Reports of the Department of Geodetic Science

Report No. 290

ERROR ANALYSIS FOR A SPACEBORNE LASER RANGING SYSTEM

by

Erricos C. Pavlis

(NASA-CR-162887) ERROR ANALYSIS FOR A SPACEBORNE LASER RANGING SYSTEM (Ohio State Univ. Research Foundation) 89 p HC A05/MF A01 CSCL 20E

N80-20575

Unclas

G3/36 47574

Prepared for

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20770

Grant No. NSG 5265
OSURF Project 711055

OSU

The Ohio State University
Research Foundation
Columbus, Ohio 43212

December, 1979

Reports of the Department of Geodetic Science

Report No. 290

ERROR ANALYSIS FOR A SPACEBORNE LASER RANGING SYSTEM

by

Erricos C. Pavlis

Prepared for

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20770

Grant No. NSG 5265
OSURF Project 711055

The Ohio State University
Research Foundation
Columbus, Ohio 43212

December, 1979

PREFACE

This project is under the supervision of Professor Ivan I. Mueller, Department of Geodetic Science, The Ohio State University, and under the technical direction of Dr. David E. Smith, Code 921, NASA, Goddard Space Flight Center, Greenbelt, Maryland 20771

ABSTRACT

The dependence (or independence) of baseline accuracies, obtained from a typical mission of a spaceborne ranging system, on several factors is investigated. The emphasis is placed on a priori station information, but factors such as the elevation cut-off angle, the geometry of the network, the mean orbital height, and to a limited extent geopotential modeling are also examined.

The results are obtained through simulations, but effort has been made to give some theoretical justification whenever possible. Guidelines for freeing the results from these dependencies are suggested for most of the factors.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor I.I. Mueller who encouraged me to perform this investigation and guided me through his lectures and valuable discussions. I would also like to thank Professor R.H. Rapp for his comments and suggestions on the original manuscript. My appreciation extends to Professor U.A. Uotila for his enlightening discussions on estimation theory. I am greatly indebted to Professor C.R. Rao for allowing me to participate in his lectures and seminars at OSU during the past academic year.

The excellent typing and editing of the report by Mrs. Rene Tesfai is greatly appreciated.

Last but not least my sincerest thanks to my family for the unconditional support at all times.

TABLE OF CONTENTS

1. INTRODUCTION

In the past decade or so laser ranging to artificial satellites proved to be one of the most precise and efficient tools for geodetic positioning. The benefits from its use were realized early enough to encourage further development of the hardware as well as extensive application to problems related to earth dynamics. One of the areas where this system will be of major importance is plate tectonics. Almost half a century after Alfred Wegener published his continental drift theory [Wegener, 1928], space geodesy provided scientists with a sound tool for measuring the relative motion of the continental plates. As our knowledge and understanding of geophysical phenomena related to plate tectonics improved, it became apparent that there exists a high correlation between the location of the plate boundaries and earthquake epicenters. Further, it has been the conviction of several scientists that geophysical activity in the region of a fault contains vital information about the actual occurrence of earthquakes. It is therefore highly desirable to be able to monitor such activities (dilatancy, strain accumulation, tilt, etc.) as they are related to seismic hazards. The regional aspect of plate tectonics, therefore, is mainly concerned with the deformation of the plates near their boundaries at the fault zones. The best way of determining this deformation is monitoring the motion of several benchmarks located near the fault relative to that of points significantly away from it. Since a fault zone can be of quite large extent, in order to be able to deduce meaningful results a large number of points is required and quick, frequent resurveying of the area.

A system that can meet all the requirements and still be cost effective is a Satellite Ranging System (SRS). The idea behind this system is the inversion of the traditional satellite trilateration scheme. Due to the large number of points whose positions must be determined, the active (and expensive) station, e.g., the laser, is placed <u>on</u> the spacecraft and the ground stations are targeted with relatively cheap reflectors. The advantages of this scheme in terms of cost are obvious.

Historical Review

The idea of the "upside-down" laser was first suggested in the late '60's at the time NASA undertook the task of improving the existing hardware so that higher accuracies could be achieved. In 1974 a research team from The Ohio State University Department of Geodetic Science undertook jointly with the Smithsonian Astrophysical Observatory the investigation of what was then called the "Close-Grid Geodynamics Measurement System" (CLOGEOS). The part investigated at SAO pertained to systems and objectives, while the one at OSU to optimum system use. The points studied under the second part included station configurations, orbital configurations, observational accuracy and data reduction techniques [Mueller et al., 1975]. Following this, a much more realistic simulation study was published [Kumar, 1976] in which the variation of several of the aforementioned factors, as well as new ones, was examined in detail. It must be mentioned here that at that time there was no final decision taken either for the type of ranging instrument to be used or for the orbit of the carrying spacecraft. The variations in orbital configuration were based therefore on theoretical arguments and the selection made was otherwise arbitrary. As for the ranging system, it was implicitly assumed to be a pulsed laser without ruling out any other suitable candidates (e.g., radio Doppler). Although these studies did not produce the final answers to the problems involved with the system, they clarified to a high degree most of them and set up guidelines for future investigations.

On the other hand, SAO produced a final report [SAO, 1977] on the investigations conducted by them pertaining to systems and objectives for CLOGEOS. The major findings of that study were the following: (1) Most promising systems for relative positioning at the 1 cm level are the pulsed laser and the radio Doppler, (2) System accuracy is hindered by insufficient knowledge of atmospheric effects and loss of information due to variable weather conditions. As a result of these two findings, further investigations on the above subjects were proposed. As for the objectives for the employment of such a system, it was emphasized that the main application of the system should be the densification of geodetic networks in the fault regions and their subsequent resurveying in order to produce the required information leading to a four-dimensional (space and time) deformation model. Secondarily, the system could also be used for studying other phenomena of interest to geodesists, geophysicists, glaciologists, and related disciplines.

The momentum acquired from these two investigations and those conducted independently at Goddard Space Flight Center [Vonbun et al., 1975; Agreen and Smith, 1973] along with the recent developments in hardware capabilities pushed the investigation into the next phase. The use of a pulsed laser was finalized and the idea of testing the prototype using the space shuttle under development gave birth to a new system, the "Spacelab Geodynamics Ranging System" (SGRS). The San Andreas fault-system zone was selected as a test area and an error analysis of this specific system was performed at Business and Technological Systems, Inc. (BTS) under the guidelines set by GSFC [Gibbs and Haley, 1978].

At this point it was felt that a variance analysis for the new system SGRS seemed proper. Our study was conducted in two phases. The initial phase is an analysis of the proposed experimental system with shuttle flights. The key issue in this analysis is the variation and dependence of the recovered baseline precision due to the variation of certain factors such as baseline length, a priori station information, observational accuracy, elevation cut-off angle and network design. As this phase reached its end, a workshop at the University of Texas at Austin, organized by GSFC, brought together the various investigators of SGRS and the candidate users of the system. The purpose of this workshop was to review the current system design and gather information from its potential users pertinent to system improvement and operational system design. The results of the discussions and the recommendations from this meeting [Report from the Workshop on the Spaceborne Geodynamics Ranging System, 1979] proved to be of major importance for the design of the of the operational system and for this reason a brief summary is given below:

1. The laser transmitter-receiver system must be designed to reduce the cost of the ground reflectors (below $1000 per unit) and still be capable of centimeter level geodesy.

2. Granted that the system is designed along the lines set during the workshop sessions, geophysicists and seismologists concluded that several unique applications of the system are possible and of great interest to the scientific world. Primarily, the system should be deployed at various areas around the globe in an attempt to "capture" a moderate-sized earthquake.

3. Tectonic plate motion monitoring can be achieved by use of a system such as SRS both for near boundary deformations as well as for relative plate motion determination.

4. In the light of the proposed system design, several—mainly of geophysical interest—experiments are proposed for the study of interplate and intraplate movements and their relationship to area seismicity.

5. The system can be used to establish global and local geodetic control networks of high quality. Mapping and resurveying of large areas with sparse or no geodetic control could be covered very rapidly, effectively and, above all, at low cost compared to classical methods.

6. Various other applications of the system are possible (glaciology, atmospheric sciences, precise time transfer, etc.) provided that the system be designed in a cost effective manner and be capable of achieving relative positioning accuracy of one to two centimeters over intersite distances of ten to fifty kilometers. Since the technology is available, it is recommended that the system be designed for a high flying (~ 1000 km mean altitude) dedicated satellite and equipped with a short pulse (0.2 ns) laser.

The analysis of the capabilities of the new system as it evolved from the guidelines set at the above meeting constitutes the second phase of the present study.

2. OUTLINE OF THE INVESTIGATION

The main objective of this study is to determine the dependence of the recovered baseline precision on the following factors:

(1) A priori station information,

(2) observational accuracy,

(3) geopotential model,

(4) elevation cut-off angle,

(5) baseline orientation (network geometry).

The reason for selecting the baselines as representative end products of the whole process is twofold: primarily, the baseline lengths and the angles between them are the only estimable quantities in the adjustment and secondarily because in most applications of this system the conclusions will be based on the baseline length variation between missions. Since only a covariance analysis is performed, several simplifications were done in the course of simulating the observations. It must be pointed out that it was mainly due to restrictions imposed by the available software that we actually had to simulate observations (GEODYN requires either real or simulated observations in order to form the normal equations and thereby compute variance-covariance matrices for the parameters or functions of the parameters such as the baselines). For a pure covariance analysis, no observations (real or simulated) are required. It was already mentioned in the Introduction that the present investigation deals with two different versions of the basic system. The differences come mainly from the carrying vehicle and the selected orbit. In the first version it is assumed that the laser station will be placed in a low orbit (mean altitude ~ 400 km) aboard the shuttle (validation experiment). The second version is based on a free-flying dedicated satellite at a mean altitude of ~ 1000 km. The investigation of only these particular versions is based on the conclusions and recommendations of the SGRS Workshop [1979] at Austin.

The characteristics of the active part of the system, the laser, were assumed to be the same in both cases. Detailed descriptions of the various components of the system appear in several reports compiled by the different agencies and companies involved in the development of the system. The only information which was actually used in this study is the rate at which the laser can operate, 10 pps, and the presently feasible resolution for the single observation, 10 cm. Although it is highly probable that the new generation pulsed lasers with 0.2 ns pulsewidth and resolution of 2 cm will be operational at the time the system flies, we felt that it was more proper to perform our investigation based on present capabilities. The results of this study can be rather easily projected to indicate the impact of such an important improvement in the hardware.

The simulation site selected covers part of the San Andreas fault system in California and Nevada. The 42 ground-based reflectors form a rectangular grid 400 km long and 200 km wide (Fig. 1). The coordinates of these stations were obtained from BTS so that a direct comparison of our results with theirs would be feasible (Table 1). The observing sequence was the same for all passes and for both orbits. The sequence that we selected is shown in Fig. 2. In reality after a short acquisition period of about 10 s, the laser points to visible stations consecutively until all of them have been observed and then cycles back to make another set of observations. As the spacecraft ascends (or descends) over the horizon, only a few of the stations are visible. This results in an increased number of observations for the peripheral stations compared to the whole. It was felt, however, that for the purpose of this simulation and being consistent with other simplifications (e.g., weather effect simulation), the fixed schedule was adequate.

The satellite orbits generated were based on a simplified force model. Two-body motion was assumed and the gravitational earth model consisted simply of GM and J_2. Only the secular variations due to J_2 were considered in the numerically integrated equations of motion. The integration stepsize was 10 s. The simulated ranges were computed from interpolated state vectors using a third-order finite difference method. This procedure was dictated by the fact that in order to obtain state

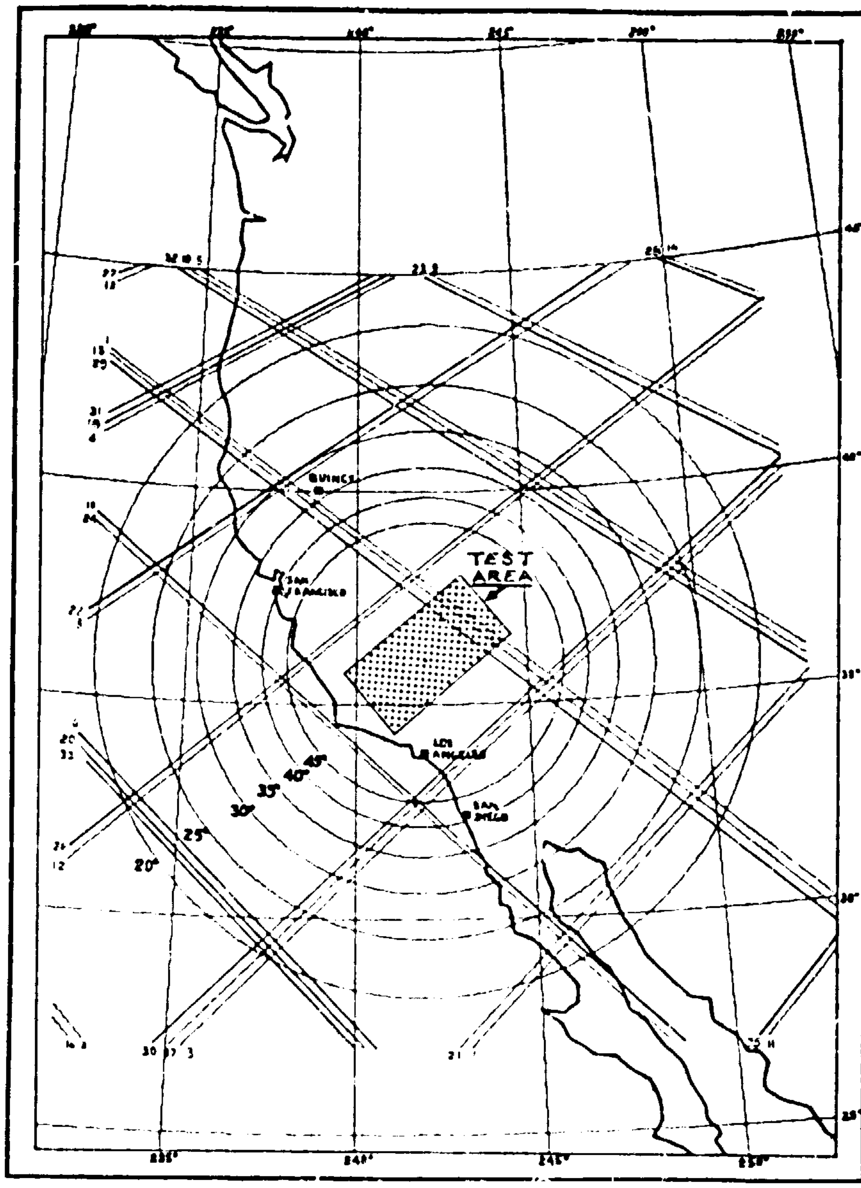


Fig. 1 Test area in California with the low orbit groundtracks superimposed. The circles are the projection of elevation rings for various elevation cut-off angles measured at the center point of the grid.

Table 1 Participating Station Coordinate List

STATION	#	LATITUDE			LONGITUDE			HEIGHT (m)	X (m)	Y (m)	Z (m)
		Reference Ellipsoid: GRS 67							Earth-fixed Geocentric Coordinates		
		°	′	″	°	′	″				
REF011	11	38	3	0.0	243	5	0.0	1524.000	-2277155.029	-4485283.629	3910768.512
REF012	12	37	42	0.0	243	26	0.0	1828.800	-2260490.906	-4520663.151	3880282.797
REF013	13	37	21	0.0	243	48	0.0	1828.800	-2241964.216	-4556272.911	3849465.890
REF014	14	37	00	0.0	244	8	0.0	1524.000	-2225555.817	-4590145.636	3818323.382
REF015	15	36	37	0.0	244	29	0.0	1066.800	-2208298.755	-4626330.838	3783983.014
REF021	21	37	29	0.0	242	11	0.0	1828.800	-2365418.249	-4483253.353	3861222.491
REF022	22	37	8	0.0	242	33	0.0	1219.200	-2347311.368	-4518766.742	3829949.440
REF023	23	36	47	0.0	242	54	0.0	914.400	-2330193.399	-4553597.171	3798723.404
REF024	24	36	25	0.0	243	15	0.0	609.600	-2313144.256	-4589194.107	3765871.815
REF025	25	36	4	0.0	243	37	0.0	1066.800	-2294096.542	-4624804.933	3734814.511
REF031	31	36	29	0.0	242	28	0.0	1219.200	-2373863.041	-4553678.307	3772185.497
REF032	32	36	19	0.0	242	39	0 0	457.200	-2364041.247	-4570455.461	3756845.271
REF033	33	36	8	0.0	242	49	0.0	914.400	-2336397.571	-4588336.267	3740732.685
REF041	41	36	21	0.0	242	15	0.0	1371.600	-2395213.617	-4552552.267	3760367.201
REF042	42	36	11	0.0	242	26	0.0	1371.600	-2385698.191	-4569893.231	3743452.739
REF043	43	36	00	0.0	242	36	0.0	1219.200	-2377854.111	-4587345.612	3728921.047
REF051	51	36	56	0.0	241	17	0.0	1524.000	-2453169.639	-4477713.330	3812410.677
REF052	52	36	38	0.0	241	49	0.0	1676.400	-2420858.734	-4518037.713	3785831.349
REF053	53	36	12	0.0	242	2	0.0	1066.800	-2416916.091	-4551944.729	3746765.581
REF054	54	36	2	0.0	242	13	6.0	1371.600	-2407535.447	-4569518.706	3732062.915
REF055	55	35	51	0.0	242	24	0.0	762.000	-2398205.844	-4587344.681	3715172.870
REF056	56	35	30	0.0	242	45	0.0	762.000	-2380509.610	-4622045.019	3682623.061
REF061	61	36	18	0.0	240	28	0.0	91.000	-2536840.668	-4477770.395	3755138.029
REF062	62	35	58	0.0	240	50	0.0	122.000	-2518774.826	-4512980.119	3725283.180
REF063	63	35	37	0.0	241	11	0.0	762.000	-2502358.038	-4548637.765	3694154.826
REF064	64	35	16	0.0	241	33	0.0	1219.000	-2484104.152	-4584687.228	3662778.128
REF065	65	34	55	0.0	241	54	0.0	762.000	-2466391.514	-4619140.840	3630738.567
REF071	71	36	1	0.0	240	3	0.0	61.000	-2571565.447	-4475255.595	3729730.267
REF072	72	35	40	0.0	240	24	0.0	91.000	-2562443.891	-4510716.991	3698272.566
REF073	73	35	19	0.0	240	45	0.0	122.000	-2544548.106	-4546719.407	3666672.665
REF074	74	34	59	0.0	241	7	0.0	305.000	-2527088.127	-4580952.188	3636539.428
REF075	75	34	38	0.0	241	29	0.0	1219.000	-2503656.707	-4617167.502	3605176.947
REF081	81	35	52	0.0	239	50	0.0	305.000	-2600498.975	-4474089.772	3716404.075
REF082	82	35	31	0.0	240	12	0.0	152.000	-2583012.139	-4510193.442	3684774.151
REF083	83	35	11	0.0	240	34	0.0	152.000	-2564611.191	-4545266.671	3654609.356
REF084	84	34	50	0.0	240	55	0.0	1829.000	-2548295.463	-4581518.993	3623762.598
REF085	85	34	29	0.0	241	17	0.0	305.000	-2528932.420	-4616001.367	3590953.043
REF091	91	35	43	0.0	239	37	0.0	457.000	-2622385.103	-4472727.624	3702991.947
REF092	92	35	22	0.0	240	00	0.0	610.000	-2603752.003	-4509830.764	3671430.288
REF093	93	35	1	0.0	240	22	0.0	914.000	-2586027.224	-4546070.990	3639918.277
REF094	94	34	41	0.0	240	43	0.0	1219.000	-2568668.776	-4580435.419	3609740.333
REF095	95	34	20	0.0	241	4	0.0	305.000	-2550919.124	-4614634.656	3577224.259

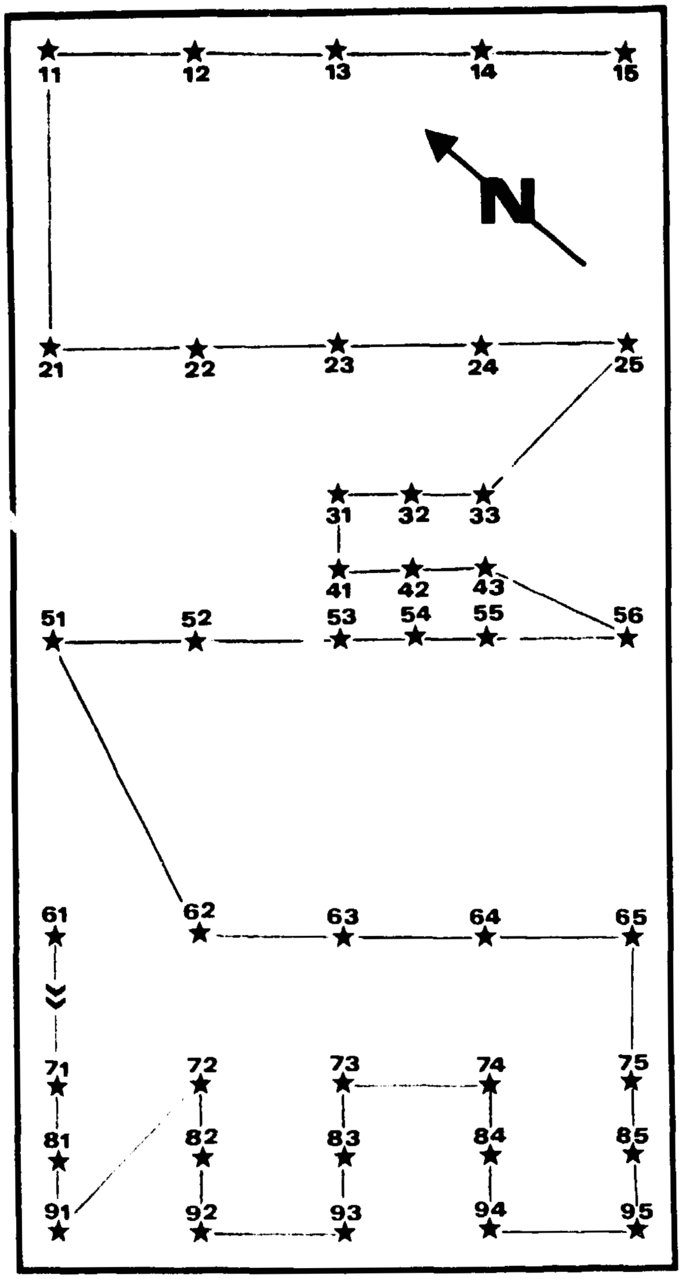


Fig. 2 Relative positions of the retroreflectors in the California test area. The lines indicate the observational sequence.

vectors at the 10 pps rate from the numerical integration, computational instabilities and unfavorable round-off accumulation introduced unacceptable biases on the results. The interpolation technique was tested against straight-forward integration and was found far more efficient in terms of computation time and accuracy. A summary of the information pertinent to the generation of the two orbits and the corresponding range observations is presented in Tables 2, 3 and 4.

In order to be able to compare our results with those of other similar investigations (like BTS), we used only nine passes (50% of the available) for the low orbit. The reason for using only one-half of the passes is that we assumed that there is a fifty-fifty chance that a pass will be observed depending on the weather conditions in the area. This, of course, is the simplest way of modeling the weather, but it was felt to be adequate for our purpose. For the high orbit only eight passes were used (25% of the available) and the observing rate was lowered to 1 pps. To compensate for this decrease in the amount of data, we increased the observational precision to 10 cm/$\sqrt{10}$ $\cong$ 3.2 cm, assuming that the ten observations within each one-second interval are independent. This assumption had also been made at BTS when developing their data set for the low orbit. We used this data set as provided by them for checking our software (GEODYN) and verifying their results. For the sake of brevity, the above three data sets will hereafter be referred to as OSUL, OSUH, and BTSL respectively.

As mentioned previously, the computer program used for the adjustment of the observations was GEODYN, obtained from NASA at an earlier time. This program is designed primarily for complex dynamic solutions and its use for the present investigation somewhat beats its purpose. However, since other investigators used this program already and a number of its capabilities simplified to a high degree the task undertaken, we decided to use it. We should mention here that in the present investigation only dynamic solutions were considered. Although in [Kumar, 1976] and [Kumar and Mueller, 1978] the geometric solutions are shown to be more promising than the dynamic, we decided that on the basis of today's technology the realization of simultaneous ranging to the required more than six (or even four) ground stations

Table 2 Data used in the simulation procedure.

Central Body Constants

$a_e = 6\,378\,160.00$ m

$1/f = 298.247\,167\,427$

$\omega_e = 0.729\,211\,514\,67\ s^{-1}$

$GM = 398\,603 \times 10^9\ m^3 s^{-2}$

$J_2 = 0.001\,082\,7$

Low Satellite Orbit

$a = 6\,778\,170.32$ m

$e = 0.00$

$i = 50°$

$\Omega = 0°$

$\omega = 0°$

$M = 0°$

Epoch T_0: 0^h UT, June 1, 1974

High Satellite Orbit

$a = 7\,378\,160.00$ m

$e = 0.00$

$i = 50°$

$\Omega = 0°$

$\omega = 0°$

$M = 0°$

Epoch T_0: 0^h UT, June 1, 1974

Table 3 Distribution of Observations per Station per Pass for the Low Orbit (OSUL)

STATION												
PASS#	REF011	REF012	REF013	REF014	REF015	REF021	REF022	REF023	REF024	REF025	REF031	REF032
1	69	65	67	80	133	73	66	65	73	99	64	65
4	84	91	101	110	124	63	68	72	75	79	65	67
23	115	106	102	97	92	77	76	74	72	69	69	68
28	67	65	66	79	130	73	66	65	73	99	65	65
31	65	92	101	110	122	64	68	72	75	79	66	68
46	47	47	47	48	50	54	53	53	54	59	55	56
50	114	107	102	97	92	77	76	75	71	69	68	68
55	66	64	65	77	127	72	66	65	72	99	65	65
58	87	93	102	109	119	65	69	73	76	78	67	67
TOTAL	733	730	753	807	989	618	608	614	641	730	584	589

STATION												
PASS#	REF033	REF041	REF042	REF043	REF051	REF052	REF053	REF054	REF055	REF056	REF061	REF062
1	69	63	64	66	79	66	63	68	65	76	78	65
4	68	64	65	67	57	61	63	68	68	68	53	62
23	67	67	66	65	68	67	66	65	65	63	69	68
28	69	64	64	66	73	66	63	64	65	77	73	65
31	68	65	66	67	59	62	64	65	66	68	59	62
46	57	57	57	59	64	53	53	59	60	70	73	68
50	67	67	66	65	69	68	66	65	64	62	70	69
55	69	64	64	67	73	65	63	64	65	78	79	65
58	69	65	66	67	59	63	65	65	67	68	60	63
TOTAL	603	576	578	589	611	577	571	575	582	630	629	587

STATION												
PASS#	REF063	REF064	REF065	REF071	REF072	REF073	REF074	REF075	REF081	REF082	REF083	REF084
1	60	60	63	75	62	53	58	59	73	62	58	57
4	64	66	69	60	63	67	70	74	61	66	69	74
23	67	65	68	72	71	69	67	65	78	75	72	69
28	61	60	63	76	64	60	58	60	74	62	59	57
31	65	67	69	61	64	67	71	73	63	66	70	74
46	64	66	[illegible]	83	72	68	69	82	93	74	69	70
50	67	64	62	73	71	69	67	64	78	75	72	68
55	62	61	64	78	64	60	59	60	76	64	60	58
58	65	68	69	62	64	63	71	73	64	67	71	74
TOTAL	575	577	602	645	595	586	590	610	662	611	600	601

STATION							
PASS#	REF085	REF091	REF092	REF093	REF094	REF095	*TOTAL*
1	58	71	60	57	55	56	2839
4	83	64	69	76	87	113	3057
23	67	101	83	79	73	70	3122
28	58	73	62	57	56	57	2849
31	82	65	69	77	87	112	3076
46	85	107	76	70	71	87	2761
50	66	103	88	79	72	69	3121
55	59	74	62	58	57	58	2859
58	81	66	70	78	87	110	3090
TOTAL	639	724	644	631	646	732	26774

Table 4 Distribution of Observations per Station per Pass for the High Orbit (OSUH)

	STATION											
PASS#	REF011	REF012	REF013	REF014	REF015	REF021	REF022	REF023	REF024	REF025	REF031	REF032
1	14	14	14	16	20	13	14	14	15	18	14	14
9	11	12	12	12	12	11	12	12	12	13	12	12
20	13	13	13	13	13	14	14	14	14	14	14	14
25	12	12	12	13	13	12	12	12	13	13	12	13
36	14	14	14	14	13	13	13	13	13	13	14	14
41	12	13	13	13	14	12	13	13	13	15	13	13
52	13	13	13	13	13	13	13	13	13	13	13	13
57	13	13	13	14	15	13	13	13	14	15	13	14
TOTAL	102	104	104	108	113	101	104	104	107	114	105	107

	STATION											
PASS#	REF033	REF041	REF042	REF043	REF051	REF052	REF053	REF054	REF055	REF056	REF061	REF062
1	14	14	14	14	13	13	13	14	14	15	13	13
9	12	12	12	12	11	11	12	12	12	13	11	12
20	14	14	14	14	14	14	14	14	14	13	14	14
25	13	12	13	13	12	12	12	12	13	14	12	12
36	14	14	14	14	14	14	14	14	14	14	14	14
41	13	13	13	13	12	12	13	13	13	15	12	12
52	13	13	13	13	13	13	14	14	14	14	14	14
57	14	13	13	14	13	13	13	13	14	15	12	13
TOTAL	107	105	106	107	102	102	105	106	108	114	102	104

	STATION											
PASS#	REF063	REF064	REF065	REF071	REF072	REF073	REF074	REF075	REF081	REF082	REF083	REF084
1	13	14	14	13	13	13	13	14	12	13	13	13
9	12	13	14	11	12	12	13	15	11	11	12	13
20	14	14	14	15	15	15	15	15	16	15	16	15
25	12	13	15	12	12	12	13	15	12	12	12	13
36	14	14	14	14	15	15	15	15	15	15	15	15
41	13	13	15	12	12	13	13	15	12	12	12	13
52	15	15	15	14	14	14	14	15	15	15	15	15
57	13	13	15	12	13	13	13	14	12	12	13	13
TOTAL	106	109	116	103	106	107	109	118	105	105	108	110

	STATION						
PASS#	REF085	REF091	REF092	REF093	REF094	REF095	*TOTAL*
1	14	12	13	13	13	13	581
9	15	11	11	12	13	16	512
20	15	16	16	16	16	16	604
25	14	11	12	12	13	15	529
36	16	16	16	17	16	17	604
41	14	12	12	12	13	14	543
52	16	15	16	16	17	19	593
57	14	12	12	13	13	14	557
TOTAL	118	105	108	111	114	124	4523

from a spaceborne laser is practically impossible. Furthermore, we chose to use the short-arc mode, primarily in order to avoid the accumulation of biases in the recovered baselines due to uncertainties in the force field description (mainly the part pertinent to earth gravity modeling), and secondarily realizing that only a small fraction of a full revolution will be covered by observations since the stations are all spread over a very limited area. Unlike other software packages which are specifically designed for short-arc adjustment, GEODYN treats the short-arc solution as a collection of simultaneously reduced short (in duration) "long arcs." There are no approximate solutions to the equations of motion nor any other approximation whatsoever. As it is explained in [Mueller et al., 1975], one cannot expect to determine anything but relative positions from such a limited mission. The geopotential model therefore must be held fixed and the optimum way for its description must be determined through suitable experiments. For our purposes we nominally used the GEM7 spherical harmonic expansion up to degree and order sixteen (16, 16). This was dictated by the fact that the same model was also used by BTS in their investigation. The groundtracks of the generated satellite passes are depicted in Fig. 3 (OSUL) and Fig. 4 (OSUH).

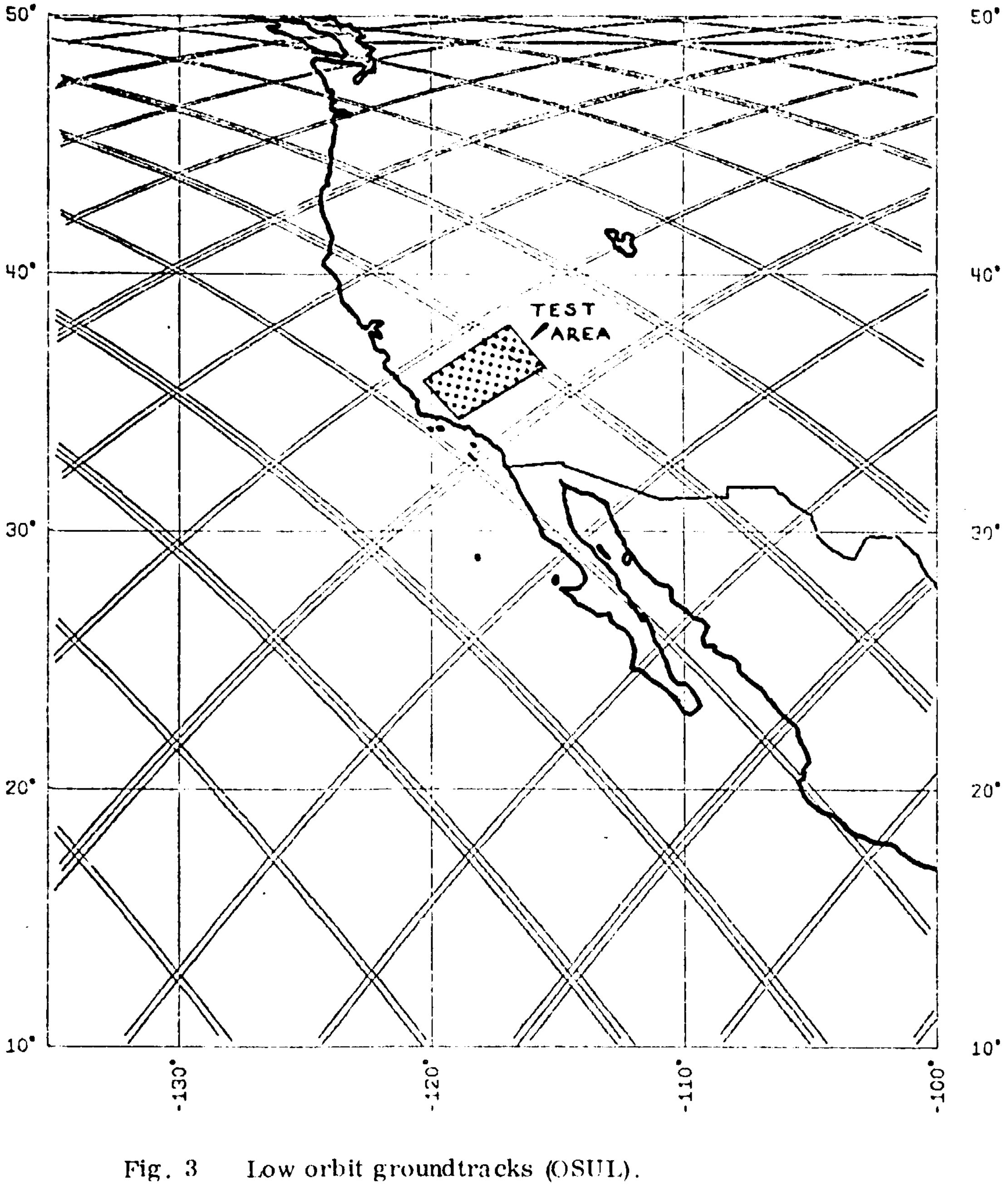


Fig. 3 Low orbit groundtracks (OSUL).
Seven day mission.

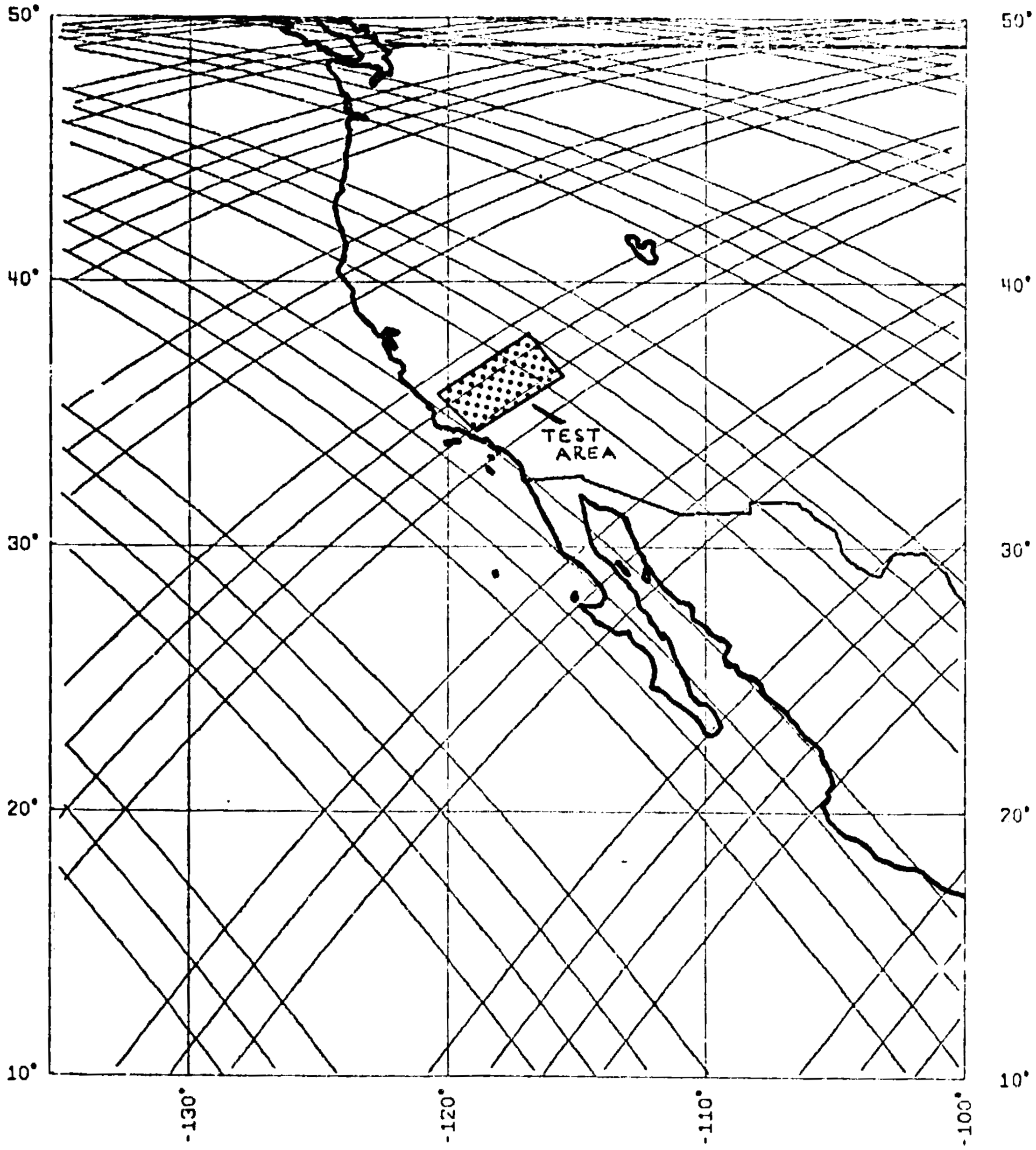


Fig. 4 High orbit groundtracks (OSUH).
Seven day mission.

3. THE ESTIMATION PROCESS

The estimation process is described in [Martin et al., 1976]. This discussion, though, is limited in theoretical details and treats mainly the technicalities and the implementation of the process in the associated computer program. A similar brief recollection of the formulas is also given in [Gibbs and Haley, 1978]. According to these references the method employed in GEODYN is Bayesian estimation. This type of estimation makes use of a priori information about the parameters in the form of a weight matrix derived from the prior distributions of the parameters. In most cases no such distributions are available and the weight matrices are derived from the assumed standard deviations of the a priori estimates for the parameters. In very rare circumstances a full variance-covariance matrix obtained from a previous solution is used for the determination of these weights. This latter procedure led several scientists into a different interpretation of the process, namely the so-called "least squares estimation with 'observed' parameters." By doing this, assuming that our approximate parameters are the outcome of some measurements, we effectively change their character from fixed quantities to stochastic ones. It is not that there exist no problems where the parameters are inherently of that nature, but in our problems, especially when dealing with parameters such as Cartesian coordinates, we cannot justify such an assumption. We should also point out that quite frequently the above procedure is used to improve the condition of the normal equations and in extreme cases to produce a full rank normal matrix for a problem which otherwise would be unsolvable in the domain of Cayleian matrix algebra. The application of such "weighted" constraints on the estimates may very well distort the results in various ways and sometimes it may even result in unacceptable answers. From this point of view, the estimation process is directly related to the concept of estimable parameters, due to R.C. Bose.

It is proved in [Rao, 1973, p. 224] that the necessary and sufficient condition for a linear function of the parameters to be estimable is that the rank of the normal equations is equal to the number of the parameters under estimation. A necessary and sufficient condition for the case where generalized inverses are used is also given in the above reference. The advantage of dealing with estimable quantities is that they are unique and unbiased for any solution of the normal equations and have minimum variance among all linear unbiased estimates. In this sense it is obvious that if we can identify the estimable parameters in the problem and if we can modify the model so that all the parameters are estimable, we have guaranteed ourselves a unique minimum variance unbiased estimate. This being the case, the need for a-priori statistical information on the parameters is alleviated, and our estimates are based only on the information provided through the observations.

In the case where we cannot find such a set of estimable parameters to describe the system, we can still by-pass the need for exterior information by use of a generalized inverse solution. If it is possible to find linear parametric functions that fulfill the estimability criteria, then the analysis of the system can be done on the basis of these results. In the case of a geometric solution in satellite goedesy, for example, it has been shown [Mueller at al., 1975] that the baseline lengths in the network and the angles formed by any three stations are estimable. For most of the applications in geodynamics, these two quantities are sufficient for inferring motions and deformations in the area. The advantages of parametrizing a problem in terms of estimable quantities have been recognized by most scientists and effort has been made recently to identify these quantities for various geodetic problems. The complex functional relationships between the observables and the parameters are the only reason that estimable parametrization has not been in wide use yet. The case of laser observations, however, has been treated extensively in [Van Gelder, 1978], and the estimable parameters have been determined for extremely simplified models as well as for more complicated ones where secular perturbations due to J_2 were included.

One therefore has all the tools to perform a proper study of a system such as the SRS, at least for simple simulations as is the case here.

If this is so, then one can naturally question the reason behind this investigation and its major concern for the effects of a priori information for nonestimable parameters. The answer to this is rather simple. Real world problems are far more complicated than an error analysis based on a simulation of the system, and although major effort is currently devoted to improving and updating our estimation procedures we are far from being able to solve our problems in the fashion described above. Our best alternative, therefore, is to study in detail the current procedures to a degree that would allow us to justify our results and to determine how optimistic or pessimistic they may be due only to biases introduced by the employed procedure.

In an effort to examine and clarify the inherent characteristics of Bayesian estimation, some of the major variations of this process are presented and compared to the well-known least squares estimation in the Appendix. It should be pointed out that the use of Bayesian estimation over least squares is an open question for statisticians as well as for scientists using these methods. Subjectivity versus objectivity is a rather philosophical question, and we feel that it is not the purpose of this study to provide the answer. One conclusion that can be drawn is that there are situations in applied science where one method is better than another. It is therefore our responsibility to choose between the two and to do this we must study both. The stand we take here is that the choice will be made on the basis of the problem requirements only, irrespective of the personal preference of the investigators.

4. CONSTRAINTS, RANK DEFICIENCY AND ILL-CONDITIONING IN SHORT-ARC SOLUTIONS

In the course of this investigation we have pointed out that we are primarily interested in determining the influence of station-related a priori information on the baselines' precision. It is only natural, however, to address the following question: What is the rationale for using such information? The answer to this question is not as simple as one might expect. If the problem is studied in depth, it will soon become apparent that this is just another way of posing the following fundamental question: Subjective (Bayes) or objective (Gauss-Markov) estimation? This is an open question for the statisticians and has a rather philosophical than mathematical nature. One should, therefore, not expect an answer from a limited study as this one. We would rather discuss the geodetic aspects of the problem and refrain from attempting to provide a conclusive answer, as that is outside the scope of this study.

In the theory of linear spaces, rank defect of a matrix is the number of linear dependencies which exist among its columns. If we consider a matrix A as a linear transformation from a space $\mathbb{R}^m$ to a space $\mathbb{R}^n$, that is, $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$, then the following example illustrates the concept of rank deficiency. From linear algebra it is known that the column rank of a matrix is the same as its row rank, where by column (row) rank we mean the maximum number of linearly independent columns (rows) of the matrix. Suppose $n \geq m$ for A and rank (A) $=$ k where the dimensions of A are n by m (rows x columns). From the above theorem it is obvious that the rank of A will be at most m, when all its columns are linearly independent. If, therefore, $k = m$, the rank deficiency of A is zero or A is of "full rank." If $k < m$, then the difference $m - k = m - \text{rank}(A)$ is the rank deficiency of A. In terms of linear operator theory this means that the null space of the operator represented by A is not just the zero element, $\text{Ker}(A) \neq \{0\}$, where the null space or kernel of a linear operator on $\mathbb{R}^m$ is defined as the subset of its domain $\mathbb{R}^m$, consisting of elements $x \neq 0$

for which $Ax = 0$ holds. Using another theorem now, we can state that the transformation is not "one-to-one" or "injective" and the equation $Ax = y$ therefore does not have a unique solution for x.

The last remark provides the link between the mathematical and the statistical and physical interpretation of rank deficiency. If y denotes the vector of observations in a system and x the vector of parameters, with A the design matrix relating the two, then given a set of n independent observations ($n \geq m$), only $k = \text{rank}(A)$ parameters out of the total m can be estimated uniquely. Physically this simply means that the given set of observations does not contain the information needed for the determination of the $m - k$ parameters. In this sense, the parameters which can be estimated from the given observations constitute the set of "estimable" parameters of the system. We can conclude therefore that for each kind of observation in a given system there is a corresponding set of "estimable" parameters. It is easy to see, for example, that observing the velocity of a vehicle is not enough to define its position in space and time. We must find, therefore, a way to remedy this handicap in order to solve the problem. This is basically done through the application of constraints on the nonestimable parameters. We can eliminate, for instance, these parameters by adopting certain values for them (absolute constraints), determined from a different approach. In this sense the "nonestimable" parameters are not parameters anymore but constants of the problem. A solution of this problem through least squares (to account for redundant observations when $n > m$) will provide minimum variance unbiased estimates for all (estimable by now) parameters [Rao, 1973].

There are situations, however, where there is reason to believe that the available statistical information about the nonestimable parameters is of poor quality and enforcing them in the estimation process would result in unreasonable distortion of the results and sometimes even indicate nonexistent inconsistency in the observations. The solution of the problem in this case can be obtained utilizing what is called a set of flexible constraints. These in turn can be either arbitrary in number (but enough to provide a solution) or they can be just enough to make the least squares normal equation matrix ($N = A^TA$) of full rank. The latter are called "minimum" or "minimal" constraints. It is obvious that these minimum constraints are not unique and the stability of the solution depends on the selection of these constraints. There is, however, a set of minimum constraints which is best in the sense that it does not discriminate among the parameters and provides a solution which is least affected by numerical instabilities in the system. This set of constraints is called "inner adjustment" constraints, and the theoretical and practical aspects of their application have been discussed extensively in [Rao, 1973] and [Blaha, 1971]. We will not attempt a discussion of the pertinent details again, but we will point out that this approach is basically the same as solving the original set of normal equations by use of a generalized inverse of N [Rao, 1973]. The advantage of the "inner constraints" is that we obtain the same results without the troublesome computation of a generalized inverse. The application of this type of constraint in the present investigation was not possible due to limitations in the available software.

Finally, another type of constraint which is widely used for the solution of rank deficient systems is the "weighted" or "relative" constraint. Since these were used in our test, we will elaborate on them and give some more details in addition to those which were already mentioned in the discussion of the estimation process. The idea of "weighted" constraints originated from the Bayesian approach of estimation in linear models. In this case, however, the a priori information which is added to the normal equations is based on an a priori known distribution of the parameters and the reason for using this information is well justified if one accepts Bayes' theorem. What we are trying to stress here is that in Bayesian estimation, the weighting of the parameters on the basis of a priori information is not intended to alleviate the rank deficiency in the problem nor the ill-conditioning of the normal equations. It merely makes use of all available information in order to arrive at estimates which are closer to the true values of the parameters, although <u>not unbiased</u> anymore. If, therefore, we know the prior distribution of the parameters, then this approach is fully justified as long as we are aware of the consequences of the Bayesian approach (biased parameters). In this sense the inclusion of a priori information on the parameters should not at all be related to the inherent rank deficiency or ill-condition of the problem.

Weighted constraints, however, are used to overcome this problem. Their application in this sense is very dangerous as far as their effects on the results are concerned. In general, the argument on which this practice is based is that prior knowledge of the parameters' range of values is available through direct observations on them or from a previous solution. In the first case we change the role of the parameters to observations, and by doing this we are effectively removing them from the parameter list altogether. If in such a case we applied the weighted constraints on the nonestimable parameters only, then we have resolved the rank deficiency of the problem since we already know that for estimable parameters the design matrix (hence the normal equations) is in general of full rank. The catch is that in several cases we are either unable to observe the parameters because of their nature (e.g., Cartesian coordinates) or we directly substitute their "observed" values with some approximate ones. In both these cases, the variance of these "measurements" is based on some personal confidence interval for the assigned values rather than what actual measurements would indicate. In the limit, as the a priori variance is decreased, the results of this adjustment are equivalent to those of absolute constraints, where we have changed the role of the nonestimable parameters to some adopted constants of the problem. For the case where we indeed have direct observations on some parameters (e.g., absolute gravity measurements in a gravity network), then we must treat them as observations with the proper variance-covariance matrix as we do for all other observations. This, of course, brings in the problem of determining the relative weights when different types of observations are simultaneously adjusted. This is a different problem, however, and we need not concern ourselves with it at this point. The important thing to note is that this type of observation should be treated in the usual manner and not used as a means for circumventing the rank deficiency in the problem.

As far as the second case is concerned, where the weighting is based on a previous solution, we can identify two subcases. If all or some of the parameters were obtained from a previous adjustment, then their full variance-covariance matrix should be used in the new one, and the result is a standard Bayesian adjustment as described previously. If, however, only the diagonal elements (variances) are used (which is common practice), then the resulting parameters (which are biased due to the use of a priori information) are not minimum variance - minimum bias estimates since the arbitrary diagonalization of the covariance matrix has changed their prior distribution from the one represented by the full matrix. Such practice is dangerous not only for the rank deficient problems but for those with full rank as well. In the case of the first, it merely provides deceiving results while in the second case it may distort the results instead of improve them. Some investigators choose to ignore the consequence of the last remark since they only judge the quality of their results from the magnitude of their a posteriori variances. These variances, however, are affected (and usually deflated) by the wrong a priori inputs and their statistical interpretation is very much questionable.

This discussion can be summarized in the following. Rank deficiency is an inherent characteristic of each observational system, and in order to overcome it we must study the system and determine the source of this deficiency. This can be done through the determination of the estimable parameters for the system and their comparison with the list of our solve-for parameters. If there are observations (direct or indirect) which can be done to determine the non-estimable parameters, then we can perform these observations and include them in our adjustment alleviating the rank deficiency. When this is not possible but we have prior information in the form of a full covariance matrix, then we can perform a minimum variance - minimum bias estimation (Bayesian approach) where we obtain a solution for all parameters at the expense of their unbiasedness. Alternatives to these are the inner constraint approach or a direct generalized inverse solution and the adoption of absolute constraints. We point out, however, that the first does not provide unbiased estimates for nonestimable parameters. The similarities of this approach to Bayesian estimation are pointed out in the Appendix. An interesting and illustrative discussion of this approach is given in [Grafarend and Schaffrin, 1974]. The implications of the use of absolute constraints are rather obvious. The results are as good as the adopted values. One only needs to consider the by now popular leveling network example where the absolute heights of the points can take different values depending on a single point's adopted height [Van Gelder, 1978].

Having discussed the nature of the various approaches to overcome rank deficiency problems in general, we proceed in examining the work and the results obtained therefrom for a satellite laser ranging system, in particular when used in the short-arc mode. The theoretical investigation of this system is given in [Van Gelder, 1978] where it is shown that for the simple case of a general elliptic orbit secularly perturbed by the J_2 harmonic, the rank deficiency is two. From the numerical tests which were performed for the case of the spaceborne laser, however, the results seem to contradict theory. Test runs with even three absolute constraints on a single station's position vector failed to provide an acceptable solution indicating serious instabilities in the normal equations. This, however, should not be taken as a proof that there is a flaw in the theoretical investigations, but rather as an indication that for a strong solution there is something more than rank deficiency to be considered. That other element was already mentioned in the estimation process discussion, and it is related to the geometry of the problem. In our investigations we are dealing with an extremely limited area (only 200 km by 400 km) and with short arcs of length which at best reach only a tenth of a full revolution. With such poor geometry any information about the coordinate system definition, which would normally come from the orbital dynamics, is so little and insufficient that the ill-conditioning of the normal equations dominates the problem rather than the inherent rank deficiency. In this case it may be that by increasing the number of observed arcs this instability is greatly reduced. This, however, must be investigated since such an increase would also introduce new parameters in the original problem. At present a definition of the origin and orientation of the coordinate system through six suitable absolute constraints seems to be our best alternative. Since these are a set of "over-constraints" (the theoretical rank deficiency still remains two) but essential in order to provide results clearly independent of numerical instabilities, we refer to them as "quasi-minimum" constraints following [Van Gelder, 1978].

5. DESCRIPTION OF THE EXPERIMENTS

The investigation was conducted in three stages which offers a natural classification for the description of the experiments. In the initial stage preliminary experiments were conducted in order to familiarize ourselves with the problem, the available software and to set up guidelines for experiments that would follow. The second stage deals with the experiments on the low orbit and the final stage with similar experiments for the high orbit.

A. Preliminary Experiments

When a simulation study is conducted, it is very important to have the ability to check the results with those obtained independently either through a different approach or a different study by another organization. Since in our case a similar study was conducted in parallel at GSFC and BTS, we followed their guidelines in setting up the experiments in order to be able to compare our results. The main purpose of the preliminary experiments was to become familiar with the problem and to test our own software (GEODYN version 7508.0) using the data provided by BTS (BTSL). Since our main concern is the effect of a priori information on the baseline precision, the preliminary test focused on the variation of this factor. These tests explored the baseline precision variations as the a priori information about the stations, the orbit and the observational accuracy varied (Table 5). The effect of an erroneous geopotential model on the baseline precision was also investigated. As is pointed out in [Gibbs and Haley, 1978], the BTSL data set was generated on the basis of GEM1 (4, 1), while in our tests the earth is modeled through GEM7 (16, 16). We felt that this inconsistency between the data generation and the reduction techniques should be cleared as to its effect on the precision of the baselines. Intuitively one expects that since we are dealing with a limited area and we are doing a short-arc adjustment, there should be no

Table 5 Baseline Standard Deviations[1]. Preliminary Test Results.

Test	OSU 1	OSU 2	OSU 3	OSU 4	OSU 5	BTS 1a
Station Constraints[2]	1 m	1 m	5 m	25 m	Quasi-Minimum Constraints	1 m
Orbital Constraints[3]	Constrained	Constrained	Constrained	Free	Free	Constrained
Standard Deviation of Observations / Baseline/Length	10 cm	100 cm	10 cm	10 cm	10 cm	10 cm
91-81 (26 km)	0.98	9.42	0.98	0.98	0.98	0.97
91-71 (51 km)	1.01	9.71	1.01	1.01	0.95	1.00
91-61 (100 km)	1.04	10.12	1.04	1.08	0.98	1.05
91-51 (202 km)	1.14	10.75	1.14	1.17	1.04	1.12
91-21 (302 km)	1.17	11.10	1.17	1.30	1.08	1.16
91-11 (403 km)	1.14	10.66	1.14	1.33	0.76	1.11
91-15 (449 km)	1.17	10.09	1.17	1.30	1.01	1.17
91-92 (52 km)	1.08	10.34	1.08	1.08	1.11	1.07
91-93 (103 km)	1.04	10.15	1.04	1.08	1.04	1.06
91-94 (152 km)	1.08	10.28	1.08	1.11	1.08	1.08
91-95 (203 km)	1.04	10.24	1.04	1.11	1.04	1.06
Average A Posteriori Standard Deviations of Station Coordinates (X, Y, Z)	~ 20 cm	~ 30 cm	~ 80 cm	~ 500 cm	~ 4 cm	—

[1]Table values are in centimeters. [2]Standard deviations in each Cartesian coordinate.
[3]Constraints as applied by BTS: 58 m in X, Y, Z and 5.8 cm/s in $\dot{X}$, $\dot{Y}$, $\dot{Z}$.

change in our precision estimates. Indeed our numerical tests showed no difference (at least in the order of millimeters) in the baseline precisions although there are appreciable changes in their lengths. The results are not tabulated since they are identical for all baselines.

We can summarize the findings of these preliminary tests in the following. An almost linear relationship exists between the precision of the observations and the precision of the recovered baselines. Although this is not a peculiarity of dynamic solutions (as it is for the geometrical ones) we can probably attribute it to the fact that there is a uniform distribution of stations in the area with approximately equal numbers of observations from each station to an optimal set of satellite passes covering it as an umbrella from all possible angles. This whole setup strengthens the geometry of the problem to such a degree that in this respect it behaves almost as a geometric problem. As can be seen from Table 5, there is some variation in the baseline precision as the constraints were varied from case to case. In connection with the estimability problem in the short-arc mode, the last entry in the table shows an average standard deviation of the recovered Cartesian coordinates of the stations. It is quite obvious that since they are nonestimable quantities their precision is strongly dictated by the input a priori information. These numbers were not included to prove that these quantities are nonestimable, but rather to show the pitfalls one faces if he were to choose these quantities as the basis of his investigation. The a priori weighted constraints on the orbital elements were introduced in accordance with the guidelines and tests conducted at BTS [Gibbs and Haley, 1978]. For reasons discussed in [Van Gelder, 1978], these constraints were never again introduced in the system in the tests that followed.

In the last column we included the results obtained at BTS in their Run # 1a [Gibbs and Haley, 1978, p. 65] which most closely resembles our Test # 1. The tabulated values in the above reference pertain to precision of the horizontal components of the baselines only, but for comparison purposes and because of the relatively short distances involved they can be taken as the precisions of the baselines' lengths themselves. It is evident from these numbers that our results are in excellent agreement with theirs.

B. The Low Orbit Experiments

The results of the preliminary tests indicated that most of the variation in the baseline precisions came from the variation of the weighting scheme and the selection of an elevation cut-off angle. The dependence of the results on the observational precision turned out to be rather straightforward, and therefore no further investigation for this factor seemed necessary. As for the network-orbit configuration effects, related to the geometric strength of the problem, little could be done since the orbit and the network design were given and assumed to be the common denominators for all experiments.

With the above in mind the investigation concentrated mainly on the effects of weighting and elevation cut-off angle and to a lesser degree on the effects of geopotential model variations. Three different weighting schemes were adopted: (1) strong weighting for all stations, assuming that the standard deviation for each of the three coordinates (X, Y, Z) is 1 m, (2) mild weighting based on a standard deviation of 25 m in each component, and (3) "quasi-minimum constraints" solutions, that is, the coordinate system is defined by holding fixed six coordinates properly distributed among three selected stations.

The "quasi-minimum" constraints for both OSUL and OSUH orbits were arranged as follows:

(i) Station REFØ11: X and Y coordinates held fixed,

(ii) Station REFØ15: Y and Z coordinates held fixed, and

(iii) Station REFØ95: X and Z coordinates held fixed.

The elevation cut-off angle was varied twice, 20° assumed to be the nominal value and 35° above horizon. The attention given to this factor is warranted by the fact that a higher cut-off angle could simplify the design of the ground reflectors and reduces their manufacturing cost. Some of the tests were repeated using a different geopotential model in order to verify the results obtained from the preliminary experiments and detect possible interactions due to the variation of the other two factors. In all cases no a priori information on the orbit was introduced. Additional discussion for the "quasi-minimum" constraints is left for a later section.

C. The High Orbit Experiments

The results obtained from the investigation of the low orbit indicated that the designed experiments were suitable enough to allow the determination of the effects of considered factors on the baseline precision. In addition to this, the fact that a comparison of the results from the two orbits would reveal their dependence on the spacecraft's mean altitude convinced us to follow the same experimental design. The description of these experiments is omitted since they are identical to those for the low orbit as described in the previous section. Additional tests were performed in this case where the main objective was the determination of the "effective" rank deficiency of the problem. By "effective" rank deficiency we mean the combined effect of the inherent rank deficiency of the short-arc mode adjustment and the deficiency arising from the ill-conditioned normal equations which is characteristic of the specific problem under study. The strategy followed in this case was to relax the number of "quasi-minimum" constraints and then examine the condition of the resulting normal equations. As one can gather from the above, the effective rank deficiency depends grossly on the design of the experiment and the quantity, structure, and quality of the collected observations. It would be useless, therefore, to try to quote for it a number such as two, three or six, while it is more appropriate to give some general guidelines that can be followed for a broad class of problems.

6. RESULTS OF THE LOW AND HIGH ORBIT EXPERIMENTS

The results which are presented and discussed in the following sections were obtained from the experiments conducted with the low orbit (OSUL) and the high orbit (OSUH) in the last two stages of the investigation. As explained in the previous section, there is a total of twelve basic experiments, six for each orbit (three weighting schemes for each of the two cut-off angles). The quantities analyzed are naturally the standard deviations of the estimated baseline lengths. For a network of say n stations, there is a maximum of $m = \frac{n(n-1)}{2}$ baselines which can be formed in all possible combinations (without repetition) among the stations. In our case n = 42, which yields m = 861; and considering the fact that there were twelve different solutions, we arrive at the total number of standard deviations to be analyzed: 10,332. Although theoretically a large data sample has several advantages, practically one can infer almost the same amount of information (not as accurately though) by restricting the analysis to a smaller sample formed on the basis of certain justifiable assumptions. We will elaborate on this a little further in the course of explaining the method of analysis, since these assumptions lay part of the foundation of our conclusions.

A. Detection of Sources of Variation in the Sample Through an Analysis of Variance (ANOVA)

The analysis of variance (hereafter referred to as ANOVA) is a statistical method of analyzing measurements depending on various kinds of effects (called factors) which simultaneously affect them, in order to make qualitative and quantitative inferences of these effects [Scheffé, 1959].

Since the method implies the existence of some measurements, it is only natural to expect that the first step—the setup of the experiment—poses an experimental design problem. The fact that inferences are to be made on the basis of the experimental results leads in turn to a problem in estimation and decision theory. For a successful experiment one should identify the factors affecting the measurements and make sure that the designed experiment will yield all possible combination treatments. Once the measurements are available they are arranged in a rectangular array which is basically a pictorial summary description of the experiment and the measurement process. Usually the factors in an experiment will be varied within a certain range of values which are of interest to the person who conducts the experiment. These variations constitute the levels for each factor. To obtain meaningful results we must have several observations (and even better, an equal number) at each level for all possible combinations. When more than one factors enter the problem, a setup such as that described above is called a complete factorial experiment. An example of such an experiment with three factors A, B, and C at I, J, K levels respectively is shown in Table 6. The simplest entity in an ANOVA table is the observation. Each observation is indexed in the following manner: one index for each factor plus one index which denotes the order of the observation within the M observations performed at each level combination (we tacitly assumed an equal number, M, of such observations in all levels). The next entity which is of interest to us is the set of the M observations for each treatment. They are easily identified in the table since they all have the same index values for the indices associated with the factors. This entity is called a cell. In the simplest setup there will be only one observation in each cell, $(M = 1)$.

There are two ways of interpreting the ANOVA table. In the most usual case we assume that given R cells with M observations in each one, each cell represents a random sample of size M drawn from R normal populations with identical variance σ^2. This means that we are only interested in the specific variations (levels) of the factors as entered in the experiment disregarding the fact that these variations may be only a subset of many more possible. The effects of these factors on the observations are therefore fixed with respect to the specific experimental setup, thereby the name of this ANOVA model: fixed effects or Model I ANOVA. In the second case the interpretation is similar to the above except that now we make the assumption that the selected levels constitute a randomly selected sample from a large normal

Table 6 Arrangement of Data for a Three-Factor Experiment
Factor A: I levels; Factor B: J levels; Factor C: K levels,
M observations per cell (treatment or level combination).

		Levels of A						
		1			...	I		
Levels of B		1	...	J	...	1	...	J
Levels of C	1	y_{1111} ⋮ y_{111M}	... ...	y_{1J11} ⋮ y_{1J1M}	...	y_{I111} ⋮ y_{I11M}	... ...	y_{IJ11} ⋮ y_{IJ1M}
	⋮		⋮		...		⋮	
	K	y_{11K1} ⋮ y_{11KM}	... ...	y_{1JK1} ⋮ y_{1JKM}	...	y_{I1K1} ⋮ y_{I1KM}	... ...	y_{IJK1} ⋮ y_{IJKM}

population of realizations of the considered factors, with an associated variance σ_R^2. As it is probably obvious by now, the names associated with this interpretation are: Random effects or Model II ANOVA. For the sake of completeness we mention here that it is possible to have an experiment where both fixed and random effects may be present simultaneously. This setup is usually referred to as the Mixed Model ANOVA. The next step, irrespective of which model we are using, is the estimation problem. Since this part is mainly computational we will restrict ourselves to discussing the procedure which should be followed for the example shown in Table 6. A generalization of the method for factorial designs of higher order than three will then be obvious without need for a general and involved presentation.

The estimation process is based on the Gauss-Markoff theorem and it involves the computation of a number of "sums of squares" (SS) of deviations about various pivot values which will be explained in the following. These sums of squares will be used then in the computation of the "mean squares" (MS) which form the input of the decision theory problem. It can be easily verified that for the setup of Table 6 the total number of observations is given by the product IJKM, and the given set of data is said to span an IJKM-dimensional space. Denote by $\{y_{ijkm}\}$ the set of observations, y_{ijkm} being the mth observation in the i,j,k "treatment combination" or the cell with "coordinates" i, j, k in Table 6. Let u_{ijk} be the expected value of the measurement on this combination of factors and $\{e_{ijks}\} \sim N(0, \sigma^2)$ (independently) the random noise component of y_{ijks}. Our linear hypothesis, which constitutes the basis of the estimation is:

$$\Omega: \begin{cases} y_{ijks} = u_{ijk} + e_{ijks} \\ \{e_{ijks}\} \sim \text{independently } N(0, \sigma^2) \end{cases}$$

Under Ω each u_{ijk} is modeled as follows:

$$u_{ijk} = \mu + a_i^A + a_j^B + a_k^C + a_{ij}^{AB} + a_{jk}^{BC} + a_{ik}^{AC} + a_{ijk}^{ABC}$$

where μ is the "grand mean" of the population, that is, the average of all the cells' u's. In order to avoid cumbersome notation involving multiple summations, averaging over a subscript is denoted by replacing the corresponding index by a dot. For example, $\mu = u_{...}$ and $\hat{u}_{ijk} = y_{ijk.}$. It turns out that under Ω, an unbiased estimate of μ is the average of all the observations, i.e., $\hat{\mu} = y_{....}$. The remaining components in the expression for u_{ijk} are the effects of the factors (a_i^A, a_j^B, a_k^C) and their "interactions" (a_{ij}^{AB}, a_{jk}^{BC}, ..., a_{ijk}^{ABC}), the effects, that is, which exist due to the simultaneous operation of these factors. The interactions cannot be attributed to a single factor but only to their coexistence. If no interactions exist, then the factors are said to be "additive." In a sense there is no correlation among them and each operates independently of the others. When the number of observations in each cell is the same, as is the case here, the observational space $\mathcal{L}$ can be decomposed into $2^F + 1$ (provided $M > 1$) mutually orthogonal subspaces with a consequent decomposition of the total sum of squares into simpler sums. The dimensions of these subspaces are the degrees of freedom (DF) associated with the corresponding SS. The estimates of the various effects and interactions are obtained from combinations of averages over different indices depending on which component we are estimating. For example,

$$\begin{aligned}
\hat{\mu} &= y_{....} \\
\hat{a}_i^A &= y_{i...} - y_{....} \\
\hat{a}_{ij}^{AB} &= y_{ij..} - y_{i...} - y_{.j..} + y_{....} \\
\hat{a}_{ijk}^{ABC} &= y_{ijk.} - y_{ij..} - y_{i.k.} - y_{.jk.} + y_{i...} + y_{.j..} + y_{..k.} - y_{....}
\end{aligned}$$

Following these example deviations, one can form all the $\hat{a}$ components involved in the expression for $\hat{u}_{ijk}$ and thereby compute their SS's. A summary of the computational procedure is shown below.

Space	Source	Spanned By	Dimension (DF)	SS
$\mathcal{L}_\mu$	—	$\hat{\mu}$	1	$SS_\mu = IJKMy^2_{....}$
$\mathcal{L}_A$	A main effects	$\hat{a}_1^A, \ldots, \hat{a}_I^A$	I-1	$SS_A = JKM \sum_i (\hat{a}_i^A)^2$
$\mathcal{L}_B$	B main effects	$\hat{a}_1^B, \ldots, \hat{a}_J^B$	J-1	$SS_B = IKM \sum_j (\hat{a}_j^B)^2$
$\mathcal{L}_C$	C main effects	$\hat{a}_1^C, \ldots, \hat{a}_K^C$	K-1	$SS_C = IJM \sum_k (\hat{a}_k^C)^2$
$\mathcal{L}_{AB}$	AB interactions	$\hat{a}_{11}^{AB}, \ldots, \hat{a}_{IJ}^{AB}$	(I-1)(J-1)	$SS_{AB} = KM \sum_i \sum_j (\hat{a}_{ij}^{AB})^2$
$\mathcal{L}_{BC}$	BC interactions	$\hat{a}_{11}^{BC}, \ldots, \hat{a}_{JK}^{BC}$	(J-1)(K-1)	$SS_{BC} = IM \sum_j \sum_k (\hat{a}_{jk}^{BC})^2$
$\mathcal{L}_{AC}$	AC interactions	$\hat{a}_{11}^{AC}, \ldots, \hat{a}_{IK}^{AC}$	(I-1)(K-1)	$SS_{AC} = JM \sum_i \sum_k (\hat{a}_{ik}^{AC})^2$
$\mathcal{L}_{ABC}$	ABC interactions	$\hat{a}_{111}^{ABC}, \ldots, \hat{a}_{IJK}^{ABC}$	(I-1)(J-1)(K-1)	$SS_{ABC} = M \sum_i \sum_j \sum_k (\hat{a}_{ijk}^{ABC})^2$
$\mathcal{L}_e$	Error e	$\{y_{ijkm} - y_{ijk.}\}$	IJK(M-1)	$SS_e = \sum_i \sum_j \sum_k \sum_m (y_{ijkm} - y_{ijk.})^2$
$\mathcal{L}$	Total about grand mean		IJKM-1	$SS_T = \sum_i \sum_j \sum_k \sum_m (y_{ijkm} - y_{....})^2$

The decomposition of $\mathcal{L}$ into its direct sum of the above nine orthogonal subspaces gives rise to the following identity:

$$SS_T = SS_\mu + SS_A + SS_B + SS_C + SS_{AB} + SS_{BC} + SS_{AC} + SS_{ABC} + SS_e$$

Once the SS's have been computed, the MS's are readily obtained by dividing with the corresponding DF's.

At this point we can answer questions concerning the significance of each factor or any of their interactions. We can solve, that is, the decision theory problem. The various hypotheses to be tested can be set up and the tests are based on comparisons of MS ratios with theoretical values of an $\mathcal{F}$ distribution. We will not go into further detail for this final step since in our case we did not do any testing, the reason being we were only concerned with the effects in a relative sense, a fact which in our case could be inferred from the values of the SS's directly. Instead we will give a detailed description of how this method was implemented in our case and present the results that we obtained.

The factors which we considered to affect the precision of the baselines were their length, the mean orbital height, the a priori weights on the station coordinates, the elevation cut-off angle, and the orientation of the baselines in the grid. Since the grid design was fixed, the same baselines would be estimated from each solution irrespective of the selection of the rest of the factors. Due to the symmetry in the grid it was possible to select a set of baselines ranging from 25 to 300 km in length, at 25 km increments (12 levels of variations), for which we could cover all possible realizations of the rest of the factors, but mainly the baseline orientation variations. It can be verified from Fig. 1 that the cardinal directions of the grid are running almost parallel and perpendicular to the satellite groundtracks. We selected as our reference the side defined by stations 91 and 11, with respect to which we determined the orientation of the various baselines (i.e., any baseline parallel to 91 - 11 has bearing 0°, any one perpendicular to it 90°, etc.). To avoid overcomplicating the setup we grouped the baselines into three groups; baselines with bearing between 0° and 30°, between 30° and 60°, and between 60° and 90° (three levels of variations). Since we were only examining two types of orbits the mean orbital height had only two variation levels, 400 km and 1000 km. The nominal elevation cut-off angle for laser observations is usually 20°. The reason we were interested in an increased cut-off angle is that this would result in a simpler retro-reflector design with a considerable decrease in cost. Two levels of variations were therefore considered, the nominal 20° and the 35° option. The last factor to be considered, the a priori station coordinate weighting, had three levels of variation. A priori standard deviations of 1 m in the three Cartesian coordinates of all stations in the grid denote strong weighting. In the second variation the standard deviations are increased to 25 m denoting medium weighting, and the third variation is the solution with "quasi-minimum" constraints. The latter were applied on stations 11 (σ_x = 0.001 m, σ_y = 0.001 m), 15 (σ_y = 0.001 m, σ_z = 0.001 m), and 95 (σ_x = 0.001 m, σ_z = 0.001 m). For the above setup of five factors with their associated levels of variation we have a total of 432 treatments, half of which pertain to the low orbit, the other half to the high orbit. From the twelve solutions which are needed in order to cover all possible combinations of factor variations, we selected the standard deviations of baselines which fell in each of the treatment categories and formed the ANOVA Table 7. Only one observation per cell (M=1) was made since we were not interested in testing any hypotheses, but only in ascertaining the relative importance of the five factors as reflected on the corresponding SS's. A summary of the setup and the notation in Table 7 is the following:

Factor		Levels	
L	baseline length	12	25 km through 300 km incrementing by 25 km
H	mean orbital height	2	400 km (OSUL) and 1000 km (OSUH)
W	weighting schemes	3	A: $\sigma_X = \sigma_Y = \sigma_Z$ = 1 m all stations B: $\sigma_X = \sigma_Y = \sigma_Z$ = 25 m all stations C: "quasi-minimum" constraints (see text)
E	elevation cut-off angle	2	20° and 35° above the horizon
O	baseline orientation	3	0°-30°, 30°-60°, and 60°-90° (see text)

Using the data from Table 7 we computed the SS's and MS's following the standard computational procedure as described previously. The numerical results are presented in Table 8. It can be readily verified from this table that the elevation cut-off angle is responsible for most of the variation in the data, followed by the rest of the factors with considerably smaller effects. In order to determine the importance of the factors in the case of the two orbits (OSUL and OSUH) independently, two more tests were performed. For each test only half of the data (pertaining to the relevant orbit) were analyzed. Numerical results are shown in Table 9 for OSUL and Table 10 for OSUH. These tests revealed some interesting facts for the effect of the orbital height on the way the rest of the factors affect the baseline precision. At first the dominance of the elevation cut-off angle was reconfirmed as well as the fact that baseline length variations are second in line as far as the precision is concerned. For the low orbit, however, the orientation of the baselines is more important than the weighting scheme, while in the case of the high orbit it is the one with the least effect. This is reversed for the weighting scheme which in the case of the high orbit is almost as important as the baseline length factor. This

Table 7 ANOVA Table of the Baseline Standard Deviations
Table Values in Centimeters

ORBIT			LOW ORBIT											
TEST A			BASELINE LENGTH (km)											
O	E	W	25	50	75	100	125	150	175	200	225	250	275	300
		A	1.2	1.2	1.2	1.2	1.1	1.2	1.2	1.2	1.2	1.2	1.3	1.3
0°	20°	B	1.2	1.2	1.2	1.3	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
:		C	1.2	1.2	1.2	1.2	1.1	1.2	1.2	1.2	1.2	1.2	1.2	1.3
		A	1.9	1.9	1.9	1.9	1.9	2.1	2.1	2.1	2.0	2.2	2.2	2.2
30°	35°	B	1.9	1.9	2.0	2.0	2.1	2.3	2.4	2.4	2.4	2.7	2.8	2.9
		C	1.9	1.9	1.9	1.9	1.9	2.0	2.0	2.1	1.9	2.2	2.2	2.2
		A	1.1	1.1	1.2	1.2	1.2	1.3	1.3	1.3	1.3	1.5	1.5	1.5
30°	20°	B	1.1	1.1	1.2	1.2	1.3	1.4	1.4	1.5	1.6	1.7	1.8	1.8
:		C	1.1	1.1	1.2	1.2	1.2	1.2	1.3	1.3	1.3	1.4	1.5	1.5
		A	1.9	2.0	2.1	2.2	2.5	2.6	3.0	3.1	3.2	3.7	3.9	4.1
60°	35°	B	1.9	2.0	2.1	2.0	2.5	2.7	3.1	3.2	3.3	3.8	4.1	4.3
		C	1.8	1.9	2.0	2.1	2.3	2.4	2.7	2.7	2.8	3.2	3.4	3.6
		A	1.2	1.2	1.2	1.2	1.2	1.4	1.2	1.2	1.3	1.3	1.6	1.5
60°	20°	B	1.2	1.2	1.3	1.3	1.4	1.6	1.5	1.5	1.7	1.7	1.9	1.7
:		C	1.2	1.2	1.2	1.2	1.2	1.4	1.3	1.2	1.3	1.3	1.4	1.4
		A	1.8	1.8	2.0	2.0	2.0	3.3	2.2	3.3	2.4	2.5	4.2	3.8
90°	35°	B	1.8	1.9	2.1	2.1	2.2	3.5	2.3	2.5	2.7	2.9	4.4	3.9
		C	1.8	1.8	2.0	1.9	2.0	3.0	2.1	2.2	2.3	2.3	3.5	3.3

ORBIT			HIGH ORBIT											
TEST A			BASELINE LENGTH (km)											
O	E	W	25	50	75	100	125	150	175	200	225	250	275	300
		A	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.1	1.1
0°	20°	B	1.0	1.0	1.0	1.1	1.1	1.2	1.3	1.4	1.4	1.5	1.6	1.7
:		C	1.3	1.3	1.2	1.3	1.1	1.2	1.2	1.3	1.2	1.3	1.2	1.3
		A	1.4	1.4	1.4	1.5	1.4	1.5	1.5	1.6	1.6	1.6	1.7	1.7
30°	35°	B	1.4	1.5	1.7	1.8	2.0	2.2	2.5	2.7	3.0	3.2	3.3	3.8
		C	3.2	3.2	2.4	3.3	2.6	2.6	2.5	3.3	2.7	2.6	2.5	3.3
		A	0.8	0.9	0.9	0.9	0.9	0.9	1.0	1.0	1.1	1.1	1.2	1.2
30°	20°	B	0.8	0.9	0.9	1.0	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7
:		C	1.0	1.1	1.1	1.0	1.3	1.1	1.2	1.2	1.3	1.3	1.3	1.5
		A	1.4	1.4	1.4	1.5	1.7	1.7	2.0	2.0	2.1	2.4	2.5	2.6
60°	35°	B	1.4	1.4	1.5	1.6	1.9	2.0	2.4	2.6	2.8	3.0	3.2	3.5
		C	4.0	2.5	4.2	2.2	2.4	4.4	3.0	4.2	3.1	3.4	3.4	3.4
		A	0.8	0.8	0.9	0.8	0.9	0.9	0.9	1.0	1.0	1.1	1.2	1.1
60°	20°	B	0.9	0.9	1.0	1.0	1.1	1.2	1.2	1.5	1.5	1.5	1.7	1.7
:		C	0.9	0.9	1.0	0.9	0.9	0.9	1.0	1.3	1.3	1.3	1.3	1.4
		A	1.5	1.4	1.6	1.5	1.5	1.5	1.6	1.7	1.9	2.4	2.7	2.4
90°	35°	B	1.5	1.5	1.8	1.9	2.2	2.4	2.2	3.1	2.8	3.0	3.3	3.0
		C	1.8	1.7	2.0	1.7	1.9	1.7	2.1	2.0	2.4	3.3	3.1	3.2

Table 8 ANOVA Results for Test A

ANALYSIS OF VARIANCE.....TEST A

LEVELS OF FACTORS

L	12
H	2
W	3
E	2
O	3

GRAND MEAN 1.81250

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	* RMS *
L	37.42806	11	3.40255	1.84
H	3.48481	1	3.48481	1.87
LH	1.17963	11	0.10724	0.33
W	8.49847	2	4.24924	2.06
LW	6.24431	22	0.28383	0.53
HW	9.63088	2	4.81544	2.19
LHW	0.93968	22	0.04271	0.21
E	144.90750	1	144.90750	12.04
LE	10.15917	11	0.92356	0.96
HE	0.01815	1	0.01815	0.13
LHE	0.85519	11	0.07774	0.28
WE	2.45097	2	1.22549	1.11
LWE	1.03069	22	0.04685	0.22
HWE	4.78949	2	2.39475	1.55
LHWE	0.84218	22	0.03828	0.20
O	3.65625	2	1.82813	1.35
LO	5.25986	22	0.23908	0.49
HO	2.09088	2	1.04544	1.02
LHO	2.63968	22	0.11999	0.35
WO	2.34986	4	0.58747	0.77
LWO	2.30569	44	0.05240	0.23
HWO	1.63051	4	0.40763	0.64
LHWO	1.83394	44	0.04168	0.20
EO	3.94681	2	1.97340	1.40
LEO	2.68819	22	0.12219	0.35
HEO	0.88532	2	0.44266	0.67
LHEO	2.38634	22	0.10847	0.33
WEO	1.16431	4	0.29108	0.54
LWEO	1.66236	44	0.03778	0.19
HWEO	1.38329	4	0.34582	0.59
LHWEO	1.65005	44	0.03750	0.19
TOTAL	269.99250	431		

Table 9 ANOVA Results for Test B (OSUL)

ANALYSIS OF VARIANCE.....TEST B

LEVELS OF FACTORS

L	12
W	3
E	2
O	3

GRAND MEAN 1.90231

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	* RMS *
L	22.18718	11	2.01702	1.42
W	2.96037	2	1.48019	1.22
LW	1.56296	22	0.07104	0.27
E	70.34116	1	70.84116	3.42
LE	7.58051	11	0.68914	0.83
WE	0.19704	2	0.09852	0.31
LWE	0.09963	22	0.00453	0.07
O	4.40431	2	2.20241	1.43
LO	5.80519	22	0.26387	0.51
WO	0.24241	4	0.06060	0.25
LWO	0.30093	44	0.00684	0.08
EO	3.68926	2	1.84463	1.36
LEO	3.78074	22	0.17185	0.41
WEO	0.10296	4	0.02574	0.16
LWEO	0.13370	44	0.00304	0.06
TOTAL	123.88884	215		

Table 10 ANOVA Results for Test C (OSUH)

ANALYSIS OF VARIANCE.....TEST C

LEVELS OF FACTORS

L	12
W	3
E	2
O	3

GRAND MEAN 1.72269

SOURCE OF VARIATION	SUMS OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	* RMS *
L	16.42051	11	1.49277	1.22
W	15.16898	2	7.58449	2.75
LW	5.62102	22	0.25550	0.51
E	74.08449	1	74.08449	8.61
LE	3.43384	11	0.31217	0.56
WE	7.04343	2	3.52171	1.83
LWE	1.77324	22	0.08060	0.28
O	1.34231	2	0.67116	0.82
LO	2.09435	22	0.09520	0.31
WO	3.73796	4	0.93449	0.97
LWO	3.83870	44	0.08724	0.30
EO	1.14287	2	0.57144	0.76
LEO	1.29380	22	0.05881	0.24
WEO	2.44463	4	0.61116	0.78
LWEO	3.17870	44	0.07224	0.27
TOTAL	142.61884	215		

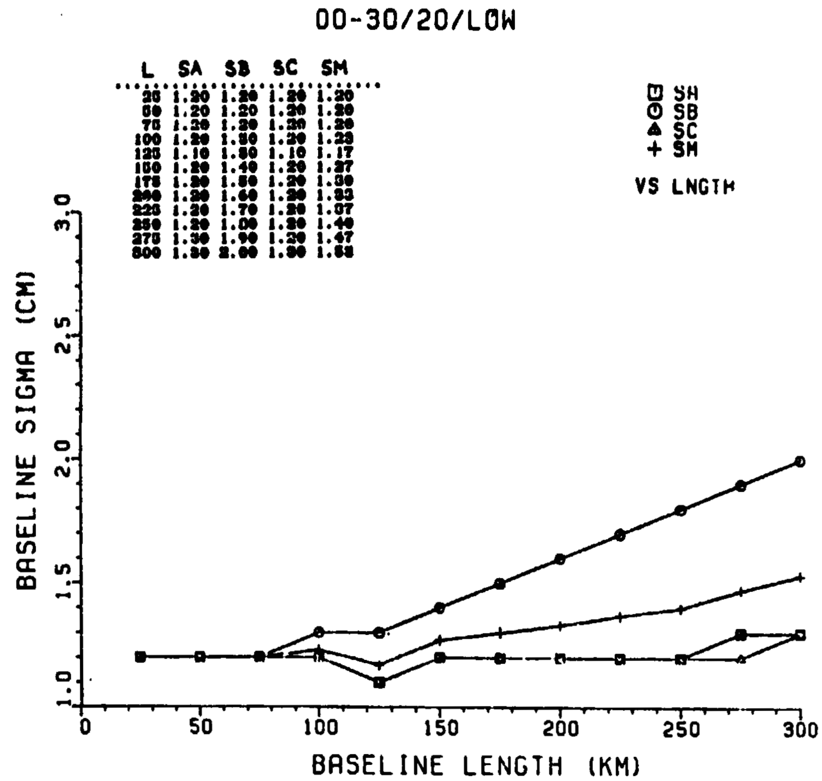
significant interaction of orbital height, weighting scheme and baseline orientation is also confirmed from the corresponding SS's (SS_{HW} and SS_{HO}) of Table 8 for the complete data set analysis. The large value for SS_{HW} in this table indicates that the applied a priori weights on the station coordinates will produce significantly different results depending on the orbital height of the spacecraft. Combining this with the results of Tables 9 and 10 we can say that the higher the orbit the larger the variation in the precision of the baselines due to identical variations of a priori station weights. These general remarks confirmed what we intuitively expected. In the following we examine the data for each of the factors individually, and we present whenever possible the theoretical explanation for the trends exhibited in them.

B. Baseline Precision Variations Due to Different A Priori Station Information

The plots which are presented and discussed in this section are graphical representations of the tabular values (Table 7) of the standard deviations and provide a more illustrative tool for the investigation of their variations. The quantities denoted by SA, SB and SC in these graphs correspond to the standard deviations obtained from the three weighting schemes A, B and C respectively. The quantity SM corresponds to the arithmetic mean of the SA, SB and SC. Figs. 5, 7 and 9 illustrate the variation of the standard deviations for both systems, for the three orientation classes and for a 20° elevation cut-off angle. Figs. 6, 8 and 10 give the same information when the elevation cut-off angle is increased to 35°. Figs. 11 and 12 finally were produced from the same set of data by averaging over the three classes of orientation.

From Figs. 5a, 7a, 9a and 11a one can conclude that the relative influence of the weights remains unchanged for differently oriented baselines, at least for the given geometry. We can reach a similar conclusion for the case of the high orbit from Figs. 5b, 7b, 9b and 12a. These conclusions seem to hold true for either choice of the elevation cut-off angle, as can be gathered from Figs. 6a&b, 8a&b, 10a&b, and 11b and 12b. Inferences, therefore, about the influence of the a priori station position information can be drawn from Figs. 11 and 12 alone.

(a)


00-30/20/LOW
L SA SB SC SM
25 1.20 1.20 1.20 1.20
50 1.20 1.20 1.20 1.20
75 1.20 1.20 1.20 1.20
100 1.20 1.30 1.20 1.23
125 1.10 1.30 1.10 1.17
150 1.20 1.40 1.20 1.27
175 1.20 1.50 1.20 1.30
200 1.20 1.60 1.20 1.33
225 1.20 1.70 1.20 1.37
250 1.20 1.80 1.20 1.40
275 1.30 1.90 1.20 1.47
300 1.30 2.00 1.30 1.53
SA
SB
SC
SM
VS LNGTH
BASELINE SIGMA (CM)
1.0
1.5
2.0
2.5
3.0
0
50
100
150
200
250
300
BASELINE LENGTH (KM)

(b)

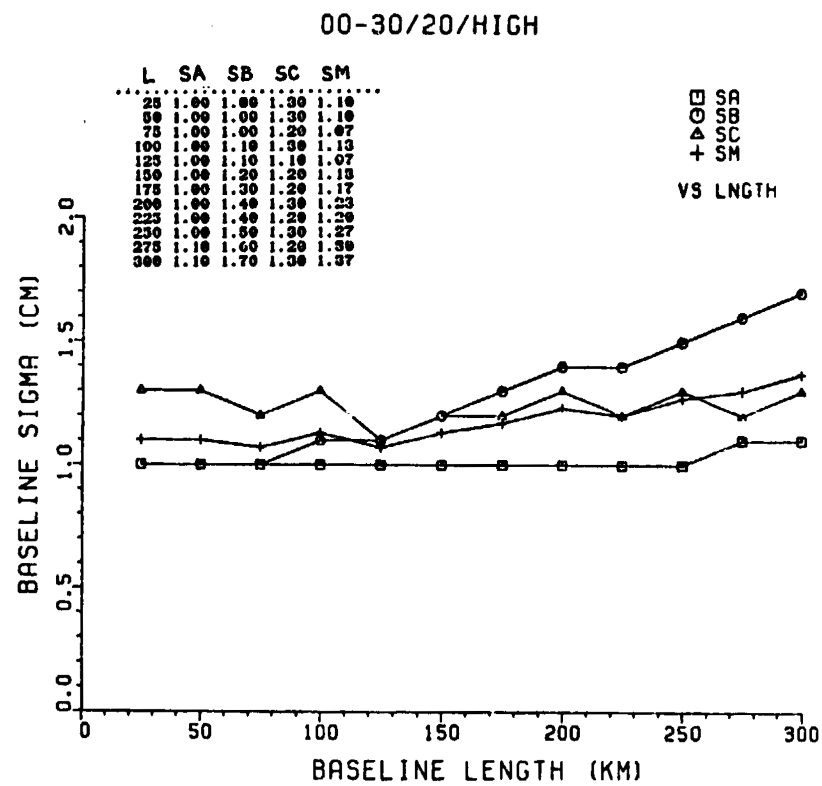
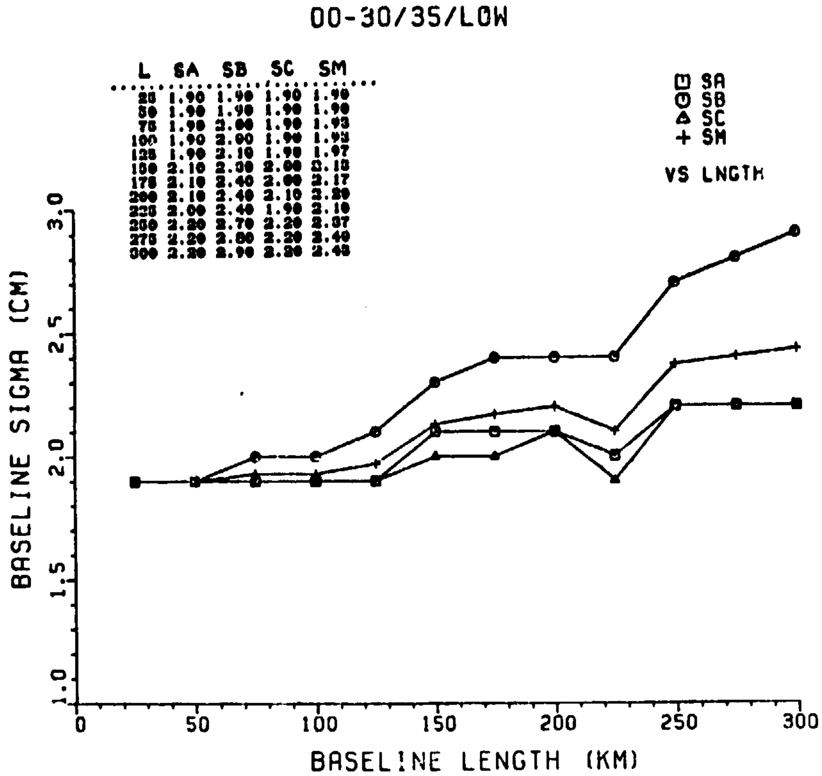
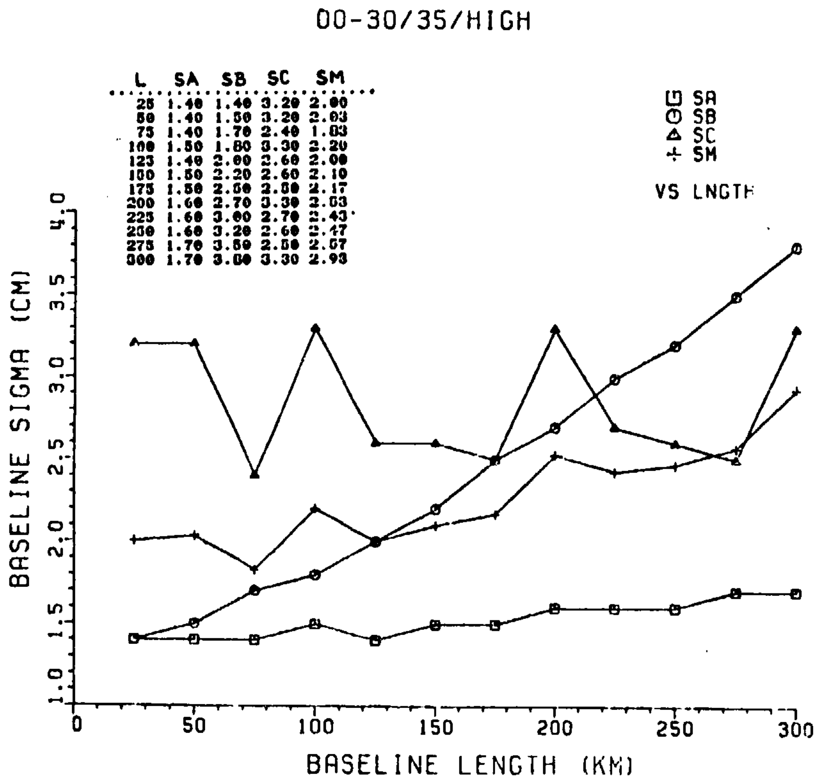

00-30/20/HIGH
L SA SB SC SM
25 1.00 1.00 1.30 1.10
50 1.00 1.00 1.30 1.10
75 1.00 1.00 1.20 1.07
100 1.00 1.10 1.30 1.13
125 1.00 1.10 1.10 1.07
150 1.00 1.20 1.20 1.13
175 1.00 1.30 1.20 1.17
200 1.00 1.40 1.30 1.23
225 1.00 1.40 1.20 1.20
250 1.00 1.50 1.30 1.27
275 1.10 1.60 1.20 1.30
300 1.10 1.70 1.30 1.37
SA
SB
SC
SM
VS LNGTH
BASELINE SIGMA (CM)
0.0
0.5
1.0
1.5
2.0
0
50
100
150
200
250
300
BASELINE LENGTH (KM)

Fig. 5



L	SA	SB	SC	SM
25	1.90	1.90	1.90	1.90
50	1.90	1.90	1.90	1.90
75	1.90	2.00	1.90	1.93
100	1.90	2.00	1.90	1.93
125	1.90	2.10	1.90	1.97
150	2.10	2.30	2.00	2.13
175	2.10	2.40	2.00	2.17
200	2.10	2.40	2.10	2.20
225	2.00	2.40	1.90	2.10
250	2.20	2.70	2.20	2.37
275	2.20	2.80	2.20	2.40
300	2.20	2.90	2.20	2.43

(a)

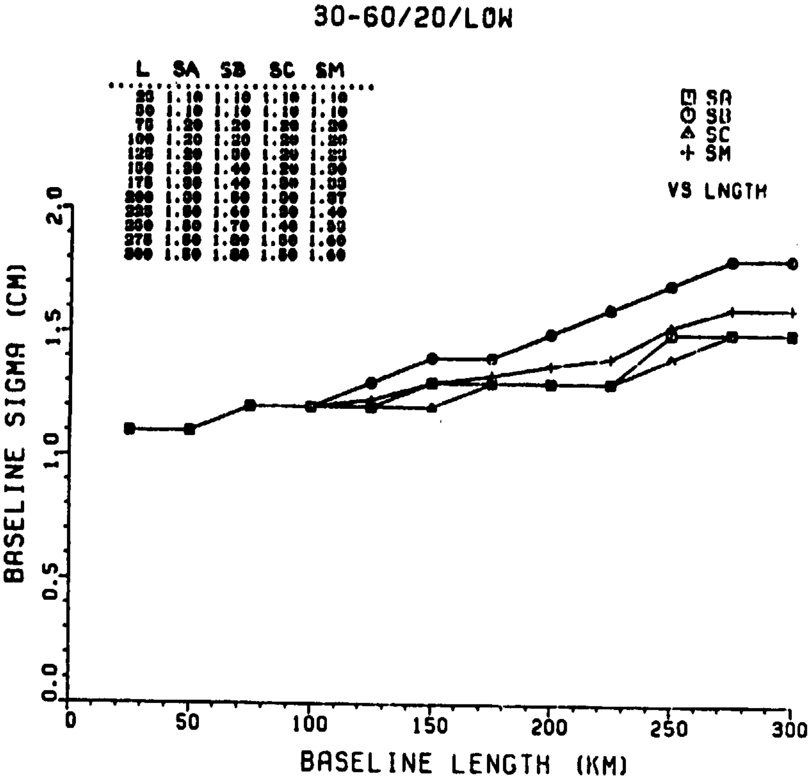


L	SA	SB	SC	SM
25	1.40	1.40	3.20	2.00
50	1.40	1.50	3.20	2.03
75	1.40	1.70	2.40	1.83
100	1.50	1.80	3.30	2.20
125	1.40	2.00	2.60	2.00
150	1.50	2.20	2.60	2.10
175	1.50	2.50	2.50	2.17
200	1.60	2.70	3.30	2.53
225	1.60	3.00	2.70	2.43
250	1.60	3.20	2.60	2.47
275	1.70	3.50	2.50	2.57
300	1.70	3.80	3.30	2.93

(b)

Fig. 6

(a)


30-60/20/LOW
L SA SB SC SM
SA
SB
SC
SM
VS LNGTH
BASELINE SIGMA (CM)
BASELINE LENGTH (KM)
0.0
0.5
1.0
1.5
2.0
0
50
100
150
200
250
300

(b)

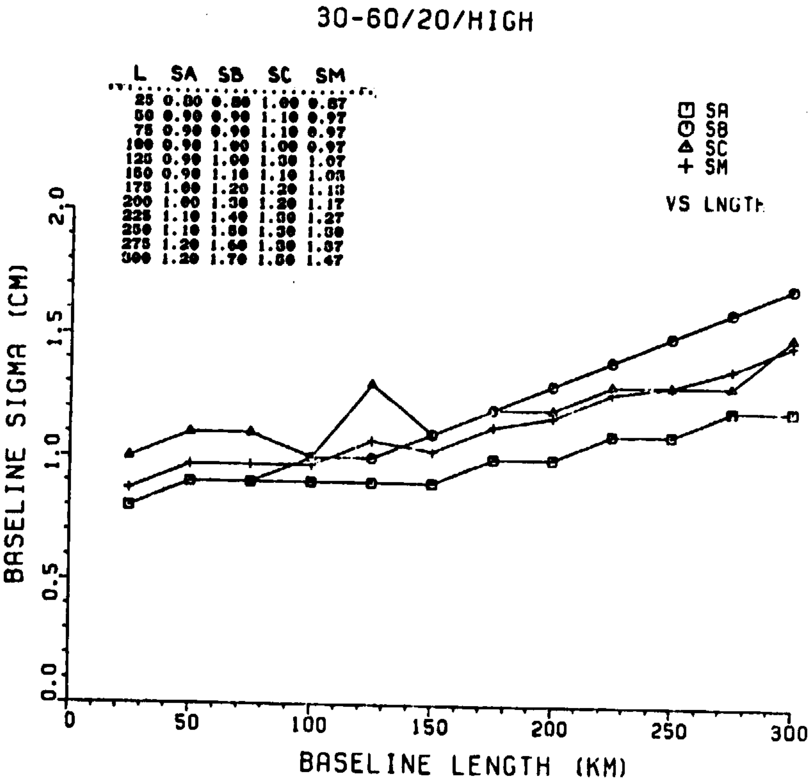
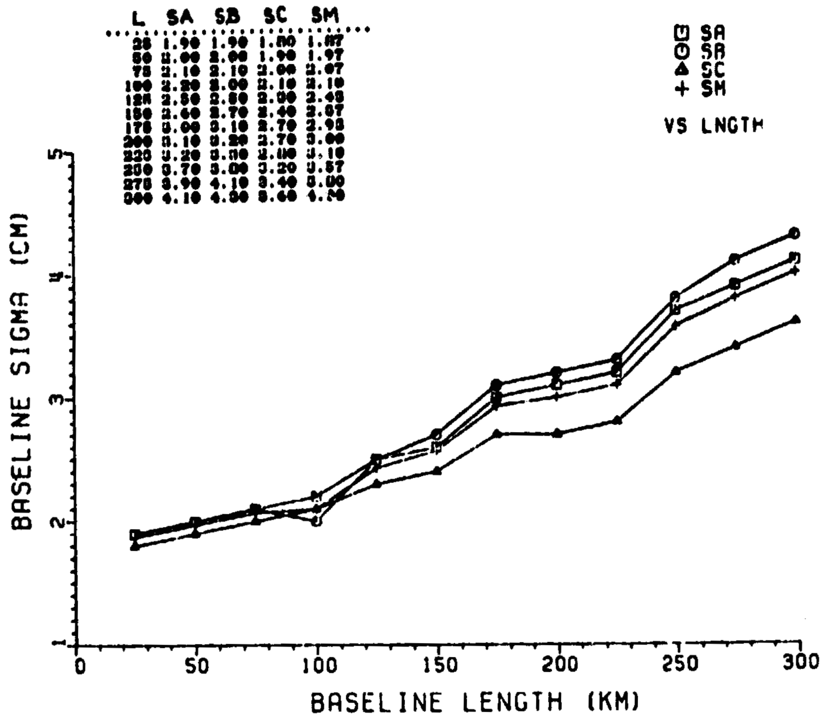

30-60/20/HIGH
L SA SB SC SM
SA
SB
SC
SM
VS LNGTH
BASELINE SIGMA (CM)
BASELINE LENGTH (KM)
0.0
0.5
1.0
1.5
2.0
0
50
100
150
200
250
300

Fig. 7

30-60/35/LOW

(a)



30-60/35/HIGH

(b)

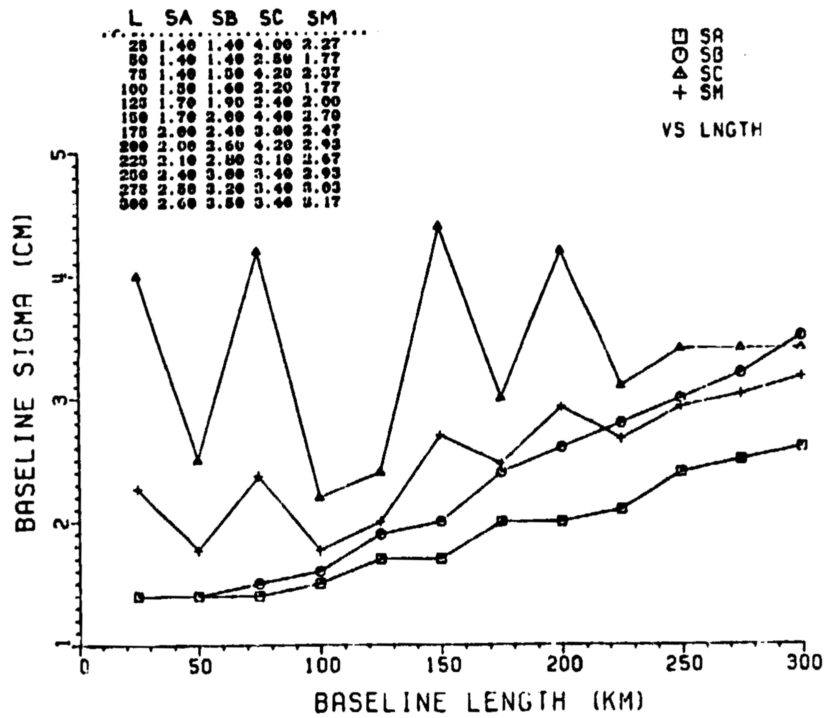
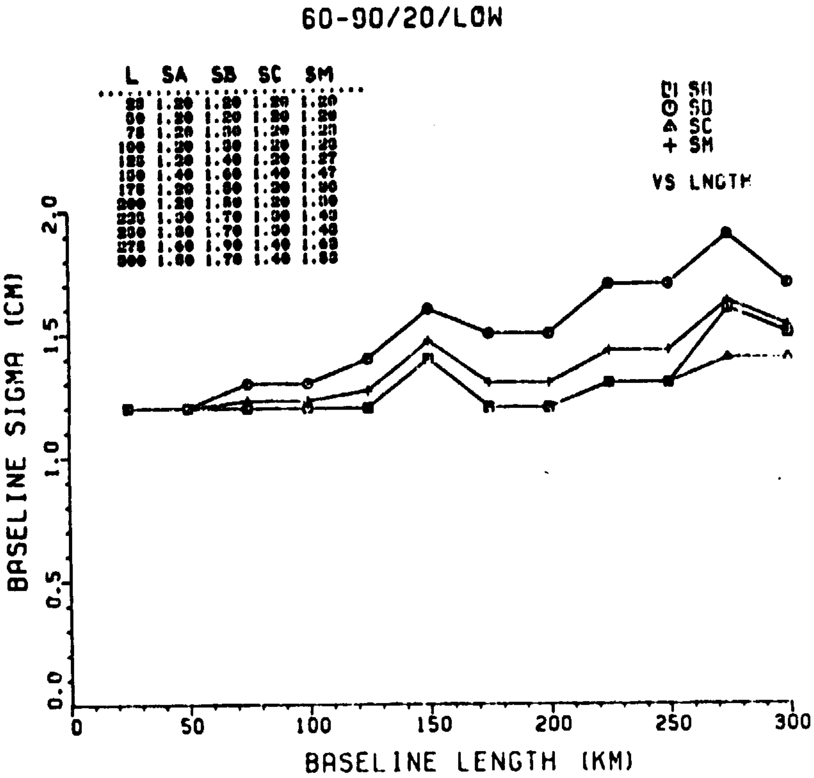


Fig. 8

(a)


60-90/20/LOW
L SA SB SC SM
SA
SB
SC
SM
VS LNGTH
BASELINE SIGMA (CM)
0.0
0.5
1.0
1.5
2.0
0
50
100
150
200
250
300
BASELINE LENGTH (KM)

(b)

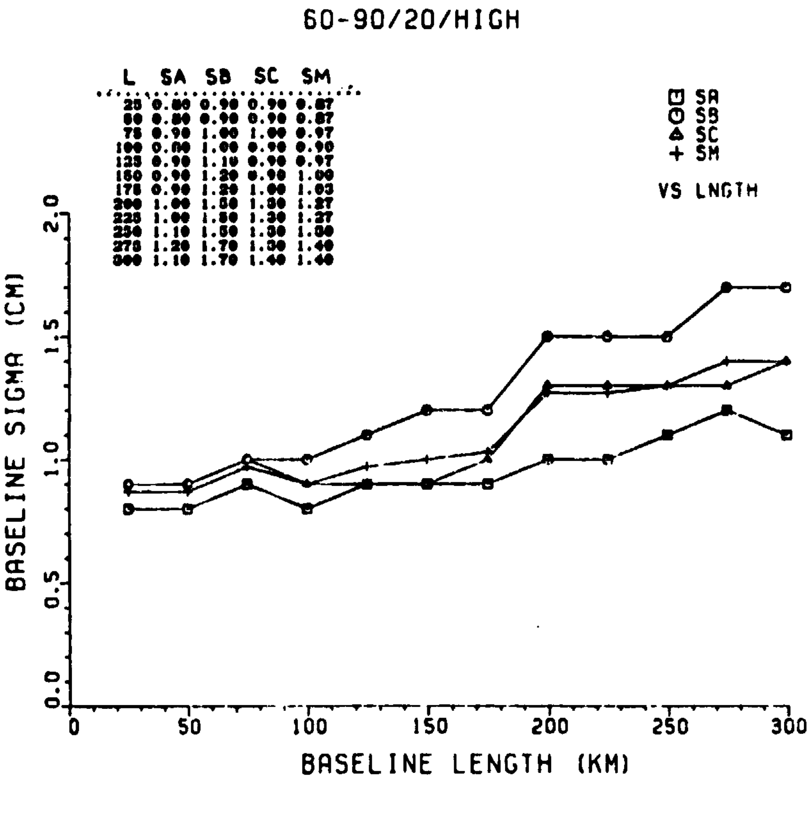

60-90/20/HIGH
L SA SB SC SM
SA
SB
SC
SM
VS LNGTH
BASELINE SIGMA (CM)
0.0
0.5
1.0
1.5
2.0
0
50
100
150
200
250
300
BASELINE LENGTH (KM)

Fig. 9

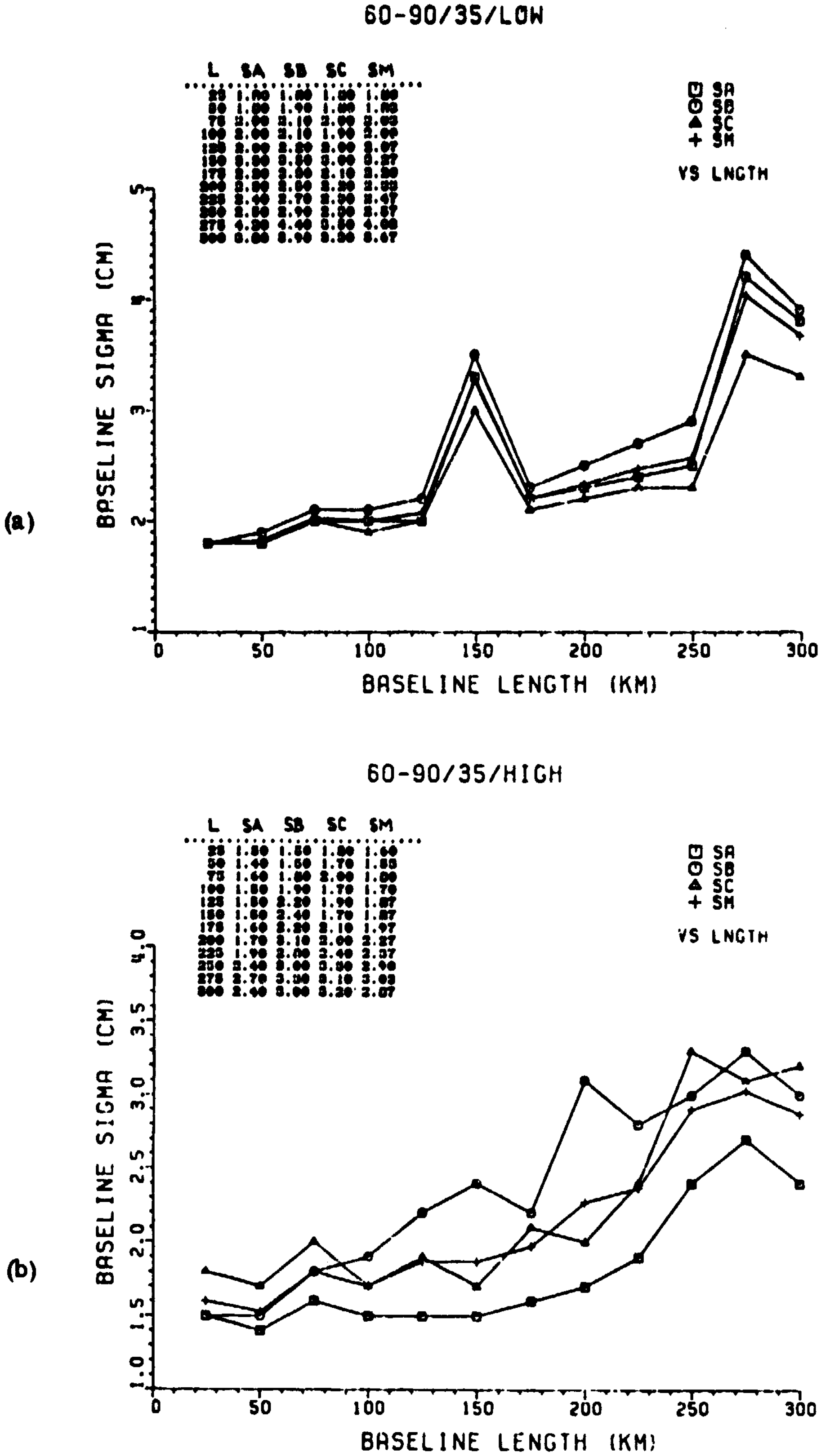
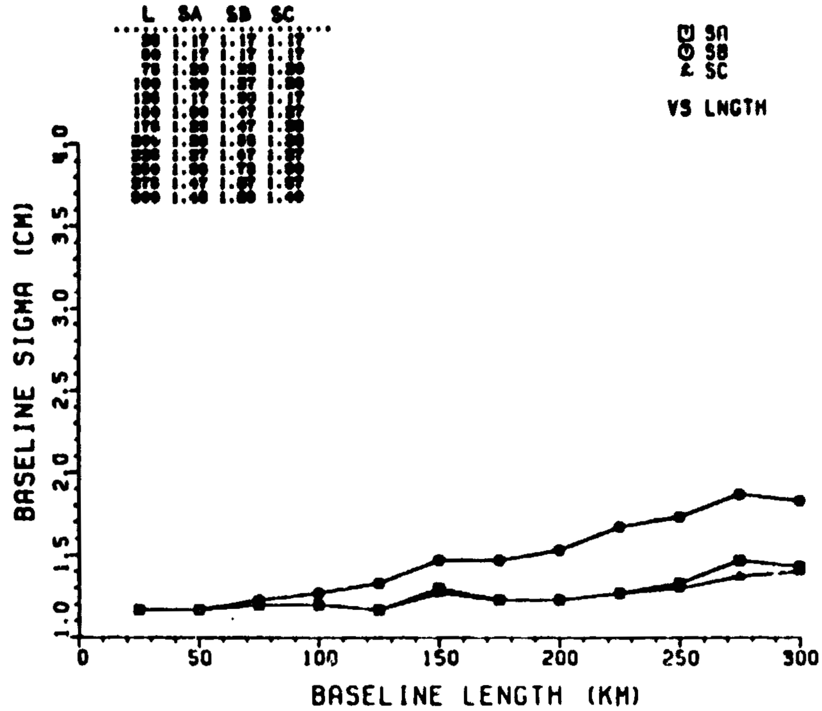

60-90/35/LOW
L SA SB SC SM
SA
SB
SC
SM
VS LNGTH
BASELINE SIGMA (CM)
BASELINE LENGTH (KM)
(a)
60-90/35/HIGH
L SA SB SC SM
SA
SB
SC
SM
VS LNGTH
BASELINE SIGMA (CM)
BASELINE LENGTH (KM)
(b)

Fig. 10

WEIGHTS : LOW /20

(a)



WEIGHTS : LOW /35

(b)

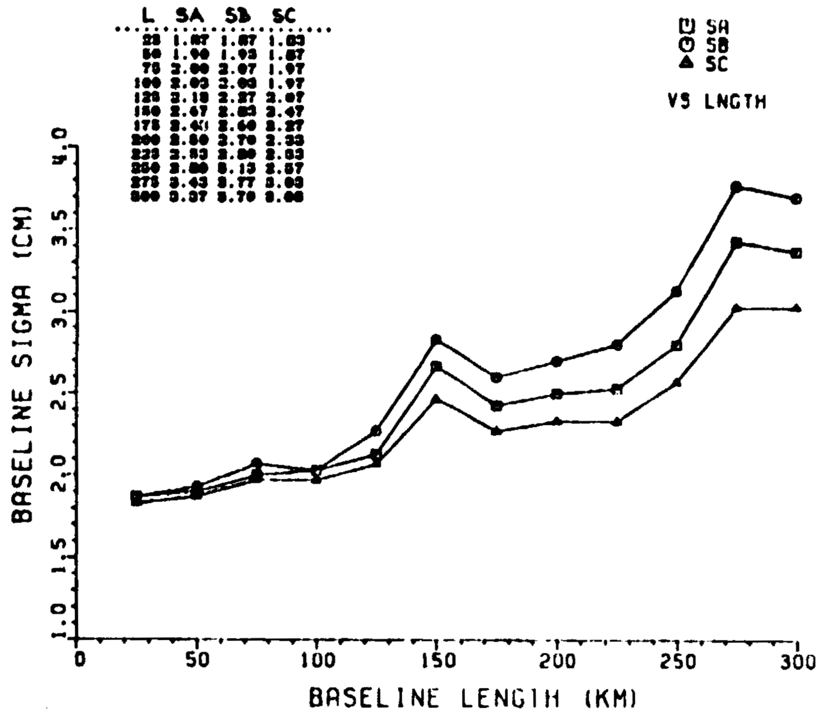


Fig. 11

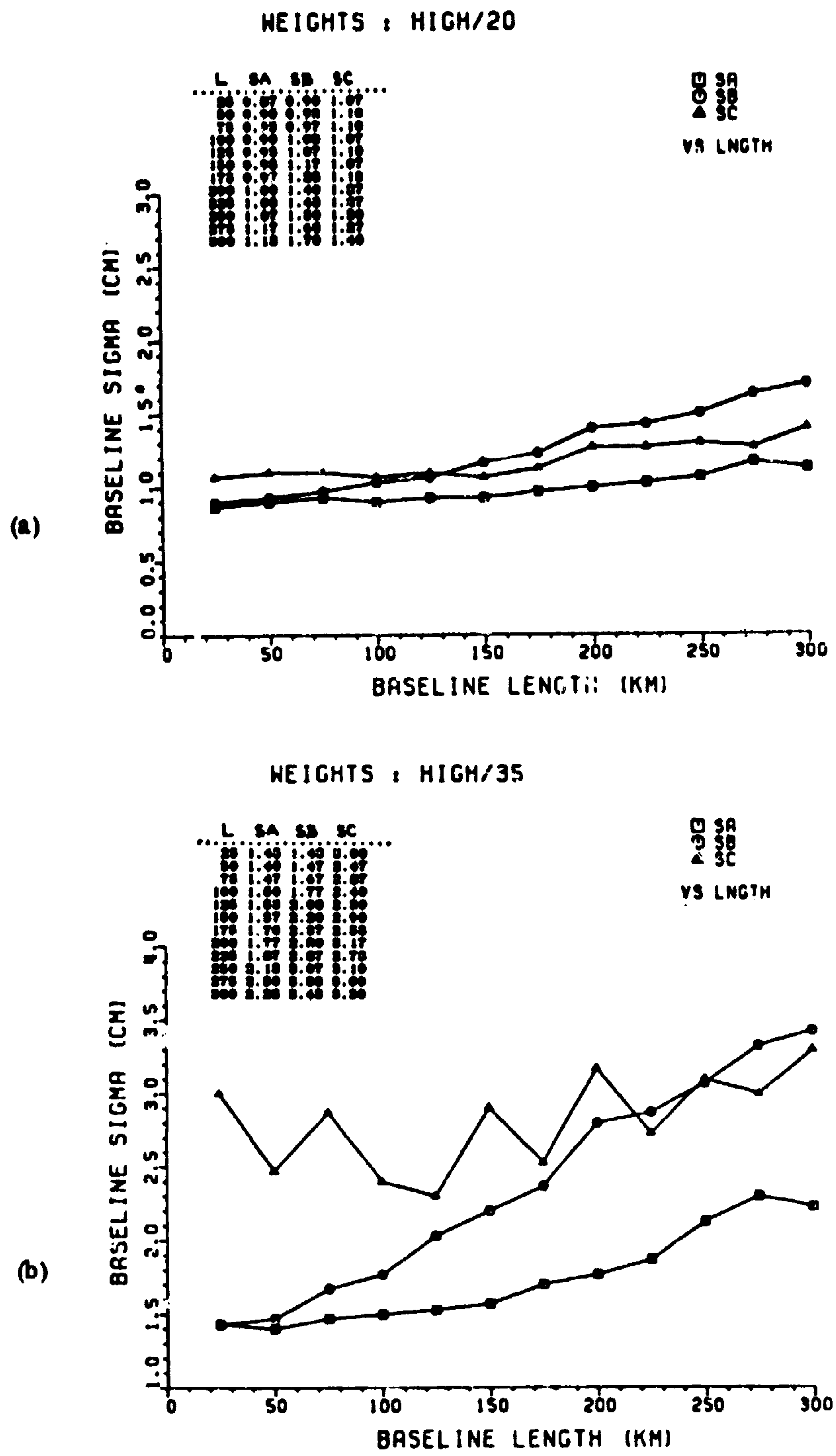

WEIGHTS : HIGH/20
L SA SB SC
SA
SB
SC
VS LNGTH
BASELINE SIGMA (CM)
BASELINE LENGTH (KM)
(a)
WEIGHTS : HIGH/35
L SA SB SC
SA
SB
SC
VS LNGTH
BASELINE SIGMA (CM)
BASELINE LENGTH (KM)
(b)

Fig. 12

Concentrating on these figures we can readily verify that the standard deviations in most of the examined cases have a nearly linear tendency to increase with increased baseline length, a fact which was intuitively expected. The only significant exception from this rule is the case of "quasi-minimum" constraints for the high orbit when the elevation cutoff angle is 35°. Possible causes behind this result will be examined later. Relatively speaking, in the case of the low orbit the results of the test with "quasi-minimum" constraints are always the best in terms of absolute magnitude, while for the high orbit they are of intermediate quality for the 20° cutoff angle test and fluctuate around 2.8 cm with no definite pattern when the cutoff angle is 35°. From Fig. 11a it is clear that for the low orbit the "quasi-minimum" constraints and the uniform weighting with a priori station standard deviations of 1 m produce almost identical results. From Fig. 11b we can see that these two schemes produce different results, the latter being about 10% above the first one in all cases. By changing the elevation cutoff angle, the number of observations was reduced equally for all three cases, A, B, and C. One would therefore expect an increase in the standard deviations proportional to the ratio $\sqrt{n_{20°}}/\sqrt{n_{35°}}$, with n being the total number of observations in each case. Such an increase would be manifested in the graphs as a shift in the "baseline sigma" scale and the relative location and shape of these graphs should remain the same. The reason they do not come out as such lies in the fact that even though all observations are of the same quality in terms of accuracy, they are not the same in terms of "geometric" quality. As it will be explained in a later section, the low elevation observations provide geometric strength in the solution which cannot possibly be compensated for by an equal number of medium or high elevation observations. Comparing the two figures (11a and b) one can verify that the loss of the low elevation observations affects the solutions with uniform weights more and within these the weaker estimates are obtained from the solution with the smaller weights. What is important from all these observations on the two graphs is that the internal structure of the observations plays a catalytic role on the performance of a given weighting scheme. The loss of observations from one case to the other is about 47% which warrants an increase of the standard deviations by a factor of about 1.37. Even if we consider this, the resulting estimates show an additional 40% deterioration which is the result of the missing low elevation observations as explained above.

These conslusions hold true for the high orbit also (Figs. 12a & 12b). The effect of the elevation cutoff angle change on the influence of the weighted constraints in the baseline standard deviations is even greater in this case. A 40% reduction in the observations implies an increase in the standard deviations by a factor of about 1.33; fromthe results, however, it is evident that even after we reduce the estimates to the same number of observations, the 35° elevation cutoff angle test yields poorer results by about 60% on the average. It is also interesting to note that strict uniform constraints produce rather optimistic standard deviations compared to a "quasi-minimum" constraints solution. As for the rather "irregular" results of the latter, in the high elevation cutoff angle test (Fig, 12b), the explanation lies in the missing observations rather than the applied constraints, the reason being that the constrained stations lie in the border of the grid and in the nominal case (20°) they collect a larger number of observations compared to the rest of the stations, especially at low elevations. With the cutoff angle increased to 35°, this advantage is lost and with it the strong "link" between the applied constraints and the parameters under estimation. This, in turn, produces a very loose system and therefore an ill-conditioned set of normal equations. An examination of the station coordinates' quality of recovery reveals some interesting facts. Although the coordinates are not estimable quantities, the application of the constraints changes their status to what is called "conditionally estimable." Since the constraints are identical for both elevation cut-off angles, a relative comparison of their standard deviations is not completely unjustifiable. The average results are shown in Table 11.

Table 11

Elevation Cutoff Angle	Average Standard Deviations (in cm) X	Y	Z
20°	1.0	1.2	1.2
35°	2.0	3.5	2.5

It is obvious from the above table that the loss of the low elevation observations has affected the three coordinates differently. The standard deviations in the Y component increased by a factor of three compared to those for X and Z which only doubled. This weakness in the Y component is also evident from the comparison of the correlations in the Y components among different stations in the grid which are approximately equal distances apart. In the 20° case these correlations show an increasing trend from 0.28 for stations near the northeast side of the grid, to 0.88 for stations at the southwest area. In the 35° case the same trend is also present, but in this case the rate of change is much steeper, from about 0.22 at the northeast side to almost 0.99 at the southwest. The increase in the correlations can only be due to the loss of the low elevation observations, since no other parameter was varied between the two tests. As for the actual trend, which is present in both cases, it is probably due to the way the Y component was constrained: the Y coordinates of stations 11 and 15. As can be seen from Fig. 2, both these stations lie at the northeasternmost side of the grid and one should therefore expect a weaker determination of the Y components as one moves away from the vicinity of stations 11 and 15. The combined effect of weak determination and high correlations in the Y components accounts for the irregular behavior of the baseline standard deviations. This view is also supported from the results depicted in Figs. 6b, 8b, and 10b. As it is seen from the first two figures, baselines with orientation angles from 0° to about 60° with respect to the direction defined by stations 11 and 91 are much more affected than those with orientation angles between 60° to 90°, the reason being that due to the specific location of the grid in space the baselines in the second group have significantly smaller components along the Y-axis compared to the baselines in the first group. The errors therefore propagate in a more favorable way for the second group of baselines (Fig. 10b).

In addition to the tests essential for the comparison of the different weighting schemes, some tests were conducted with different "quasi-minimum" constraints in connection to the rank deficiency problem in the short-arc mode. Description of this problem and a discussion of the results was presented in the section on rank deficiency and ill-conditioning in short-arc solutions.

C. The Effect of the Elevation Cut-off Angle on the Baseline Precision

From the ANOVA tests on both orbits (OSUL and OSUH), it is obvious that the elevation cut-off angle is the factor responsible for the largest variation in the baseline precision. Here we examine these variations in a more detailed manner, accounting for the fact that an increase in the cut-off elevation results in a simultaneous decrease in the number of observations in the problem. With the frequency of the observations held fixed for any choice of the cut-off elevation, the loss of observations is proportional to the percent reduction of the originally "observable" portion of a given pass.

Based on simple geometrical relationships between the orbit and the observing station, we can derive the following formula for the length of the observable arc, given the maximum elevation (E_{max}) that the satellite reaches with respect to the station and the minimum elevation (E_{min}) beyond which no observations are permissible:

$$\tan S = \frac{\sqrt{\sin[(E_{max} + E_{min}) + (P_1 + P_0)]\ \sin[(E_{max} - E_{min}) + (P_1 - P_0)]}}{\sin(E_{min} + P_0)}$$

where:

S is the length of the arc in geocentric angle measure

P_0 is the parallactic angle at the satellite, subtended by the geocentric radius of the station when the satellite elevation is equal to E_{min}

P_1 is the parallactic angle (as P_0) for satellite elevation E_{max}.

The parallactic angles are computed from the following formulae given the radius of the orbit a (assumed circular) and the mean earth radius a_e:

$$\sin P_0 = \frac{a_e}{a} \cos E_{min} \qquad \text{and} \qquad \sin P_1 = \frac{a_e}{a} \cos E_{max}$$

The arc lengths can be given in time measure also once the period of the satellite is computed from Kepler's third law:

$$p = 2\pi \sqrt{\frac{a^3}{GM}} \quad ; \qquad GM = 398\,603 \times 10^9 \ m^3/s^2$$

The periods for the low and the high orbits in our test are for example:

OSUL: $p \cong 92$ min; OSUH: $p \cong 105$ min

Based on the above, Table 12 gives some numerical examples for the arc length of overhead passes ($E_{max} = 90°$) for three different choices of the minimum cut-off angle. In addition to the angular length and the duration of each arc, the percent reduction is also given for comparison purposes.

Table 12

Cut-off Angle	Orbit			
	OSUL		OSUH	
	Arc Length	Duration	Arc Length	Duration
0°	40°	~ 10 min	60°	~ 18 min
	↓ 63%		↓ 50%	
20°	15° 75%	~ 4 min	30° 67%	~ 9 min
	↓ 33%		↓ 33%	
35°	10°	~ 3 min	20°	~ 6 min

The conclusion that can be drawn on the basis of these percentages is that the reduction in length, hence in observations too, is very significant (33%) for the two cut-off angles considered in our tests. The additional implication is that we are not decreasing our observations uniformly (such would be the case if we had decreased the frequency of the observations) but only at the beginning and the end of each pass. As pointed out already in the previous section, even if all our observations are of the same precision, their geometric quality depends on the relative positions of the observer and the target. It is rather simple to realize that when a distance is to be estimated indirectly, the best measurements to do are the distances between its end-points and a third point on the same line. The low elevation observations are the "third" points for our problem. It is conjectured in [Van Gelder, 1978] that each of these low observations contains as much geometric information as two observations in medium elevations.

To get an idea of the geometric quality of these observations, we have plotted in Fig. 13 the average standard deviations for our baseline sample for the two different cut-off angles. At first glance one might think that indeed the precision is [illegible] by a factor of two going from the 20° (S20 curve) to the 35° (S35 curve). We must consider though that the loss of the observations should be accounted for first and is depicted by the dashed curve. So this curve gives the expected precision if in the 20° angle case we had decreased (uniformly) our observations to as many as we had in the case with 35°. On the basis of this curve, the deterioration of the results due to purely worse geometry is about 40%. In order to overcome this we would have to increase observational frequency (already 10 pps) to unrealizable rates. It is, therefore, recommended that we include as many low observations as possible. This, of course, should be further examined when systematic effects (like measurement biases, atmospheric refraction model inadequacies, etc.) are included in future simulation studies. It is expected that the accuracy of these observations (although their precision may be the same) will in general be poor due to biases and unmodelled effects. One should therefore try to find the "golden cut" so that with proper weighting of these observations or even setting limits as to their number per pass, an improved geometry can be achieved with tolerable biases at the same time.

To these direct effects of the cut-off angle we must add its interaction with two other important factors. We already discussed the first one, the weighting schemes. In the case where "quasi-minimum" constraints are applied, it must be established beforehand that the stations which are to be constrained have enough

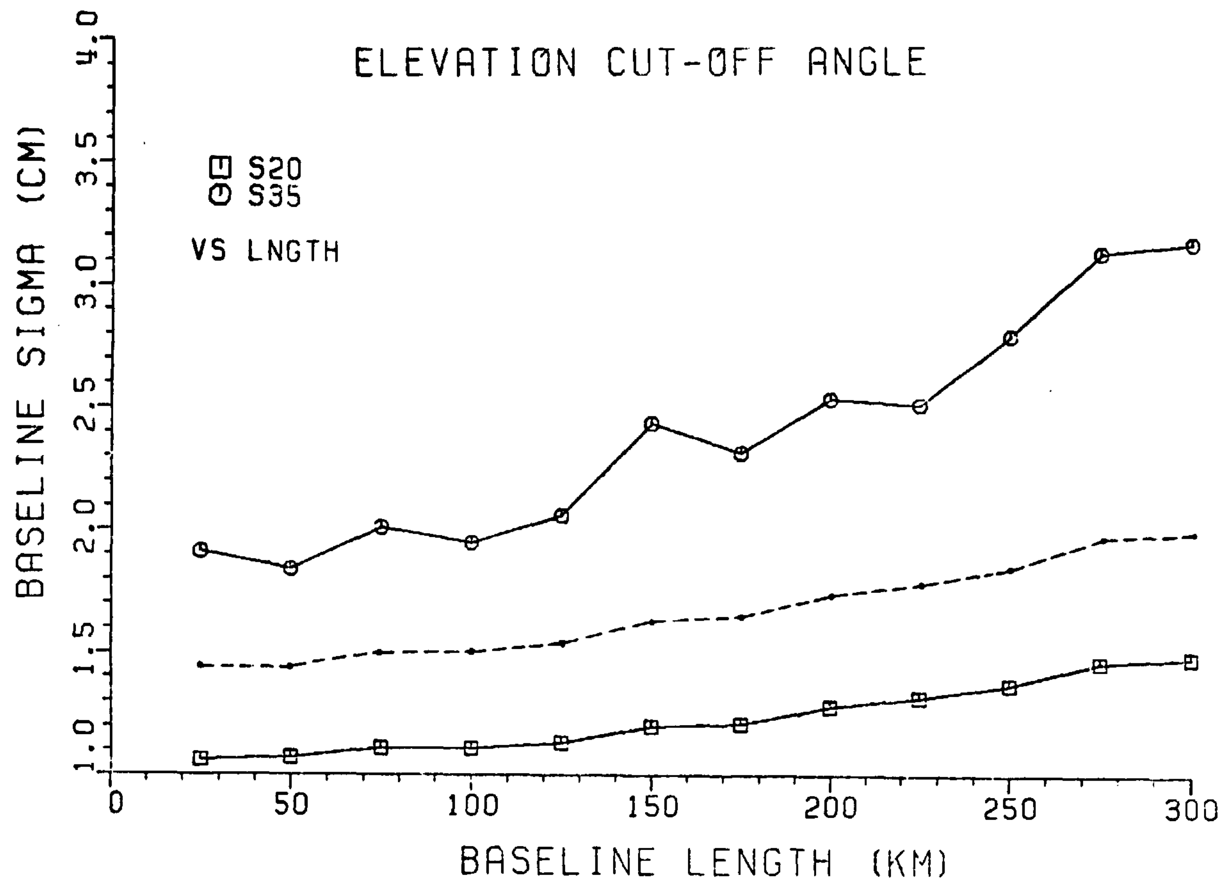


Fig. 13

and uniformly distributed observations on all passes in the solution. It should always be kept in mind that the constraints "flow" through these observations in order to determine the orbit, based on which the observations from the unknown stations determine their position.

The other factor which seriously interacts with the cut-off angle is the baseline orientation relative to the satellite passes. Again the reason lies behind the geometric quality of the low observations. More details will be given in the next section where the orientation factor is examined.

D. Variations of the Baseline Precision Due to Different Network-Satellite Pass Configuration

One of the factors which affect the precision of the recovered baselines in satellite ranging networks is the relative orientation of these baselines with respect to the satellite pass(es). Because of the dependence of this effect on the adopted cut-off angle, we have already given some hints on the source of the problem in the previous section. These remarks, however, were based on purely intuitive geometric considerations. In this section the problem is examined more systematically, and the use of a simple example will clarify the situation and provide some justification for the results of the numerical tests.

Our simple setup is shown in Fig. 14. A satellite pass lies on the plane defined by the axes X and Z. Disregarding time for the moment and denoting by

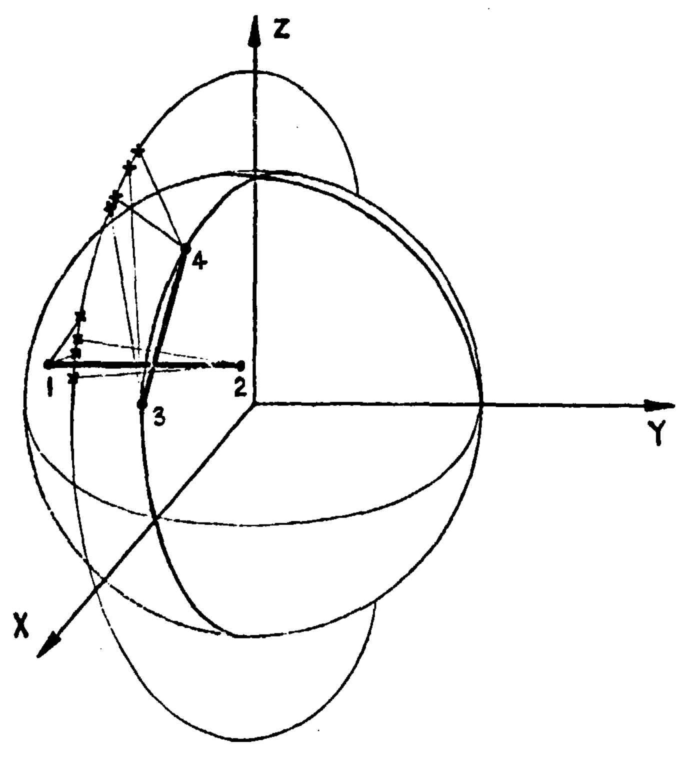


Fig. 14

subscripts i the i^{th} satellite point and j the j^{th} station, the geometric distance between them is

$$r_{ji} = [(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^{\frac{1}{2}}.$$

Now we add one more station in the picture, station k, which also observes the satellite at a different epoch (close to the i^{th}) and measure the range r_{ki}. From two sets of this type of observation we want to estimate the distance (baseline length) between stations j and k. For the sake of simplicity let us assume that the orbit is perfectly known, so the only unknowns will be the station coordinates (X_j, Y_j, Z_j) and (X_k, Y_k, Z_k). The design matrix of partial derivatives with respect to the parameters will typically look like the one below:

Parameter:	X_j	Y_j	Z_j	X_k	Y_k	Z_k
r_{ji}:	$-\frac{X_i - X_j}{r_{ji}}$	$-\frac{Y_i - Y_j}{r_{ji}}$	$-\frac{Z_i - Z_j}{r_{ji}}$	0	0	0
r_{ki}:	0	0	0	$-\frac{X_i - X_k}{r_{ki}}$	$-\frac{Y_i - Y_k}{r_{ki}}$	$-\frac{Z_i - Z_k}{r_{ki}}$

Examination of this matrix indicates the sensitivity of the system in each parameter under estimation. We consider two different cases of baseline-pass configuration. In the first case we examine the baseline 1-2 which is perpendicular to the plane of the pass. Since all satellite points have zero Y coordinates, the derivatives $\partial r_{1i}/\partial Y_1$ and $\partial r_{2\ell}/\partial Y_2$ are equal to Y_1/r_{1i} and $Y_2/r_{2\ell}$ respectively. Unfortunately, the ranges do not vary too much in a short interval such as a ten-minute pass and therefore neither do these partials. This set-up therefore is very insensitive with respect to Y_1 and Y_2, and such a solution would yield very poor estimates for these parameters. For this case the baseline length b_{12} is simply $|Y_1 - Y_2|$, so obviously the poor results will propagate in the determination of b_{12}. We now examine the other extreme case where the stations 3 and 4 define a baseline on the plane of the pass. In this case the derivatives for Y_1 and Y_2 are zero, since $Y_1 = 0 = Y_2$; so their determination is impossible. The baseline b_{34} is now given by

$$b_{34} = [(X_3 - X_4)^2 + (Z_3 - Z_4)^2]^{\frac{1}{2}}$$

and the result therefore will be independent of the estimates Y_1 and Y_2. Between the two extreme cases discussed above, the quality of the recovered baseline changes from poor to best as its orientation varies from perpendicular to parallel to the plane of the pass.

One must realize that these examples are indeed oversimplified, and in reality the situation is much more complicated. They are adequate enough, however, to illustrate how poor geometry can affect the results of equally precise observations. The regular grid design of the network which we investigated and the convenient orientation of the satellite passes with respect to its sides provided an excellent set-up for numerical tests. In Fig. 15 we plotted the average standard deviations from our standard sample of baselines, for the three different orientation classes which we considered as representative for this problem. We observe that in general the baselines which belong to the (0°-30°) and (60°-90°) orientation classes are better determined than these which lie in the middle. This, of course, happens

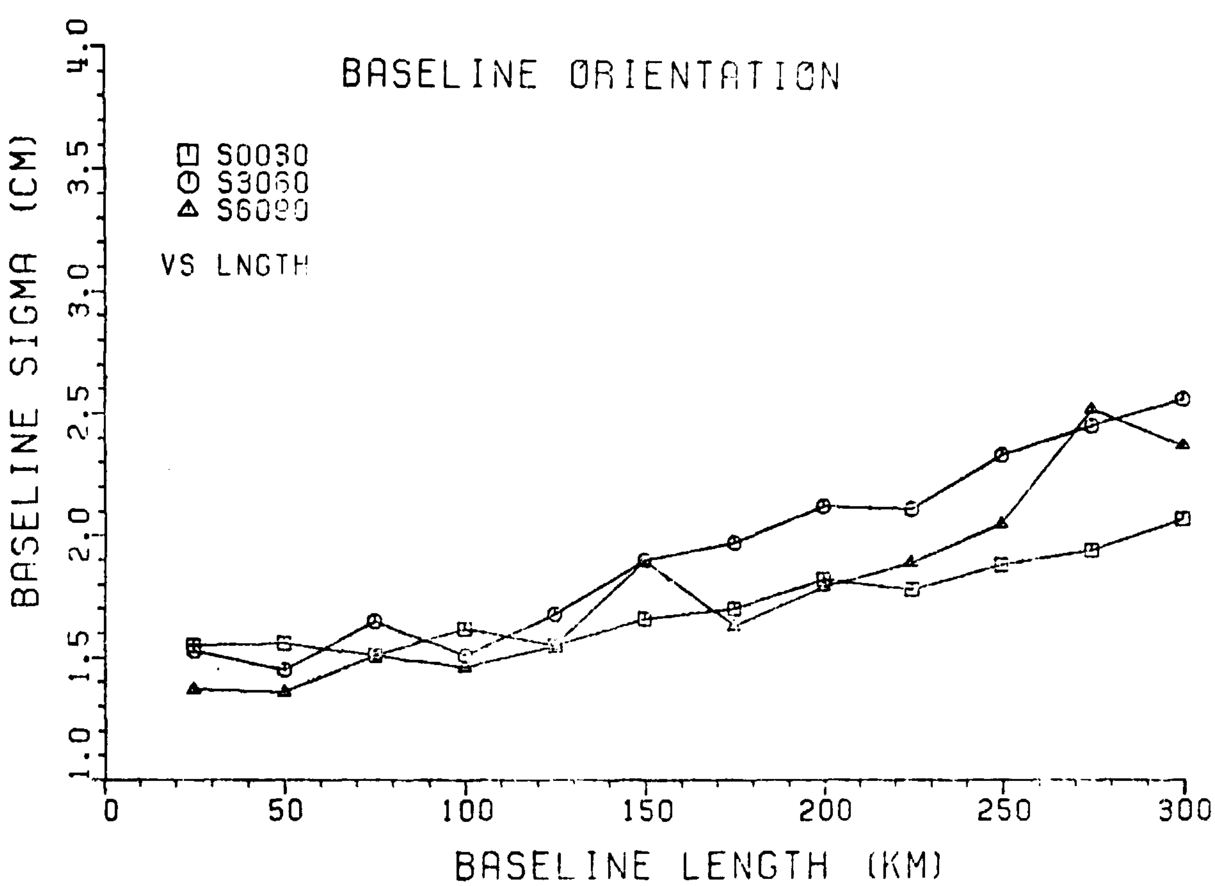


Fig. 15

because these sets of baselines are more "parallel" to the satellite passes than the others. A slight variation in the results for the (60° - 90°) class is most probably due to a slightly unbalanced collection of observations from ascending and descending passes.

Finally we must point out again that although relatively speaking the baselines which are nearly aligned with passes will have smaller standard deviations than the rest, their absolute quality will depend highly on the adopted cut-off elevation as already explained in the previous section.

E. Baseline Bias Due to "Erroneous" Geopotential Model

When we outlined the purpose of this investigation we stressed that we were mainly interested in the determination of the degree of dependence (or independence) of the baseline standard deviations on certain factors. In this sense we are unable toquote absolute numbers for the expected accuracy of the results except possibly for the component due to noise in the observations only. Average results for both orbits (OSUL and OSUH) are depicted in Fig. 16. Most systematic effects (e.g., refraction) have been disregarded throughout the course of study. It is known, however, that almost all models used to correct for these systematic effects are imperfect and an uncertainty is always attached to their results. It is therefore expected that our baseline accuracy will be affected by these uncertainties and in fact worsened. Results on the magnitude of these components of the total variance were recently given by [Smith, 1978].

In addition to the inflation of the total baseline variance, imperfect corrections for systematic effects introduce biases also in the actual baseline length estimates. We alread· ›inted out that in our investigation the only case where such effects were of concern to us was in deriving the baseline precision from an estimation process based on a different geopotential model from the one used to generate the observations. Various cases were rerun using GEM7 (3,1), GEM7 (8,8), and GEM9 (16,16); and the baseline precisions were compared to the original tests with GEM7 (16,16). In all cases the results were identical down to the millimeter level. We noticed, however, very significant differences in the recovered

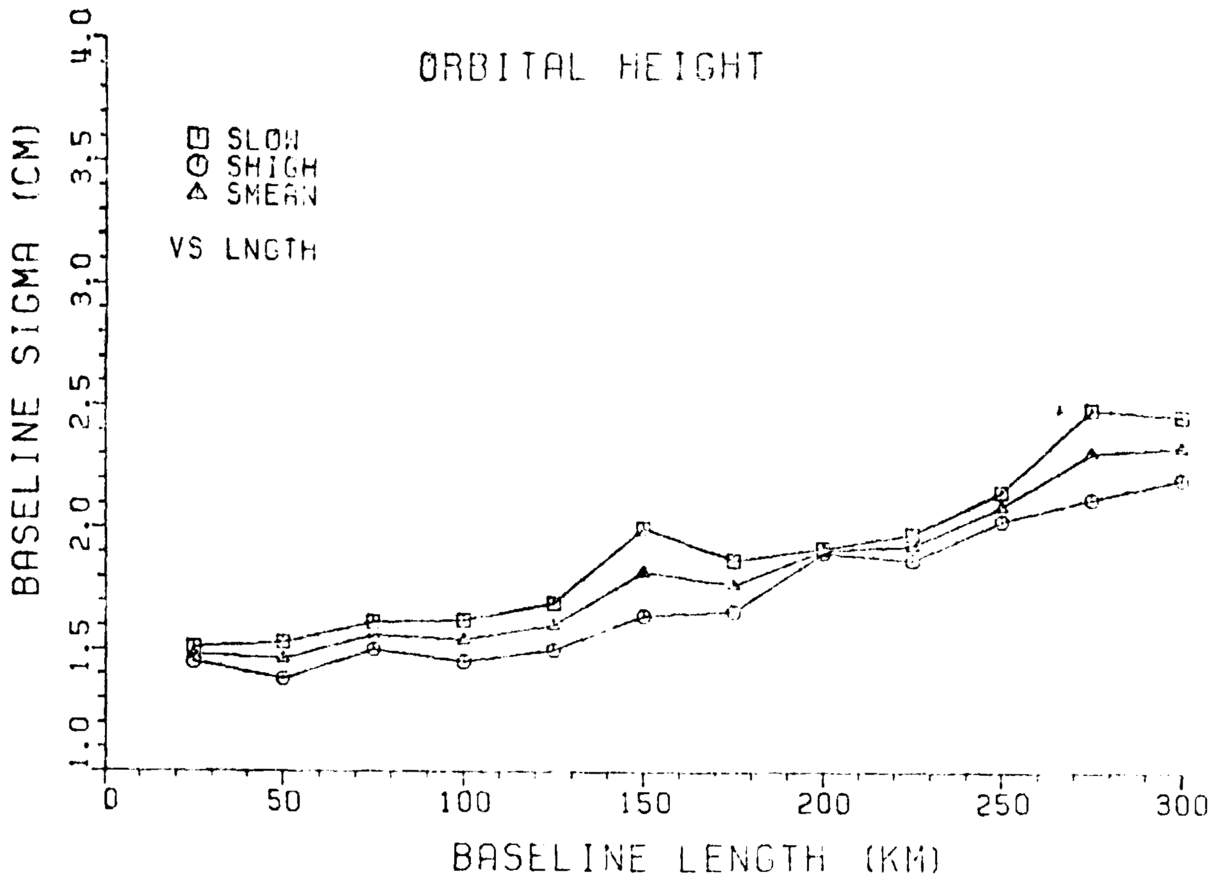


Fig. 16

baseline lengths and in the height differences between stations. This was not further pursued since it is a major problem in itself and another study quite different from this one should deal with it. We do want to stress that this must be cleared to satisfaction before concepts such as "repeatability" are employed in order to determine station motions from baseline length variations.

ORIGINAL PAGE IS
OF POOR QUALITY

7. SUMMARY: CONCLUSIONS AND RECOMMENDATIONS

The objectives of the conducted investigation were to determine the variation of the baseline precision in an SRS system due to various factors such as the a priori weighted constraints on station positions, the increase (or decrease) of the elevation cut-off angle, the relative orientation of the baselines and satellite passes, and the accuracy of the observations.

Through ANOVA tests for both of the orbits considered, it has been established that these factors produce significant variations when our accuracy goal is 1 cm. Aside from their direct effects, it has been shown that some of these factors interact with each other producing very undesirable results. The decrease of the elevation cut-off angle is shown to be the factor responsible for the largest variations which in this case are due both to the decrease in the number of observations as well as to the degradation of the geometric strength of the system as a whole. The effect of changing the a priori station information was examined in parallel with the problem of constraints and rank deficiency in the system. From numerical tests it is established that for the specific problem examined, the effective rank deficiency (inherent rank deficiency plus ill conditioning in the normal equations) is at least four and on the basis of theoretical considerations at most six. The necessity therefore for "quasi-minimum" type of constraints is clear and so is the fact that their number and arrangement in the network depends highly on the geometry of each individual problem. As far as Bayesian (biased) estimation techniques are concerned, it has been explained that no meaningful results will be obtained unless the a priori covariance matrix for the station positions reflects reality to a high degree of approximation. In view of the mixed emotions in the statistical world for the Bayesian estimation techniques, further detailed study of the theoretical basis as well as the appropriateness and the consequences of such techniques applied in geodetic problems must be undertaken. The usefulness of estimable parametrization of our problems was pointed out at several instances. This should be considered when future software is developed or the existing software is grossly modified. The geometric strength of the network-satellite passes configuration plays an important role in obtaining uniform results throughout the surveyed area. In the design of the ground networks therefore we should always consider the kind of satellite coverage that the specific area has (assuming that this is dictated by the adopted orbit which will not change drastically in the short time frame of the survey). Obtaining the optimum network in this sense will not be feasible in all cases since other factors (such as the direction of expected motion or the actual location of the area) will also impose restrictions on the design. A reasonable compromise should always be feasible though. In fact since most of the future potential users of the system have already indicated the primary areas of interest, it would probably be beneficial if larger simulations were conducted for all these areas. In these future simulation studies it is important that we include all known systematic error sources in terms of the best currently available models. For all study areas previous weather records must be examined and some realistic weather model must be developed for each one. It is rather awkward to expect that either all stations are visible or that none of them is due to unfavorable weather conditions. The quality, quantity, and frequency of ground collected weather data must be established and the sensitivity of the adopted atmospheric refraction model must be examined in terms of the resulting biases due to residual refraction. A decision must be taken with respect to the type of the laser to be used in the operational system and instrumental biases from laboratory calibrations should be included in the future simulations. The problem of the geopotential model used in these simulations seems to be much more complicated than what was originally expected. Although the precision of the results is insensitive to any changes in the assumed model, the resulting biases may be orders of magnitude larger than the quoted standard deviations. Although this investigation was not concerned with the accuracy of the baseline lengths but rather with their precision, some of the tests which we performed using different goepotential models indicate that a more detailed examination of the problem is in order. The argument of "repeatability" which is so much used in recent publications on SRS must be re-examined. We do not actually know whether slow processes (such as strain accumulation) will leave the local field unchanged between regular resurveys. It is dangerous therefore to make such an assumption because the results may show motions which have nothing to do with reality. In view of the use of this system by scientists in different disciplines, a warning must be given along with the results as to their suitability for the various applications. Attempts to establish strain models on the basis of the changes in the Cartesian coordinates of the ground reflectors have already taken place. In order to associate accuracy estimates with such a model we must first prove that strain is an estimable parametric function of the nonestimable coordinates. If this proves to be true, then such a treatment can be justified provided the coordinates are obtained from the proper estimation process (e.g., inner constraints least squares adjustment).

We hope that the results obtained herein and those which are to come from the proposed further investigations of the system will provide sound arguments for the appropriateness and the capabilities of a satellite laser ranging system.

REFERENCES

Agreen, W.R. and D.E. Smith. (1973). "A Simulation of the San Andreas Fault Experiment," Preprint X-592-73-216, Goddard Space Flight Center, Greenbelt, Maryland.

Blaha, G. (1971). "Inner Adjustment Constraints with Emphasis on Range Observations," Dept. of Geodetic Science Rept. No. 148, The Ohio State Univ., Columbus.

Cohen, S.C. and G.R. Cook. (1979). "Determining Crustal Strain Rates with Spaceborne Geodynamics Ranging System Data," manuscripta geodaetica, 4, 3, 245-260.

Counselman, C.C. and I.I. Shapiro. (1978). "Miniature Interferometer Terminals for Earth Surveying," in Applications of Geodesy to Geodynamics, Proc. of the 9th GEOP Conference, Dept. of Geodetic Science Rept. No. 280, The Ohio State Univ., Columbus.

Gibbs, B.P. and E.M. Haley. (1978). "Error Analysis of the Spacelab Geodynamics Laser Ranging System," BTS-TR-78-52, Business and Technological Systems, Inc., Seabrook, Maryland.

Grafarend, E. and B. Schaffrin. (1974). "Unbiased Free Net Adjustment," Survey Review, XXII, No. 171.

Kahn, W.D., F.O. Vonbun, D.E. Smith, T.S. Englar and B.P. Gibbs. (1979). "Performance Analysis of the Spaceborne Laser Ranging System," NASA Technical Memorandum 80330, Goddard Space Flight Center, Greenbelt, Maryland.

Kumar, M. (1976). "Monitoring of Crustal Movements in the San Andreas Fault Zone by a Satellite-Borne Ranging System, " Dept. of Geodetic Science Rept. No. 243, The Ohio State Univ., Columbus.

Kumar, M. and I.I. Mueller. (1978). "Detection of Crustal Motion Using Space-Borne Laser Ranging Systems," Bulletin Geodesique, 52, 2.

Martin, T.V. and N.E. Mullins. (1972. "GEODYN Operations Description," III, Wolf Research and Development Corp., Riverdale, Maryland.

Martin, T.V., I.H. Oh, W.F. Eddy, J.A. Kogut. (1976). "GEODYN System Description," I, Part 2, Wolf Research and Development Corp., Riverdale, Maryland.

Mueller, I.I., B.H.W. van Gelder, M. Kumar. (1975). "Error Analysis for the Proposed Close Grid Geodynamic Satellite Measuring System (CLO-GEOS)," Dept. of Geodetic Science Rept. No. 230, The Ohio State Univ., Columbus.

Rao, C.R. (1973). Linear Statistical Inference and Its Applications, 2nd ed., John Wiley & Sons, New York.

Scheffé, H. (1959). The Analysis of Variance. John Wiley & Sons, New York.

SGRS Workshop Report. (1979). "The Report from the Workshop on the Spaceborne Geodynamics Ranging System," Institute for Advanced Study in Orbital Mechanics, University of Texas at Austin.

Smith, D. (1978). "Spaceborne Ranging System," in Applications of Geodesy to Geodynamics, Proc. of the 9th GEOP Conference, Dept. of Geodetic Rept. No. 280, The Ohio State Univ., Columbus.

Smithsonian Astrophysical Observatory Staff. (1977). "Study of a Close-Grid Geodynamic Measurement System," SAO Reports in Geoastronomy No. 5.

Van Gelder, B.H.W. (1978). "Estimability and Simple Dynamical Analyses of Range (Range-Rate and Range Difference) Observations to Artificial Satellites," Dept. of Geodetic Science Rept. No. 284, The Ohio State Univ., Columbus.

Vonbun, F.O., W.D. Kahn, P.D. Argentiero, D.W. Koch, K.J. Eng. (1975). "Spaceborne Earth Applications Ranging System SPEAR," Preprint X-920-75-299, Goddard Space Flight Center, Greenbelt, Maryland.

Wegener, A. (1966). The Origin of Continents and Oceans. Translated by J. Biram. Dover Publications, New York.

APPENDIX

The following quotations serve as an introduction to the subject of this Appendix.

> In this name [mathematical statistics], "mathematical" seems to be intended to connote rational, theoretical, or perhaps mathematically advanced, to distinguish the subject from those problems of gathering and condensing numerical data that can be considered apart from the problem of inductive inference, the mathematical treatment of which is generally relatively trivial. The name "statistical inference" recognizes that the subject is concerned with inductive inference. [Savage, 1972]

> Subjective expectations, valuations and preferences and their changes from person [to person] or in the course of time can and should be investigated by means of "objective" statistical methods. Trying to use them as a basis of statistics is like trying to gauge a fever thermometer by means of the patient's shivers. [van Dantzig, 1957]

> I shall call them "Bayes" probabilities because, frequency or not, they are the ones needed for insertion into Bayes's theorem. Savage argues that they are "personalistic", that is, they are a property of the individual and not of society. I would dispute this myself, and agree with Jeffreys in saying that in scientific questions they are objective. They only differ between individuals because the individuals are differently informed; but with common knowledge we have common Bayesian probabilities. We can ignore this side-issue in the present account. [Lindley, 1958]

> I have to comment on the sentence: "It is the greatest strength of the Bayesian argument that it provides a formal system within which any inference or problem can be described". I would like to turn it around and say: "It is the greatest defect of the Bayesian argument that it provides a formal system according to which you can believe what you wish and, furthermore, without any data". I believe the search for the sort of panacea envisaged is a false one, which is based on a total misunderstanding of the nature of language and the nature of knowledge. Here again I believe some homework is desirable. [Kempthorne, 1972]

Biased Linear Estimation

The problem of estimation arises in most sciences today and although some still consider it as the primary and exclusive area of interest to statisticians, many theoretical developments can be credited to other scientists as well. The problem can be briefly stated as the determination under certain conditions, of quantities which are related through a known functional relationship to a given set of observations. By conditions we mean a set of criteria that we establish in order to obtain optimum estimates in the sense implied by these criteria. A set of such criteria which is most often used in physical sciences and engineering is the following:

(1) linear

(2) unbiased,

(3) minimum variance estimators.

For a detailed discussion of these and other possible criteria, one may consult [Rao, 1973]. Since the choice of the criteria is more or less subjective and depends on the nature of the problem in hand, a good deal of controversy and confusion is evident from the literature whenever a comparison of different estimators is attempted. Most of this is due to variations of the second property—the unbiasedness. A number of statisticians, for instance, substitute this by "method consistency," a concept proposed by Fisher and Haldane. An even greater number of scientists follow the approach proposed by the late Professor L.J. Savage [1954 (1972 ed.) and 1962], whereby they attempt to uniformly minimize the variance of the estimators at the expense of unbiasedness. A deeper study of the problem reveals the origin of the problem as being the definition of probability adopted by each of the parties. An elegant and extensive presentation on the four different definitions of probability is given in [Papoulis, 1965]. They are as follows:

A. Axiomatic (measure theory),

B. Relative frequency (Von Mises),

C. Classical (favorable outcomes over total "equally likely" alternatives),

D. Measure of belief (inductive reasoning).

We will not argue here which of the above is most suitable as a definition, although we personally believe that the foundation of statistics lies in the axiomatic definition rather than any empirically or intuitively conceived definitions. The purpose of the following sections is to present the structure of certain biased estimators which are often used in our areas of interest and to compare them whenever possible with the unbiased estimators pointing out advantages of the one over the other. To the best of our knowledge, the first thorough examination of Bayesian estimation techniques in connection with geodetic problems is [Bossler, 1972].

The argument on which the application of biases estimation is based is that if we have prior information on our parameters we should use it in order to obtain more accurate a posteriori estimates. On the other hand, unbiasedness is essential in our problems if we want to make correct inferences from our results. We should always keep in mind that geodetic problems are mainly dynamic (the earth is not rigid!). The majority of our estimates therefore are estimates of the true averages of the parameters over the time span of the observational data set. These averages change with time and the use of biased estimation techniques does not guarantee that the introduced biases between different solutions will be the same. This being the case we can readily conclude that any model for the rate of change of the parameters in question will be biased too. Considering that the accuracy of our parameters can be improved by improving the quality of our observations—which is possible in most cases—it seems unreasonable to seek this improvement at the expense of unbiased results. We can probably justify the use of biased estimation at preliminary stages of our research when we only want to obtain a rough picture of the problem with a limited number of observations. When we proceed, though, to explore the fine structure of the problem, such techniques should be avoided at all costs.

Best Linear Estimation

Let us consider the model $(Y, X\beta, \sigma^2 V)$:

$$Y = X\beta + \epsilon ; \qquad \epsilon \sim N(0, \sigma^2 V)$$

and the estimable parameteric function $a^T\beta$ to be determined. For the development of the theory we need not make any assumptions on the ranks of X and V. It will simplify the derivations though if we assume that both are of full rank. A general treatment of the problem is given in [Rao, 1971]. From the above set up, we may find a best linear unbiased estimator (BLUE) of $a^T\beta$; here, however, we are interested in seeing the results obtained when we drop the restriction for unbiasedness. Assume that the new estimator of $a^T\beta$ is $b^T Y$. The estimation of b is based on the minimization of the mean square error (MSE) of $b^T Y$:

$$MSE(b^T Y) \equiv E[b^T Y - a^T\beta]^2 = \text{a minimum} \tag{1}$$

After some algebraic manipulation, we can reduce the above to the following form:

$$MSE(b^T Y) = \sigma^2 [b^T Vb + (b^T X - a^T)(\beta/\sigma)(\beta/\sigma)^T (X^T b - a)] \tag{2}$$

It is obvious that we have one equation containing both unknowns β and σ, and its minimization for b presupposes some knowledge for both of them or at least for their ratio β/σ. We have at least three choices to circumvent this problem:

a) Use some a priori value for β/σ which we base either on prior experience (if any) or on what seems reasonable to us,

b) Consider β as a random variable with a priori mean dispersion $E[\beta\beta^T]$. Note that we need not know the actual distribution of β, only the mean dispersion matrix is required. In this case (1) must be modified:

$$MSE(b^T Y \mid \beta) \equiv \underset{\beta}{E}\,\underset{Y}{E}\,[(b^T Y - a^T\beta)^2 \mid \beta] \quad \text{(Bayes' risk)} \tag{3}$$

c) Close inspection of (2) reveals the significance of the individual terms inside the brackets. The first term represents the variance while the second the bias squared. The matrix $\left(\frac{\beta}{\sigma}\right)\left(\frac{\beta}{\sigma}\right)^T$ then can be interpreted as the relative weight that we may associate with the bias compared to the variance. One can therefore select this matrix according to which of the two quantities is more important.

Irrespective of which of the above or any other conceivable approach we select for the determination of $\left(\frac{\beta}{\sigma}\right)\left(\frac{\beta}{\sigma}\right)^T$, the resulting equations are identical in appearance. To follow a unified approach we adopt the notation $\left(\frac{\beta}{\sigma}\right)\left(\frac{\beta}{\sigma}\right)^T = W$ with the proper interpretation implied. Minimization of the MSE leads to the following set of equations:

$$(V + XWX^T)\,b = XWa \tag{4}$$

$$b = (V + XWX^T)^{-1} XWa \tag{5}$$

$$b^T Y = a^T WX^T (V + XWX^T)^{-1} Y \quad \text{(Bayes Linear Estimator, BLE)} \tag{6}$$

$$b^T Y = a^T (X^T V^{-1} X + W^{-1})^{-1} X^T V^{-1} Y \tag{6'}$$

with:

$$\hat{\beta}* = (X^T V^{-1} X + W^{-1})^{-1} X^T V^{-1} Y \tag{7}$$

The superscript * denotes a biased estimate. We can write the analogous expression for the least squares estimate (LSE) of β as:

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \tag{8}$$

Comparing (7) with (8) we see that $\hat{\beta}^* \rightarrow \hat{\beta}$ as $W^{-1} \rightarrow 0$. The matrix W^{-1} however cannot be a null matrix except in special cases. If we choose for instance $W = \frac{1}{k^2} I$, then $W^{-1} = k^2 I$ so that as $k \rightarrow 0 \Rightarrow W^{-1} \rightarrow 0$ in the limit. In this sense we can state that the LSE is the limit of the BLE. The above choice of W leads to a special type of BLE, introduced by Hoerl and Kennard who called them "ridge estimators" due to their similar mathematical structure to methods used for ridge analysis of second-order response surfaces [Hoerl and Kennard, 1970a and 1970b]. These estimators will be discussed in more detail later.

There are two points of interest that we would like to examine, namely, the bias in $\hat{\beta}*$ and its mean error dispersion. Both will be compared to the LSE counterparts. We denote:

$$T = (X^T V^{-1} X + W^{-1})^{-1} \tag{9}$$

Then (7) is written as

$$\hat{\beta}* = T X^T V^{-1} Y \tag{10}$$

To obtain the bias:

$$
\begin{aligned}
E[\hat{\beta}* \mid \beta] &= E[TX^T V^{-1} Y \mid \beta] \\
&= TX^T V^{-1} X\beta \\
&= T(X^T V^{-1} X + W^{-1})\beta - TW^{-1}\beta \\
&= T T^{-1}\beta - TW^{-1}\beta \\
&= \beta - TW^{-1}\beta \qquad (11)
\end{aligned}
$$

In the LSE case, since $\hat{\beta}$ is unbiased,

$$E[\hat{\beta}] = \beta \qquad (12)$$

The BLE is therefore negatively biased. It can be shown that this bias is toward the origin in the sense that the norm of $\hat{\beta}*$ is smaller than that of $\hat{\beta}$.

We derive now the mean error dispersion matrix for $\hat{\beta}*$:

$$
\begin{aligned}
\underset{\beta}{E}\ \underset{Y}{E}[(\hat{\beta}*-\beta)(\hat{\beta}*-\beta)^T] &= \underset{\beta}{E}\Big\{\underset{Y}{E}[\hat{\beta}*-\underset{Y}{E}(\hat{\beta}*)][\hat{\beta}*-\underset{Y}{E}(\hat{\beta}*)]^T + \\
&\qquad [E(\hat{\beta}*)-\beta][E(\hat{\beta}*)-\beta]^T\Big\} \\
&= \underset{\beta}{E}[\sigma^2(T - TW^{-1}T) + TW^{-1}\beta\beta^T W^{-1}T] \\
&= \sigma^2(T - TW^{-1}T) + TW^{-1}E(\beta\beta^T)W^{-1}T \\
&= \sigma^2 T - \sigma^2 TW^{-1}T + \sigma^2 TW^{-1}WW^{-1}T \\
&= \sigma^2 T \qquad (13)
\end{aligned}
$$

where without loss of generality we have selected $\sigma^2 W = E(\beta\beta^T)$.

For the LSE $\hat{\beta}$ we have:

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T] = \sigma^2(X^T V^{-1} X)^{-1} \qquad (14)$$

Recalling the definition of T we may establish the following inequality between (13) and (14):

$$(X^T V^{-1}X + W^{-1})^{-1} \leq (X^T V^{-1} X)^{-1} \qquad (15)$$

Two matrices A and B are said to fulfill A > B if A - B is non-negative definite. To obtain the above we must further assume that W is non-negative definite which is true for all three choices of W previously described. In the case that we make a different choice of W, the mean error dispersion for $\hat{\beta}*$ is

$$D(\hat{\beta}*) \equiv E[(\hat{\beta}* - \beta)(\hat{\beta}* - \beta)^T] = \sigma^2(T - TW^{-1}T) + TW^{-1}\beta\beta^T W^{-1} T \qquad (16)$$

If $\beta = 0$, then:

$$D(\hat{\beta}*) = \sigma^2 (T - TW^{-1}T) < \sigma^2 (X^T V^{-1} X)^{-1} \qquad (17)$$

For $\beta \neq 0$, then, and for every choice of W, there exists a region of the parameter space in which (16) produces results smaller than (14). In general, therefore, we can state that there is a region, including the origin ($\beta = 0$), for which the BLE is a uniformly smaller mean square error estimator of β compared to the LSE and another region for which the converse is true. The more general case where the estimate of $a^T\beta$ is not a homogeneous function of the observations Y, i.e., it has the form $b^T Y + c$ with c a vector of constants, can be found in [Rao, 1976].

We conclude this section giving an expression which relates the BLE with the LSE:

$$\begin{aligned} \hat{\beta}* &= (X^T V^{-1} X + W^{-1})^{-1} X^T V^{-1} Y \\ &= (X^T V^{-1} X + W^{-1})^{-1} (X^T V^{-1} X)(X^T V^{-1} X)^{-1} X^T V^{-1} Y \\ &= (X^T V^{-1} X + W^{-1})^{-1} (X^T V^{-1} X)\hat{\beta} \qquad (18) \end{aligned}$$

or setting $G = (X^T V^{-1} X + W^{-1})^{-1} X^T V^{-1} X$, we have

$$\hat{\beta}* = G\hat{\beta} \qquad (19)$$

It is obvious from (19) that the BLE can be considered as a linear transformation of the LSE. This is a striking similarity of this type of estimator with "shrinkage estimators" used to uniformly improve unbiased estimators. We point out that it can easily be shown that the BLE "pulls" the estimate $\hat{\beta}*$ towards the origin as a whole, i.e., $\| \hat{\beta}* \| \leq \| \hat{\beta} \|$, and not each of its components individually.

Ridge Estimation

As pointed out earlier, ridge estimation is a special type of biased estimation where $V \equiv I$ and $W \equiv \frac{1}{k^2} I$. It is rather easy to visualize the results when either or none of the above is true. In [Hoerl and Kennard, 1970a] for instance, the case where $W \equiv K = [\delta_{ij} k_i^2]$ is also treated under the same name. All these variations can be categorized as methods using "uniform" prior distributions of the parameters (even though this is not explicitly stated). We examine in what follows the properties of these estimators, and we compare them whenever possible with the BLUE.

Under the LSE theory the expectation of the squared length (L^2) of the distance between the true β and its BLUE estimate $\hat{\beta}$ is

$$E[L^2] = E[(\hat{\beta} - \beta)^T (\hat{\beta} - \beta)] = \sigma^2 \operatorname{tr} (X^T X)^{-1} \tag{20}$$

We therefore obtain

$$E[\hat{\beta}^T \hat{\beta}] = \beta^T \beta + \sigma^2 \operatorname{tr} (X^T X)^{-1} \tag{21}$$

and assuming that the errors ϵ are normally distributed:

$$\operatorname{Var}[L^2] = 2\sigma^4 \operatorname{tr} (X^T X)^{-2} \tag{22}$$

We are interested in the dependence of the above quantities on the condition of our normal equations $X^T X$. Let the eigenvalues of $X^T X$ be $\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = \lambda_{min} > 0$, where p is the number of parameters. Then:

$$E[L^2] = \sigma^2 \sum_{i=1}^{p} (1/\lambda_i) \quad \text{and} \quad \operatorname{Var}[L^2] = 2\sigma^4 \sum_{i=1}^{p} (1/\lambda_i)^2 \tag{23}$$

On the basis of (23) we can see that the lower bounds for the average and the variance of L^2 are:

$$\sigma^2/\lambda_{min} \quad \text{and} \quad 2\sigma^4/\lambda^2_{min} \tag{24}$$

respectively. If our experiment is carefully set up so that it fulfills the requirements of a complete orthogonal design, then $X^T X \sim I$, and we have no problem in obtaining stable estimates. In geodesy, however, most of our problems are nonlinear and we very rarely have the chance to "design" the setup. These facts result in extremely non-orthogonal problems which in serveral cases produce "numerically singular" normals due to the high correlations among the parameters. The condition of a symmetric matrix is defined [Forsythe and Moler, 1967] as:

$$\mathrm{cond}\,(A) = \|A\|\,\|A^{-1}\|$$

for any selected norm $\|\cdot\|$. Specifically, for the Euclidean norm [ibid]:

$$\mathrm{cond}\,(A) = \lambda_{max}/\lambda_{min} \geq 1$$

The equality holds for orthogonal matrices always. The farther the condition number is from unity, the more ill-conditioned the matrix is. When λ_{min} is very small, we see from (24) that the distance L will tend to be large and it will vary more intensively than a slight change in our design warrants. Although our results are unbiased, they may be too "far" from the true values. To quote [Hoerl and Kennard, 1970a]: "The least squares estimate suffers from the deficiency of mathematical optimization techniques that give point estimates; the estimation procedure does not have built into it a method for portraying the sensitivity of the solution to the optimization criterion" (minimum sum of squares of the residuals). In support of this, [Marquardt and Snee, 1975] go one step further in identifying the cause of this insensitivity: "The 'fly in the ointment' with least squares is its requirement for unbiasedness." We do not comment on this since we have already pointed out the relevance of unbiasedness in geodetic problems. Instead we will investigate some interesting properties of ridge estimators. The usual form of a ridge estimate can be obtained from (7) by direct substitution of $V = I$ and $W^{-1} = k^2 I$, $k^2 \geq 0$:

$$\hat{\beta}* = (X^T X + k^2 I)^{-1} X^T Y = RX^T Y \tag{25}$$

To make matters simpler we assume that $X^T X$ has been properly scaled so that it is already in correlation form and the proper transformation has been appl[ied] on $X^T Y$ also. Equation (25) can also be written as

$$\hat{\beta}* = [I + k^2 (X^T X)^{-1}]^{-1} \hat{\beta} = Z\hat{\beta} \tag{26}$$

If (λ_i) denote the eigen values of $X^T X$, then the eigenvalues of R and Z, (ξ_i) and (η_i) respectively, are:

$$\xi_i = 1/(\lambda_i + k^2) \tag{27}$$

$$\eta_i = \lambda_i/(\lambda_i + k^2) \tag{28}$$

Since Z is symmetric positive definite, $Z^T Z = Z^2$, and since the eigenvalues of Z^2 are (η_i^2) we have:

$$(\hat{\beta}*)^T(\hat{\beta}*) = \hat{\beta}^T Z^T Z\hat{\beta} = \hat{\beta}^T Z^2 \hat{\beta} \le \eta^2_{max} \hat{\beta}^T \hat{\beta} \tag{29}$$

Now by (28) $\eta_{max} = \lambda_1/(\lambda_1 + k^2) \le 1$; hence:

$$\| \hat{\beta}* \| \le \eta_{max} \| \hat{\beta} \| \le \| \hat{\beta} \| \tag{30}$$

The inequality (30) establishes the fact that ridge estimators are "shorter" in a global sense than their LSE counterparts. The fact that $\hat{\beta}*$ depends on k^2 makes the residual sum of squares also a function of k^2:

$$\begin{aligned} \varphi*(k^2) &= (Y - X\hat{\beta}*)^T (Y - X\hat{\beta}*) \\ &= Y^T Y - (\hat{\beta}*)^T X^T Y - k^2 (\hat{\beta}*)^T (\hat{\beta}*) \end{aligned} \tag{31}$$

Since our criterion of optimization is to find a $\hat{\beta}*$ that gives the minimum $\varphi*$ we see that a different minimum will be obtained for each choice of k^2. This will be examined next.

The Ridge Trace

Most of the problems that we are faced with are large both in terms of the amount of data and the number of parameters involved. In addition to this most of these problems are non-linear and we very rarely see the interrelations between the parameters unless some geometrical approach is feasible. Ridge estimates claim to be the answer to such problems (at the expense of unbiasedness, of course).

If we denote by $\tilde{\beta}$ an estimate of β, then the residual sum of squares is:

$$\begin{aligned}\varphi &= (Y - X\tilde{\beta})'(Y - X\tilde{\beta}) \\ &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) + (\tilde{\beta} - \hat{\beta})'(X'X)(\tilde{\beta} - \hat{\beta}) \\ &= \varphi_{min} + \varphi(\tilde{\beta}) \qquad (32)\end{aligned}$$

In the above φ_{min} denotes the minimum obtained from the LSE theory. From the above we see that $\varphi(\tilde{\beta})$ is a continuous function of $\tilde{\beta}$ and the loci of φ = constant are concentric hyperellipsoids centered at $\hat{\beta}$. The continuity implies that for a given $\varphi(\tilde{\beta}) = \varphi_0 > 0$ there will always be a $\tilde{\beta}_0$ which produces this φ_0. Obviously we are interested in determining a specific $\tilde{\beta}_0$ which would fulfill an optimality criterion in some sense. A natural choice is for the one that produces the minimum bias. We state the problem formally:

Find a $\tilde{\beta}$ that minimizes $\tilde{\beta}'\tilde{\beta}$ under the constraint that $(\tilde{\beta} - \hat{\beta})'(X'X)(\tilde{\beta} - \hat{\beta}) = \varphi_0$. The solution of this problem through Lagrangian minimization yields:

$$\tilde{\beta} = (X'X + k^2 I)^{-1} X'Y \qquad \hat{\beta}^* \qquad (33)$$

So the ridge estimator is the answer to this problem. The value of k^2—which in the above is the inverse of the Lagrange multiplier for the φ_0 constraint—is determined so that it is consistent with the prescribed φ_0 value. Actually it is simpler to choose a certain value for $k^2 > 0$ and compute the φ_0 value later. The previous approach, however, sheds some light on the role of k^2 in ridge estimation. In [Hoerl and Kennard, 1970a] another interpretation which leads to identical results is also discussed. A more general discussion which allows for rank deficient X and V (error variance-covariance matrix) is given in [Rao, 1971].

An examination of the likelihood function for $\hat{\beta}*$ reveals that the loci of constant likelihood are also concentric hyperellipsoids centered at $\hat{\beta}$. Based on this and the similar fact that holds for the φ = constant surfaces, Hoerl and Kennard introduced the concept of the "ridge trace" which is the path described by the estimate in the likelihood space. This is practically obtained from plots of the components of $\hat{\beta}*$ for various choices of k^2. Their justification for the use of this concept as a means of studying these estimators is the fact that although "long" and "short" $\hat{\beta}*$'s are equally likely, the "long" ones, which will probably be farther away from the true value β, may not always have equal physical meaning. This is where although not mathematically or statistically stated, the use of Bayes' approach is implied. The mathematical expression that provides useful information for the ridge trace is the mean square error $E[L^2(k^2)]$ which is considered as the loss function. The following expression can be derived by use of the expectation operator:

$$
\begin{aligned}
E[L^2(k^2)] &\equiv E[(\hat{\beta}* - \beta)^T (\hat{\beta}* - \beta)] \\
&= \sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k^2)^2 + k^4 \beta^T (X^T X + k^2 I)^{-2} \beta \\
&= L_1(k^2) + L_2(k^2) \qquad (34)
\end{aligned}
$$

The first of the terms in (34) represents the total variance of the parameters, while the second is the square of the bias introduced when $\hat{\beta}*$ is used in place of $\hat{\beta}$. To obtain the first result we can use the definition of Z from (26) and its associated eigenvalues from (28) and the fact that:

$$\mathrm{Var}[\hat{\beta}*] = \sigma^2 Z(X^T X)^{-1} Z^T \qquad (35)$$

In [Hoerl and Kennard, 1970a] it is shown that there exists a $k^2 > 0$ for which $E[L^2(k^2)] < E[L^2(0)]$, $k^2 = 0$ being the case for the LSE unbiased estimate $\hat{\beta}$. This existence theorem can be proved by examining the behavior of $L_1(k^2)$ and $L_2(k^2)$. In the same reference the following key results are derived:

Theorem 1: The total variance $L_1(k^2)$ is a continuous, monotonically decreasing function of k^2.

Theorem 2: The squared bias $L_2(k^2)$ is a continuous, monotonically increasing function of k^2.

On the basis of the above it is shown that:

$$\lim_{k^2 \to 0^+} \frac{dL_1(k^2)}{d(k^2)} = -2\sigma^2 \sum_{i=1}^{p} (1/\lambda_i^2) \quad (36)$$

$$\lim_{k^2 \to 0^+} \frac{dL_2(k^2)}{d(d^2)} = 0 \quad (37)$$

These results are very important since they indicate that for an ill-conditioned system ($\lambda_{min} \approx 0$) the total variance will tend to be too large as $k^2 \to 0^+$ while the bias will be zero irrespective of the condition of $X^T X$. As we move a little from the origin and $k^2 > 0$, we introduce a very small amount of bias (the derivative of L_2 is nearly flat around the origin) and at the same time we reduce the total variance tremendously. This is better understood from Fig. 1 which is reproduced from [Hoerl and Kennard, 1970a]. It is a graphical representation of L_1, L_2 and their sum $E[L^2 (k^2)]$. As it can be seen from this figure, the ridge estimate produces a uniformly smaller mean square error than that of the LSE, for $0 < k^2 < 0.6$. For $k^2 \cong 0.06$, the ridge trace achieves its minimum. This point corresponds to the "minimum variance - minimum bias" estimate of β, the one that Rao refers to as BLIMBE in [Rao, 1973]. The determination of the k^2 value that will produce this estimate for β is as follows. We treat the general case that for each component of β a different k^2 is adopted.

Let $P^T \Lambda P = X^T X$ where Λ is the eigenvalue matrix and P the matrix of the eigenvectors of $X^T X$. Denoting $\alpha = P\beta$ it can be shown that starting with an approximation of β and choosing each $k_i^2 = \sigma^2 / \hat{\alpha}_i^2$, the iterative solution until $\hat{\alpha}^T \hat{\alpha}$ stabilizes will produce the desired set of k^2's and the corresponding ridge estimate $\hat{\beta}*$. In the case of large systems where this procedure may be very tedious and expensive, the direct solution using the pseudoinverse of X^T as shown in [Rao, 1973] might be more efficient to use. Numerical examples for the iterative approach can be found in [Hoerl and Kennard, 1970b; Marquardt and Snee, 1975]. In the second reference the results of the ridge estimator are

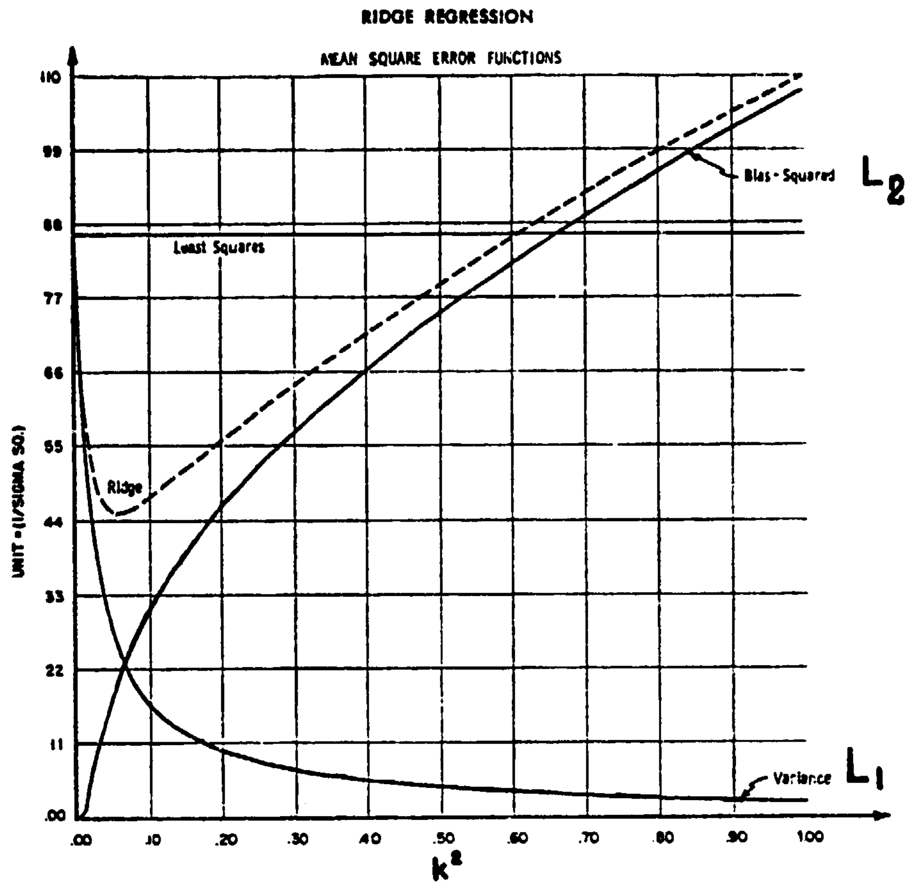


Fig. 1

compared to the generalized inverse estimates also. Marquardt [1970] has also given an excellent comparison of ridge estimators, generalized estimators and least squares estimators. This reference is particularly valuable, for it compares the ridge estimates with those obtained from an "inner constraint" adjustment; a procedure very popular with geodesists. The ridge estimator should by no means be confused with the generalized inverse counterpart, even though the two have several properties in common. A numerical example solved analytically [ibid.] with both methods provides a number of interesting results and an illustrative comparison of these estimators.

Summary

The concept, the foundation, and some of the most popular techniques for biased linear estimation were presented. The emphasis placed on ridge estimators is not without justification. One can very easily see their structural resemblance to what geodesists call "weighted constraints" adjustment. A better understanding of these estimators will probably help in their optimal utilization in our problems rather than outright rejection. It is, for instance, true that in a short-arc solution even when the inherent rank deficiency is taken care of, critical geometry or too short passes or any combination of such factors may result in a very ill-conditioned and unstable system. If there is no way to obtain a linear unbiased estimate (hence a BLUE) or if we agree that we can tolerate a certain amount of bias, then a ridge estimator with controlled bias may very well be the answer to our problem. A last remark though should be made concerning the unbiasedness of parametric functions. If a minimum norm least squares g-inverse is used to obtain the BLIMBE of a parametric function $a^T\beta$, then the resulting estimate will be unbiased if we erroneously assumed that it did not admit a LUE initially and if $a^T\beta$ is estimable. For the ridge estimator this is not true. In this case, therefore, all parametric function will be assigned biased estimates irrespective of their status under the classical least squares theory.

The epilog:

"The science of statistics is essentially a branch of Applied Mathematics, and may be regarded as mathematics applied to observational data." [Fisher, 1925]

References

Bossler, J.D. (1972). Bayesian Inference in Geodesy, PhD dissertation, The Ohio State University, Columbus.

Fisher, R.A. (1925). Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

Forsythe, G. and C.B. Moler. (1967). Computer Solution of Linear Algebraic Systems. Prentice Hall, Englewood Cliffs, New Jersey.

Hoerl, A.E. and R.W. Kennard. (1970a). "Ridge Regression: Biased Estimation for Nonorthogonal Problems," Technometrics, 12, 1.

Hoerl, A.E. and R.W. Kennard. (1970b). "Ridge Regression: Application to Nonorthogonal Problems," Technometrics, 12, 1.

Kempthorne, O. (1972). "Discussion on the paper by Prfr. Lindley and Dr. Smith: Bayes Estimates for the Linear Model," Journal of Royal Statistical Society, B24.

Lindley, D.V. (1958). "Professor Hogben's 'Crisis'—A Survey of the foundations of statistics," Applied Statistics, 7.

Marquardt, D.W. (1970). "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation," Technometrics, 12, 3.

Marquardt, D.W. and R.D. Snee. (1975). "Ridge Regression in Practice," The American Statistician, 29, 1.

Papoulis, A. (1965). Probability, Random Variables and Stochastic Processes. McGraw-Hill Book Ccmpany, New York.

Rao, C.R. (1971). "Unified Theory of Linear Estimation," Sankhyā, Series A, 33, Part 4.

Rao, C.R. (1973). Linear Statistical Inference and Its Applications, 2nd ed. John Wiley & Sons, Inc., New York.

Rao, C.R. (1976). "Estimation of Parameters in a Linear Model," The Annals of Statistics, 4, 6.

Savage, L.J. (1962). "Subjective Probability and Statistical Practice," The Foundations of Statistical Inference, M.S. Bartlett, ed. Methuen & Co. Ltd., London.

Savage, L.J. (1972). The Foundations of Statistics, revised edition of the original, 1954. Dover Publications Inc., New York.

Van Dantzig, D. (1957). "Statistical Priesthood (Savage on Personal Probabilities)," Statistica Neerlandica, 2.