RESULTS OF TEST 2 ON GROSS ERROR DETECTION OF ISP WG III/1 AND OEEPE*

Wolfgang Förstner
Photogrammetric Institute, Stuttgart University
Germany, Federal Republic

Commission III

Abstract: In 1980 an empirical test on gross error detection procedures was started by ISP WG III/1 and OEEPE. Using simulated data the aim was to get basic information about existing strategies for blunder detection in aerial triangulation and about the ability of these procedures to distinguish between small gross errors and systematic errors. The paper presents the results of the second phase of the test.

1. Introduction

"Only errors exist." Roberts' Axiom

The quality of block adjustment results highly depends on the ability of the procedure to detect gross errors especially small ones and to give information about the stability of the solution with respect to nondetected gross and systematic errors. Whereas the theory for handling systematic errors has reached a high and practicable standard which is proved by numereous controlled tests during the last decade there is no commonly accepted strategy nor theory for the blunder detection problem. This is due to the great variety of types of gross errors and the unability of the theory to predict the efficiency of even simple strategies in the presence of more than two or three gross errors.

During the XIIIth ISP congress in Hamburg it therefore was decided to start an empirical test on the efficiency of existing error detection procedures. In order to increase the possible number of participants it was decided to run the test as a joint programm of OEEPE* and WG III/1.

The scope of the test is twofold:

1. To find out the present status of strategies for gross error detection, especially to develop information on how efficiently large gross errors can be found.
2. To find out the sensitivity of existing error detection procedures to separate small gross errors on one hand and random and systematic errors on the other hand.

In order to achieve clear statements is was decided to split the test into two phases 1 and 2 resp.. In both phases several blocks are to be generated with errors which are only known to the distributer (Institute for Photogrammetry, Stuttgart University). These data are distributed to the participants who clean the blocks using their standard procedure.

This paper is intended to give a preliminary report on the results of phase 2 of the test. It also includes the summarized results of phase 1 (cf. Förstner, 1982). The report is based on the work of the following organizations:

---

* Organisation Européenne d'Études Photogrammétriques Expérimentales

	phase 1	phase 2	Organisation
1.	x		Surveyor General, Adelaide, Australia
2.	x	x	Lands & Survey Dept., Perth, Australia
3.	x	x	National Research Council, Ottawa, Canada
4.	x	x	Laboratoriet for Fotogrammetri og Landmaling, Aalborg, Denmark
5.	x		Hunting Survey Ltd., England
6.	x	x	Institute for Photogrammetry, Helsinki, Finland
7.	x		Fachgebiet Photogrammetrie und Kartographie, Darmstadt, FRG
8.	x		Institut für Angewandte Geodäsie, Frankfurt/M., FRG
9.	x		Lehrstuhl für Photogrammetrie, München, FRG
10.	x		Technische Hogeschool, Delft, The Netherlands
11.	x	x	Rijkswaterstaat, Delft, The Netherlands
12.	x	x	ITC, Enschede, The Netherlands
13.	x		Norges Geografiske Oppmaling, Hønefoss, Norway
14.	x		Universitet i Trondheim, Norway
15.	x	x	National Land Survey, Sweden
16.	x		Institut de Photogrammetrie, Lausanne, Suisse
17.	x		Lands & Survey Dept., New Zealand

Remark: Parts of this report are identical or at least overlap with the one given on phase 1 (cf. Förstner, 1982), as the datageneration and partially the scopes of both phases are similar. It was, however, decided to put up with some repetitions in order to keep this report readable for itself.

2. Design of test

Phase 2 was designed according to the following line of thought:

1. Control on detected gross errors

   In order to keep control on the detected gross errors, simulated data are used. Whereas the point distribution was chosen as realistic as possible random, systematic and gross errors were artificial.

2. Types of blocks

   Blocks with bundles and with independent models were generated, both with sparse and dense tie point distribution in order to simulate blocks for topographic mapping and for point densification. Thus 4 blocks BI/2, BII/2, MI/2 and MII/2 were generated.

3. Number of gross errors

   As in phase 1 the number of gross errors was chosen as high as it may occur in the worst case. In phase 2, however, only small gross errors were inserted, except for a few medium sized ones to simulate the situation during the last stage of the error detection procedure.

4. Documentaion of criteria for error detection

   The participants were asked to describe the criteria they used for rejecting observations in order to be able to compare the empirical and the theoretical efficiency.

5. Estimated size of gross errors

   The estimated size of the gross errors compared with the true size gives an indication whether the estimated size can be used for classification or even correction of the gross errors.

6. Accuracy

The accuracy of the cleaned blocks is a decisive check on the quality of the error detection procedure. Therefore the adjusted coordinates of all points are compared with the true coordinates, yielding the absolute accuracy in terms of a root mean square and a maximum error. The participants were also asked to tell how accurate they guess the result is in order to compare it with the empirical one. This part of the analysis is only finished for the bundle blocks up to now.

No attempt was made to compare the economy of the procedures as this topic already was discussed in phase 1.

3. Test performance

3.1 Data generation

The simulation of the data was based on the adjusted observations of a real bundle block. These data lead to error free coordinates of the new points and are used as a reference for the evaluation.

Two bundle blocks are generated by selecting appropriate points, leading to blocks BI/2 and BII/2 with sparse and dense tie point distribution. Similarily two model blocks MI/2 and MII/2 are generated by building independent models using a relative orientation. The blocks had 4 strips with 13 images and 12 models resp.. The sidelap was 20 %, the overlap 60 %. Thus the size of the blocks was 52 images and 48 models resp..

In contrary to phase 1 the true observations were contaminated by non-normally distributed random errors. Their distribution F was a mixture of two normal distributions N:

$$F = 0.95\ N(0,\sigma^2) + 0.05\ N(0,\ (2\sigma)^2\ ).$$

Thus on an average every 2oth observation was assumed to have double the standard deviation than the others. The standard deviation σ was constant for each block.

All images were deformed systematically using a combination of Brown's and Ebner's set of additional parameters. In contrary to phase 1 these deformations were not block invariant. Actually the additional parameters $\underline{p}_i$ were assumed to be random variables with constant, i. e. block invariant expectation $E(\underline{p}_i)$ and standard deviation σ_{p_i}. $E(\underline{p}_i)$ varied between 0 μm and 14 μm, σ_{p_i} between 0.5 μm and 1.5 μm. They were taken and adapted from the empirical results obtained by Schroth (1982).

On the other hand the coordinates of the model block MI/2 with sparse tie point density were <u>not</u> contaminated by systematic errors, in order to compare the efficiency of the error detection procedures more simply with theory. The systematic errors introduced into block MII/2 were constant for all models and consisted in a general deformation of 2nd degree.

The random and systematic errors were identical for all participants. Table 1 contains general information about the simulated data.

3.2 Inserted gross errors

The idea of phase 2 was to determine the efficiency of the practical procedures to detect small gross errors. The efficiency can be described by the probability of finding an error of a given size. As known from theory gross errors can only be found if they are larger than a certain lower bound $\nabla_o l_i$. This bound depends on the precision σ_{l_i} and the redundancy number r_i of the observation and on the statistical parameter δ_o which has been assumed to be 4 in this test. It corresponds to a critical value of appr. 3 and a minimum power , i.e. efficiency of 80 %. The size of the inserted gross errors is referred to the lower bound of the observation in concern.

Four types of errors were inserted into the blocks (cf. columns 2-4 in tables 2-5):

1. small gross errors in the photogrammetric data. Their size varied between 0.7 $\nabla_o l_i$ and 2 $\nabla_o l_i$. Always 4-7 errors of the same type were generated to be able to estimate the empirical efficiency. Due to the different local redundancy the actual size of the errors in µm varied within each group. The model block MII/2 with dense tie point distribution gave the oportunity for groups of gross errors which are treated as one error in the analysis. The dense bundle block BII/2, however, was mainly distorted with single gross errors as even adjacent points within one image do not really control each other.
2. small gross errors in the control points. Their size varied between 1 $\nabla_o l_i$ and 4 $\nabla_o l_i$.
3. medium sized gross errors up to 150 $\nabla_o l_i$. They partly were supposed to be correctable.
4. miscellaneous errors such as point exchanges or grouped errors.

Types 3 and 4 were meant to keep the data realistic.

4. Results

4.1 General information

The result is based on 18 blocks; 5 MI/2, 5 MII/2, 4 BI/2 and 4 BII/2. The used programms may be subdivided into the following catagories:

a. independent models

- check of model connections only
- rigorous strip adjustment only
- iterative least squares block adjustment (planimetry-height)
- rigorous block adjustment

Only one programm compensated for systematic errors using an analysis of the residuals. Only one programm did not use data snooping technique. One programm includes an automatic procedure for data cleaning adapting the weights to the residuals of the previous iteration (cf. table 6a).

b. bundles

All programms have applied selfcalibration with 9-12 additional parameters. Two programms have used the facility of data snooping technique. One used an automatic error detection procedure (cf. table 6b).

4.2 Detected gross errors

Tables 2-5 show the reaction of the individual participants on each gross error introduced into the data. Specifically the estimated size of the error and the evaluation of the response is given.

In all cases, except those denoted by the minus sign "-", it is assumed that the error has been found. Scanning for the different symbols the following statements can be made:

\# the error has been found, correctly located and corrected. Only few small gross errors have been corrected.

● the error has been found, correctly located but corrected by a wrong amount.
Quite some errors have been introduced by the participants using a weak estimation of the size of the gross errors. A comparizon of the estimated and the true sizes of the errors clearly shows that there is no real chance to get a reliable basis for error correction in case of small gross errors. This is in full agreement with the theory as the standard deviation of the estimated size $\widehat{\nabla l}_i = -v_i / r_i$ is

$$\sigma_{\widehat{\nabla l}_i} = \sigma_{l_i} / \sqrt{r_i}$$

Thus the standard deviation never is better than the precision of the observation, on an average it is 2-3 times larger. Only for medium sized gross errors the relative accuracy of the estimated size is high enough to draw reliable conclusions.

\+ the error has been found, correctly located and the observation(s) have been eliminated.
This sign only is given for reactions which are correct and justified, i. e. if there is a reason for the decision. Wrong tie points lying in 2 models or tie points with a non acceptable x-parallax lying in 3 images had to be eliminated completely. Otherwise the reaction was correct by chance (→ O), used the non-ideal geometry of the strip (→ +!) or was made arbitrarily (→ O^1) being aware of possibly commit an erroneous decision.

O the error has been found but not correctly located.
This situation mainly occured at the above mentioned points in 2 models or 3 images but also at control points.

? in this case the reaction was not quite clearly described. For the analysis it was assumed that this error has been found.

x only one of the errors within a group of errors has been found.

There are quite some gross errors where all participants reacted the same way. Thus, with respect to these errors the strategies, namely testing the residuals or applying a rigorous test, are equivalent. On the other hand there are gross errors which have not been found by one participant but correctly located by another.A comparison of the reactions to the errors in the model blocks suggests strategies 4 and 5 to be superior to the others while mutually showing minor differences. Concerning the model blocks, however, participant 4 was the only one who did not use a statistical test. This suggests data snooping in pratice not beeing as effective as to be expected from theory and contradicts the results of phase 1 where strategies with data snoopung showed to be superior to strategies without rigorous test. The same observations can be made concerning the bundle blocks.

This statement is also proved by the analysis of the performance index I_p, which has been used in phase 1 for the evaluation of the efficiency of the procedures. It is the number of wrong decisions, namely the number n_c of deleted but correct observations and the number n_m of the gross errors not found by the procedures (cf. tables 2-5). As the gross errors were small and wrong decisions might have only little influence on the final result we will not discuss the performance index but rather analyse the power of the tests and the obtained absolute accuracy.

4.3 Efficiency

The efficiency of the procedures is estimated from the results listed in tables 2-5. Table 6 contains the probabilities with which the gross errors of the different sizes have been found by the participants. The extreme values (min, max) and the average values (Ø) of these probabilities are shown in figure 1.

The efficiency or power of a test depends on the size ∇l_i of the gross error and is set into relation to the lower bound $\nabla_o l_i$. Theoretically gross errors of this size can be found with a probability of approximately 80 % if a statistical test with a critical value of 3.3 (corresponding to a significance level of 99.9 %) is used. The probability for detecting larger errors increases, smaller errors can be found with a lower probability. In figure 1 the theoretical efficiency is represented by the smooth curve.

The comparison of the empirical and the theoretical efficiency for model block MI/2 shows that the power of the practical error detection procedures can be predicted quite reliably. The maximum values for the efficiency are not reached by the same participant (namely 4 and 5, cf. sect. 4.2). The minimum values are reached by the strip adjustment control.

The proximity of the empirical findings to the theoretical values is mainly due to the absence of any systematic errors in MI/2. This is proved by the results of model block MII/2 where the same procedures were used to clean a block with systematic errors but where no selfcalibration was applied in all cases.

The results of the bundle blocks confirm this as the empirical and the theoretical efficiency do not differ so much as for MII/2. The empirical and the theoretical curves have a similar shape. The selfcalibration applied in all 4 cases obviously was capable to compensate parts of the systematic errors. Remember, that the systematic errors are varying from image to image leaving at least rests of the image deformations in the data. The results from BII/2 are closer to the theoretical expectation than those of BI/2 probably because of the higher stability of the block BII/2 which allows a more reliable determination of the additional parameters.

If one would use an average of the estimated $\hat{\sigma}_o$-values instead of the true value for determining the theoretical efficiency the difference to the empirical efficiency would become negligible, suggesting that the uncompensated systematic errors are the main source for the reduced power of the error detection procedures. Again, there seems to be no significant difference between those procedures which use a rigorous test and those which rely in the analysis of the residuals.

4.4 Absolute accuracy

The absolute accuracy of the cleaned blocks can be determined by comparing the adjusted with the true coordinates. Table 7 contains the maximum ε and the r.m.s. errors μ for the bundle blocks including also the estimates $\hat{\mu}$ for the accuracy and the empirical precision $\hat{\sigma}_o$ of the image coordinates provided by the participants.

The maximum errors obtained by the 4 participants for the bundle blocks are nearly identical suggesting the results to be of similar quality. This actually is true for the planimetry (μ_{xy}) but not for the heights (μ_z), especially for BI/2 where the r.m.s. errors vary up to a factor 3. (The columns in table 7 are sorted according to the achieved absolute accuracy.) The absolute values are also very high if one takes the scales 1 : 15 000 and 1 : 3 000 for BI/2 and BII/2 resp. and the precision (σ_o) into account. This is confirmed by the optimistic estimates $\hat{\mu}$ for the absolute accuracy given by the participants. The actual r.m.s.e. prove to be at least a factor 2 (up to a factor 10) larger than presumed. The maximum discrepancies on an average are also larger than one would expect from pure error propagation ranging up to 6 times the r.m.s.e. values $\hat{\mu}$. They are due to undetected small gross errors, specifically errors in the x-coordinates of points lying in only three images. This proves the external reliability measures ($\delta_o \cdot \sigma_x$) to be a useful approximation for the maximum error in the result of an adjustment.

5. Conclusions

The result of this empirical test gives a clear picture of the status of existing error detection procedures. The classical procedures which rely solely on the analysis of the residuals without taking into account the local geometry of the block are capable to produce coordinates of a high quality provided the data cleaning is done by skilled personnel and with care. The implementation of rigorous tests is helpful for the localization of even medium sized gross errors and in extreme cases for the detection of gross errors which are hardly larger than the lower bounds for detectable errors. There is, however, no evidence that in photogrammetric blocks with a more or less regular shape rigorous tests would improve the results significantly nor that statistical methods might be a surrogate for experience. Only for irregular shaped blocks with weak control the local geometry must be taken into account, which then will of course document the instability of the block.

The following recommendations can be derived from the analysis of the test:

- Pre-adjustment error detection procedures are necessary to grasp large gross errors. On-line procedures, strip formation or automatic checks of conditions may be used to advantage. The separate checking of photogrammetric observations and ground control is recommendable at this stage (1).
- Automatic elimination of observations reduces the number of runs considerably. Weighting down bad observations seems to be the appropriate way, as erroneously deleted correct observations are reintroduced automatically into the adjustment. The weighting may be based on the residuals, if the local redundancy is not available (phase 1).
- A statistical test (e. g. data-snooping) in general leads to the best results with respect to the localization of medium sized gross errors and the detection of small gross errors. A plot of the residuals then does not seem to be necessary (phase 1)
- Observations only should be corrected if the error can be identified based on listings etc. or if point numbers are evidently wrong. The size of a gross error estimated from an adjustment is inaccurate up to several times the standard deviation of the observation and therefore only can be used for the classification of medium sized gross errors (larger 1 mm at image scale; phases 1 and 2).

- If gross errors are <u>not locatable</u> the whole point should be taken out of the adjustment or appropriately renumbered in order to identify the point to be unreliable (phases 1 and 2).
- The <u>efficiency</u> of the tests with respect to small gross errors can reliably be predicted by theory. Unmodelled or uncompensated systematic errors seem to be the main effect reducing the efficiency of the error detection by increasing the estimate of the variance factor $\hat{\sigma}_0^2$ (phase 2)
- The <u>absolute accuracy</u> derived from error propagation using the estimated varinace factor usually is too optimistic. A correction factor of 2 seems to be realistic. The maximum errors can be reliably estimated from the external reliability of the block (phase 2).

Acknowledgements: The author wishes to thank all participants in the test for their great support and especially the patience in dealing with the artificial data. Without their contribution this test would have not been possible.

"One man's errors is another man's data." Berman's Corollary to Roberts' Axiom

Literature:

Bloch, A. (1982): Murphy's Law Book Two - more reasons why things go ƃuoɹʍ Price/Stern/Sloan Publishers, Inc. Los Angelos, 1982

Förstner, W. (1982): Results of Test I on Gross Error Detection of ISP WG III/1 and OEEPE, Pres. Paper to ISP Comm. III Symposium, Helsinki, 1982

Schroth, R. (1982) On the Stochastical Properties of Image Coordinates, Pres. Paper to ISP Comm. III Symposium, Helsinki, 1982

Table 1: Generated blocks of phase 2

	MI/2	MII/2	BI/2	BII/2
n^1	1421	2844	1227	2739
u	909	1641	887	1608
r	512	1203	340	1131
r/n	0.36	0.42	0.28	0.41
Scale	1:20 000	1:8 000	1:15 000	1:3 000
No. units	48	48	52	52
σ_{oHO}[m]	0.20	0.05	0.05	0.015
σ_{oVE}[m]	0.40	0.12	0.01	0.05
σ_o[µm]	-	-	4	2
σ_{oxy}[µm]	8	4	-	-
σ_{oz}[µm]	12	7	-	-
No. gross errors	34	36	34	35
% of wrong points	2.8%	2.3%	4.3%	2.5%
Systematic errors	no	const.	variable	variable

n=no. of observations, u=no. of unknowns, r=redundancy
σ_{oHO}, σ_{oVE}, σ_o, σ_{oxy}, σ_{oz} standard deviations of horizontal and vertical control and of photogrammetric observations

Table 2: Reactions on errors in block MI/2

	ERROR NO.	TYPE	SIZE	PARTICIPANT 1 [1]	2 [2]	3	4	5
1	6	0.7	60 y	-	-	-	-	-
2	7	0.7	90 z	-	-	-	-	-
3	8	0.7	40 y	-	-	-	-	-
4	9	0.7	45 x	-	-	-	-	-
5	10	1.0	90 x	-	-	-	-	100 ●
6	11	1.0	90 z	130 +	+	110 +	100 0	100 ●
7	12	1.0	130 z	-	-	-	-	-
8	13	1.0	95 y	-	-	-	-	80 +
9	14	1.0	50 x	-	-	-	50 0	50 +
10	15	1.0	90 z	-	-	82 +	100 +	80 +
11	16	1.0	60 x	-	-	-	-	-
12	17	1.3	110 z	130 +	+	134 +	150 0	140 +
13	18	1.3	80 y	90 0	-	77 +	80 #	80 +
14	19	1.3	75 y	70 0	-	82 +	75 +	80 +
15	20	1.3	90 x	-	-	112 +	75 +	80 +
16	21	1.3	160 z	160 +	+	164 +	160 #	170 +
17	22	1.3	75 y	-	-	-	50 0	-
18	23	1.6	140 x	-	-	-	-	-
19	24	1.6	150 z	160 0	0	150 +	100 +	160 +
20	25	1.6	90 y	90 +	-	112 +	100 0	100 ●
21	26	1.6	90 x	75 +	-	-	50 0	80 +
22	27	1.6	250 z	240 +	+	262 +	250 Ø	230 +
23	31	2 HO	2.8 X	-	-	-	3. +	-
24	32	2 VE	6.5 Z	-	-	5. ●	10. +	7. +
25	33	4 HO	5.0 Y	-	-	-	5. +	6. +
26	34	4 VE	12.0 Z	-	-	10. ●	10. ●	10. ●
27	28	20	1000 x	1000 +	+	1000 +	1000 +	1000 #
28	29	50	5000 y	5000 +	+	5000 +	5000 +	5000 #
29	30	150	20000 z	20000 +	?	20000 +	20000 #	20000 #
30	1	ex	-	0	#	#	#	#
31	2	ex	-	0	-	+	#	#
32	3	ex	-	0	0	#	#	#
33	4	gr	100 z	100 B	-	-	x	x
34	5	gr	80 x	-	-	-	-	-
no. of deleted correct observ. n_c				0	3	4	5	1
no. of missed gross errors n_m				18	25	16	10	10
Performance index $I_p = n_c + n_m$				18	28	20	15	11

[1] only check of model connections [2] only strip adjustment

Explanation to Tables 2-5

Type:
- number gives size of error in units of $\nabla_o l_i$ ($\delta_o = 4$)
- HO = horizontal control point
  VE = vertical control point
  ex = exchange of point numbers
  gr = group of errors
  S = single observation (default)
  D = double, two adjacent observations
  T = triple, three adjacent observations
  PC = projection center

Size:
- size in µm or m
- coordinate (x ≙ flight direction)

Participant:
- estimated size of errors in µm or m
- code

Table 3: Reactions on errors in block MII/2

	ERROR NO.	TYPE	SIZE	PARTICIPANT 1[1]	2[2]	3	4	5
1	4	0.7	20 y	-	-	-	-	-
2	5	0.7	40 z	-	-	-	-	-
3	6	0.7	25 y	-	-	-	-	-
4	7	0.7 D	40 x	-	-	-	-	-
5	8	0.7 D	40 x	-	-	-	-	-
6	9	0.7 D	45 z	55 0	-	-	+	-
7	10	1.0	50 z	70 +	+	-	+	-
8	11	1.0	25 y	-	-	-	-	-
9	12	1.0	22 x	-	-	-	-	-
10	13	1.0 PC	170 x	-	-	-	-	170 +
11	14	1.0 D	60 z	-	-	-	-	-
12	15	1.0 D	30 x	-	+	-	-	-
13	16	1.3	60 z	-	-	-	-	-
14	17	1.3	40 y	-	-	+	45 ?	-
15	18	1.3	30 x	-	-	-	-	-
16	19	1.3	30 x	-	-	-	-	-
17	20	1.3 D	75 y	-	-	●	-	-
18	21	1.3 D	35 x	-	-	-	+	-
19	22	1.6	70 z	80 0	-	-	0	-
20	23	1.6	35 x	-	-	-	0	40 +
21	24	1.6	40 y	55 0	-	+	35 ?	50 +
22	25	1.6 PC	230 x	-	-	+	-	210 +
23	26	1.6 D	40 x	-	-	-	-	-
24	27	1.6 D	70 z	70 0	-	-	0	-
25	28	2	50 y	+	-	-	0	60 +
26	29	2 PC	250 x	-	-	-	-	-
27	30	2 T	120 z	80 0	-	-	-	-
28	33	2 HO	2.0 y	-	-	-	0	-
29	34	2 VE	1.4 z	-	-	-	+	1.4 +
30	35	4 HO	2.0 x	-	-	0.6 0	+	0.6 +
31	36	4 VE	3.0 z	-	-	●	+	3.6 +
32	31	6 D	250 z	135 0	0	250 +	+	260 +
33	32	50 D	1500 x	-	-	+	+	1500 +
34	1	ex		0	0	#	#	#
35	2	ex		#	#	#	#	#
36	3	ex		0	0	#	?	#
no. of deleted correct observ. n_c				2	0	8	2	1
no. of missed gross errors n_m				25	30	25	18	23
Performance index $I_p = n_c + n_m$				27	30	33	20	24

[1] only check of model connections [2] only strip adjustment

Table 4: Reactions on errors in block BI/2

	ERROR NO.	TYPE	SIZE	PARTICIPANT 1	2	3	4
1	7	1.0	20 y	50 +	42 +	40 +	40 +
2	8	1.0	35 x	-	-	-	40 0
3	9	1.0	20 x	-	-	-	-
4	10	1.0	20 x	?	-	-	-
5	11	1.0	35 y	-	-	-	-
6	12	1.0	30 x	-	-	-	-
7	13	1.0	20 y	-	-	-	-
8	14	1.3	24 y	80 +	-	-	40 +
9	15	1.3	40 x	-	-	-	-
10	16	1.3	28 x	-	-	-	-
11	17	1.3	70 x	-	-	-	-
12	18	1.3	30 y	65 +	-	-	30 +
13	19	1.6	28 y	50 +	-	-	50 +
14	20	1.6	40 y	-	-	-	30 0
15	21	1.6	52 x	130 +!	-	75 0[1]	40 +!
16	22	1.6	60 x	-	-	-	(70)0
17	23	2.0	60 x	50 +	39 +	80 0[1]	60 +!
18	24	2.0	90 y	80 +	90 +	100 +	100 +
19	25	2.0	45 y	80 +	-	-	60 +
20	26	2.0	60 x	50 +	58 +	50 +	60 +
21	5	2.0	40 x	50 +	45 +	(50)0[1]	60 0
22	6	2.0	70 x	50 +	-	75 0[1]	(80)0
23	27	2.0 VE	1.8 Z	4.0 ●	-	-	2.4 +
24	28	2.0 HO	1.2 X	.8 ●	-	-	-
25	29	4.0 VE	3.0 Z	-	3.0 #	2.0 +	3.0 +
26	30	4.0 HO	1.5 Y	1.0 ●	-	-	1.0 +
27	31	6	150 x	300 +	155 +	150 +	150 #
28	32	18	500 y	1000 +	488 +	500 +	500 #
29	33	50	1000 x	1086 0	525 +	1000 0[1]	(1000)0
30	34	150	4000 y	4000 +	4000 #	4000 +	4000 #
31	1	ex	-	#	#	#	#
32	2	ex	-	#	#	#	#
33	3	gr	50 x	-	-	-	-
34	4	gr	100 x	65 x	50 x	100 x	(120)x
no. of deleted correct observ. n_c				7	0	3	6
no. of missed gross errors n_m				12	21	19	11
Performance index $I_p = n_c + n_m$				19	21	22	17

[1] being aware to possibly reject an error free observation

Table 5: Reactions on errors in block BII/2

	ERROR NO.	TYPE	SIZE	PARTICIPANT 1	2	3	4
1	7	.7	10 x	45 0	-	-	-
2	8	.7	10 x	-	-	-	-
3	9	.7	11 x	-	-	-	-
4	10	.7	8 x	-	-	-	-
5	11	1.0	13 x	-	-	-	-
6	12	1.0	12 y	-	-	-	-
7	13	1.0	10 y	70 +	-	-	-
8	14	1.0	14 y	-	-	-	-
9	15	1.3	20 y	20 +	-	20 +	20 +
10	16	1.3	30 x	+	22 +	35 0[1]	(40)0
11	17	1.3	18 y	30 +	22 +	-	20 +
12	18	1.3	25 y	-	25 0	20 +	-
13	19	1.3	20 y	40 +	33 +	30 +	30 +
14	20	1.6	36 x	(50)0	-	45 0[1]	(40)0
15	21	1.6	24 y	-	35 +	35 +	30 +
16	22	1.6	20 y	50 +	21 +	21 +	20 +
17	23	1.6	22 x	20 +	15 +	20 +	30 +
18	24	1.0 VE	.14 Z	●	0	-	-
19	25	1.5 VE	.12 Z	●	.26 +	.25 +	.23 +
20	26	2.0 VE gr	.30 Z	-	-	-	-
21	27	1.5 VE	.12 Z	-	-	-	-
22	28	1.0 HO	.08 X	●	.12 +	.11 +	.12 +
23	29	2.0 HO	.30 Y	.30 #	.27 +	.25 +	.28 +
24	30	1.5 HO gr	.12 X/Y	-	-	-	-
25	31	2.	50 x/y	-	34 0!	-	-
26	32	6.	120 x	100 +	(118)+	100 0[1]	100 0
27	33	18.	400 x	400 +	(402)+	400 0[1]	400 #
28	34	50.	1414 x/y	? +	1400 +	? +	1414 #
29	35	150.	3000 y	3000 #	3000 +	3000 +	3000 #
30	1	ex	-	#	#	+	#
31	2	ex	-	#	#	#	#
32	3	gr	15 y	-	-	18 +	-
33	4	gr	20 x	-	-	-	-
34	5	gr	20 y	-	-	-	-
35	6	gr	45 x	-	-	-	-
no. of deleted correct observations n_c				23	2	0	1
no. of missed gross errors n_m				16	17	17	18
Performance index $I_p = n_c + n_m$				39	19	17	19

[1] being aware of possibly rejecting a correct observation

Table 6: Empirical efficiency and features of error detection procedures

a) Model Blocks

	MI/2								MII/2							
SIZE	1	2	3	4	5	MIN	φ	MAX	1	2	3	4	5	MIN	φ	MAX
0.7	0	0	0	0	0	0	0	0	.17	0	0	.17	0	0	.06	.17
1.0	.14	.14	.29	.43	.71	.14	.34	.71	.17	.33	0	.17	.17	0	.17	.33
1.3	.67	.33	.83	1.0	.83	.33	.73	1.0	0	0	.33	.33	0	0	.13	.33
1.6	.80	.40	.60	.80	.80	.40	.68	.80	.50	0	.33	.67	.50	0	.40	.67
2.0	-	-	-	-	-	-	-	-	.67	0	0	.33	.33	0	.27	.67
D	+	+	+	-	+				+	+	+	-	+			
S	-	-	-	-	-				-	-	-	-	-			
A	-	-	+	-	-				-	-	+	-	-			

b) Bundle Blocks

	BI/2							BII/2						
SIZE	1	2	3	4	MIN	φ	MAX	1	2	3	4	MIN	φ	MAX
0.7	-	-	-	-	-	-	-	.25	0	0	0	0	.08	.25
1.0	.29	.14	.14	.29	.14	.21	.29	.25	0	0	0	0	.08	.25
1.3	.40	0	0	.40	0	.20	.40	.80	.80	.80	.80	.80	.80	.80
1.6	.50	0	.25	1.0	0	.44	1.0	.75	.75	1.0	1.0	.75	.88	1.0
2.0	1.0	.67	.83	1.0	.67	.87	1.0	-	-	-	-	-	-	-
D	+	-	+	-				+	-	+	-			
S	+	+	+	+				+	+	+	+			
A	-	+	-	-				-	+	-	-			

D: data-snooping
S: self calibration
A: automatic weight reduction

Table 7 Absolute accuracy of cleaned bundle blocks

	BLOCK BI/2				BLOCK BII/2			
	1	2	3	4	1	2	3	4
Data-Snooping	-	-	D	D	D	-	D	-
Points Not Listed	6	7	5	10	7	18	9	13
Points Compared	159	158	160	155	369	358	367	363
ε_{xy} max [m]	1.43	1.41	1.75	1.44	.14	.19	.17	.15
ε_{z} max [m]	2.71	2.61	2.62	3.04	.36	.40	.38	.29
μ_{xy} [m]	.27	.36	.39	.37	.038	.040	.045	.043
μ_{z} [m]	.44	.56	.64	1.22	.076	.072	.083	.108
ε_{xy} max / μ_{xy}	5.3	3.9	4.4	3.9	3.7	4.8	3.8	3.5
ε_{z} max / μ_{z}	6.2	4.7	4.1	2.5	4.7	5.6	4.6	2.7
$\hat{\mu}_{xy}$ [m]	-	.18	.13	.05	.012	-	.005	.025
$\hat{\mu}_{z}$ [m]	-	-	.28	.10	.027	-	.010	.060
$\hat{\mu}_{xy} / \mu_{xy}$	-	.50	.33	.14	.32	-	.11	.48
$\hat{\sigma}_{o}$ [µm]	5.7	7.9	6.5	3.5	3.3	3.5	2.7	3.8
$\hat{\sigma}_{o} / \sigma_{o}$	1.4	2.0	1.6	0.9	1.6	1.7	1.3	1.9

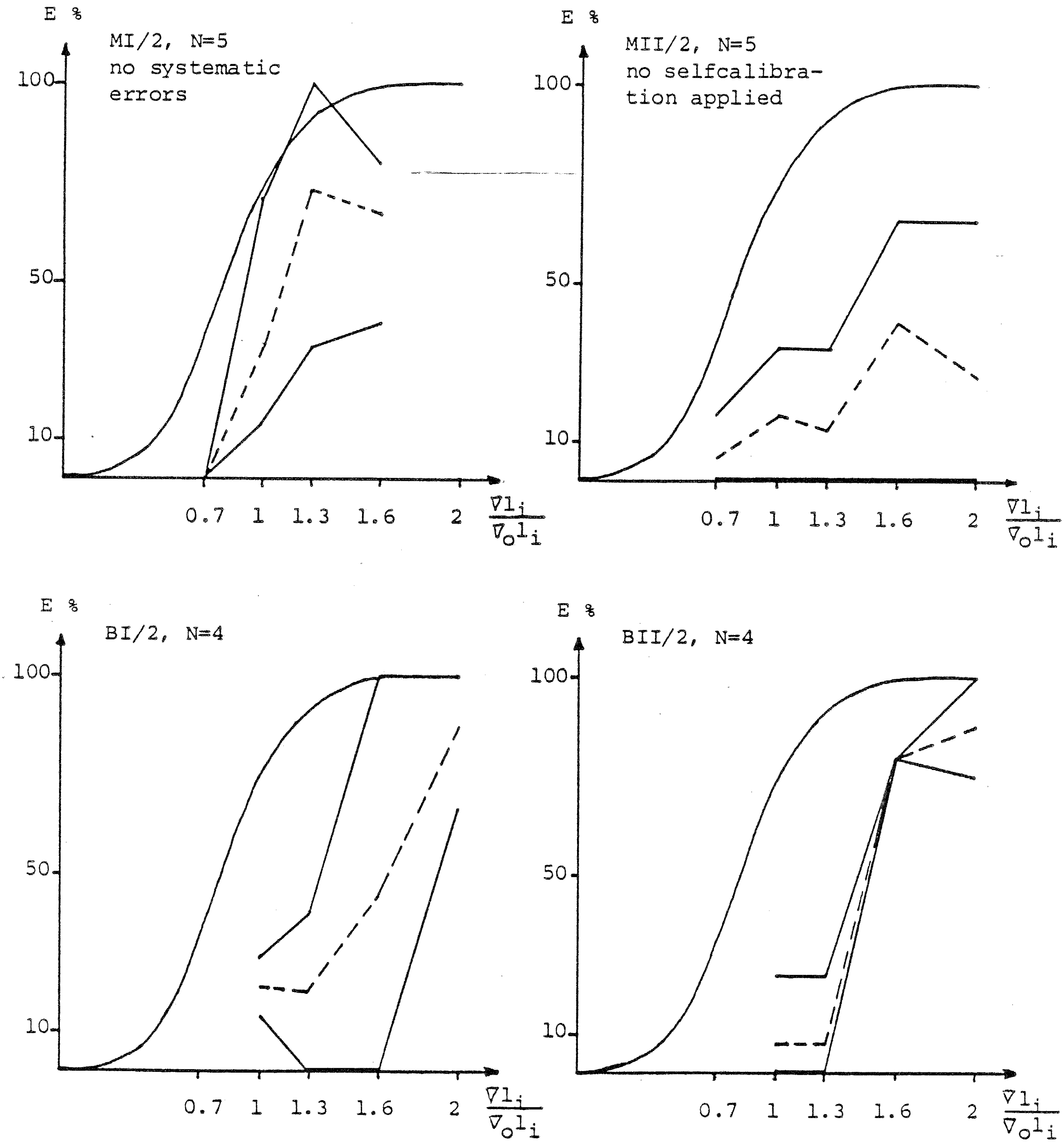

Fig.1 Empirical and theoretical efficiency of practical procedures (cf. table 6)

Extreme (——) and mean (----) values for probability E (%) of error detection

Curve = theoretical efficiency