HIERARCHICAL DATA FUSION WITH PHOTOGRAMMETRIC APPLICATIONS

Burkhard Schaffrin
Jackson Cothren

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

ABSTRACT

When two datasets are fused using least-squares adjustment, usually all the results will be affected by some change, even the reference data that are meant to provide information of such high quality that they should remain stable. In order to avoid this effect, the sequential adjustment is to be replaced by a strictly hierarchical method in which the estimation procedure is designed to reproduce everything that belongs to a "higher category" and to perform an adjustment in the least-squares sense for everything else. After presenting such a suboptimal estimator, but with the "reproducing property," this technique is applied to the integration of photogrammetric networks of substantially different scales.

Keywords: Hierarchical, least-squares, photogrammetric

1. INTRODUCTION

Let us assume that photogrammetric images were taken at two substantially different scales from the same scene. To merge the available information 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.

Out of these procedures, only the third one would truly reproduce the results from the first adjustment, but at the cost of neglecting the corresponding variance-covariance matrix. The first and second procedures, however, would provide us with "up-dates" of the first adjustment, thus not fulfilling the requirements of a *hierarchical data fusion method* that ought to keep the so-called "reference information" unchanged.

On the other hand, the third procedure will rarely be optimal among all possible hierarchical data fusion methods. This was shown by B. Schaffrin (1997), who derived the "optimal reproducing estimator" in the context of geodetic network densification by employing non-Bayesian techniques. Hierarchical Bayesian estimators 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.

In the following we shall give a brief review of alternative procedures, including the "optimal hierarchical data fusion method." Afterwards, we shall compare them with each other in view of a (synthetic) photogrammetric example (where three close-range images provide the reference frame for three other closest-range images) before we draw some conclusions and give an out-look on further research.

In this context we also want to draw attention to the previous work of *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, mostly without reference to hierarchical procedures for which the "Helmert transformation" was traditionally used. Now we know that the latter approach is non-optimal, in general; see, e.g., *Schaffrin (2002)*.

2. OPTIMAL DATA FUSION – NON-HIERARCHICAL MODE

Let us start from a standard situation in which we assume the (higher level) "reference information" to be given in the form of estimated parameters which, in turn, appear in the (linearized) observation equations for the second (lower level) dataset. Consequently, we basically have an *Extended Gauss-Markov Model*

$$y = \underset{n\times l}{A_1}\,\xi_1 + \underset{n\times(m-l)}{A_2}\,\xi_2 + e, \quad rkA_2 = m - l, \tag{1}$$

$$e \sim (0, \Sigma = \sigma_0^2 P^{-1}), \quad \Sigma \text{ positive-definite}, \tag{2}$$

$$\hat{\xi}_1 \sim (\xi_1, \Sigma_1^0 = \sigma_0^2 Q_1^0), \quad C\{\hat{\xi}_1, e\} = 0, \tag{3}$$

where

- y is a $n \times 1$ vector of observational increments,
- ξ_1 is a $l \times 1$ vector of parameters with prior information,
- ξ_2 is a $(m-l) \times 1$ vector of parameters without prior information,
- $A := [A_1, A_2]$ is the $n \times m$ coefficient matrix,
- e is a $n \times 1$ vector of random observation errors,
- $\Sigma := D\{e\}$ is the corresponding $n \times n$ dispersion matrix,
- P is the corresponding $n \times n$ weight matrix,
- σ_0^2 is the (typically unknown) variance component,
- $\hat{\xi}_1$ is the $l \times 1$ (given) vector of unbiased "reference" (prior) information,
- $\Sigma_1^0 := D\{\hat{\xi}_1\}$ is the corresponding $l \times l$ dispersion matrix,
- Q_1^0 is the corresponding $l \times l$ cofactor matrix,
- $P_1^0 := (Q_1^0)^{-1}$ may be used as $l \times l$ weight matrix if Q_1^0 is positive-definite and thus invertible,
- C denotes "covariance" while E stands for "expectation" and D for "dispersion."

Note that we introduced the same variance component σ_0^2 for both the new observations and the reference data.

Following the standard approach, we may rephrase (3) and give it the form of *stochastic constraints*

$$\hat{\xi}_1 = \underset{l \times m}{K} \xi + e_1^0 = I_l \cdot \xi_1 + 0 \cdot \xi_2 + e_1^0, \quad rk(K = [I_l, 0]) = l, \tag{4}$$

$$e_1^0 \sim (0, \Sigma_1^0), \quad C\{e_1^0, e\} = 0, \tag{5}$$

from which we arrive at the same formulas as given by *H.J. Buiten (1978)*, for instance; namely

$$\hat{\hat{\xi}} = \begin{bmatrix} \hat{\hat{\xi}}_1 \\ \hat{\hat{\xi}}_2 \end{bmatrix} = (N + K^T P_1^0 K)^{-1}(c + K^T P_1^0 \hat{\xi}_1) \tag{6}$$

$$= N^{-1}c + N^{-1}K^T(Q_1^0 + KN^{-1}K^T)^{-1}(\hat{\xi}_1 - KN^{-1}c) \tag{7}$$

if $N := A^T P A$ is invertible, $c := A^T P y$,

for the estimated parameters, and at

$$D\{\hat{\hat{\xi}}\} = \sigma_0^2 (N + K^T P_1^0 K)^{-1} = \sigma_0^2 N^{-1} - \sigma_0^2 N^{-1} K^T (Q_1^0 + KN^{-1}K^T)^{-1} KN^{-1} \tag{8}$$

for their dispersion matrix. Obviously, $\hat{\hat{\xi}}_1$ does not have the "reproducing property" since the corresponding residual vector

$$\tilde{e}_1^0 = \hat{\xi}_1 - K\hat{\hat{\xi}} = \hat{\xi}_1 - \hat{\hat{\xi}}_1 = (I_l + KN^{-1}K^T P_1^0)^{-1}(\hat{\xi}_1 - KN^{-1}c) \tag{9}$$

will *not vanish*, in general (unless $\hat{\xi}_1 = KN^{-1}c$, which would rarely be fulfilled). The corresponding dispersion matrix is then readily derived as

$$D\{\tilde{e}_1^0\} = \sigma_0^2 (I_l + KN^{-1}K^T P_1^0)^{-1}(Q_1^0 + KN^{-1}K^T)(I_l + P_1^0 KN^{-1}K^T)^{-1} = \sigma_0^2 Q_1^0 (Q_1^0 + KN^{-1}K^T)^{-1} Q_1^0 = \sigma_0^2 (P_1^0 + P_1^0 KN^{-1}K^T P_1^0)^{-1} \tag{10}$$

$$= \sigma_0^2 Q_1^0 - \sigma_0^2 K(N + K^T P_1^0 K)^{-1} K^T. \tag{11}$$

We may also introduce the residual vector for the observations as

$$\tilde{e} = y - A\hat{\hat{\xi}} = y - A_1\hat{\hat{\xi}}_1 - A_2\hat{\hat{\xi}}_2 \tag{12}$$

with the dispersion matrix

$$D\{\tilde{e}\} = D\{y\} - D\{A\hat{\hat{\xi}}\} = \sigma_0^2 (P^{-1} - AN^{-1}A^T) + \sigma_0^2 AN^{-1}K^T (Q_1^0 + KN^{-1}K^T)^{-1} KN^{-1}A \tag{13}$$

$$= \sigma_0^2 P^{-1} - \sigma_0^2 A(N + K^T P_1^0 K)^{-1} A^T \tag{14}$$

and the covariance matrix

$$C\{\tilde{e}, \tilde{e}_1^0\} = C\{y, \tilde{e}_1^0\} = -\sigma_0^2 AN^{-1}K^T (Q_1^0 + KN^{-1}K^T)^{-1} Q_1^0 \tag{15}$$

$$= -\sigma_0^2 A(N + K^T P_1^0 K)^{-1} K^T, \tag{16}$$

knowing that $\hat{\hat{\xi}}$ and $\tilde{e}_1^0$ (as well as $\tilde{e}$) will be *uncorrelated*:

$$C\{\hat{\hat{\xi}}, \tilde{e}_1^0\} = 0,, \quad C\{\hat{\hat{\xi}}, \tilde{e}\} = 0. \tag{17}$$

Obviously, the normal equations from which the solution (6) was obtained essentially represent *orthogonality relations* which in terms of the residual vectors can be rewritten as:

$$A_1^T P\tilde{e} + P_1^0 \tilde{e}_1^0 = 0, \quad A_2^T P\tilde{e} = 0. \tag{18}$$

If we now turn to the estimation of the variance component σ_0^2, the *best invariant quadratic uniformly unbiased estimate* of it is readily given by

$$\hat{\sigma}_0^2 = [\tilde{e}^T P\tilde{e} + (\tilde{e}_1^0)^T P_1^0 \tilde{e}_1^0]/(n - m + l) = (\Omega + R)/(n - m + l), \tag{19}$$

$$\Omega := y^T P y - c^T N^{-1} c = \tilde{e}^T P \tilde{e} + c^T N^{-1} K^T P_1^0 \tilde{e}_1^0, \tag{20}$$

$$R := (\hat{\xi}_1 - KN^{-1}c)^T (Q_1^0 + KN^{-1}K^T)^{-1} (\hat{\xi}_1 - KN^{-1}c) = (\tilde{e}_1^0 - KN^{-1}c)^T P_1^0 \tilde{e}_1^0, \tag{21}$$

where R and Ω can be shown to be statistically independent under the assumption of normal distributions for e and e_1^0. Thus, we may *check the consistency* of the "reference information" $\hat{\xi}_1$ with the new observations in y by looking at the *test statistic* (if $rkA = m$)

$$T := \frac{R/l}{\Omega/(n-m)} \sim F(l, n-m) \tag{22}$$

which happens to be *F*-distributed under the null hypothesis

$$H_0 : \hat{\xi}_1 \sim \mathcal{N}(\xi_1, \sigma_0^2 Q_1^0). \tag{23}$$

Of course, T would follow a non-central *F*-distribution if the alternative hypothesis

$$H_a : \hat{\xi}_1 \sim \mathcal{N}(\gamma_1 \neq \xi_1, \sigma_0^2 Q_1^0) \tag{24}$$

holds true where the non-centrality parameter is then given as

$$\vartheta := (\gamma_1 - \xi_1)^T (Q_1^0 + KN^{-1}K^T)^{-1} (\gamma_1 - \xi_1)/2. \tag{25}$$

For later applications, let us mention that under certain conditions the regular inverse N^{-1} may be replaced by a *g-inverse* N^- in the above formulas, and m by $q := rkA$ if the latter is smaller than m.

Note that, due to the special structure of $K = [I_l, 0]$, we may always replace $KN^{-1}K^T$ by the inverse of the "*first Schur complement*" of the matrix N, namely

$$KN^{-1}K^T = [I_l, 0] \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} \begin{bmatrix} I_l \\ 0 \end{bmatrix} = S_1^{-1}, \tag{26}$$

$$S_1 := N_{11} - N_{12} N_{22}^{-1} N_{21}, \quad N_{ij} := A_i^T P A_j. \tag{27}$$

Similarly, we have

$$KN^{-1} = \left[S_1^{-1}, \quad -S_1^{-1} N_{12} N_{22}^{-1} \right], \tag{28}$$

$$KN^{-1}c = S_1^{-1}(c_1 - N_{12}N_{22}^{-1}c_2), \quad c_i := A_i^T P y. \tag{29}$$

This concludes our short review of the optimal data fusion procedure in the non-hierarchical mode.

3. OPTIMAL DATA FUSION – HIERARCHICAL MODE

Now let us turn to the *hierarchical mode*, but still using the original Extended Gauss-Markov Model as defined in (1)-(5). In this case we require the prior information $\hat{\xi}_1$ to be reproduced along with its covariance matrix Σ_1^0, or at least with its cofactor matrix Q_1^0. Let us call the new estimate $\bar{\xi}$; then the "reproducing property" reads:

$$\bar{\xi}_1 = \hat{\xi}_1, \quad D\{\bar{\xi}_1\} = \sigma_0^2 Q_1^0. \tag{30}$$

Following the derivations in *B. Schaffrin (1997)*, the "optimal linear uniformly unbiased estimate of ξ, with the reproducing property for ξ_1," is readily obtained as

$$\bar{\xi} = \hat{\hat{\xi}} + K^T (KK^T)^{-1} \tilde{e}_1^0 \tag{31}$$

where $\hat{\hat{\xi}}$ and $\tilde{e}_1^0$ are to be taken from (6), (7) and (9), respectively. Because of (17), we immediately get its dispersion matrix as

$$\begin{aligned} D\{\bar{\xi}\} &= D\{\hat{\hat{\xi}}\} + D\{K^T(KK^T)^{-1}\tilde{e}_1^0\} \\ &= \sigma_0^2 (N + K^T P_1^0 K)^{-1} + \sigma_0^2 K^T (KK^T)^{-1} Q_1^0 (KK^T)^{-1} K \\ &\quad - \sigma_0^2 K^T (KK^T)^{-1} K (N + K^T P_1^0 K)^{-1} K^T (KK^T)^{-1} K \end{aligned} \tag{32}$$

$$= \sigma_0^2 \begin{bmatrix} N_{11} + P_1^0 & N_{12} \\ N_{21} & N_{22} \end{bmatrix}^{-1} + \sigma_0^2 \begin{bmatrix} Q_1^0 - (S_1 + P_1^0)^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \tag{33}$$

thereby exploiting the special pattern of $K := [I_l, 0]$.

The corresponding residual vectors then result in

$$\bar{e}_1^0 = \hat{\xi}_1 - \bar{\xi}_1, \quad D\{\bar{e}_1^0\} = 0, \quad C\{\bar{e}_1^0, \bar{e}\} = 0, \tag{34}$$

and in

$$\bar{e} = y - A\bar{\xi} = y - A\hat{\hat{\xi}} - AK^T(KK^T)^{-1}\tilde{e}_1^0 = \tilde{e} - A_1 \tilde{e}_1^0 \tag{35}$$

with the dispersion matrix

$$\begin{aligned} D\{\bar{e}\} &= D\{\tilde{e}\} - C\{\tilde{e}, A_1\tilde{e}_1^0\} - C\{A_1\tilde{e}_1^0, \tilde{e}\} + D\{A_1\tilde{e}_1^0\} \\ &= \sigma_0^2 P^{-1} - \sigma_0^2 A(N + K^T P_1^0 K)^{-1} A^T + \sigma_0^2 A (N + K^T P_1^0 K)^{-1} K^T A_1^T \\ &\quad + \sigma_0^2 A_1 K (N + K^T P_1^0 K)^{-1} A^T - \sigma_0^2 A_1 K (N + K^T P_1^0 K)^{-1} K^T A_1^T \\ &\quad + \sigma_0^2 A_1 Q_1^0 A_1^T \\ &= \sigma_0^2 (P^{-1} + A_1 Q_1^0 A_1^T) \\ &\quad - \sigma_0^2 (A - A_1 K)(N + K^T P_1^0 K)^{-1} (A - A_1 K)^T \end{aligned} \tag{36}$$

$$\begin{aligned} &= \sigma_0^2 (P^{-1} + A_1 Q_1^0 A_1^T - A_2 N_{22}^{-1} A_2^T) \\ &\quad - \sigma_0^2 N_{22}^{-1} N_{21} (S_1 + P_1^0)^{-1} N_{12} N_{22}^{-1} \\ &\neq D\{y\} - D\{A\bar{\xi}\} \end{aligned} \tag{37}$$

after using formulas (10)-(16). Note that $\bar{\xi}$ and $\bar{e}$ are *not uncorrelated* due to

$$\begin{aligned} C\{\bar{\xi}, \bar{e}\} &= C\{\hat{\hat{\xi}} + K^T(KK^T)^{-1}\tilde{e}_1^0, \tilde{e} - A_1\tilde{e}_1^0\} \\ &= C\{\hat{\hat{\xi}}, \tilde{e}\} + C\{K^T(KK^T)^{-1}\tilde{e}_1^0, \tilde{e}\} - C\{\hat{\hat{\xi}}, A_1\tilde{e}_1^0\} - C\{K^T(KK^T)^{-1}\tilde{e}_1^0, A_1\tilde{e}_1^0\} \\ &= -\sigma_0^2 K^T (KK^T)^{-1} K (N + K^T P_1^0 K)^{-1} (A - A_1 K)^T \\ &\quad - \sigma_0^2 K^T (KK^T)^{-1} Q_1^0 A_1^T \end{aligned} \tag{38}$$

$$= -\sigma_0^2 K^T (KK^T)^{-1} [Q_1^0 A_1^T - (S_1 + P_1^0)^{-1} N_{12} N_{22}^{-1} A_2^T], \quad (39)$$

in contrast to (17) for the optimal (non-reproducing) linear estimate of ξ. Furthermore, the orthogonality relations (18) turn into

$$A_1^T P\bar{e} = A_1^T P\tilde{e} - N_{11}\tilde{e}_1^0 = -(N_{11} + P_1^0)\tilde{e}_1^0, \quad A_2^T P\bar{e} = -N_{21}\tilde{e}_1^0, \quad (40)$$

which again shows the central role that the "original" residual vector $\tilde{e}_1^0$ has to play.

For the variance component σ_0^2, we now derive an *invariant quadratic uniformly unbiased estimate* based on the sum of weighted squared residuals $\bar{e}$ and $\bar{e}_1^0 = 0$, namely

$$\begin{aligned} \Omega + \bar{R} &:= \bar{e}^T P\bar{e} + (\tilde{e}_1^0)^T P_1^0 \tilde{e}_1^0 = (\tilde{e} - A_1\tilde{e}_1^0)^T P(\tilde{e} - A_1\tilde{e}_1^0) \\ &= \tilde{e}^T P\tilde{e} - 2\tilde{e}^T P A_1 \tilde{e}_1^0 + (\tilde{e}_1^0)^T N_{11}\tilde{e}_1^0 \\ &= [\tilde{e}^T P\tilde{e} + (\tilde{e}_1^0)^T P_1^0 \tilde{e}_1^0] + (\tilde{e}_1^0)^T (N_{11} + P_1^0)\tilde{e}_1^0 \end{aligned} \quad (41)$$

where

$$\bar{R} - R := (\tilde{e}_1^0)^T (N_{11} + P_1^0)\tilde{e}_1^0 \quad (42)$$

defines the increase due to the additional requirement of "reproducing $\hat{\xi}_1$". Consequently, the expectation of $\Omega + \bar{R}$ can be split into

$$E\{\Omega + R\} = E\{\tilde{e}^T P\tilde{e} + (\tilde{e}_1^0)^T P_1^0 \tilde{e}_1^0\} = (n - m + l)\sigma_0^2 \quad (43)$$

and

$$\begin{aligned} E\{\bar{R} - R\} &= tr(N_{11} + P_1^0) \cdot E\{\tilde{e}_1^0 (\tilde{e}_1^0)^T\} = tr(N_{11} + P_1^0) \cdot D\{\tilde{e}_1^0\} \\ &= tr(N_{11} + P_1^0)[Q_1^0 - (S_1 + P_1^0)^{-1}] \cdot \sigma_0^2 \\ &= \sigma_0^2 \cdot tr[(I_l + N_{11}Q_1^0)(I_l + S_1 Q_1^0)^{-1} S_1 Q_1^0], \end{aligned} \quad (44)$$

thus leading to the *new estimate*

$$\bar{\sigma}_0^2 = \bar{e}^T P\bar{e} / [n - m + l + tr(I_l + N_{11}Q_1^0) S_1 (S_1 + P_1^0)^{-1}]. \quad (45)$$

We may now form the *new test statistic*

$$\bar{T} := \frac{(\bar{R} - R) / tr[(I_l + N_{11}Q_1^0) S_1 (S_1 + P_1^0)^{-1}]}{(\Omega + R)/(n - m + l)} \quad (46)$$

in order to test the *null hypothesis*

$$H_0 : e_1^0 = 0 \quad (47)$$

against the alternative

$$H_a : e_1^0 \neq 0 \quad (48)$$

while still keeping (5) unchanged, *i.e.* $e_1^0 \sim (0, \Sigma_1^0)$.

On the other hand, if $\bar{T}$ ends up being smaller than the appropriate fractile of the *F*-distribution, then we should also be allowed to conclude that it would be smaller than the fractile of the actual (but unknown) distribution, thus enabling us to *at least accept* the null hypothesis H_0 in (47). It will be harder to argue the opposite direction, namely when to reject H_0, without the knowledge of approximate fractiles for $\bar{T}$.

Note that any test for consistency in the form of the null hypothesis

$$H_0 : e_1^0 \sim \mathcal{N}(0, \Sigma_1^0 = \sigma_0^2 Q_1^0), \quad (49)$$

based on the comparison of $\bar{R} / E\{\bar{R}\}$ with $\Omega/(n - rkA)$ would lead to a decision less powerful than the original test (22).

4. <u>THE APPROACH BASED ON HELMERT'S TRANSFORMATION</u>

In this chapter we try to follow the "traditional" approach, which can be interpreted as a *free adjustment* of the new dataset y, followed by a so-called "*Helmert transformation*" with respect to the reference data $\hat{\xi}_1$ which, at the end, will not be changed. This is a two-step procedure in which the stochastic constraints (4)-(5) are used only to identify a unique solution within the solution space of the normal equations coming from the model (1)-(2) alone. Nevertheless, in this contribution, we shall derive all corresponding formulas on the basis of the full model (1)-(5), that explains some of the variations from the established formalism.

In accordance with *B. Schaffrin (1984)*, for instance, the above described estimation formulas can equivalently be based on the following set of *minimum constraints* as they refer to our original prior information:

$$\begin{aligned} z_0 &:= EK^T (KK^T)^{-1} \hat{\xi}_1 = \bar{K}\xi + e_0^0, \\ \bar{K} &:= EK^T (KK^T)^{-1} K, \\ rk\bar{K} &= m - rkA, \end{aligned} \quad (50)$$

$$\begin{aligned} &e_0^0 := EK^T (KK^T)^{-1} e_1^0 \sim (0, \sigma_0^2 EK^T Q_1^0 KE^T =: \sigma_0^2 (P_0^0)^{-1}), \\ &C\{e_0^0, e\} = 0, \end{aligned} \quad (51)$$

where E is any $(m - q) \times m$ matrix with

$$AE^T = 0, \quad rkA + rkE = m, \quad q := rkA, \quad (52)$$

and $K^T (KK^T)^{-1} K = K^T K$ represents a "selection matrix." In combination with (1)-(2), we therefore obtain

$$\bar{\bar{\xi}} = (N + \bar{K}^T P_0^0 \bar{K})^{-1} (c + \bar{K}^T P_0^0 z_0) \quad (53)$$

$$= [N + K^T (KE^T P_0^0 EK^T) K]^{-1} [c + K^T (KE^T P_0^0 EK^T) \hat{\xi}_1] \quad (54)$$

where, in comparison to (6), only the matrix P_1^0 is to be replaced by $KE^T P_0^0 EK^T$. The result would not even change if we replaced P_0^0 here by the identity matrix I_{m-q}:

$$\overline{\overline{\xi}} = [N + K^T (KE^T EK^T) K]^{-1} [c + K^T (KE^T EK^T) \hat{\xi}_1]. \quad (55)$$

Note that $\overline{\overline{\xi}}$ would *not possess* the "reproducing property", in general; this is why we have to modify this solution by using the residual vector

$$\overline{e}_1^0 = \hat{\xi}_1 - K\overline{\overline{\xi}} = \hat{\xi}_1 - \overline{\overline{\xi}}_1$$
$$= [I_l - K(N + \overline{K}^T P_0^0 \overline{K})^{-1} \overline{K}^T P_0^0 EK^T] \hat{\xi}_1 - K(N + \overline{K}^T P_0^0 \overline{K})^{-1} c \quad (56)$$

in order to obtain

$$\hat{\overline{\overline{\xi}}} = \overline{\overline{\xi}} + K^T (KK^T)^{-1} \overline{e}_1^0 = \overline{\overline{\xi}} + K^T \overline{e}_1^0 \quad (57)$$

$$= K^T \hat{\xi}_1 + (I_m - K^T K)(N + \overline{K}^T P_0^0 \overline{K})^{-1} (c + \overline{K}^T P_0^0 EK^T \hat{\xi}_1) \quad (58)$$

in analogy to (31). It is now easy to check that

$$K\hat{\overline{\overline{\xi}}} = K\overline{\overline{\xi}} + \overline{e}_1^0 = \hat{\xi}_1, \quad D\{K\hat{\overline{\overline{\xi}}}\} = D\{\hat{\xi}_1\} = \sigma_0^2 Q_1^0, \quad (59)$$

holds true, indeed.

Its dispersion matrix will now be derived stepwise. First, we get

$$D\{\overline{\overline{\xi}}\} = \sigma_0^2 (N + \overline{K}^T P_0^0 \overline{K})^{-1} \cdot$$
$$[N + \overline{K}^T P_0^0 (EK^T Q_1^0 KE^T) P_0^0 \overline{K}] \cdot (N + \overline{K}^T P_0^0 \overline{K})^{-1} \quad (60)$$

$$= \sigma_0^2 [N + K^T (KE^T P_0^0 EK^T) K]^{-1} = \sigma_0^2 (N + \overline{K}^T P_0^0 \overline{K})^{-1} \quad (61)$$

which obviously resembles (8). Note that the cofactor matrix is, in fact, a *non-singular g-inverse* of N for which we have

$$N(N + \overline{K}^T P_0^0 \overline{K})^{-1} N = N - N[(N + \overline{K}^T P_0^0 \overline{K})^{-1} \overline{K}^T P_0^0] \overline{K}$$
$$= N - N[E^T (\overline{K}E^T)^{-1}] \overline{K} = N. \quad (62)$$

But it is *no longer a reflexive g-inverse* since we introduced $\hat{\xi}_1$ as random vector. Furthermore, we obtain the relations

$$N(N + \overline{K}^T P_0^0 \overline{K})^{-1} \overline{K}^T = N[E^T (\overline{K}E^T)^{-1}](P_0^0)^{-1} = 0 \quad (63)$$

and

$$\overline{K}(N + \overline{K}^T P_0^0 \overline{K})^{-1} \overline{K}^T = \overline{K}[E^T (\overline{K}E^T)^{-1}](P_0^0)^{-1} = (P_0^0)^{-1} \quad (64)$$

by exploiting the *important identity*

$$(N + \overline{K}^T P_0^0 \overline{K})^{-1} \overline{K}^T P_0^0 = E^T (\overline{K}E^T)^{-1} \quad (65)$$

which holds true because of $NE^T = 0$. Consequently, the dispersion matrix of $\overline{\overline{\xi}}$ can be rewritten from (60) as

$$D\{\overline{\overline{\xi}}\} = \sigma_0^2 (N + \overline{K}^T P_0^0 \overline{K})^{-1} N (N + \overline{K}^T P_0^0 \overline{K})^{-1}$$
$$+ \sigma_0^2 E^T (\overline{K}E^T)^{-1} (P_0^0)^{-1} (E\overline{K}^T)^{-1} E \quad (66)$$

$$= \sigma_0^2 N_{rs}^- + \sigma_0^2 E^T (E\overline{K}^T P_0^0 \overline{K}E^T)^{-1} E \quad (67)$$

with the reflexive symmetric g-inverse N_{rs}^- defined by

$$N_{rs}^- := (N + \overline{K}^T P_0^0 \overline{K})^{-1} N (N + \overline{K}^T P_0^0 \overline{K})^{-1} \quad (68)$$

in agreement with *B. Schaffrin (1984, formula (51)).*

In a second step, we obtain the dispersion matrix of the residual vector $\overline{e}_1^0$, again using (65), as

$$D\{\overline{e}_1^0\} = D\{[I_l - KE^T (\overline{K}E^T)^{-1} EK^T] \hat{\xi}_1\} + D\{K(N + \overline{K}^T P_0^0 \overline{K})^{-1} c\}$$
$$= \sigma_0^2 [I_l - KE^T (\overline{K}E^T)^{-1} EK^T] Q_1^0 [I_l - KE^T (\overline{K}E^T)^{-1} EK^T]$$
$$+ \sigma_0^2 K(N + \overline{K}^T P_0^0 \overline{K})^{-1} N (N + \overline{K}^T P_0^0 \overline{K})^{-1} K^T$$
$$= \sigma_0^2 \{Q_1^0 - KE^T (\overline{K}E^T)^{-1} EK^T Q_1^0$$
$$- Q_1^0 KE^T (E\overline{K}^T)^{-1} EK^T + K(N + \overline{K}^T P_0^0 \overline{K})^{-1} K^T\}. \quad (69)$$

Finally, the covariance matrix between $\overline{\overline{\xi}}$ and $\overline{e}_1^0$ can be computed as

$$C\{\overline{\overline{\xi}}, \overline{e}_1^0\} = C\{(N + \overline{K}^T P_0^0 \overline{K})^{-1} (c + \overline{K}^T P_0^0 EK^T \hat{\xi}_1), \overline{e}_1^0\}$$
$$= -\sigma_0^2 (N + \overline{K}^T P_0^0 \overline{K})^{-1} N (N + \overline{K}^T P_0^0 \overline{K})^{-1} K^T$$
$$+ \sigma_0^2 (N + \overline{K}^T P_0^0 \overline{K})^{-1} \overline{K}^T P_0^0 EK^T Q_1^0 [I_l - KE^T P_0^0 \overline{K}(N + \overline{K}^T P_0^0 \overline{K})^{-1} K^T]$$
$$= \sigma_0^2 [E^T (\overline{K}E^T)^{-1} EK^T Q_1^0 - (N + \overline{K}^T P_0^0 \overline{K})^{-1} K^T]. \quad (70)$$

Combining (60)-(61) with (66)-(70) now yields the *dispersion matrix of the modified estimator* $\hat{\overline{\overline{\xi}}}$ via

$$D\{\hat{\overline{\overline{\xi}}}\} = D\{\overline{\overline{\xi}}\} + C\{\overline{\overline{\xi}}, K^T \overline{e}_1^0\} + C\{K^T \overline{e}_1^0, \overline{\overline{\xi}}\} + D\{K^T \overline{e}_1^0\}$$
$$= \sigma_0^2 [(I_m - K^T K)(N + \overline{K}^T P_0^0 \overline{K})^{-1} (I_m - K^T K) + K^T Q_1^0 K$$
$$+ (I_m - K^T K) E^T (\overline{K}E^T)^{-1} E (K^T Q_1^0 K)$$
$$+ (K^T Q_1^0 K) E^T (E\overline{K}^T)^{-1} E (I_m - K^T K)] \quad (71)$$

which ought to turn out having *larger trace* if compared with (32)-(33).

It is worth noting that the *"reduced" residual vector* becomes

$$\overline{e}_0^0 = (EK^T) \overline{e}_1^0$$
$$= [I_{m-q} - \overline{K}(N + \overline{K}^T P_0^0 \overline{K})^{-1} \overline{K}^T P_0^0](EK^T) \hat{\xi}_1 - \overline{K}(N + \overline{K}^T P_0^0 \overline{K})^{-1} c$$
$$= [I_{m-q} - (P_0^0)^{-1} P_0^0](EK^T) \hat{\xi}_1 - \overline{K}(N + \overline{K}^T P_0^0 \overline{K})^{-1} N (N_{rs}^- c) = 0 \quad (72)$$

after applying (63)-(64), and its dispersion matrix will thus be

$$D\{\bar{\bar{e}}_0^0\} = D\{(EK^T)\bar{\bar{e}}_1^0\}$$
$$= \sigma_0^2[(EK^TQ_1^0KE^T) - \bar{K}E^T(\bar{K}E^T)^{-1}(EK^TQ_1^0KE^T)$$
$$-(EK^TQ_1^0KE^T)(E\bar{K}^T)^{-1}E\bar{K}^T + \bar{K}(N+\bar{K}^TP_0^0\bar{K})^{-1}\bar{K}^T]$$
$$= \sigma_0^2[(P_0^0)^{-1} - (P_0^0)^{-1} - (P_0^0)^{-1} + (P_0^0)^{-1}] = 0. \qquad (73)$$

On the other hand, formula (71) confirms the second part of (59), indeed, as we get

$$D\{K\hat{\bar{\bar{\xi}}}\} = \sigma_0^2 KK^TQ_1^0KK^T = \sigma_0^2Q_1^0 \qquad (74)$$

since

$$K(I_m - K^TK) = (I_l - KK^T)K = 0, \quad KK^T = I_l. \qquad (75)$$

We may now study the residual vector for the observations as defined by

$$\tilde{\bar{e}} := y - A\hat{\bar{\bar{\xi}}} = (y - A\bar{\bar{\xi}}) - AK^T(KK^T)^{-1}\bar{\bar{e}}_1^0 =: \bar{\bar{e}} - A_1\bar{\bar{e}}_1^0, \qquad (76)$$

and the residual vector for the reference data given by

$$\tilde{\bar{e}}_1^0 := \hat{\xi}_1 - K\hat{\bar{\bar{\xi}}} = \hat{\xi}_1 - K\bar{\bar{\xi}} - KK^T(KK^T)^{-1}\bar{\bar{e}}_1^0 = 0 \qquad (77)$$

which vanishes following (56) with

$$D\{\tilde{\bar{e}}_1^0\} = 0, \quad C\{\tilde{\bar{e}}_1^0, \tilde{\bar{e}}\} = 0. \qquad (78)$$

In order to compute the dispersion matrix $D\{\tilde{\bar{e}}\}$ we first determine

$$D\{\bar{\bar{e}}\} = D\{y - A\bar{\bar{\xi}}\}$$
$$= D\{[I_n - A(N+\bar{K}^TP_0^0\bar{K})^{-1}A^TP]y - A(N+\bar{K}^TP_0^0\bar{K})^{-1}\bar{K}^TP_0^0EK^T\hat{\xi}_1\}$$
$$= \sigma_0^2\{[I_n - A(N+\bar{K}^TP_0^0\bar{K})^{-1}A^TP]P^{-1}[I_n - PA(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T]$$
$$+ A(N+\bar{K}^TP_0^0\bar{K})^{-1}\bar{K}^TP_0^0(EK^TQ_1^0KE^T)P_0^0\bar{K}(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T\}$$
$$= \sigma_0^2[P^{-1} - A(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T], \qquad (79)$$

and secondly

$$C\{\bar{\bar{e}}_1^0, .\bar{\bar{e}}\} = -C\{K(N+\bar{K}^TP_0^0\bar{K})^{-1}c, y - A(N+\bar{K}^TP_0^0\bar{K})^{-1}c\}$$
$$- C\{\hat{\xi}_1 - K(N+\bar{K}^TP_0^0\bar{K})^{-1}\bar{K}^TP_0^0EK^T\hat{\xi}_1, A(N+\bar{K}^TP_0^0\bar{K})^{-1}\bar{K}^TP_0^0EK^T\hat{\xi}_1\}$$
$$= -\sigma_0^2\{K(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T - K(N+\bar{K}^TP_0^0\bar{K})^{-1}N(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T$$
$$+ Q_1^0KE^TP_0^0\bar{K}(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T$$
$$- K(N+\bar{K}^TP_0^0\bar{K})^{-1}\bar{K}^TP_0^0(EK^TQ_1^0KE^T)P_0^0\bar{K}(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T\}$$
$$= -\sigma_0^2\{K(N+\bar{K}^TP_0^0\bar{K})^{-1}\bar{K}^TP_0^0\bar{K}(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T$$
$$+ Q_1^0KE^T(E\bar{K}^T)^{-1}EA^T$$
$$- K(N+\bar{K}^TP_0^0\bar{K})^{-1}\bar{K}^TP_0^0\bar{K}(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T\} = 0 \qquad (80)$$

using (52) in combination with (65). Therefore, we can readily conclude that the dispersion matrix of $\tilde{\bar{e}}$ is represented by

$$D\{\tilde{\bar{e}}\} = D\{\bar{\bar{e}}\} + D\{A_1\bar{\bar{e}}_1^0\}$$
$$= \sigma_0^2[P^{-1} - A(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T + A(K^TQ_1^0K)A^T$$
$$- A\bar{K}^T(E\bar{K}^T)^{-1}E(K^TQ_1^0K)A^T - A(K^TQ_1^0K)E^T(\bar{K}E^T)^{-1}\bar{K}A^T$$
$$+ A(K^TK)(N+\bar{K}^TP_0^0\bar{K})^{-1}(K^TK)A^T]$$

$$= \sigma_0^2[P^{-1} - A(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T] \qquad (81)$$
$$+ \sigma_0^2A[I_m - \bar{K}^T(E\bar{K}^T)^{-1}E]K^TQ_1^0K[I_m - E^T(\bar{K}E^T)^{-1}\bar{K}]A^T$$
$$+ \sigma_0^2A(K^TK)(N+\bar{K}^TP_0^0\bar{K})^{-1}(K^TK)A^T - \sigma_0^2A\bar{K}^T(E\bar{K}^TP_0^0\bar{K}E^T)^{-1}\bar{K}A^T$$

$$= \sigma_0^2[P^{-1} - A(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T + A(K^TKN_{rs}^-K^TK)A^T] \qquad (82)$$
$$+ \sigma_0^2A[I_m - \bar{K}^T(E\bar{K}^T)^{-1}E]K^TQ_1^0K[I_m - E^T(\bar{K}E^T)^{-1}\bar{K}]A^T,$$

and the covariance matrix between $\hat{\bar{\bar{\xi}}}$ and $\tilde{\bar{e}}$ by

$$C\{\hat{\bar{\bar{\xi}}}, \tilde{\bar{e}}\} = C\{\bar{\bar{\xi}} + K^T\bar{\bar{e}}_1^0, \bar{\bar{e}} - AK^T\bar{\bar{e}}_1^0\}$$
$$= C\{\bar{\bar{\xi}}, \bar{\bar{e}}\} + C\{K^T\bar{\bar{e}}_1^0, \bar{\bar{e}}\} - C\{\bar{\bar{\xi}}, AK^T\bar{\bar{e}}_1^0\} - D\{K^T\bar{\bar{e}}_1^0\}\cdot A^T$$
$$= 0 - C\{\bar{\bar{\xi}}, \bar{\bar{e}}_1^0\}\cdot KA^T - K^T\cdot D\{\bar{\bar{e}}_1^0\}\cdot KA^T$$

$$= \sigma_0^2[(N+\bar{K}^TP_0^0\bar{K})^{-1}K^TKA^T - E^T(\bar{K}E^T)^{-1}E(K^TQ_1^0K)A^T - (K^TQ_1^0K)A^T$$
$$+ \bar{K}^T(E\bar{K}^T)^{-1}E(K^TQ_1^0K)A^T + (K^TQ_1^0K)E^T(\bar{K}E^T)^{-1}\bar{K}A^T$$
$$- K^TK(N+\bar{K}^TP_0^0\bar{K})^{-1}K^TKA^T]$$

$$= \sigma_0^2\{(I_m - K^TK)(N+\bar{K}^TP_0^0\bar{K})^{-1}K^TKA^T$$
$$-[I_m - \bar{K}^T(E\bar{K}^T)^{-1}E]K^TQ_1^0K[I_m - E^T(\bar{K}E^T)^{-1}\bar{K}]A^T$$
$$+ \bar{K}^T(E\bar{K}^TP_0^0\bar{K}E^T)^{-1}\bar{K}A^T - E^T(\bar{K}E^T)^{-1}E(K^TQ_1^0K)A^T\} \qquad (83)$$

using the fact that $\bar{\bar{\xi}}$ and $\bar{\bar{e}}$ are *uncorrelated* according to

$$C\{\bar{\bar{\xi}}, \bar{\bar{e}}\} = C\{\bar{\bar{\xi}}, y\} - D\{\bar{\bar{\xi}}\}\cdot A^T$$
$$= \sigma_0^2(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T - \sigma_0^2(N+\bar{K}^TP_0^0\bar{K})^{-1}A^T = 0. \qquad (84)$$

Apparently the following orthogonality relation holds true as well:

$$A^TP\bar{\bar{e}} = c - N\bar{\bar{\xi}} = c - N(N+\bar{K}^TP_0^0\bar{K})^{-1}(c + \bar{K}^TP_0^0z_0)$$
$$= c - N(N+\bar{K}^TP_0^0\bar{K})^{-1}c - NE^T(\bar{K}E^T)^{-1}z_0 = 0 \qquad (85)$$

due to (54), (62)-(64) and (65), which nicely complements the earlier result in (72).

Applying (76), we then obtain

$$A_1^TP\tilde{\bar{e}} = A_1^TP(\bar{\bar{e}} - A_1\bar{\bar{e}}_1^0) = -N_{11}\bar{\bar{e}}_1^0, \quad A_2^TP\tilde{\bar{e}} = -N_{21}\bar{\bar{e}}_1^0, \qquad (86)$$

which is somewhat in contrast to formulas (40) for the *optimal* linear estimate with the reproducing property. Consequently, we may derive an *invariant quadratic uniformly unbiased estimate* σ_0^2 from the sum of weighted squared residuals $\tilde{\bar{e}}$ and $\tilde{\bar{e}}_1^0 = 0$, which is easily calculated as

$$\begin{aligned}\Omega + \tilde{\bar{R}} &:= \tilde{\bar{e}}^T P \tilde{\bar{e}} + (\tilde{\bar{e}}_1^0)^T P_1^0 \tilde{\bar{e}}_1^0 = (\bar{\bar{e}} - A_1 \bar{\bar{e}}_1^0)^T P (\bar{\bar{e}} - A_1 \bar{\bar{e}}_1^0) \\ &= \bar{\bar{e}}^T P \bar{\bar{e}} + (\bar{\bar{e}}_1^0)^T N_{11} \bar{\bar{e}}_1^0 = y^T P \bar{\bar{e}} + (\bar{\bar{e}}_1^0)^T N_{11} \bar{\bar{e}}_1^0 \\ &= [y^T P y - c^T (N + \bar{K}^T P_0^0 \bar{K})^{-1} c] - c^T E^T (\bar{K} E^T)^{-1} z_0 \\ &\quad + (\bar{\bar{e}}_1^0)^T N_{11} \bar{\bar{e}}_1^0 \\ &= \Omega - 0 + (\bar{\bar{e}}_1^0)^T N_{11} \bar{\bar{e}}_1^0 \end{aligned} \tag{87}$$

with

$$E\{\Omega\} = tr(P \cdot D\{\bar{\bar{e}}\}) = \sigma_0^2 (n - rkA) = \sigma_0^2 (n - q) \tag{88}$$

and

$$\begin{aligned} E\{\tilde{\bar{R}}\} &= tr(N_{11} \cdot D\{\bar{\bar{e}}_1^0\}) = tr(KNK^T \cdot D\{\bar{\bar{e}}_1^0\}) \\ &= \sigma_0^2 \cdot tr\{KNK^T Q_1^0 - KN\bar{K}^T (E\bar{K}^T)^{-1} EK^T Q_1^0 \\ &\quad - KNK^T Q_1^0 KE^T (\bar{K}E^T)^{-1} EK^T + KNK^T K (N + \bar{K}^T P_0^0 \bar{K})^{-1} K^T\} \\ &= \sigma_0^2 \cdot tr\{N[I_m - \bar{K}^T (E\bar{K}^T)^{-1} E](K^T Q_1^0 K)[I_m - E^T (\bar{K}E^T)^{-1} \bar{K}]\} \\ &\quad + \sigma_0^2 \cdot tr\{NK^T K (N + \bar{K}^T P_0^0 \bar{K})^{-1} [(N + \bar{K}^T P_0^0 \bar{K}) - \bar{K}^T P_0^0 \bar{K}] \\ &\quad (N + \bar{K}^T P_0^0 \bar{K})^{-1} K^T K\} \\ &= \sigma_0^2 \cdot \{trN_{11} [I_l - KE^T (\bar{K}E^T)^{-1} EK^T] Q_1^0 [I_l - KE^T (\bar{K}E^T)^{-1} EK^T] \\ &\quad + trN_{11} (KN_{rs}^- K^T)\} \end{aligned} \tag{89}$$

after applying (69) and further simplifications. Thus, our *estimated variance component* will turn out as

$$\hat{\hat{\sigma}}_0^2 = (\Omega + \tilde{\bar{R}}) / [n - q + E\{\tilde{\bar{R}}\}], \tag{90}$$

and the *appropriate test statistic* as

$$\tilde{\bar{T}} := \frac{(\tilde{\bar{R}} - R) / [E\{\tilde{\bar{R}}\} - l + (m - q)]}{(\Omega + R) / [(n - q) + l - (m - q)]} \tag{91}$$

by means of which we may check the *null hypothesis*

$$H_0 : e_1^0 = 0 \tag{92}$$

against the alternative

$$H_a : e_1^0 \neq 0 \tag{93}$$

while still keeping (5) unchanged, *i.e.* $e_1^0 \sim (0, \Sigma_1^0)$. It is an open question whether $\tilde{\bar{R}} - R$ is *statistically independent* of $\Omega + R$; it rather looks unlikely so that $\tilde{\bar{T}}$ will not strictly follow a *F*-distribution. We may, nevertheless, use the *F*-fractiles as *lower bounds* for those of the true, but unknown, distribution. In this way we are able to at least make a decision of *accepting the null hypothesis* (92) in case the lower bound is not surpassed by $\tilde{\bar{T}}$. Unfortunately, we cannot reject (92) simply because $\tilde{\bar{T}}$ turns out to be larger than this threshold.

Again, as in Chapter 3, we would *not* use a comparison of $\tilde{\bar{R}} / E\{\tilde{\bar{R}}\}$ and $\Omega/(n - q)$ to check consistency in the sense of (49).

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

To illustrate the theory of non-hierarchical and hierarchical data fusion in a photogrammetric application, let us consider the object space point field shown in Figure 1. The field consists of two distinct arrangements of points:

1) a ring, centered at the origin and consisting of 20 points labeled R1-R20, representing points to be added to, and
2) a box-like framework, also centered at the origin and consisting of 27 points labeled C21-C47.

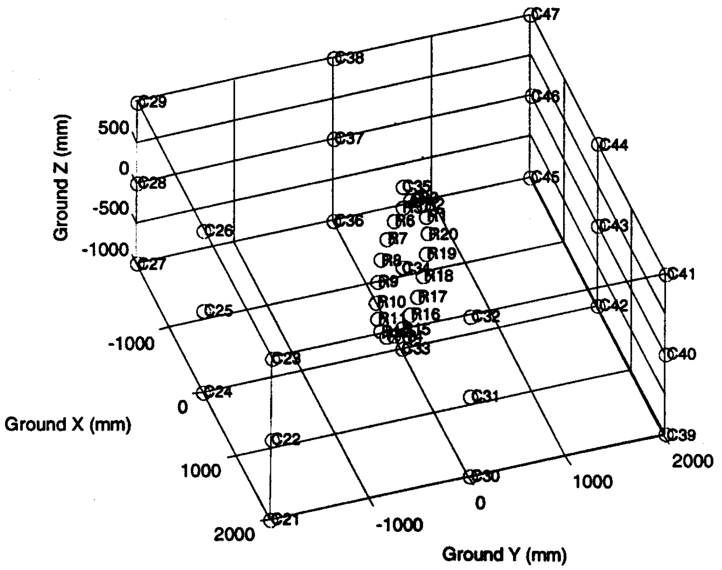

Fig. 1. - Object Space Point Field

The small scale network, which we denote FAR3, consists of three images with a total of 81 observations of points C21-C47 (see Table 1). This network contained no observations of the ring points. The mean scale of the images in this network is approximately 235:1. We will consider FAR3 our reference network.

TABLE 1 - FAR3 EXTERIOR ORIENTATION PARAMETERS

FAR3	Xo (mm) / omega (deg)	Yo (mm) / phi (deg)	Zo (mm) / kappa (deg)	Range (mm)
far1	6000	-10000	8000	14,142
	51.35	25.10	18.74	
far2	6000	10000	8000	11,313
	-51.35	25.10	-18.74	
far3	8000	0	8000	14,142
	0	45.00	-90.00	

The large scale network, which we denote NEAR3, also consists of three images but with a mean scale of approximately 88:1 (see Table 2). The NEAR3 photographs contain observations of points R1-R20 and, because of their larger scale, only points C24, C25, C27, C33, C34, C35, C36 and C43 of the reference field. These eight points are the only common parameters between the two networks.

TABLE 2 – NEAR3 EXTERIOR ORIENTATION PARAMETERS

NEAR3	Xo (mm) / omega (deg)	Yo (mm) / phi (deg)	Zo (mm) / kappa (deg)	Range (mm)
near1	2000	2500	2000	3,775
	-51.00	32.00	-22.99	
near2	2000	-4500	2000	5,315
	66.00	22.10	9.51	
near3	2000	4500	2000	5,315
	-66.00	22.00	-9.51	

We applied the following, normally distributed, random errors to generate image coordinates of the object space point field:

Exterior Orientation Parameters, rotation angles:
$\sigma = \pm 5$ arc seconds
Exterior Orientation Parameters, camera position:
$\sigma = \pm 100$ mm
Interior Orientation Parameters, (x_p, y_p, c):
$\sigma = \pm 0.01$ mm
Image Coordinates after projection:
$\sigma = \pm 0.005$ 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 lens distortion or field unflatness.

The experiment consisted of one adjustment followed by three updates.

1. The free network adjustment of FAR3. This provided a reference framework of points C21-C47 (i.e. the *a priori* information contained in $\hat{\xi}_1$ and $D\{\hat{\xi}_1\}$).
2. The update of the FAR3 reference framework with observations from NEAR3 using the non-hierarchical procedures described by equations (6) and (8) to obtain $\hat{\hat{\xi}}$ and $D\{\hat{\hat{\xi}}\}$, respectively.
3. The update of the FAR3 reference framework with the optimal hierarchical procedure described by equations (31) and (32) to obtain $\bar{\xi}$ and $D\{\bar{\xi}\}$, respectively.
4. The update of the FAR3 reference framework with the non-optimal hierarchical procedure described by equations (57) and (71) to obtain $\hat{\bar{\xi}}$ and $D\{\hat{\bar{\xi}}\}$, respectively.

We consider all involved object space points and exterior orientation elements as parameters to be estimated. All adjustments treated interior orientation elements (x_p,y_p, and c) and parameters with pseudo-observations and variance of 0.0001mm^2, essentially fixing them at their *a priori* values. This small variance, and associated large weight, did not significantly affect the condition number of the normal matrices. We assumed the *a priori* image coordinate observation precision to be ±0.005mm. We further assumed the *a priori* variance component to be 1.00.

The results of the three update adjustments are shown in Table 3.

TABLE 3 - VARIANCE COMPONENT COMPUTATIONS

Update Type	Sum of Residuals Squared / Variance Component Estimate	Expectation σ_0^2 / Test Statistic	F-fractile $\alpha = 0.01$
(computed from each update)	$\Omega = 53.145$	65	N/A
	N/A	N/A	
Stochastic Update	$R = 7.421$	102	0.60
	0.363	T = 0.089	
Optimal Reproducing Update	$\bar{R} - R = 47.613$	112.442	0.66
	0.387	$\bar{T} = 1.166$	
Helmert Reproducing Update	$\bar{\bar{R}} - R = 126.188$	208.439	0.73
	0.4973	$\bar{\bar{T}} = 1.668$	

The point-by-point errors for each adjustment are shown in Figures 2, 3, 4, and 5. We define RMS as the distance between the true point locations and the estimated location, and show it in the graph as a triangle. The square root of the trace of the cofactor matrix of each point is shown as an open circle. Finally, the Helmert point error is shown as a closed circle.

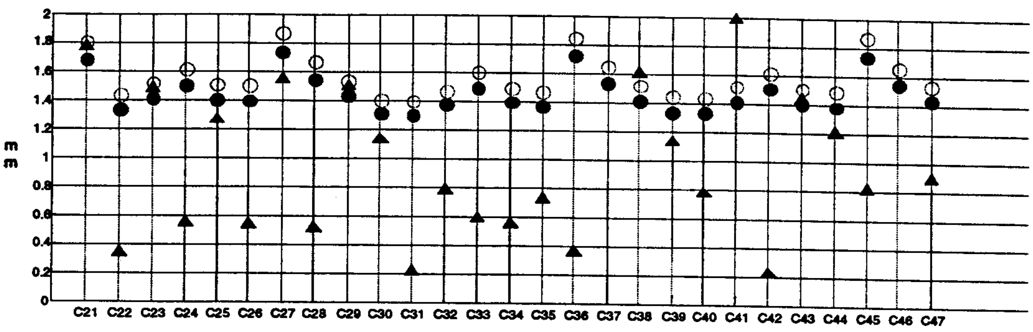

Fig. 2 - Reference Network

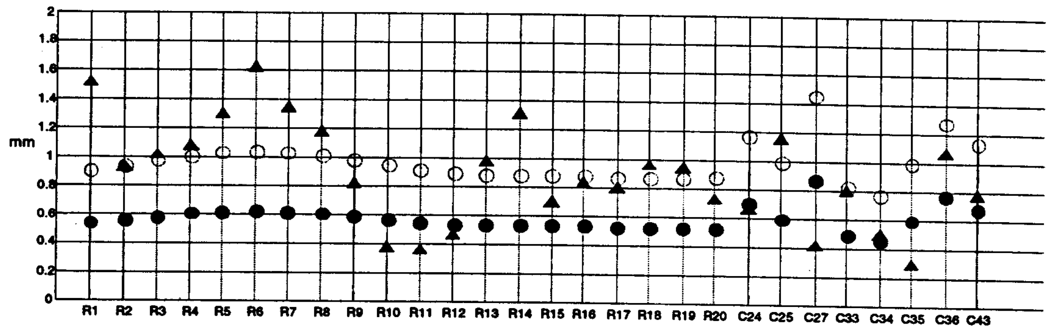

Fig. 3 - Stochastic Update (Chapter 2)

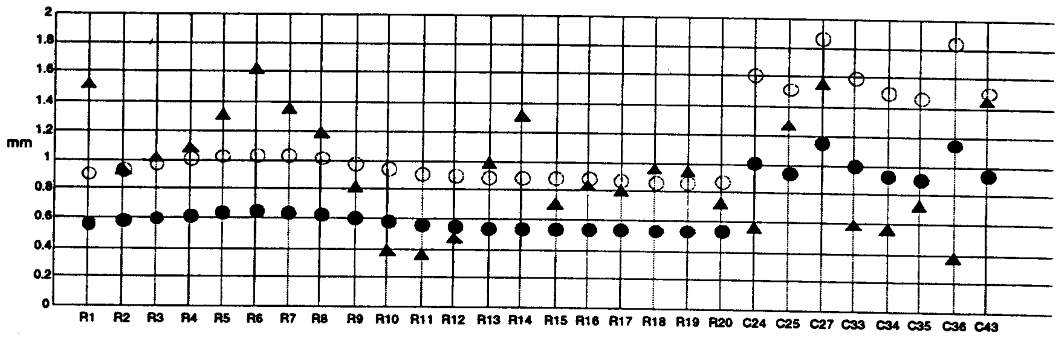

Fig. 4 - Optimal Reproducing Method (Chapter 3)

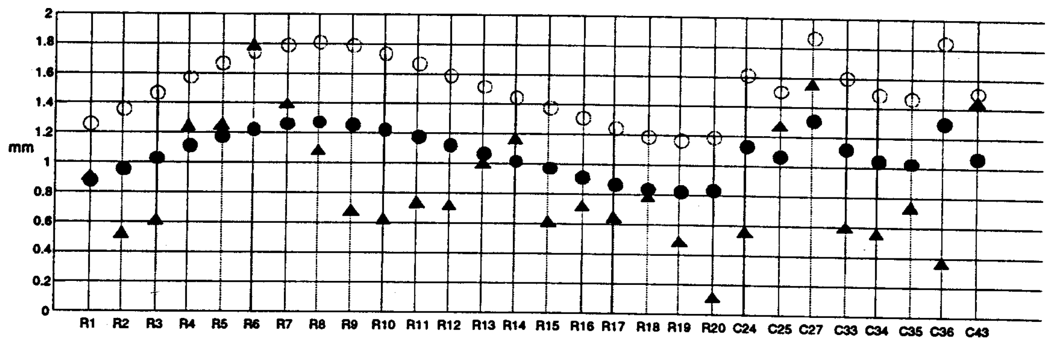

Fig. 5 - Helmert Reproducing Method (Chapter 4)

6. CONCLUSIONS

In this paper we have studied data fusion in a hierarchical mode. The traditional approach based on Helmert's transformation has been found non-optimal (in the sense of minimum mean square error); the optimum, reference-data reproducing estimator has been presented and compared with the above as well as with the optimal non-hierarchical estimator. The following conclusions may be drawn with particular view of the photogrammetric example in Chapter 5.

(i) The *non-hierarchical* approach yields Helmert point errors that tend to *underestimate* the true deviations in the non-reference points quite drastically. This is particularly surprising as the test clearly *confirms the hypothesis of consistency* between new and reference data. Consequently, the reference points would experience a certain improvement, both in terms of true deviations as well as Helmert point errors.

(ii) In the *hierarchical* approach, none of the reference points involved will be affected by the adjustment. The *optimal reproducing method* will provide the same results for the non-reference points as the non-hierarchical approach except for the slightly enlarged variance component estimate. Thus, the true deviations will still be mostly *underestimated* by the Helmert point errors.

(iii) In contrast, for the *Helmert reproducing method* we find a much better agreement between true deviations and Helmert point errors of the non-reference points, while the mean square errors have grown considerably. It also seems that the true deviations themselves have been *reduced somewhat* through the process.

(iv) The tests for both hierarchical procedures, although (strictly speaking) inconclusive, would *hint toward rejecting* the hierarchical approach in favor of the non-hierarchical one. At this point it is still unclear why this should be so, in view of the previously established *consistency*.

(v) Further studies are also necessary to *explain the mechanism* that lets the Helmert reproducing method perform so well in comparison, although on paper it is known to be "outperformed" by the optimal reproducing method.

(vi) It would also be nice to learn more about the true (or at least approximate) *distributions of the test statistics* in the hierarchical mode. However, this must be postponed for future investigations.

7. REFERENCES

ADUOL, F.W.O. A static-dynamic model for densification of geodetic networks, **AVN-Int'l. Edition**, 10, pp.20-36, 1993.

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

BUITEN, H.J. Junction of nets by collocation, **Manus. Geodaet.** 3, pp.253-297, 1978.

GRAFAREND, E. AND B. SCHAFFRIN. From the static to the dynamic treatment of geodetic networks (in German), **ZfV** 113, pp.79-103, 1988.

KOCH, K.R. The choice of the datum of a trigonometric network for the determination of new points (in German), **ZfV** 108, pp.104-111, 1983.

SCHAFFRIN, B. On the use of prior information in free networks (in German); in: K.Rinner et al. (eds.), **Ingenieurvermessung '84**, Dümmler: Bonn, B6/1-15, 1984.

SCHAFFRIN, B. Fiducial versus fixed points in the GPS network approach; in: **Proc. of the 5th. Int'l. Geodetic Symp. on Satellite Positioning**, Las Cruces, NM, Vol. I, 1989, pp.500-511.

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

SCHAFFRIN, B. Reproducing estimates via least squares: An optimal alternative to Helmert transformation. In: E. W. Grafarend; F.W. Krumm and V.S. Schwarzs (eds). **Geodesy – The Challenge of the 3rd. Millennium**, Springer, Berlin/Heidelberg/New York, pp.387-392, 2002.

WIKLE, C.K., L.M. BERLINER AND N. CRESSIE. Hierarchical Bayesian Space**-Time Models, Environmental and Ecological Statistics**, 5, pp.117-154, 1998.

8. ACKNOWLEDGEMENT

Essential parts of this paper were developed during the visit of the first author at the Institute of Photogrammetry, University of Bonn, Germany with Professor Wolfgang Förstner as his host. This is gratefully acknowledged.

Received: 02 February 2004 – Accepted: 17 March 2004.