Paper [1]:

Lange, A. (1988): “A high-pass filter for optimum calibration of observing systems with applications.” In: Simulation and Optimization of Large Systems, edited by Andrzej J. Osiadacz, Oxford University Press/Clarendon Press, Oxford, pages 311-327. The paper is reprinted by a written permission of the publisher.

Simulation and optimization of large systems

Based on the proceedings of a conference
organized by the Institute of Mathematics and its Applications
on Simulation and Optimization of Large Systems, held at
the University of Reading, in September 1986

Edited by

ANDRZEJ J. OSIADACZ
University of Manchester
Institute of Science and Technology

CLARENDON PRESS · OXFORD · 1988

A HIGH-PASS FILTER FOR OPTIMUM CALIBRATION OF OBSERVING SYSTEMS WITH APPLICATIONS

A.A. Lange
(Finnish Meteorological Institute, Helsinki, Finland)

The solution reported here is based on the standard statistical theory of Best Linear Unbiased Estimates (BLUE). All observational data relevant to the calibration problem are listed and the underlying physical background of each measurement is represented by an appropriate algebraic equation. The resulting normal equation system is bordered block diagonal. The sparse structure of the matrix has been exploited through an analytical solution method. This solution of the calibration problem can be regarded as a "filter" to suppress the "noise" caused by the calibration errors which are constant or almost constant in time (i.e. low frequency). The computer solution together with the formulae have been reported by the author (IEEE PLANS, 1982) in connection with its application to the multipath propagation problem of the Omega VLF signals used in upper-air windfinding. Another application has been a windfinding comparison experiment with several types of tracking sensors, also reported by the author (WMO TECEMO, 1984). The filtering method is currently being used in the attempt to solve the problem of systematic errors in the thermodynamical weather reports from the Global Observing System (WMO CIMO/WG-UA/ Doc. 6, 1986). Here the global coverage of the space-based observing systems is expected to provide a sufficient degree of overdetermination for the solution, similar preliminary results being reported by McMillin and Gelman (WMO CIMO/WG-UA/Doc. 13, 1986).

1. INTRODUCTION

1.1 Scope of paper

This paper specifies a general method of adjusting the impact of calibration errors in observational data from an overdetermined observing system. Optimality of the filtering method will be discussed and three meteorological applications briefly described.

1.2 Historical background

Since World War II, upper-air measurements from the earth's surface up to stratospheric heights of 10-30 km have provided a scientific basis for weather forecasting. Errors in these measurements have posed a difficult problem because better observational reference data are remarkably costly to acquire in any meaningful amount. Several radiosonde and windfinding intercomparison experiments have been performed in past years always yielding results with respect to one observing system chosen more or less arbitrarily. Quite recently efforts have been made to remove subjectivity from the results of intercomparisons through the use of sophisticated statistical methods.

Satellite-based observing systems have, since the late 1970's greatly improved both the spatial and diurnal coverage of the Global Observing System (GOS) of the World Weather Watch (WWW). More redundancy has also been introduced into the GOS network to make it possible to compare different observing systems on an operational basis. The advent of supercomputers and extensive global data archival systems like that of the European Centre for Medium Range Weather Forecasts (ECMWF) now allow the use of sophisticated statistical methods. Thus, it becomes possible to identify those upper-air observing systems which suffer most from systematic errors and to diagnose their calibration problems for an immediate computational adjustment (i.e. the filter).

2. FORMULATION OF PROBLEM

2.1 Regression equation system

Consider an arbitrary set of measuring devices for example for windfinding: balloon-tracking radars, radio and optical theodolites, balloon-borne altimeters, Navaid signal retransmitters, remote-sensing profilers, etc.; or, for geopotential height measurements of a given atmospheric pressure level: radiosonde-based pressure, temperature, and humidity sensors, satellite- and ground-based vertical sounders, etc.

In the course of an intercomparison the values of observable variables obey the following statistical model:

$$y_{i,k} = A_{i,k}(\underline{b}_k) + S_{i,k}(\underline{b}_k, \underline{c}) + e_{i,k} \quad \text{for } i=1,\ldots,I_k \;\&\; k=1,\ldots,K \tag{2.1.1}$$

where each basic measurement is a synthesis of three components: atmospheric response (A), systematic error (S) and random error (e). More precisely, the notation is as follows:

- y : an observable variable (slant-range, azimuth, elevation, etc.)
- i : an index for each device participating at observation point k, i.e. $i=1,\ldots,I_k$
- j : an index of each atmospheric state variable pertinent to observation point k, i.e. $j=1,\ldots,J_k$
- k : an index for each observation point k, i.e. $k=1,\ldots,K$
- l : an index for each calibration error l, i.e. $l=1,\ldots,L$
- $\underline{b}$: a vector representing atmospheric state variables (unknowns) i.e. $\underline{b}_k=(b_{1,k},\ldots,b_{J_k,k})'$ for $k=1,\ldots,K$
- $\underline{c}$: a vector representing calibration errors (unknowns, assumed constant in time and space) i.e. $\underline{c}=(c_1,\ldots,c_L)'$
- A : a function defined for each index i giving the physical relationship between an observable variable and the atmospheric state variables of observation point k assuming no calibration error
- S : a function defined for each index i giving the physical relationship between the systematic error of an observable variable and the atmospheric state variables of observation point k and the calibration errors
- e : a random error of a basic measurement, assumed Gaussian about a zero mean and covariance matrix V with zero correlations between different observation points i.e. $E(\underline{e}) = \underline{O}$ and $V = \mathrm{Cov}(\underline{e}) = E(\underline{e}\,\underline{e}') = \mathrm{diag}(V_1,\ldots,V_K)$

  where $\underline{e}$ is the vector of all random errors $e_{i,k}$.

These equations (2.1.1) constitute a typical nonlinear system which must be highly overdetermined to have a solution which is stable enough statistically and which can also be reliably determined by numerical means. The quality of a solution, and indeed the existence of a meaningful solution depends on the information content of the data. Thus, the number of equations often becomes significantly larger than the total number of the unknown parameters. We may end up with a huge system with millions of equations and thousands of unknowns. The maximum-likelihood estimates are utilized here because of their suitable properties, see subsection 2.3. The optimization problem now becomes the minimization of the sum of squared residuals after they have been normalized with respect to the error variances and covariances.

2.2 Solution of the regression problem

The nonlinear least-squares solution can normally be established by using some iterative numerical method as outlined for example by Chambers (1977). It may often be necessary to use various Taylor, Fourier, spline or chapeau expansions for cost-effective approximation of the A and S functions in equation system (2.1.1). After the equations have been linearized through differentiation at an initial or any subsequent guess of $\underline{b}$ and $\underline{c}$ they may be written in the following vectorial form:

$$d\underline{y} = \text{Jacobian}(A+S,\underline{b})\ d\underline{b} + \text{Jacobian}(S,\underline{c})\ d\underline{c} + d\underline{e} \tag{2.2.1}$$

which takes the following sparse structure when using matrix notation:

$$\begin{bmatrix} d\underline{y}_1 \\ d\underline{y}_2 \\ d\underline{y}_3 \\ \cdot \\ \cdot \\ \cdot \\ d\underline{y}_K \end{bmatrix} = \left[\begin{array}{ccccc|c} X_1 & 0 & 0 & \ldots & 0 & G_1 \\ 0 & X_2 & 0 & \ldots & 0 & G_2 \\ 0 & 0 & X_3 & & 0 & G_3 \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & & & X_K & G_K \end{array}\right] \cdot \begin{bmatrix} d\underline{b}_1 \\ d\underline{b}_2 \\ d\underline{b}_3 \\ \cdot \\ \cdot \\ \cdot \\ d\underline{b}_K \\ d\underline{c} \end{bmatrix} + \begin{bmatrix} d\underline{e}_1 \\ d\underline{e}_2 \\ d\underline{e}_3 \\ \cdot \\ \cdot \\ \cdot \\ d\underline{e}_K \end{bmatrix} \tag{2.2.2}$$

where the vector elements are defined as follows:

- $d\underline{y}_k$: the vector of the differences between the basic measurements and the values of the A + S functions with a current guess of the atmospheric state and calibration parameters at observation point k, k=1,...,K i.e. $d\underline{y}_k = (dy_{1,k},\ldots,dy_{i,k},\ldots,dy_{I_k,k})'$
- $d\underline{b}_k$: the vector of the Gauss-Newton iterative adjustments of the atmospheric state parameters at observation point k, k=1,...,K i.e. $d\underline{b}_k = (db_{1,k},\ldots,db_{j,k},\ldots,db_{J_k,k})'$

$d\underline{c}$: the vector of the Gauss-Newton iterative adjustments of the calibration parameters i.e. $d\underline{c}=(dc_1,\ldots,dc_l,\ldots,dc_L)'$

$d\underline{e}_k$: the vector of the differences between the basic measurements and the values of the A + S functions with the updated guess of the atmospheric state and calibration parameters at observation point k, k=1,...,K i.e. $d\underline{e}_k = (de_{1,k},\ldots,de_{i,k},\ldots,de_{I_k,k})'$

and the matrix elements as follows:

X_k : the I_k by J_k Jacobian matrix of partial derivatives of A + S functions with respect to the atmospheric state parameters with a current guess of the atmospheric state and calibration parameters at observation point k, k=1,...,K

G_k : the I_k by L Jacobian matrix of partial derivatives of S functions with respect to the calibration parameters with a current guess of the atmospheric state and calibration parameters at observation point k, k=1,...,K.

The solution of this equation system (2.2.2) can be exceedingly demanding in computer resources if no attention is paid to the sparse structure of its coefficient matrix, see Golub and Plemmons (1980). Fortunately, the equation system can be partitioned into K subsystems, see Lange (1982). A subsystem is assigned to each observation point k, k=1,...,K. Using the terminology of the Stratified Sampling Theory we may regard the observation points as strata, the components of vector $\underline{b}_k$ as the separate regression parameters of stratum k (k=1,...,K) and the components of vector $\underline{c}$ as the combined regression parameters, see Cochran (1977) or for a more detailed discussion Lange (1973). The Gauss-Newton adjustments of the atmospheric state variables for an iterative step are then computed:

$$d\underline{b}_k = (X_k'\ V_k^{-1}\ X_k)^{-1}\ X_k'\ V_k^{-1}\ (d\underline{y}_k - G_k d\underline{c}) \tag{2.2.3}$$

for k=1,...,K.

The calibration error adjustments ($d\underline{c}$) are computed from:

$$\underline{dc} = \left(\sum_{k=1}^{K} G_k' R_k G_k \right)^{-1} \sum_{k=1}^{K} G_k' R_k \underline{dy}_k \qquad (2.2.4)$$

where R_k is the "residual operator" that extracts the normalized residuals out of the basic measurements and is computed by:

$$R_k = V_k^{-1} \left\{ I - X_k (X_k' V_k^{-1} X_k)^{-1} X_k' V_k^{-1} \right\} \qquad (2.2.5)$$

for k=1,...,K.

Some initial guesses of the regression parameters ($\underline{b}$ and $\underline{c}$) are needed. These guesses are then improved with the consecutive adjustments derived from formulae (2.2.3-4.) The process will normally converge to the required solution after a few iterative steps.

2.3 Optimality and stability

Under quite general conditions this method of solution leads to the maximum-likelihood estimation of both the atmospheric state and the calibration parameters, see Chambers (1977). Consequently, the resulting estimates are efficient, asymptotically at least. An estimator is called efficient if it is unbiased and has the smallest variance among the entire class of unbiased estimates, for example see Rao (1973). Meteorological applications deal with increasingly large samples so that the favourable asymptotic behaviour of derived estimates can certainly be effectively exploited. Thus, the filtering method leads to an optimum calibration because one cannot find essentially better estimates for both the atmospheric state and calibration parameters.

In practice, there may exist major uncertainties in the magnitudes of the random error variances and covariances. Rao's Minimum Norm Quadratic Unbiased Estimates (1972) and Horns' Almost Unbiased Estimates (1975) have both been used for calculating approximate magnitudes. Where the random error distribution is not Gaussian, the derived least-squares estimates are not maximum-likelihood estimates. The asymptotic efficiency of the least-squares estimates has in any case been derived by Cox and Hinkley (1968). The optimality of the method may also be degraded due to a significant time variation occurring in the calibration parameters. This problem is overcome by using a time-series analysis expansion based on coefficients that can be taken as constants. A longer learning period will then certainly be required in order to obtain statistically stable estimates for the increased number of calibration parameters.

The joint covariance matrix of the estimated regression parameters is readily obtained from the solution of the last iterative step to a good approximation, see Lange (1982). These variances and covariances indicate how accurately the atmospheric state and the calibration parameters are being estimated:

$$\mathrm{Cov}(\underline{b}_1,\ldots,\underline{b}_K,\underline{c}) \simeq \mathrm{Cov}(d\underline{b}_1,\ldots,d\underline{b}_K,d\underline{c})$$

$$= \left|\begin{array}{cccc|c} C_1+D_1SD_1' & D_1SD_2' & \cdots & D_1SD_K' & -D_1S \\ D_2SD_1' & C_2+D_2SD_2' & & D_2SD_K' & -D_2S \\ \vdots & & \ddots & & \\ D_KSD_1' & D_KSD_2' & & C_K+D_KSD_K' & -D_KS \\ \hline -SD_1' & -SD_2' & & -SD_K' & S \end{array}\right| \qquad (2.3.1)$$

where

$$C_k = (X_k'\, V_k^{-1}\, X_k)^{-1}$$

$$D_k = (X_k'\, V_k^{-1}\, X_k)^{-1}\, X_k'\, V_k^{-1}\, G_k$$

$$S = (\sum_{k=1}^{K} G_k'\, R_k G_k)^{-1}$$

and, G_k and R_k are as in equations (2.2.2) and (2.2.5), respectively.

Lack of overdetermination for all or some subsets of the observing systems causes the solution to become statistically unstable. Under such circumstances the least-squares problem is ill-posed and it is indicated by some unreasonably large variances and covariances (2.3.1). One could then try to remove suspect observing system subsets from the equation system (2.1.1) in order to determine whether the Gauss-Newton iteration also suffers from numerical instability.

For a realistic solution, other sources may provide additional information on the regression parameters. If the ancillary data are in the form of basic measurements then they should have been included in the equation system (2.1.1) in the first place. Direct measurements of a shifted calibration are to be accommodated by introducing an auxiliary stratum k, say k=K+1, with $J_{K+1}=0$.

The ancillary information may also be in the form of a physical relationship between the parameters to be estimated, see O'Leary and Bert (1986). The following three techniques are used:

(a) Regularization favours smooth solutions at the expense of fidelity to the basic measurements. The physical relationship must be a mathematical equality. It can be catered for by using ancillary X_k and G_k matrices with auxiliary strata added into the equation system (2.2.2). The regularization parameters are taken in as artificial residual variances.

(b) Projection techniques make use of a mathematical or statistical equality derived on the basis of the physical relationship. The number of regression parameters to be estimated is reduced, and this increases the stability of the solution. Cluster Analysis can be used for combining similar strata into homogeneous groups. The X_k and G_k matrices need to be adjusted accordingly.

(c) Side constraints are prescribed in order to eliminate undesirable solutions. These conditions are represented by mathematical inequalities. Stoer's method (1971) could be used for the partitioned solutions (2.2.3).

Thus, under quite general conditions the solution of an optimum calibration can be determined by using the filtering method. The current software package has been used in combination with techniques of the first two items (a) and (b) for inclusion of ancillary information.

3. NUMERICAL RESULTS

3.1 Multi-path problem with VLF Navaids

The International Omega Navigation System (ONS) was successfully used for upper-air windfinding during the Global Weather Experiment in 1978-79, see Lange (1986). The Vaisala RS21-12CN navaidsondes retransmitted the 13.6 kHz Omega signals from ascending weather balloons to data-acquisition and recording units aboard 33 Tropical Wind Observing Ships (TWOS). Serious discrepancies were observed between some computed winds as they were crucially dependent on the selection of the Omega stations. It became obvious that the short-path signal was not always dominant. Fortunately, there was a sufficient degree of overdetermination in the Navaid signals from the 8-station Omega network for simultaneous estimation of a few signal bearings and the entire wind profile of a sounding.

Numerical results from optimum calibration of the signal bearings are demonstrated here by a typical case from the eastern Pacific. The sounding was made on June 17, 1979 at 20.00 GMT on board U.S. ship Knorr. The latitude and longitude were 10.20° N and 113.49° W. Figure 1 shows the Omega signal data. Each station has its own vertical channel. The ship was maintaining constant speed and heading, cruising eastwards at 5 metres per second. Therefore the phase values were changing linearly in time. Liberia and Reunion were parallel to Japan and Hawaii. Had their signals taken the short paths, their phases should have been more or less parallel to Trinidad! The inclusion of Liberia and Reunion in the standard hyperbolic wind computation gave incorrect results as seen by comparing the wind profiles in figures 2 and 3. However, when the filtering method was used for the correction of the two arrival directions of the signals from Liberia and Reunion, the resulting profile in figure 4 exhibited remarkable similarity to the best available reference given in figure 2. Validity of the filtered solution in figure 4 was confirmed through an investigation into the internal consistency of the Navaid signals from all eight Omega stations. The Gauss-Newton iteration needed several steps due to high non-linearity in the two bearings.

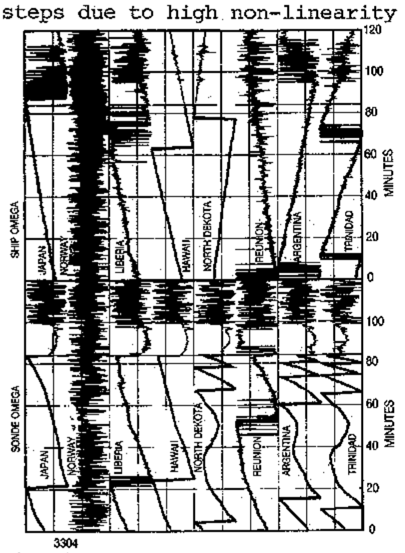

Fig. 1 Raw phase data (thin lines) and smoothed phases (thick lines) from the Omega stations (Japan, Norway, Liberia, Hawaii, North-Dakota, Reunion, Argentina and Trinidad). Ship data is above and sonde data below.

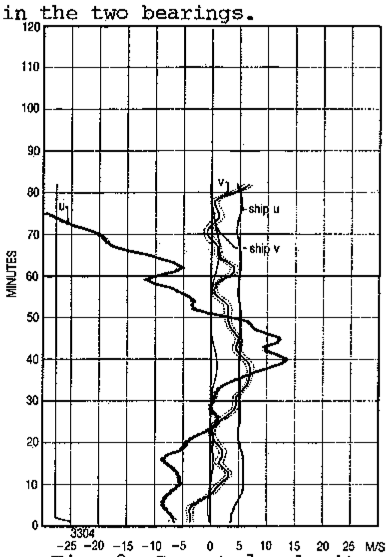

Fig. 2 Computed velocity profiles using all available signals except long-path signals of Liberia and Reunion.

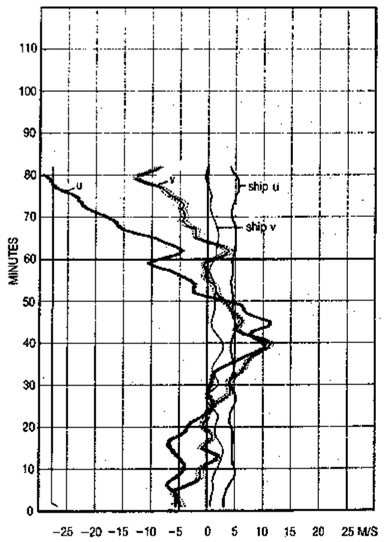

Fig. 3 Computed velocity profiles using all available signals without any bearing corrections.

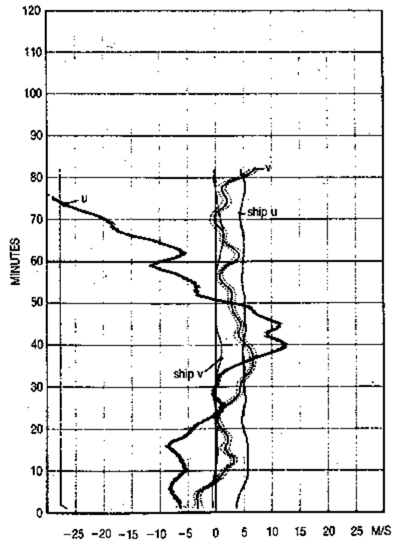

Fig. 4 Computed velocity profiles using all available signals and making the bearing corrections for the two long-path signals.

3.2 Wind-tracking with a hybrid system

A windfinding experiment involving several radars, theodolites, Omega Navaids and laser distance measurements was performed in Finland in 1972. This was to be one of the first intercomparisons without a subjectively chosen reference system, see Lange (1984). It was essential that the errors of the basic measurements were properly attributed to the two components (systematic and random) as represented by equation (2.1.1) and that the hybrid system of all the different tracking devices provided a sufficient degree of overdetermination. Calibration errors, for example inaccurate levelling of theodolites, were assumed constant throughout the whole sounding period, whereas the resulting systematic errors of elevation and azimuth measurements always become more or less time-dependent. The A and S functions and their linearization were also discussed by Lange (1984). Only a few steps were needed in the numerical iteration.

3.3 Global upper-air measurements

For many radiosonde stations the monthly bias profiles of geopotential heights tend to tilt either to the left or to the right whether it is a night-time or day-time monthly mean, see figure 5. Atmospheric tides also play a role and their amplitudes must be estimated as well.

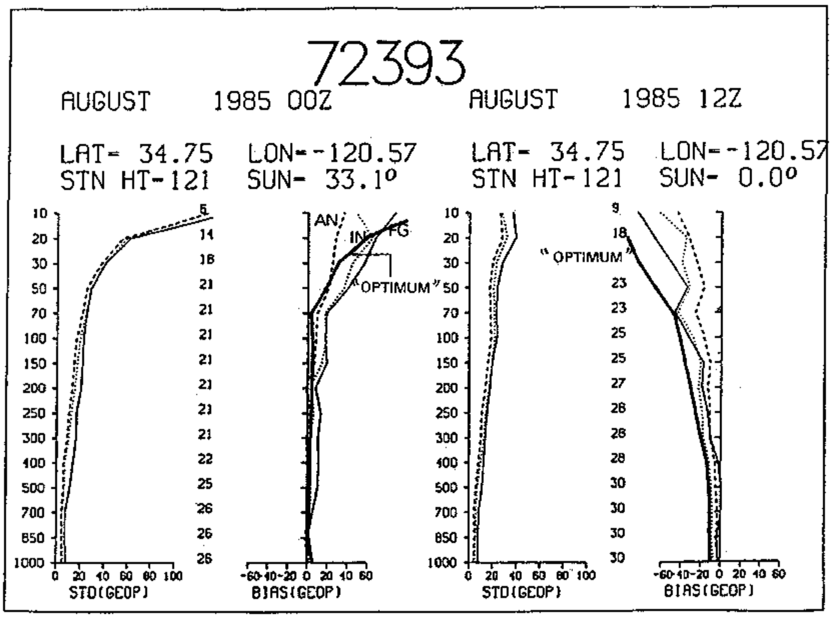

Fig. 5 Vertical monthly statistics of height departures from the first guess (FG, solid line), analysis (AN, dashed line) and initialized analysis (IN, dotted line) of the numerical weather prediction system used at ECMWF for North American station 72393 in August 1985. The random (STD) and systematic (BIAS) components of the departures at 00 GMT (left) and 12 GMT (right) were depicted. Units are shown in metres and number of reports received for each pressure level is given in the columns between the STD and BIAS curves. The mean of the solar elevation angles are in degrees.

A huge equation system emerges. Its linearized form is outlined in the Appendix. The normal equation system cannot be solved on any existing computer or supercomputer system without making use of the sparse structure of its matrix. The algorithm which was developed for the optimum calibration procedure can now be applied. A proper distinction between the separate and combined regression parameters must be made.

Numerical results from limited experimentation are available. The first two equation sets of the joint equation system given in the Appendix have been programmed thus far. An independent sample of four previous months was used for the prediction of radiosonde biases. Two of the predicted "optimum" calibration profiles have been superimposed in figure 5.

A complete-linkage Cluster Analysis of all radiosonde stations was carried out and 700 stations were grouped into 412 homogeneous clusters (NAG, 1982). The number of the separate regression parameters was reduced from 11,184 to 7,440. Three combined regression parameters were involved which related to tidal effects. The number of the equations and the estimated parameters was 2,200,000 and 7,443, respectively. The solution of the equation system (2.1.1) required 135 seconds and 11 Mbytes on a Cray XMP-48 system including the accumulation of the necessary sums and cross-products as well as the clustering. Double precision arithmetic was used for some matrix operations. All equations were sufficiently linear to require no iteration.

4. CONCLUSIONS

The optimum calibration procedure is finding increasing use in the handling of systematic errors of various overdetermined observing systems. Powerful number-crunching and archival/retrieval capabilities are required from the computer system, be it then a supercomputer or a microprocessor, depending on the application. They are undergoing rapid development and costs are coming down which makes the sophisticated computations increasingly cost-effective.

ACKNOWLEDGEMENTS

Part of the work has been carried out on a consultancy basis at the ECMWF.

REFERENCES

Chambers, J.M., (1977) "Computational Methods for Data Analysis". John Wiley & Sons, New York, Chichester, Brisbane and Toronto.

Cochran, W.G., (1977) "Sampling Techniques, third edition", John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore.

Cox, D.R. and Hinkley, D.V., (1968) "A Note on the Efficiency of Least Squares Estimates". *J. R. Stat. Soc.* B. Vol. 30, 284-289.

Golub, G.H. and Plemmons, R.J., (1980) "Large-Scale Geodetic Least-Squares Adjustment by Dissection and Orthogonal Decomposition". *Lin. Alg. Appl.* **34**, 3-27.

Horn, S.D., Horn, R.A. and Duncan, D.B., (1975) "Estimating Heteroscedastic Variances in Linear Models". *J. Am. Stat. Assoc.*, Vol. 70, No. 350, 380-385.

Lange, A.A., (1973) "Statistical Surface Wind Prediction in Finland". Technical Report No. 6, Finnish Meteorological Institute, Helsinki.

Lange, A.A., (1982) "Multipath Propagation of VLF Omega Signals". IEEE PLANS '82 - Position Location and Navigation Symposium Record, 302-309, IEEE AES Society, Atlantic City, NJ.

Lange, A.A., (1984) "Integration, Calibration and Intercomparison of Windfinding Devices". Papers Presented at the WMO Technical Conference on Instruments and Cost-Effective Meteorological Observations (TECEMO), WMO Instruments and Observing Methods Report No. 15, 41-45, Secretariat of the World Meteorological Organization, Geneva.

Lange, A.A., (1986) "Navaid Vertical Sounding System". International Conference on the Results of the Global Weather Experiment and Their Implications for the World Weather Watch, Volume I, GARP Publications Series No. 26, WMO/TD - No. 107, 157-175. Secretariat of the World Meteorological Organization, Geneva.

NAG FORTRAN Library Manual, Mark 9 (1982) Numerical Algorithms Group, Oxford.

O'Leary, D.P. and Rust, B.W. (1986) "Confidence Intervals for Inequality-Constrained Least Squares Problems, with Applications to Ill-Posed Problems". *SIAM J. Sci. Stat. Comput.*, Vol. 7, No. 2, 473-489.

Rao, C.R., (1972) "Estimation of Variance and Covariance Components in Linear Models". *J. Am. Stat. Assoc.*, Vol. 67, Number 337, 112-115.

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

Stoer, J., (1971) "On the Numerical Solution of Constrained Least-Squares Problems". *SIAM Journal of Numerical Analysis*, Vol. 8, 382-411.

APPENDIX

THE JOINT EQUATION SYSTEM FOR GLOBAL UPPER-AIR DATA-COMPATIBILITY

i) Equations for departures of radiosonde heights from reference (FG,AN,IN) heights:

$dy_{k,stn,ref,p,hr}$; radiosonde height - reference height.

$= b_{k,stn}$; station height error (barotropic bias at a station).

$+ (\underline{b}_{k,nocrs} - \underline{b}_{k,nocref})'\underline{f}_p$; nocturnal/infrared rsonde bias - noc ref bias.

$+ (\underline{b}_{k,dayrs} - \underline{b}_{k,dayref})'\underline{f}_p$ * solarradiation(lat,lon,dat,p,hr) ; solar bias.

$+ (c_{wv} - c_{wvref})$ * watervapourtidalforcing (lat,lon,dat,p,hr) ; tidal bias.

$+ (c_{oz} - c_{ozref})$ * ozonetidalforcing (lat,lon,dat,p,hr) ; tidal bias.

$+ (c_{nm} - c_{nmref})$ * nonmigratorytide (lat,lon,dat,p,hr) ; tidal bias.

$+ (c_{sd} - c_{sdref})$ * semidiurnaltide (lat,lon,dat, hr) ; tidal bias.

$+ e_{k,stn,ref,p,hr}$; random error.

ii) Equations for 00-12z (day-night) differences of radiosonde heights:

$dy_{k,stn,stn,p}$; 00z radiosonde height - 12z radiosonde height.

$= \underline{b}_{k,dayrs}'\underline{f}_p$ * solarradiationdifference (lat,lon,dat,p,00z,12z) ; solar bias.

$+ c_{wv}$ * watervapourtidalforcingdifference(lat,lon,dat,p,00z,12z) ; tidal bias.

$+ c_{oz}$ * ozonetidalforcingdifference (lat,lon,dat,p,00z,12z) ; tidal bias.

$+ c_{nm}$ * nonmigratorytidedifference (lat,lon,dat,p,00z,12z) ; tidal bias

$+ e_{k,stn,stn,p}$; random error.

iii) Equations for radiosonde intercomparison height differences (or heights):

$dy_{k,stn,com,p,hr}$; radiosonde height - comparison height (or 0).
$= \underline{c}_{com,hr}{}'\underline{f}_p$; comparison height bias (or true height).
$+ \underline{b}_{k,nocrs}{}'\underline{f}_p$; nocturnal/infrared radiosonde bias.
$+ \underline{b}_{k,dayrs}{}'\underline{f}_p$ * solarradiation(lat,lon,dat,p,hr); solar radiosonde bias.
$+ e_{k,stn,com,p,hr}$; random error.

iv) Equations for departures of radiosonde heights from precision radar heights:

$dy_{k,stn,pre,p,hr}$; radiosonde height - precision height.
$= \underline{b}_{k,nocrs}{}'\underline{f}_p$; nocturnal/infrared radiosonde bias.
$+ \underline{b}_{k,dayrs}{}'\underline{f}_p$ * solarradiation(lat,lon,dat,p,hr); solar radiosonde bias.
$+ e_{k,stn,pre,p,hr}$; random error.

v) Equations for acceleration vectors of tidal winds (00,06,12,18z TEMP/PILOT):

$\underline{dv}_{stn,p,hr}$; wind increment/time increment Coriolis force
$= c_{wv}$ * watervapourtidalwindforcingvector(lat,lon,dat,p,hr)
$+ c_{oz}$ * ozonetidalwindforcingvector (lat,lon,dat,p,hr)
$+ c_{nm}$ * nonmigratorytidalwindvector (lat,lon,dat,p,hr)
$+ c_{sd}$ * semidiurnaltidalwindvector (lat,lon,dat, hr)
$+ \underline{e}_{stn,p,hr}$; random error vector.

vi) Equations for differences between radiosonde and satellite radiances:

$dr_{k,stn,sch,hr}$; radiosonde "radiance" - satellite radiance.
$= \underline{b}_{k,nocrs}{}'\underline{d}_{sch}$; nocturnal/infrared radiosonde bias.
$+ \underline{b}_{k,dayrs}{}'\underline{d}_{sch}$ * solarradiation(lat,lon,dat, ,hr); solar radiosonde bias.
$- \underline{c}_{sch,vwf}{}'\underline{t}_{k,stn,hr}$; vertical weighting function biases of a satellite channel
$- \underline{c}_{sch,nlr}{}'\underline{l}_{radlevel}$; nonlinear response of a satellite channel.
$+ e_{k,stn,sch,hr}$; random error.

vii) <u>Equations for differences between reference (FG,AN,IN) and satellite radiances:</u>

$dr_{k,ref,sch,hr}$; reference "radiance" -satellite radiance.

$= \underline{b}_{k,nocref}{}'\underline{d}_{sch}$; nocturnal reference bias.

$+ \underline{b}_{k,dayref}{}'\underline{d}_{sch}$ * solarradiation(lat,lon,dat, ,hr) ; "solar" reference bias

$- \underline{c}_{sch,vwf}{}'\underline{t}_{k,ref,hr}$; vertical weighting function biases of a satellite channel.

$- \underline{c}_{sch,nlr}{}'\underline{l}_{radlevel}$; nonlinear response of a satellite channel.

$- (c_{wv} - c_{wvref})$ * watervapourtidalforcing(lat,lon,dat,sch,hr) ; tidal bias.

$- (c_{oz} - c_{ozref})$ * ozonetidalforcing (lat,lon,dat,sch,hr) ; tidal bias.

$- (c_{nm} - c_{nmref})$ * nonmigratorytide (lat,lon,dat,sch,hr) ; tidal bias.

$- (c_{sd} - c_{sdref})$ * semidiurnaltide (lat,lon,dat,sch,hr) ; tidal bias.

$+ e_{k,ref,sch,hr}$; random error.,

where the separate (b) and combined (c) regression parameters and parameter vectors ($\underline{b}$ and $\underline{c}$, respectively) to be estimated have been subscripted in a descriptive fashion (see "; comments,"). The other parameters are as follows:

k =1,...,K i.e. total number of radiosonde and satellite sounding clusters 500-1000
stn=1,...,total number of radiosonde stations and comparison sites 700-1000
ref=first guess (FG), analysis (AN) or initialized analysis (IN) height
com=derived data from radiosonde intercomparisons used as a comparison standard
pre=precision radar (Wallops Island, etc.)
p =1000,850,700,500,400,300,250,200,150,100,70,50,30,20,10 hPa
hr =00,06,12,18z, etc. as available (usually 00 and 12z only)
lat=latitude of a station or a radiosonde experiment site
lon=longitude (as above)
dat=Julian date (1-365)
sch=1,2,3,4,5,6,7,13,14,15,16,17(HIRS),2,3,4(MSU)1,2,3(SSU) for each satellite
radlevel=radiance level in a satellite radiometer channel (see above)

$\underline{d}_{sch}$= vector of heightbiases-to-radiance conversion for a satellite channel

$\underline{f}_p$ = vector of vertical basis function values for pressure level p

$\underline{l}_{radlevel}$ = vector of values of some basis functions for a radiance level

$\underline{t}_{k,stn,hr}$ = vector of Planck's emittances from the pressure levels for radiosonde data

$\underline{t}_{k,ref,hr}$ = vector of Planck's emittances from the pressure levels for reference data.

Some "a priori" estimates of the variances of the random errors are used as weights according to the linear estimation theory in order the minimize adverse effects of data from unstable sources. Rao's MINQUE (Minimum-Norm-Quadratic-Unbiased-Estimation) and related variance component estimation theories are to be used for improved "a posteriori" estimates of the random error variances. These results can be used in the quality monitoring of observational data e.g. radiances from each satellite channel etc. In addition, this results in an unbiased estimation of accuracies of different sounding systems and first guess (FG) analysis fields to be used for the determination of the optimal weights of Optimum Interpolation (OI).