Published in final edited form as:

J Struct Biol. 2007 April ; 158(1): 19–32. doi:10.1016/j.jsb.2006.10.003.

Bootstrap Resampling for Voxel-wise Variance Analysis of Three-dimensional Density Maps Derived by Image Analysis of Two-dimensional Crystals

Anchi Cheng^{*,#} and Mark Yeager^{*,‡}

^{*} The Scripps Research Institute Department of Cell Biology 10550 North Torrey Pines Road La Jolla, CA 92037, USA.

[‡] Division of Cardiovascular Diseases Scripps Clinic 10666 North Torrey Pines Road La Jolla, CA 92037, USA.

Abstract

Difference density maps are commonly used in structural biology for identifying conformational changes in macromolecular complexes. For interpretation of the results, it is essential to estimate the variance or standard deviation of the difference density and the distribution of errors in space. In order to compare three-dimensional density maps of gap junction channels with and without the C-terminal regulatory domain, we developed a bootstrap-resampling method for estimation of the voxel-wise standard deviation. The bootstrap approach has been successfully used for estimating the sampling distribution from a limited data set and for estimating the statistical properties of the derived quantities [Efron, B. *Ann. Statist.* 7, 1 (1979)]. In our application, the standard deviation map can be estimated by bootstrapping the images. Our result showed that, apart from the symmetry axes and small regions bordering the lumen of the extracellular vestibule, difference maps normalized by the mean of the standard deviation map can be used as a good approximation of the *t*-test map of the gap junction crystals.

Keywords

error analysis; statistics; image processing; electron microscopy; crystallography

INTRODUCTION

Calculation of the variance or standard deviation of three-dimensional (3D) density maps derived by electron microscopy and image analysis is useful to estimate the confidence interval of a map, and the variation in different voxels can provide information on local structural flexibility. A standard deviation map can be easily generated for a density map if a complete 3D map can be derived from an individual electron microscopic (EM) image, as is the case for helical objects (Milligan and Flicker, 1987). For single particle analysis, methods have also been developed to estimate the variance from nearest neighbors (Liu *et al.*, 1995; Liu and Frank,

© 2006 Elsevier Inc. All rights reserved.

corresponding author .

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1995). Unfortunately, neither of these approaches is suitable for two-dimensional (2D) crystallography where 3D maps are derived by merging several EM images at varying tilt axis and tilt angles, in which there is often relatively sparse sampling of orientations.

Difference map analysis of reconstructed structures derived from analysis of EM images of 2D crystals have yielded important information about the location of bound ligands, substrates (Tate *et al.*, 2003) and subcomponents (Smit *et al.*, 2001) within macromolecular complexes. In these examples, the difference maps were calculated by scaling the two individual maps to the same density range, and the analysis of the difference map assumed that the pixel or voxel variations of the standard deviation for each structure were insignificant. Since the mean variance was not calculated nor estimated, the significance of the difference could not be established. In our comparison of the 3D density maps of gap junction channels with and without the C-terminal regulatory domain, there were only small changes in density, necessitating careful statistical evaluation of the difference peaks. Therefore, we investigated the use of the bootstrap resampling technique to estimate the voxel-wise variation in the final density maps.

The bootstrap resampling method developed by Efron (1979) has been used widely in statistical problems. Examples can be found in finance and management science (Govindarajulu, 1999), environmental sciences (Dixon, 2002), as well as bioinformatics (Molinara *et al.*, 2005) and medicine (Walters and Campbell, 2005). In the field of structural biology, the bootstrap resampling method has also been recently applied to single particle reconstruction (Penczek *et al.*, 2006a; Penczek *et al.*, 2006b). When a quantity is derived from a set of samples or measurements of a population, a single measurement of the quantity gives no information about its distribution properties such as variance. The bootstrap method estimates the distribution properties of the quantity by a Monte Carlo simulation, in which a distribution of the quantity is created by sampling the set of measurements with replacement many times. This is shown schematically in Fig. 1(d-f), where a limited number of colored spheres represent a set of samples from the population distribution of the spheres.

The main advantage of using the bootstrap resampling approach is that good estimates can be obtained, regardless of the complexity of the data processing. In this study we show that the bootstrap resampling technique is well suited for estimating the voxel-wise variance map and converges in a practical number of cycles. As expected, the variance is maximal at symmetry axes, whereas the variance at most other locations is minimal.

METHOD

Variance, standard deviation, and standard error of the mean

In standard statistical practice (for example Samuels, 1989; Devore, 2001), for N repeated measurements (or samples) of ψ , the mean, $\bar{\psi}$, is defined as

$$\bar{\psi} = \frac{1}{N} \sum_{n=1}^N \psi_n \quad (1)$$

and is an estimate of the population mean. The standard deviation of the measurements, σ_{ψ} , is

$$\sigma_{\psi} = \sqrt{\frac{\sum_{n=1}^N (\psi_n - \bar{\psi})^2}{N - 1}} \quad (2)$$

and is an estimate of the population standard deviation. The variance, v_ψ , of the measurements is the square of the standard deviation, i.e.,

$$v_\psi = \sigma_\psi^2 \quad (3)$$

and the standard error of the mean, δ_ψ , is

$$\delta_\psi = \sigma_\psi / \sqrt{N}. \quad (4)$$

Standard deviation combination and t test

When the difference is considered between two data sets and if the significance level of the difference is of interest, the variance (or standard deviation, or standard error of the mean) for two data sets have to be combined to obtain the standard error of the difference. The method of combination depends on whether the variances of the two data sets are equal. Statistical tests exist for determining the equality of the two populations. For example, the F -test for equality of variance requires calculation of F -statistics, F , by

$$F = \frac{v_{\psi_1}}{v_{\psi_2}} \equiv \frac{v_1}{v_2} \quad (5)$$

where we use both v_{ψ_i} and v_i 's, to express the variance of the measurements ψ_i 's for the i th data set. For two sets of N_i measurements having N_1-1 and N_2-1 degrees of freedom, critical values of the F distribution at a given significance level α can be found in tables in elementary statistics books (Samuels, 1989; Devore, 2001).

For an unequal population distribution for the standard deviations, estimated by σ_ψ , the unpooled standard error of the difference, $\delta_{\Delta\psi}$, is

$$\delta_{\Delta\psi} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} = \sqrt{\delta_1^2 + \delta_2^2} \quad (6)$$

where the δ_i 's are the standard error of the means. The unpooled combination is also recommended when the sample number is small, in which case the F -test is not reliable.

For equal population distributions, the pooled standard error of the difference, $\delta_{\Delta\psi}$, is

$$\delta_{\Delta\psi} = \sqrt{\frac{\sigma_c^2}{N_1} + \frac{\sigma_c^2}{N_2}} \quad (7)$$

where

$$\sigma_c^2 = \frac{(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2}{N_1 + N_2 - 2}$$

$$= \frac{(N_1 - 1)N_1\delta_{-1}^2 + (N_2 - 1)N_2\delta_{-2}^2}{N_1 + N_2 - 2} \quad (8)$$

Therefore, t statistics for the difference of the two data sets, ψ_1, ψ_2 , are estimated as

$$\hat{t}_{\Delta\psi} = \frac{\Delta\bar{\psi}}{\delta_{\Delta\psi}} \quad (9)$$

The degrees of freedom are required for calculation of the critical value of t statistics at a given level of significance. We will adopt a conservative estimate for the degrees of freedom, which is the smaller of the two data sets (Devore, 2001).

Sampling distribution and bootstrap resampling of the mean

To understand the bootstrap technique, it is useful to consider how the sampling distribution of the mean can be obtained. The standard error of the mean for N samples corresponds to the standard deviation of the means in a meta-experiment (Samuels, 1989) in which N samples are drawn from the population with replacement to calculate the mean of each N samples for an infinite number of repetitions (Fig. 1a,b). That is,

$$\delta_{\psi} = \sigma_{\psi} \quad (10)$$

The sampling distribution of the mean is narrower than the population distribution for any sample size $N > 1$ as described in Eq. 4, and shown in Figure 1 (compare gray and red curves).

The bootstrap resampling (Efron, 1979) of the mean for a given set of samples is a simulation of such a meta-experiment. The “population” distribution in such a bootstrap resampling experiment is the distribution of the N given samples (compare Figs. 1a and 1d), and the standard deviation of all the means obtained from a large number (Q) of resamples will approach the sampling distribution of the mean for the distribution of the N samples (compare Figs. 1c and 1f). Because the distribution of the N samples is an estimate of the original population distribution, the bootstrap resampling distribution becomes an estimate of the sampling distribution of the mean for the N samples. Specifically, the bootstrap estimate of the

standard error of the mean, $\widehat{\delta}_{\psi}^B$, is defined as the standard deviation of the resampled means, σ_{ψ}^B .

$$\widehat{\delta}_{\psi}^B = \sigma_{\psi}^B \quad (11)$$

For a more general application, instead of $\bar{\psi}$, the distributions of much more complicated quantities can be estimated by the same procedure. Examples are sample variance, regression and correlation coefficients, ratio estimators and smooth transforms of these (Babu and Rao, 1993).

The bootstrap resampling method estimates the statistical properties based on the distribution of the samples. For a given number of resamplings, Q , the error of the estimate, e , for a particular

statistical quantity is given by the ratio of σ_{ψ}^2 to the square of the variance, σ_{ψ}^4 :

$$e = \frac{\sigma_{\psi}^2}{\sigma_{\psi}^4} = \frac{2}{Q} \quad (12)$$

assuming a normal distribution (Penczek *et al.*, 2006b). In this formula, the quantity σ_{ψ}^2 is the variance of the variance. We performed a simulation with a normal distribution of numbers to confirm that an error of 0.01 can be achieved in 200 bootstrap loops (data not shown).

Simple bootstrap resampling for density maps

Specific algorithms are used to reconstruct both 2D projection and 3D density, $\tilde{\rho}$, maps from a set of 2D images. That is, they are both quantities that can be derived from the samples (measured images) in Fig 1c through a regression model and smooth transformations as stated above. We use the symbol $\sim$ to describe the density, $\tilde{\rho}$, to indicate that these functions are not the average density, as we will discuss later, but are quantities also derived from the collected information provided by the samples. Therefore, the above general case can be extended to 2D and 3D density maps where we are interested in the standard error of $\tilde{\rho}$ and the difference between $\tilde{\rho}$ s, as well as the t -statistics for each pixel and voxel, respectively.

The following is used to estimate the variance and the standard error of $\tilde{\rho}$, $v_{\tilde{\rho}}$ and $\delta_{\tilde{\rho}}$, respectively, from N images:

For a given image, j , the reflection structure factor vector list is

$$F_j = \left[\vec{f}_{jh_1, k_1, z_{j,1}^*}, \dots \right] \quad (13)$$

The reconstructed 2D or 3D density, $\tilde{\rho}$, is obtained from $F_1 \dots F_N$ using standard 2D crystallography reconstruction algorithms (Crowther, 1971) as previously described (Unger *et al.*, 1999; Cheng *et al.*, 2003). The process involves fitting of lattice lines to the unevenly sampled measurements (Agard and Stroud, 1982), resampling the lattice lines at equal z^* separation based on the crystal thickness, and performing an inverse Fourier transformation to obtain the density map. For the case of a 2D projection map, the fitted lattice line is restricted to $z^* = 0$ before the inverse Fourier transformation is performed.

To apply the bootstrap technique to calculate the distribution of the voxel- (or pixel-)wise density map, the following bootstrap loops are constructed: For each bootstrap loop, q , we randomly draw, from the available images, N samples, named $j_{1,q}^B \dots j_{N,q}^B$ that have a corresponding reflection list of structure factor vectors $F_{1,q}^B \dots F_{N,q}^B$. Note that each bootstrap loop considers all the images in the dataset. After selection of a particular image, it is returned to the data set, and a random selection is made again, a process known as sampling with replacement, until there are N images to complete loop q . For example, it is possible that both $j_{1,q}^B$ and $j_{N,q}^B$ sample image 1 so that $F_{1,q}^B = F_{N,q}^B = F_1$ for loop q . The reconstructed density $\tilde{\rho}_q^B$ for

loop q is then calculated from the values of $F_{1,q}^B \dots F_{N,q}^B$ using the exact algorithm that was used to reconstruct the original $\tilde{\rho}$ from $F_1 \dots F_N$. In the extreme case where all images in the similar tilt geometry are resampled by others, a z^* bin may become so sparsely sampled that it no longer contributes to the reconstruction. However unusual it is, such a condition must be allowed in the process since the frequency for it to occur is also a result of distribution of the image samples. With reference to Figure 1d, we are interested in the distribution of $\tilde{\rho}$ calculated from the N projection images. In this case, each of the N images would be a colored circle as shown in Figure 1d. Therefore, for a total of Q bootstrap loops, the estimate of the standard error of $\tilde{\rho}$, $\widehat{\sigma}_{\tilde{\rho}}^B$, following Eqns. 2 and 11, is

$$\widehat{\sigma}_{\tilde{\rho}}^B = \sigma_{\tilde{\rho}_q^B} = \left(\sum_{q=1}^Q (\tilde{\rho}_q^B - \overline{\tilde{\rho}^B})^2 / (Q - 1) \right)^{1/2} \quad (14)$$

The average of $\tilde{\rho}^B$ from the samples, $\overline{\tilde{\rho}^B}$, is an estimate of $\tilde{\rho}$, and is given by

$$\overline{\tilde{\rho}^B} = \sum_{q=1}^Q \tilde{\rho}_q^B / Q \quad (15)$$

Residual bootstrap resampling estimate for 3D density maps

Since the 3D density reconstruction utilizes lattice line fittings of the measurements in Fourier space, we also considered residual bootstrapping for regression models (Härdle and Bowman, 1988). The assumption for residual bootstrapping is that while the population means follow the regression model of the lattice lines along z^* , the shape and width of the population distribution is unchanged, or homoscedastic along z^* . Therefore, the residual of the fits can be resampled at various z^* positions where the measurements exist. The complete description of its application to the estimate of 3D variance is presented in the Supplementary Material.

Calculation

The bootstrap samples of the images were generated by a program written in PYTHON (<http://www.python.org>). The input data were the scaled structure factors, which had been merged to a common phase origin. The outputs were Q sets of resampled lists of reflections. For bootstrapping, all reflections from the n th image were replaced by its bootstrap resampled selection. The image processing and lattice line fittings were performed with the MRC 2D crystal image processing software (Crowther *et al.*, 1996). The density reconstruction from the fitted structure factor list was performed with the standard CCP4 (Collaborative Computational Project, 1994) packages. The statistical parameter calculations of the density maps were performed by minor modifications of MAPSIG in the CCP4 package (e.g., to calculate the square root of a map). All programs were generalized for all 17 two-sided plane group symmetries, and the programs and example scripts are available upon request. We did not attempt to optimize the efficiency of the calculation. The longest standard error map calculation that contained 400 cycles of sampling for 62 images with 9,675 reflections required 2.3 hr on a 2.8 GHz Intel Xenon processor running the Linux operation system.

Analysis of bootstrap resampling for projection data was based on a total of 17 images with small tilt angles ($< 4^\circ$) of gap junction crystals formed by Cx43 that had been treated with trypsin to remove the C-tail (designated Cx43-TR). For bootstrap resampling of the 3D data, we separately analyzed 48 images of crystals formed by full-length Cx43 (designated Cx43-

WT) and 62 images of crystals of Cx43-TR. Scaling of the 3D structure factors and the maps was based on the slope of a linear fit to all corresponding structure factor amplitudes present in both data sets. It is common to sharpen 3D maps derived by EM image analysis using an inverse temperature factor (Scherter *et al.*, 1993). The results we obtained by scaling with a temperature factor of $B = -350$, were consistent with our previous studies (Unger *et al.*, 1997; Unger *et al.*, 1999; Cheng *et al.*, 2003).

Independence between voxels is required for a quantitative analysis of the t -map. Previous calculations by Liu *et al.* (Liu *et al.*, 1995) showed that independence is achieved when the map is sampled at the FWHH of the point spread function. In our case, these values are 2.58, 2.58, and 7.04 Å in the x , y , and z directions, respectively. We note that the FWHH in the z direction is larger because of the missing cone in the z^* direction, due to the maximum tilt angle of 35°. However, for an in-plane resolution limit of 7.5 Å, the Shannon sampling theorem requires sampling at better than 3.75 Å in all directions. As a result, we sampled the map using values of 2.74, 2.74, and 3.33 Å on a grid of 28, 28, and 90 pixels, to divide the unit cell in the x , y , and z directions, respectively.

3D surface views were rendered using Chimera (Pettersen *et al.*, 2004). The error map calculation was performed on grids of a p6 lattice from a p6 symmetrized density map. However, we did not impose p6 symmetry on the calculated error map. The presence of good p6 symmetry in the projection map indicates that the round-off errors were small. Nevertheless, the p6 symmetry for the 3D error maps was not perfect in the final display due to the coarse sampling and the interpolations.

Structure variation simulation in projection

To generate artificial crystals of molecules of two conformations, we considered a simple molecule consisting of two Gaussian density peaks (Fig. 2). The projection density in the unit cell sampled by grids of 45 by 45 was defined by:

$$\rho_{A,B}(x,y) = 20^* \exp \left[-\frac{((x-22)^2 + (y-22)^2)}{8^2} \right] + 30^* \exp \left[-\frac{((x-34)^2 + (y-22 \pm 5)^2)}{4^2} \right] \quad (16)$$

The two conformations were different in the choice of the sign of the y offset in the second Gaussian function (Figs. 2a and b). Seventeen 2D crystals were generated by assigning a unit cell density of either conformation randomly to the square matrix of a lattice containing 40 lattice points in each direction. The density calculation and crystal image generation were written in PYTHON (<http://www.python.org>) using functions in numarray (http://www.stsci.edu/resources/software_hardware/numarray) and those written for Legimon (Suloway *et al.*, 2005). The simulated crystals were treated as CTF-corrected projection images. Standard error calculation and estimation were then performed as with real data. Because the simulated data lacked variations from other sources such as background white noise, the reflection phase IQ (Henderson *et al.*, 1986) values were mostly 1 except for those that were strongly affected by variation in the conformation distribution.

Effect of noise level to the bias of $\overline{\rho_i}$ and δ_{ρ_i} as estimate of $\widetilde{\rho}$ and $\delta_{\widetilde{\rho}}$

To generate simulated crystal images with a defined noise level, random values were drawn from a normal distribution $N(0, N_L^2)$ centered at 0. The standard deviation N_L defined the noise

level, which was added to the pixel values of each simulated crystal image used in Figure 3. As in the simulation without noise, 17 crystals with molecules of two conformations were created. The bootstrap estimation used 64 loops. The simulation was repeated 64 times so that $\delta_{\bar{\rho}}$ could be calculated from the sampling distribution. The average values of $\delta_{\bar{\rho}_i}$ and $\delta_{\bar{\rho}_q^B}$ from the 64 simulations are reported.

Testing of the jackknife estimation method

The jackknife method (Quenouille, 1949) is also a popular approach for estimating the distribution properties of parameters that are either derived indirectly or obtained directly from multiple measurements (Govindarajulu, 1999). In its simplest form, the jackknife method estimates the variance of a given dataset by examining the variance of synthetic datasets, each created by removing one of the measurements in turn. Therefore, the jackknife-estimated standard error of the reconstructed projection map, $\widehat{\delta}_{\bar{\rho}}^J$, is obtained from

$$\left(\widehat{\delta}_{\bar{\rho}}^J\right)^2 = (N-1) \sum_{n=1}^N \left(\bar{\rho}_{(-n)}^J - \bar{\rho}\right)^2 / N \quad (17)$$

where $\bar{\rho}_{(-n)}^J$ is the projection map with n th image omitted from the calculation. We applied the jackknife method to the simulated map using a PYTHON program to iterate the omitting process.

RESULTS

Test of standard error estimation of the reconstructed projection-simulated data

Our first test for verifying the approach was to confirm that the bootstrap estimate of the standard error of the projection map, $\widehat{\delta}_{\bar{\rho}}^B$, could be used to detect pixel-wise variation when variations in the structure factor values were the only source of variation. We performed this test on a simulated projection data set that contained artificial molecules of two conformations (Figs. 2a and b). The simulated images contained no phase contrast or white noise background and had identical phase origins, lattice parameters, and tilt geometry. A total of 17 simulated images were used to match the number of images in the projection data set of gap junction crystal images. For comparison, we also calculated the mean of the individually reconstructed maps $\bar{\rho}_i$, and the results indicate that $\bar{\rho}^B$ and $\bar{\rho}_i$ are both good estimate of $\bar{\rho}$. The bootstrap-estimated standard error, $\widehat{\delta}_{\bar{\rho}}^B = \sigma_{\bar{\rho}_q^B}$, (Fig. 2e) of this simulated data set faithfully reproduced $\widehat{\delta}_{\bar{\rho}_i}$ (Fig. 2d) with minimal bias. We repeated the calculations for different simulated image sets, and the results were consistent. Recall that the ultimate purpose of bootstrap resampling is not just to estimate the statistical parameters based on the existing samples but those of the population the samples were drawn from. With the simulated population as here, we can perform directly the sampling experiment as illustrated in Fig 1a-c to obtain the standard error of $\bar{\rho}$, $\delta_{\bar{\rho}}$. The comparisons of the three maps showed that both $\sigma_{\bar{\rho}_q^B}$ and $\delta_{\bar{\rho}_i}$ were good estimates of $\delta_{\bar{\rho}}$ (from 64 sampling experiments) in the absence of noise (Fig. 3).

Agreements among various estimates broke down when noise was added to the simulated crystal images. Figure 3 shows the comparison at three noise levels. Two pixels were chosen to show the two extremes for the behavior of various standard errors. As shown in the insert of Figure 3b, one of the pixels was located at the center of the invariant structure, and the other was at the center of one of the lobes with maximal variance. Without noise, $\delta_{\bar{\rho}}$ was well

estimated by both $\delta_{\rho_i}^-$ and $\sigma_{\rho_q^B}$. With noise, $\sigma_{\rho_q^B}$ became the better estimate as $\delta_{\rho_i}^-$ increased at both pixels (Fig. 3a, b). For high noise levels $\bar{\rho}$ was also better estimated by $\bar{\rho}^B$ rather than $\bar{\rho}_i$ (Fig. 3c, d).

We also used the jackknife method to estimate the same quantity (Fig. 2f). The map showed spurious peaks at various locations where they were not expected. Therefore, we concluded that the bootstrap method was more appropriate for our analysis.

Description of gap junction structure

Our real test case for the bootstrap method used two gap junction Cx43 structures, a map of the full-length wild type channel (Cx43-WT) and a map of crystals treated with trypsin to remove the C-tail (Cx43-TR). Both of these maps are similar in architecture to the Cx43-263T variant previously published (Unger *et al.*, 1997; Unger *et al.*, 1999; Fleishman *et al.*, 2004). Gap junction hemichannels, called connexons, assemble as hexamers in the plasma membrane. The dodecameric channel (Fig. 4a) is formed by the end-to-end docking of hexamers of closely apposed cells, thereby forming an intercellular conduit across the extracellular gap, from which the name is derived. The 2D crystals are formed by hundreds of channels, each of which displays pseudo non-crystallographic p622 symmetry, reflecting its homododecameric nature (Fig. 4). The crystals show only p6 two-sided plane group symmetry. Within each membrane the central aqueous channel of the connexon is surrounded by 24 α -helices, 4 per connexin subunit (Figs. 4b and d). In the extracellular region, the protein density forms a tight seal to prevent exchange of ions and metabolites with the extracellular space (Figs. 4a and c). Because this continuous wall of density is superimposed with most of the tilted helices in the membranes, the projection density map has the appearance of a continuous ring of density at a radius of 25 Å from the center of the channel (Fig. 4e). Overlapping densities from the helices of the two hemichannels form the 6 smeared inner densities at 17 Å radius and 6 strong circular densities at 33 Å radius. The latter strong densities arise from good vertical alignment by a set of straight helices along the channel axis. The inner ring of 6 smeared densities arises from overlapping ends of tilted helices.

Test of standard error estimation in projection – Analysis of real data

To demonstrate the validity of bootstrap resampling on real data, we compared maps of the standard error of the mean of individually reconstructed projection density maps, $\delta_{\rho_i}^-$, with bootstrap estimates, $\sigma_{\rho_q^B}$ (Fig. 4). The maps were based on 17 images of gap junction crystals formed by Cx43-TR. The maximum tilt angle was 3.7°, assuming equal weighting. Figure 4a shows $\delta_{\rho_i}^-$ where each single map, derived from one of the 17 images, was treated as a measurement in the pixel-wise standard error calculation using Eqns. (1), (2) and (4). Figure 4b and c show two independent maps of $\sigma_{\rho_q^B}$, each based on 200 cycles of bootstrap resampling of the images. The near equivalence of the maps in Figures 5b and c demonstrates that 200 cycles was sufficient for convergence of the results. The regions with higher variance were similar in all 3 maps, with the strongest peak at the 6-fold rotation axis. Secondary peaks were located on the 6 circular densities at a radius of 33 Å. However, the values of the estimated standard error in Figures 4b and c were consistently lower than the actual values in Figure 4a (19% on average). This result suggests that the bootstrap resampling estimate of the density variance is not the same as the variance of individually reconstructed maps for noisy data, exemplified by the simulation in Figure 3.

Test for ability to detect strong voxel-wise variation in 3D data set

Variation in the structure accounts for only part of the variance obtained from analysis of EM images. In order to verify that our approach was capable of detecting variations in 3D structure,

we considered the case where the structure variation was expected to dominate. To this end, we generated an artificial 3D data set that had strong variation by merging the whole set of Cx43-WT images with their mirrored counterpart. Because the dodecameric channel is formed by the end-to-end docking of two hexamers, the mirror-merged data set was expected to generate a density map with non-crystallographic pseudo-6/mmm point group symmetry, as would the voxel-wise variation. Figure 6 displays the results of the test as a pie section of the density map, as indicated in Figure 4b. The merged features in the mirrored structures from Figures 6a and b are shown as a solid contoured density map in Figure 6c and as a wire-framed density map in Figure 6d. The asymmetric elongated density arms from two sloped α -helices in Figure 6a and b merge into symmetric lobes at a radius of 17 Å with respect to the 6-fold axis, while the merging of the rest of the helices gives rise to continuous lobe of density at the outer radii. The mirror plane in Figure 6d clearly indicates the pseudo mirror symmetry in the merged map (Fig. 6d). The bootstrap estimated voxel variation shown as a surface rendering of $\sigma_{\tilde{\rho}_q}^B$ in Figure 6e and f also displays the same mirror symmetry, with one of the mirror planes shown in panel f. The prominent pair of variations, marked by the red stars in Figure 6e, were located within the sloped helices where they disappeared in the merged map at a comparable contour level.

Standard error map in 3D

The bootstrap estimated standard error map for both Cx43-WT and Cx43-TR, as well as that of the difference map, shared similar features, except that the scales were different in each. Therefore, we will present only the bootstrap estimated standard error map of the difference, $\widehat{\sigma}_{\Delta\rho}^B$. Critical value of F at $\alpha=0.05$ for the size of the two data sets is 1.56. *F*-statistics of the two variance maps gave values in the range of 0.61 to 5.40 and 87% of the voxels have *F* values higher than the critical value of *F*. This result suggests that for most voxels of the two maps, the variance in Cx43-TR was larger than Cx43-WT possibly due to more conformational flexibility in the trypsin digested sample. Therefore, the estimated standard errors were combined in an unpooled fashion using Eqn. 6. The values of $\widehat{\sigma}_{\Delta\rho}^B$, estimated from both simple and residual bootstrap resampling, are shown parallel to the bilayers in Figures 7c and d, respectively, and as a slab within the extracellular region of the channel in Figures 7a and b, respectively. For both simple and residual bootstrapping, the voxel-wise values of $\widehat{\sigma}_{\Delta\rho}^B$ were quite uniform, with an rms deviation that was only 23-24%. The highest values of standard error were located at the 3- and 6-fold symmetry axes and extended throughout and beyond the thickness of the protein density. Apart from the peaks at the symmetry axes, the only peaks that could be associated to the protein at 1.5 rms deviation above the $\text{mean}_{sp}[\widehat{\sigma}_{\Delta\rho}^B]$ are at the extracellular gap, and at the boundary of the lumen of the pore where the density gradient was highest (Figs. 7). Although the variance peaks were located at similar positions for simple and residual bootstrapping, the significant features at the pore lumen in the $\widehat{\sigma}_{\Delta\rho}^B$ maps were elongated due to the low *z* resolution (see discussion).

Using the estimated $\widehat{\sigma}_{\Delta\rho}^B$ from simple bootstrapping, a *t*-statistics map was calculated (Figs. 8a and c) and compared with the difference map $\Delta\tilde{\rho}$ (Figs. 8b and d). The detailed interpretation of the 3D *t* map will be published elsewhere. The main observation at this point is that due to the relative uniformity of $\widehat{\sigma}_{\Delta\rho}^B$ and the general mismatch of regions of high variance and high difference, the $\widehat{t}_{\Delta\rho}$ map and $\tilde{\rho}$ map were quite similar at well matched contour cutoffs. Only the difference peaks at symmetry axes were significantly lower in the *t* map due to high values of the local $\widehat{\sigma}_{\Delta\rho}^B$.

We also note that, in the 3D maps, the $SD_{sp}[\Delta \tilde{\rho}] \approx 1.2 * \text{mean}_{sp}[\tilde{\delta}_{\Delta \rho}^B]$, implying that only a small percentage of voxels contained significant difference peaks, in contrast to the projection maps, where $SD_{sp}[\Delta \tilde{\rho}_{xy}] \approx 1.8 * \text{mean}_{sp}[\tilde{\delta}_{\Delta \rho_{xy}}^B]$.

DISCUSSION

Estimation of variance in density maps has been attempted in several different settings. Blow and Crick (1959) estimated the mean variance in an X-ray crystallographic density map derived by the isomorphous replacement method. Henderson and Moffat (1971) extended the error analysis to a difference Fourier synthesis. Their estimate of the mean variance requires

$\Delta F > \sigma_{\Delta F}^-$ and $\Delta F \ll |F|$ where ΔF is the difference in structure amplitude, $|F|$, and, $\sigma_{\Delta F}$ is the r.m.s. error in measurement of ΔF . Since there is considerable error in EM structure factor amplitudes derived by Fourier transformation of images, we did not attempt a similar calculation. However, our results confirm that variation of the variance is small throughout the map even with large errors in the structure factor amplitudes. Another attempt to estimate the projection variance map was made by Mitra et. al. (1993) who synthesized one single error map with the standard deviation of each structure factor randomly flipped in direction. We found that such an estimator was inconsistent and note that it does not reflect the fact that each image, not reflection, is a sample of the distribution.

In this study, we bootstrapped the images in order to estimate the variation in the density maps. We demonstrated through analysis of simulated (Fig. 2 and 3) and real data (Figs. 5 and 6) that the calculation is practical and that the standard error map so estimated can display the expected properties when structure variation is present. Interestingly, the variance of the voxel-wise standard error was surprisingly low except at the symmetry axes (Fig. 7). Variation in the structure accounts for only part of the variance obtained from analysis of EM images. There are several other sources of systematic error that contribute to the variance in producing the reconstruction that we will consider.

The most obvious source of variation is the artifact of symmetry averaging. We observed that the variances were stronger at the special positions of the crystallographic symmetry elements: The higher the symmetry, the more dominant the variation. We rationalize this observation as follows. In the unit cell of a crystal with p6 symmetry, the density value at any non-special position is, in fact, the mean of the 6 symmetrically related positions. On the other hand, density at a position on the 6-fold axis has no symmetry pair to average. Therefore, assuming that, before symmetrization, all positions in the unit cell have the same statistical property, the p6 symmetrized map would find the pixel-wise standard deviation at non-special position at a value square-root of 6 times that at the 6-fold axis, as suggested by Eqn. 4. Indeed, when we divide the standard deviation value at the 6-fold axis from any of our maps by $\sqrt{6}$, the resulting values are all within $sd_{sp}[\tilde{\delta}_{\rho}^B] + \text{mean}_{sp}[\tilde{\delta}_{\rho}^B]$. Therefore, we should always look beyond the obvious variance peaks at the symmetry axes for the true representation of the map error.

The analysis of 2D crystals is unique in its reconstruction method. Translational alignment error is found at two places. During image processing a reinterpretation method is used to “unbend” the crystal to correct for lattice distortions. This process in effect adjusts the translational alignment of the molecules so that information at the reciprocal lattice is all that is passed on in further analysis, thereby generating error. Other sources of error in the amplitude and phase values for each reflection are rotational variation of molecules within each unit cell, white-noise of the background, and variation in imaging conditions such as specimen drift and charging. Phase origin alignment between images is the other translational alignment that may

introduce variation in the reconstruction, although probably at a lower level due to the strong symmetry constraint. In addition, tilt geometry and the CTF are refined by the best fit of the image data against the fitted lattice lines, and they too have their own uncertainty. Since the scaling among images is based on the very simple model of total scattering, its variation may also account for variance in the reconstruction. Unfortunately, many of these sources of systematic error such as drift and any of the alignment and fitting errors cannot be separated from structure variation in the same data set, and each may affect the pixel-wise variation in different ways. For example, drift, translational alignment error, the CTF, and scaling probably add variance uniformly throughout the map. The effect of tilt geometry error, on the other hand, is probably not isotropic over the 3D volume. The consistent difference we observed between $\sigma_{\rho_q}^{-B}$ and $\delta_{\rho_i}^-$ was puzzling at first (compare Figs. 4b and c to 4a). We expect that, in whatever ways the variance map is affected by the factors described above, the estimate of the map by bootstrap resampling should be affected equally. This difference is not present when structure variance was the only source of the variance as in the simulated data set. Defocus distribution in the data set, which is the most likely parameter that is present in a distorted or polarized distribution, was found to be approximately symmetric and mono-modal. We also checked the bootstrap distribution of the pixels for possible strong deviations from a normal distribution and ruled it out as the source of the bias. When we examined $\bar{\tilde{\rho}_i}$, where $\tilde{\rho}_i$ is the reconstructed projection map from a single image, we found that $\bar{\tilde{\rho}_i}$ deviates from $\tilde{\rho}$ on average almost 5 times that of $\tilde{\rho}^B$ from $\tilde{\rho}$. We therefore suspect that the complexity of the reconstruction algorithm plays an important role in this discrepancy. For example, in the MRC algorithm AVGAMPHS that is used to average the projected structure factors in reciprocal space, the amplitudes are combined separately from the phases. Therefore, the combination is not truly vectorial. As a result, adding two maps together in real space is not equal to combining the structure factors through AVGAMPHS. Similar separation of amplitude and phase components is also used in LATLINE. Therefore, we believe that it is likely that our comparison between $\delta_{\rho_i}^-$ and $\sigma_{\rho_q}^{-B}$ is not meaningful for obtaining the bias level of the bootstrap method as an estimate of $\delta_{\rho_i}^-$. The effect is not apparent in the simulated structure because the lack of noise resulted in a very narrow distribution of the phase for a given (h, k) reflection. By adding different levels of Gaussian noise to the simulated crystal images, we confirmed that the deviation of $\bar{\tilde{\rho}_i}$ from $\tilde{\rho}$ increases with the increase of noise level and that bias of $\delta_{\rho_i}^-$ increases faster than that of $\sigma_{\rho_q}^{-B}$ as an estimate of $\delta_{\rho_i}^-$.

In our residual bootstrap attempt, many of the standard deviation peaks are elongated extensively in the z^* direction (red arrows in Fig. 7d). We believe that the cause of the elongation is the incorrect assumption that the measured structure factor vectors along z^* is homoscedastic. Plotting the fitting error against z^* for the lattice lines shows that data along some of the lattice lines are heteroscedastic (Fig. S2). In contrast, the simple bootstrap estimate does not require such an assumption of directional homoscedastic behavior but requires only that the variance in each voxel in the 3D map contributes to all projection images. Therefore, the latter estimate produced a better voxel independency.

Apart from the bootstrap estimator, the jackknife method is also commonly used in complicated systems. Not being a Monte Carlo simulation, the jackknife method has the obvious advantage of speed in comparison to the bootstrap technique. Which one is a better estimator of the variance is often determined by the parameter that is being estimated. Density map reconstruction involves Fourier transformation, filtering, and complex model fittings. As a result, predicting which method yields a better estimate is difficult. The best way to choose the

better estimator is to perform tests on simulated data of known variance. Our simulation (Fig. 2) showed that the bootstrap method outperforms the jackknife method in our particular application.

The recent paper by Penczek *et al.* describes the successful application of the bootstrap technique to estimate the 3D variance map reconstructed from single particle projection images (Penczek *et al.*, 2006b). In addition, they showed that the variance map can be used to focus further classification in the volume that displays high variance (Penczek *et al.*, 2006a). In this way there is potential to distinguish different conformational states of molecules within a sample. Molecules in 2D crystals are usually tightly packed so that the projection of one molecule often overlaps with another, especially when tilted. Therefore, it is unlikely that we can use the variation information for separation of heterogeneous molecules within a crystal.

Conclusions

We used the method of bootstrap resampling to estimate the standard error of the reconstructed density derived from image analysis of 2D crystals. We showed that the method can be applied to 3D data in a practical computation time. Due to the nature of the reconstruction algorithm and the high noise level, the estimates derived by bootstrap resampling are closer to the true values of the standard error of the reconstructed projection map than the standard error of the mean of individual reconstructions. High variance peaks at symmetry axes can be found as an artifact of symmetry averaging. While the local variation of the variance in gap junction 2D crystals was small, we showed from several simulations that the method can be reliably used for locating regions of high variance and to assess the significance of 3D difference maps.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Pawel Penczek for valuable discussions, James Pulokas for help in PYTHON programming, and Carolyn Lanigan for comments on the manuscript. This work was supported by NIH RO1 HL48908 (to MY). The computation resource from the National Center for Research Resources at NIH (RR17573) is gratefully acknowledged. Chimera is developed at the Resource for Biocomputing Visualization, and Informatics at the University of California, San Francisco (NIH P43 RR-01081).

abbreviations

2D	two-dimensional
3D	three-dimensional
CTF	contrast transfer function
Cx43-WT	gap junction crystals formed by full-length, connexin43
Cx43-TR	gap junction crystals formed by Cx43 that have been treated with trypsin
EM	electron microscopic
IQ	intensity quality of a reflection

REFERENCES

1. Agard DA, Stroud RM. Linking regions between helices in bacteriorhodopsin revealed. *Biophys. J* 1982;37:589–602. [PubMed: 7074187]

2. Babu, GJ.; Rao, CR. Bootstrap Methodology. In: Rao, CR., editor. Computational statistics. Vol. 9. Elsevier Science; Amsterdam: 1993. p. 627-659.
3. Blow DM, Crick FHC. The treatment of errors in the isomorphous replacement method. *Acta Cryst* 1959;12:794–802.
4. Cheng A, Schweissinger D, Dawood F, Kumar NM, Yeager M. Projection structure of full length connexin 43 by electron cryo-crystallography. *Cell Communication and Adhesion* 2003;10:187–191. [PubMed: 14681014]
5. Collaborative Computational Project. The CCP4 Suite: Programs for protein crystallography. *Acta Cryst. D* 1994;50:760–763. [PubMed: 15299374]
6. Crowther RA. Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. *Philos. Trans. R. Soc. Lond. B Biol. Sc* 1971;261:221–230. [PubMed: 4399207]
7. Crowther RA, Henderson R, Smith JM. MRC image processing programs. *J. Struct. Biol* 1996;116:9–16. [PubMed: 8742717]
8. Devore, J. L. Statistics: the exploration and analysis of data. Duxbury; Pacific Grove: 2001.
9. Dixon, PM. Bootstrap resampling. In: El-Shaarawi, AH.; Piegorsch, WW., editors. Encyclopedia of Environmetrics. Vol. 1. John Wiley and Sons; New York: 2002. p. 212-219.
10. Efron B. Bootstrap methods: Another look at the jackknife. *Ann. Statist* 1979;7:1–26.
11. Fleishman SJ, Unger VM, Yeager M, Ben-Tal N. A C^α model for the transmembrane α helices of gap junction intercellular channels. *Molecular Cell* 2004;15:879–888. [PubMed: 15383278]
12. Govindarajulu, Z. Elements of Sampling Theory and Methods. Prentice-Hall; Upper Saddle River, NJ: 1999.
13. Härdle W, Bowman AW. Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc* 1988;83:102–110.
14. Henderson R, Baldwin JM, Downing KH, Lepault J, Zemlin F. Structure of purple membrane from halobacterium halobium: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. *Ultramicroscopy* 1986;19:147–178.
15. Henderson R, Moffat JK. The difference Fourier technique in protein crystallography: Errors and their treatment. *Acta Cryst* 1971;B27:1414–1420.
16. Liu W, Boisset N, Frank J. Estimation of variance distribution in three-dimensional reconstruction. II. Applications. *J. Opt. Soc. Am. A* 1995;12:2628–2635.
17. Liu W, Frank J. Estimation of variance distribution in three-dimensional reconstruction. I. Theory. *J. Opt. Soc. Am. A* 1995;12:2615–2627.
18. Milligan RA, Flicker PF. Structural relationships of actin, myosin, and tropomyosin revealed by cryo-electron microscopy. *J. Cell Biol* 1987;105:29–39. [PubMed: 3611188]
19. Mitra AK, Miercke L, Turner GL, Shand RF, Betlach MC, Stroud RM. Two-dimensional crystallization of Escherichia coli-expressed bacteriorhodopsin and its D96N variant: High resolution structural studies in projection. *Biophys. J* 1993;65:1295–1306. [PubMed: 8241409]
20. Molinara AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005;21:3301–3307. [PubMed: 15905277]
21. Penczek P, Frank J, Spahn CMT. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J. Struct. Biol* 2006a; 154:184–194. [PubMed: 16520062]
22. Penczek PA, Yang C, Frank J, Spahn CMT. Estimation of variance in single particle reconstruction using the bootstrap technique. *J. Struct. Biol* 2006b;154:168–183. [PubMed: 16510296]
23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem* 2004;25:1605–1612. [PubMed: 15264254]
24. Quenouille M, H. Approximate tests of correlation in time series. *J. Roy. Statist. Soc* 1949;B11:68–84.
25. Samuels, ML. Statistics for the life sciences. Dellen Publishing; San Francisco: 1989.
26. Schertler GFX, Villa C, Henderson R. Projection structure of rhodopsin. *Nature* 1993;362:770–772. [PubMed: 8469290]

27. Smit E, Oling F, Demel R, Martinez B, Pouwels PH. The S-layer protein of *Lactobacillus acidophilus* ATCC 4356: Identification and characterisation of domains responsible for S-protein assembly and cell wall binding. *J. Mol. Biol* 2001;305:245–257. [PubMed: 11124903]
28. Suloway C, Pulokas J, Fellmann D, Cheng A, Guerra F, Quispe J, Stagg S, Potter CS, Carragher B. Automated molecular microscopy: The new Leginon system. *J. Struct. Biol* 2005;151:41–60. [PubMed: 15890530]
29. Tate CG, Ubarretxena-Belandia I, Baldwin JM. Conformational changes in the multidrug transporter EmrE associated with substrate binding. *J. Mol. Biol* 2003;332:229–242. [PubMed: 12946360]
30. Unger VM, Kumar NM, Gilula NB, Yeager M. Projection map of a gap junction channel at 7 Å resolution. *Nature Struct. Biol* 1997;4:39–43. [PubMed: 8989321]
31. Unger VM, Kumar NM, Gilula NB, Yeager M. Three-dimensional structure of a recombinant gap junction membrane channel. *Science* 1999;283:1176–1180. [PubMed: 10024245]
32. Walters SJ, Campbell MJ. The use of bootstrap methods for estimating sample size and analysing health-related quality of life outcomes. *Stats. in Medicine* 2005;24:1075–1102.

Figure 1.

Illustration of the sampling distribution meta experiment (a,b,c) and its correspondence to the bootstrap resampling method (d,e,f) for a population distribution (gray curve) of colored balls. In a, b and c the sample mean is the quantity of interest, represented by the green ellipse. (a) With 7 samples taken from a known population, a mean can be calculated. (b) By repeating the sampling process without changing the population, a mean is calculated for each of the 7 sample sets. The value of the mean changes with each sampling, as shown by the different colors (blue and purple) for the sample ellipses. The sampling distribution shown by the red curve in (c) is then the distribution of these means. Note that this distribution is narrower compared with the population distribution (gray). (d) When the population distribution is not

known, bootstrap resampling can be used to estimate the distribution. As in (a), a single value of the quantity of interest can be calculated from the 7 samples. (e) The bootstrap resampling estimate is generated using the 7 sampled balls shown in the gray rectangular box. Samples can be repeatedly drawn with replacement so that the probability of sampling a particular ball is not altered by the resampling process. For each loop, the resampling is repeated 7 times to match the number of the original sampling from the population. Since the resampling follows the distribution of the samples, the many calculated values for the quantity of interest from a large number of loops will distribute, shown in different colors (blue and purple), according to the propagated distribution that originated from the same distribution. This process yields a good estimate for the distribution of the quantity, shown by the red curve in (f).

Figure 2.

Application of the bootstrap method to projection images of simulated crystals that contain only structure variation. The unit cells in (a) and (b) contain two possible conformations of an artificial Gaussian molecule. Shown in (c) is the reconstructed projection map of 17 simulated crystals displaying an average of the two conformations. The inset shows a portion of a simulated 2D crystal. (d) Standard error map of the mean for an individual reconstruction, $\delta_{\rho_i}^-$, calculated from Eqns. 1-4. (e) and (f) show the Bootstrap estimation, $\sigma_{\rho_q}^B$, and Jackknife estimation, $\widehat{\delta}_{\rho}^J$, respectively, of the standard error map, at the same contour level as in (d).

Figure 3.

Noise level dependence of the bias of the estimation for reconstructed map mean and standard error from simulated crystals. (a,b) Standard error of the reconstruction, $\delta_{\tilde{\rho}}$ (blue), is estimated by $\delta_{\tilde{\rho}_i}$ (red) or $\sigma_{\tilde{\rho}_q^B}$ (green). (c,d) Reconstructed density, $\tilde{\rho}$, is estimated by $\tilde{\rho}_i$ or $\tilde{\rho}^B$, shown in the same color scheme. The two positions of the unit cell where the plots are indicated on the inset map in (b). Results at position **a** are shown in (a,c), and those at **b** are in (b,d).

Figure 4.

Surface-shaded views (a-d) of the 3D density map of Cx43-WT contoured at $1.5 \, sd_{sp} \left[\tilde{\rho}_{WT} \right]$. The boundaries shown in the side view (a) indicate the locations of sections for (b-d). The arrows indicate the viewing directions. Shown in (b), (c) and (d), respectively, are the helices within the top connexon, a cross-section showing the continuous wall of protein density that lines the extracellular vestibule, and a cross-section showing the helices within the bottom connexon.. The wedge lines in (b) demarcate a pie slice used for the representations in Figure 5. The crystallographic and one of the non-crystallographic 2-fold symmetry axes are indicated

in (c). Shown in (e) is a projection map contoured at $sd_{sp}[\delta_{\Delta\rho}]$. We note that the viewing direction of this map and those shown in Fig. 5 are consistent with the viewing direction of the 3D maps in our (Unger *et al.*, 1999). This is opposite to that of the projection maps we published previously (Unger *et al.*, 1997; Cheng *et al.*, 2003) Scale bars: 20 Å

Figure 5.

Maps of the standard error of the mean projection density derived from 2D crystals of Cx43-TR. Seventeen images with tilt angle values less than 4° were used in the analysis. (a) Standard

error, $\delta_{\rho_i}^-$, of the mean of the individually reconstructed map calculated according to Eqns. 1-4.

(b) and (c) Estimated standard error of the reconstruction, $\widehat{\delta}_{\rho}^B$, from two independent bootstrap resamplings, each consisting of 200 cycles. The red outlines show the approximate boundary of the projected protein density. All maps are scaled and contoured equally, with each contour corresponding to the rms deviation of the pixel values in $\delta_{\rho_i}^-$. The similarity between (b) and (c) confirms that the number of cycles of bootstrap resampling was adequate.

Figure 6.

Simple bootstrap estimation test on a 3D map with large variations in structure. All maps shown are asymmetric unit pie slices around the top connexon as indicated in Figure 4b. (a) and (b) are the mirrored pair of Cx43-WT in yellow and white, respectively. (c) surface-shaded representation of the 3D reconstruction of the combined data set (light purple), contoured at a higher value than the wire-frame rendering (white and yellow) from (a) and (b). (d) 3D reconstruction of the combined data in wire-frame rendering (purple), shown with a line indicating one of the mirror planes. The map has approximate 6/mmm point group non-crystallographic symmetry. (e) and (f) Bootstrap estimated standard error map (green) superimposed on the same density in (d). The line in (f) indicates the location of the mirror plane passing through the middle of the helical bundle. The pseudo mirror symmetry is apparent after all the processing that used only p6 symmetry. Scale bar: 20 Å.

Figure 7.

Bootstrap estimated standard error maps of the difference, $\widehat{\delta}_{\Delta\rho}^B$ between Cx43-WT and Cx43-TR. The green maps in (a) and (c) show the results for simple boot strapping, surface rendered at $1.5 \text{ } sd_{sp}[\widehat{\delta}_{\Delta\rho}^B] + \text{mean}_{sp}[\widehat{\delta}_{\Delta\rho}^B]$. The green maps in (b) and (d) show the results for residual bootstrapping, scaled to the peak on the 6-fold axis in (a) and (c). As a reference, the error map is superimposed on the surface-shaded 3D map of Cx43-WT (gold), rendered at 1.5

$sd_{sp}[\tilde{\rho}_{WT}]$. The top views in (a) and (b) are through a section of the extracellular gap, indicated by the white boxes in (c) and (d). This section contains the largest non-crystallographic error peaks, which border the lumen of the extracellular vestibule. The dashed line in (a) indicates the slicing plane for the side views shown in (c) and (d). Scale Bar: 20 Å.

Figure 8.

Similarity of the 3D difference and t -density maps between Cx43-WT and Cx43-TR. Top (a) and side views (c) of the difference map, contoured at $2.9 \text{ mean}_{sp}[\widehat{\delta}_{\Delta\rho}^B]$. The positive difference, i.e., $\Delta \tilde{\rho} = \tilde{\rho}_{WT} - \tilde{\rho}_{TR} > 0$, is colored in red and negative difference in sky blue. Top (b) and side views (d) of the t -test map surface rendered at $t = \pm 2.681$ or 99% confidence level for 47 degree of freedom. Positive t values are colored in orange and negative values in violet. As a reference, the maps are superimposed on the surface-shaded 3D map of Cx43-WT (gold), rendered at $1.5 \text{ sd}_{sp}[\tilde{\rho}_{WT}]$. The sections and viewing direction are identical to that of Figure 7.