D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

TOWARDS OPTIMIZING HIERARCHICAL DATA REVISIONS

Burkhard Schaffrin and Jackson Cothren
Geodetic Science and Surveying
The Ohio State University
2070 Neil Ave., Columbus, OH 43210-1275, USA

ABSTRACT

The revision of existing data must always be considered when new data are collected which have known relations with the old data, thereby taking into account that the two datasheets in question may belong to one and the same, or to two different hierarchical levels. In the *first case*, optimal data fusion would amount to a joint adjustment and, as a result, to modifications of the existing data which may then be checked for their significance. In the *second case*, the situation turns out to be somewhat trickier since, after the integration, the old dataset with a higher position in the hierarchy should still be unaffected, *including the corresponding dispersion matrix*. Here we shall explore the optimal procedure for the second case and present a unifying algorithm which would allow us to go ahead with the revisions until (only in the last step) we have to decide about the hierarchical behaviour.
Although some of the more theoretical questions must be left unanswered at this point, we do include an example in which *two photogrammetric networks of substantially different scales* are to be integrated.

1 INTRODUCTION

To illustrate the issue of hierarchical data revisions let us consider the example of photogrammetric imagery taken at two substantially different scales from the same scene. Then we could apply one of the following procedures:

- perform a *simultaneous* bundle adjustment,
- use object space coordinates from one adjustment as *stochastic* constraints on the second,
- use object space coordinates from one adjustment as *fixed* constraints on the second.

Only the third procedure would truly fulfill the requirements of a *hierarchical* data revision method insofar as it would leave the first adjustment results unchanged; the estimated object space coordinates will be *reproduced* during the second adjustment indeed, but possibly not their variance-covariance or dispersion matrix.

But even if we modify the „error propagation“ for the adjustment formulas with fixed constraints in accordance with the real stochastic constraints at hand, we could never be certain that this is the *optimal* way to integrate the two datasets in a hierarchical manner. We shall therefore try in the following to find the „best“ among those (sequential) estimates which reproduce the „hierarchically superior“ estimates along with their dispersion matrix (*i.e.* the „*reproducing property*“).

In the next chapter we shall present a summary of key results as they have been derived in the context of geodetic network fusion by *B. Schaffrin (1997)* without employing Bayesian techniques. Hierarchical Bayesian methods have, in contrast, been proposed *by L.M. Berliner (1996)* for time series and by *C.K. Wikle/L.M. Berliner/N. Cressie (1998)* for space-time models. We would also like to draw the attention to the previous work *by K.R. Koch (1983), E. Grafarend/B. Schaffrin (1988), B. Schaffrin (1989)* and *F.W.O. Aduol (1993)* among many others who have discussed the „dynamic“ network densification problem before, hierarchical or not. In the subsequent chapter we will then show the effect of the optimal hierarchical technique on the above-mentioned photogrammetric situation before we draw some (preliminary) conclusions, combined with an outlook on further research.

2 THE THEORY OF HIERARCHICAL DATA REVISION

For an easy access let us consider the new dataset be given in the form of (linearized) observation equations, namely

$$y = \underset{nxr}{A_1}\,\xi_1 + \underset{nx(m-r)}{A_2}\,\xi_2 + e\ , \qquad rk\,A_2 = m - r\ , \qquad e \sim (0, \Sigma)\ , \tag{1.1}$$

while the existing data are available from

$$\hat{\xi}_1 \sim \left(\xi_1\ ,\ \Sigma_1^0\right), \tag{1.2}$$

D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

thus representing unbiased information about ξ_1 with uncertainties as shown in the dispersion matrix Σ_1^0. In the *hierarchical* data revision problem, we try to find *optimal estimates* $\bar{\xi}_1$ and $\bar{\xi}_2$ from the new dataset y such that the *reproducing property*

$$\hat{\xi}_1 = \bar{\xi}_1, \; E\{\hat{\xi}_1\} = \xi_1 = E\{\bar{\xi}_1\}, \; D\{\hat{\xi}_1\} = \Sigma_1^0 = D\{\bar{\xi}_1\}, \tag{1.3}$$

holds true if the index 1 denotes parameters of a higher hierarchical level. The corresponding adjusted observation vector would then, of course, be given by

$$\overline{E\{y\}} := A_1 \bar{\xi}_1 + A_2 \bar{\xi}_2 = y - \bar{e} \tag{1.4}$$

with the appropriate residual vector $\bar{e}$.

For data from equal hierarchical levels, however, we would have to allow changes in $\hat{\xi}_1$ as a consequence of the new information in y. The optimal revisions may then be obtained from a joint adjustment with the *stochastic constraints*

$$\hat{\hat{\xi}}_1 = \underset{rxm}{\mathrm{K}}\, \xi + e_1^0 = \mathrm{I}_r \cdot \xi_1 + 0 \cdot \xi_2 + e_1^0, \qquad rk\ \mathrm{K} = r, \tag{1.5a}$$

$$e_1^0 \sim \left(0, \Sigma_1^0\right), \quad C\{e_1^0, e\} = 0, \qquad \left(\hat{\xi}_1 =: z_0\right), \tag{1.5b}$$

obviously resulting in *reduced uncertainties* for the revised estimates of ξ_1, say $\hat{\hat{\xi}}_1$. Following the standard approach we arrive at the same formulas as given by *H.J. Buiten (1978)*, e.g., namely

$$\hat{\hat{\xi}} = \begin{bmatrix} \hat{\hat{\xi}}_1 \\ \hat{\hat{\xi}}_2 \end{bmatrix} = \left[\mathrm{N} + \mathrm{K}^T \left(\Sigma_1^0\right)^{-1} \mathrm{K}\right]^{-1} \left[c + \mathrm{K}^T \left(\Sigma_1^0\right)^{-1} \hat{\xi}_1\right], \tag{1.6a}$$

with $[\mathrm{N}, c] := A^T \Sigma^{-1} [A, y]$, and if $rk\ A = rk\ [A_1, A_2] = m$, at

$$\hat{\hat{\xi}} = \mathrm{N}^{-1} c + \mathrm{N}^{-1} \mathrm{K}^T \left(\Sigma_0 + \mathrm{K}\mathrm{N}^{-1} \mathrm{K}^T\right)^{-1} \left(\hat{\xi}_1 - \mathrm{K}\mathrm{N}^{-1} c\right) \tag{1.6b}$$

for the estimated parameters (*without* the „reproducing property"), and at

$$D\left\{\hat{\hat{\xi}}\right\} = \left[\mathrm{N} + \mathrm{K}^T \left(\Sigma_1^0\right)^{-1} \mathrm{K}\right]^{-1} = \mathrm{N}^{-1} - \mathrm{N}^{-1} \mathrm{K}^T \left(\Sigma_1^0 + \mathrm{K}\mathrm{N}^{-1} \mathrm{K}^T\right)^{-1} \mathrm{K}\mathrm{N}^{-1} \tag{1.7}$$

for their dispersion matrix. Note that we have not decomposed the dispersion matrix into a so-called „cofactor matrix" and a variance component.

We may as well identify the *residual vector* for the stochastic (or fiducial) constraints as

$$\tilde{e}_0 = \hat{\xi}_1 - \mathrm{K}\hat{\hat{\xi}} = \hat{\xi}_1 - \hat{\hat{\xi}}_1 = \left[\mathrm{I}_r + \mathrm{K}\mathrm{N}^{-1} \mathrm{K}^T \left(\Sigma_1^0\right)^{-1}\right]^{-1} \left(\hat{\xi}_1 - \mathrm{K}\mathrm{N}^{-1} c\right) \tag{1.8}$$

with its dispersion matrix

$$D\{\tilde{e}_0\} = \left[\mathrm{I}_r + \mathrm{K}\mathrm{N}^{-1} \mathrm{K}^T \left(\Sigma_1^0\right)^{-1}\right] \left(\Sigma_1^0 + \mathrm{K}\mathrm{N}^{-1} \mathrm{K}^T\right) \left[\mathrm{I}_r + \left(\Sigma_1^0\right)^{-1} \mathrm{K}\mathrm{N}^{-1} \mathrm{K}^T\right] =$$

$$= \Sigma_1^0 \left(\Sigma_1^0 + \mathrm{K}\mathrm{N}^{-1} \mathrm{K}^T\right)^{-1} \Sigma_1^0, \tag{1.9}$$

D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

knowing that $\hat{\hat{\xi}}$ and $\tilde{e}_0$ will be *uncorrelated*:

$$C\left\{\hat{\hat{\xi}}, \tilde{e}_0\right\} = 0. \tag{1.10}$$

Now, as soon as we concentrate ourselves on the class of linear estimates with the „*reproducing property*“, however, the optimal estimates will be slightly different from (1.6a – b), but can be easily found via updates using the above results. As shown previously by *B. Schaffrin (1997)*, we obtain

$$\bar{\xi} = \hat{\hat{\xi}} + \mathrm{K}^T \left(\mathrm{K}\mathrm{K}^T\right)^{-1} \tilde{e}_0 \tag{1.11}$$

for the best „hierarchically estimated“ parameters with

$$D\left\{\bar{\xi}\right\} = D\left\{\hat{\hat{\xi}}\right\} + K^T \left(\mathrm{K}\mathrm{K}^T\right)^{-1} \cdot D\{\tilde{e}_0\} \cdot \left(\mathrm{K}\mathrm{K}^T\right)^{-1} \mathrm{K} \tag{1.12}$$

as corresponding dispersion matrix. Introducing our particular matrix $\mathrm{K} = \left[\mathrm{I}_r\ ,\ 0\right]$, these formulas can be further simplified using the identities

$$\mathrm{K}\mathrm{K}^T = \mathrm{I}_r\ , \qquad \mathrm{K}^T\mathrm{K} = \begin{bmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{bmatrix}, \tag{1.13}$$

so that we eventually get the following estimates

$$\bar{\xi} = \hat{\hat{\xi}} + \begin{bmatrix} \mathrm{I}_r \\ 0 \end{bmatrix} \left(\hat{\xi}_1 - \hat{\hat{\xi}}_1\right) = \begin{bmatrix} \hat{\xi}_1 \\ \hat{\hat{\xi}}_2 \end{bmatrix}, \tag{1.14}$$

clearly exhibiting the „reproducing property“, and their dispersion matrix

$$D\left\{\bar{\xi}\right\} = D\left\{\hat{\hat{\xi}}\right\} + \left[\begin{array}{c|c} \Sigma_1^0 \left(\Sigma_1^0 + \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T\right)^{-1} \Sigma_1^0 & O \\ \hline O & O \end{array}\right] \tag{1.15}$$

with

$$\begin{aligned} D\left\{\bar{\xi}_1\right\} = D\left\{\mathrm{K}\hat{\hat{\xi}} + \tilde{e}_0\right\} &= \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T - \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T \left(\Sigma_1^0 + \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T\right)^{-1} \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T + \\ &\quad + \Sigma_1^0 \left(\Sigma_1^0 + \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T\right)^{-1} \left(\Sigma_1^0 + \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T - \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T\right) = \\ &= \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T - \left(\mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T + \Sigma_1^0\right)\left(\Sigma_1^0 + \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T\right)^{-1} \mathrm{K}\mathrm{N}^{-1}\mathrm{K}^T + \Sigma_1^0 \ = \ = \Sigma_1^0 \end{aligned} \tag{1.16}$$

in the upper left hand corner as expected, whereas all the other three blocks remain the same as in $D\left\{\hat{\hat{\xi}}\right\}$.

As a *consequence*, we see that the best hierarchical data revision is exactly the same as the optimal (non-hierarchical) data revision unless a higher level is involved in which case the old information is reproduced. This, however, should not be done without properly testing the difference between $\bar{\xi}_1 = \hat{\xi}_1$ and $\hat{\hat{\xi}}_1$ via the null hypothesis

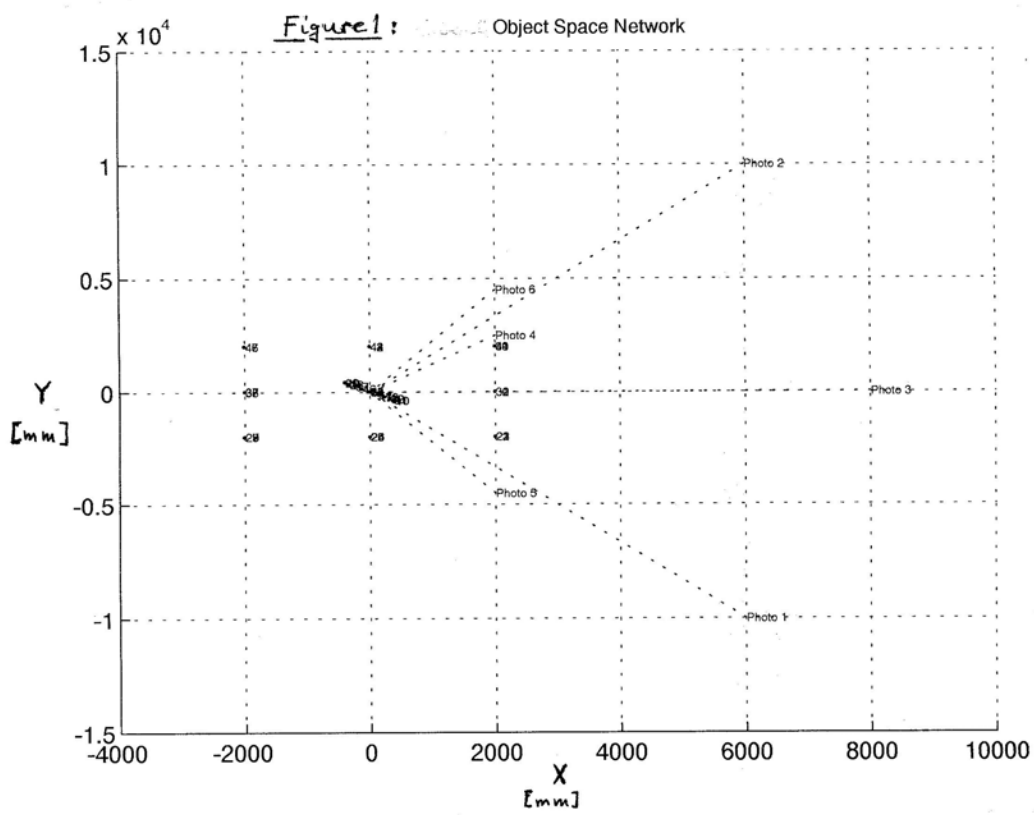
D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

$$H_0 : E\left\{\hat{\xi}_1 - \hat{\tilde{\xi}}_1\right\} = E\left\{\tilde{e}_0\right\} = 0$$

which is not too difficult to do, but will be left for a subsequent paper. Instead, let us see how the above theory works in a (simulated) photogrammetric example.

3 AN EXAMPLE: THE JOINT ANALYSIS OF PHOTOGRAMMETRIC IMAGERY WITH TWO DIFFERENT SCALES

In the following let us describe the simulated datasets along with the parameters of the exterior orientation. The object space point field consisted of two distinct arrangements of three-dimensional points (see Figure 1).



1. 27 points equally spaced within a cube centered on the point (0,0,2000) mm, 4000 mm on a side. The points on the cube are numbered 21-47.
2. 20 points forming the perimeter of a circle with a 200 mm radius, rotated about its vertical axis, also centered on the point (0,0,2000). The points on the circle are numbered 1-20 and should be considered densification points of the cube network, i.e. lower in the hierarchy.

We generated two synthetic sets of photogrammetric images using this object space field, FAR3 and NEAR3. Each set was composed of three images with a focal length of 60 mm, a principal point at the image coordinate system origin, and a format of 36 mm x 24 mm. These parameters are typical of an off-the-shelf 35 mm camera and lens arrangement.

FAR3 consisted of images of the cube only, with exterior orientation parameters given in Table 1. The average scale of the images in this set was approximately 235:1.

FAR3	Xo (mm)	Yo (mm)	Zo (mm)	omega (deg)	phi (deg)	kappa (deg)	Range (mm)
Image 1	6000	- 10000	8000	51.35	25.10	18.74	14,142
Image 2	6000	10000	8000	- 51.3	25.10	- 18.74	11,313
Image 3	8000	0	8000	0	45	- 90	14,142

Table 1, FAR3 Data, exterior orientation parameters

D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

NEAR3 consisted of images of the entire visible field (cube and circle) with the exterior orientation parameters given in Table 2. The average scale of the images in this set was approximately 88:1.

NEAR3	Xo (mm)	Yo (mm)	Zo (mm)	omega (deg)	phi (deg)	kappa (deg)	Range (mm)
Image 1	2000	2500	2000	- 51.35	32.00	- 22.99	3775
Image 2	2000	- 4500	2000	66.00	22.10	9.51	5315
Image 3	2000	4500	2000	- 66.00	22.00	- 9.51	5315

Table 2, NEAR3 Data, exterior orientation parameters

The closer range of these photographs resulted in images of all the points on the circle (points 1-20), but the appearance on multiple images of only eight points on the cube (points 24, 25, 27, 33, 34, 35, 36, and 43). These eight points are thus common to both datasets.

We applied the following random errors to compute image coordinates of the projected object space points visible in each image:

Exterior Orientation Parameters, rotation angles: $\sigma = \pm 1$ arc second
Exterior Orientation Parameters, camera position: $\sigma = \pm 10$ mm
Interior Orientation Parameters, (x_p, y_p, c): $\sigma = \pm 0.01$ mm
Projected Image Coordinates: $\sigma = \pm 0.004$ mm

In order to isolate network geometry effects and keep the model parameters to a minimum, the simulation did not model other interior orientation parameters such as radial lens distortion or film unflatness.

The experiment consisted of four separate adjustments:

1. The free network adjustment of the NEAR3 images,
2. The free network adjustment of the FAR3 images, providing the „prior information“,
3. The free network adjustment of both, FUSE6, following (1.6a – b), and
4. The adjustment of the NEAR3 with stochastic constraints using the optimal reproducing estimator, which we designate NEAR3K reflecting the inclusion of the constraint matrix K. This adjustment recomputed adjustment 1, but updated the parameter vector with the constraint matrix K and constraint observation vector z_0 at each iteration, eventually leading to the solution (1.11) or (1.14), respectively.

The first two adjustments serve as basis from which to measure the effects of the network integration. In all adjustments the interior orientation parameters (x_p, y_p, and c) were added to the adjustment as pseudo-observations with a variance of 0.0001 mm^2. We did not consider any other interior orientation parameters. The exterior orientation of each image and all multi-ray object space points were the parameters to be estimated in each adjustment. We assumed the *a priori* image coordinate observation precision to be ± 0.005 mm. Note that the *a priori* reference variance, the so-called „variance component“, was 1.00.

The results of the four adjustments are shown in Table 3:

	σ_0^2	RMS X (mm)	RMS Y (mm)	RMS Z (mm)	Max Var X (mm^2)	Max Var Y (mm^2)	Max Var Z (mm^2)
NEAR3 20 points	0.92	± 0.61	± 0.88	± 0.52	7.38 @ pt 25	12.13 @ pt 20	6.92 @ pt 20
FAR3 27 points	1.00	± 0.58	± 0.71	± 0.86	1.50 @ pt 26	1.37 @ pt 21	1.23 @ pt 36
FUSE6 47 points	0.80	± 0.32	± 0.06	± 0.17	0.97 @ pt 36	1.11 @ pt 21	0.95 @ pt 21
NEAR3K 28 points	3.54	± 0.36	± 0.30	± 0.39	6.91 @ pt 25	3.35 @ pt 34	5.72 @ pt 36

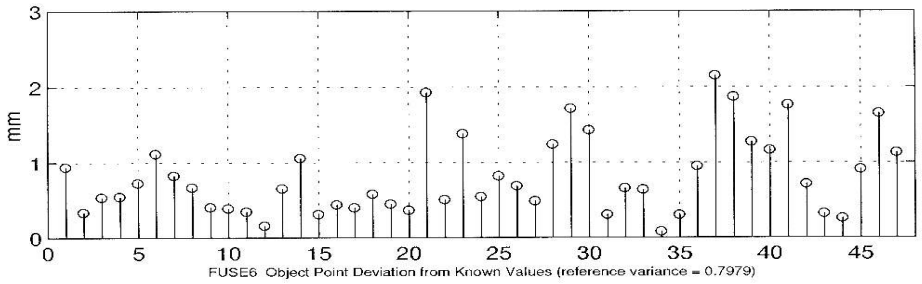
It should be noted that all the adjustments resulted in exterior orientation parameters that differed from the synthetic values by no more than one degree (angles) and four millimetres (camera position). The variance-covariance matrix of all exterior orientation parameters can be provided upon request.

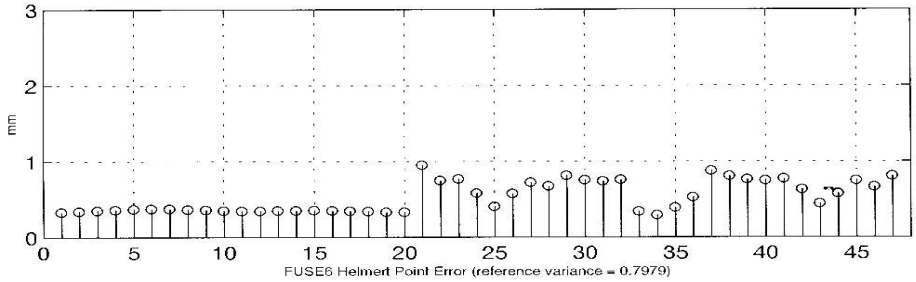
It is clear from an analysis of the summarised results that the FUSE6 network yielded the best precision of all the adjustments. We thought the inclusion of the large-scale dataset would increase the variances of some points imaged only in the FAR3 network. However, this was not the case.

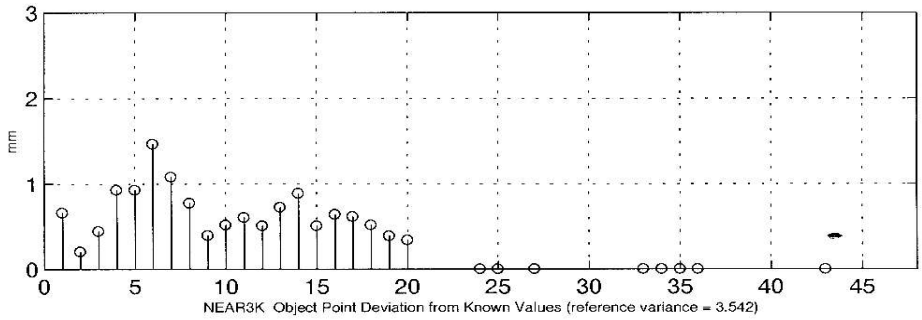
One measure of the precision of the estimated object space coordinates is the Helmert point error. A comparison of the FUSE6 and FAR3 point errors reveals that the precision of the entire network (by this measure) was improved in the

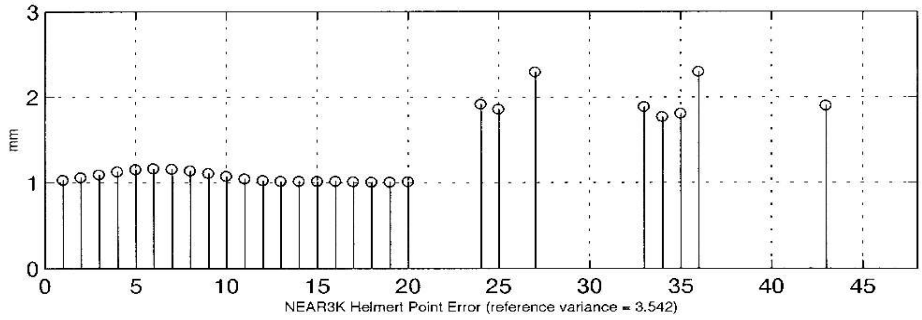
D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

combined adjustment. Points 1-20 improved slightly over the NEAR3 adjustment, as did the common points. The other points on the cube, however, where we expected to see larger point errors, were either unchanged or slightly improved. The image observations in all six images were equally weighted at ± 0.005 mm. It might be that applying larger weights to the NEAR3 images (as might occur with real data) would have this effect.









A comparison of the point errors of the NEAR3K and FUSE6 networks shows a definite loss of precision using the optimal reproducing estimator. Point errors in the NEAR3K adjustment were 2 to 2.5 times larger than in the FUSE6 adjustment. This is to be expected from the optimal reproducing estimator because it is only optimal in the class of estimators that result in no change on the constrained parameters.

D. Fritsch, M. Englich & M. Sester, eds, 'IAPRS', Vol. 32/4, ISPRS Commission IV Symposium on GIS - Between Visions and Applications, Stuttgart, Germany.

As second quality measure of the adjustments might be considered, especially in the case of synthetic data, the Root Mean Square (RMS) error. The RMS errors in each coordinate axis are shown in Table 3 above. A comparison of NEAR3K and FUSE6 using this measure shows a very small difference in the X-axis RMS error, a substantial difference in the Y-axis RMS error, and a substantial difference in the Z-axis RMS error. These computations in the NEAR3K adjustment include the zero differences of the common point parameters.

The RMS errors for points 1-20, the points that densify the FAR3 network, are shown in Table 4.

	RMS X (mm)	RMS Y (mm)	RMS Z (mm)
FUSE6	0.26	0.52	0.26
NEAR3K	0.43	0.36	0.46

Table 4, RMS for points 1-20 only

This provides a slightly different view of the adjustment. The RMS difference between the estimates of the densification points in the two adjustments is less pronounced than the difference shown in Table 3.

4 Conclusions and Outlook

A significant loss of overall precision resulted from the use of the optimal reproducing estimator compared with a simultaneous adjustment when integrating the two datasets, FAR3 and NEAR3. This loss of precision must be weighed against the amount of change in the common points that is a result of a simultaneous adjustment, an effect that the optimal reproducing estimator is designed to eliminate.

Future experiments might proceed as follows:

1. Design other networks with different geometries. The data used in this experiment simulated a close range network. An aerial or space scene could be used to simulate different space and airborne sensors currently in use.
2. Apply different weight to the image coordinate observations in the different scale images. This is perhaps a more realistic situation.
3. Compute the mapping function from image observations to residuals and compare the inner and outer reliability of the two integration methods.
4. Perform the experiment with real data from one or more close range scenes and aerial scenes. This, along with more simulations, will provide a more robust measure of the effects of the sub-optimal estimator.

Acknowledgement:

This paper was finalized while the first author was on sabbatical leave at the Institute for Applied Geodesy of the Technical University of Graz, Austria, with Prof. Fritz K. Brunner as his host and with financial support through the Faculty of Civil Engineering. This is gratefully acknowledged.

References:

Aduol, F.W.O.: A static-dynamic model for densification of geodetic networks, AVN – Intl. Edition 10 (1993), 20 – 36.

Berliner, L.M.: Hierarchical Bayesian time series models; in: K. Hanson/R. Silver (eds.), Maximum Entropy and Bayesian Methods, Kluwer: Dordrecht 1996, 15 – 22.

Buiten, H.J.: Junction of nets by collocation, Manus. Geodaet. 3 (1978), 253 – 297.

Grafarend, E. and Schaffrin, B.: From the static to the dynamic treatment of geodetic networks (in German), ZfV 113 (1988), 79 – 103.

Koch, K.R.: The choice of the datum of a trigonometric network for the determination of new points (in German), ZfV 108 (1983), 104 – 111.

Schaffrin, B.: Fiducial versus fixed points in the GPS network approach; in: Proc. of the 5th Intl. Geodetic Symp. on Satellite Positioning, Las Cruces, New Mexico, U.S.A., 1989, Vol. I, 500 – 511.

Schaffrin, B.: On suboptimal geodetic network fusion, IAG General Meeting, Rio de Janeiro/Brazil, Sept. 1997 (to be published in J. of Geodesy).

Wikle, C.K., Berliner, L.M. and Cressie, N.: Hierarchical Bayesian Space-Time Models, Environmental and Ecological Statistics, to appear in 1998.