Exploring DRI’s replicate carbon analyses

Warren H. White, Durango 9/2007

DRI replicates its carbon analyses at the rate of one per group of ten samples. Samples for reanalysis are randomly selected and randomly assigned.

DRI has produced a file of 5103 replicate analysis pairs from its new instruments, covering September 2005 to March 2007. The data come from 10 different instruments, with identification numbers from 6 to 15.

number of pairs

replicate instrument

original instrument

	6	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
6	89	87	96	97	81	74	82	18	51	25
7	90	101	105	70	65	61	77	19	44	24
8	78	72	95	99	82	52	64	23	53	17
9	81	73	83	100	89	80	75	12	53	15
10	84	81	87	86	107	64	61	13	43	18
11	71	61	57	69	56	69	69	15	32	18
12	71	72	72	80	56	77	70	10	38	20
13	19	14	27	13	15	21	14	26	15	17
14	36	33	52	57	44	35	46	13	59	10
15	11	16	24	19	11	14	16	15	14	18

Question: are the different instruments indistinguishable?

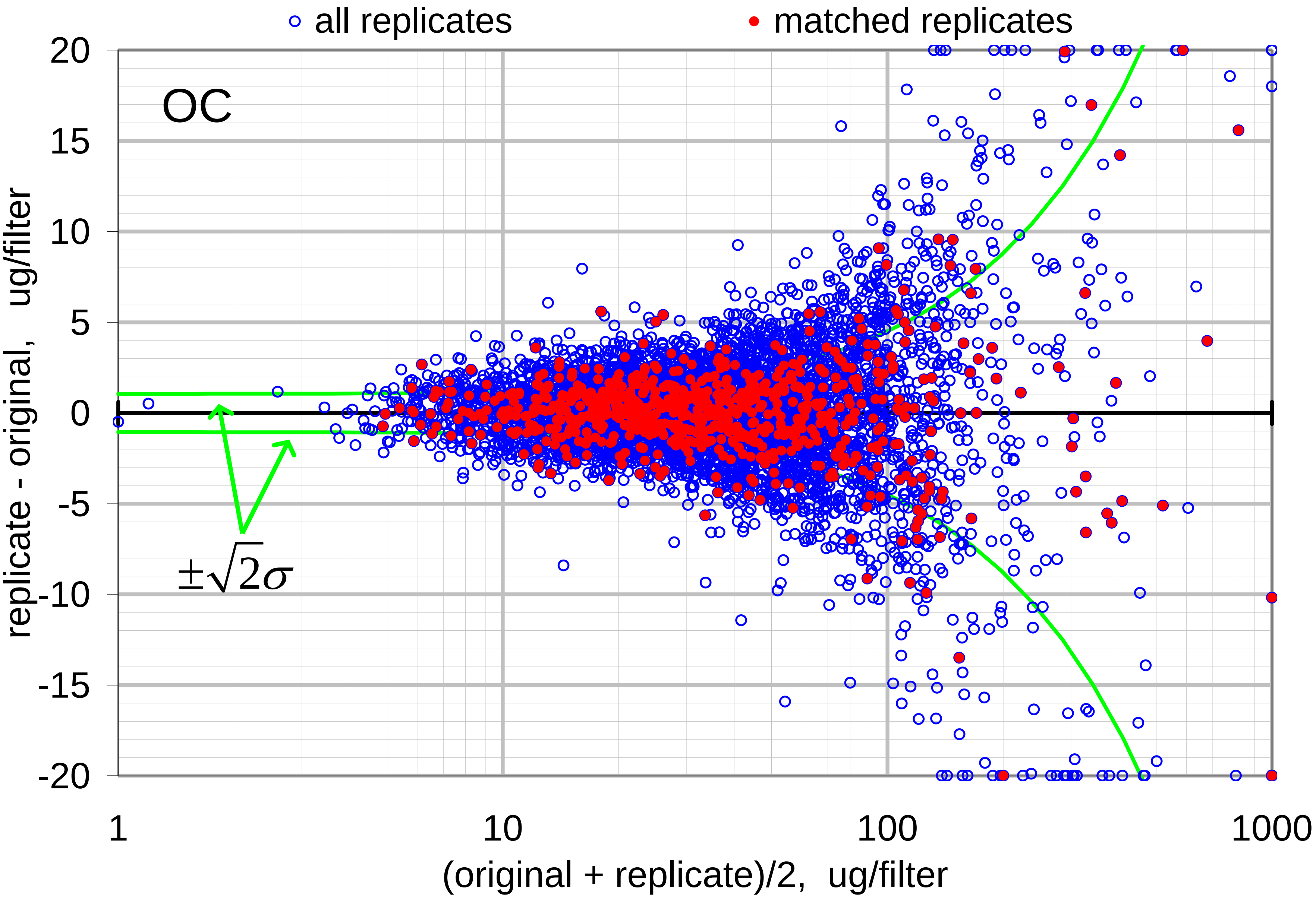
Each replicate analysis pair involves a different filter, no two sampled on the same day at the same location.

We can expect, *a priori*, that more-heavily-loaded filters will tend to yield larger absolute differences when reanalyzed, and that less-heavily-loaded filters will tend to yield larger relative differences.

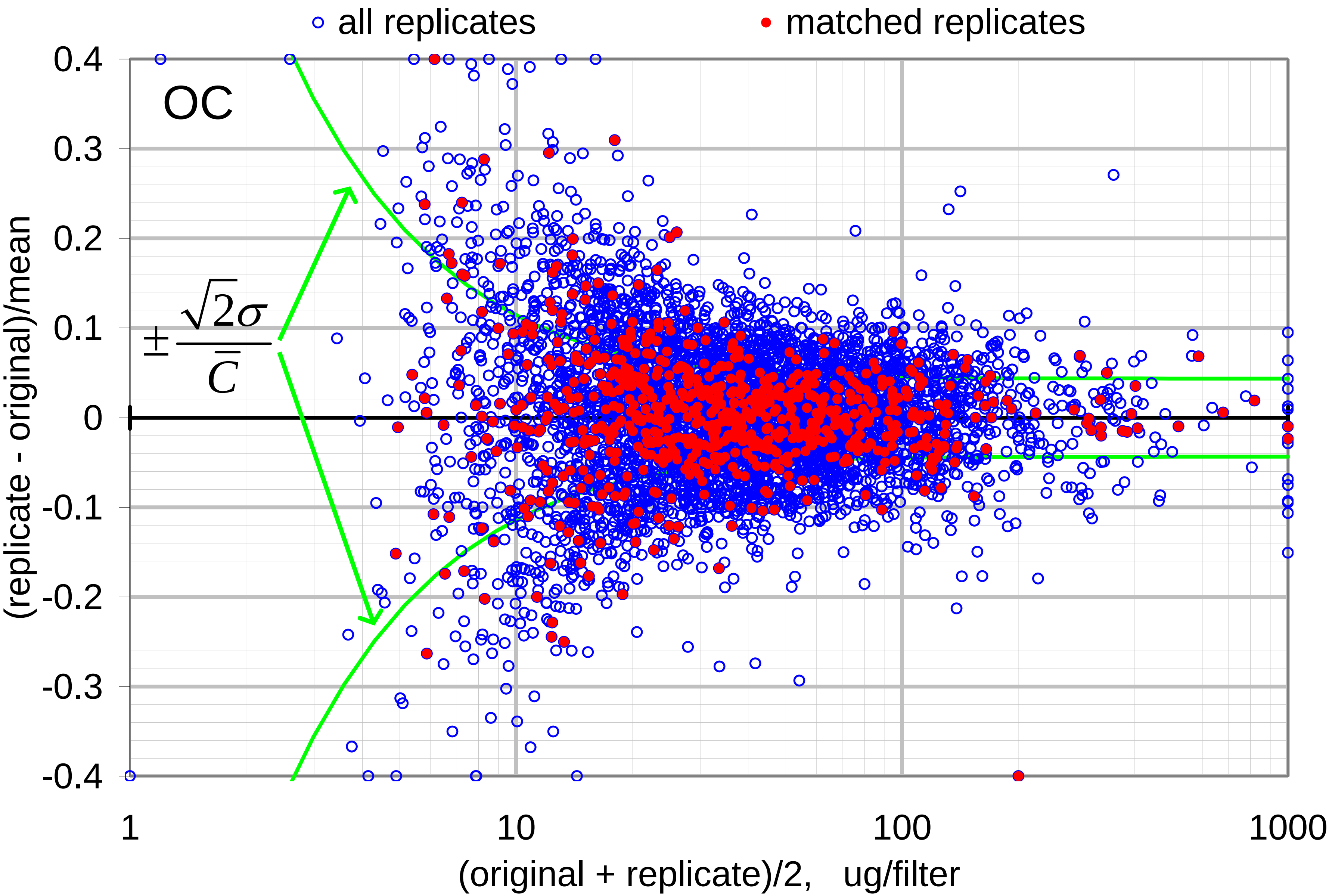
The observed reanalysis differences can be described by the following formula for measurement uncertainty (sigma) as a function of concentration (*C*):

$$\sigma\left(C\right)=\sqrt{\alpha^{2}+\beta^{2}C^{2}}.$$

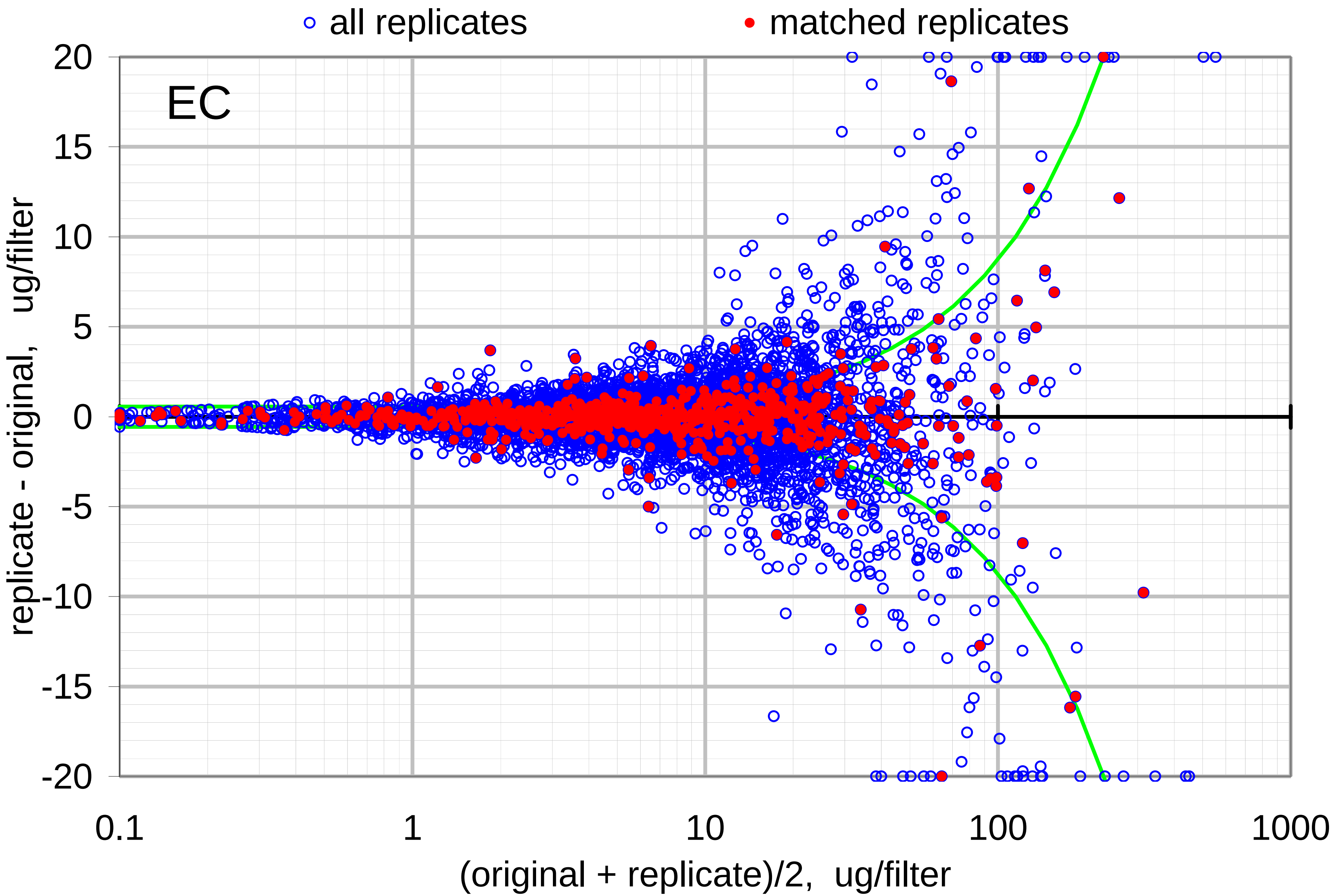
The following plots show this formula for OC and EC, with parameters fitted to data from reanalyses on the same instrument.



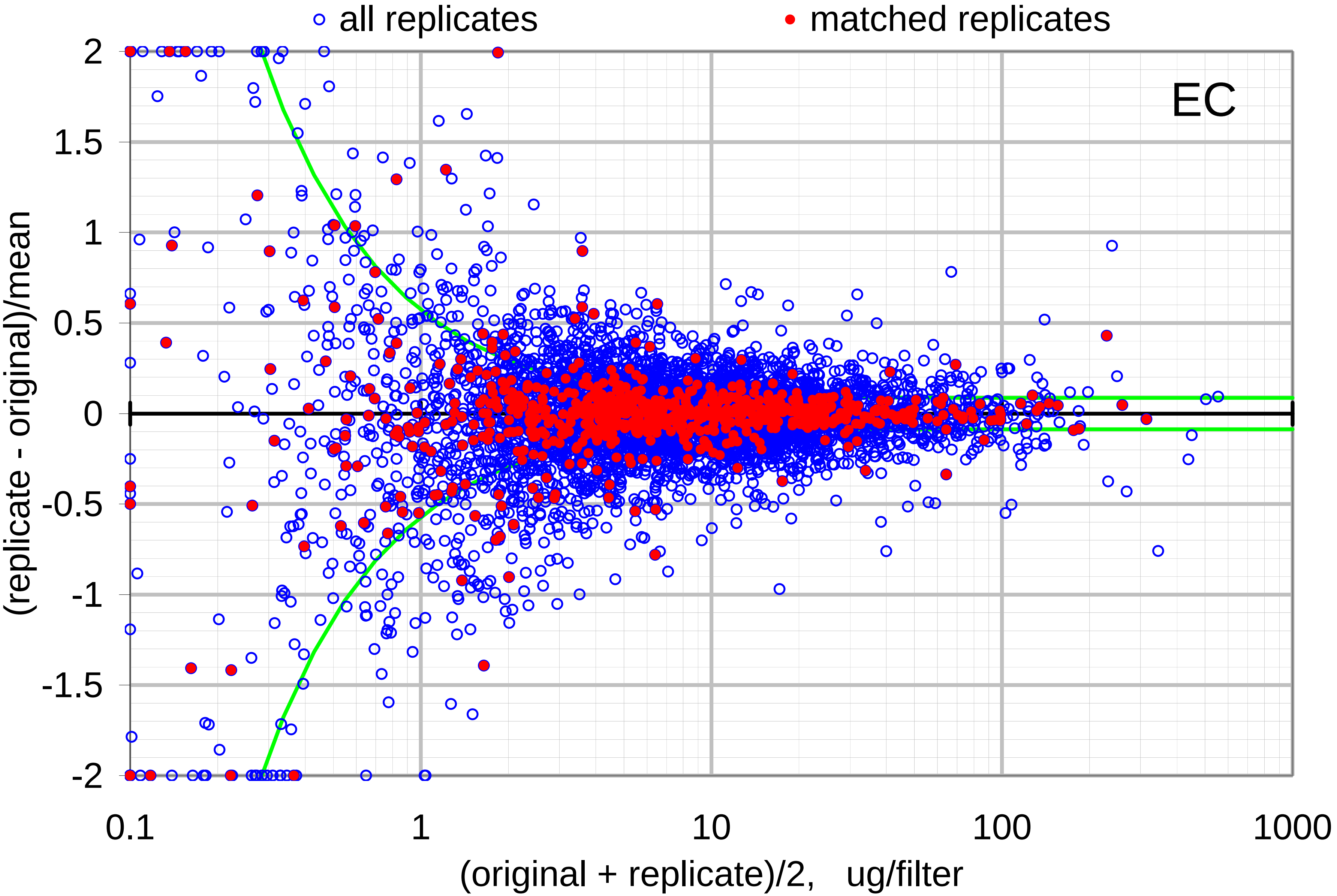
$\sigma_{OC}\left(\overline{C}\right)=\sqrt{0.75^2+0.031^2\overline{C}^2}$ was fitted by ML estimation to **matched** data



$$\sigma_{OC}\left(\bar{C}\right) = \sqrt{0.75^2 + 0.031^2\,\bar{C}^2}$$



$$\sigma_{EC}\left(\overline{C}\right) = \sqrt{0.40^2 + 0.061^2\,\overline{C}^2}$$



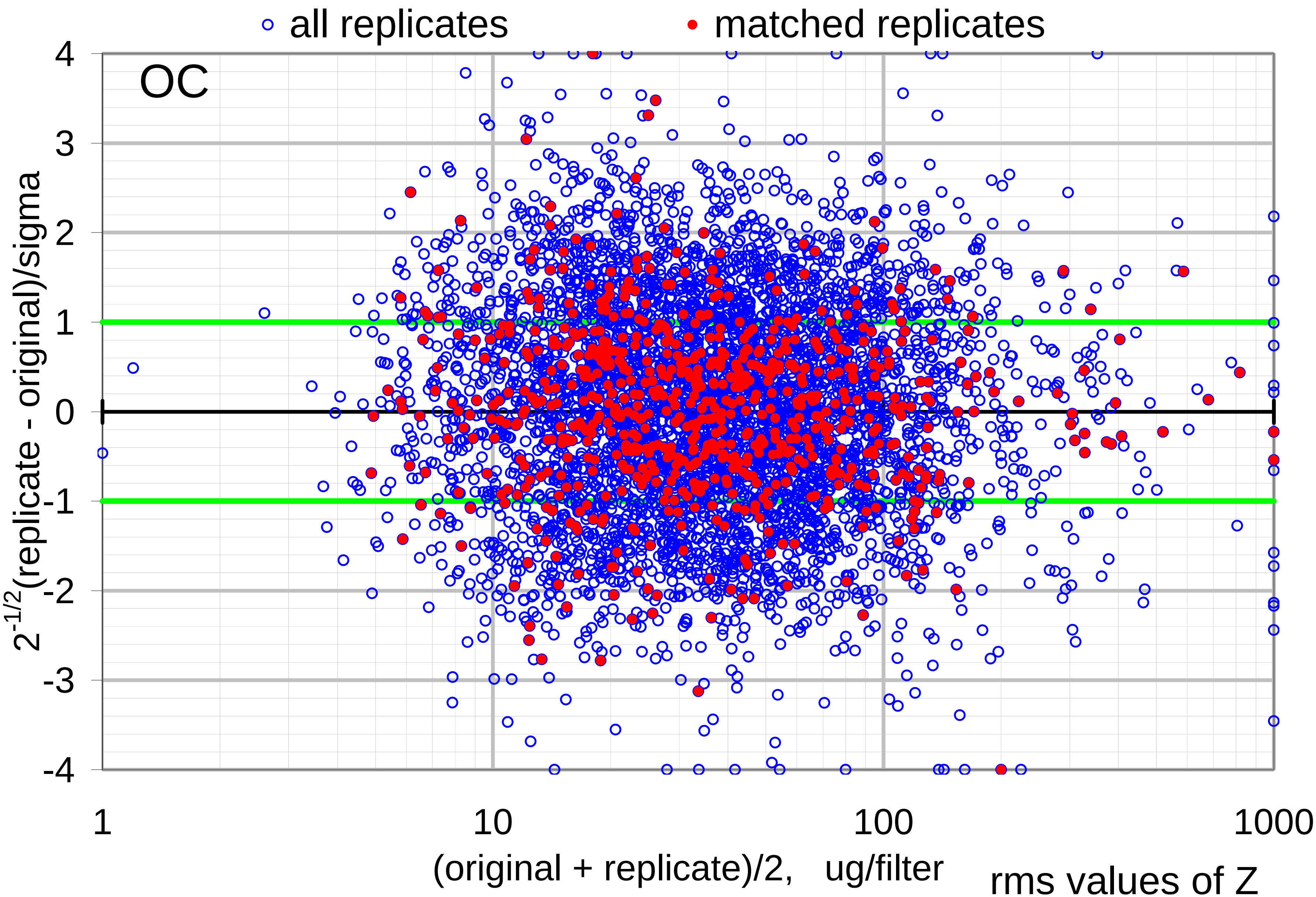
$$\sigma_{EC}\left(\overline{C}\right)=\sqrt{0.40^2+0.061^2\,\overline{C}^2}$$

Rather than try to account for these dependences in our statistical analysis, we will normalize them out. We divide the observed differences by $2^{1/2}$sigma to obtain a variable Z whose distribution is independent of concentration:

$$Z = \frac{1}{\sqrt{2}} \frac{C_{replicate} - C_{original}}{\sigma\left(\bar{C}\right)}.$$

All of the replicates are now on an equal statistical footing.

(The factor $2^{1/2}$ accounts for the presence of uncertainty in each of 2 independent measurements.)

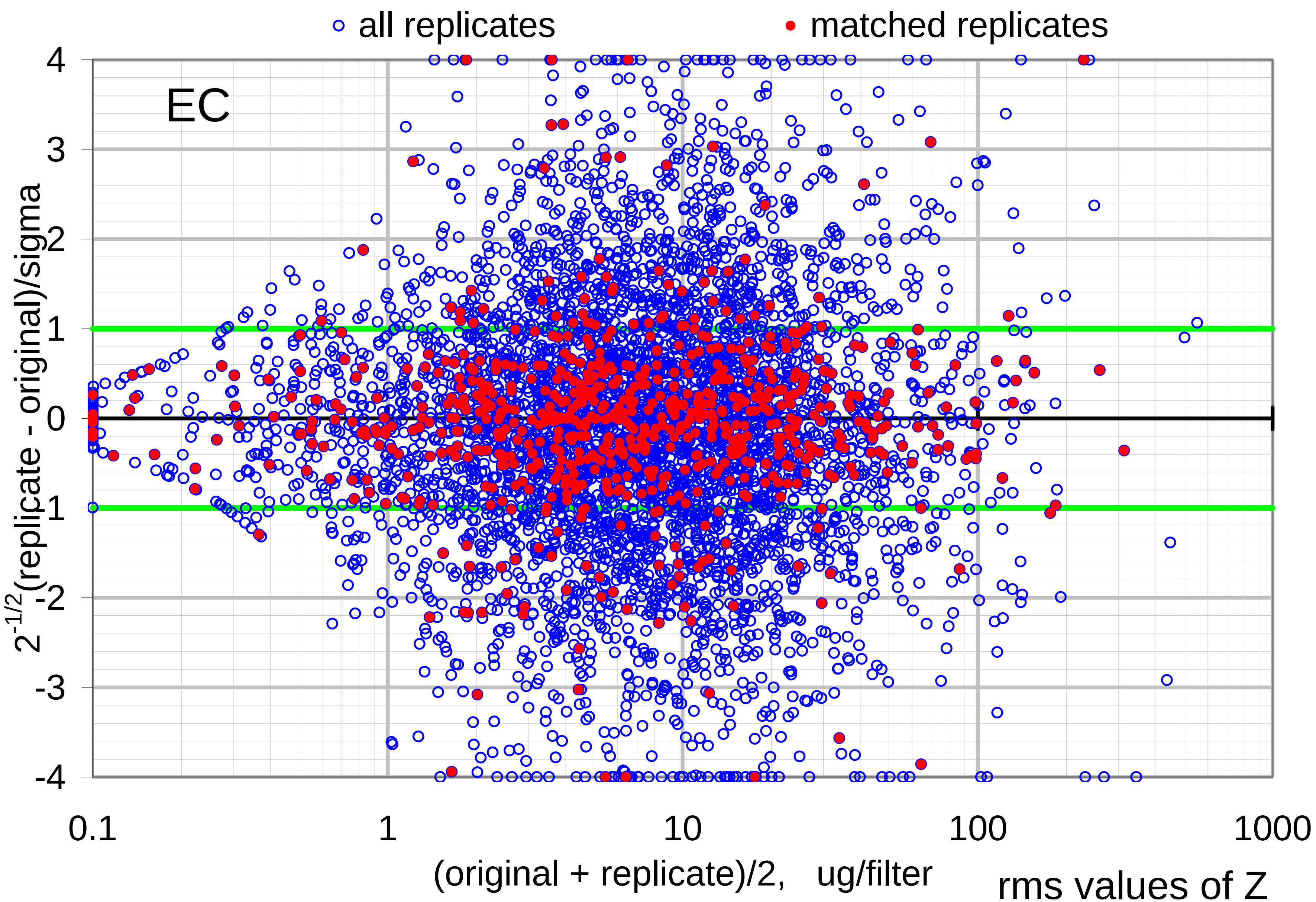


$$\sigma_{OC}\left(\bar{C}\right)=\sqrt{0.75^2+0.031^2\bar{C}^2}$$

rms values of Z

matched: 1.00

unmatched: 1.25



$$\sigma_{EC}\left(\overline{C}\right)=\sqrt{0.40^2+0.061^2\,\overline{C}^2}$$

rms values of Z

matched: 1.00

unmatched: 1.56

OC

$$Z = 2^{-1/2}(\text{replicate} - \text{original})/\text{sigma}$$

						replicate					
	OC signif mean Z	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
original	**6**			-0.28	-0.60	-0.58	0.37				
	7	0.30			-0.38	-0.35	0.32	0.42			
	8				-0.34	-0.51	0.43				
	9	0.78	0.55	0.44			0.65	1.09			
	10	0.77	0.92	0.64			0.70	1.11	1.19	0.45	0.82
	11	-0.33	-0.40	-0.45	-0.81	-0.65			-0.53		
	12		-0.40	-0.37	-0.85	-0.89		0.27	-0.71		-0.53
	13				-0.76	-1.39		1.44			
	14	0.59		0.90				0.57			
	15				-0.46	-1.28					

						replicate					
	OC signif rms Z	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
original	**6**				1.22	1.22	1.29	1.22		1.50	
	7	1.17	1.41	1.14	1.38	1.40		1.30		1.47	
	8			0.76		1.36				1.25	
	9	1.19	1.38			1.17	1.26	1.45	1.40	1.42	
	10	1.29	1.39	1.15	1.22		1.39	1.45	1.43	1.47	
	11		1.45		1.34	1.42				1.50	
	12				1.24	1.29	1.26	0.82		1.57	
	13					1.57		2.19		1.63	
	14	1.64	1.37	1.46	1.29	1.24	1.45	1.45	1.93		2.33
	15	1.46									0.54

Here are the instrument-specific statistics for the normalized differences, Z. For completely equivalent instruments, mean Z should be 0, and rms Z should be 1. The mean Z over all 5103 replicate analyses is +0.057 +/- 0.017, a statistically significant but substantively trivial bias.

$Z = 2^{-1/2}$(replicate – original)/sigma

OC

						replicate					
	OC signif mean Z	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
original	**6**			-0.28	-0.60	-0.58	0.37				
	7	0.30			-0.38	-0.35	0.32	0.42			
	8				-0.34	-0.51	0.43				
	9	0.78	0.55	0.44			0.65	1.09			
	10	0.77	0.92	0.64			0.70	1.11	1.19	0.45	0.82
	11	-0.33	-0.40	-0.45	-0.81	-0.65			-0.53		
	12		-0.40	-0.37	-0.85	-0.89		0.27	-0.71		-0.53
	13				-0.76	-1.39		1.44			
	14	0.59		0.90				0.57			
	15				-0.46	-1.28					

						replicate					
	OC signif rms Z	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
original	**6**				1.22	1.22	1.29	1.22		1.50	
	7	1.17	1.41	1.14	1.38	1.40		1.30		1.47	
	8			0.76		1.36				1.25	
	9	1.19	1.38			1.17	1.26	1.45	1.40	1.42	
	10	1.29	1.39	1.15	1.22		1.39	1.45	1.43	1.47	
	11		1.45		1.34	1.42				1.50	
	12				1.24	1.29	1.26	0.82		1.57	
	13					1.57		2.19		1.63	
	14	1.64	1.37	1.46	1.29	1.24	1.45	1.45	1.93		2.33
	15	1.46									0.54

Some individual instruments stand out from the crowd. For example, instrument #10 is biased low for OC relative to other instruments.

Mean(replicate – original) < 0 when it provides the original analysis, and mean(replicate – original) > 0 when it provides the replicate analysis.

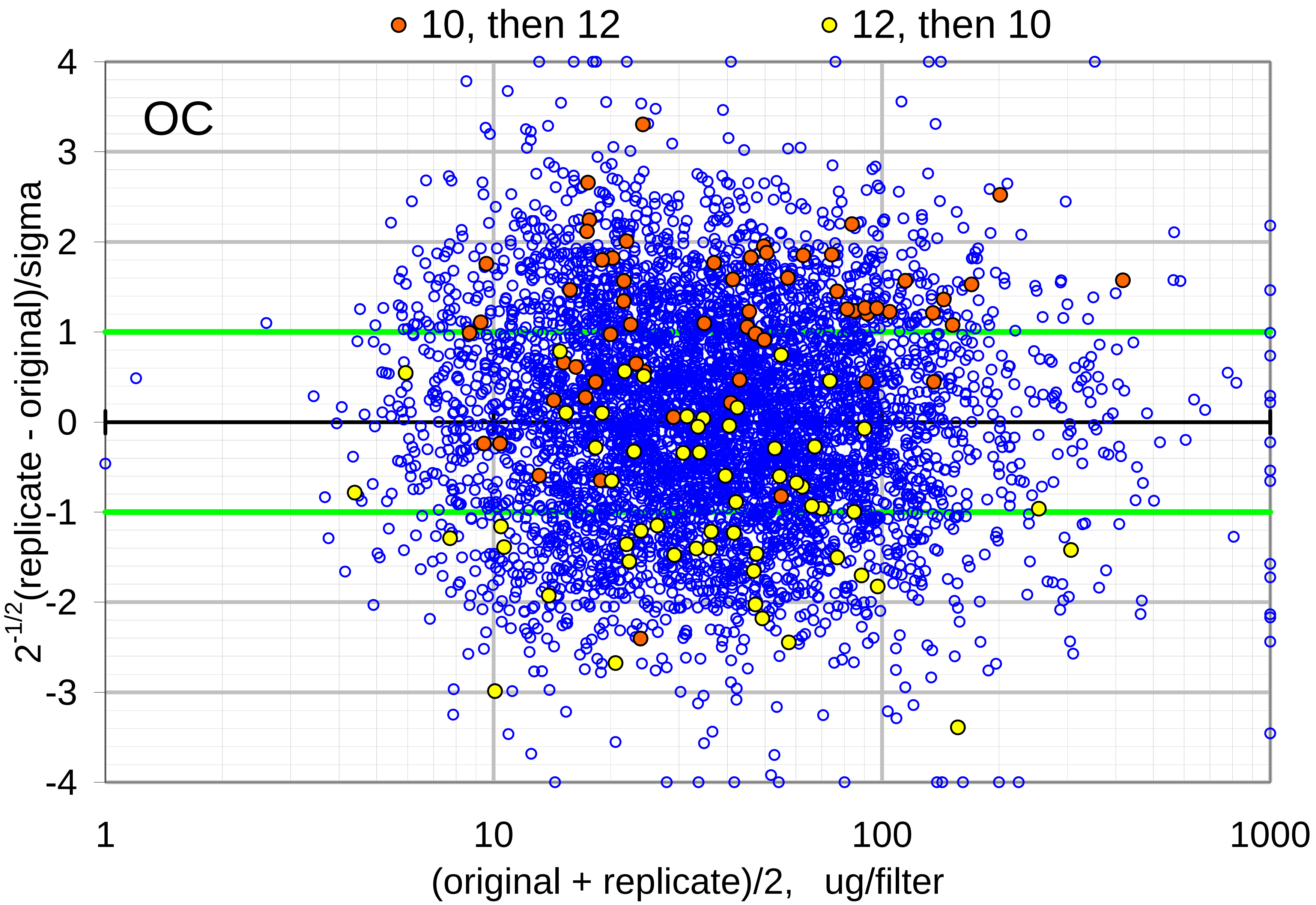
$Z = 2^{-1/2}(\text{replicate} - \text{original})/\text{sigma}$

EC

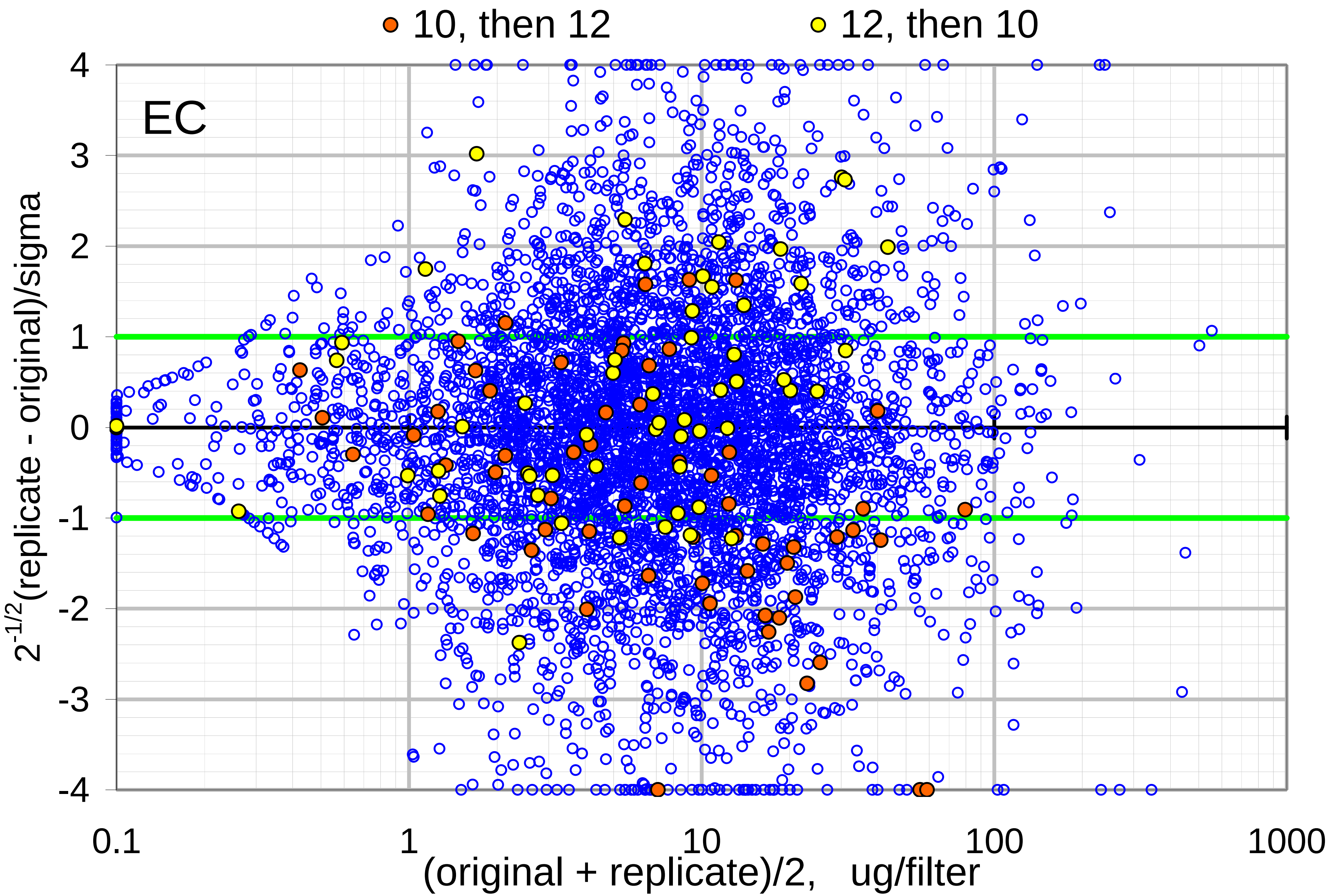
						replicate					
	EC signif mean Z	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
original	**6**		0.75	0.32	0.47		-0.62	-0.31			
	7	-1.07	-0.22	-0.50	-0.57	-0.73	-1.18	-1.39		-1.10	
	8	-0.52	0.49		-0.36		-0.75	-0.45		-0.85	0.56
	9	-0.43	0.68				-1.17	-0.46		-0.44	0.71
	10	-0.40		-0.33			-1.12	-0.85			
	11	1.13	1.52	1.22	1.21	1.05					1.53
	12		0.62	0.36	0.42	0.36					0.57
	13				-1.23		-1.67				
	14		0.70					-0.57			
	15				-0.76		-2.28	-0.94		-2.26	

						replicate					
	EC signif rms Z	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
original	**6**		1.50	1.20	1.47	1.30	1.85	1.18	2.21	1.34	
	7	1.61		1.19	1.32	1.34	1.96	1.92		1.84	
	8	1.34	1.26		1.43	1.44	1.52			1.73	
	9	1.35	1.47		0.78	1.37	1.80	1.21		1.44	
	10	1.25	1.83	1.42	1.44	1.22	2.26	1.84	2.04	1.78	1.52
	11	1.99	2.36	1.94	1.77	2.59	0.69	1.78	1.51	1.45	1.96
	12	1.58	1.45		1.47	1.21	1.42	0.61		1.67	
	13				1.54	1.91	2.29	1.78		2.41	
	14	1.52	1.77	1.30	1.49	1.43	1.68	1.70	1.54	1.28	1.95
	15	1.46					3.16	1.59		2.59	1.51

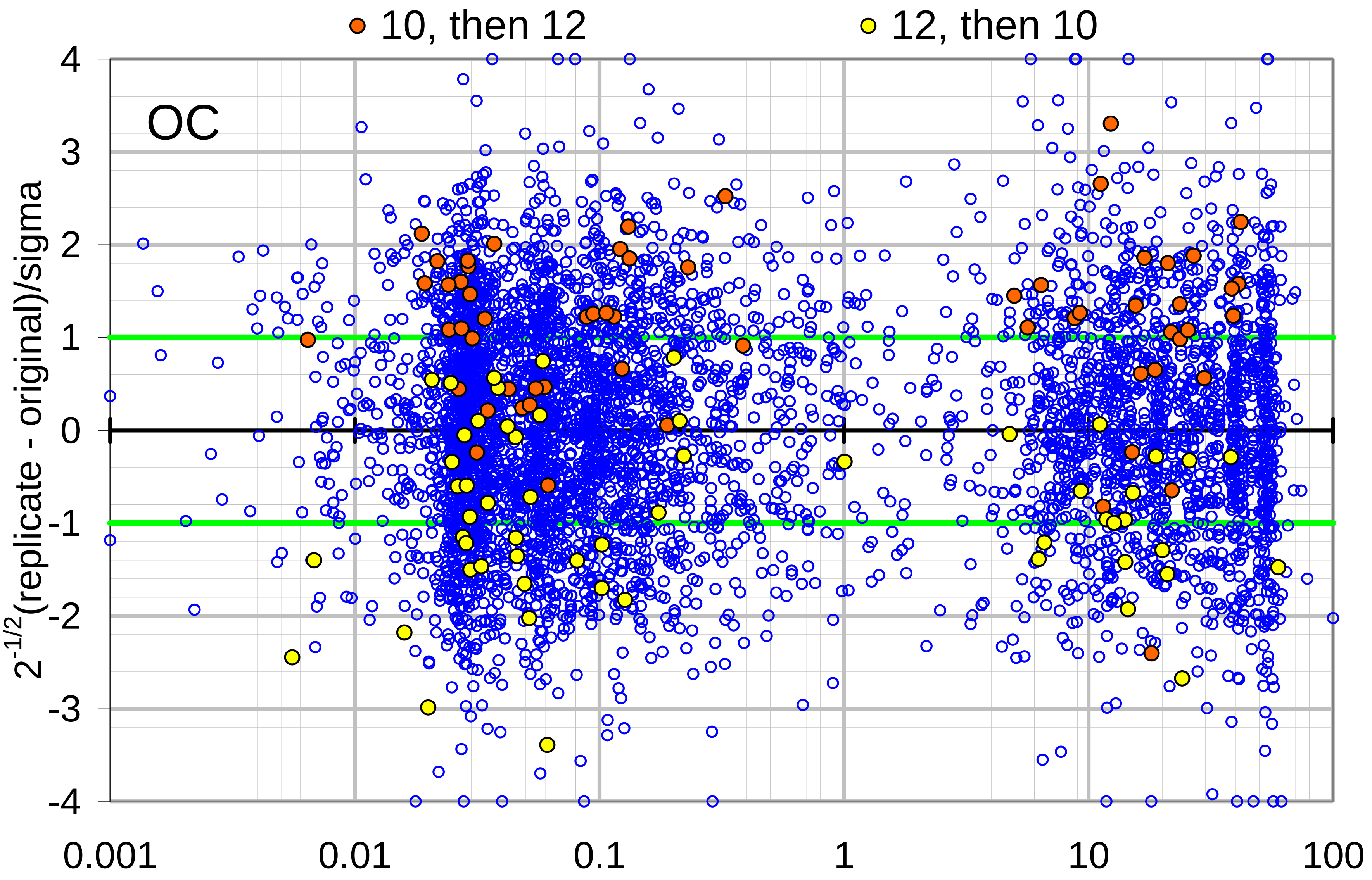
The mean Z for EC over all 5103 replicate analyses is -0.094 +/- 0.021, another statistically significant but substantively trivial bias.



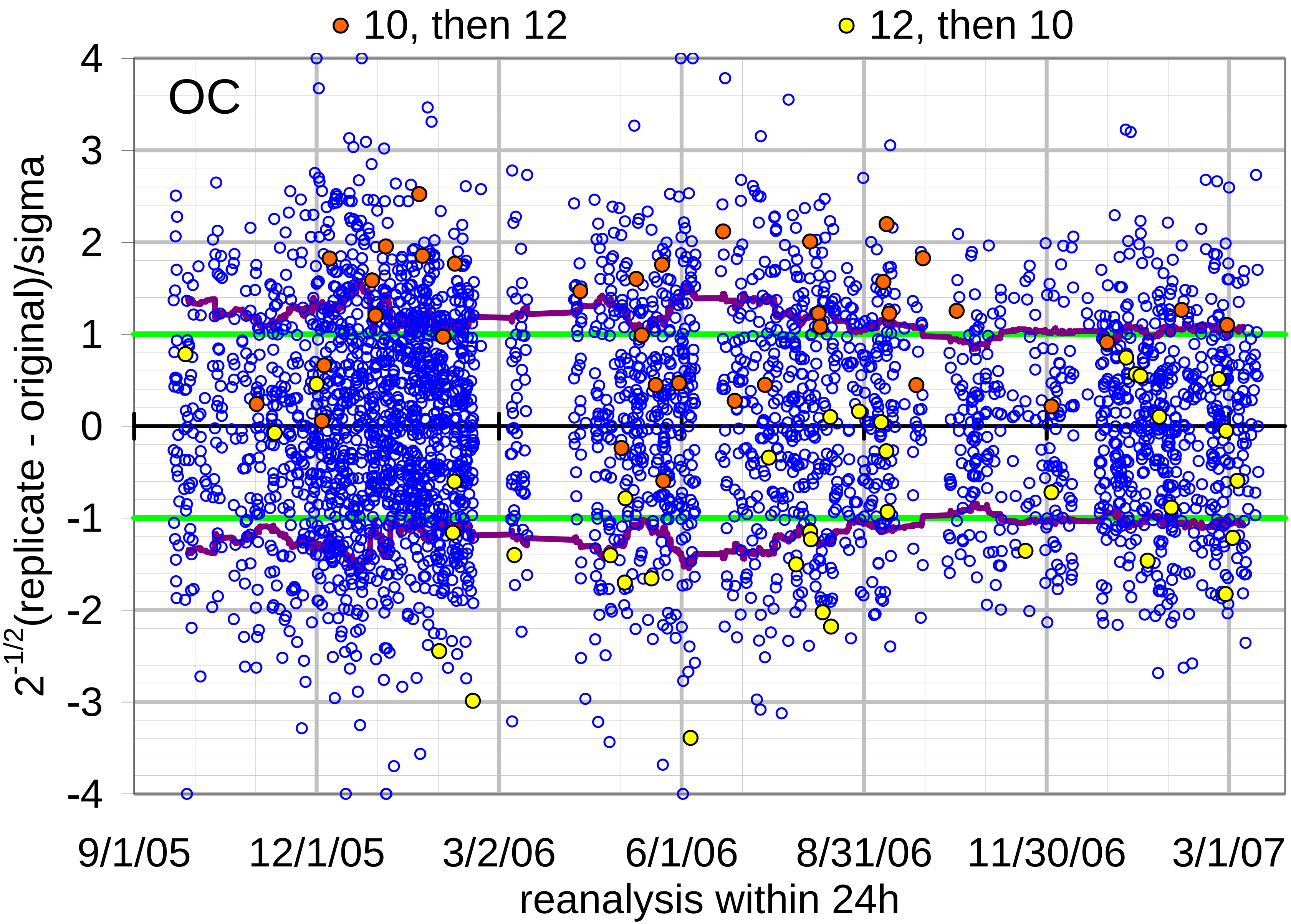
Here is a comparison of specific instruments in the context of all replicate analyses. This example was chosen specifically to highlight differences.



For EC, these two instruments' biases are reversed.



The time interval between the original and replicate analyses has no evident effect on observed differences.



Nor is there an evident trend over time, except that the Z values tighten (i.e., improve) somewhat with experience. The purple curves show the running rms over all differences in a centered 91-pair window.