CrossMark

electronic reprint

Journal of
Applied Crystallography
ISSN 1600-5767

Update for *BayesApp*: a web site for analysis of small-angle scattering data

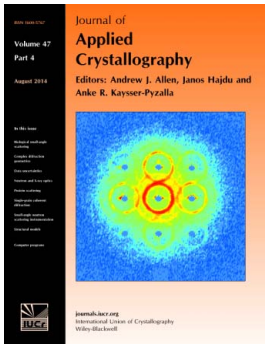
Steen Hansen

J. Appl. Cryst. (2014). **47**, 1469–1471

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see http://journals.iucr.org/services/authorrights.html



Many research topics in condensed matter research, materials science and the life sciences make use of crystallographic methods to study crystalline and non-crystalline matter with neutrons, X-rays and electrons. Articles published in the *Journal of Applied Crystallography* focus on these methods and their use in identifying structural and diffusion-controlled phase transformations, structure-property relationships, structural changes of defects, interfaces and surfaces, *etc*. Developments of instrumentation and crystallographic apparatus, theory and interpretation, numerical analysis and other related subjects are also covered. The journal is the primary place where crystallographic computer program information is published.

Crystallography Journals Online is available from **journals.iucr.org**

Journal of
Applied Crystallography
ISSN 1600-5767

Received 4 April 2014
Accepted 5 June 2014

Update for *BayesApp*: a web site for analysis of small-angle scattering data

Steen Hansen

Niels Bohr Institute, University of Copenhagen, Faculty of Science, Universitetsparken 5, DK-2100 KBH Ø, Denmark. Correspondence e-mail: slh@nbi.dk

An update for *BayesApp*, a web site for analysis of small-angle scattering data, is presented. The indirect transformation of the scattering data now includes an option for a maximum-entropy constraint in addition to the conventional smoothness constraint. The maximum-entropy constraint uses an ellipsoid of revolution as a prior, and the dimensions of the ellipsoid as well as the overall noise level of the experimental data are estimated using Bayesian methods. Furthermore, a correction for slit smearing has been added. The web site also includes options for calculation of the scattering intensity from simple models as well as the estimation of structure factors for polydisperse spheres and nonspherical objects of axial ratios between 0.4 and 2.5.

© 2014 International Union of Crystallography

1. Introduction

The web site http://www.bayesapp.org is intended for analysis of small-angle scattering data from scatterers in solution. The main aim of the analysis of such data is to obtain structural information about the scatterers. This can be done either by comparing the scattering intensity of a model of the scatterer with the experimental data or by indirect estimation of a real-space distribution characteristic for the scatterer (*e.g.* Glatter & Kratky, 1982). The web site *BayesApp* allows for both approaches.

As the information content of small-angle scattering data is usually quite low, the indirect method requires the introduction of some form of regularization in order to obtain a unique solution for the distribution of interest. This is conventionally done by adding a smoothness constraint to the indirect transformation of the scattering data (Glatter, 1977), leading to a final result which is the smoothest distribution in accordance with the experimental data (*i.e.* yielding a fit of the data having a sensible χ^2 value or a similar misfit function).

The low information content of the data implies that the regularization influences the final estimate. Therefore, it is relevant to test the outcome of the analysis using different regularization constraints to see which features of the estimate may be attributed to the constraint and which features may be considered to be real. For this purpose, the maximum-entropy method offers an additional constraint which is transparent as the bias is given in the form of a prior which is easily compared with the final estimate. Therefore, the maximum-entropy constraint has been implemented at the *BayesApp* web site for indirect transformation.

As an additional option, the direct calculation of the scattering intensity from simple models has been included at the web site. Furthermore, for nondilute solutions appropriate structure factors have to be taken into account, which is now also possible. For calculation of the structure factor it is necessary that the scatterer has an axial ratio between 0.4 and 2.5. Finally, a correction for slit smearing has been added. This has been done by the addition of an extra smearing matrix in the transformation from the distance distribution functions to the scattering intensity. Details are given by Glatter & Kratky (1982, p. 132) and at the web site, where the width of the Gaussian beam profile is to be specified by the user.

2. Indirect transformation

With regard to the input and output, the web site has been described by Hansen (2012). Except for the experimental data, no user input is necessary, and with the default settings the estimation of the distribution of interest is carried out in just a few seconds of CPU time.

2.1. Maximum entropy

Regularization by maximum entropy selects the distribution that is as close as possible to some prior distribution while simultaneously fitting the experimental data adequately. The distance between the prior distribution and the final estimate is measured by the entropy functional (*e.g.* Hansen & Pedersen, 1991) and the optimal χ^2 value is determined using Bayesian methods (Hansen, 2000).

In the form implemented at the web site, the prior is an ellipsoid of revolution having dimensions that are also estimated by Bayesian methods. This means that the final estimate of the distance distribution function will be as close as possible to that of an ellipsoid of revolution. The ellipsoidal shape is fairly general and corresponds well to a large class of molecules.

In addition to providing an additional method for regularization, this maximum-entropy constraint may, for globular structures, lead to a more realistic estimate of the distribution than the smoothness constraint which simply minimizes the curvature of the estimated distribution. By nature the smoothness constraint favors linearity, which is not in agreement with the true shape of the distance distribution $p(r)$ for either short or large distances, where the smoothness constraint may lead to a truncation of $p(r)$. This means that the smoothness constraint often underestimates $d_{\max}$, especially for the frequent case of globular scatterers, without any substantial penalty for the misfit function. For a sphere, only about one percent of the total area of the distance distribution functions is found in the last ten percent of $p(r)$ and only about 0.1 percent is found in the last five percent, *i.e.* in the interval from $0.95d_{\max}$ to $d_{\max}$, where $d_{\max}$ is the maximum dimension of the scatterer.

For globular scatterers, this systematic error may be avoided using maximum entropy with an ellipsoidal prior, thus ensuring that the shape of $p(r)$ will be biased towards a smooth and physically correct transition to zero at $r = 0$ and $r = d_{\max}$.

electronic reprint

2.2. Examples

For illustration of the influence of regularization, *BayesApp* was used for simulation of scattering intensities (as described below in §3.1). Subsequently, the distance distribution function was estimated using *BayesApp* with a smoothness constraint as well a maximum-entropy constraint as described above.

For the examples shown below, 50 points in the q range of 0.015–0.15 nm^{-1} were used. Gaussian noise was added with a standard deviation σ at each point i according to $\sigma_i = 0.05I(q_i) + 0.001I(0)$. For estimation of the distance distribution $p(r)$, 100 points were used.

In Fig. 1(*a*), the simulated intensity for a sphere is shown. The corresponding distance distribution function estimated using the smoothness constraint is shown in Fig. 1(*b*). The most likely solution selected by the Bayesian approach is shown as well as an example of a probability weighted average of 1000 different solutions. The calculation of 1000 solutions requires about 30 s of CPU time, but it is sufficient to illustrate the difference between the optimal and the average solution.

In spite of the improvement of the average solution, it is clear from the overall shape of the estimates of $p(r)$, as well as the behavior around $d_{\max}$, that the smoothness constraint leads to estimates which differ systematically from the distance distribution of a sphere. Using the maximum-entropy constraint with an ellipsoidal prior gives a solution that is indistinguishable from the original distribution.

In Fig. 1(*c*), the simulated intensity from a multishell structure is shown. The corresponding distance distribution functions estimated using the smoothness constraint and the maximum-entropy constraint are both shown in Fig. 1(*d*) (average of 1000 solutions).

It is evident that the smoothness constraint has a propensity to smear out the structure in the solution. By contrast, the maximum-entropy constraint allows the features of the distance distribution function to be reproduced. It may be noted that the simulated data were fitted to the correct value (*i.e.* given by the simulated noise) for both the smoothness and the maximum-entropy constraint; in reciprocal space there was no difference in the quality of the fits for the two methods (both having a reduced χ^2 of 0.88).

Furthermore, the smoothness constraint also impairs the estimate at short distances as may be seen from the insert in Fig. 1(*d*). Once more, the maximum-entropy constraint gives a better estimate for the globular structure.

In Fig. 1(*e*), the simulated intensity for two overlapping spheres is shown. The corresponding distance distribution functions estimated using the two different constraints are shown in Fig. 1(*f*). Even though this particular structure is not similar to that of an ellipsoid of revolution, the maximum-entropy constraint leads to a better overall estimate of the original distribution.

However, in spite of these improvements, the main advantage of the introduction of an extra constraint is that for globular structures it provides a means for an easy evaluation of the influence of the regularization on the final estimate.

3. Direct modeling

3.1. Calculation of the intensity

The scattering object may be composed from a number of simple subunits (maximum ten): spheres, ellipsoids, cylinders or shells as described by Hansen (1990). The scattering length density of each subunit may specified. From the model, the scattering intensity is calculated to about four orders of magnitude in just a couple of seconds of CPU time.

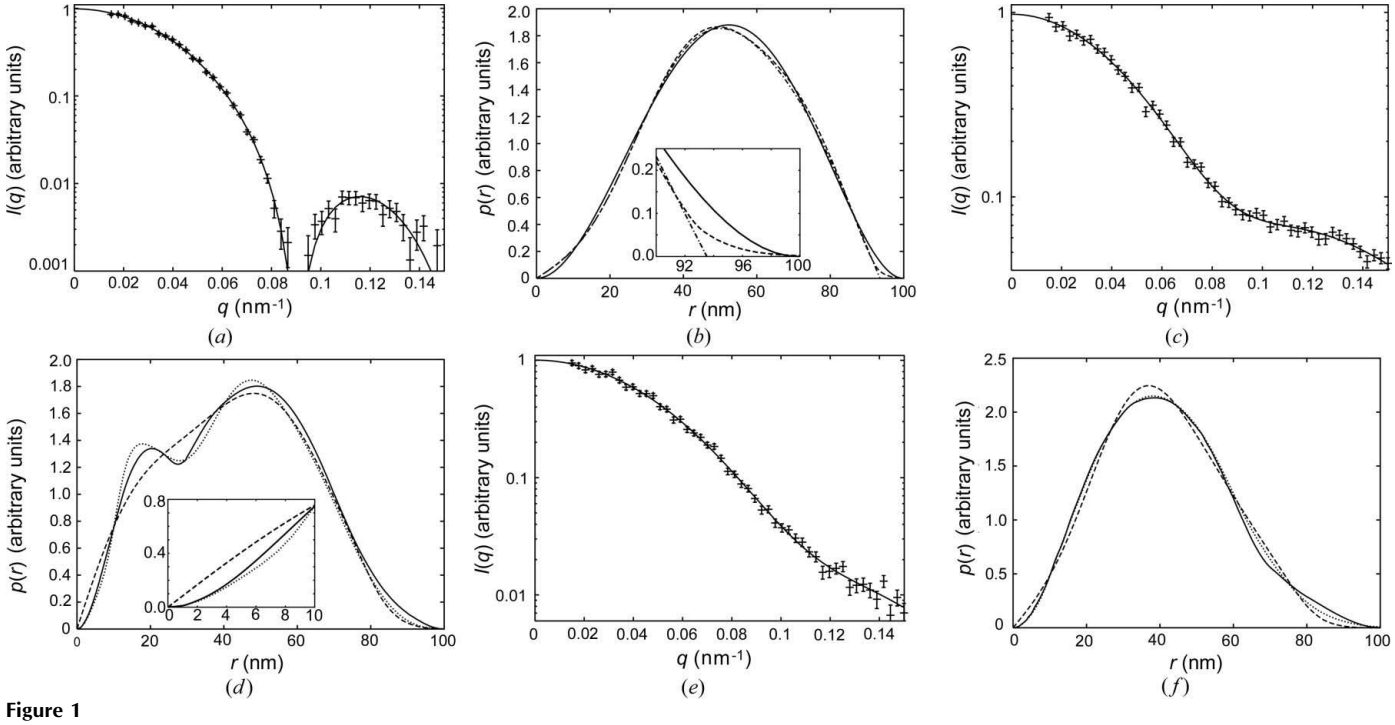


Figure 1
Simulated intensities and estimates of the distance distribution functions. For the intensities, the error bars show the simulated data and the full line indicates the fit. For the distance distribution functions, the full line is the original distribution, the dash–dotted line is the estimate using the smoothness constraint, the dashed line is the estimate using the smoothness constraint and average of 1000 solutions, and the dotted line is the estimate using the maximum-entropy constraint. In (*a*) and (*b*), the sphere is of radius 50 nm. The insert in (*b*) shows the details near $d_{\max}$. The multishell structures (*c*) and (*d*) comprise a sphere of radius $r = 20$ nm with scattering length density $\rho = 1$, a shell of radii 20–30 nm with $\rho = 0.2$ and a shell of radii 30–50 nm with $\rho = 0.1$. The insert in (*d*) shows the details at the shortest distances. The simulations in (*e*) and (*f*) use two overlapping spheres. The scattering length density of the overlapping region was made equal to the scattering length density of the individual spheres. The radius for both spheres was 33.3 nm, which was also their center-to-center distance.

electronic reprint

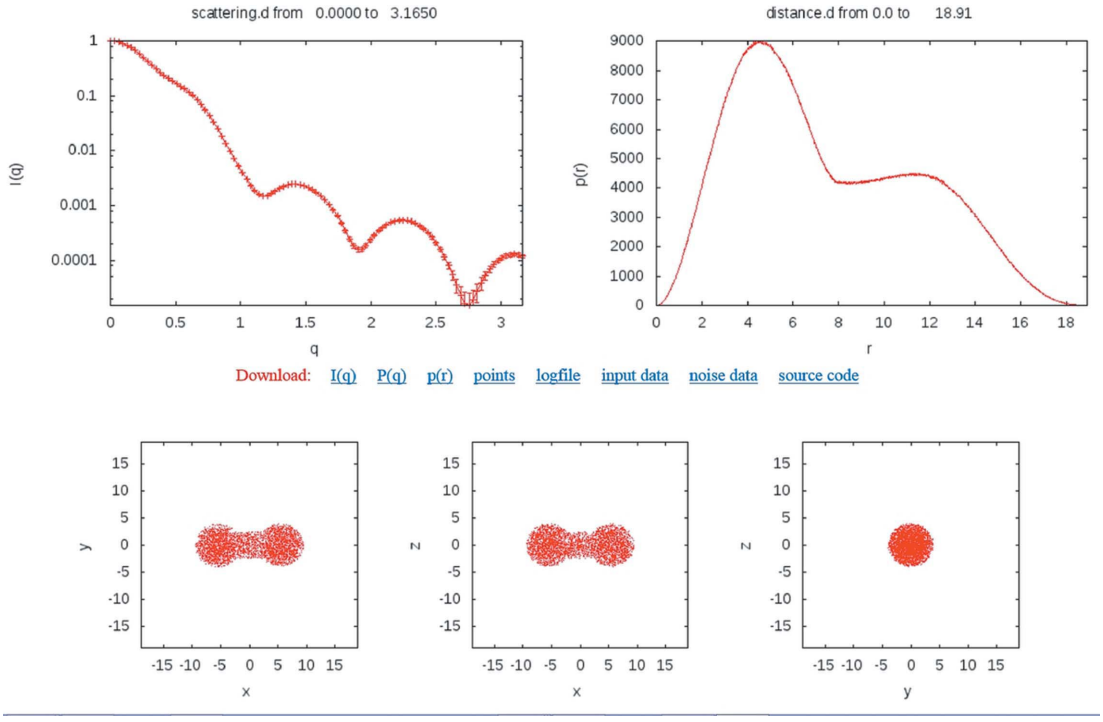


Figure 2
Example of a screen output from *BayesApp*. Calculation of the intensity from two spheres connected by a rod as shown in the figure.

This is shown in the program output in Fig. 2, which in addition to the calculated intensity also shows the distance distribution functions as well as the model visualized from three different angles.

For simulation purposes, random Gaussian noise, background and smearing may be added to the calculated intensity at a number of data points and in a q range specified by the user. It may be noted that the result of such a simulation is used as the default input for estimation of the distance distribution at the web site.

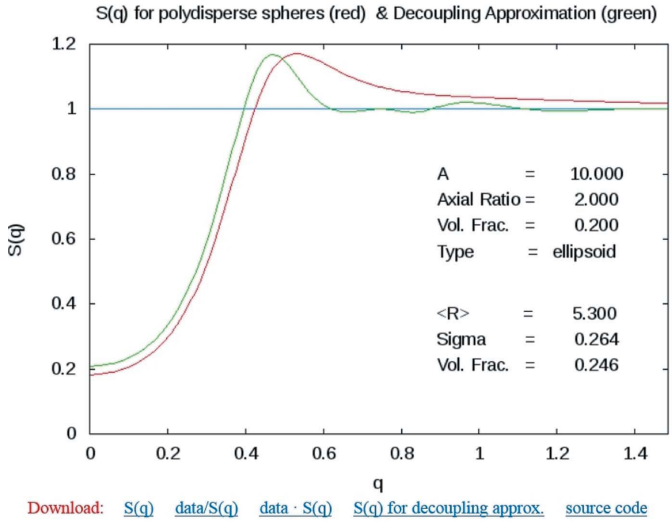


Figure 3
Example of a screen output from *BayesApp*. Calculation of the structure factor for a prolate ellipsoid of revolution of maximum dimension 20 nm and axial ratio 2.0 at a volume fraction of 0.2. The green line is the decoupling approximation and the red line is the approximation using polydisperse spheres.

3.2. Calculation of the structure factor

For nondilute solutions the structure factor should be included in the modeling. For this purpose the structure factors for polydisperse spheres, cylinders or ellipsoids of relatively low axial ratio and volume fraction may be estimated (Vrij, 1979; Hansen, 2013) and compared with the results of the decoupling approximation (Kotlarchyk & Chen, 1983).

An example of the output of this part of the web site is shown in Fig. 3, where the structure factor for a spheroid of axial ratio 2.0 is calculated using an approximation of polydisperse spheres as well as the decoupling approximation.

4. Availability

The programs described are all written in Fortran and contain subroutines given by Press *et al.* (1992). The source codes are all freely available for download, modification and re-use without restrictions from the web site http://www.bayesapp.org.

References

Glatter, O. (1977). *J. Appl. Cryst.* **10**, 415–421.
Glatter, O. & Kratky, O. (1982). Editors. *Small Angle X-ray Scattering*. London: Academic Press.
Hansen, S. (1990). *J. Appl. Cryst.* **23**, 344–346.
Hansen, S. (2000). *J. Appl. Cryst.* **33**, 1415–1421.
Hansen, S. (2012). *J. Appl. Cryst.* **45**, 566–567.
Hansen, S. (2013). *J. Appl. Cryst.* **46**, 1008–1016.
Hansen, S. & Pedersen, J. S. (1991). *J. Appl. Cryst.* **24**, 541–548.
Kotlarchyk, M. & Chen, S.-H. (1983). *J. Chem. Phys.* **79**, 2461–2469.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in Fortran 77. The Art of Scientific Computing*, 2nd ed. New York: Cambridge University Press.
Vrij, A. (1979). *J. Chem. Phys.* **71**, 3267–3270.

electronic reprint