value (including all scaling and noise terms), and estimated covariance matrix parameters for each iteration of each Monte-Carlo experiment were recorded.

The image patch size used in the evaluation was $\pm 30$ voxels around the landmark; this was chosen from visual inspection of the data as a patch size that encompassed local but not global structure. From this, a border of two voxels was ignored during cost function calculation, in order to provide a margin of data from which to calculate derivatives by finite differencing. During registration, the Monte-Carlo noise was added to the patches, they were smoothed using a Gaussian filter with a kernel size of 1 voxel, and horizontal and vertical derivatives were taken. The smoothing step was omitted when calculating the final $\chi^2$ per degree-of-freedom, in order that it did not have to be taken into account when estimating the effective noise on the $\chi^2$.

A second set of experiments were performed in which the effects of patch size were investigated. These were performed at one level of added noise equal to $1\sigma$ of the image noise, and the patch size was varied from $\pm 5$ to

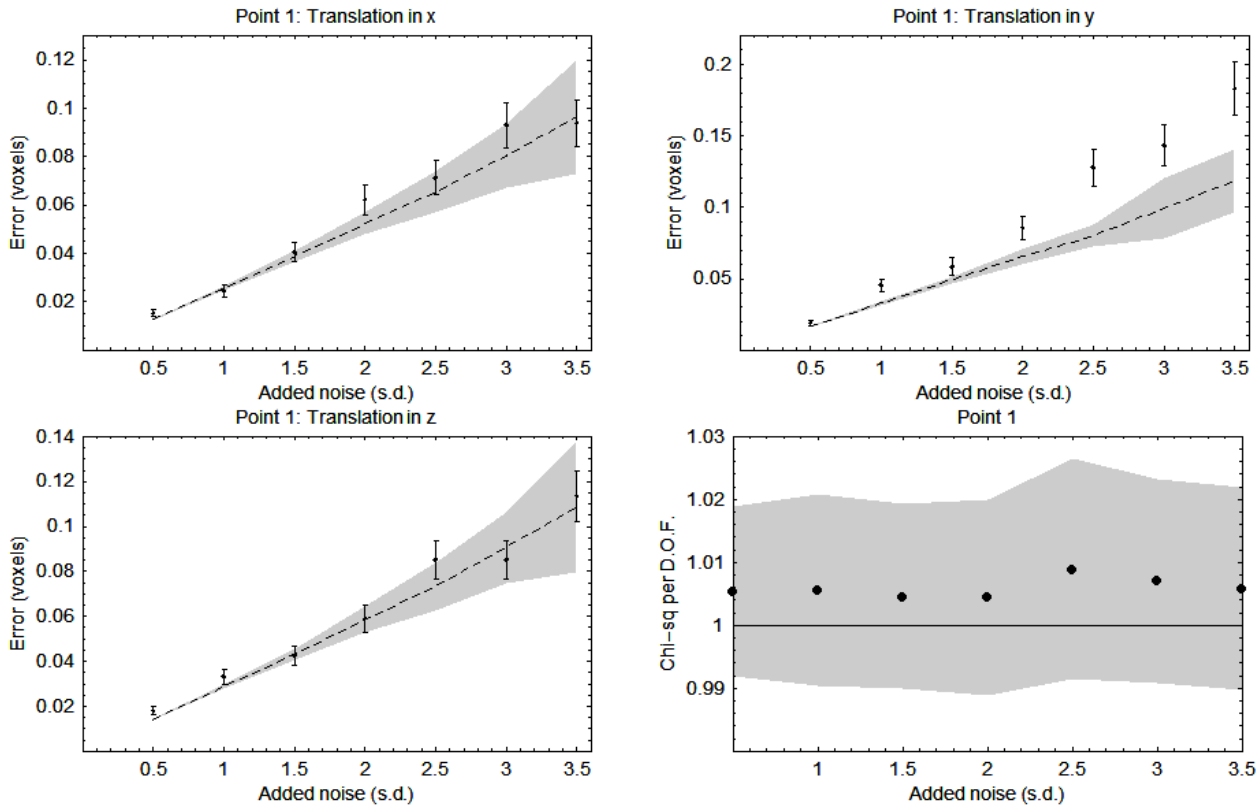


Figure 3: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 1, and the $\chi^2$ per degree of freedom at the optimum, for varying added noise.

$\pm 60$ voxels in steps of 5 voxels. All other experimental details remained the same.

The results of each Monte-Carlo experiment (i.e. each noise level and patch size) were processed by calculating the standard deviation of the optimised transformation model parameters in the usual way; this gave a measurement of the registration error. The error on this number (i.e. the standard deviation of the standard deviation) was estimated by assuming it to be proportional to $1/\sqrt{N}$ where $N$ was the number of voxel terms summed in the cost function. The mean and standard deviation of the final $\chi^2$ values for the 100 iterations were calculated in the usual way and divided by $N$ to produce the $\chi^2$ per degree of freedom. Finally, the mean and standard deviation of each of the main diagonal elements of the estimated covariance matrix were calculated in the usual way to give the estimated error.

# 5 Results

Figures 3 to 7 show the results from the first set of Monte-Carlo simulations, in which the level of added noise was varied. Estimated and measured covariances on the transformation parameters for each point at each level of added noise are shown, together with the $\chi^2$ per degree of freedom at the optimum. Since each registration result consists of three optimised parameters, the covariance matrix calculated from these has three degrees of freedom, and so analysis of three of its components is sufficient to provide a complete analysis; the diagonal elements i.e. the variances are obviously the most informative, and so these are the parameters shown. The graphs show the square-root of the variance i.e. the standard deviation. Since both the source and target patches were drawn from the same image volume, the intrinsic noise on this image volume is identical between the patches, and so acts as structure rather than random noise. Therefore, only the noise added during the Monte-Carlo simulation acts as random noise in the optimisation, and so a linear relationship between the error on the registration parameters, quoted as a standard deviation, and the standard deviation of the added noise would be expected (rather than a dependence in which the original image noise adds in quadrature). Such a relationship can be seen in the results.

Figures 8 to 12 show the results from the second set of Monte-Carlo simulations, in which the level of added noise was fixed at $1\sigma$ of the image noise and the patch size was varied. This symbology is the same as that used in Figs. 3 to 7.

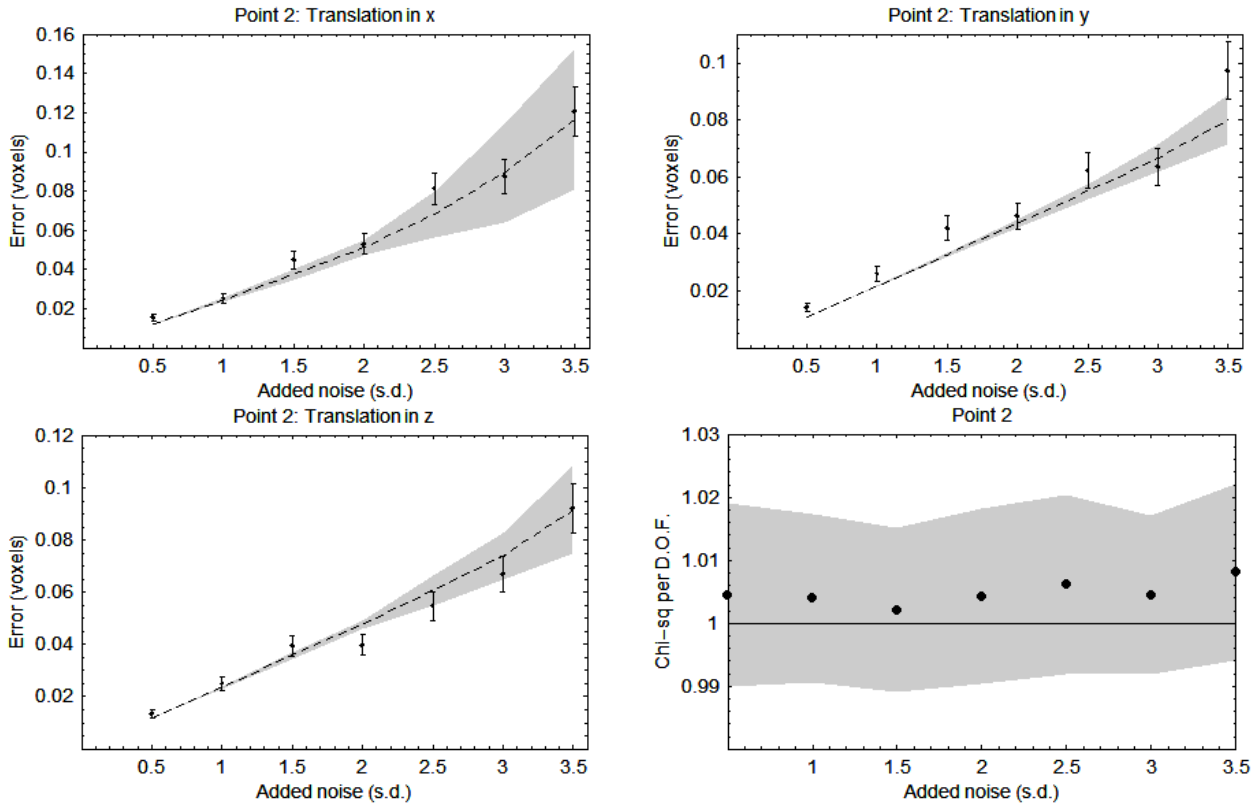


Figure 4: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 2, and the $\chi^2$ per degree of freedom at the optimum, for varying added noise.

# 6 Discussion and Conclusion

The main aim of the work presented in this document was to check the numerical stability and implementation of the proposed algorithm. The results of the Monte Carlo simulations in which added noise was varied, shown in Figs. 3 to 7, demonstrate that the method behaves as expected. The algorithm reliably converges on the correct solution to sub-voxel accuracy, although the absolute value of the accuracy is a reflection of the fact that identical volumes were being matched; in more realistic applications where the source and target volumes are different the accuracy will be correspondingly lower (see Section 6.1). The standard deviation of the errors on the optimised transformation model parameters depends linearly on the standard deviation of the added noise as expected. The $\chi^2$ per degree-of-freedom also behaves as expected; it is consistently higher than the value of unity that would result from a perfect registration with infinite accuracy, but is within about 1% of this value, demonstrating its suitability for use as a goodness-of-fit measurement. The measured errors also match the estimated errors closely and show the same functional dependency on added noise. There is some indication that the measured errors are systematically higher than the estimated errors for some parameters and points but, since the estimated error is a lower bound, this is to be expected. In interpreting these results it must be remembered that standard deviations of parameters are being compared, not parameters themselves; in general, an accuracy of within a factor of 2 is acceptable for estimated errors, and the results presented here are comfortably within that bound. The registration process is expected to destabilise at the point where the added noise dominates and the information content of the data is no longer sufficient to produce a stable cost function; the results indicate that this is beginning to occur at around 3 to $3.5\sigma$.

Results from the patch size experiments are similar; the measured and estimated errors match closely, with some indication that the measured errors are consistently higher, as would be expected from the experimental design (see Section 6.1). However, the functional dependency on patch size is the same, and the $\chi^2$ per degree of freedom is consistently higher than, but within 1% of, unity as expected. Again, the registration process is expected to destabilise as the patch size is reduced to the point where there is insufficient data to produce a stable cost function; the results indicate that this occurs at patch sizes lower than around $\pm 20$ voxels on points 1 and 2 and $\pm 10$ voxels on the other points. This difference reflects the local bone structure; points 1 and 2 are on the coronoid processes, thin spikes of bone, where there is little local structure to provide information for the registration, whereas the

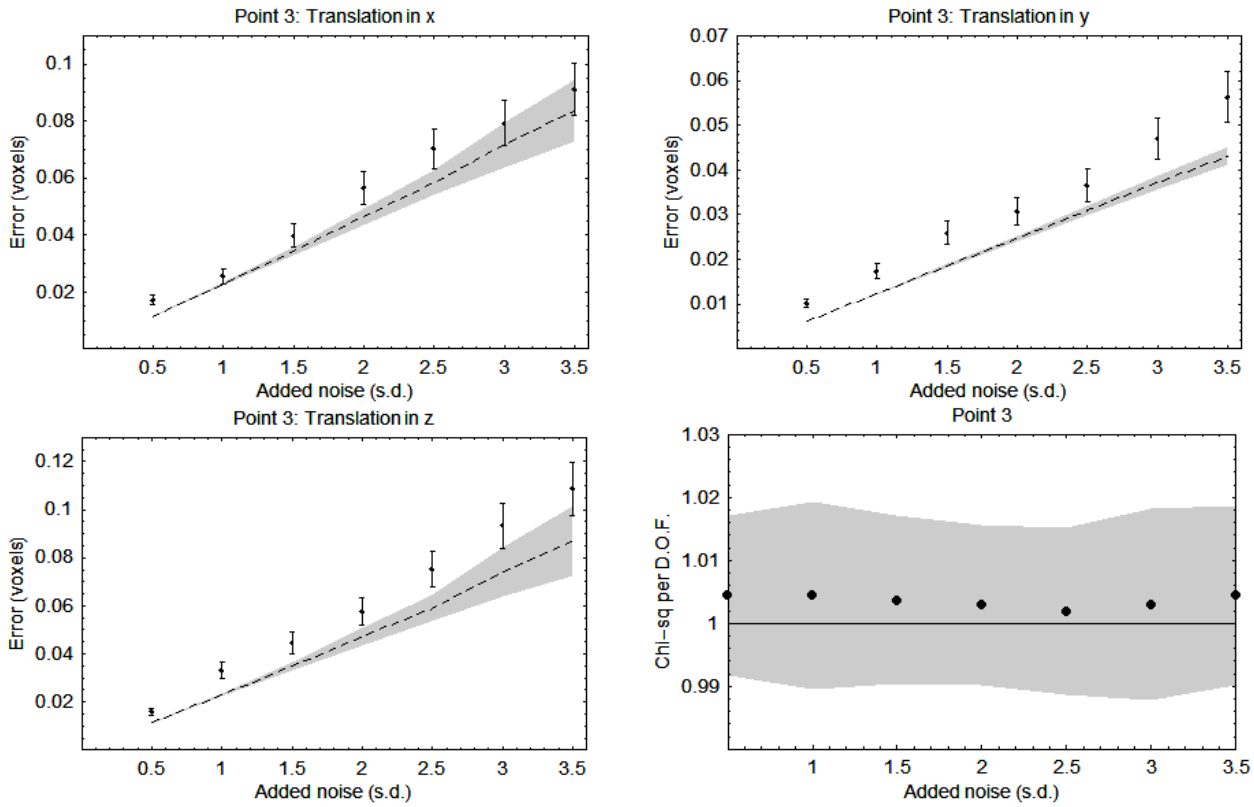


Figure 5: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 3, and the $\chi^2$ per degree of freedom at the optimum, for varying added noise.

remaining points are on smooth bone surfaces, providing more local information for the registration cost function.

On the basis of these results, we conclude that the algorithm is operating as expected, providing reliable parameter, error and goodness-of-fit estimates. The evaluation process must therefore now proceed to more realistic data sets, incorporating skulls that exhibit shape variation, in order to test its suitability for use in the intended application of automatic landmark location.

## 6.1 Limitations of the Method

The aim of image registration is to take two images of the same scene (i.e. the same underlying structures) that have been acquired with a spatial transformation between them, and to work out the parameters of that transformation such that the images are brought into alignment. In order to do this, a cost function describing image similarity is optimised over the parameters of the transformation model. A typical cost function, such as the $\chi^2$ used here, performs a direct voxel-to-voxel comparison between the source and target images. In order to do this, the voxels must be drawn from the same spatial positions. However, since a transformation is being optimised the boundaries of the voxels in the two images will not usually be the same (i.e. the voxel grids of the two images will not be aligned). Therefore, an interpolation algorithm is used to resample one of the images on the voxel grid of the other. Interpolation inevitably has an effect on image noise, as it applies some form of averaging over neighbouring voxels around the position of the voxel being interpolated, and this effect will vary according to the mis-alignment between the voxel grids.

The obvious experimental technique to use in the experiments described here would have been to add noise to the entire image volume prior to each iteration of the Monte-Carlo experiments. However, this would lead to a problem. The parameter-dependent effect of the interpolation algorithm on the added noise would have to be taken into account. The noise could be measured from the interpolated patches, but this measurement would include the effect of the intrinsic image noise (i.e. the noise added to the data during image acquisition). However, in the experiments performed here the same volume was used as both the source and the target, in order to ensure that the correct solution was known and that the images matched perfectly (apart from the added noise) at the solution. Therefore, the intrinsic noise is not random variation between the source and target patches; instead, it

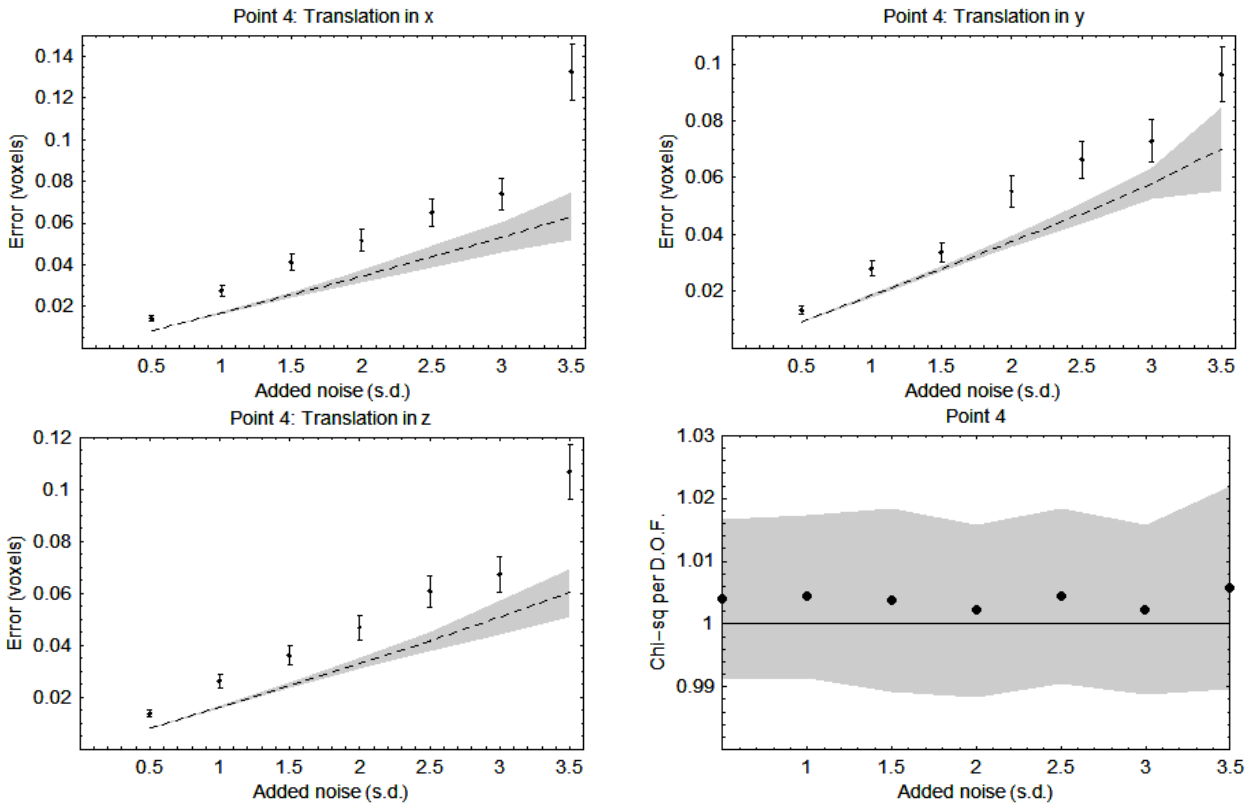


Figure 6: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 4, and the $\chi^2$ per degree of freedom at the optimum, for varying added noise.

acts like structure. Estimating the noise from the interpolated patches would therefore provide an over-estimate compared to the effective noise in the registration. Trying to take account of this effect would bring us back to the original problem; the effect of interpolation on noise is parameter-dependent.

In order to avoid this problem, noise was added to the image patches after interpolation. Error propagation then became straightforward, since the only further processing applied to each patch was differentiation. This allowed reliable estimation of the $\chi^2$ per degree-of-freedom at the solution. However, it also resulted in a numerical instability in the cost function, since the noise field added to the patches was unique for each patch generation i.e. each cost-function evaluation performed by the optimiser. The simplex optimiser used here maintains a record of the lowest cost function values it has obtained; this introduces a bias, where the optimiser can evaluate the same set of parameters multiple times and choose the lowest value of the cost function. The result will be a reduction in registration accuracy, since accuracy is dependent on the gradient of the cost function, i.e. the effect is like an additional noise term on the cost function.

Another effect of using the same volume as both the source and target, and the consequence that the intrinsic noise acts as structure rather than noise, is that the information content of the data in the experiments presented here is unrealistically large, as the algorithm can match the intrinsic noise, as well as the structures in the images. This, combined with the fact that there is no underlying shape variation between the source and target, means that the covariance estimates are unrealistically low compared to those that would be seen when matching image patches from two different image volumes. This is not a drawback given the aim of these experiments; to demonstrate that the image patch matching algorithm and its implementation were numerically stable, and that good estimates of the transformation model parameters, the covariance matrix of those parameters, and the $\chi^2$ per degree of freedom could all be obtained. However, the registration used in automatic landmark location is likely to be significantly less accurate. Furthermore, the shape of the covariance matrix would be expected to depend on the information content of the structure in the data e.g. a point on a flat, featureless bone surface will be significantly more accurately located along the direction perpendicular to the surface that along the directions parallel to it. This type of behaviour cannot be seen in the experiments presented here, due to the information content provided by the intrinsic image noise.

To conclude, the experiment performed here is the one which allows us to evaluate the ability of the algorithm to

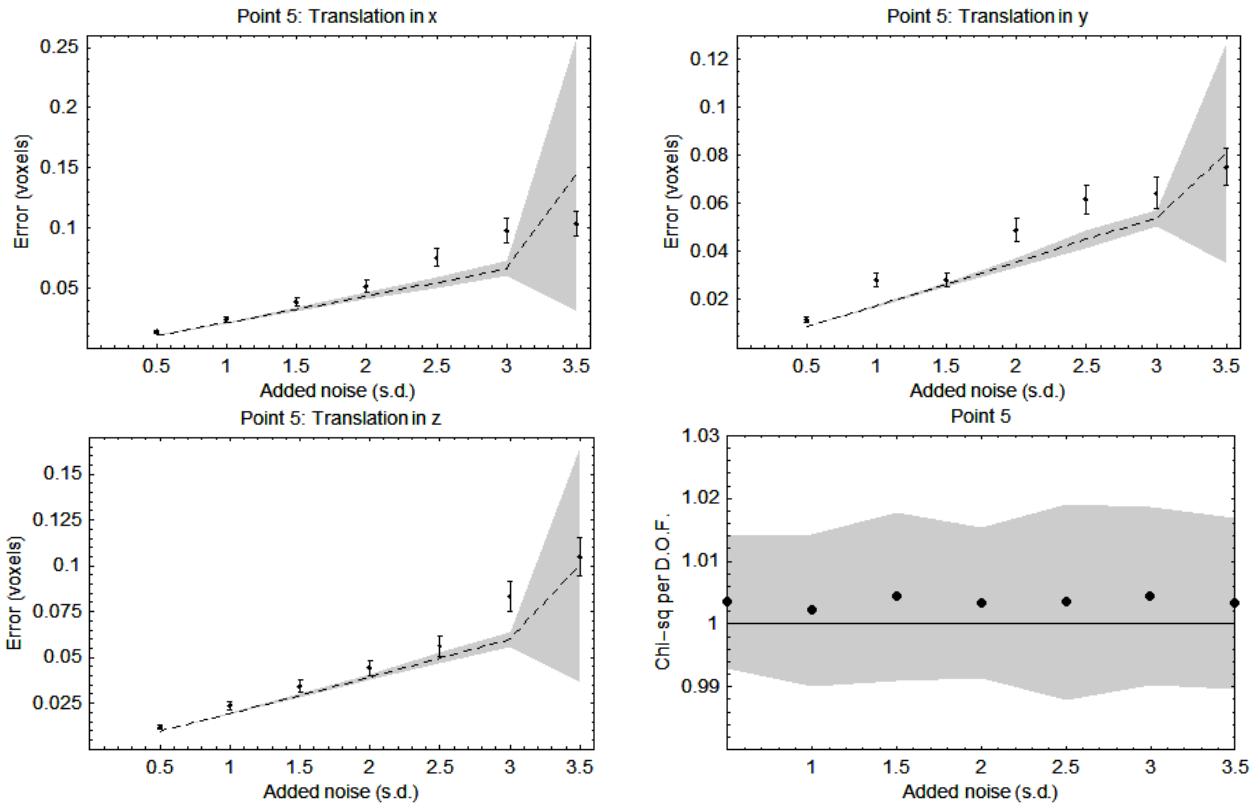


Figure 7: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 5, and the $\chi^2$ per degree of freedom at the optimum, for varying added noise.

find correct landmark locations, estimated errors and $\chi^2$ per degree-of-freedom at the solution. A better experiment would involve the use of repeated acquisitions of the same image volume, but that data is not available at this time. Finally, when the algorithm is used to match landmark locations between different skulls, a noise estimation routine will have to be added in order to directly estimate the noise on the interpolated image patches and thus the $\chi^2$ per degree of freedom; at that point, there will be no problem with the intrinsic image noise as it will not be identical across the source and target patches.

# References


[1] R Bansal, L H Staib, A F Laine, D Xu, J Liu, L F Posecion, and B S Peterson. Calculation of the confidence intervals for transformation model parameters in the registration of medical images. *Med Image Anal*, 13(2):215–233, 2009.

[2] R J Barlow. *Statistics: A Guide to the use of Statistical Methods in the Physical Sciences*. John Wiley and Sons Ltd., UK, 1989.

[3] J Berkson. Minimum chi-square, not maximum likelihood. *Ann Stat*, 8(3):457–487, 1980.

[4] P A Bromiley. TINA Memo No. 2010-007: The TINA manual landmarking tool. Technical report, Imaging Sceince and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, 2010. http://www.tina-vision.net/docs/memos/2010-007.pdf.

[5] P A Bromiley, M Pokrić, and N A Thacker. Computing covariances for mutual information coregistration. In *Proc. MIUA'04*, pages 77–80, 2004.

[6] P A Bromiley, M Pokrić, and N A Thacker. Emprical evaluation of covariance estimates for mutual information coregistration. In *Proc. MICCAI'04*, pages 607–614, 2004.

[7] P A Bromiley and N A Thacker. TINA Memo No. 2003-002: Computing covariances for mutual information coregistration 2. Technical report, Imaging Sceince and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, 2003. http://www.tina-vision.net/docs/memos/2003-002.pdf.

[8] P A Bromiley and N A Thacker. TINA Memo No. 2004-001: Empirical validation of covariance estimates for mutual information coregistration. Technical report, Imaging Sceince and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, 2004. http://www.tina-vision.net/docs/memos/2004-001.pdf.

[9] T F Cootes, G J Edwards, and C J Taylor. Active appearance models. *IEEE Trans Pattern Anal Mach Intell*, 23(6):681–685, 2001.

[10] J A Nelder and R Meade. A simplex method for function minimisation. *Computer Journal*, 7:308–313, 1965.

[11] S I Olsen. Estimation of noise in images: An evaluation. *CVGIP: Graphical Models and Image Processing*, 55:319–323, 1993.

[12] S Palaniswamy, N A Thacker, and C P Klingenberg. TINA Memo No. 2006-002: A statistical framework for detection of connected features. Technical report, Imaging Sceince and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, 2006. http://www.tina-vision.net/docs/memos/2006-002.pdf.

[13] S Palaniswamy, N A Thacker, and C P Klingenberg. Automatic identification of morphometric landmarks in digital images. In *Proc. BMVC'07, 10-13 September, Warwick, U.K.*, page 112, 2007.

[14] S Palaniswamy, N A Thacker, and C P Klingenberg. TINA Memo No. 2007-007: Automatic identification of morphometric landmarks in digital images. Technical report, Imaging Sceince and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, 2007. http://www.tina-vision.net/docs/memos/2007-007.pdf.

[15] S Palaniswamy, N A Thacker, and C P Klingenberg. Automated landmark extraction in digital images - performance evaluation. In *Proc. VIE'08, July 19 - Aug 1, Xi'an, China*, 2008.

[16] S Palaniswamy, N A Thacker, and C P Klingenberg. TINA Memo No. 2008-006: Automated landmark extraction in digital images. Technical report, Imaging Sceince and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, 2008. http://www.tina-vision.net/docs/memos/2008-006.pdf.

[17] S Palaniswamy, N A Thacker, and C P Klingenberg. TINA Memo No. 2008-007: Automatic identification system for morphometric landmarks. Technical report, Imaging Sceince and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, 2008. http://www.tina-vision.net/docs/memos/2008-007.pdf.

[18] J P W Pluim, J B Antoine Maintz, and M A Viergever. Interpolation artefacts in mutual information-based image registration. *Computer Vision and Image Understanding*, 77:211–232, 2000.

[19] W H Press, B P Flannery, S A Teukolsky, and W T Vetterling. *Numerical Recipes in C*. Cambridge University Press, New York, 2nd edition, 1992.

[20] N A Thacker, P A Bromiley, and M Pokric. TINA Memo No. 2001-013: Computing covariances for mutual information co-registration. Technical report, Imaging Sceince and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, 2001. http://www.tina-vision.net/docs/memos/20101-013.pdf.

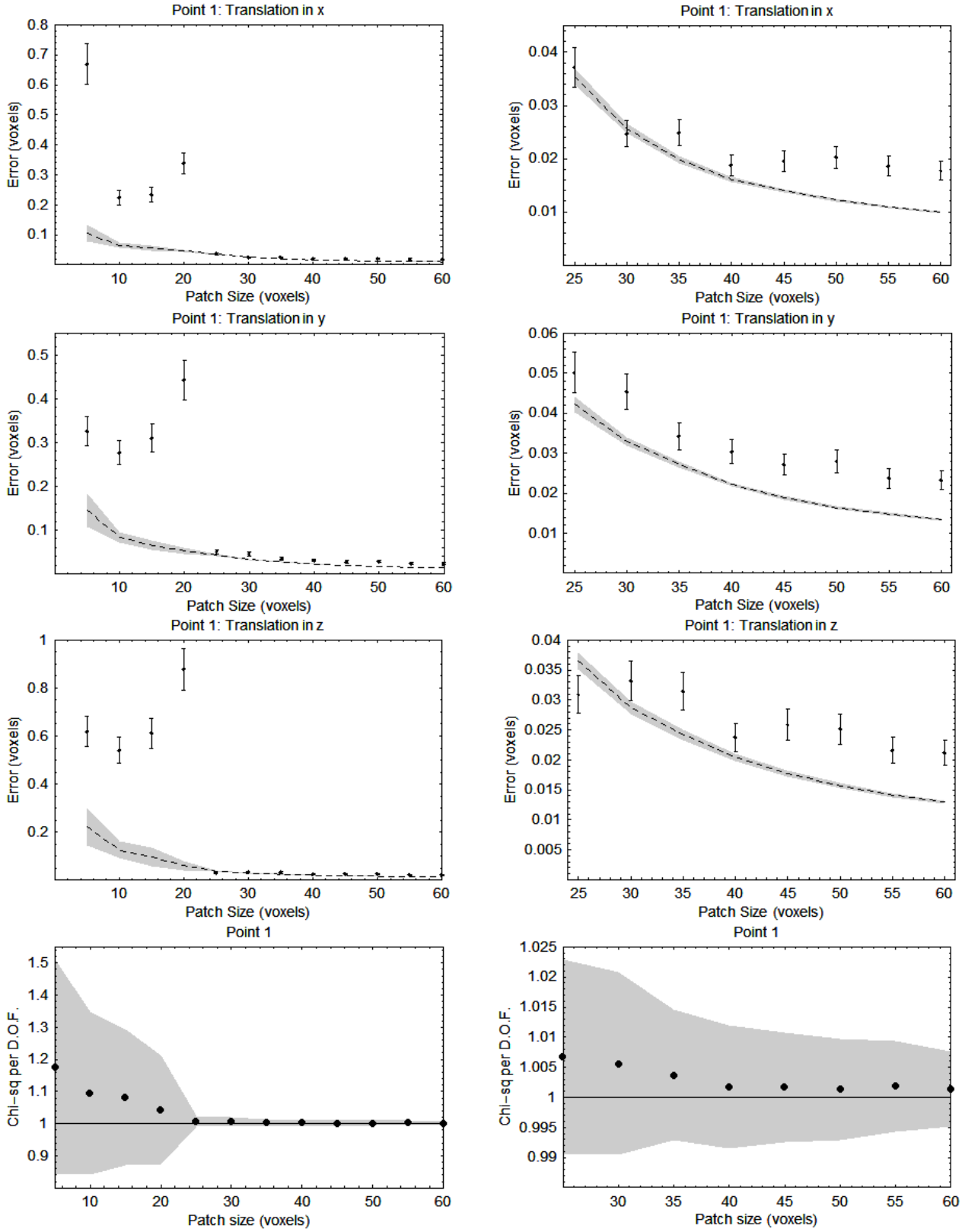


Figure 8: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 1, and the $\chi^2$ per degree of freedom at the optimum, for varying patch size. The right-hand images show expanded views of the higher patch size regions of the left-hand images.

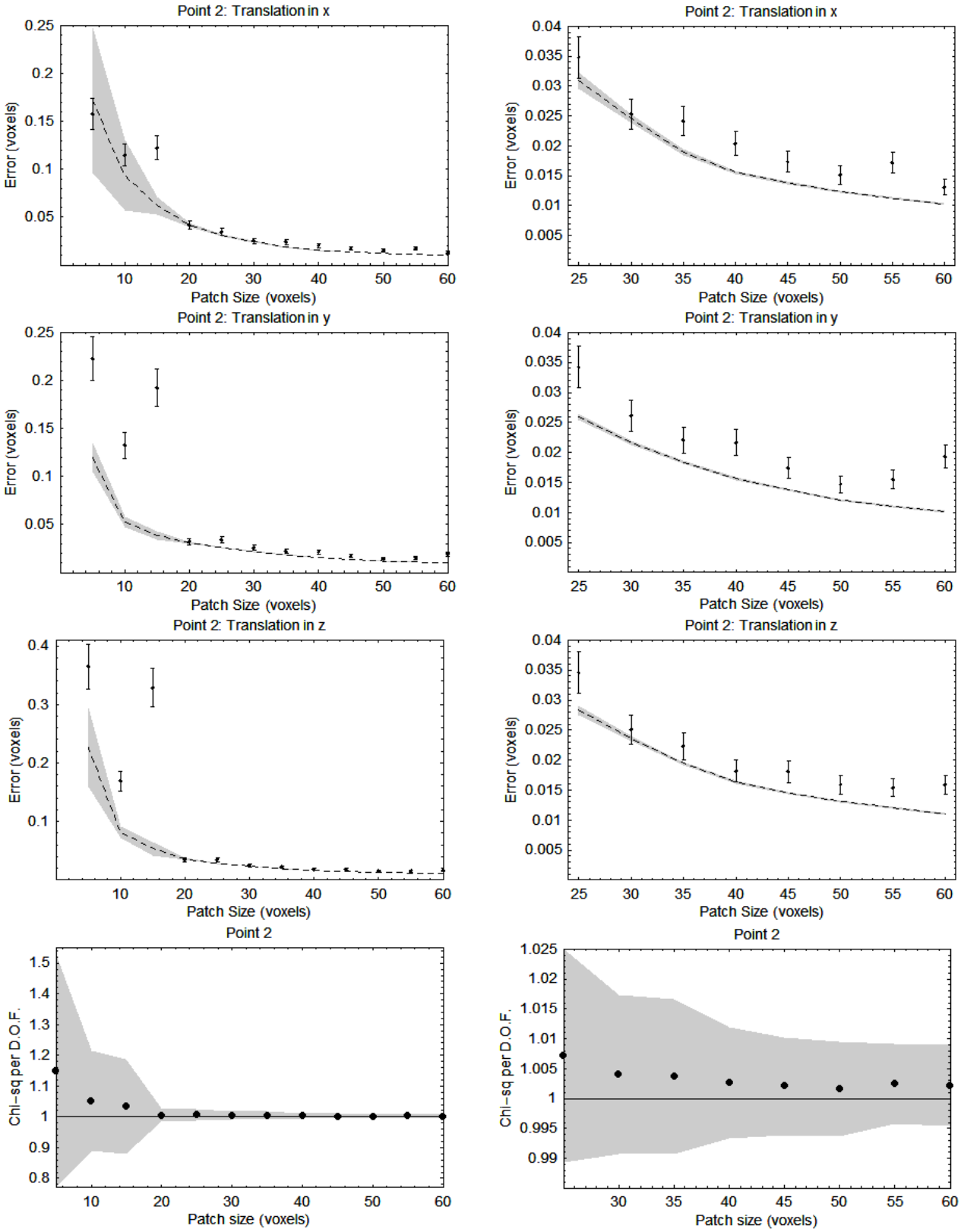


Figure 9: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 2, and the $\chi^2$ per degree of freedom at the optimum, for varying patch size. The right-hand images show expanded views of the higher patch size regions of the left-hand images.

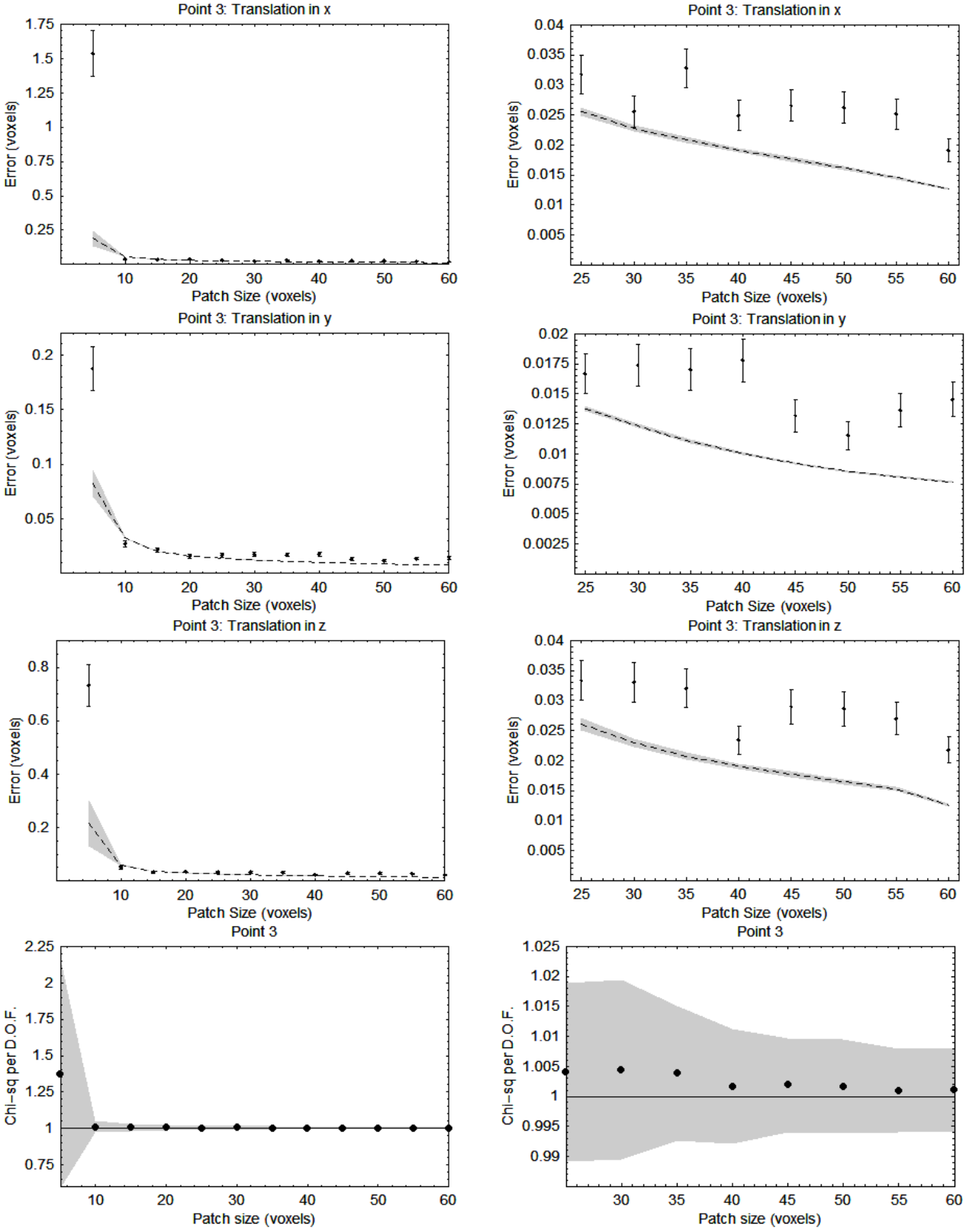


Figure 10: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 3, and the $\chi^2$ per degree of freedom at the optimum, for varying patch size. The right-hand images show expanded views of the higher patch size regions of the left-hand images.

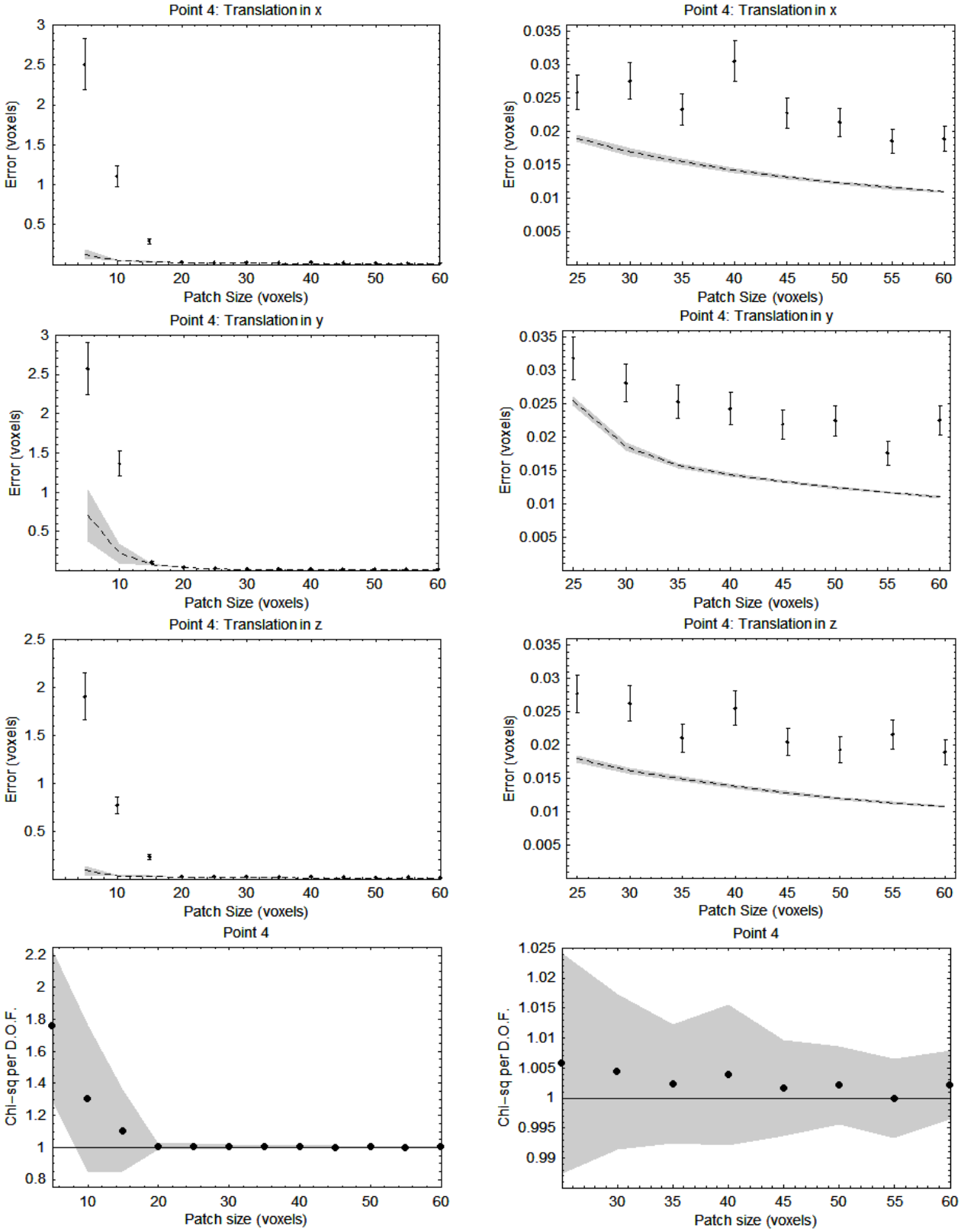


Figure 11: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 4, and the $\chi^2$ per degree of freedom at the optimum, for varying patch size. The right-hand images show expanded views of the higher patch size regions of the left-hand images.

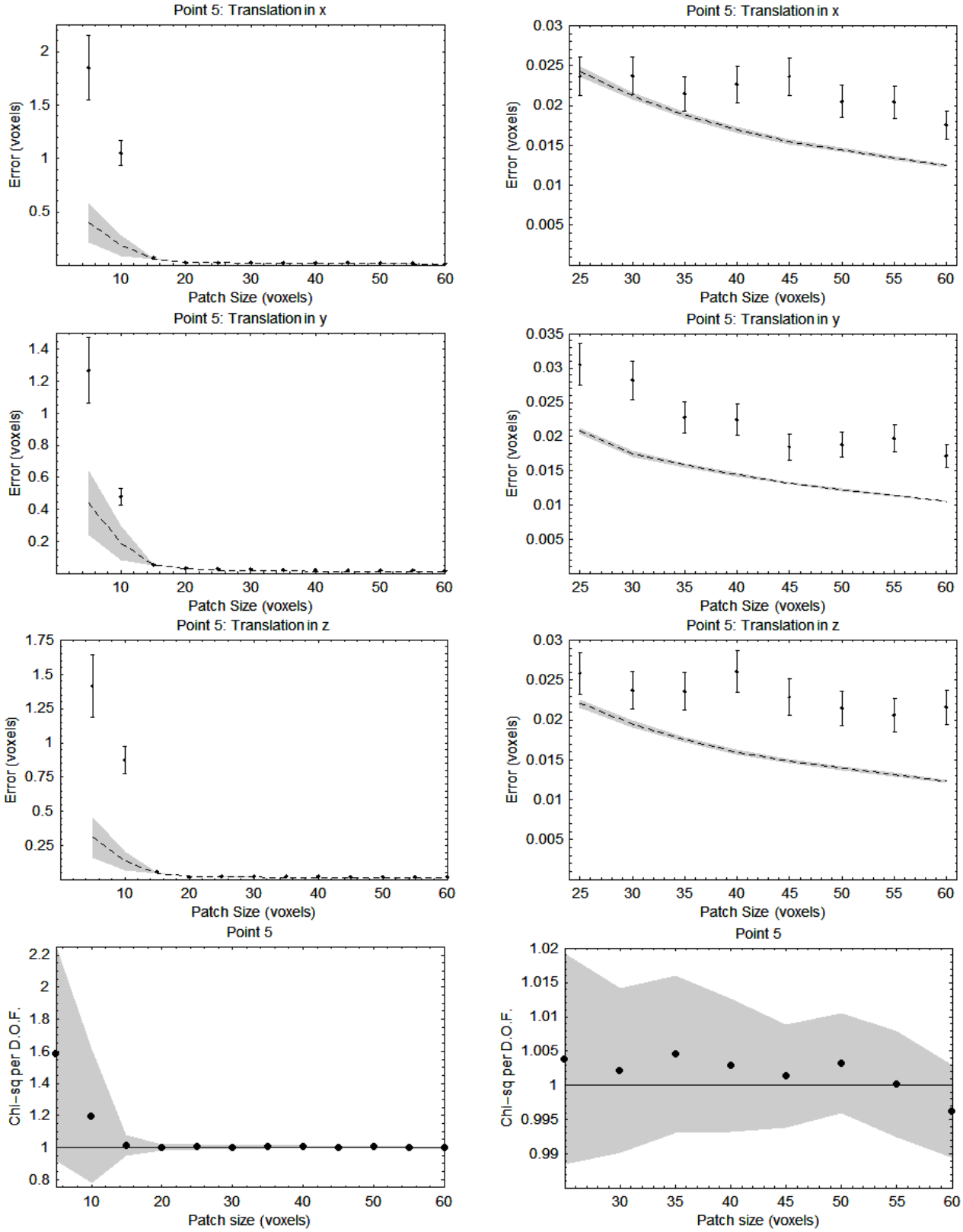


Figure 12: Measured (points with $1\sigma$ error bars) and estimated (line; grey region shows $1\sigma$ envelope) errors on the translation parameters for point 5, and the $\chi^2$ per degree of freedom at the optimum, for varying patch size. The right-hand images show expanded views of the higher patch size regions of the left-hand images.