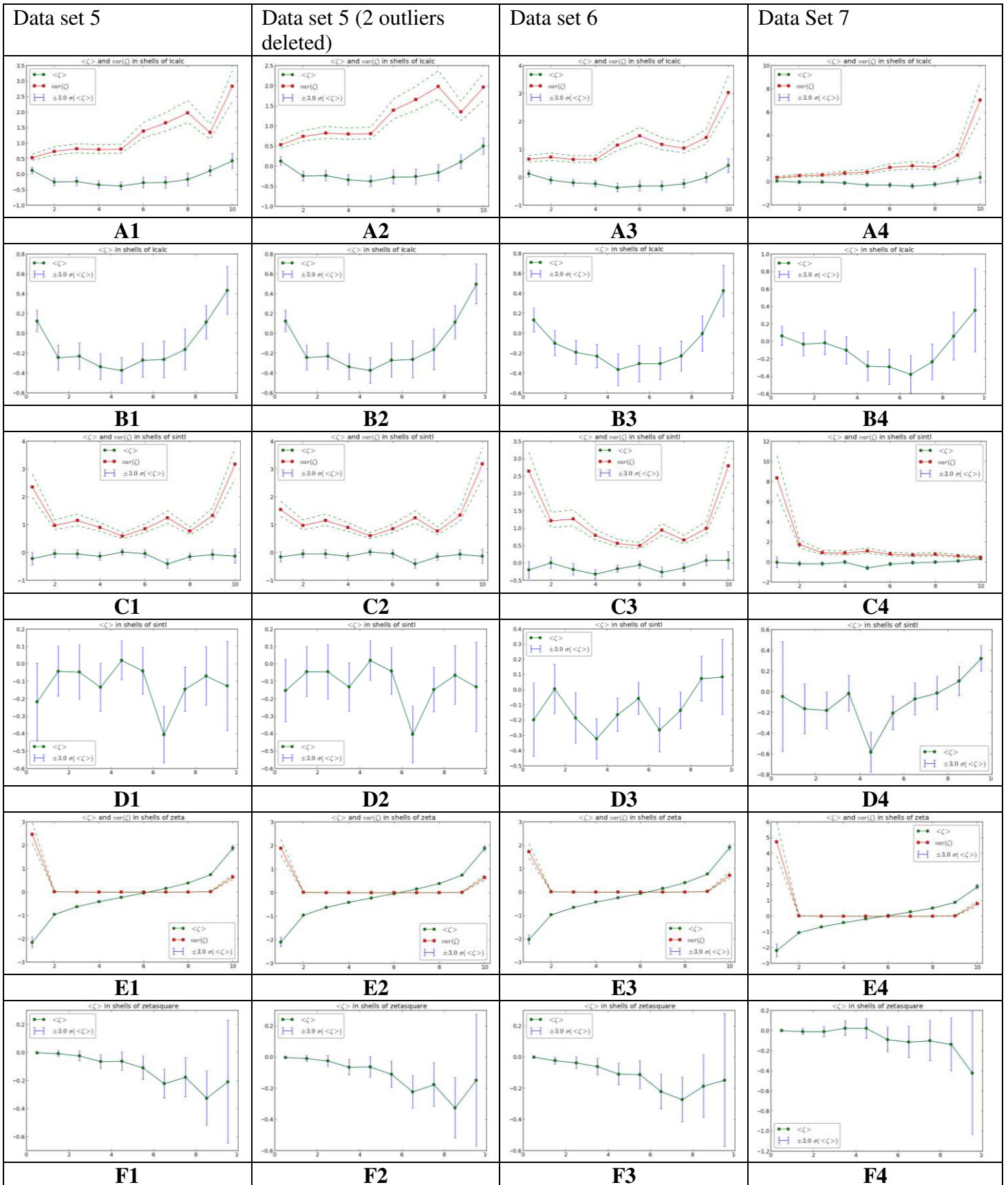# Statitical tests against systematic errors in data sets based on the equality of residual means and variances from control samples: Theory and Applications

Julian Henn[*a] and Kathrin Meindl[b]

Emil-Warburg Weg 6, 95447 Bayreuth, Germany and Instituto de Biologia Molecular de Barcelona (IBMB-CSIC), Barcelona Science Park, Baldiri Reixach, 08028 Barcelona, Spain. E-mail: Julian.Henn@uni-bayreuth.de

The supplementary material contains 9 diagnostic plots (with to some extent redundant information) for each of the data sets 5, 6, 7, 12, 13, 24 and additionally for data set 5 with 2 omitted outliers, data set 24 with cut off, data set 24 with manipulated s.u.'s, and data set 24 with 5 intentionally flawed reflections. A Figure is included that shows plots of residuals vs resolution and of observed versus calculates intensities for artificial data known to be free of systematic errors and contaminated with systematic errors due to inappropriate standard uncertainties in one case and due to 5 intentionally flawed intensity observations in the other case.
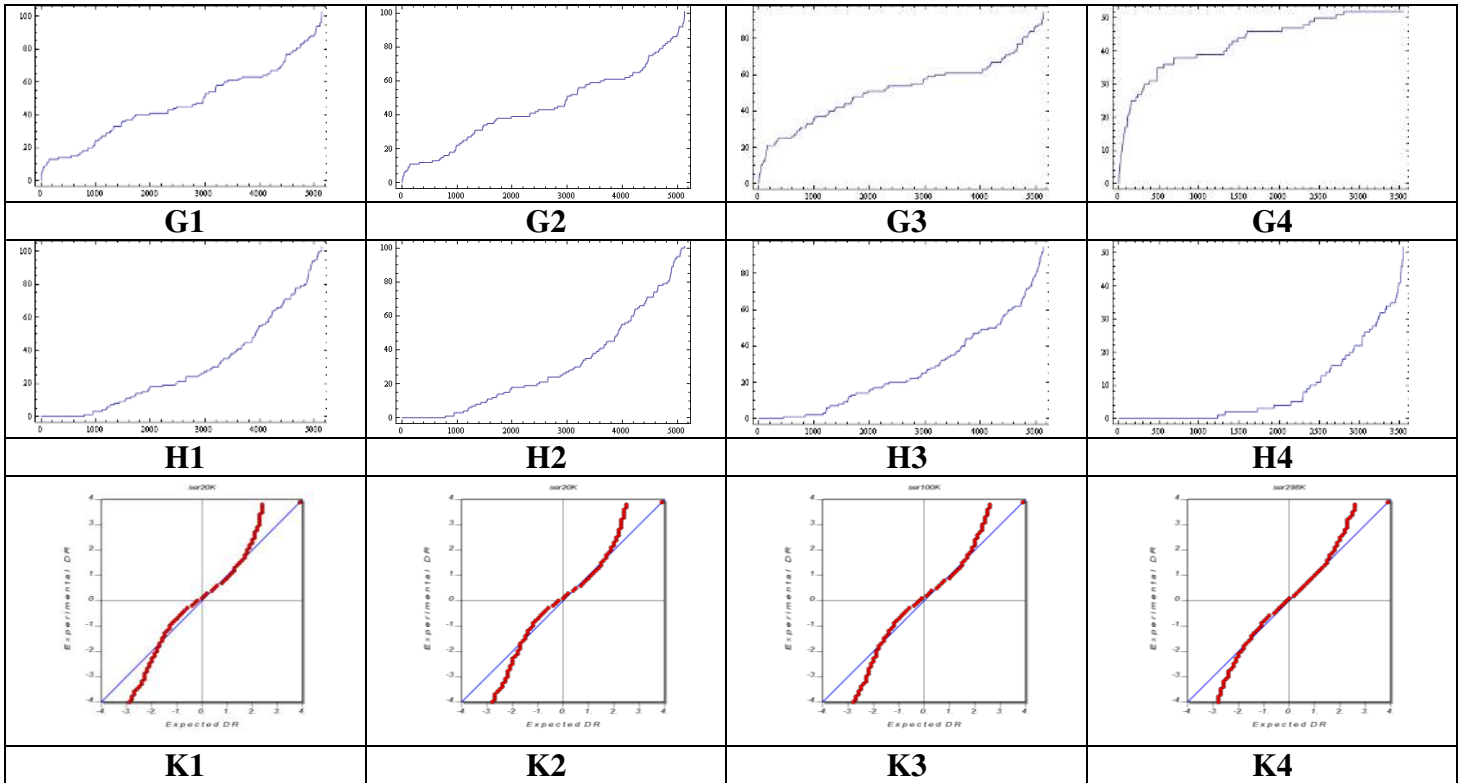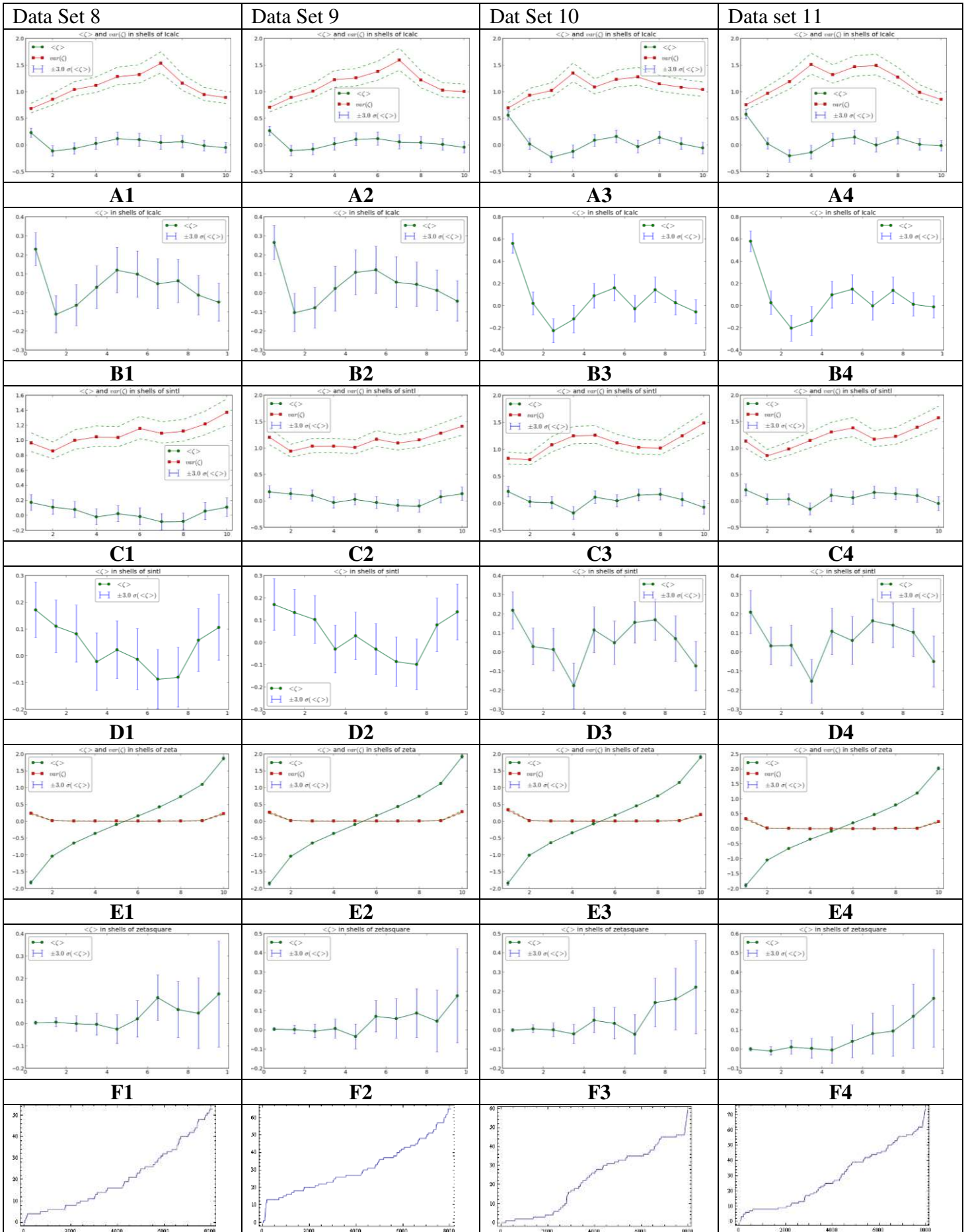
Data set 5	Data set 5 (2 outliers deleted)	Data set 6	Data Set 7
A1	**A2**	**A3**	**A4**
B1	**B2**	**B3**	**B4**
C1	**C2**	**C3**	**C4**
D1	**D2**	**D3**	**D4**
E1	**E2**	**E3**	**E4**
F1	**F2**	**F3**	**F4**

Fig. 1: Residual mean values and residual's mean variances sorted by increasing intensity, (A), residual mean values sorted by increasing intensity,(B), Residual mean values and residual's mean variances sorted by increasing resolution, (C), residual mean values sorted by increasing resoution,(D), residual mean values and residual's mean variances sorted by residuals, (E), residual mean values sorted by squared residuals, (F), cumulative rare events sorted by increasing resolution, (G), and by increasing intensity,(H), and normal probability plots, (K) for data sets 5-7.
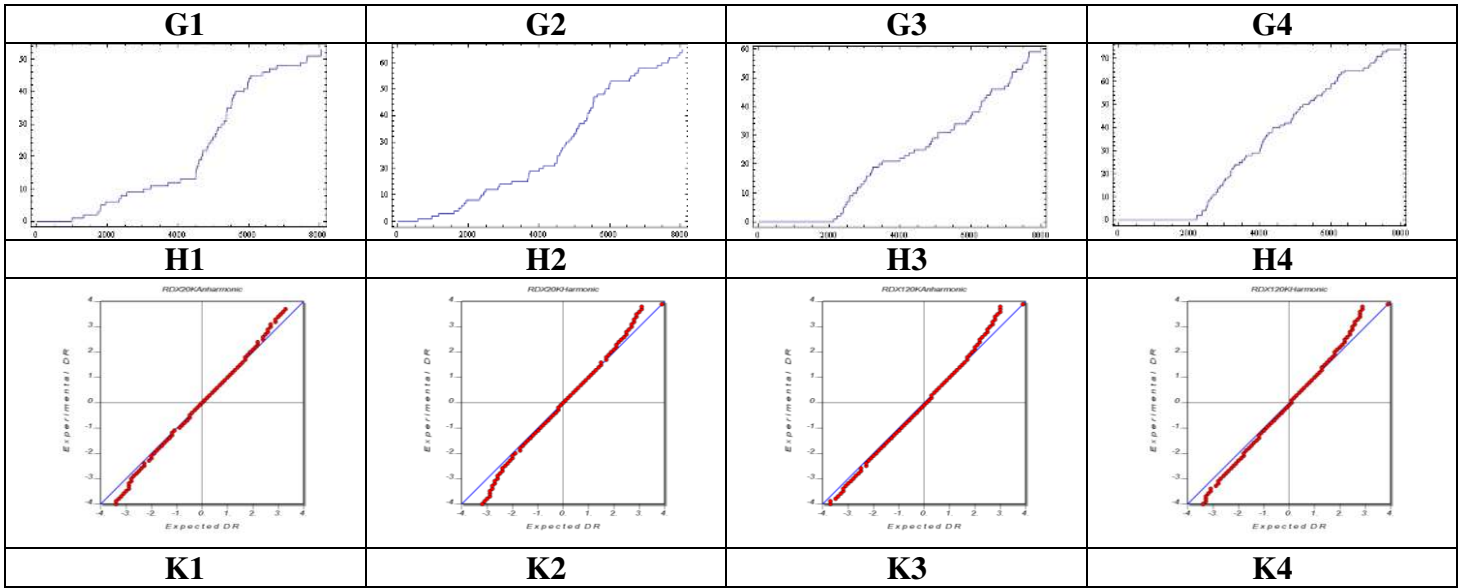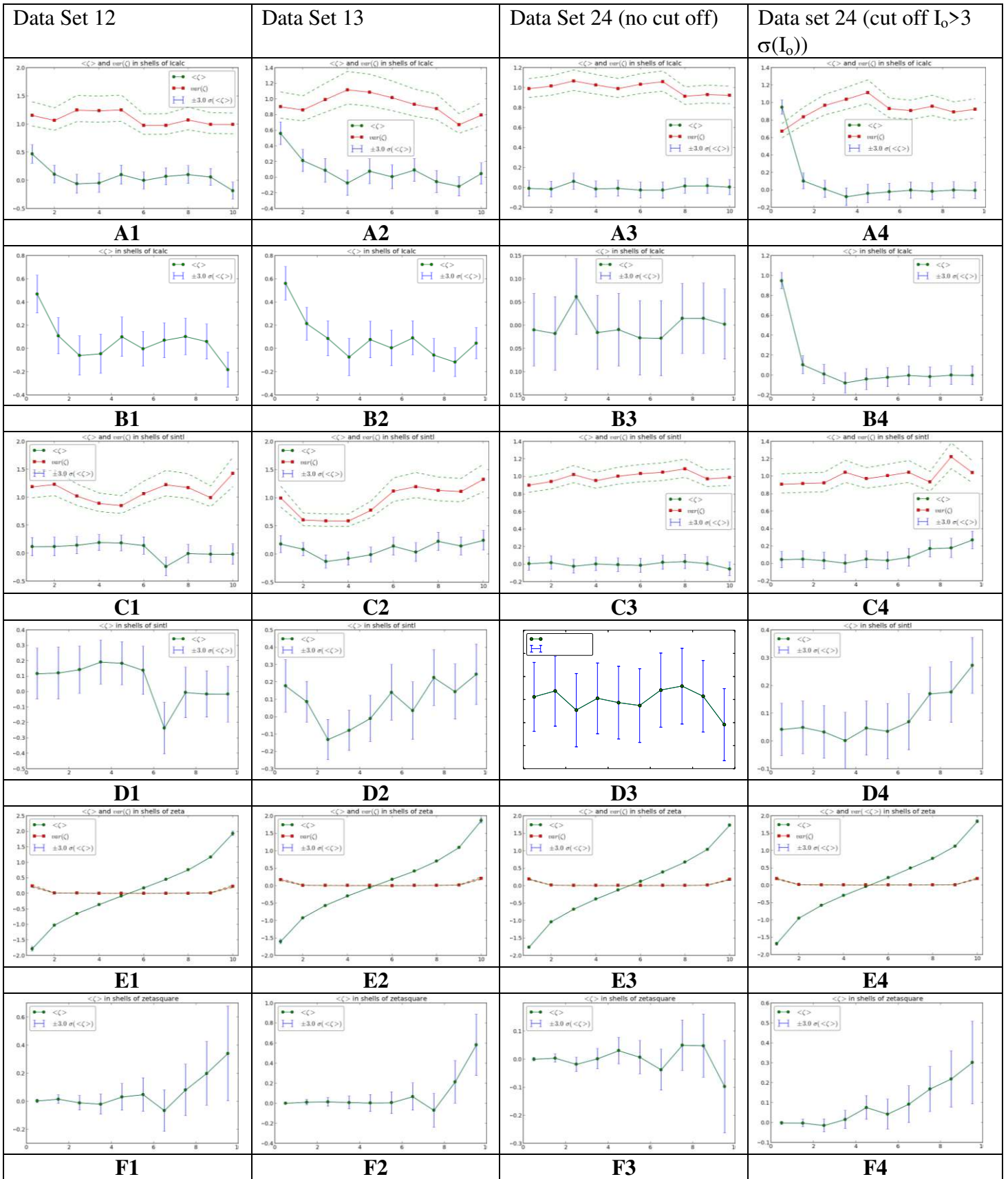
Data Set 8	Data Set 9	Dat Set 10	Data set 11
A1	**A2**	**A3**	**A4**
B1	**B2**	**B3**	**B4**
C1	**C2**	**C3**	**C4**
D1	**D2**	**D3**	**D4**
E1	**E2**	**E3**	**E4**
F1	**F2**	**F3**	**F4**

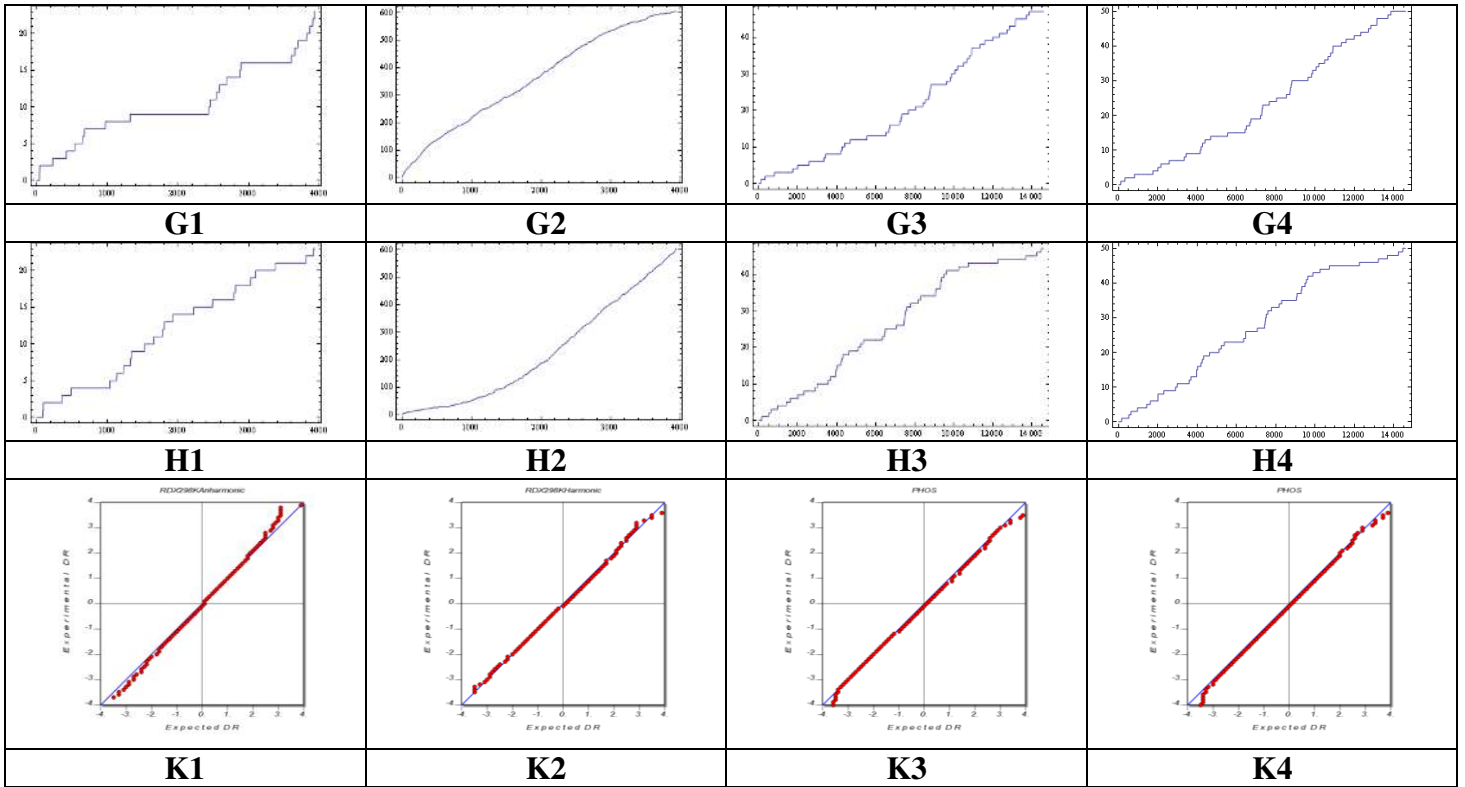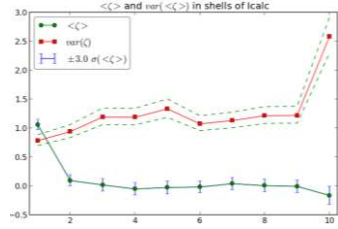| G1	G2	G3	G4
H1	**H2**	**H3**	**H4**
K1	**K2**	**K3**	**K4**

Fig. 2: Residual mean values and residual's mean variances sorted by increasing intensity, (A), residual mean values sorted by increasing intensity,(B), Residual mean values and residual's mean variances sorted by increasing resolution, (C), residual mean values sorted by increasing resoution,(D), residual mean values and residual's mean variances sorted by residuals, (E), residual mean values sorted by squared residuals, (F), cumulative rare events sorted by increasing resolution, (G), and by increasing intensity,(H), and normal probability plots, (K) for data sets 8-11.
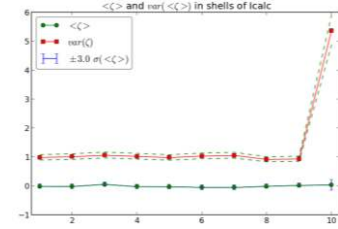
Data Set 12	Data Set 13	Data Set 24 (no cut off)	Data set 24 (cut off $I_o > 3\ \sigma(I_o)$)
A1	**A2**	**A3**	**A4**
B1	**B2**	**B3**	**B4**
C1	**C2**	**C3**	**C4**
D1	**D2**	**D3**	**D4**
E1	**E2**	**E3**	**E4**
F1	**F2**	**F3**	**F4**

Fig. 3: Residual mean values and residual's mean variances sorted by increasing intensity, (A), residual mean values sorted by increasing intensity,(B), Residual mean values and residual's mean variances sorted by increasing resolution, (C), residual mean values sorted by increasing resoution,(D), residual mean values and residual's mean variances sorted by residuals, (E), residual mean values sorted by squared residuals, (F), cumulative rare events sorted by increasing resolution, (G), and by increasing intensity,(H), and normal probability plots, (K), for data sets 12, 13, and the artificial data set 24 without (column 3) and with significance cut off (column4).

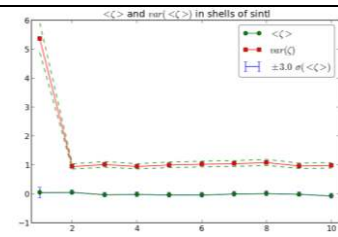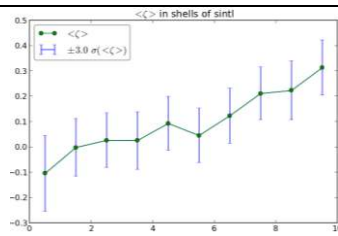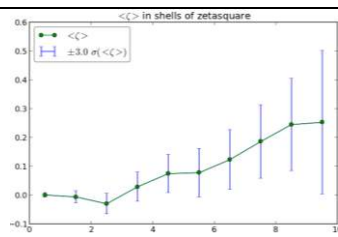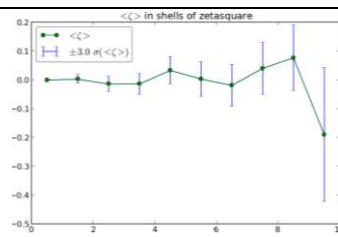| Data Set 24, manipulated σ values according to eq. (2)	Data set 24, 5 intentionally flawed reflections according to Table 1		
A1	**A2**		
B1	**B2**		
C1	**C2**		
D1	**D2**		
E1	**E2**		
F1	**F2**		
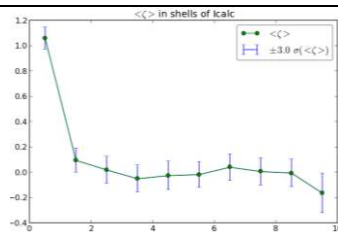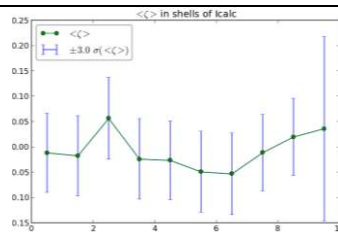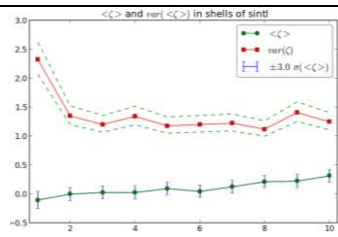
G1

G2

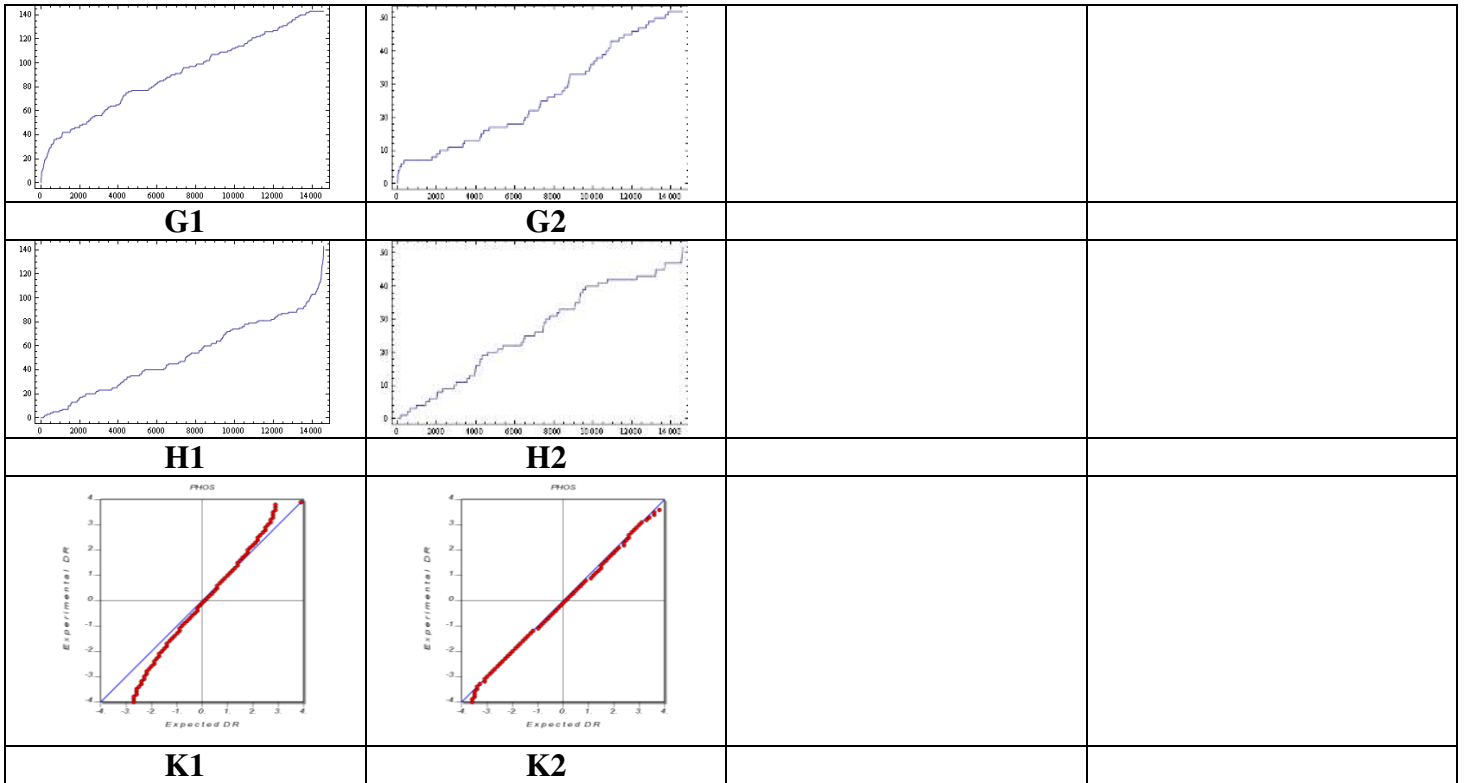**H1**

H2

K1
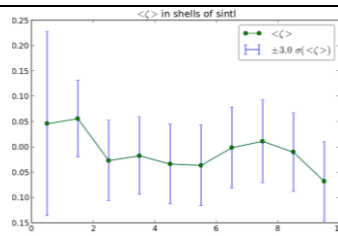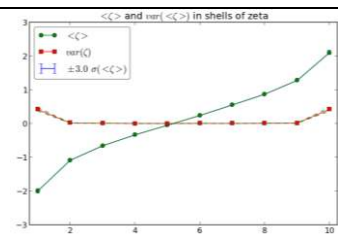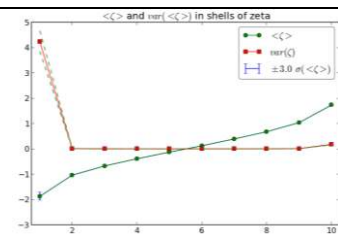
K2

Fig. 4: Residual mean values and residual's mean variances sorted by increasing intensity, (A), residual mean values sorted by increasing intensity,(B), Residual mean values and residual's mean variances sorted by increasing resolution, (C), residual mean values sorted by increasing resoution,(D), residual mean values and residual's mean variances sorted by residuals, (E), residual mean values sorted by squared residuals, (F), cumulative rare events sorted by increasinf resolution, (G), and by increasing intensity,(H), and normal probability plots, (K), for variations of the artificial data set 24 with intentionally introduced systematic errors.
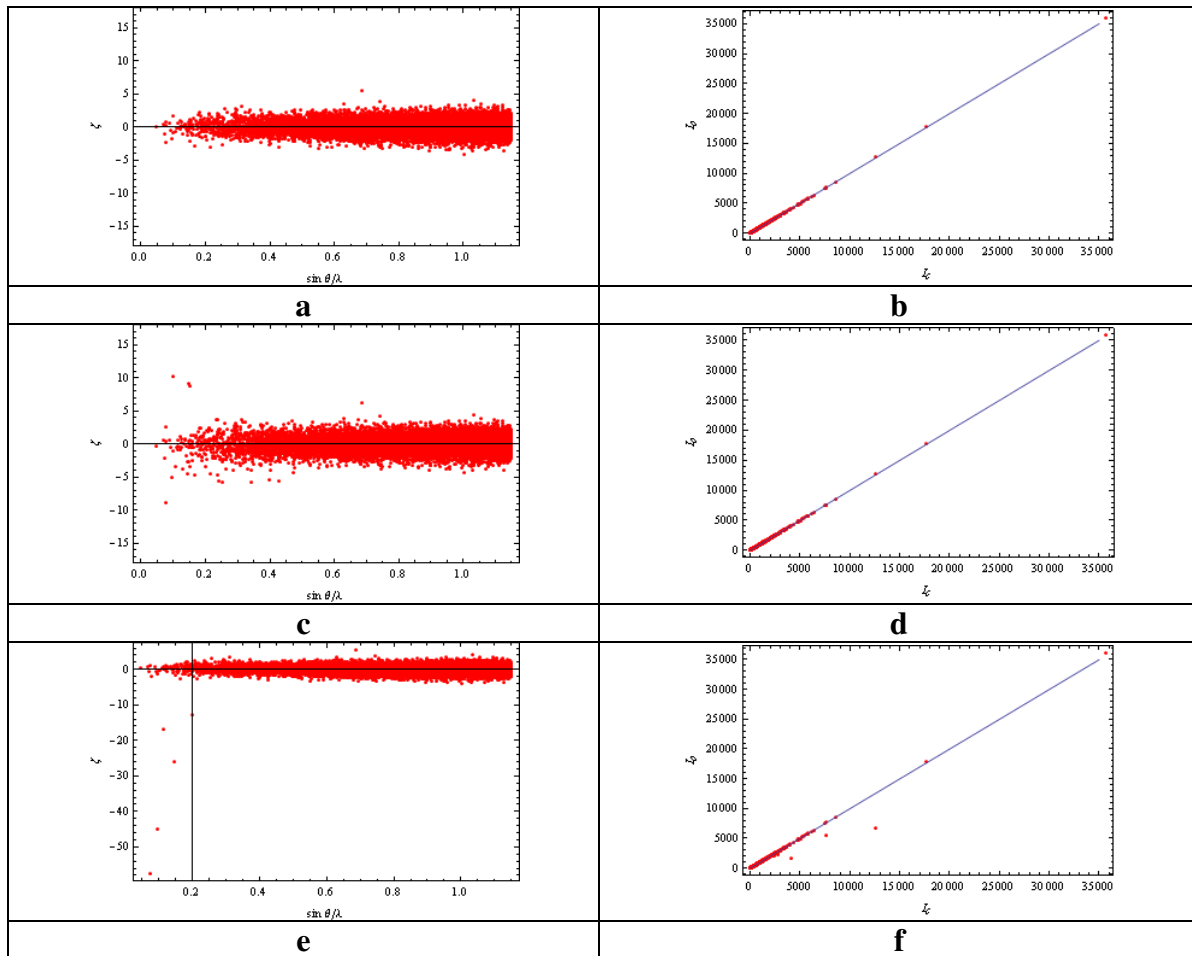
Fig. 5: Plots of residuals versus resolution (left column) and of observed versus calculated intensities (right) of artificial data set 24 without systematic error (top), with manipulated s.u.s according to eq. (2) in the manuscript and with 5 intentionally flawed observations (bottom).