Electronic Supplementary Material (ESI) for Journal of Environmental Monitoring

APPENDIX

Interlaboratory evaluation of trace element determination in workplace air filter samples by inductively coupled plasma mass spectrometry

Kevin Ashley,*[a] Stanley A. Shulman,[a] Michael J. Brisson[b] and Alan M. Howe[c]

[a] US Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, M.S. R-7, Cincinnati, OH 45226-1998, USA. E-mail: KAshley@cdc.gov; Tel.: +1.513.841.4402

[b] Savannah River Nuclear Solutions, Savannah River Site 772-F, Aiken, SC 29808, USA

[c] Health and Safety Laboratory, Harpur Hill, Buxton, Derbyshire SK19 9JN, UK

Data distribution

The linear mixed model[1] was fitted to the relative bias data separately for each metal and produced estimates of random lab deviations from the overall mean and of residual deviations of the three lab measurements from the lab mean. Deviations are assumed to be normally distributed. Box plots are presented in Figs. A1 and A2 for random and residual effects of deviations for each metal, respectively. In the top plot of each pair of figures, all the data are plotted. In the bottom plot of each pair, random effects with absolute value > 5 are omitted. Although there is some skewness, there is considerable evidence that potential outliers can be either at the low or high end for either deviation. This is the justification for using the normal theory.

Robust methods

Robust estimates were obtained via the statistical package R.[2] More sophisticated procedures[1] rely on achieving what is called a breakdown point, which specifies the fraction of data that can be contaminated by outliers without having much effect on estimates. For the results presented here, the breakdown point was chosen as 0.50, since there seemed to be a large number of outliers based on the box plots (Figs. A1 and A2). Results from different methods[1,2] were compared: 1. a minimum covariance matrix determinant method, 2. constrained M estimator, 3. MM method, and 4. the Stahel-Donoho estimator. The first was found to overestimate the variance, based on simulation results. The third method sometimes gave unstable estimates for smaller data sets. Although some of the estimates for the Stahel-Donoho method differ considerably from the constrained M and MM estimators, the overall conclusions are similar. Results are presented here based on the constrained

M estimator; there were not usually large differences between the constrained M estimates and MM estimates.

There are recommendations for adequate sample size for use of these methods: n=15 when there are three determinations by each lab.[1] For the median substitution data, for 14 elements there were at least 14 labs, which is quite close to the recommended number. For the four elements that had six to eight labs (Ag, Se, V, U), the number of labs seemed too small. For these metals, up to two sequential computations of the ASTM E691[3] h and k statistics were used to remove outliers.

Median substitution vs. no substitution

For each element, two versions of the data were evaluated. The first entailed omission of labs that did not report MDLs nor values for their media blank samples. The second entailed substitution of the median blank value for these labs. With substitution, 14 of the 21 metals had at least 14 labs evaluated, compared with only 6 for no substitution. With substitution, no metals had fewer than 11 labs, whereas without substitution five metals had fewer than 11 labs.

The differences in bias estimates (substitution model estimate – no substitution model estimate) are plotted in Fig. A3 for elements spiked at low levels. When the median used for substitution approaches the target value, the bias estimates for those metals have large negative bias compared to no substitution bias. Five metals had bias estimate differences of about -0.2 or lower, and four of these occurred for median substitution 0.02 or greater. (Note that three of the five metals for which the bias difference was about -0.2 or less had the fewest labs in the study, i.e., 6 or 7.) This was a main reason to present the data for the model without median substitution.

When either the median substitution is not large relative to the target, or the number of labs is 10 or greater, there is considerable agreement between the substitution and no-substitution estimates. If we exclude the four labs with large median substitutions, and note that Al and Cr did not have any labs that required median substitution, then there were just two metals that had absolute value of bias difference larger than 0.1: Ag and Zn. Silver had seven substitutions of median value (though the median value was low (0.0053), a possible reason for this).

For elements with large median substitutions, there were large differences between substitution and no-substitution estimates of within and between lab RSDs. For other metals and metalloids, there was good agreement. Only Ag, Zn and Fe had differences that exceeded 0.11 in absolute value. Substitution gave between 0.2 and 0.3 larger RSD for the between component for Ag and Zn. The difference was larger for Fe (greater than 0.4). However, for Fe the within component was much smaller for the substitution estimate; the total RSD does not differ much for iron between the two data version, which suggests that the difference relates to where to assign the variability. For the within-lab RSD of other metals, only Ag has a difference of estimates that exceeds 0.11 in absolute value. Like Fe, the issue is again one of assigning a large total RSD to the two components.

In the context of robust estimation, the differences discussed above are relatively small. That there is substantial agreement between the substitution and no substitution methods is reassuring.

Simulation results

The sample size studied was 10 labs, in a study of simulated data that was meant to be similar to the observed relative bias data (B_i'). The median values of the robust RSD (= relative standard deviation of relative bias data) for between- and within-labs for the low spike levels were about 0.24 and 0.16, respectively. Simulations were carried out at these and higher levels for the standard deviations, with mean of zero. Situations were studied with biases of 0.10 for the between lab component and 0.20 for within lab, and vice versa. The bias was either a normal variable with standard deviation 1, or the absolute value of this variable (to get at one-sided influence of contamination, for example). The total RSD tended to be overestimated by about a factor of 1.25 and 1.5, respectively. Even if similar biases in the RSDs apply to the real data, multiplication of the sample results by a factor of 1/1.4 (to adjust for bias in total RSD) would still give larger variability than is desirable.

Example calculations

For an example of the robust statistical treatment described in ref. [2] as applied to reported ILS data, we consider low-level ICP-MS results for Be.[4] Table A(1) presents raw reported results for Be for media/reagent blanks and for low-spiked samples from the 20 participating laboratories. Laboratories reporting no method detection limit (MDL) cannot be used for blank-corrected data; also, laboratories reporting values <MDL for low-level spikes cannot be used either. This leaves eleven laboratories for further consideration, as shown in Table A(2). For laboratories reporting <MDLs for given entries, the value given by MDL/√2 was substituted. (For example: for Laboratory 1 the reported MDL was 0.008 µg Be, thus the substituted value = 0.008/1.414 = 0.0057 µg.) In Table A(2) the average of the blank level values (in µg) is subtracted from each of values reported for the low-level spikes. These values are then divided by the target value of 0.025 µg, and 1 is subtracted from the ratios to obtain the estimated biases (Table A(3)).

In the robust treatment the breakdown point value used was 0.50. The distance of each laboratory's data from the estimated mean, scaled by variance-covariance estimates, is computed. Those distances are then compared to a value which depends on the breakdown point, and weights are assigned to each lab's data. The weighted data are then used to compute robust estimates of the means of the three samples and of the variance-covariance matrix.

The overall estimate of bias for the blank-corrected low-level Be data after robust treatment is then the average of the three averages for the reported Values 1, 2 and 3 from Table A(3): (-0.201 – 0.154 – 0.213)/3 = -0.189. The RSD_b estimate is obtained by taking the square root of the average of the six off-diagonal elements of the covariance matrix: $[2\times(0.0623 + 0.0552 + 0.0604)/6]^{½}$ = 0.244. The estimate of RSD_w is then given by [(average of three diagonal elements of the variance matrix) – RSD_b]^{½} = 0.053. Finally $RSD_{tot} = (RSD_w^2 + RSD_b^2)^{½}$ = 0.25. 34.

References

1. S. Heritier, E. Cantoni, S. Copt and M. Victoria-Feser, *Robust Methods in Biostatistics*, Wiley, New York, 2009; p. 100.
2. www.r-project.org; Package "rrcov".
3. ASTM E691, *Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method*, ASTM International, West Conshohocken, PA, 2005; www.astm.org .
4. K. Ashley, M. J. Brisson, A. M. Howe and D. L. Bartley, *J. Occup. Environ. Hyg*., 2009, **6**, 745.

Fig. A1. Random effects. Top figure: all data; bottom figure: deviations of absolute value < 5.

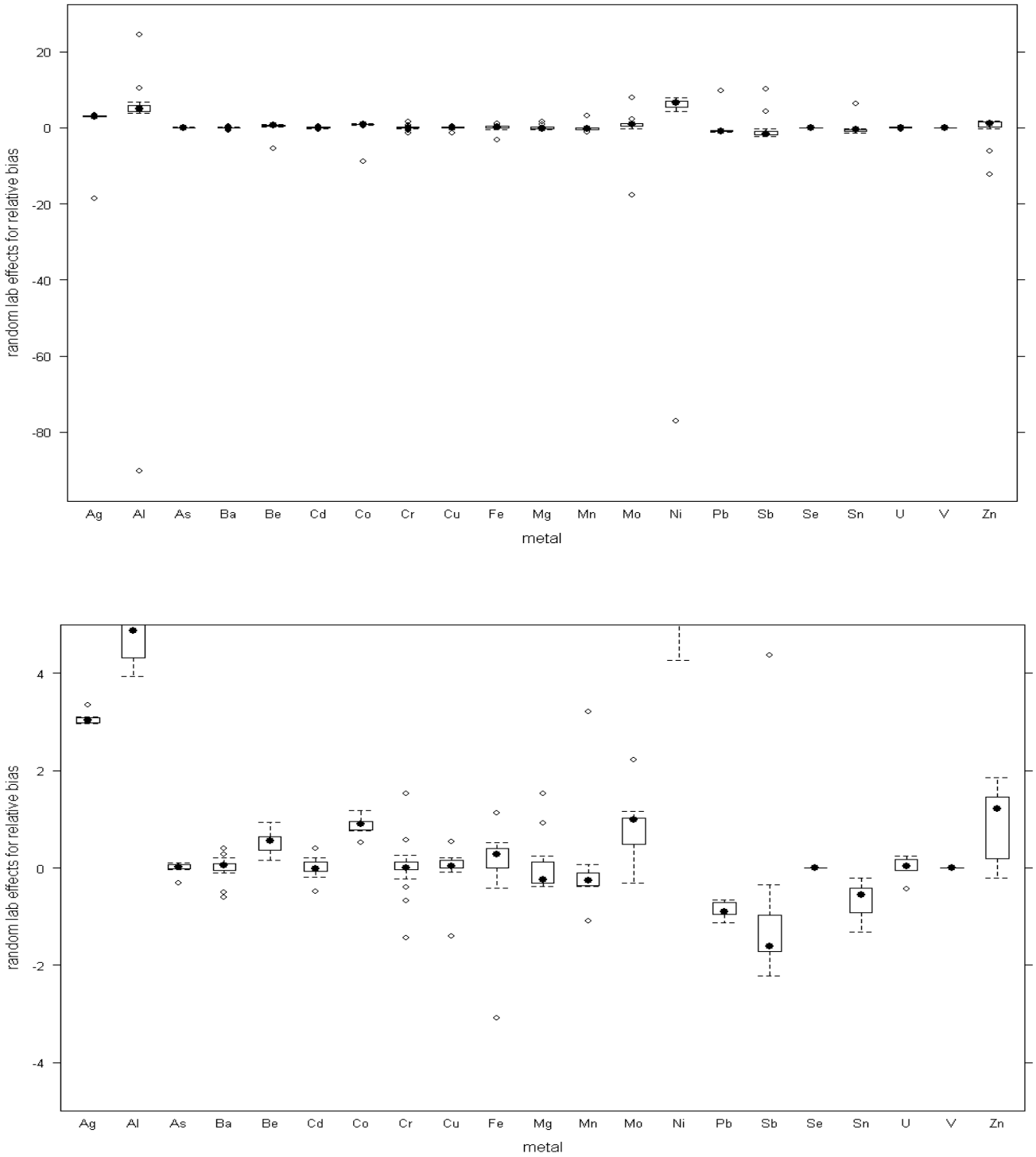
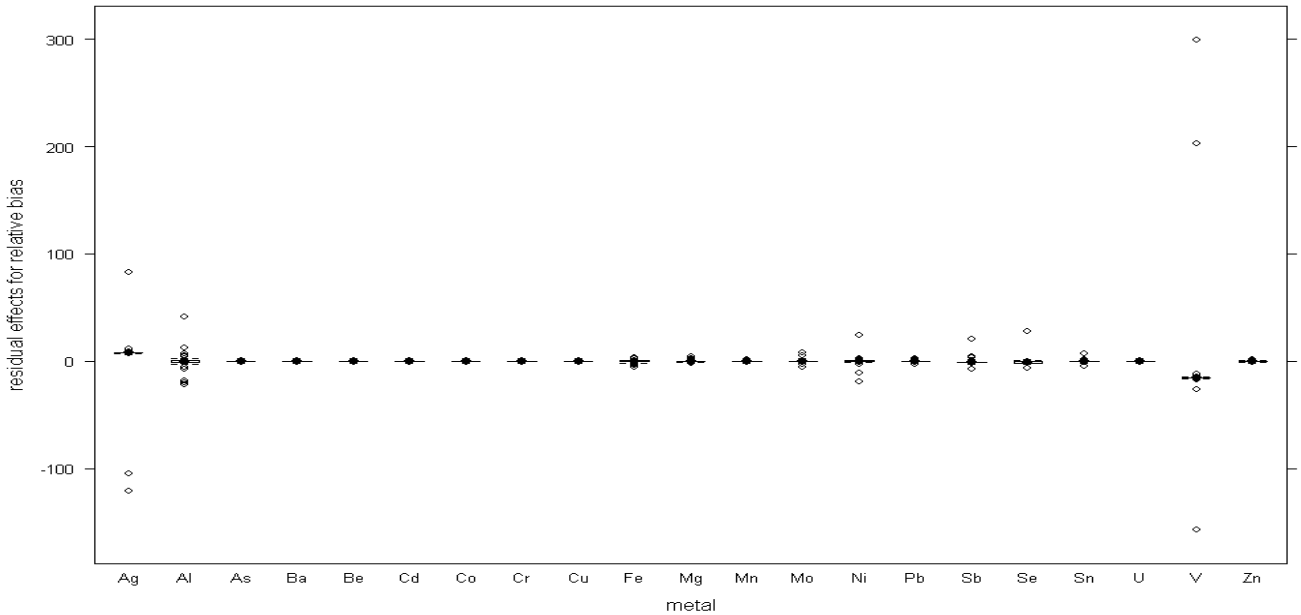


Fig. A2. Residual effects. Top figure: all data; bottom figure: deviations of absolute value < 5.



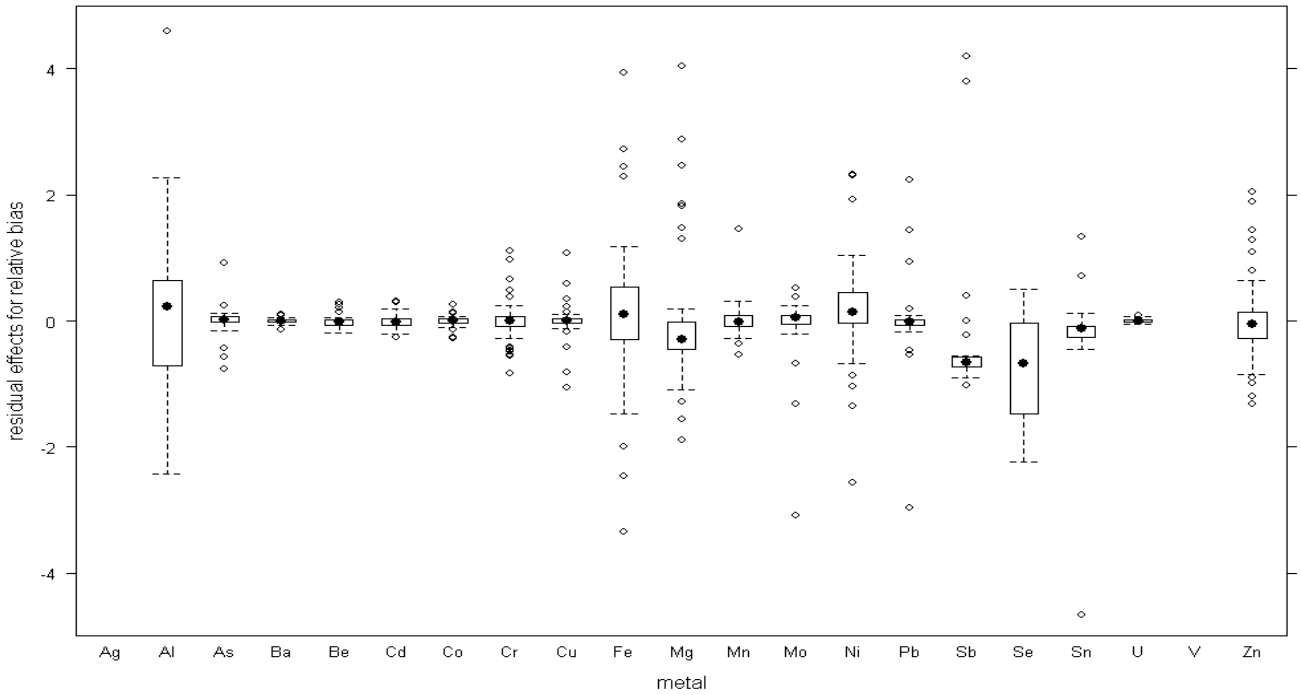


Fig. A3. Bias difference (median substitution – no substitution) vs. median value substituted for low-spike target value.

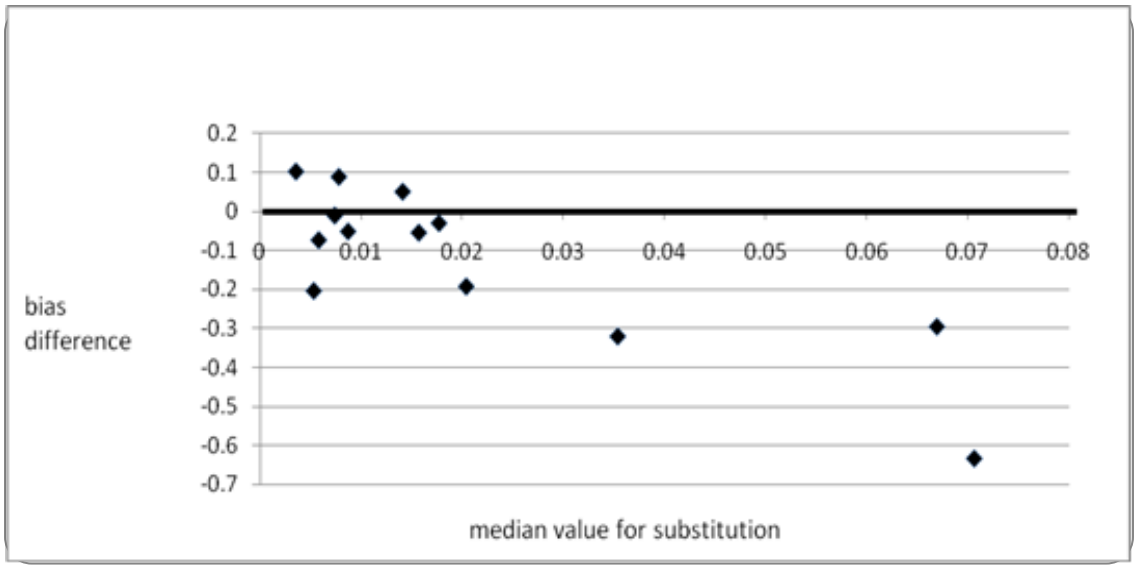


Table A(1) Raw reported ILS blank and low-level spike data for Be.

	Media blanks (µg Be sample^{-1})			Low-level spikes[a] (µg Be sample^{-1})		
Laboratory No.	**Value 1**	**Value 2**	**Value 3**	**Value 1**	**Value 2**	**Value 3**
1	<MDL[b]	<MDL	<MDL	0.0184	0.0177	0.0185
2	<MDL	<MDL	<MDL	0.0264	0.0278	0.0260
3	<MDL	<MDL	<MDL	<MDL	<MDL	<MDL
4	<MDL	<MDL	<MDL	<MDL	<MDL	<MDL
5	<MDL	<MDL	<MDL	0.03	0.02	0.02
6	NA[c]	NA	NA	0.023	0.022	0.023
7	0.0156	0.0243	0.0214	0.0414	0.0427	0.0409
8	<MDL	<MDL	<MDL	<MDL	<MDL	<MDL
9	NA	NA	NA	0.023	0.022	0.022
10	0.00036	0.00032	0.00016	0.0251	0.0253	0.0250
11	0.027	0.028	0.031	0.068	0.056	0.058
12	<MDL	<MDL	<MDL	0.025	0.030	0.025
13	<MDL	<MDL	<MDL	0.032	0.032	0.031
14	<MDL	<MDL	<MDL	0.021	0.022	0.021
15	<MDL	<MDL	<MDL	<MDL	<MDL	<MDL
16	<MDL	<MDL	<MDL	<MDL	<MDL	<MDL
17	NA	NA	NA	NA	NA	NA
18	0.001[d]	-0.001[d]	0.002[d]	0.030	0.024	0.020
19	NA	NA	NA	NA	NA	NA
20	0.48	0.05	0.04	0.06	0.06	0.07

[a] Be spike level = 0.025 µg filter^{-1}; [b] Value is below the laboratory's reported method detection limit (MDL); [c] No value reported, and laboratory-reported MDL available; [d] Laboratory reported blank-corrected values

Table A(2) Subset of low-level ILS Be data with <MDL values (if MDLs reported by laboratories) substituted values for blanks.

	Media blanks (µg Be sample^{-1})			Low-level spikes (µg Be sample^{-1})		
Laboratory No.	**Value 1**	**Value 2**	**Value 3**	**Value 1**	**Value 2**	**Value 3**
1	0.0057[a]	0.0057	0.0057	0.0184	0.0177	0.0185
2	0.0018	0.0018	0.0018	0.0264	0.0278	0.0260
5	0.0005	0.0005	0.0005	0.03	0.02	0.02
7	0.0156	0.0243	0.0214	0.0414	0.0427	0.0409
10	0.00036	0.00032	0.00016	0.0251	0.0253	0.0250
11	0.027	0.028	0.031	0.068	0.056	0.058
12	0.0042	0.0042	0.0042	0.025	0.030	0.025
13	0.0099	0.0099	0.0099	0.032	0.032	0.031
14	0.0071	0.0071	0.0071	0.021	0.022	0.021
18	0.001	-0.001	0.002	0.030	0.024	0.020
20	0.48	0.05	0.04	0.06	0.06	0.07

[a] For <MDL values, value of MDL/√2 is substituted;

Table A(3) Bias estimates for low-level ILS Be data after blank correction.

Bias Estimates			
Laboratory No.	**Value 1**	**Value 2**	**Value 3**
1	-0.49	-0.52	-0.49
2	-0.01	0.04	-0.03
5	0.18	-0.22	-0.22
7	-0.16	-0.11	-0.18
10	-0.01	0.0	-0.01
11	0.57	0.09	0.17
12	-0.17	0.03	-0.17
13	-0.12	-0.12	-0.16
14	-0.44	-0.40	-0.44
18	0.19	-0.05	-0.21
20	-6.2	-6.2	-5.8