GΦL
Global Phasing Limited

StarAniso: Automation and evaluation of tests for hemihedral twinning

Ian J. Tickle

Global Phasing Consortium Meeting

June 21-24, 2021

Summary (1)

- The aim initially was to evaluate hemihedral twinning tests for robustness using deposited data.

- 41 datasets were analysed where merohedral twinning had been reported and images were available.

- All datasets were re-processed with autoPROC, using the reported space group and cell.

- Given that the reported twin fraction(s) (a.k.a. 'volume fractions') may not be entirely trustworthy, all the test structures were re-refined with REFMAC full-structure twinned refinement ('TWIN' option). This also means we are comparing results like-for-like.

Summary (2)

- Existing implementations of hemihedral twinning tests require manual analysis of graph plots to obtain the twin fraction: this is quite subjective.
- I am also proposing here simplification of 2 existing twinning tests: now 8 twinning tests in total.
- It would clearly be generally useful to automate determination of the twin fraction, if only to avoid having to perform ~300 twinning tests manually with repeats for various input options!
- Automated twinning tests are now implemented in StarAniso; a single plot of the twin fraction *vs*. all test scores is now output.
- Serious inconsistencies between test results were found for a significant number of datasets.

GΦL
Global Phasing Limited

Hemihedral twinning tests

1. Britton (1972)	non-distributional
2. Murray-Rust (1973)	“
3. Fisher & Sweet, a.k.a. Britton (1980)	“
4. Rees F(Z) (1982)	distributional
5. Rees F($\|Z_1–Z_2\|$)$_{\mathrm{related}}$ (1982)	“
6. Yeates F($\|I_1–I_2\|/ (I_1+I_2)$)$_{\mathrm{related}}$ (1988)	“
7. Padilla & Yeates F($\|I_1–I_2\|/ (I_1+I_2)$)$_{\mathrm{unrelated}}$ (2003)	“
8. P($\|Z_1–Z_2\|$)$_{\mathrm{unrelated}}$ (unpublished)	“

Britton-related tests (1972, 1973 & 1980)

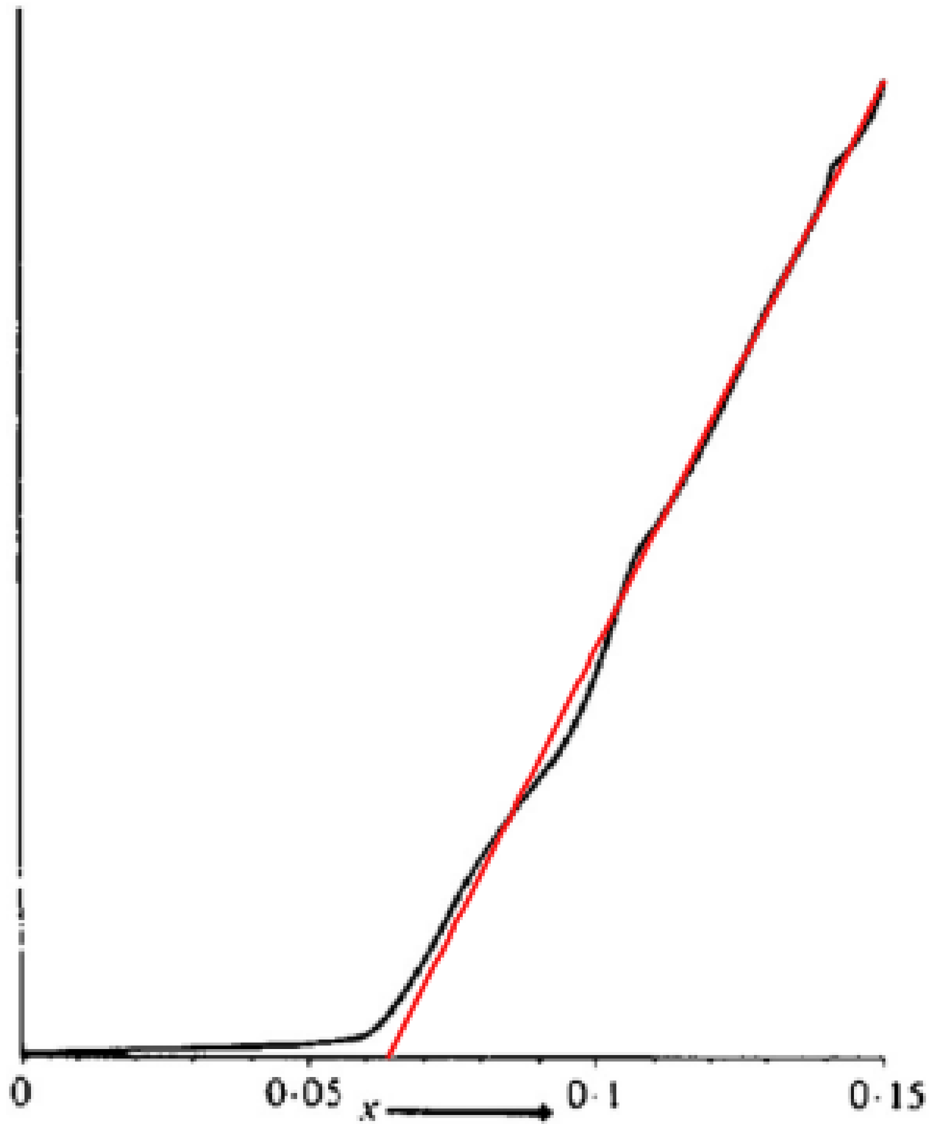
- The non-distributional Britton (1972), Murray-Rust (1973) and Fisher & Sweet (1980) tests all rely on the fact that if the data are detwinned using increasing trial values of the twin fraction, for some twin fraction additional significant negative detwinned intensities will be generated. They differ only in the method of analysis of the results.

- These tests (and others that depend on pairs of reflections related by the twin operator) obviously require that the twin operator be known, or at least that all possible twin operators are tested.

- Also, in the case of perfect twinning it is essential to process the data in the correct sub-group, so that the twin-related intensities are not averaged!

- The Fisher & Sweet (1980) test is usually referred to in the literature as the 'Britton plot' and although based on it, it is not the same as Britton's original version of the test, so henceforth it is referred to as the 'Fisher & Sweet plot'. However it is a recognized as a significant improvement on the original version, and also on the Murray-Rust test. The latter two tests will therefore not be discussed further.

Fisher & Sweet (1980) test
(a.k.a. ‘Britton plot’)

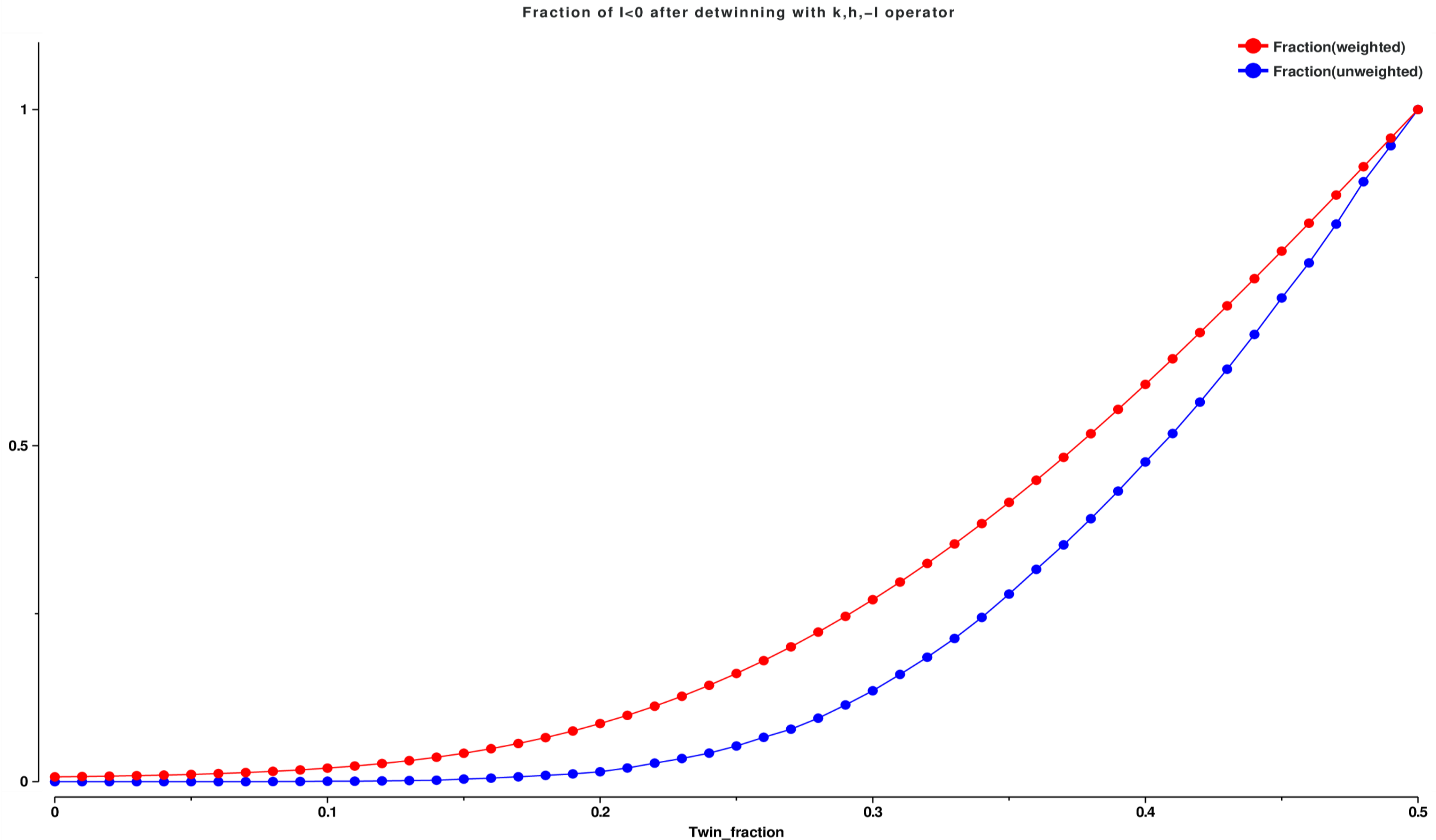


GΦL
Global Phasing Limited

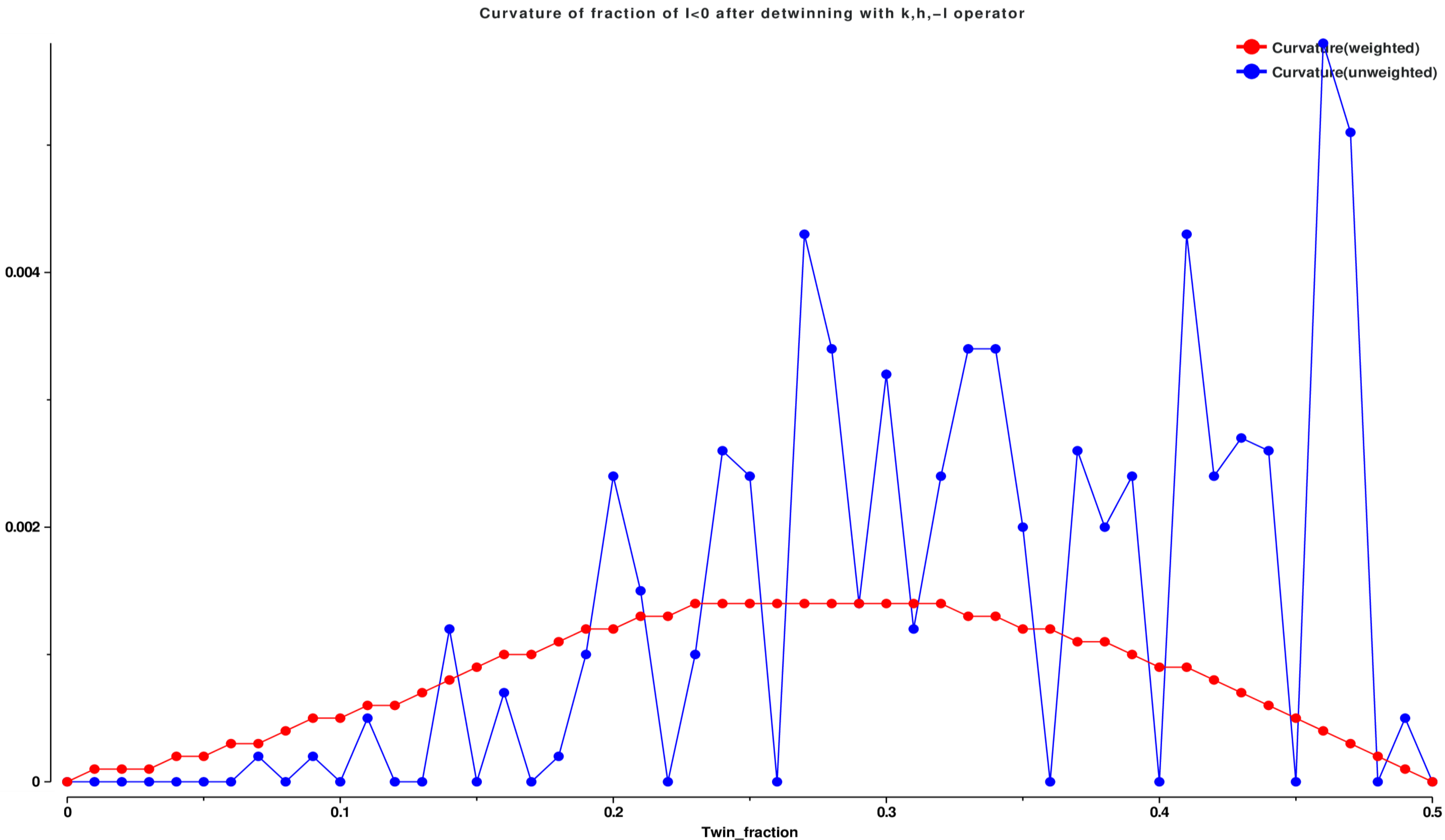
Fisher & Sweet ('Britton') plot

- The point where a straight-line fit of the negative intensity count intersects the *x* axis gives the twin fraction.
- This is fine for low values of the twin fraction but for values near 0.5 (perfect twin) there are insufficient points to perform the fitting and the plots are more curved.
- A better method that is also more suited for automation is to locate the maximum curvature of the negative intensity count.
- The intensities of weak reflections are much more likely to go negative after detwinning due to random error so it is necessary to exclude these from the count.
- However it's better to include them but apply a smooth weighting function that reduces the contribution of the weak reflections to the negative intensity count.

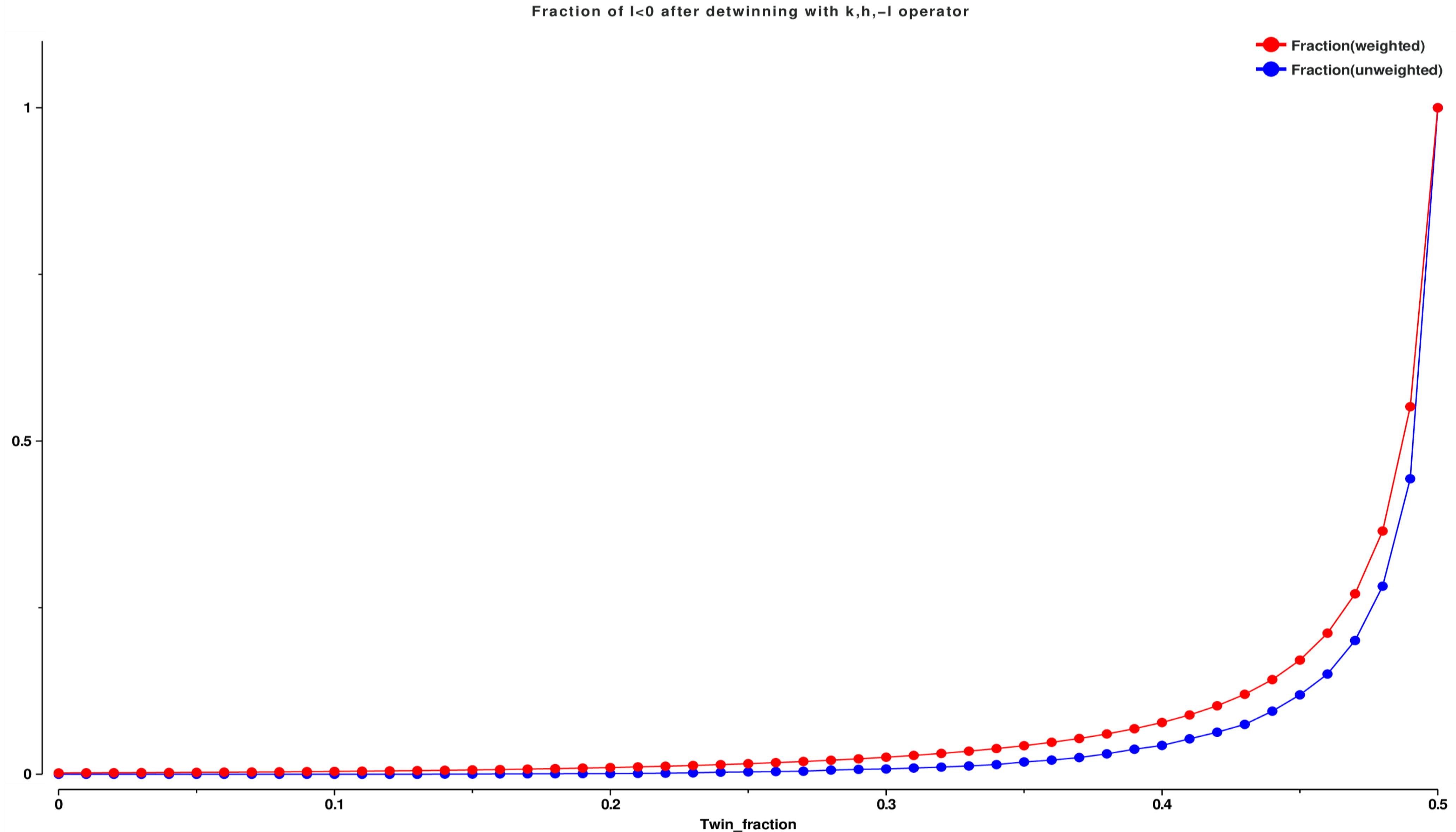
3RRO: Original Fisher & Sweet (Britton) plot



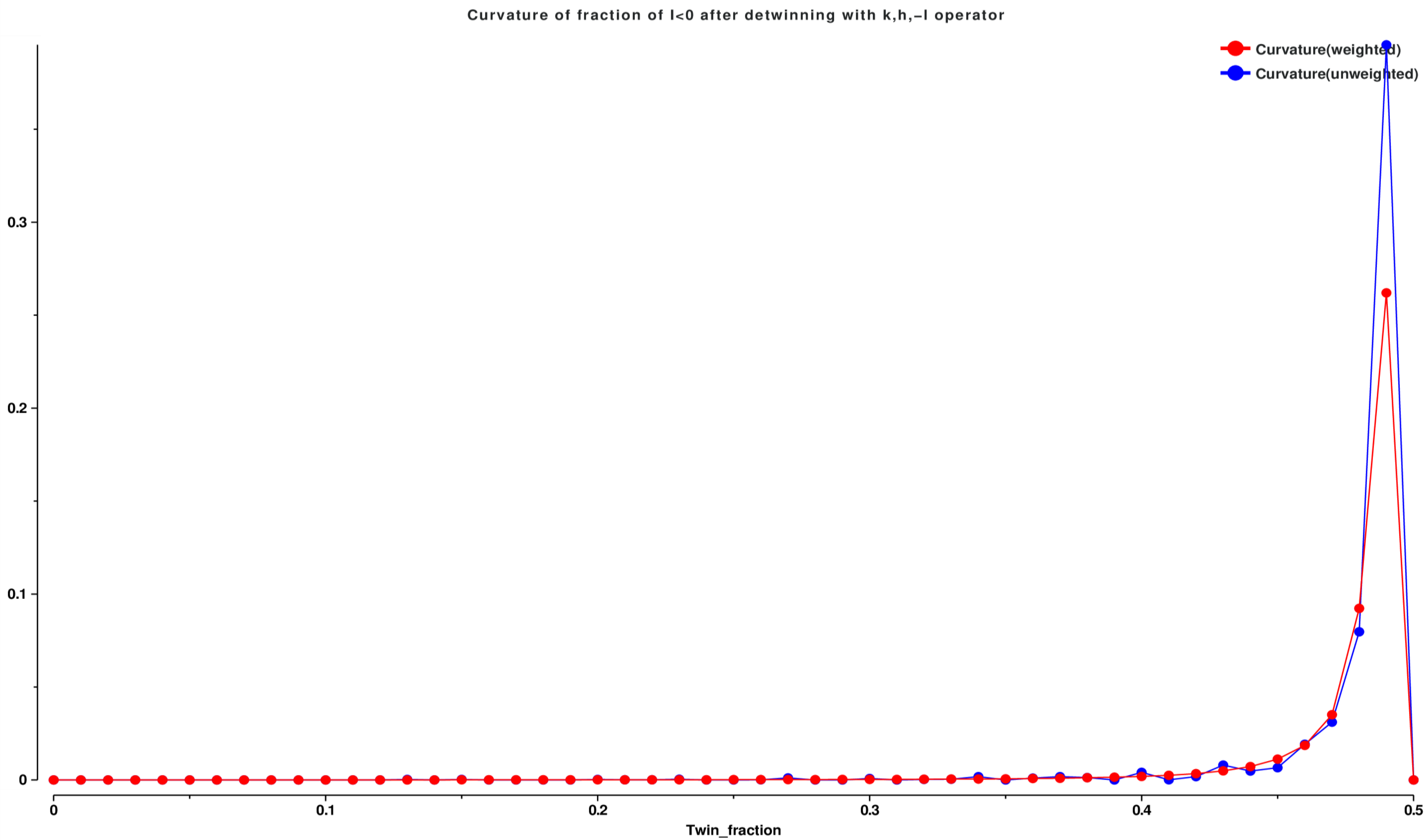
3RRO: Curvature of F & S test score



4JD1: Original Fisher & Sweet (Britton) plot



4JD1: Curvature of F & S test score



Z distribution tests (1)

- A time-honoured test for twinning is the cumulative distribution function (CDF) F(Z) plot (Rees, 1980), where Z is the normalized intensity: $Z = I \,/\, \mathrm{E}[I]$.
- The observed plot of F(Z) is compared with theoretical plots for untwinned centric and acentric, and perfectly twinned centric and acentric reflections (the distribution of twinned centric is the same as that for untwinned acentric).
- Knowledge of the twin operator (or even that the data have been processed in the correct sub-group) is not required for this test.
- A drawback is that an NCS axis parallel to the twinning axis tends to cancel the effect of twinning on the distribution functions; for this reason the F(Z) test is not considered reliable in the presence of NCS.
- In fact pNCS itself will increase the chance of twinning, since it lowers the free-energy penalty of the lattice defects that give rise to twinning by exchanging crystallographic lattice contacts with non-crystallographic ones.

Z distribution tests (2)

- For this, instead of using the cumulative distribution, I compute the Kullback-Leibler divergence D_{KL} of the theoretical probability density function (PDF: the gradient of the CDF) from the observed one as a function of the twin fraction.

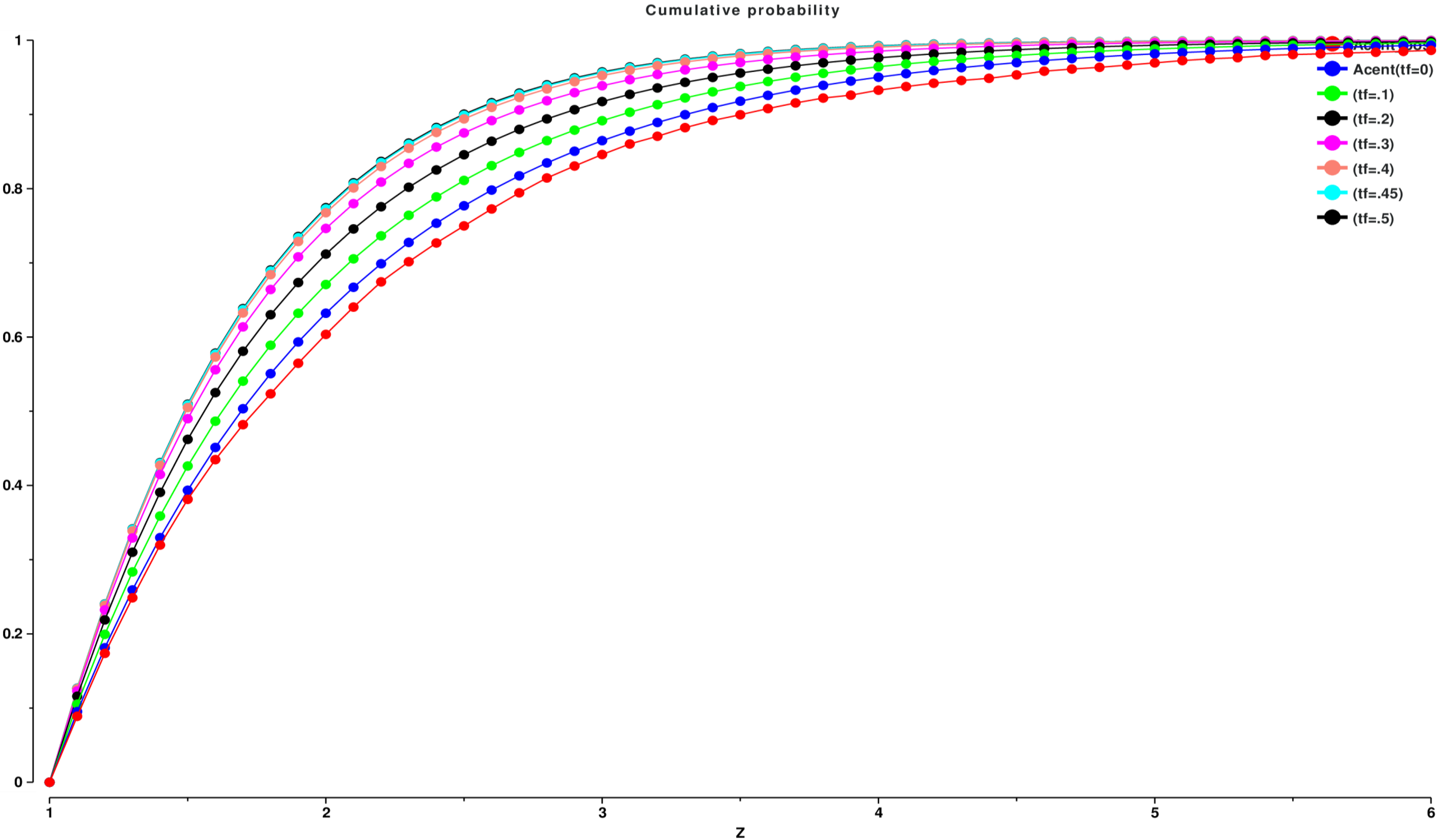
- The true twin fraction is then located at the minimum of D_{KL}. This is easily automated compared with the classical manual graphical method.

- The use of the individual moments of the distribution as tests for twinning is also not ideal because it is the complete set of moments that specify the distribution, so that any individual moment is an incomplete representation of the distribution.

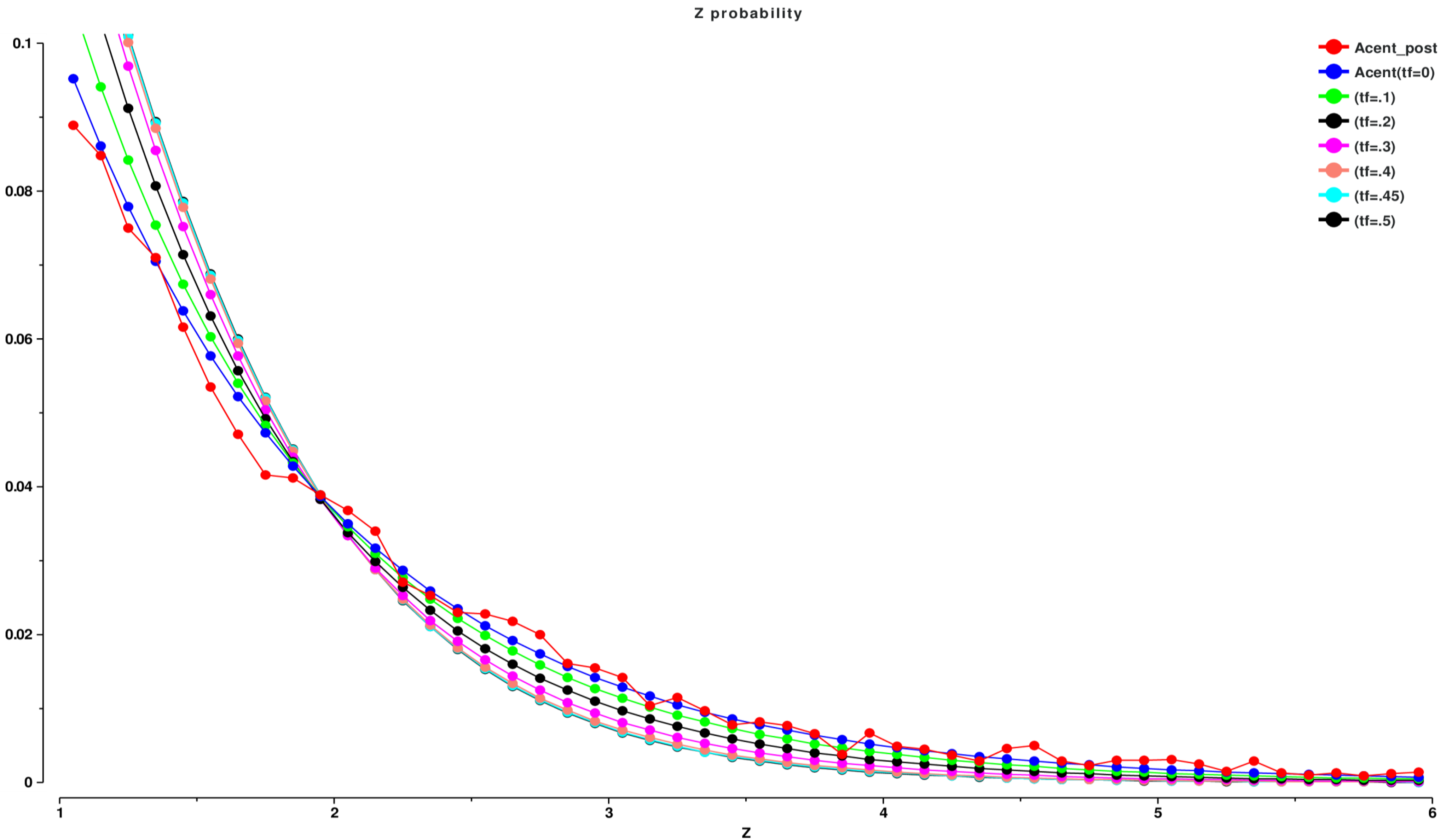
- Also there's no need for separate tests of the acentric and centric reflections: one simply uses the appropriate form of the PDF in the calculation of D_{KL} , obtaining a single estimate of the twin fraction directly.

GΦL
Global Phasing Limited

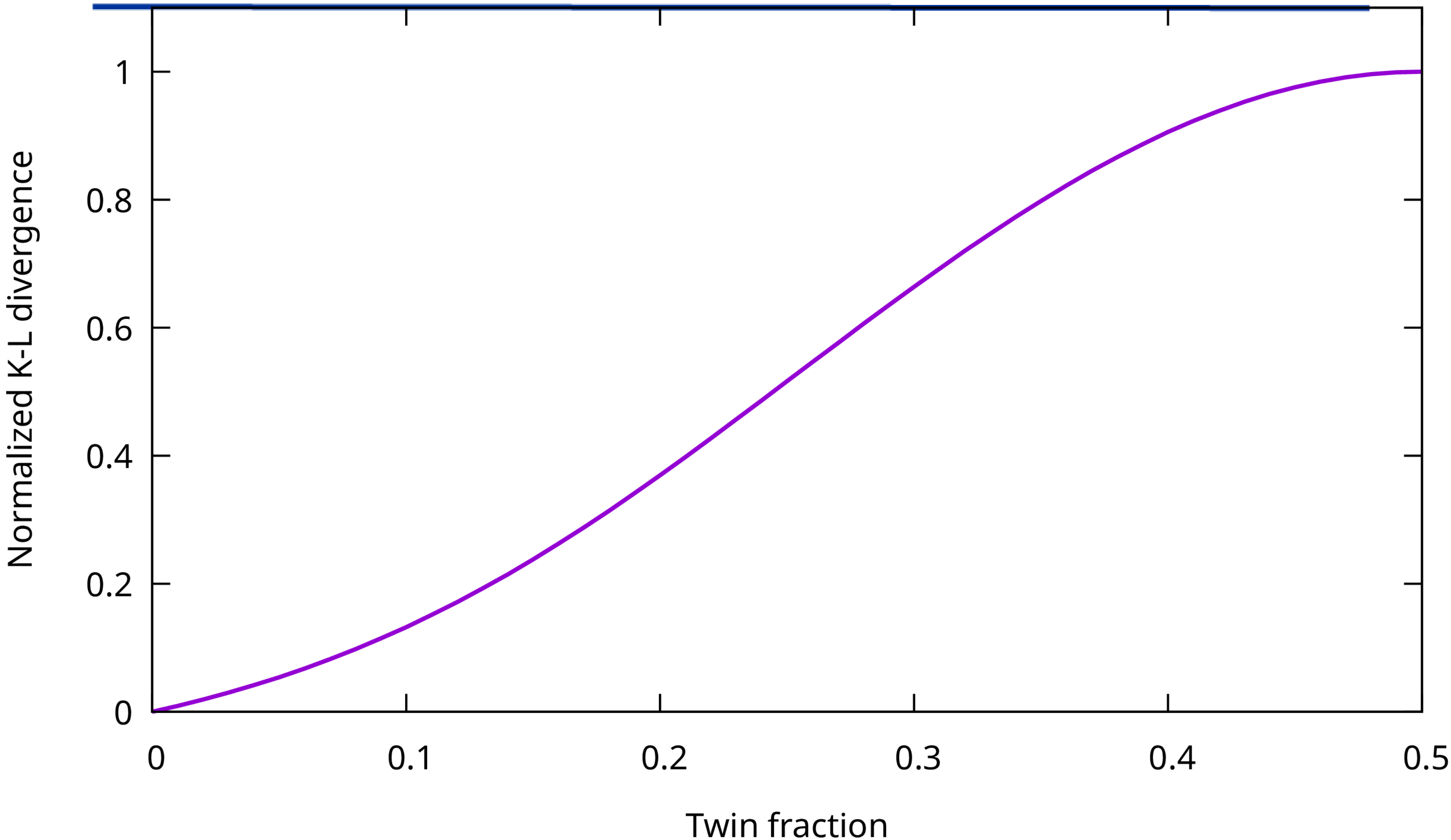
4JD1: Acentric CDF F(*Z*) *vs.* *Z*



4JD1: Acentric PDF P(*Z*) *vs*. *Z*



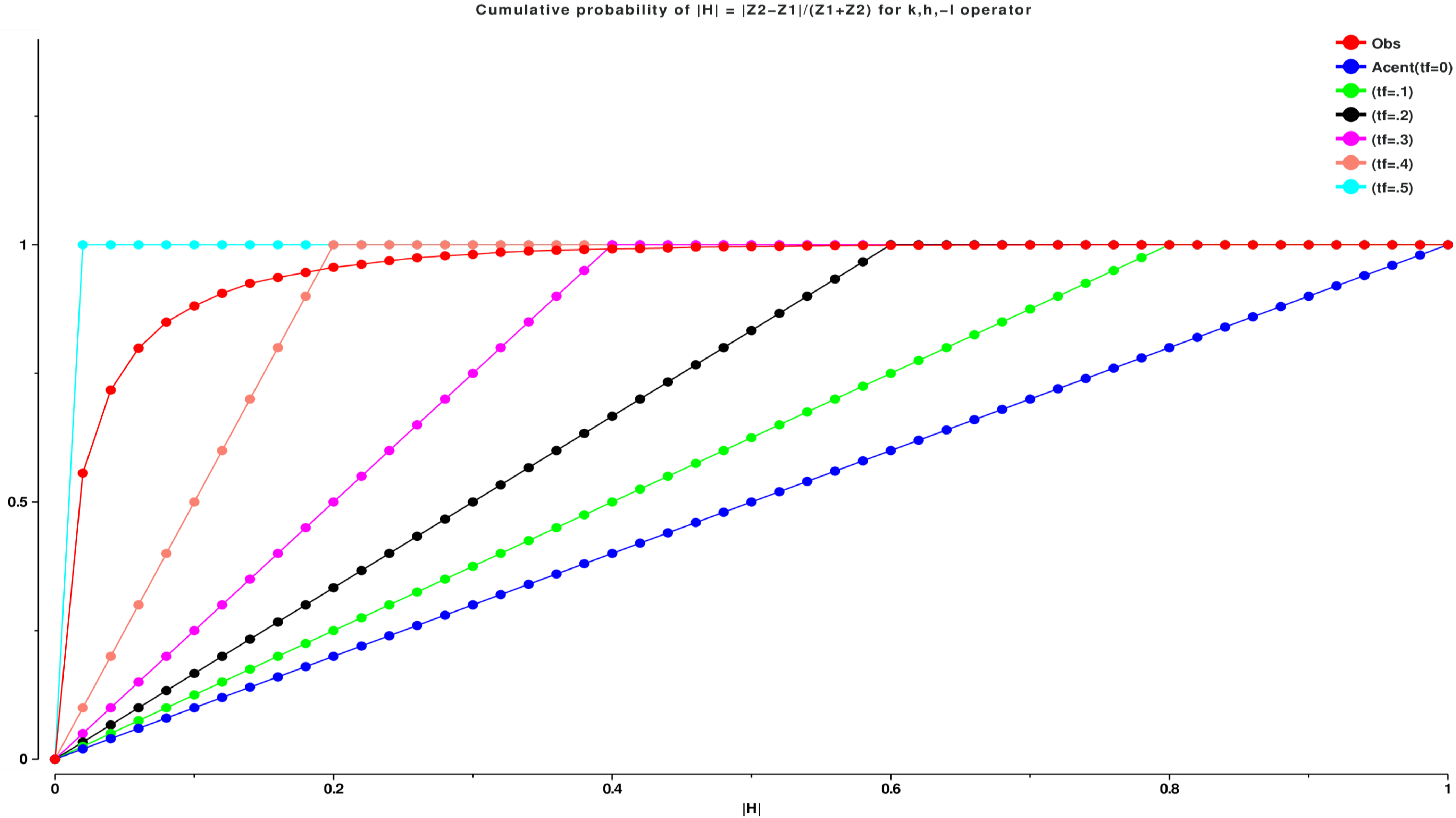
4JD1: K-L divergence of P(*Z*)



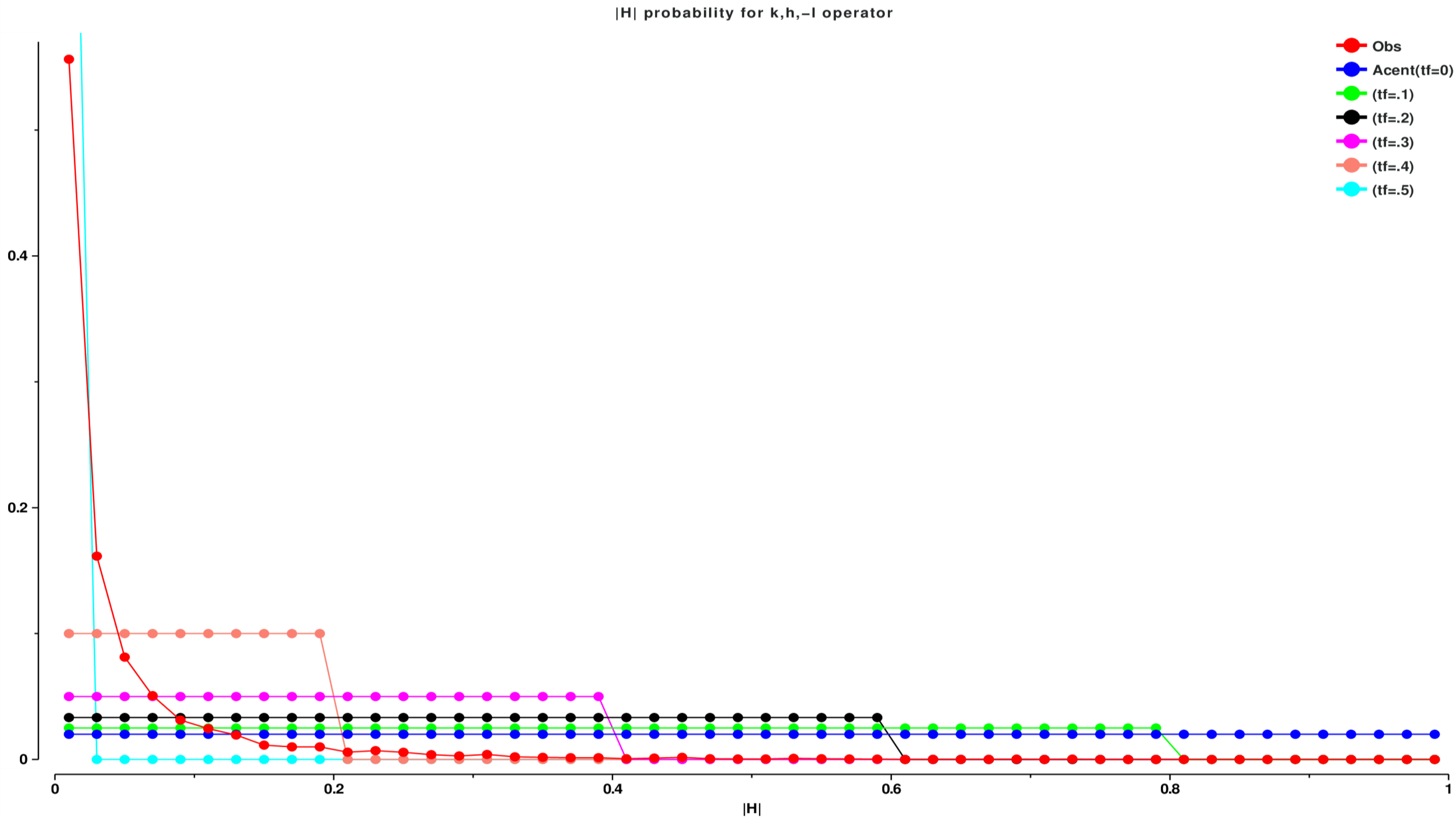
Twin-operator related difference tests (1)

- Yeates (1988) proposed a test based on the CDF F(|H|) of the |H| function.
- $H = (I_1 - I_2) / (I_1 + I_2)$ and the Is are the Bayesian posterior estimates of the pairs of twin-operator related intensities (so can never be ≤ 0).
- Otherwise one has to deal with the possibility of a zero or negative denominator in H.
- Obviously, knowledge of the twin operator is required for this test (and again the processing must have been in the correct space group!).
- This functional form was chosen in order not to have to deal with normalization of the intensities, the assumption being that the expectations of the numerator and denominator will always cancel.
- In the isotropic case and in the case of twinning by merohedry, I_1 and I_2 have equal expectations so the above assumption is valid.

4JD1: Yeates H test F(|H|) $vs.$ |H|



4JD1: P(|H|) vs. |H|



|H| probability for k,h,−l operator
Obs
Acent(tf=0)
(tf=.1)
(tf=.2)
(tf=.3)
(tf=.4)
(tf=.5)
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
|H|

GΦL
Global Phasing Limited

Twin-operator related difference tests (2)

- In the case of twinning by pseudo-merohedry (where the twin symmetry is not an element of the point group), the expectations in the H test will differ in the presence of anisotropy, so normalization of the intensities is necessary.

- The H test then no longer has an advantage over using the PDF of the absolute difference $|Z_1 - Z_2|$ for twin-related pairs.

- Use of the Debye scattering formula for the spherical intensity distribution solves the normalization problem.

- It would seem more sensible to fix the effects of anisotropy at source (*i.e.* StarAniso), rather than try to compensate for it after the fact.

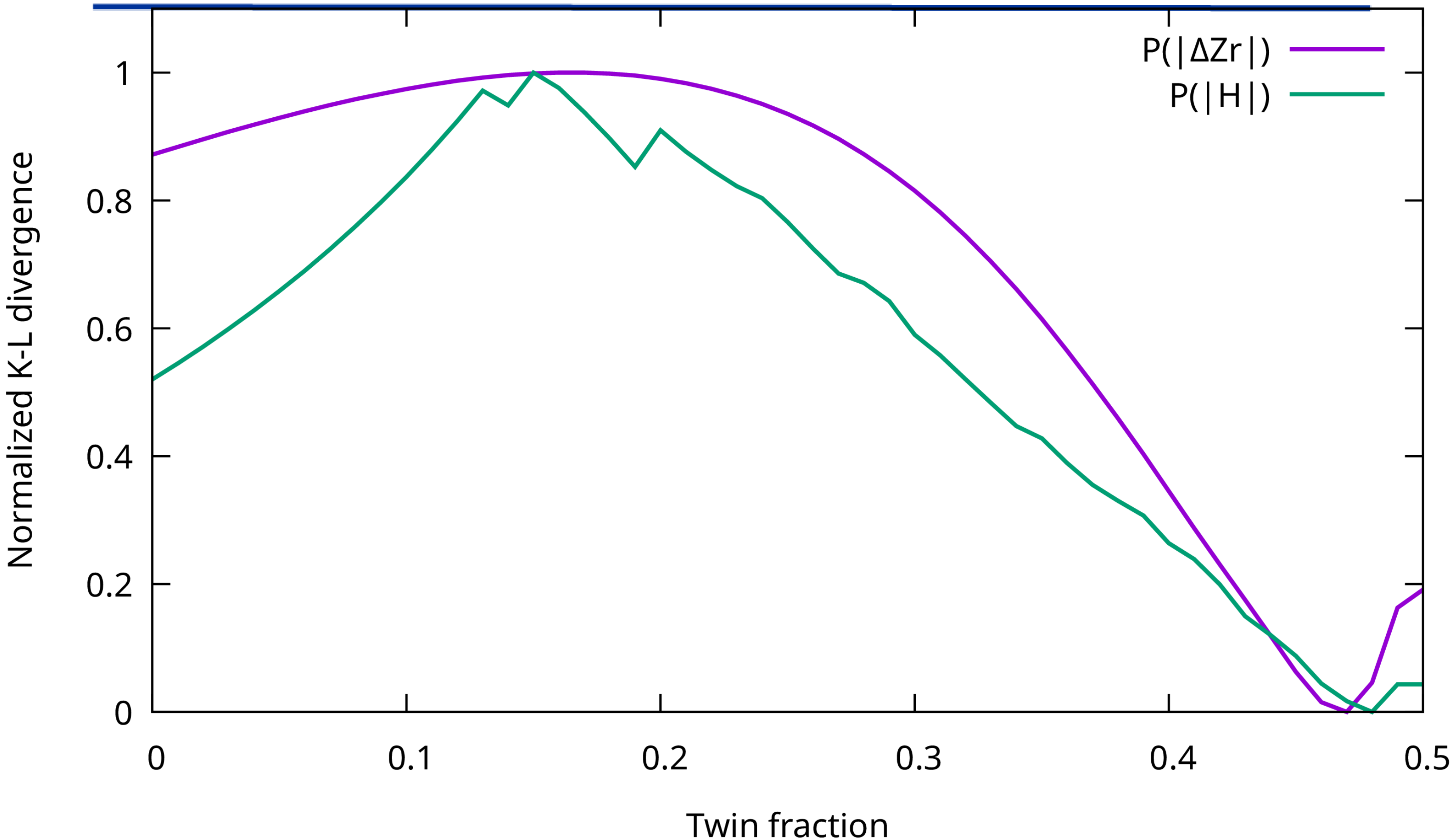
- For automated determination of the twin fractions we again use the minimum K-L divergence method.

Global Phasing Limited

Twin-operator related difference tests (3)

- The preceding assumes that I_1 and I_2 (or Z_1 and Z_2) are uncorrelated.
- This will not hold if the twin operator is also a 2-fold NCS operator (the very common pNCS case), so these tests are unreliable in this situation.
- The correlation of Z_1 and Z_2 will depend on the exactness of the NCS, so these tests are expected to give an incorrect result in this situation.

4JD1: K-L divergences of P(|ΔZ_r|) and P(|H|)



Non-twin-operator related difference tests (1)

- Analogous to the $|H|$ function, Padilla & Yeates (2003) proposed a test based on the CDF $F(|L|)$.

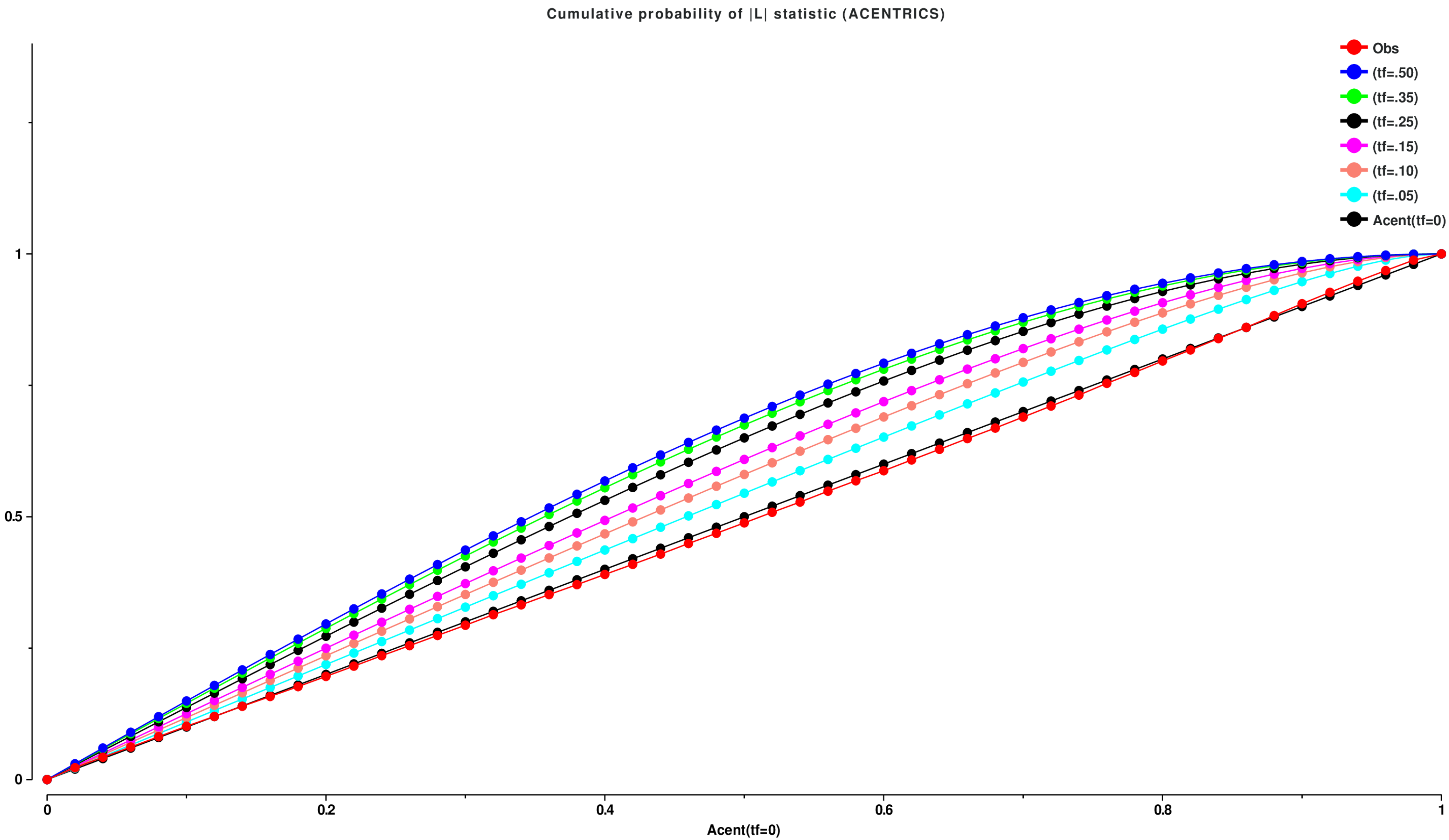
- $L = (I_1 - I_2) / (I_1 + I_2)$ and the *I*s are the Bayesian posterior estimates of pairs of neighbouring intensities that are not related by the twin operator (so knowledge of the twin operator and processing in the sub-group are not required).

- The indices of I_1 and I_2 typically differ by 0 or 2 in all three directions in an attempt to avoid interference by translational NCS (this assumes that the tNCS vector has components exactly 0 or 0.5).

- The drawback of the L test as originally proposed by Padilla & Yeates is that it is insensitive for twin fractions above ~0.25, so they suggested that it be used only to detect perfect twinning and use an alternative method to estimate the twin fraction.

- The problem with this approach is that it is being used precisely because the other methods may not be accurate.

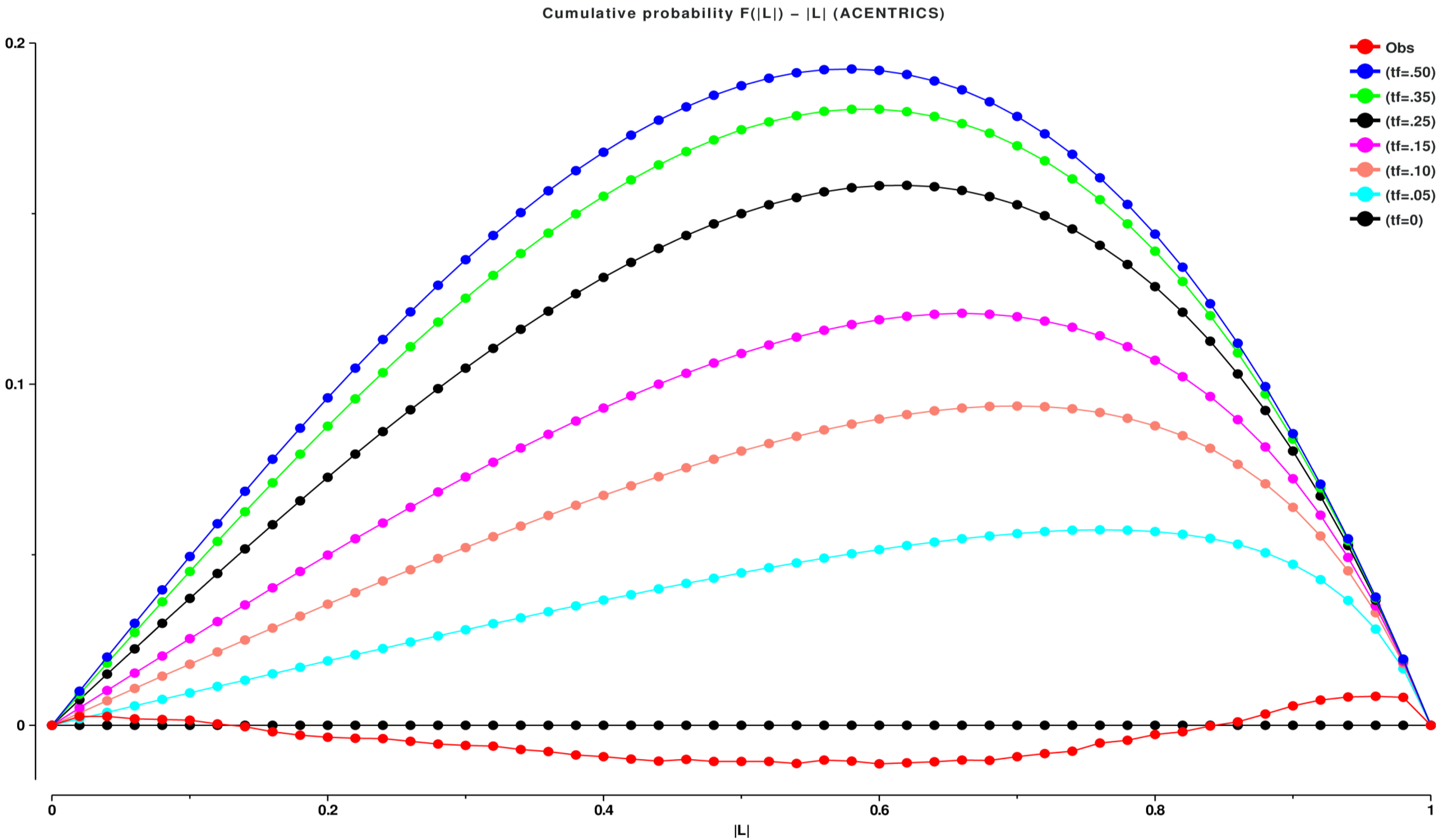
4JD1: Padilla & Yeates L test F(|*L*|) vs. (*L*)



Non-twin-operator related difference tests (2)

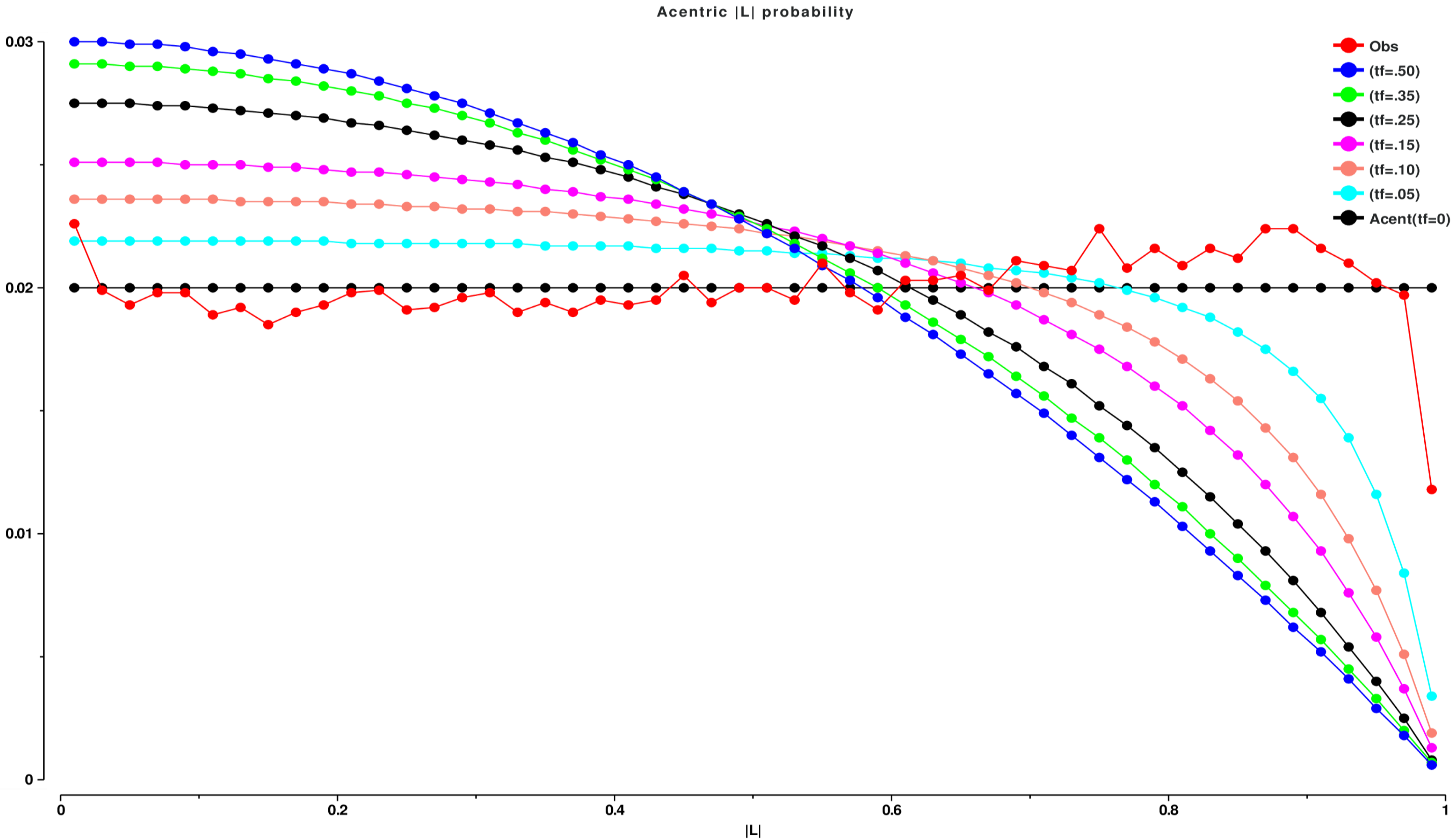
- To get around the insensitivity of the L test, we can use instead the function $F(|L|) - |L|$, *i.e.* rotate the baseline clockwise through 45°.
- We can reduce the effect of tNCS on the L test by using tNCS-corrected expectations in the intensity normalization.
- As for the twin-related case, normalization is not an issue and there's no longer an advantage of the L test over the absolute difference ($|Z_1 - Z_2|$) test.
- For determination of the twin fractions we can again locate the minimum of the K-L divergence of the theoretical from the observed PDF.

4JD1: F(|L|) - |L| vs. (|L|)

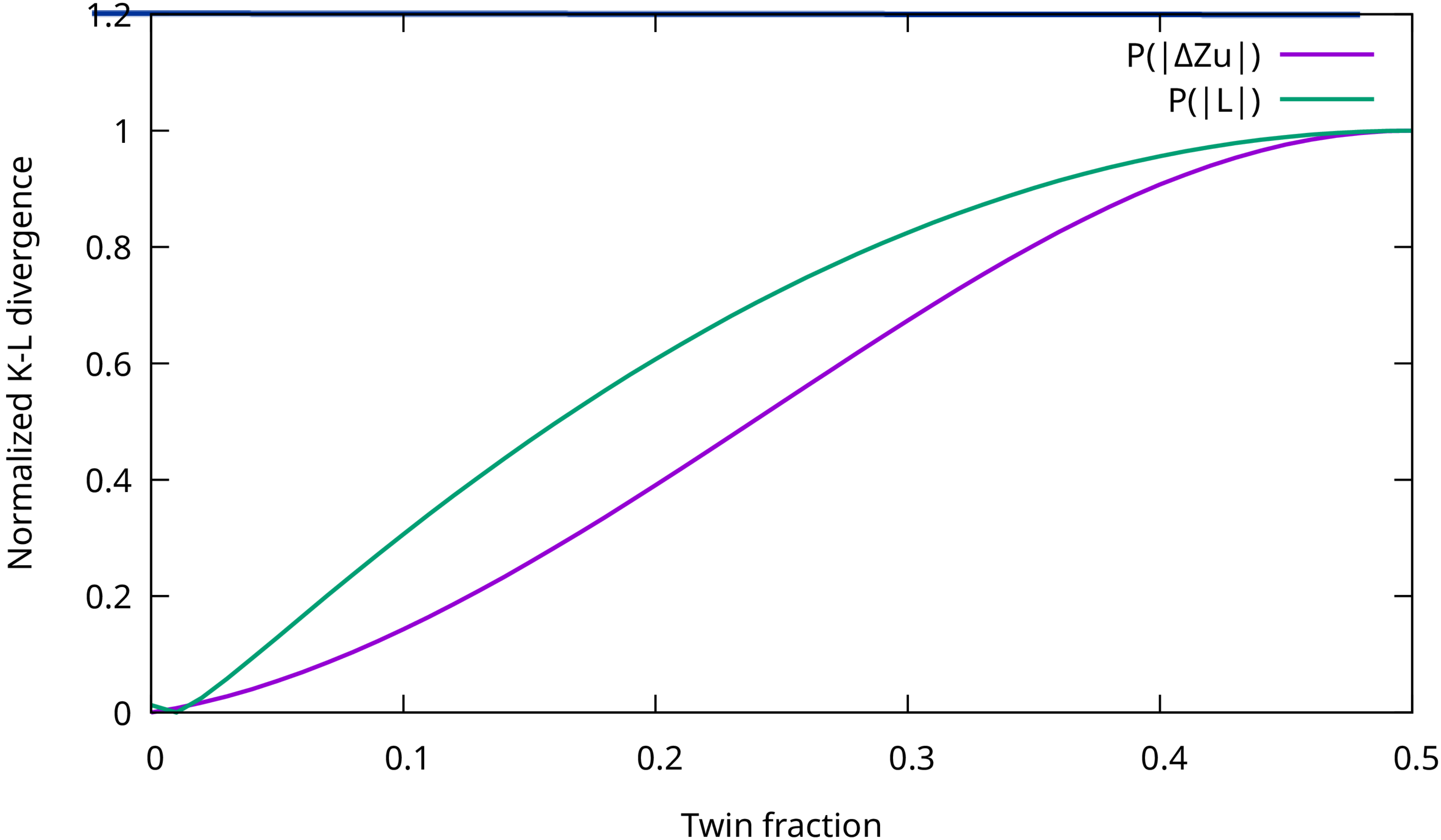

Cumulative probability F(|L|) – |L| (ACENTRICS)
Obs
(tf=.50)
(tf=.35)
(tf=.25)
(tf=.15)
(tf=.10)
(tf=.05)
(tf=0)
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
|L|

GΦL
Global Phasing Limited

4JD1: P(|*L*|) vs. (|*L*|)

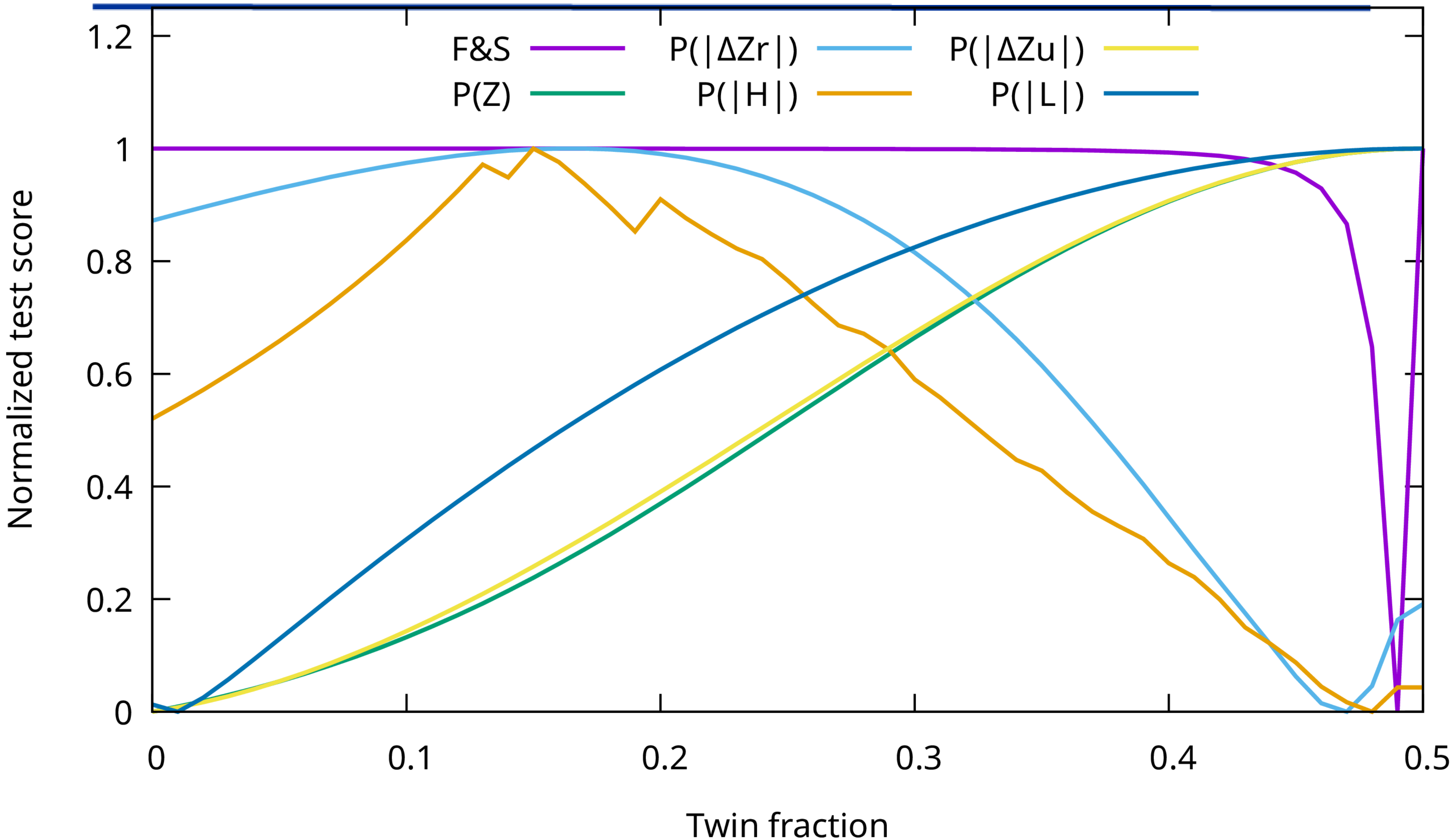


4JD1: K-L divergences of $P(|\Delta Z_u|)$ and $P(|L|)$



GΦL
Global Phasing Limited

4JD1: Summary of automated twinning test scores



GΦL
Global Phasing Limited

Summary of results (1)

- Among the 41 datasets used in the twinning tests, a large number exhibited significant inconsistencies between the results.
- All but 4 datasets have NCS, and a few also translational NCS.
- A quarter showed serious discrepancies, with some tests indicating zero twin fraction while others indicated perfect twinning (t.f. = 0.5) for the identical data!
- All of these serious cases had parallel NCS, *i.e.* a NCS 2-fold rotation axis parallel to the twinning axis indicated in the PDB entry.
- There are too few cases of translational NCS to conclude whether this has a significant effect.
- This rather throws doubt on the usefulness of some of the tests!

Global Phasing Limited

Summary of results (2)

- Keeping in mind the rather limited sample and the fact that parallel NCS is present in the majority of datasets, the twinning tests appear to fall into two clear groups:
    - The L test (Pointless), the P(*Z*) test, the unrelated difference test and the StarAniso optimization result.
    - The Britton plot, the H test, the related difference test and the twinned refinement result.
- The twin fractions are reasonably consistent within each group but in some cases differ hugely between the groups.
- The greatest inconsistencies occur where group #1 gives a zero twin fraction, and group #2 indicates perfect twinning.

GΦL
Global Phasing Limited

Inconsistent results of twinning tests (1)

ID	PG	NCS	tNCS	pNCS		P-L	P(Z)	$P(\lvert\Delta Z_u\rvert)$	SA-ML		F & S	$P(\lvert\Delta Z_r\rvert)$	Refmac	PDB
4EJO	4	2	F	T		0.10	0.07	0.09	0.05		0.47	0.38	0.39	0.41
4IR0	4	2	F	T		0.01	0.00	0.00	0.00		0.49	0.46	0.49	0.50
4JD1	4	2	F	T		0.01	0.00	0.00	0.00		0.49	0.47	0.50	0.50
4NP6	4	4	F	T		0.02	0.00	0.00	0.00		0.48	0.46	0.49	0.50
5TK4	4	2	F	T		0.02	0.00	0.00	0.00		0.49	0.46	0.49	0.49
5UOG	3	4	T	T		0.03	0.00	0.00	0.00		0.49	0.46	0.50	0.49

PG	Point group.
NCS	Nmol/a.u.
tNCS	T = translational NCS present.
pNCS	T = parallel NCS present.
P-L	Twin fraction from Pointless L test.
SA-ML	Twin fraction from StarAniso ML optimization.
Refmac	Twin fraction from Refmac full-structure twinned re-refinement.
PDB	Twin fraction from deposited data.
Others	Twin fractions from automated twinning tests in StarAniso (see main text).

GΦL
Global Phasing Limited

Inconsistent results of twinning tests (2)

ID	PG	NCS	tNCS	pNCS		P-L	P(Z)	$P(\lvert\Delta Z_u\rvert)$	SA-ML		F & S	$P(\lvert\Delta Z_r\rvert)$	Refmac	PDB
3RRI	4	2	F	F		0.50	0.50	0.50	0.50		0.26	0.34	0.35	0.35
3U09	6	2	F	T		0.50	0.50	0.50	0.50		0.31	0.31	0.16	0.17
4GZE	3	6	T	T		0.41	0.50	0.50	0.50		0.13	0.25	0.26	0.28
4JOQ	321	2	F	F		0.24	0.34	0.34	0.29		0.43	0.42	0.43	0.43
4R8O	321	2	F	T		0.14	0.27	0.28	0.14		0.48	0.44	0.50	0.50
4WJZ	6	4	T	T		0.50	0.39	0.40	0.35		0.43	0.32	0.24	0.24
5DXD	4	2	F	T		0.50	0.35	0.35	0.35		0.45	0.43	0.46	0.47
5HW2	4	4	F	T		0.49	0.30	0.32	0.32		0.48	0.45	0.49	0.49
6PI7	6	6	F	T		0.50	0.50	0.50	0.50		0.41	0.39	0.43	0.43

PG	Point group.
NCS	Nmol/a.u.
tNCS	T = translational NCS present.
pNCS	T = parallel NCS present.
P-L	Twin fraction from Pointless L test.
SA-ML	Twin fraction from StarAniso ML optimization.
Refmac	Twin fraction from Refmac full-structure twinned re-refinement.
PDB	Twin fraction from deposited data.
Others	Twin fractions from automated twinning tests in StarAniso (see main text).

Global Phasing Limited

Conclusions

- It would appear that parallel NCS, which occurs in all but two test cases, not unsurprisingly invalidates the tests that depend on twin-related intensities (group #2), and rather surprisingly the twinned refinement (and therefore the deposited twin fraction).

- Among the 41 test datasets that were successfully processed, the majority had twin fractions from the unrelated difference test that were either:

  - too small (< 0.25) to give good discrimination in the tests, and/or
  - hugely overestimated in the second group due to parallel NCS.

- This left only 9 datasets with twin fractions that were large enough to be useful for evaluation of the tests.

- Instead of performing the tests only on PDB entries for which twinning was reported, a better strategy would be to analyse all entries for twinning using the unrelated difference test (perhaps limited to entries with deposited images).

Acknowledgements

STARANISO and autoPROC development team

Claus Flensburg — Andrew Sharff

Peter Keller — Ian J. Tickle

Clemens Vonrhein — Wlodek Paciorek

Gérard Bricogne

… and all users who have provided feedback and test data.