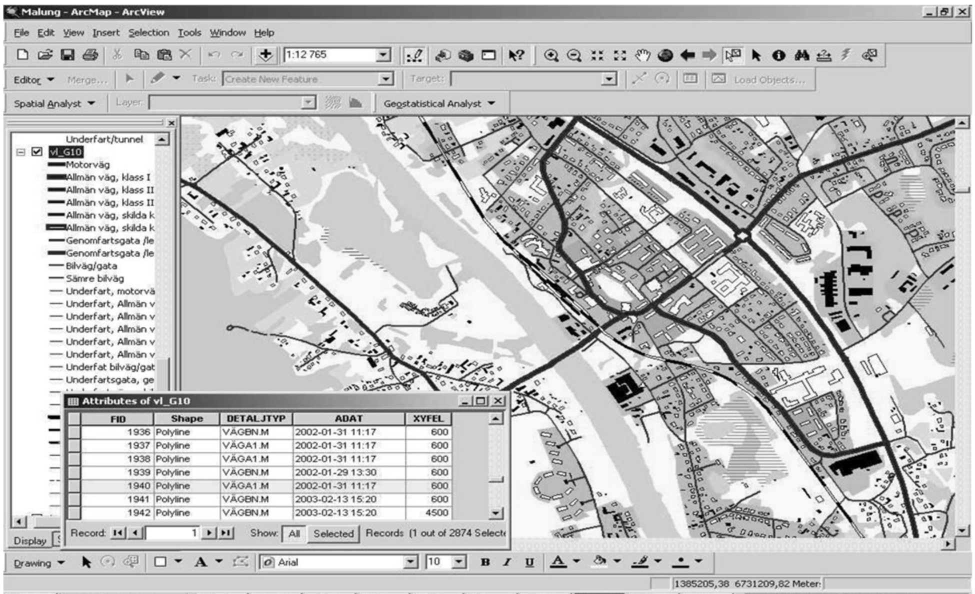
AI1149 Geodata quality and time series analysis

Huaan Fan

hfan@kth.se, 08-7907340, 0739-442712

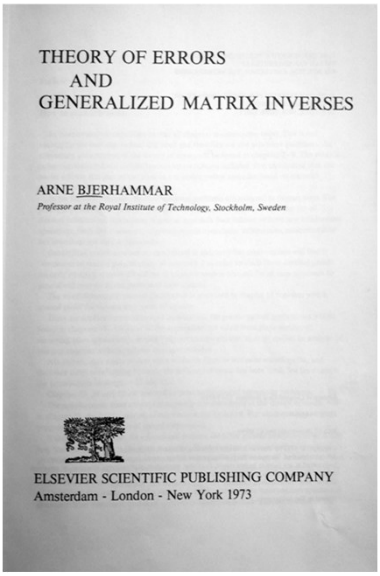
Geodata: maps and/or digital info



Geodata are derived from field measurements

”Felteori”, Arne Bjerhammar, KTH, 1964

Theory of Errors and Generalized Matrix Inverses, Elsevier, 1973



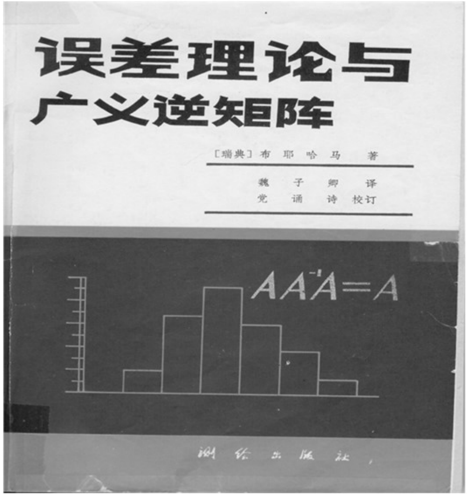
THEORY OF ERRORS
AND
GENERALIZED MATRIX INVERSES

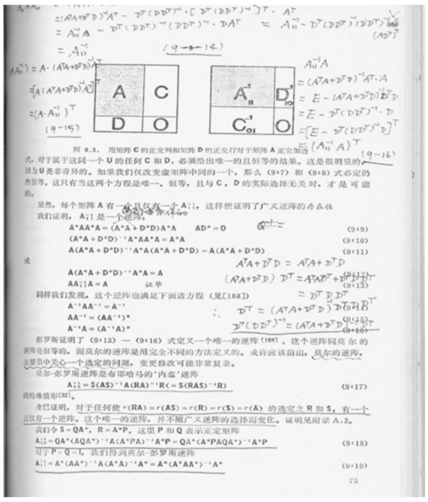
ARNE BJERHAMMAR
Professor at the Royal Institute of Technology, Stockholm, Sweden

ELSEVIER SCIENTIFIC PUBLISHING COMPANY
Amsterdam - London - New York 1973

Arne Bjerhammar

Theory of Errors and Generalized Matrix Inverses, Elsevier, 1973





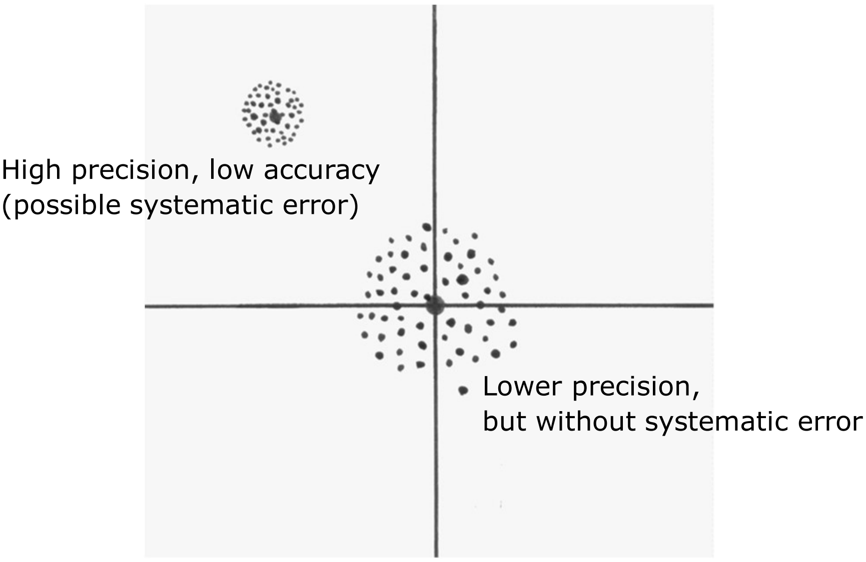
Quality of geodata

- Geo-databases: formats, algorithms, accessibility
  - raster/vector data, advantages/disadvantages
  - algorithms - efficient storage/retrieval
  - access to geo-databases
  - ownership, security, regulations
  - National Spatial Data Infrastructure (NSDI)
- Geographic aspects
  - geographic coverage, resolution
  - georeferencing: geodetic reference/coordinate systems
  - completeness & correctness of attribute data
- Positional information (coordinates)
  - *errors from original measurements to computed quantities*
  - processing: reformatting, interpolation, extrapolation
  - temporal changes

Quality of measurements and geodata

- Precision
  - internal agreement among the measurements
  - a measure of repeatability of the measurements under same measurement condition
- Accuracy
  - a measure of agreement with the true value
- Reliability
  - capability against gross errors or systematic errors
  - reliable measurements are said to be *robust*, *stable*

The example of shooting

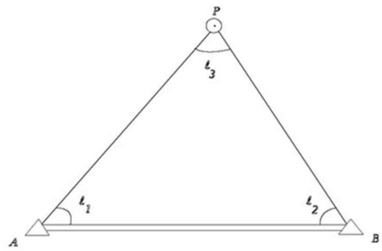


Objectives of theory of errors

- How to treat measurement errors ?
- How to assess quality of the measurements ?
- How to carry out **least squares adjustment** ?
  - eliminate the disagreements (*misclosures*) among the measurements **and** obtain optimal results from given measurements based on the least squares principle
- How to use theoretical insights to design good methods and practical procedures ?

How to detect existence of errors ?

- Compare repeated meaurements
- Check meaured values with theoretical (*mathematical* or *physical*) relations



$$w = \ell_1 + \ell_2 + \ell_3 - 180^0$$

$\longrightarrow$ Misclosure (*slutfel*)

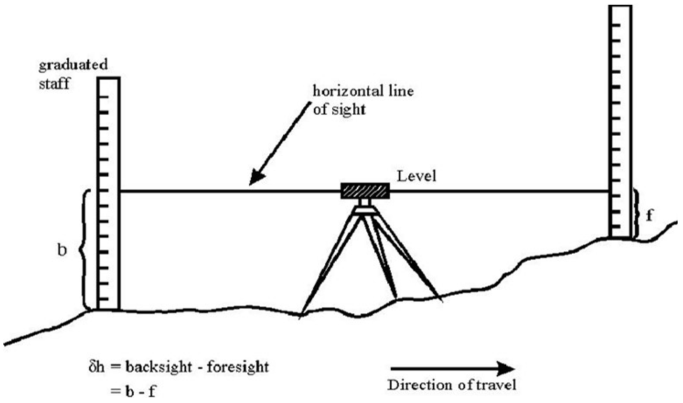
Classification of measurement errors

- Systematic errors
  - influences of instruments, environment or surveyors etc
  - eliminate the causes, reduce/eliminate the effect
  - apply theoretical correction or automatic detection
- Random errors (Gauss)

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} \varepsilon_i}{n} = 0$$

  - arithmatic mean approaches zero
  - equal chance for positive and negative errors
  - more small errors than larger errors
  - magnitude of errors is limited
  - use statistical methods: normal distribution
- Gross errors
  - mistakes which should be avoided. Automatic detection

Example of systematic errors: collimation error



Collimation error: the horizontal line of sight is not parallel to the horizontal axis of the levelling bubble.

Reduce the collimation error:
- calibrate instruments
- appy correction if the error is well known
- design good surveying procedures: e.g. put the level in the middle between the backsight and foresight staffs

Random errors: *Statistical way of thinking*

- Quality is defined by the overall measurement condition:
  - instruments
  - physical (e.g. weather) conditions
  - surveyor's skills
  - other specific conditions

- Statistical way of thinking
  - not meaningful to describe each individual random error
  - same quality under same measurement conditions
  - one or a fewer *collective* index numbers
  - statistical error index: mean square error (*medelfel*), standard error (*standardfel*), standard deviation (*standard avvikelse*)
  - variance (*varians*) = standard error squared

New ISO terminologies: **Uncertainty**

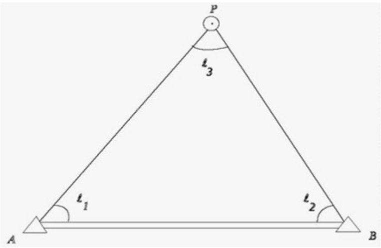
Joint Committee for Guides in Metrology (JCGM):
Evaluation of measurement data – Guide to the expression of uncertainty in measurement

- No true value, due to errors and incomplete description of the quantity to be measured
- Measurand, value of the measurand
- Uncertainty, instead of SE
- Uncertainty computed as SE
- Uncertainty as 2 or 3 times SE

True error, relative error vs standard error

- Absolute true error = measured – true value $\varepsilon = \ell - \widetilde{\ell}$
- Relative error = error divided by the quantity $\gamma = \left|\frac{\varepsilon}{\ell}\right| = \left|\frac{\ell - \widetilde{\ell}}{\ell}\right|$
- Standard error σ (squared), *theoretical* $\sigma^2 = \lim_{n \to \infty} \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n}$
- Standard error σ (squared), *estimated* $\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n}$

An example: triangulation misclosures



$$w = (\ell_1 + \ell_2 + \ell_3) - 180^0$$

$$\widehat{\sigma}_w{}^2 = \frac{1}{30} \sum_{i=1}^{30} w_i^2 = \frac{25.86}{30} \qquad ('''^2)$$

$$\widehat{\sigma}_w = 0.93''$$

Table 1.1: List of Triangular Misclosures

i	w_i ('')	i	w_i ('')	i	w_i ('')
1	+1.5	11	-2.0	21	-1.1
2	+1.0	12	-0.7	22	-0.4
3	+0.8	13	-0.8	23	-1.0
4	-1.1	14	-1.2	24	-0.5
5	+0.6	15	+0.8	25	+0.2
6	+1.1	16	-0.3	26	+0.3
7	+0.2	17	+0.6	27	+1.8
8	-0.3	18	+0.8	28	+0.6
9	-0.5	19	-0.3	29	-1.1
10	+0.6	20	-0.9	30	-1.3

$$\widehat{\sigma} = \frac{\widehat{\sigma}_w}{\sqrt{3}} = 0.54''$$

Definition of Weights

- Weight is inversely proportional to variance:

$$p_i = \frac{c_0}{\sigma_i^2}$$

- C_0 is an arbitrary positive number
- C_0 is the variance of a quantity with weight 1
- C_0 is called unit-weight standard error (σ_0) squared or variance factor

$$p_i = \frac{\sigma_0^2}{\sigma_i^2}$$

Empirical weighting

- *levelling:* $p_i = c_0$ divided by the length of the levelling line
- *distance measurement:* $p_i = c_0$ divided by the distance or distance squared
- *direction (angle) measurement:* $p_i =$ number of whole rounds measured divided by c_0

- *levelling:* $c_0 =$ length of the levelling line with weight 1 Often taken as 1 km
- *distance measurement:* $c_0 =$ distance or squared distance of weight 1
- *direction (angle) measurement:* $c_0 =$ number of whole rounds by which the direction (angle) of unit weight is measured

Computation of the variance factor σ_0^2

- Variance factor, *theoretical* $\sigma_0^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} p_i \varepsilon_i \varepsilon_i$
- Variance factor, *from finite number of errors* $\widehat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^{n} p_i \varepsilon_i \varepsilon_i$
- Variance factor, *computed from estimated errors* $\widehat{\sigma}_0^2 = \frac{1}{f} \sum_{i=1}^{n} p_i \widehat{\varepsilon}_i \widehat{\varepsilon}_i$

? $\widehat{\varepsilon}_i = \ell_i - \widehat{\ell}_i$ (= estimated error of ℓ_i);

f = number of redundant observations (statistical degrees of freedom)

Statistical terminologies: random variables

- Distribution function *F(x)* $P(\varepsilon \leq x) = F(x) = \int_{-\infty}^{x} f(\varepsilon) d\varepsilon \qquad (-\infty < x < +\infty)$
- Density function *f(x)* $\frac{\partial F(x)}{\partial x} = f(x)$
- Expectation (*average*) $E(\varepsilon) = \int_{-\infty}^{+\infty} \varepsilon \, f(\varepsilon) d\varepsilon$
- Variance $var(\varepsilon) = \sigma^2 = E\{[\varepsilon - E(\varepsilon)]^2\} = \int_{-\infty}^{+\infty} [\varepsilon - E(\varepsilon)]^2 \cdot f(\varepsilon) \cdot d\varepsilon$
- Covariance between 2 variables $\sigma_{12} = E\{[\varepsilon_1 - E(\varepsilon_1)][\varepsilon_2 - E(\varepsilon_2)]\}$
- Correlation coefficient $\rho_{12} = E\left[\frac{\varepsilon_1 - E(\varepsilon_1)}{\sigma_1} \frac{\varepsilon_2 - E(\varepsilon_2)}{\sigma_2}\right] = \frac{\sigma_{12}}{\sigma_1 \cdot \sigma_2}$

$-1 \leq \rho_{12} \leq +1$

Variance-covariance matrix of a random vector

$$\underset{n\cdot 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix}$$

Random errors often have zero mean (expectation)

$$E(\varepsilon_i) = 0$$
$$E\left\{[\varepsilon_i - E(\varepsilon_i)]^2\right\} = E(\varepsilon_i^2) = \sigma_i^2 \qquad (i,\ j = 1, 2, 3, \cdots, n)$$
$$E\left\{[\varepsilon_i - E(\varepsilon_i)]\,[\varepsilon_j - E(\varepsilon_j)]\right\} = E(\varepsilon_i \cdot \varepsilon_j) = \sigma_{ij}$$

$$\underset{n\cdot n}{C_{\varepsilon\varepsilon}} = E\left\{[\varepsilon - E(\varepsilon)]\,[\varepsilon - E(\varepsilon)]^\top\right\} = E\left\{\varepsilon\varepsilon^\top\right\} = E\left(\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} [\varepsilon_1,\ \varepsilon_2,\ \cdots,\ \varepsilon_n]\right)$$

$$= \begin{bmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \cdots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & \cdots & E(\varepsilon_2\varepsilon_n) \\ \cdots & \cdots & \cdots & \cdots \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \cdots & E(\varepsilon_n^2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} \quad !$$

Cofactor Matrix *Q and* Weight Matrix *P*

- An error vector of n componnets: $\underset{n\cdot 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix}$
- Variance-covariance matrix *C* $\quad \underset{n\cdot n}{C_{\varepsilon\varepsilon}}$
- Co-factor matrix *Q* $\quad \underset{n\cdot n}{Q_{\varepsilon\varepsilon}} = \frac{1}{\sigma_0^2} \cdot \underset{n\cdot n}{C_{\varepsilon\varepsilon}}$

$$\underset{n\cdot n}{C_{\varepsilon\varepsilon}} = \sigma_0^2\, \underset{n\cdot n}{Q_{\varepsilon\varepsilon}}$$

- Weight matrix *P* $\quad P_{\varepsilon\varepsilon} = (Q_{\varepsilon\varepsilon})^{-1}$

Summary of Lecture 1, Basic Concepts

- Classification of errors
- Quality aspects: *precision, accuracy, reliability*
- Standard errors, statistical way of thinking
- Weights. Variance factor.
- Variance-covariance matrix
- Co-factor matrix and weight matrix