How to interpret the BUSTER reciprocal space correlation coefficients plot

Gérard Bricogne

These graphs display the correlation coefficients CC(X,Y) between pairs of quantities X and Y attached to unique reflections in given resolution bins, as well as the completeness of the observed data in those bins.

The four distinct quantities being correlated are all structure factor amplitudes. They are defined and denoted as follows (the underlying complex-valued structure factors being denoted in bold).

Fo is the observed structure factor amplitude, with σ(Fo) denoting its observational e.s.u. (estimated standard uncertainty).

Ffrg is the amplitude of the complex structure factor **F**frg computed from the atomic model, without the contributions from the continuous distribution of missing atoms (if any) nor from the bulk solvent.

Fc is the amplitude of the calculated total complex structure factor **F**c, including the contributions from the atomic model, from the continuous distribution of missing atoms (if any) and from the bulk solvent.

Fxpct is the expectation value of the observable structure factor amplitude for the total structure, whose distribution incorporates both the error model relating the true but unknown structure factor **F**true to the current **F**c, and the knowledge that |**F**true| would be measured with a scaled-corrected error of σ(Fo). It is this distribution that is used to form the likelihood function on which the refinement is based.

We also define the quantity δ as being σ(Fo) x N(0,1), N(0,1) being a normal (i.e. Gaussian) random variate with mean 0 and standard deviation 1.

The four correlation coefficients displayed on the CC plot are as follows.

CC(Fo,Ffrg) Measures the agreement between the observed data and the atomic model devoid of any contribution from continuously distributed missing atoms or bulk solvent. The absence of bulk solvent correction make this agreement predictably bad at resolutions worse than 5Å.

CC(Fo,Fc) Measures the agreement between the observed data and the total real-space model, including contributions from the atomic model, from the continuously distributed missing atoms (if any) and the bulk solvent. The improvement over CC(Fo,Ffrg) at resolutions worse than 5Å measures the adequacy of the bulk solvent correction.

CC(Fc,Fxpct) Measures the expected loss of correlation when the belief that the total structural model is error-free and the amplitude of its transform is measured exactly is downgraded to the belief that both the structural model and the measurement process are affected by the currently estimated level of error.

CC(Fo,Fo+δ) Measures the expected loss of correlation between Fo and itself after it has been perturbed by a typical observational error with e.s.u. σ(Fo).

Rationale of the RecSCC plot.

The examination of the RecSCC plot affords a check on the adequacy of both the experimental data and the current structural model, and of the error models associated with them, on the basis of the following, admittedly rather simplistic, arguments.

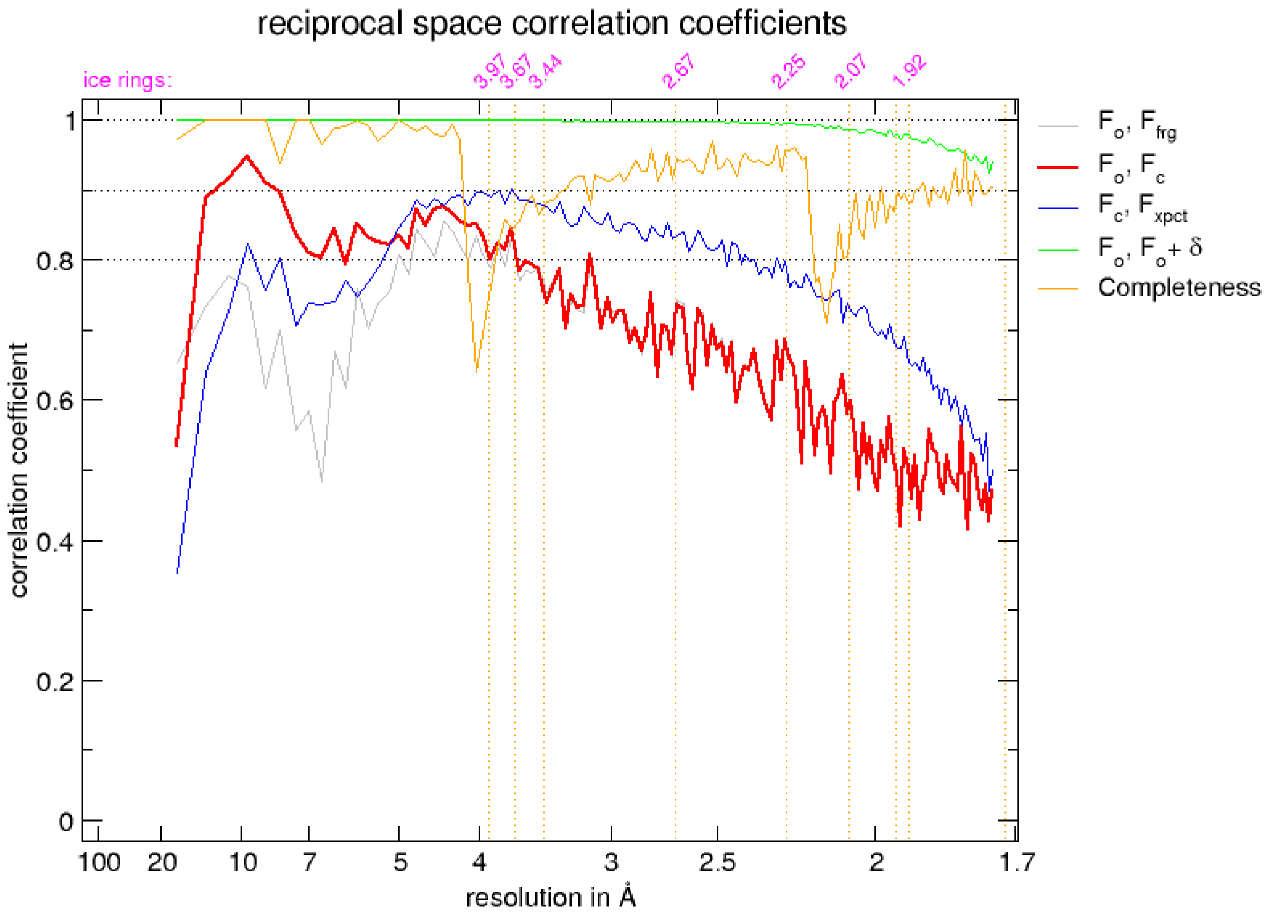
Let us assume that the atomic model, with its bulk solvent correction, were perfect and gave Fc values that coincided exactly with those of the true Fobs. As Fo is a measurement of Fobs with e.s.u. σ(Fo), this implies that CC(Fo,Fc) must be less than CC(Fo,Fo+δ): the CC(Fo,Fo+δ) curve is therefore a “ceiling” for the CC(Fo,Fc) curve.

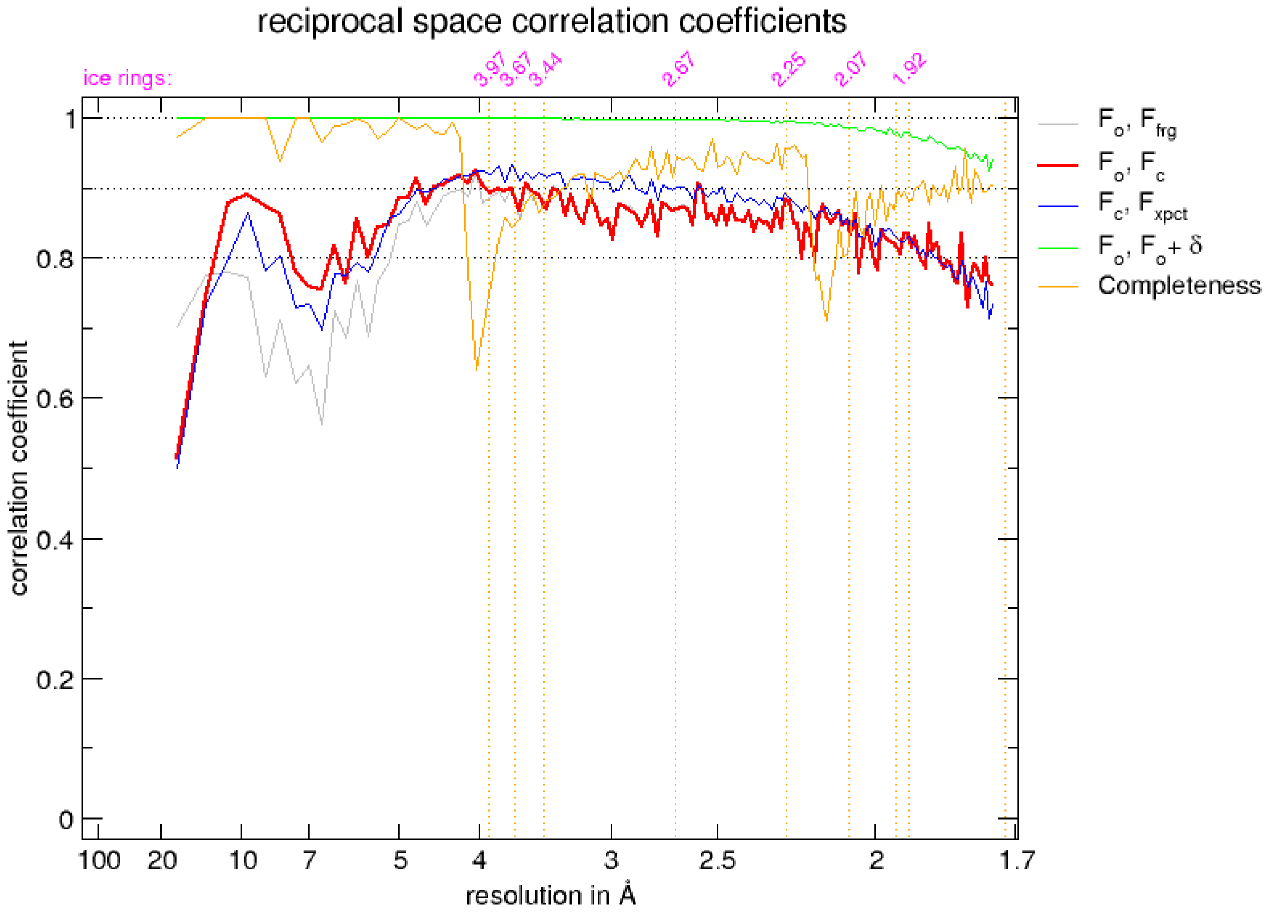
Let us now assume that our estimate of the uncertainty on the atomic model from which Fc is calculated is adequate, as is our error estimate σ(Fo) on the the observed amplitude Fo. The true structure factor being a typical member of the ensemble of perturbed atomic models describing the structural mode uncertainty, and its amplitude being measured within σ(Fo), the final observation Fo will be decorrelated from Fc in exactly the same way as is Fxpct.

This is the most important prediction of this analysis: if the error models for the uncertainty of the structural model and the observational e.s.u.'s and are adequate, then we should have CC(Fo,Fc)≈CC(Fc,Fxpct), and these two CC values should have CC(Fo,Fo+δ) as a ceiling.

Main use of the RecSCC plot: monitoring progress.

Typical example (internal project) of a pair of initial and final plots:


reciprocal space correlation coefficients
ice rings:
3.97
3.67
3.44
2.67
2.25
2.07
1.92
correlation coefficient
1
0.8
0.6
0.4
0.2
0
100
20
10
7
5
4
3
2.5
2
1.7
resolution in Å
F_o, F_{frg}
F_o, F_c
F_c, F_{xpct}
F_o, $F_o+\delta$
Completeness



Ice rings and “icicles”.

The presence of ice rings on diffraction images will result in a variety of prominent features in some of the graphs, in the form of sharp drops in the plotted values within narrow resolution ranges, which on account of their visual aspect we call “icicles”.

For convenience in confirming this diagnosis, the CC plot indicates the resolutions at which ice rings are know to occur. The main categories of symptoms manifested through the presence of icicles are as follows.

1. If the reflections potentially affected by ice rings have been rejected during data processing as a precautionary measure, there will be icicles in the Completeness graph in the corresponding resolution bins.

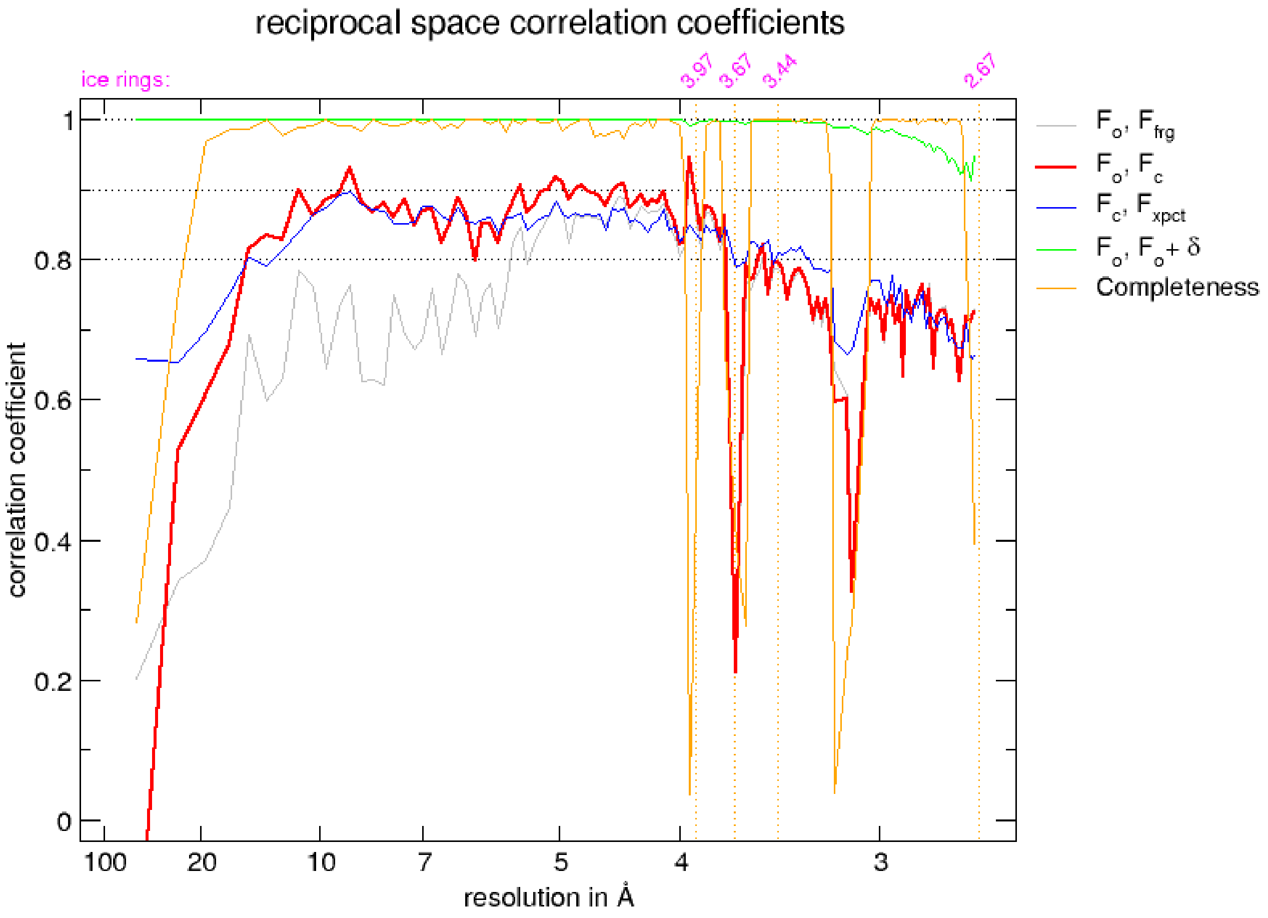
2. If no rejection has taken place, then there are two possibilities, depending on whether or not the value of σ(Fo) has been correctly estimated to reflect the potential contamination by an ice ring.

a) If σ(Fo) is suitably large, there will be icicles in the plots for

- CC(Fo,Fo+δ), by definition of the latter;
- CC(Fo,Fc) because the Fc do not predict ice ring contamination;
- CC(Fc,Fxpct) because the observational σ(Fo) is taken into account in the calculation of Fxpct.

b) If σ(Fo) fails to indicate potential ice-ring contamination, CC(Fo,Fo+δ) and CC(Fc,Fxpct) will not show icicles, but CC(Fo,Fc) will.

Example: 3lwq showing ice-ring problems.

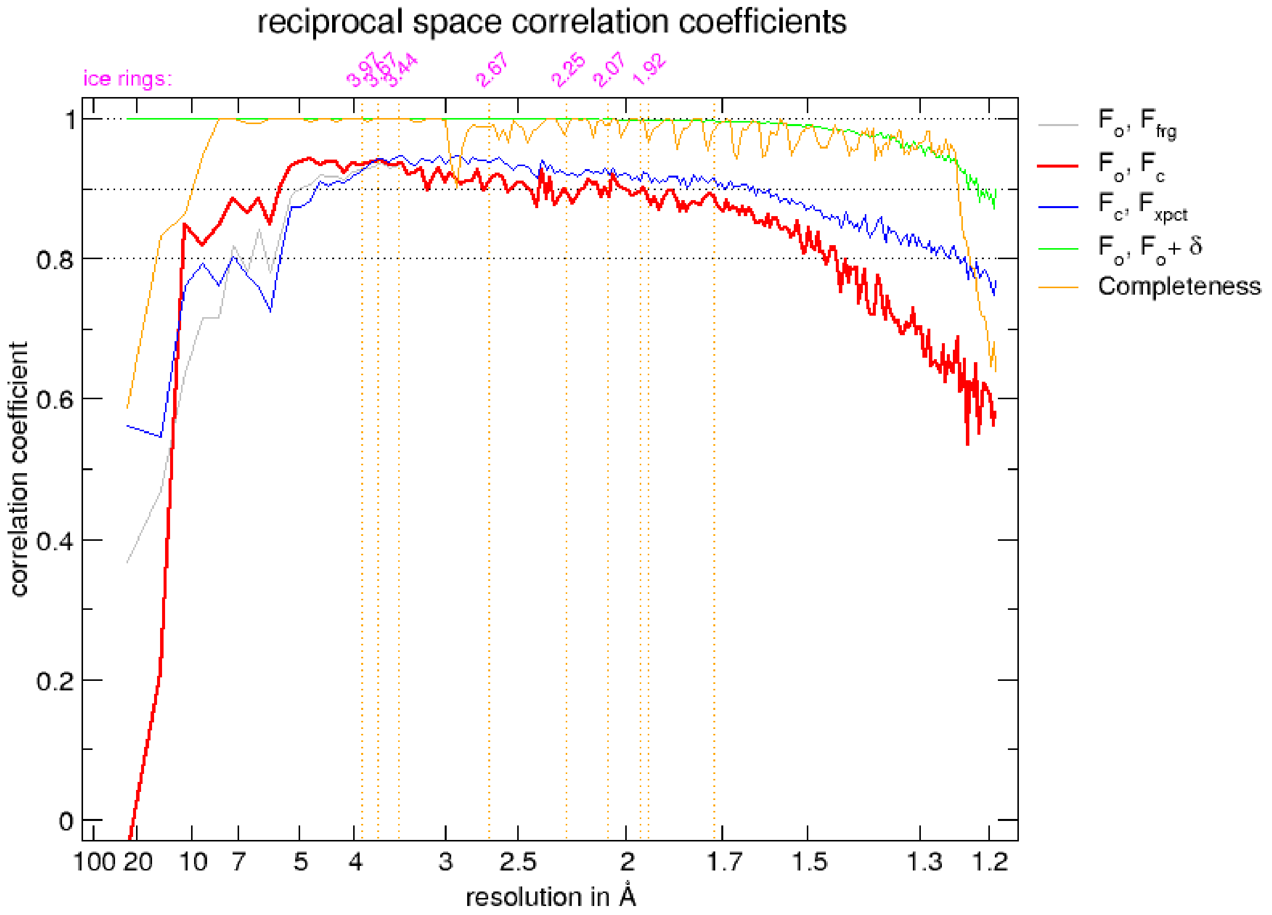


Discrepancies between CC(Fo,Fc) and CC(Fo,Fxpct).

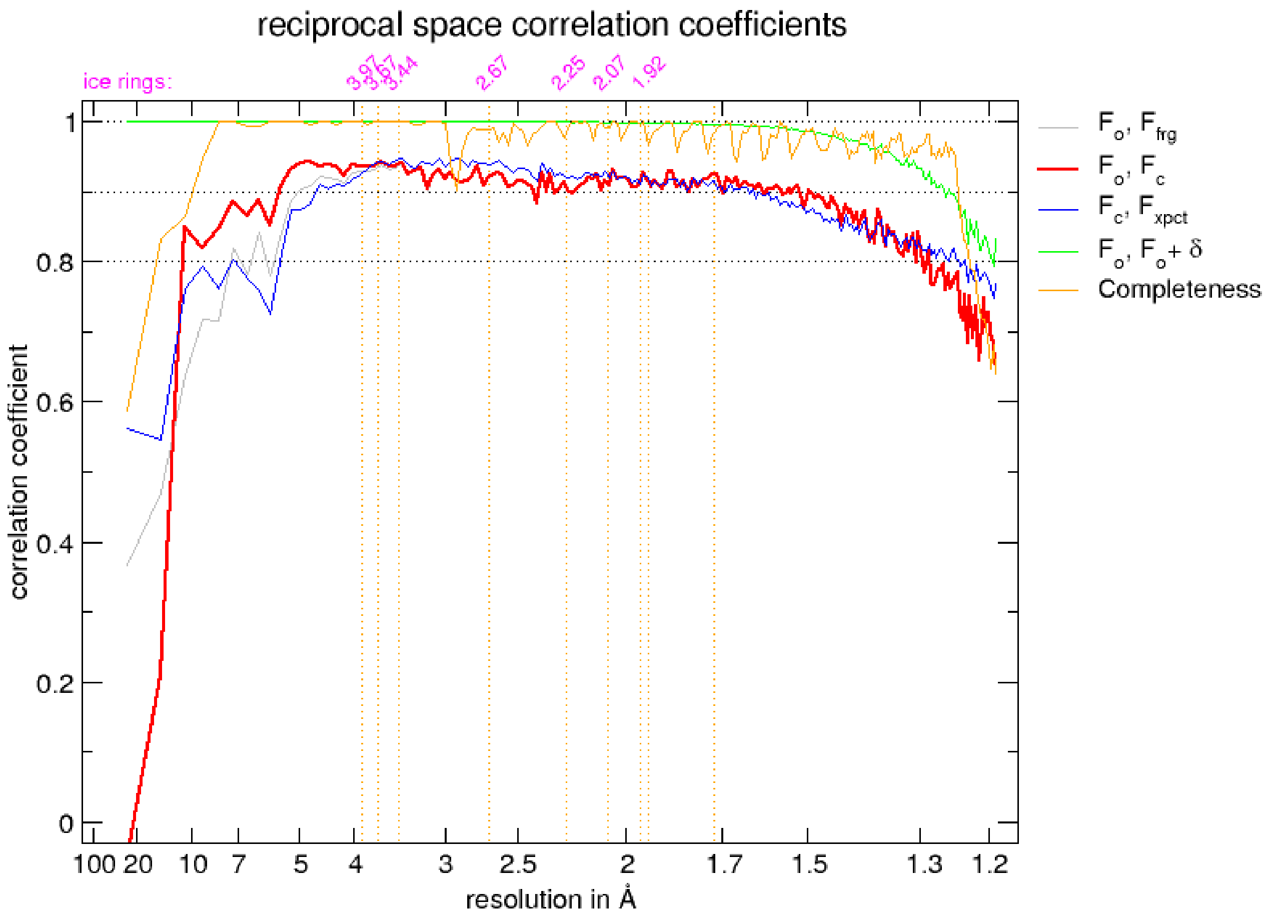
Such discrepancies indicate that the current estimate of the errors in the atomic model is inadequate, since Fo is not a typical member of the ensemble of structure factor amplitudes representing the uncertainties on the underlying atomic model and the observational errors affecting their measurement.

A frequent cause of discrepancy is a marked **anisotropy in the diffraction limits of the data**, because the error model on which the calculation of CC(Fc,Fxpct) is based is isotropic. Submitting the data file to the UCLA Diffraction Anisotropy Server will often correct the situation and yield a much closer similarity in the two CC plots.

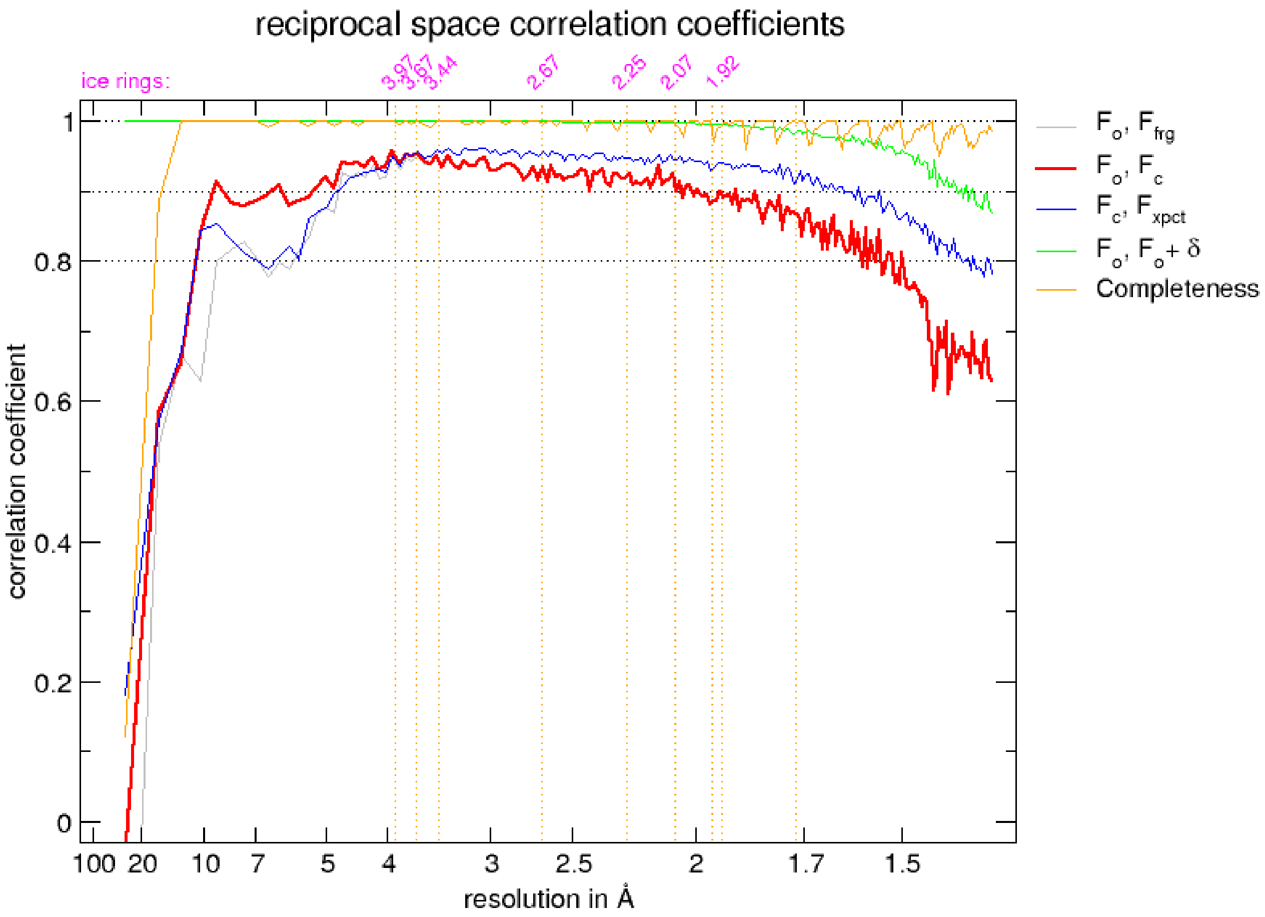
Example: 2xio before anisotropy correction



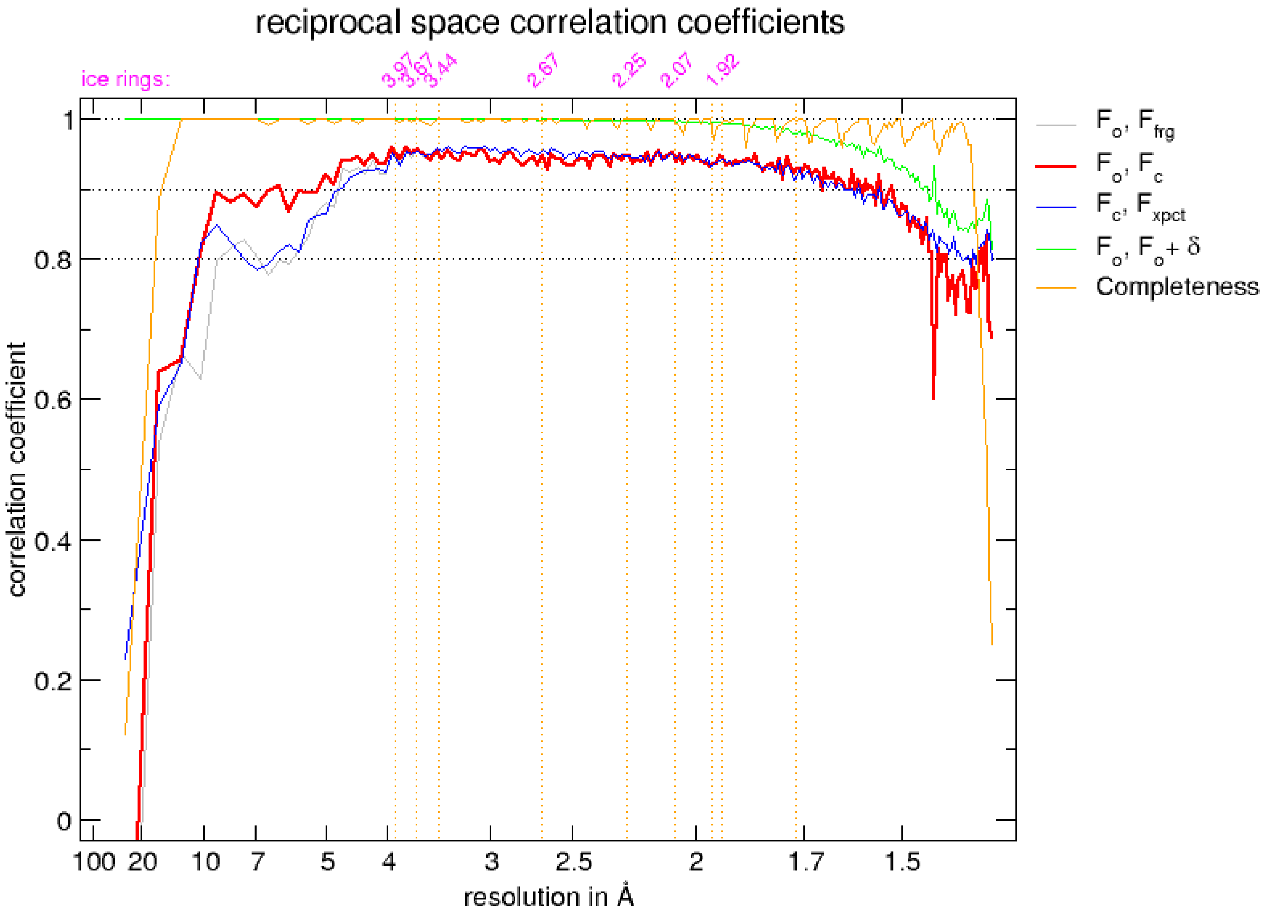
Example: 2xio after anisotropy correction.



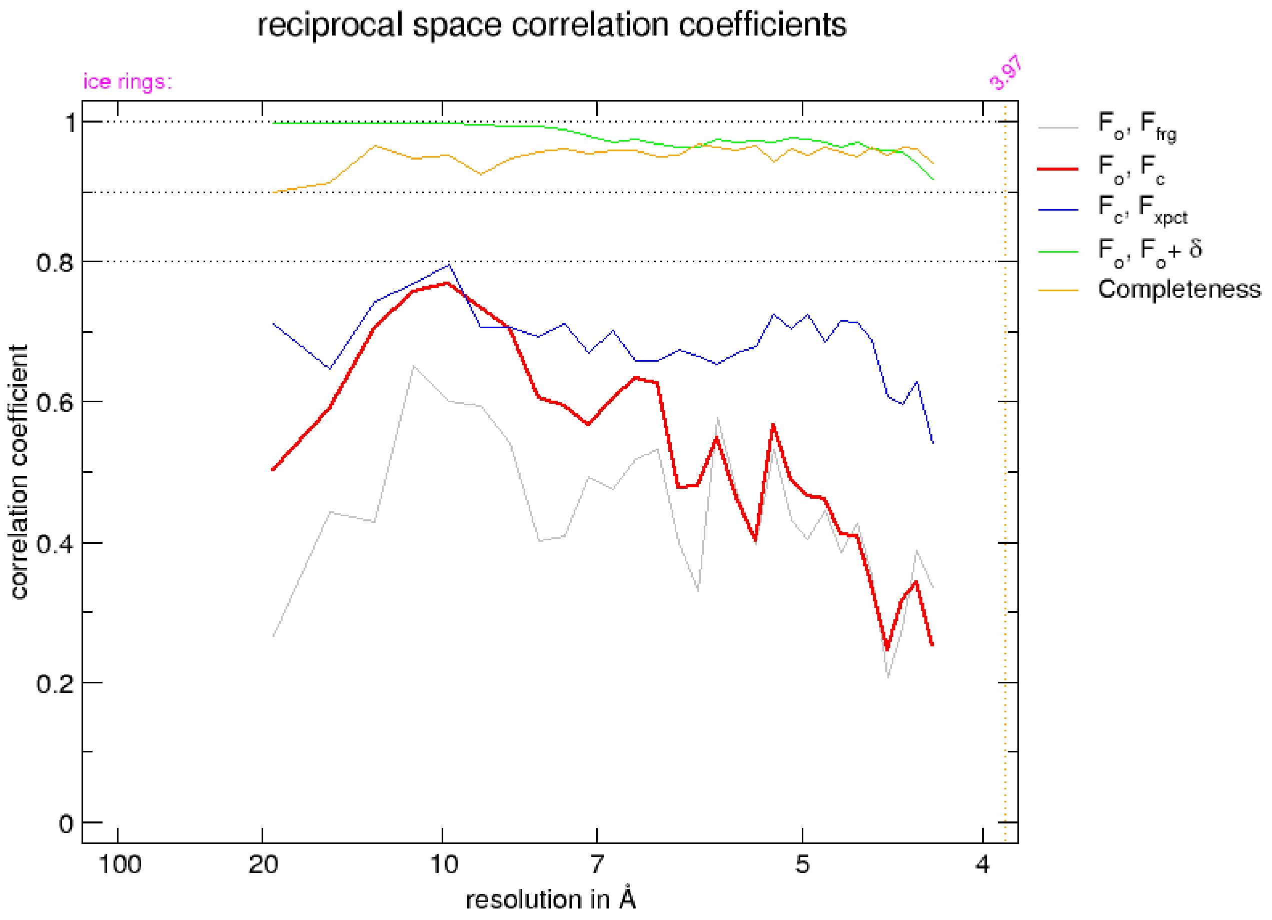
Example: 3ip5 before anisotropy correction.



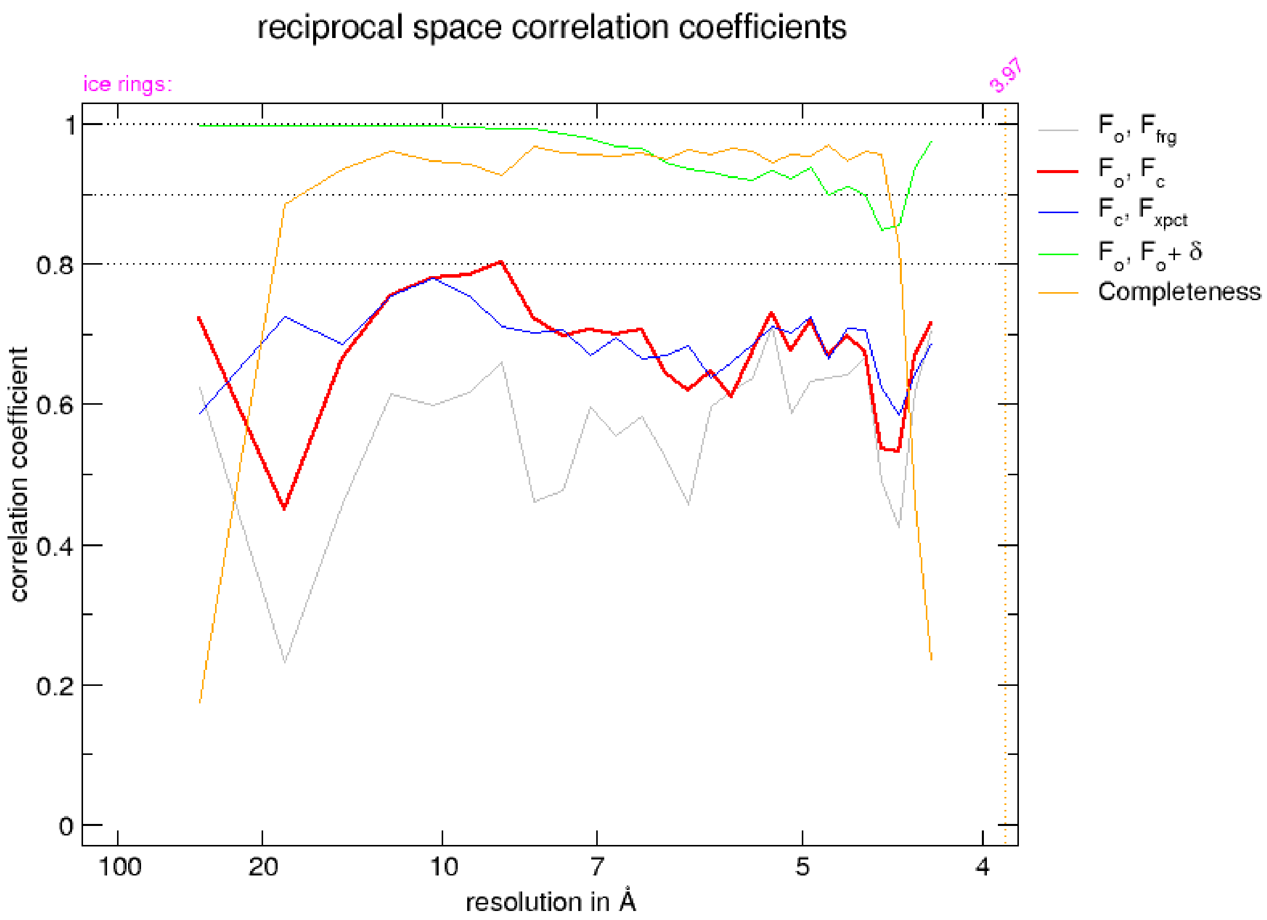
Example: 3ip5 after anisotropy correction.



Example: ChikE before anisotropy correction.



Example: ChikE after anisotropy correction.



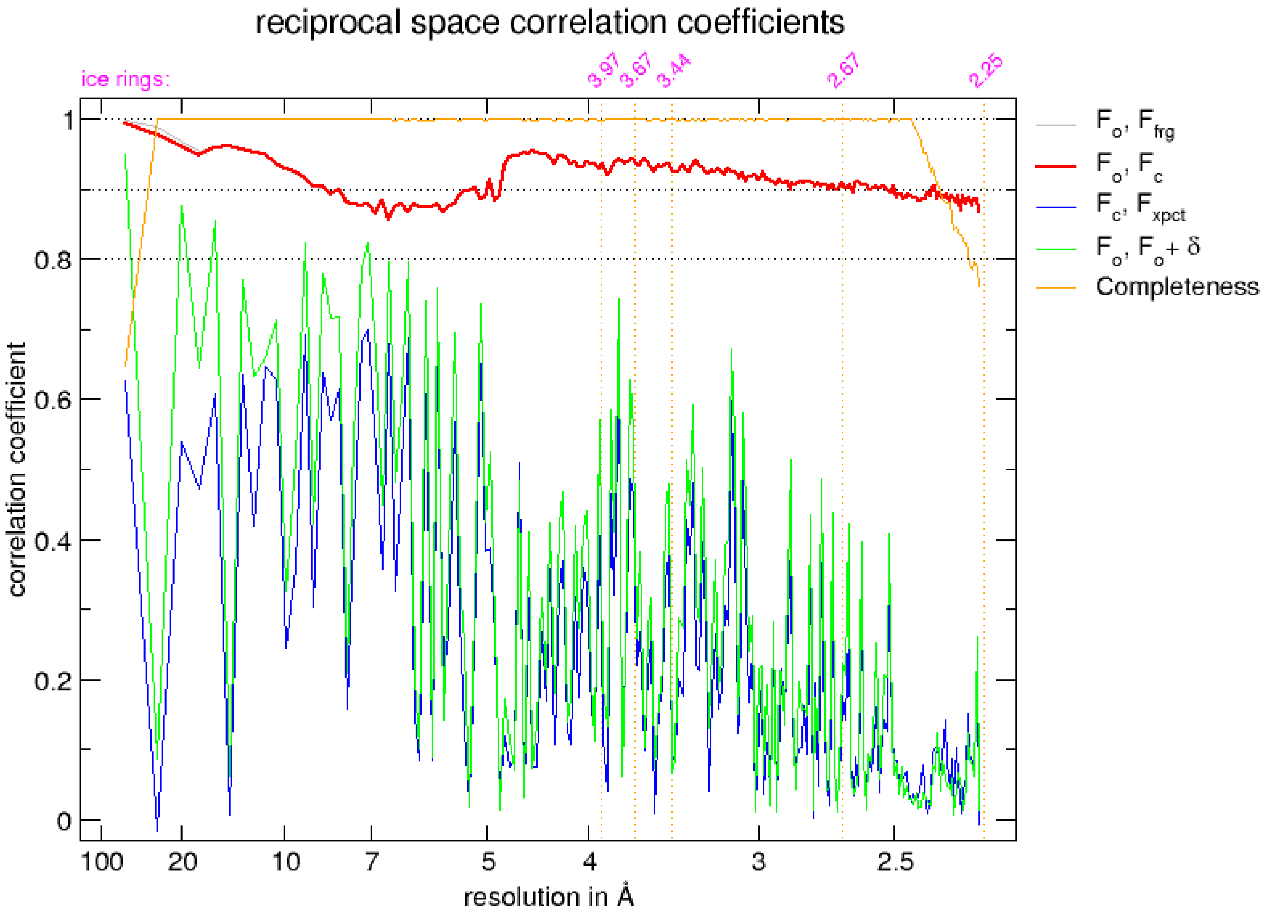
Diagnosis of erroneous information

If CC(Fc,Fxpct)>CC(Fo,Fc), then the error estimate is too optimistic, i.e. the model is worse than it is thought to be; and conversely in the opposite case.

The case of erroneous sigmas.

BUSTER takes the σ(Fo) seriously in the construction of the (log-) likelihood function it strives to maximise in the structure refinement process. If erroneous and excessive values have inadvertently been placed in the corresponding column of the input mtzfile, the CC(Fo,Fo+δ) and CC(Fc,Fxpct) curves will be disfigured. The CC(Fo,Fc) curve may look normal at the start of the refinement if the input model was refined with another program that doesn't use σ(Fo), but it will deteriorate as the model progressively “unrefines” as a consequence of the resulting nonsensical under-weighting of the X-ray data caused by that error.

Example: the “cause célèbre” faked data for 2hr0.


reciprocal space correlation coefficients
ice rings:
correlation coefficient
resolution in Å
F_o, F_{frg}
F_o, F_c
F_c, F_{xpct}
$F_o, F_o + \delta$
Completeness