Available online at www.sciencedirect.com
SCIENCE DIRECT®
Pattern Recognition 38 (2005) 2587–2601
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Multiobjective algorithm parameter optimization using multivariate statistics in three-dimensional electron microscopy reconstruction

C.O.S. Sorzano[a,b,*], R. Marabini[a,c], G.T. Herman[d], J.M. Carazo[a,c]

[a] *National Center for Biotechnology (CSIC), Campus Universidad Autónoma s/n, 28049 Cantoblanco, Madrid, Spain*
[b] *Engineering Division, Univ. San Pablo-CEU, Campus Urb. Montepríncipe, 28668 Madrid, Spain*
[c] *Department of Computer Science, ETSII, Campus Univ. Autónoma de Madrid s/n, 28049 Cantoblanco, Madrid, Spain*
[d] *Department of Computer Science, The Graduate Center, The City University of New York, New York, NY 10016-4309, USA*

Received 16 March 2004; received in revised form 7 March 2005; accepted 7 March 2005

Abstract

Many algorithms require the tuning of parameters in order to achieve optimal performance. Usually the best values of these parameters depend on both the particular conditions under which the experimental data have been acquired and the kind of information that we aim to obtain. The performance of an algorithm can be measured by means of numerical observers called Figures of Merit (FOMs). Usually there are no analytical formulas expressing the dependence of the FOMs on the parameters, but the nature of such dependence can be observed by the use of computational experiments. This article proposes a methodology for assigning values to the algorithmic parameters in the presence of a high number of FOMs. A multiobjective optimization framework is provided that identifies a set of optimal parameter values whose performance, from several points of view based on the initial FOMs, is statistically indistinguishable. This methodology is illustrated by applying it to the three-dimensional reconstruction (using an algebraic reconstruction technique) of single particles in electron microscopy.

Keywords: Multiobjective optimization; Multivariate statistics; Electron microscopy; Tomography; Figures of merit

1. Introduction

The problem of tuning parameters of algorithms arises in many applications including image processing [1,2], machine learning [3], statistics [4], numerical methods [5], etc. The problem is usually stated as one of finding a set of parameters that optimizes one or several objective functions. Sometimes the objective functions are provided analytically, but more typically they are expressed in terms of Figures of Merit (FOMs) that measure the usefulness of the algorithm for obtaining some specific information when applied to a given data set.

In such a case, good values of the parameters can be identified by simulation experiments in which the algorithm is applied, with some values of the parameters, to a representative data set and the dependence of the usefulness of the output on the parameter values is evaluated using the FOMs [6]. In the last sentence the phrase "representative data set" is very important. The range of good values of the parameters is usually dependent on the experimental conditions under which the data are obtained. This is illustrated for example in Ref. [7]. There, the problem of reconstructing a three-dimensional model of a macromolecule from projection images acquired by an electron microscope is addressed. The

* Corresponding author. CSIC, Biotechnology Unit, Campus Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain. Tel.: +34 91 585 4510; fax: +34 91 585 4506.
E-mail address: coss@cnb.uam.es (C.O.S. Sorzano).

0031-3203/$30.00
doi:10.1016/j.patcog.2005.03.013

reconstruction algorithm used has a free parameter called *relaxation parameter*. It is shown that for some level of noise in the data collection the optimal value (for some FOM) of the relaxation parameter is around 0.020 while for a higher level of noise it is around 0.015 (see plots 1a and 1b in Ref. [7]). However, even for a particular set of experimental conditions, there is an infinite ensemble of data sets which may be produced and, so, the reliability of the simulation experiments for specifying the values of the parameters mandates a proper sampling of this ensemble.

Inappropriate selection of algorithm parameters can lead to incorrect conclusions, as is shown in Ref. [8]. In that work, a blurring in the vertical direction (related to a preferred particle projection orientation) was found. However, it can be avoided by a proper selection of the values of the parameters used in the reconstruction process, as is demonstrated in Ref. [9].

The major novelty in this paper is a treatment of the general approach indicated above in the presence of a large number of FOMs. This is motivated by the fact that the output of an algorithm may be used for achieving a number of different (and different kind of) tasks and it is desirable to select the parameters in such a way that performance is satisfactory from the combined point of view of all the FOMs. Notice that if only one task must be addressed, a single FOM would be measured and the problem loses its multiobjective nature.

This problem can be formulated as one of multiobjective optimization for which many algorithms have been devised [10,11]. However, we are unable to make use of them for the following reasons: some of the algorithms need objective functions that are expressed by a formula in the parameters, for us the number of FOMs is high (around 30) as compared to the two or three usually dealt with by the standard multiobjective optimization algorithms, the FOMs for a given combination of parameters are randomly distributed due to the random noise present in the data set, the FOMs are evaluated by time-consuming simulation experiments (this in practice excludes methods based on genetic algorithms), and the solutions sought by us are not any Pareto optima (this concept is defined below) while many of the existing multiobjective optimization techniques are specially designed with that criterion in mind. The method we propose in the following provides a statistically based optimization framework that is applicable to the specifics of our problem.

Our procedure can be briefly described as follows. For each simulation (a specific choice from the ensemble of data sets and a specific choice of the parameters) a vector is formed whose components are the values of the various FOMs. A dimensionality reduction is performed on these vectors by identifying a more compact set of FOMs, called principal trends, that are linear combinations of the original ones. Then a least-squares curve is fitted to the average of the trends (an average is used since no a priori information is available about the relative importance of the original FOMs). The maxima of this curve are marked. Finally, optimal intervals of the parameters are computed that, in a statistical sense that will be made clear below, capture the optima of the principle trends.

A secondary novel contribution in our paper is the following. As we have discussed already, our basic approach is appropriate for identifying optimal ranges of the parameters for one set of experimental conditions (this situation will be referred to as *intra-experiment* in the rest of the paper). In what follows, we will also discuss what we call *inter-experiment* analysis. By this we mean the investigation of what aspects of the experimental conditions actually influence the optimal range of parameters and, if there is such an influence, the mathematical nature of its form. (For example, one can ask for some analytic expression that approximately indicates the dependence of the upper limit of the optimal range of some parameter on the level of noise in the data.) We use analysis of variance (ANOVA) to identify the aspects of the experimental conditions that influence the optimal range and nonlinear regression to find analytical expressions for such an influence. The formulas so obtained can be used to predict optimal ranges of the parameters for additional experimental conditions without explicitly simulating them.

Avcibas et al. [12,13] proposed a statistical multivariate approach in the presence of multiple of FOMs in the context of steganalysis similar to the first stages (although with different statistical techniques) of our intra-experiment analysis. One of the goals of in these two papers, as well as in ours during the intra-experiment phase, is to condense as much information as possible into a few number of FOMs (referred to as supermetrics in Ref. [12]). The steps to produce these representative FOMs can be summarized as: (1) remove the non-relevant FOMs, (2) cluster the remaining FOMs and (3) build a representative FOM. The detection of non-relevant FOMs is done by Avcibas et al. and in this paper using ANOVA. Avcibas et al. cluster the remaining FOMs using self-organizing maps (SOMs) while we use hierarchical classification. Finally, Avcibas et al. build their supermetrics using linear regression while we employ principal component analysis for this task. However, our final goal (optimization of an algorithm performance) is different and, therefore, our intra-experiment case performs a further step of interval selection using Student's *t*-tests.

We illustrate our methodology by applying it to the algebraic reconstruction technique with blobs (ART) [14] as it is used in 3D electron microscopy (3D EM). 3D EM is an experimental technique for obtaining the 3D structure of biological macromolecules [15,16]. Projection images are acquired using an electron microscope and ART produces a spatial Coulomb potential distribution of the particle under study. Previously published approaches to parameter selection for ART [6,9,17–20] tune the parameters according to one particular FOM. However, the selection of a single training FOM produces a bias in the parameter choice, since the optimization process only regards one aspect of the reconstructed volume. This can be observed in Ref. [9], where a

wide optimal range for the reconstruction algorithm's relaxation parameter was found, but some of the reconstructions were clearly inferior to other volumes reconstructed using a relaxation parameter picked from that same "optimal" range. This is related to the fact that the training FOM used in that work was unable to detect these differences (while they can be captured by other FOMs). Therefore, the proposed multiobjective optimization methodology provides a mean for selecting the parameters considering all possible points of view expressed by an inclusive FOM set.

In Section 2 we establish the problem framework: a general overview of the two optimization problems to be solved as well as the necessary background for the particular case of 3D reconstruction in electron microscopy are provided. Section 3 describes our optimization algorithm for the intra-experiment and inter-experiment problems. Results obtained by applying these algorithms to the particular case of the ART+blobs in 3D EM are presented as the different algorithmic steps are introduced. Finally, Section 4 discusses the results and draws important parameter-selection rules specific to the reconstruction algorithm used as an example.

2. Problem description

In this section, a general overview of the problem to be solved, as well as of its specific particularization to the case of 3D EM of single particles, is presented.

2.1. General overview

2.1.1. Intra-experiment (minor changes)

The problem addressed by an "intra-experiment" optimization can be formulated as follows: which are the values of the parameters of a given algorithm that achieve a certain trade-off among a set of different quality measures (FOMs) under various experimental conditions? This problem is called intra-experiment because the experimental conditions are fixed while some random nature of the experiment (minor changes) is considered.

To answer this question a number of simulation experiments are run following the general methodology proposed in Ref. [6], which can be outlined as follows. First a number of tasks are defined; these, in the case of 3D EM, should be related to the type of information that we intend to extract from the reconstructed volume. Then several realizations from a statistically defined set of phantoms (in the case of 3D EM, artificial volumes) are created. These phantoms must resemble, in some way, the real objects of interest. After that, the algorithm under study is run several times varying randomly the set of noise variables (i.e., those variables that cannot be controlled in a real-life experiment) and using each time a different set of parameters. In the case of 3D reconstruction, the process is simulated by projecting the phantoms and making a 3D reconstruction from the simulated data. Finally, the degree of accomplishment of the defined tasks is measured using numerical observers called FOMs. These FOMs compare the input of the simulation process (the phantoms) with its outputs (in the case of 3D EM, the reconstructed volumes). Several FOMs can be defined to measure the usefulness of the reconstruction for various tasks.

Let us now consider a fixed set of experimental conditions. Let $\mathbf{e}$ denote a specific choice from the corresponding ensemble of data sets and $\mathbf{f}$ denote a specific choice of the parameters. We use $\Phi(\mathbf{e}, \mathbf{f})$ to denote the vector whose components are the values of the various FOMs for this $\mathbf{e}$ and $\mathbf{f}$. Let $\Phi(\mathbf{f}) = (\Phi_1(\mathbf{f}), \Phi_2(\mathbf{f}), \ldots, \Phi_n(\mathbf{f}))$ be the average over $\mathbf{e}$ of $\Phi(\mathbf{e}, \mathbf{f})$; this is an approximation of the expected value of $\Phi(\mathbf{e}, \mathbf{f})$ for a fixed $\mathbf{f}$. We seek an $\mathbf{f}$ which "optimizes" $\Phi(\mathbf{f})$. One possible interpretation is to find the Pareto-optimal set of those $\mathbf{f}$ that are not inferior to any other choice of the parameters (a point $\mathbf{f}_1$ is inferior to $\mathbf{f}_2$ if and only if, for $1 \leqslant i \leqslant n$, $\Phi_i(\mathbf{f}_1) \leqslant \Phi_i(\mathbf{f}_2)$ and, for some i, $\Phi_i(\mathbf{f}_1) < \Phi_i(\mathbf{f}_2)$). However, for our kind of problem, it is usually the case that all reasonable choices of the parameters $\mathbf{f}$ are Pareto-optimal (see Section 3.1). It seems appropriate to select among these $\mathbf{f}$ only those that lead to a certain compromise value in all objective functions. Such solutions are known in the literature as *middling* solutions and are typically avoided by existing algorithms [11]. Our solution is presented in Section 3.1.

2.1.2. Inter-experiment (major changes)

The intra-experiment optimization addresses the problem of optimizing the algorithm parameters under fixed experimental conditions. However, if the latter are changed (major changes) the optimal algorithm parameters will be different. Two questions arise at this point: what are the experimental variables having an effective influence on the optimal algorithm parameters? And, is it possible to establish a model of the dependency of the optimal algorithm parameters with the experimental variables?

The answer to the first question is a well-established field of the multivariate statistical analysis. It is called "experiment design" [21] and, in particular, we will make use of the technique known as MANOVA [22]. The second question is handled by nonlinear regression analysis [21–23].

2.2. 3D electron microscopy

The objective of 3D EM is to elucidate the 3D structure of a macromolecular complex under study based on its projection images that have been acquired by an electron microscope. These images are the input to a reconstruction algorithm [24] that produces a reconstructed volume. The objective functions in this problem are related to the fidelity and resolution achieved by the reconstruction.

As discussed above, we distinguish between two kinds of changes in the experimental variables: we may think of them as *major* and *minor*. In practice, experimental variables may differ in a major way from each other; for example, in

the number of projections, in the typical size of the particle to be reconstructed, in the desired number of voxels in the reconstruction (either because of a change in the sampling rate or because of a real size difference in the particle being reconstructed), in the sample preparation technique (cryomicroscopy or negative staining) that mainly affects the correlation structure of the noise, in the microscope defocus, as well as in the data collection geometry (random conical tilt, single axis series, uniform angular coverage, etc.) [15,16]. It is not our aim to solve the parameter optimization problem for situations that involve major changes of the experimental variables (inter-experiment case). Rather, our attitude is that different ranges of the parameters are likely to be optimal in the presence of major changes of the experimental variables. On the other hand, if an ensemble of data sets contains only minor changes in the experimental variables (small differences in particle size, in the level of noise, or in the angular distribution of projections), then it is reasonable to search for a single optimal range (intra-experiment situation).

The simulation process that we present in this paper for this problem takes a phantom volume (a protein known at atomic resolution), makes projections of it and applies the reconstruction algorithm with various values of one of its parameters. The FOMs compare the discrepancy between the reconstructed volume and the phantom. Our aim is to find an optimal range of the parameter in the presence of multiple FOMs.

2.3. ART: one of the approaches to 3D reconstruction

We illustrate our ideas on parameter optimization on a particular 3D reconstruction algorithm, namely an algebraic reconstruction technique with blobs (ART+blobs) as is applied in electron microscopy [14]. This method is a series expansion method [24]; i.e., it is assumed that a volume $f(\mathbf{r})$ ($\mathbf{r} \in \mathbb{R}^3$) can be approximated by a linear combination of a finite set of known basis functions b_j, each one of which is just the same function b shifted to one of J grid points (denoted by $\mathbf{g}_j$), as in

$$f(\mathbf{r}) \approx \sum_{j=1}^{J} c_j \cdot b_j(\mathbf{r}) = \sum_{j=1}^{J} c_j \cdot b(\mathbf{r} - \mathbf{g}_j) \tag{1}$$

and the task of the algorithm is to estimate the unknown coefficients c_j. For a review on ART, see Ref. [25]. A consequence of the volume series expansion is an image formation model of the form:

$$y_i \approx \sum_{j=1}^{J} l_{i,j} c_j, \tag{2}$$

where y_i is the ith measurement of the volume to be reconstructed (that is, a pixel value in the experimental data) and $l_{i,j}$ is the corresponding line integral of the basis function b_j. The values y_i and c_j form a MN-dimensional vector and a J-dimensional vector, respectively (which we will denote by $\mathbf{y}$ and $\mathbf{c}$), where N is the number of projections and M is the number of pixels per projection. The y_i elements are arranged in such a way that all pixels belonging to a projection are consecutive.

Following Lewitt [26] and Matej and Lewitt [27] we select b to be a generalized Kaiser–Bessel window function (also referred to as a *blob*) and for the grid points we use a finite subset of the so-called body-centered cubic grid. Such a representation was found useful in electron microscopy applications (see for example Ref. [14]). The specific choice that we adopted for the blob and the grid is the one referred to as the "standard blob" in Ref. [28].

Notice that $\mathbf{c}$ is the array of coefficients weighting the various blob functions. These coefficients are the ones estimated by the reconstruction algorithm described below. However, in order to evaluate the FOMs defined in Section 2.4, we evaluate the reconstructed volume in a set of *voxels* (little cubes used as basis functions) distributed on a cubic grid. We use Eq. (1) to convert to voxels the volume expressed in blobs.

The particular variant of ART that we use as the basis of our algorithm operates as follows. Starting with a J-dimensional zero vector for the estimate of $\mathbf{c}$, we update this estimate of $\mathbf{c}$ iteratively. In an iterative step we make use of data from one projection only; we cycle through all the projections once in the complete algorithm. The update of the jth component of the estimate of $\mathbf{c}$ is done as described by

$$\mathbf{c}_j^{(k+1)} = \mathbf{c}_j^{(k)} + \frac{\lambda}{N_j^{(k)}} \sum_{i=kM+1}^{(k+1)M} \frac{y_i - \langle \mathbf{l}_i, \mathbf{c}^{(k)} \rangle}{\|\mathbf{l}_i\|^2} l_{i,j}, \tag{3}$$

where λ is a real number called the relaxation parameter that controls the magnitude of each update, $\mathbf{l}_i$ is the J-dimensional vector whose jth component is $l_{i,j}$, $\langle \cdot, \cdot \rangle$ denotes the inner product (dot product) between two vectors, and $N_j^{(k)}$ is the number of i's satisfying $kM + 1 \leqslant i \leqslant (k+1)M$ for which $l_{i,j}$ is different from 0. The output of the algorithm is $\mathbf{c}^N$.

In this algorithm, the only parameter that is not specified is the relaxation parameter λ. We defined things in this fashion to simplify the illustration of our approach to parameter optimization. (We could have avoided specifications of parameters defining the blob and the number of cycles through the data; this would have resulted in having to optimize all these parameters simultaneously. Our approach is capable of doing that, but its description and its illustration would become more cumbersome.)

2.4. FOMs

It is common to measure the goodness of a reconstruction by evaluating the average squared error between the reconstruction and the phantom. Calling p_i and r_i the values

assigned to the ith voxels of the phantom and the reconstruction, respectively, this FOM can be defined as

$$scL2 = 1 - \frac{1}{N_V}\sum_{i=1}^{N_V}\left(\frac{p_i - r_i}{2}\right)^2,$$

where N_V is the total number of voxels in the volume. The name $scL2$ comes from the fact that we are measuring the *structural consistency* by evaluating the l_2-norm of the error. However, we could have measured this error only within the surface defining our region of interest (also called feature or particle) or outside this surface. In this way, we can get an idea of how this error is distributed over the volume (whether it concentrates inside or outside the particle). Two new FOMs can thus be defined:

$$scL2(B) = 1 - \frac{1}{N_B}\sum_{i\in B}\left(\frac{p_i - r_i}{2}\right)^2,$$

$$scL2(F) = 1 - \frac{1}{N_F}\sum_{i\in F}\left(\frac{p_i - r_i}{2}\right)^2,$$

where B is the set of indices of the voxels in the *background* and F is the set of indices of the voxels within the *feature*. Furthermore, we could be interested in knowing whether the error outside the reconstructed particle is concentrated close to the particle or far from it. For this purpose the error can be weighted according to the distance of the ith voxel in the phantom from the main feature. This distance is referred to as d_i. This FOM should be able to detect extra masses appearing far from the desired particle, as in

$$scap = 1 - \frac{1}{N_B}\sum_{i\in B} d_i\left(\frac{p_i - r_i}{2}\right)^2.$$

Much more complicated FOMs can be defined. In Ref. [9] we presented 24 reasonable FOMs; all of them are used in this work. In addition, we also considered the correlation between the two volumes (called *sccorr*), the mutual information (*scinf*, this FOM has been widely used as a similarity function for bringing volumes into registration [29]) and resolution measured as the frequency at which the Fourier shell correlation (*scresol*) [30] falls below 0.5 (this is a standard measure of quality in the 3D EM field). We evaluated each reconstruction by each of these FOMs, thus obtaining 27 FOMs per reconstruction.

3. Parameter optimization and its results

For the sake of clarity, in the following description of the optimization procedure, each of its steps is immediately followed by the experimental results obtained in that step. In the experiments reported below, we will consider major changes in the sampling rate, in the number of projections, and in the characteristics of noise. In addition, with these experimental variables fixed at their various values, we will illustrate our parameter optimization methodology under circumstances that allow minor changes in the angular distribution of the projection directions, the size of the phantoms, and the noise realization. Section 3.1 shows the application of the proposed methodology through an example of *intra-experiment* optimization, while Section 3.2 describes the *inter-experiment* analysis.

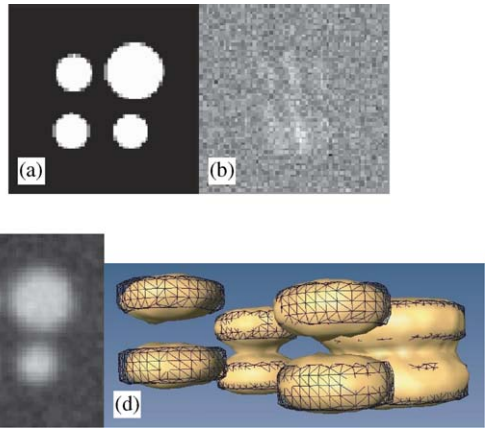

Fig. 1. (a) A horizontal slice through one of the phantoms used in the experiments. (b) One of the noisy projections of this phantom. (c) Corresponding slice through the reconstruction from the noisy projections. (d) Three-dimensional display of the noisy reconstruction: yellow indicates the surface of the reconstruction, the black wire-mesh indicates the surface of the phantom.

3.1. Intra-experiment optimization (minor changes)

Intra-experiment optimization deals with the problem of selecting optimal parameters for the algorithm performance for a set of experimental conditions that involves only minor changes of the experimental variables. Due to these changes, uncertainties remain even with the choice of a representative data set, and so rather than selecting a single optimal value for the parameter of the reconstruction algorithm, a range of values is chosen. The performance associated with the values within this range are statistically indistinguishable from each other. This idea was already suggested in Ref. [19] and was further developed in Ref. [9].

The data of the example shown in this section corresponds to the optimization of the ART relaxation parameter when the ART algorithm is applied to the reconstruction of phantoms of size $65 \times 65 \times 65$, taking 2010 projection images with white Gaussian additive noise (these parameters correspond to the particular combination of major changes selected for the experimental variables). The phantoms used were pairs of cylinders of density 1 aligned with the vertical axis (see Fig. 1 and, for further details, Ref. [9]). Minor changes are applied in the cylinders' size, the angular distribution and the realization of the noise in the projection gray

levels. In particular, every phantom had four pairs of cylinders vertically aligned. Each pair was vertically separated by a random distance between 3 and 6 voxels. Each cylinder had a random radius between 5 and 10 voxels, and a random height between 3 and 5 voxels. The projections were evenly and randomly distributed along all possible directions. Gaussian noise with standard deviation 9 was added to the pixel values.

Ten simulations changing the noise parameters (phantom realization, projection direction set, and noise realization) were performed for each λ in the range from 0.015 to 0.150 in steps of 0.015. Twenty seven FOMs (see Section 2.2) were computed for every simulation resulting in a table with 100 experiments (rows) and 28 columns: one column with the relaxation parameter used and 27 with the corresponding FOM vector.

The proposed optimization methodology is based on statistical data analysis and it can be summarized as follows:

(1) *FOM normalization*: Normalize the FOMs so that they have comparable values.
(2) *Removal of irrelevant FOMs*: Remove all those FOMs that cannot detect differences among the various values of the parameters.
(3) *FOM clustering*: Cluster all those FOMs showing a similar dependency with the parameters.
(4) *Cluster dimensionality reduction*: Reduce the dimensionality of the clusters obtaining a single representative of each cluster.
(5) *Interval selection*: Select an optimal region for the parameters using the information contained in the cluster representatives.

The various data analysis steps are described in more detail below and have been carried out using the program Statistica [31].

3.1.1. FOM normalization

The different FOMs need not have comparable values and by normalizing we prevent high values from dominating the analysis. We are working under the assumption of no a priori knowledge about the relative importance among FOMs. One common way to achieve this goal is by normalizing the input data to become of zero-mean and unit variance (by subtracting the average of the FOM and dividing by its standard deviation). This normalization is applied independently to the columns of the data table corresponding to the FOMs. Depending on the data set, it might be necessary to perform an outlier rejection step before the FOM normalization. In the data used in this example, there were no severe outliers and, therefore, no outlier treatment was done. However, if outliers turned out to be an important source of error they might have been removed using a large variety of techniques ranging from simple stem-and-leaf plots [32] to complicated robust multivariate detectors [33].

3.1.2. Removal of irrelevant FOMs

For a particular experiment there might be FOMs that do not provide any information about the optimal performance. In other words, they do not detect any difference among the different λ's and should be removed from further analysis. These irrelevant FOMs can be identified using 1-way ANOVA studies [22] in which the independent variable is λ and the dependent variable is each one of the FOMs. If an ANOVA analysis indicates that the independent variable does not significantly affect the level of the dependent one, then the corresponding FOM is irrelevant to the particular experiment.

When this analysis is applied to each of the FOMs in this example, it reveals that there are four FOMs for which dependence on λ cannot be observed with a significance greater than 90%. The corresponding columns are removed from the data table. Fig. 2 shows one of these FOMs, the line drawn represents a distance-weighted least-squares fit of the data [34]. It should be noticed that these FOMs are rejected for this particular study but they may be accepted for optimization under other experimental conditions.

3.1.3. FOM clustering

It is quite possible that some of the remaining FOMs express the same kind of information, i.e., that several of them show the same behavior with respect to the parameter. In this step we aim at identifying the number of different behaviors within the FOMs and at clustering these according to their behavior. The behavior or tendency of a FOM with respect to the parameter can be represented in a plot of the FOM values versus the parameter value. It is interesting to recognize these groups as different tendencies as they probably point to different optimal intervals. In order to perform this group identification, the different FOM columns are used as input to a hierarchical classifier [35] and, in this case, the classification results can be validated by principal components analysis (PCA) [36]. The output of this step is a partitioning of the FOMs into clusters according to their tendency with the parameter. Although the FOMs have been clustered in this work using hierarchical classification, any other clustering method or combination of methods could have been used. The choice of the most appropriate clustering procedure may be particular to each problem. There is active research, whose discussion falls beyond the scope of this paper, on how to estimate the number of clusters in a data set. There are situations in which the number of clusters can be easily identified because the clusters are sufficiently apart. However, this is not always the case and, in most situations, no matter what algorithm was employed to estimate the number of clusters, it must be subjectively validated by the user.

In the example considered in this section, the FOMs can be grouped into two main classes: those whose value decreases with λ and those whose value increases with λ (see Fig. 3). These two groups are correctly separated using

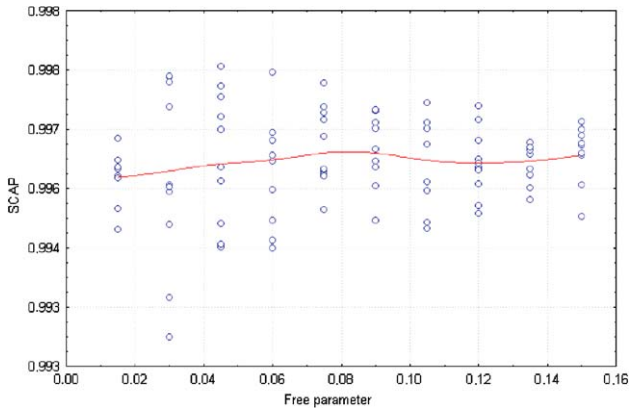

Fig. 2. Plot of the *scap* FOM (see Section 2.4) for the study described in Section 3.1. The line corresponds to the distance-weighted least-squares fit of the samples. Notice that the FOM appears to be insensitive to the choice of the parameter.

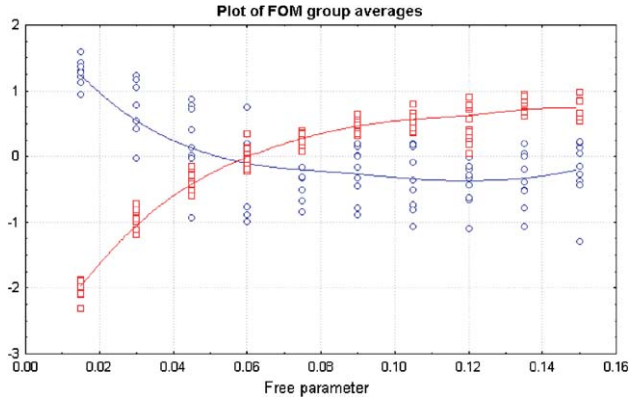

Fig. 3. Plot of the averages of the two FOM groups identified by hierarchical classification. One of the averages is represented by circles while the other by squares. To help visualization two continuous curves have been fitted to the data using DWLS [34]. Each circle (respectively, square) indicates the average FOM in one of the groups for one of the experiments (thus there should be ten circles, respectively squares, for each value of the free parameter; not all of these are visible, due to overlap.

hierarchical analysis with weighted pair-group average [37] (see Fig. 4). The existence of two groups is confirmed, in this case, by PCA whose first factor using varimax rotated axes [38] accumulates more than 79% of the total variance (two factors accumulate 87%). The existence of a factor with so much variance seems to indicate a binary separation of data. Fig. 5 represents the so-called factor loadings of the FOMs being clustered. The groups at different levels identified in the hierarchical classification are marked with the same type of line on both Figs. 4 and 5.

In this particular case, each group has a clear meaning. The FOMs that increase their value with higher relaxation factors (for example, $scL2(F)$) reflect an increase of the contrast in the macromolecular reconstruction, while those FOMs decreasing with higher relaxation factors detect an increase of the noise level in the reconstruction (for instance, $scL2(B)$). These two effects of the relaxation factor are well known within the 3D reconstruction field: if we keep the relaxation parameter low, then the reconstructions will not be very noisy but will have a low contrast, on the other hand if the parameter is high then the contrast is increased at the price of getting higher noise.

3.1.4. Cluster dimensionality reduction

At this stage the FOM groups may have a relatively high number of members resulting in a high dimensionality group (a vector containing all the information gathered by the FOMs assigned to a cluster has as many components as FOMs in the cluster). However, since the FOMs are grouped according to their similarity, the group is formed basically by some common trends disturbed by "noise." The goal of our next step is to reduce as much as possible the number of components. Eventually, we will reduce all FOMs in a cluster to a single representative of their behavior.

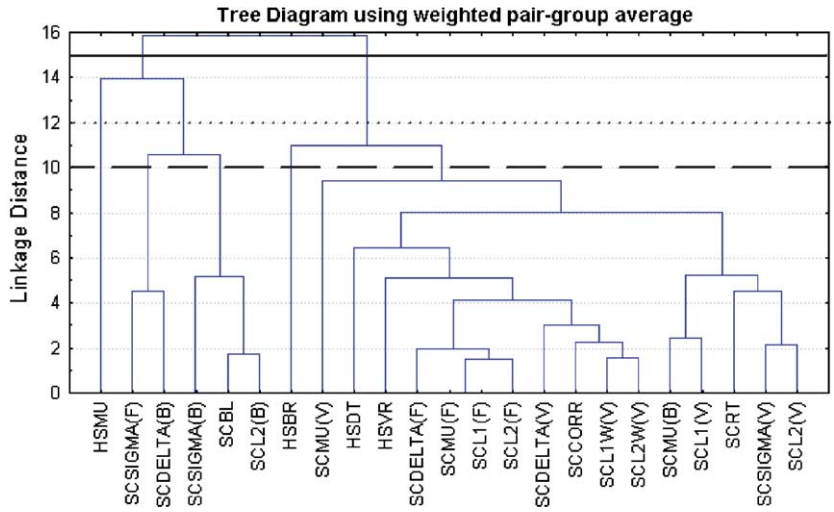

Fig. 4. Tree diagram of the FOM hierarchical classification. Three different cut levels reveal the existence of either 5 (dashed line), 3 (dotted line) or 2 (solid line) groups. Further data exploration reveals that two groups are sufficient for these FOMs.

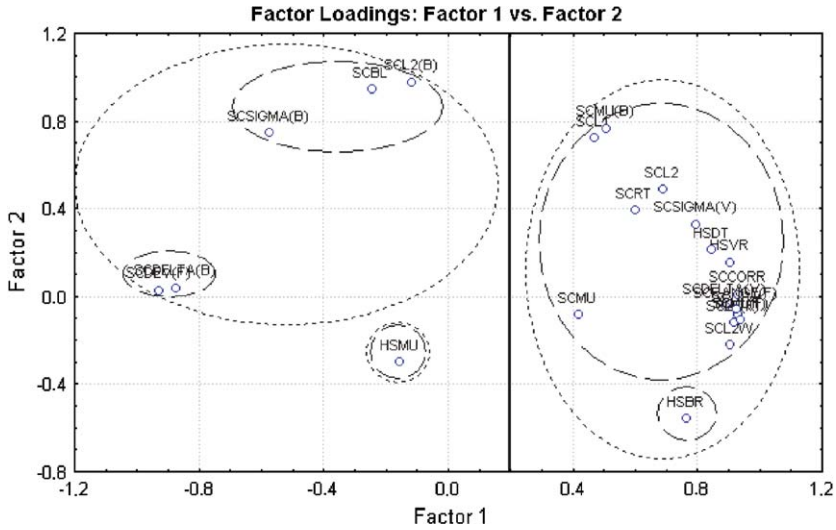

Fig. 5. Plot of the factor loadings in the first and second PCA factors. The factor loading in a PCA factor indicates how important is that PCA factor to explain the variance of a particular FOM. The groups identified by hierarchical classification at various levels are marked with the same line type as in Fig. 4.

PCA is well known for its linear dimensionality reduction properties [39]. Data subtables are formed for each group with the FOM values corresponding to FOMs belonging to that group. In the example shown in this section, two subtables with 100 experiments and 6 and 17 FOMs, respectively, are extracted. PCA is applied to each group independently. This analysis finds new basis vectors (called principal components) in the 6 and 17 dimensional spaces so that the first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The principal components are the eigenvectors of the covariance matrix of the input data. The corresponding eigenvalues reflect the amount of variance of the total data variance explained by each principal component. A common practice in data dimensionality reduction via PCA is to select as many principal components as to account for a certain amount of total data variance.

In the example shown, after analyzing each group separately, it can be found that three components on each group are enough to explain 98% of the total group variance in the decreasing FOMs group and 94% of the total group variance

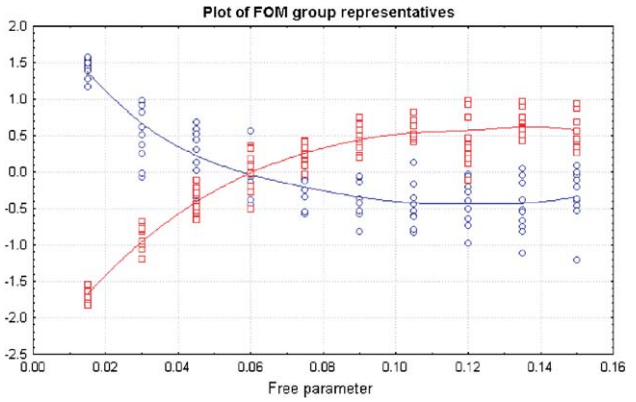

Fig. 6. Plot of the two FOM group representatives (one using circles and the other using squares). To help visualization two continuous curves have been fitted to the data using DWLS. Their behavior is similar to the average behavior within each FOM group (compare Fig. 3).

in the increasing FOMs group. Thus, the number of FOMs on each group can be safely reduced to three that capture nearly all the information.

At this stage a second reduction is proposed by combining the information of the three components according to their explained variance into a single FOM group representative. Let us assume that the P first principal components are needed to account for the selected amount of variance (in the example, 94%), and that the eigenvector pc_i explains a w_i% of the total variance. The FOM group representative is constructed as

$$FOM_G = \sum_{i=1}^{P} w_i\, pc_i.$$

In this way the number of objective functions is reduced to the number of distinct FOM tendencies. Fig. 6 shows the so-combined FOMs for both groups. It can be seen that they essentially show the average behavior of the FOMs within their group (see Fig. 3) but with a slightly smaller variance for each λ. Special care must be taken with the sign of the eigenvectors pc_i since $-pc_i$ is also an eigenvector. The sign of the eigenvectors must be chosen in such a way that FOM_G minimizes the distance between the group representative and the represented FOMs.

3.1.5. Interval selection

At this point, we have reduced the problem of selecting the parameter value achieving a compromise in the algorithm performance as measured by 27 FOMs to only two opposing objective functions. A multiobjective optimization searching for the Pareto-optimal solution within these two objective functions would say that all values of the parameters are non-inferior to the rest and, therefore, any value of the parameter would be valid. However, this is not the case since if the relaxation parameter is very high or very low then very noisy or poorly contrasted reconstructions are obtained.

The goal is to pick one of the *middling* solutions for which a reasonable performance is reflected by each of the FOM groups. For doing so, in the case of the relaxation parameter of ART, we found useful to make use of the distance-weighted least-squares (DWLS) curves [34] fitted to each of the FOM representatives (see Fig. 6). Since they show two opposing behaviors, we picked as middling solution the point where the two DWLS curves intersect.

However, it is not always the case that there are just two groups with opposing tendencies. For instance, the method of iterative data refinement (IDR) [40] has been recently applied in electron microscopy to correct for microscope aberrations [41,42]. That algorithm also has a relaxation factor to be optimized. When this was done using the statistical data analysis approach advocated above, four FOM groups were identified. In this case, we found useful to simply average the FOM group representatives producing a new FOM. Then, we select as the middling point the one that maximizes the DWLS fit of this new FOM as is shown in Fig. 7. This idea of aggregating the different objective functions is not new in multiobjective optimization [11].

An alternative procedure for selecting the middling point in the presence of more than two trends could be to chose the free-parameter that minimizes the variance of the values of the trend DWLS fits. (To clarify: for each fixed parameter value, there are as many DWLS fits as there are trends, and these values have a variance that changes with the parameter; we suggest the selection of that value for the parameter for which the variance is minimal.) Notice that our criterion of selecting the crossing point of these DWLS fits when only two trends are present is a particular case (in which the variance is zero) of this alternative procedure. The analysis of the distribution of the values of the DWLS fits at the middling point can also identify those trends that are particularly benefited or disadvantaged from the selection of that middling point. In the case of having outliers in that distribution, the reason of their existence should be checked. In any case, the selection of the middling

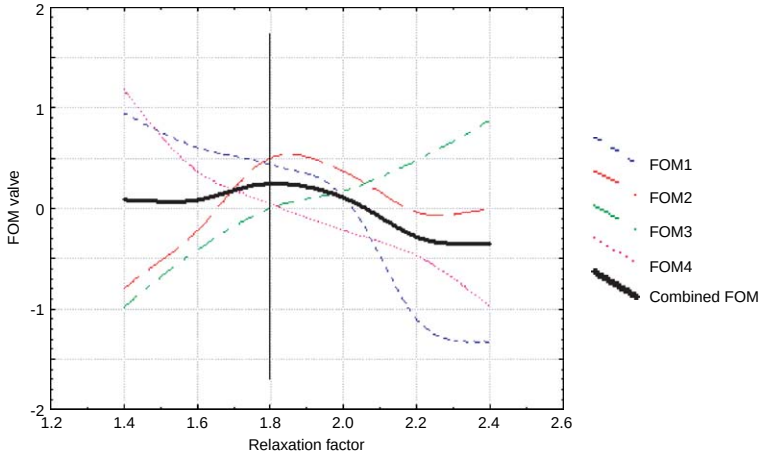

Fig. 7. Plots of the DWLS curves fitting the four FOM group representatives and the combined average FOM for the IDR application [41,42].

point is situation dependent and must be done under expert supervision.

Once the middling solution is picked, it is checked whether the surrounding solutions have a significantly different performance. For each FOM group, G, an interval, (a_G, b_G), is constructed (containing the middling point) formed by all those solutions whose representative FOM is not significantly different from the representative FOM at the middling point. For this comparison, Student's t-tests are performed among the available samples as was done in Refs. [3,9]. Finally, the *middling interval* is defined as the intersection of all the group intervals. The application of this criterion to the ART relaxation parameter in the example shown in this section gives the interval (0.045, 0.075).

3.2. Inter-experiment analysis (major changes)

The process outlined in the previous section describes a method for selecting an interval of parameters whose performance is statistically indistinguishable from the optimal middling performance found in the experiments. However, this interval is dependent on the particular selection from the possible choices of the major changes for the experimental variables. In the example of 3D EM applied to single particles, some instances of major changes are the data collection geometry, the particle size or sampling rate, the number of projection images, and the noise nature. These variables can introduce a large enough change in the simulated conditions so that the optimal interval computed under other conditions may be invalid.

Two questions arise from this point: first, given a set of major changes in the experimental variables, which ones have an important influence on the optimal interval? And, second, is the dependency of the optimal interval predictable? These two questions will be addressed in an experimental manner by determining the optimal interval for the relaxation parameter λ under various experimental conditions.

Table 1
Volume size and number of projections simulated

(65^3, 773)	(65^3,2010)	(65^3,3246)
(97^3,1153)	(97^3,2999)	(97^3,4844)
(129^3,1533)	(129^3,3988)	(129^3,6443)

In the case of the relaxation parameter of ART used in 3D EM, the experimental variables selected for major changes are the sampling rate, the number of projections and the noise nature (i.e., its correlation structure). Instead of working directly on the sampling rate we work with a related quantity: the size in voxel-units of the reconstructed volume (for a given volume size, increasing the sampling rate results in a volume with a higher number of voxels). The volume size and number of projections were chosen as shown in Table 1 and two kind of gray-level noises were studied: unfiltered white Gaussian noise (simulated up to the maximum frequency 0.5) and white Gaussian noise low pass filtered at a frequency of 0.2. The unfiltered noise gives an image aspect similar to that experimentally obtained under low defocusing and the filtered noise resembles that acquired when the sample is strongly defocused. In both cases the signal to noise ratio (SNR) is 0.33. These simulation values were validated by microscopy experimentalists.

The whole test amounts to 18 different experiments. Each experiment is treated with the intra-experiment methodology described in Section 3.1. The optimal interval limits for each experiment are shown in Table 2.

Once the optimal intervals are determined for each experiment the two questions formulated at the beginning of this section can be answered.

Table 2
ART relaxation parameters optimal intervals

Size (voxels)	Number of projections	Noise	λ_{min}	λ_{max}
65^3	773	Unfiltered	0.060	0.080
65^3	2010	Unfiltered	0.050	0.070
65^3	3246	Unfiltered	0.030	0.050
97^3	1153	Unfiltered	0.060	0.080
97^3	2999	Unfiltered	0.050	0.070
97^3	4844	Unfiltered	0.035	0.055
129^3	1533	Unfiltered	0.040	0.020
129^3	3988	Unfiltered	0.025	0.045
129^3	6443	Unfiltered	0.015	0.030
65^3	773	Filtered	0.050	0.070
65^3	2010	Filtered	0.055	0.075
65^3	3246	Filtered	0.030	0.050
97^3	1153	Filtered	0.055	0.075
97^3	2999	Filtered	0.050	0.070
97^3	4844	Filtered	0.025	0.045
129^3	1533	Filtered	0.050	0.070
129^3	3988	Filtered	0.020	0.040
129^3	6443	Filtered	0.015	0.035

The question of which changes have a significant effect on the optimal interval is addressed by a *n*-way ANOVA study [22], where *n* is the number of experimental variables with major changes being considered (three, in the example shown). This analysis can tell which changes have a significant influence on the interval limits. Furthermore, it can detect whether the combination of any of two, three, or more variables affects the interval limits. Those experimental variables whose changes do not affect the optimal region of the algorithmic parameter can be dropped in further analysis.

An ANOVA study of the results of the ART relaxation parameter reveals that the filtered or unfiltered nature of the noise is irrelevant for the optimal interval selection. However, the number of projection images and the volume size have a significant influence on the optimal parameter's limits. Between these two variables the number of projection images is responsible for twice as much of the variance than the volume size. No significant association is found among any of the three experimental variables.

The question of whether the dependency is predictable is addressed via nonlinear regression analysis [23]. This kind of analysis provides information about the validity of a given nonlinear model of the relevant experimental variables according to the data. The model itself serves as a heuristic selection rule for those experimental cases that have not been simulated as long as they can be reliably "interpolated" from the cases simulated.

In the example considered, nonlinear regression was used to establish a model for the upper limit of the optimal interval. Since the interval length is nearly constant due to the sampling effects in the parameter (we have measured the function at evenly distributed λ's with a separation of 0.2), the lower limit can be easily computed. Fig. 8 represents the dependency of the upper limit of the optimal interval of λ on the volume size and the number of projections. Table 3 shows the models that we tried, as well as their fits. R^2 is a standard measure of the regression fit and it represents the amount of variance in the data explained by the model; therefore, a good fit is reflected by a large R^2 value. One common way to produce models that account for interactions between two variables, N (the number of projections) and S (volume size), is by introducing terms of the form NS. However, as our previous 3-ANOVA analysis did not detect any significant interaction between these variables, such terms have not been included in the models shown.

Rounding the coefficients of model $f_1(N, S)$ in Table 3, it can be reasonably approximated by

$$\lambda_{max} = 0.09 - 10^{-6}(8N - 6(S - 96)^2),$$
$$\lambda_{min} = \lambda_{max} - 0.02$$

which accounts for the 89% of the total variance of the upper bound of the optimal interval. However, it must be kept in mind that the validity of this model is restricted to the particular conditions simulated (volume size from 65^3 to 129^3, a number of projections ranging from 800 to 6000, and random projection directions uniformly distributed).

4. Discussion

Our described procedure for intra-experiment analysis provides a methodology for recognizing different FOM tendencies, decreasing the complexity of the FOM optimization problem and, finally, finding a trade-off among them. Multivariate statistics serves as a solid base for this task. Furthermore, following this methodology one can concentrate on the extraction of middling solutions that are usually avoided by multiobjective optimization methods.

The fact that the measures of the FOMs are noisy forces us to perform many experiments. Moreover, the cost of each simulation in computation time is very high. For this reason, common multiobjective genetic algorithms searching for the whole Pareto optimal set are prohibitive in this context. However, the presented statistical framework provides a simple way of extracting the performance information from the data available.

The recognition of statistical associations among variables is a good feature of this analysis, since it points out which are the main problem characteristics captured by the different FOMs. For instance, in the example of ART in 3D EM two main tendencies have been recognized (one of the FOM groups is related to the contrast and the other to the noise in the reconstructions).

The inter-experiment study allows us to identify those major changes in the experimental variables that significantly

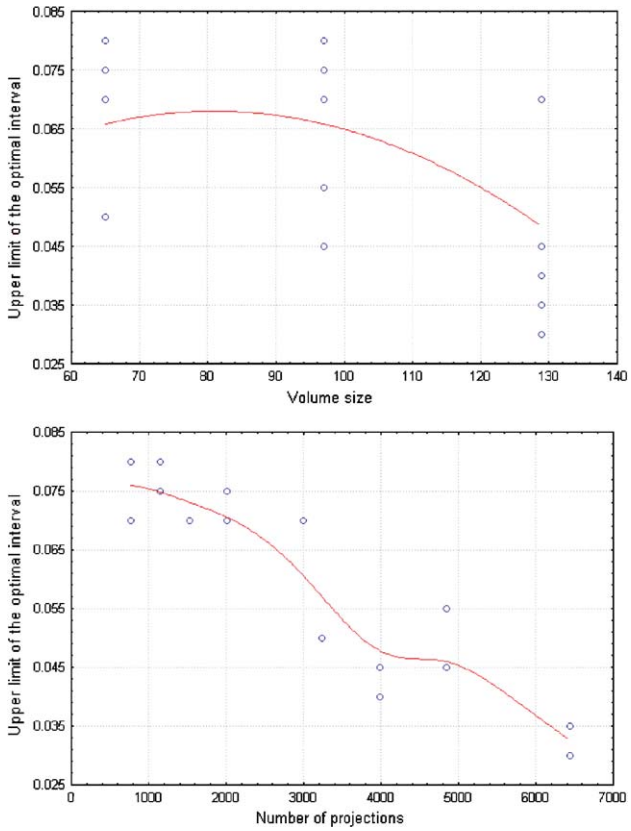

Fig. 8. Plot of the dependency of the upper limit of the optimal interval of the ART relaxation factor with the volume size and the number of projections. The lines correspond to the distance-weighted least-squares fit of the samples. It can be seen in the plots that the volume size is less important (the variance of the upper limit of the interval is larger) than the number of projections to select the ART relaxation factor.

Table 3
Models for the upper limit of the ART relaxation parameter

Model	Formula	R^2
Constant	0.074	0
$f_1(N)$	$0.0842 - 8 \times 10^{-6} N^{1.008}$	0.82
$f_2(N)$	$f_1(N) - 0.391 \times 10^{-12}(N-1021)(N-2697)(N-4643)(N-6501)$	0.85
$f_1(N,S)$	$f_1(N) + 5.83 \times 10^{-3} - 6 \times 10^{-6}\lvert S - 95.76\rvert^{2.12}$	0.89
$f_2(N,S)$	$f_2(N) + 5.44 \times 10^{-3} - 0.8 \times 10^{-6}\lvert S - 94.65\rvert^{2.66}$	0.92

N stands for the number of projection images and S for the volume size. The R^2 column reflects the fitness of the regression. The first column shows the name of the model in the second column. These names help to simplify the model formulation, for instance, $f_1(N,S)$ makes use of $f_1(N)$.

affect the parameters. Furthermore, nonlinear regression can be applied to establish a model of dependency. However, the utility of this model is strictly constrained by the experiments actually run and cannot be extrapolated to other situations that have not been considered.

The optimization of the ART+blobs relaxation factor (as well as that of the iterative data refinement (IDR) relaxation factor [42]) demonstrates the validity of our approach. Although visual appearance of reconstructions is too subjective to form the basis of definitive judgments, we note that in our opinion the reconstructed volumes using relaxation factors within the optimal interval showed the best visual appearance at the same value of the parameter at which the FOM trade-off (see Fig. 6) is achieved. Previous parameter

selection approaches in 3D EM reconstruction concentrated on the optimization of a single training FOM and the paradox was found that reconstructed volumes using parameters with good performance on the training FOM resulted in bad behavior from the point of view of other FOMs and this was reflected in inferior visual appearance.

Several important rules for the selection of the ART relaxations parameter can be drawn from our experiments, although it should be noted that these ideas may only apply to the particular test conditions simulated:

- The filtered or unfiltered nature of the additive noise does not affect the parameter optimal interval. This fact indicates that we need not distinguish between strongly and weakly defocused images in 3D EM reconstructions as far as the selection of the relaxation parameter is concerned.
- The most important parameter determining the optimal relaxation factor is the number of projection images. This result is logical, since more projections mean more volume updates and consequently a smaller iterative step should be used in order to reach an optimal reconstruction.
- The size of the volume being reconstructed is the other important factor when computing the optimal ART parameter. In our implementation of ART, the volume side length is the same as the projection side length. Thus, increasing the volume side length increases quadratically the number of equations specified by the projections. As the number of equations increases the iterative step should be made smaller in order to reach the same optimal reconstruction.

5. Conclusions

A methodology for the optimization of algorithmic parameters when it is possible to make experiments in which the true solution is known and there exist FOMs to measure the algorithmic performance has been provided. This methodology makes extensive use of multivariate statistical analysis and is capable of recognizing several FOM tendencies and achieving a trade-off among them. The approach proposed has been tested in the case of the optimization of the ART+blobs relaxation parameter. Finally, a guide has been provided for the selection of the ART relaxation factor in situations similar to the ones simulated.

Acknowledgements

We acknowledge partial economic support by the Spanish CICYT Grants TIC2002-00228 and BFU2004-00217/BMC, by the Comunidad de Madrid GR/SAL/0342/2004, by the Spanish FIS Grant G03/185, by the EU Grant FP6-502828 and by the NIH Grant HL70472.

References

[1] S.Z. Li, Close-form solution and parameter selection for convex minimization based edge-preserving smoothing, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 916–932.
[2] K.N. Otto, E.N. Antonsson, Design parameter selection in the presence of noise, Res. Eng. Des. 6 (1994) 234–246.
[3] R. Kohavi, G.H. John, Automatic parameter selection by minimizing estimated error, in: Machine Learning: Proceedings of the 12th International Conference, San Francisco, CA, 1995, pp. 304–312.
[4] K.B. Kulasekera, J. Wang, Smoothing parameter selection for power optimality in testing of regression curves, J. Am. Stat. Assoc. 92 (1997) 500–511.
[5] J.A.C. Weideman, Algorithms for parameter selection in the Weeks method for inverting the Laplace transform, SIAM J. Sci. Comput. 21 (1999) 111–128.
[6] S.S. Furuie, G.T. Herman, T.K. Narayan, P.E. Kinahan, J.S. Karp, R.M. Lewitt, S. Matej, A methodology for testing for statistically significant differences between fully 3D PET reconstruction algorithms, Phys. Med. Biol. 39 (1994) 341–354.
[7] R. Marabini, G.T. Herman, J.M. Carazo, Fully three-dimensional reconstruction in electron tomography, in: C. Borgers, F. Natterer (Eds.), Computational Radiology and Imaging: Therapy and Diagnostics, Springer, New York, 1999, pp. 251–281.
[8] N. Boisset, P. Penczek, J.C. Taveau, V. You, F. deHaas, J. Lamy, Overabundant single-particle electron microscope views induce a three-dimensional reconstruction artifact, Ultramicroscopy 74 (1998) 201–207.
[9] C.O.S. Sorzano, R. Marabini, N. Boisset, E. Rietzel, R. Schröder, G.T. Herman, J.M. Carazo, The effect of overabundant projection directions on 3D reconstruction algorithms, J. Struct. Biol. 133 (2001) 108–118.
[10] C. Coello Coello, A comprehensive survey of evolutionary-based multiobjective optimization techniques, Knowl. Inform. Syst. 1 (1999) 129–156.
[11] C.M. Fonseca, P.J. Fleming, An overview of evolutionary algorithms in multiobjective optimization, Evol. Comput. 3 (1995)
1–16.
[12] I. Avcibas, N. Memon, B. Sankur, Steganalysis using image quality metrics, IEEE Trans. Image Process. 12 (2003) 221–229.
[13] I. Avcibas, B. Sankur, K. Sayood, Statistical evaluation of image quality measures, J. Electron. Imag. 11 (2002) 206–223.
[14] R. Marabini, G.T. Herman, J.M. Carazo, 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs), Ultramicroscopy 72 (1998) 53–65.
[15] J. Ruprecht, J. Nield, Determining the structure of biological macromolecules by transmission electron microscopy, single particle analysis and 3D reconstruction, Prog. Biophys. Mol. Biol. 75 (2001) 121–164.
[16] M. van Heel, B. Gowen, R. Matadeen, Single-particle electron cryo-microscopy: towards atomic resolution, Quart. Rev. Biophys. 33 (2000) 307–369.
[17] I. García, P.M. Ortigosa, L.G. Casado, G.T. Herman, S. Matej, Multidimensional optimization in image reconstruction

from projections, in: I.M. Bomze, T. Csendes, R. Horst, P.M. Pardalos (Eds.), Developments in Global Optimization, Nonconvex Optimization and Applications Series, Kluwer Academics Publishers, Dordrecht, 1997, pp. 289–300.

[18] R. Marabini, E. Rietzel, R. Schröder, G.T. Herman, J.M. Carazo, Three-dimensional reconstruction from reduced sets of very noisy images acquired following a single-axis tilt schema: application of a new three-dimensional reconstruction algorithm and objective comparison with weighted backprojection, J. Struct. Biol. 120 (1997) 363–371.

[19] S. Matej, S.S. Furuie, G.T. Herman, Relevance of statistically significant differences between reconstruction algorithms, IEEE Trans. Image Process. 5 (1996) 554–556.

[20] S. Matej, G.T. Herman, T.K. Narayan, S.S. Furuie, R.M. Lewitt, P.E. Kinahan, Evaluation of task-oriented performance of several fully 3D PET reconstruction algorithms, Phys. Med. Biol. 39 (1994) 355–367.

[21] D.C. Montgomery, Design and Analysis of Experiments, fifth ed., Wiley, New York, 2000.

[22] R.G. Miller, R.J. Miller, B.W. Brown, Beyond ANOVA: basics of applied statistics, Texts in Statistical Science Series, Chapman & Hall, London, 1997.

[23] D.M. Bates, D.G. Watts, Nonlinear Regression Analysis and its Applications, Wiley, New York, 1988.

[24] G.T. Herman, Image Reconstruction from Projections: The Fundamentals of Computerized Tomography, Academic Press, New York, 1980.

[25] G.T. Herman, Algebraic reconstruction techniques in medical imaging, in: C.T. Leondes (Ed.), Medical Imaging, Systems Techniques and Applications, vol. 6: Computational Techniques, Gordon and Breach Science Publishers, Amsterdam, 1998, pp. 1–42.

[26] R.M. Lewitt, Alternatives to voxels for image representation in iterative reconstruction algorithms, Phys. Med. Biol. 37 (1992) 705–716.

[27] S. Matej, R.M. Lewitt, Efficient 3D grids for image reconstruction using spherically-symmetric volume elements, IEEE Trans. Nucl. Sci. 42 (1995) 1361–1370.

[28] S. Matej, R.M. Lewitt, Practical considerations for 3-D image reconstruction using spherically symmetric volume elements, IEEE Trans. Med. Imag. 15 (1996) 68–78.

[29] P. Viola, W.M. Wells, Alignment by maximization of mutual information, Int. J. Comput. Vision 24 (1997) 137–154.

[30] M. van Heel, Similarity measures between images, Ultramicroscopy 21 (1987) 95–100.

[31] Inc. Statsoft. Electronic statistics textbook. http://www.statsoftinc.com/textbook/stathome.html, 2001.

[32] J.W. Tukey, L.T. Fernholz, S. Morgenthaler, The Practice of Data Analysis, Princeton University Press, Princeton, NJ, 1997.

[33] D. Peña, F.J. Prieto, Multivariate outlier detection and robust covariance matrix estimation, Technometrics 43 (2001) 286–310.

[34] I.K. Lee, Curve reconstruction from unorganized points, Comput. Aided Geom. Des. 17 (2000) 161–177.

[35] S.C. Johnson, Hierarchical clustering schemes, Psychometrika 32 (1967) 241–254.

[36] W.R. Dillon, M. Goldstein, Multivariate Analysis: Methods and Applications, Wiley, New York, USA, 1984.

[37] P.H.A. Sneath, R.R. Sokal, Numerical Taxonomy, Freeman, San Francisco, CA, USA, 1973.

[38] H.F. Kaiser, The varimax criterion for analytic rotation in factor analysis, Pyrometrical 23 (1958) 187–200.

[39] I.T. Jolliffe, Principal Component Analysis, Springer, Berlin, 1986.

[40] Y. Censor, T. Elfving, G.T. Herman, A method of iterative data refinement and its applications, Math. Methods Appl. Sci. 7 (1985) 108–123.

[41] C.O.S. Sorzano, J.J. Fernández, R. Marabini, G.T. Herman, Y. Censor, J.M. Carazo, Transfer function restoration in 3D electron microscopy via iterative data refinement, in: Proceedings of the Sixth International Meeting on Fully Three-dimensional Image Reconstruction in Radiology and Nuclear Medicine, 2001, pp. 133–136.

[42] C.O.S. Sorzano, R. Marabini, G.T. Herman, Y. Censor, J.M. Carazo, Transfer function restoration in 3D electron microscopy via iterative data refinement, Phys. Med. Biol. 49 (2004) 509–522.

About the Author—C.O.S. SORZANO was born in Málaga (Spain, 1973) where he studied Electrical Engineering (M.Sc., Honors) and Computer Science (B.Sc.). He then joined the BioComputing Unit at the National Center for BioTechnology (CNB) of the Spanish National Council of Scientific Research (CSIC), Madrid, Spain, where he obtained a Ph.D. (Honors). He worked as a Research Assistant in the Biomedical Imaging Group (EPFL, Switzerland), and currently he is a Lecturer in University San Pablo-CEU (Madrid, Spain). His research interests include image processing, tomography, multiresolution approaches, and electron microscopy.

About the Author—ROBERTO MARABINI received the M.S. (1989) and Ph.D. (1995) degrees in Physics from the University Autonoma de Madrid and University of Santiago de Compostela, respectively. He was a Ph.D. student at the BioComputing Unit at the National Center for BioTechnology (CNB) of the Spanish National Council of Scientific Research (CSIC), Madrid, Spain.
He worked at the University of Pennsylvania and the City University of New York from 1998 to 2002. At present he is an Associate Professor in the "Escuela Superior Politecnica" at the University Autónoma de Madrid. His current research interests include inverse problems, image processing and high performance computing.

About the Author—GABOR T. HERMAN is a pioneer in the field of X-ray computerized tomography and the author of several books and over 100 articles, including many classic works in the field. He is recognized internationally for his major contributions to image processing and its biomedical applications. His Ph.D. is in Mathematics. He was the leader of successful image processing groups at SUNY Buffalo and at the University of Pennsylvania and has garnered millions of dollars in research funding. He is a highly accomplished scientist of international distinction and has been awarded honorary degrees from the universities of Haifa in Israel, Szeged in Hungary, and Linköping in Sweden. He is currently a Distinguished Professor of Computer Science at The Graduate Center of the City University of New York.

About the Author—J.M. CARAZO received the M.Sc. degree in Physics (Honors) from the Granada University, Spain, in 1981, to be followed by a Ph.D. in Molecular Biology at the University Autónoma of Madrid (UAM) in 1984. He left for Albany, New York, in 1986, coming back to Madrid in 1989 to set up the BioComputing Unit of the newly established National Center for Biotechnology (CNB), nowadays the largest center of the Spanish National Council of Scientific Research (CSIC), located in the campus of the UAM. Currently he is involved in the Spanish Ministry of Science and Technology as Deputy General Director for Research Planning while he keeps engaged in his activities at the CNB, the UAM, the newly established Scientific Park of Madrid (PCM), and the start-up company Integromics.