

PERTURBING EFFECTS IN PAST OPTICAL ASTROMETRIC DATA. STATISTICAL MODELISATION AND SEPARATION

M. L. BOUGEARD⁽¹⁾⁽²⁾

(1) URA 1125 CNRS, I.E.R.S., Observatoire de Paris
61 Av. de l'Observatoire, F-75014 Paris, France

(2) Paris X University:I.U.T.
1 Chemin Desvallieres, F-92410 Ville d'Avray, France

ABSTRACT. *Past optical astrometric observations are to be re-reduced in order to get a new evaluation of the Earth Orientation Parameters in the more accurate Hipparcos reference frame. Among the selected instruments is the Paris astrolabe, considered here. In this Paper, through multivariate statistical procedures, we deal with the preliminary step of the new evaluation which consists in detection and separation of sources of significant inconsistency and outstanding errors in the Paris reductions so far performed. The analysis is performed over two test periods -one per instrumental setting- that are compared in particular in terms of observer effects, magnitude, colour and sidereal time effects. The Paper also gives a synthetic overview on the statistical methods we used to obtain the main results so that they can be applied to other astrometric data.*

1. INTRODUCTION

Past optical astrometric observations are to be re-considered in order to obtain a new 1900-1985 EOP (Earth Orientation Parameters) evaluation in a single and more accurate reference frame. For this purpose, an IAU Working Group on Earth Rotation in the (forthcoming) Hipparcos reference frame (in brief ERHRF-WG) was formed under the auspices of Commission 19 at the 20th IAU General Assembly (1988). A recent paper by J. Vondrák (1991), chairman of the ERHRF-WG, gives the general outlines of the prepared new EOP solution. It will be based on the

astrometric observations of *individual* stars. Vondrák and Ron (1992, this issue) give the present status of the work. The final list of selected instruments is to be found in (Vondrák *et al.*, 1992). Among them is the Paris astrolabe, considered here.

The Paris astrolabe data under analysis were obtained between 1956 and 1982. They concern about 180 000 star observations and 7352 "group-reductions". For the purpose of future revision of this long series, a careful analysis of the usual regression model that was applied in the reductions so far performed in a ground based celestial reference frame may prove useful to investigate potential perturbing effects. Here, through multivariate statistical procedures, we deal with the preliminary step, initiated in 1982, which consists in detection and separation of sources of significant inconsistency and outstanding errors in the reductions so far performed.

Section 2 is devoted to a brief review of the most striking sources of heterogeneity in the data under analysis. In Section 3, an extended linear regression model is introduced . Its statistical analysis is performed over two test periods, one per instrumental setting (1965-67: Danjon astrolabe / 1977-79: full pupil astrolabe). The respective results are studied and compared in particular in terms of observer effects, magnitude, colour and sidereal time effects. The last Section is devoted to mathematical validations of the extended linear model introduced before. They are obtained through a new analytic reconsideration of the *astrolabe equation of condition* and by use of S.M.A.R.T. projection pursuit regressions.

At each step, the Paper also gives a synthetic overview on the statistical methods we used to obtain the main results so that they can be applied or transposed to other astrometric data.

2. DETECTING A PRIORI SOURCES OF INCONSISTENCY

2.1 Introduction

Among the factors that are known to be acting on the precision of the optical astrometric reductions, people commonly invoke the effects of:

- (a) - a priori catalog errors
- (b) - the level of modelling of the physical system in use
- (c) - atmospheric and geophysical models (refraction, seasonal effects,...)
- (d) - instrumental effects and geometrical parameters of the instrument
- (e) - clock instability
- (f) - observer effects for the non-automatized instruments.

Here, we will be mainly concerned with the points (d) and (f) in application to the Paris astrolabe past observations of stars.

2.2 Main sources of inconsistency

The 1956-1982 data under analysis here were computed in the IAU System of Constants in use in 1982. We gratefully acknowledge the Paris astrolabe team that allowed the use of these data. The *a priori* main outstanding sources of inconsistency consist in:

(H1) - two instrumental periods: the 1956-70 *prism astrolabe* (refractive optical system) and the 1970-1982 *full pupil instrument* (reflective system),

(H2) - a variation of the observer team,

(H3) - a non-constant observer effect on the error variance.

This last point is brought into evidence by considering, per observer, the mean per year of the RMS errors computed from the classical group-reductions of his observations. On (Bougeard, 1987, Fig.1), it is clear that, for several observers, the variations of such a rough estimator are not only due to a sampling effect since for instance a large and significant decrease can be observed at the beginning of what was named *an apprenticeship observational curve*.

By pooling the individual observation equations in the forthcoming global EOP reduction, one will have to take into account the above sources of *heteroscedasticity* (significant inequality of the error variances) through at least, say, an *observer-adaptive Weighted Least Squares* fit.

3. DETECTION AND SEPARATION OF SYSTEMATIC EFFECTS: TOWARDS A PROPOSED STRATEGY

3.1 Choice of two test-periods

In order to adapt a statistical strategy for the analysis of potential perturbing effects in the reductions so far performed, one test-period per instrumental setting has to be chosen. Furthermore, these periods have to be:

- long enough to permit the use of statistical tests of significance,
- if possible, distant of 6 or 12 years to preserve the estimations from annual and Chandler periodic terms so that they can be compared,
- stable enough in terms of observer effects on the error variance.

In the Paris-astrolabe records, the following periods proved to satisfy these constraints (observers with too few observations were deleted):

- 1965-67 (prism astrolabe): 14 observers, about 21000 observations/900 group-reductions,

- 1977-79 (full pupil one): 9 observers, about 17000 observations/700 group-reductions.

3.2 Taking into account possible perturbing effects: method

Astrolabe observations are based on the equal altitude method. The observation equation of condition is usually expressed as a linear model in terms of two predictors ($\sin Z$, $\cos Z$, where Z is the star azimuth). The coefficients (parameters) to be estimated are the differential corrections in time, latitude and instrumental zenith distance. The stars (about 28) of a given group being successively observed during the night, the related system of equations is usually solved by a least squares fit.

To take into account some possible perturbing effects, in terms of colour and magnitude, instrumental transmission or instability of the instrumental zenith distance, an extended linear model with 5 predictors is considered. This model includes the 3 following *additional predictors*: the visual magnitude of the star, a B-V colour index, its sidereal time of transit.

In a first step, this group of predictors is confirmed to be potentially appropriate through exploratory multivariate *Principal Component Analyse*. In a second step, statistical tests of significance of each predictor (F-test at a chosen level) are applied by means of *stepwise regressions*.

3.3 Outliers

Since such tests are very sensitive to departure from a Gaussian law of the error distribution, a systematic identification and elimination of the outliers is needed. For the data under study, this problem was tackled by using simultaneously:

- the Belsley-Kuh-Welsch influence statistics (based on iterating the least squares fit after having dropped a lot of observations),
- an L_1 fit that is known to be more robust than the least squares process.

It had led to the detection of nine *atypical stars* in the Astrolabe program (stars very influential in the estimations and with often systematically ill-behaved residuals). They were proved to have contaminated the *internal corrections* in use.

As soon as the Hipparcos catalog will be available, this effect can be expected to disappear. Nevertheless, some difficulties are likely to be still introduced in the forthcoming global EOP evaluation by those of these atypical stars that were found to be *double* or *suspected variable*.

3.4 Results and interpretation

In order to obtain an overview on the results, separate chi-square statistics are not applied. To identify the factors that are responsible of proximities, we resort to the use of a global *multiple correspondence analysis* for each test period in which the

individuals are given by the different group-regressions, the *variables* are the indicators of the variables that were found significant in the stepwise approach. Indicators of groups, observers, mean hour of observation are also considered. This method has enabled us to prove that the perturbing effects in time, colour, magnitude act differently according to the instrumental set-ups (for details, see: Bougeard, 1992b). The Paris Danjon prism astrolabe appears very sensitive mainly to time and magnitude effects. As far as the full-pupil astrolabe is concerned, attention also has to be paid to the time effect (variation of the instrumental zenith distance) that was verified to increase in summer and to show significant dependence on the observer for the period 1977-79.

A priori calibrations of the instruments are needed in terms of these effects. This point will be questionable in the treatment of past observations. How a posteriori taking into account such perturbing effects? The above extended linear model may be an answer.

4. EXTERNAL VALIDATIONS OF AN EXTENDED MODEL, CONCLUSIONS

This Section gives complementary validations of the above linear model.

4.1 Effects of catalog errors on the temporal variation

In (Bougeard, 1991), a more complete analytical decomposition of the *astrolabe equation of condition* is obtained by expressing the errors on the adopted equatorial coordinates of the star at the time of the observation in terms of *catalog errors* and of *errors on the astronomical constants* in use. On the resulting *new equation of condition*, it appears that the observational residual has to be expressed linearly not only in terms of the two classical predictors but also in terms of *six* other ones that are the sine and the cosine of the *sidereal time of transit* and of their *respective products* with the sine and cosine of the azimuth.

These six predictors that are missing in the classical astrolabe model may have caused systematic seasonal effects in the estimations so far performed. In addition, a part of the *time variation effect* detected as significant in the above statistical approach can be partially ascribed to these catalog errors.

The last point is to estimate how approximated can be the use of a linear model and of the usual Least Squares resolution that relies on the error Gaussianity assumption. For this purpose, we resort to the use of *projection pursuit regressions* (PPR) based on Friedman (1984,87)'s SMART (Smooth Additive Regression Technique) approach.

4.2 Testing the linearity of perturbing effects

In PPR modeling, to describe the relation between the response variable Y (observational residual) and a set of predictors $X = (X_i) \ i = 1 - p$ according to a loss criterion, we model Y as a weighted linear combination of M predictor functions f_m that are respectively taken as a smooth function of a linear combination $(\alpha_m \cdot X)$.

For the data under study, it was proved (Bougeard 1990,1992a) that the main term in the PPR modeling is *approximately linear*. It closely coincides with the LS (least squares) solution of the above extended 5 predictor scheme only for data reduced in the FK4 reference system. When one uses the FK5 reference system, it shows departure from the LS estimation and confirms the importance of the sidereal time effect for the full pupil instrumental period. Such a fact is not quite surprising according to the analytic approach, since we proved that this effect can be partially ascribed to catalog errors. Such points will have to be tested again as soon as the Hipparcos star coordinates will be available.

5. REFERENCES

- Bougeard M.L.,1987, *Astron. and Astroph.*,n°173, p191-203 and n°187, p156-166
 Bougeard M.L.,1988: in *Compstat: short communications*, eds. Edwards and Raun, Danish Comput. Center, Copenhagen, p5-6
 Bougeard M.L.,1990: in *Journees Systemes de reference*, Paris Obs., p69-73
 Bougeard M.L.,1991: *Etude analytique de l'effet temps sidereal aux astrolabes*, Techn. report, Paris X University: Cahiers Math., vol 18, march
 Bougeard M.L.,1992a: Doctor thesis es Mathematical Sciences, Paris 6 University
 Bougeard M.L.,1992b, *Astron. and Astroph.*,n°255, p388
 Friedman J.H.,1984, *SMART's User Guide* (Techn. report, Stanford)
 Friedman J.H.,1987,*JASA*, vol 82, p249-266
 Vondrak J.,1991, *Bull. Astron. Inst. Czechosl.*, 42 n°5, 283-294
 Vondrak J., Feissel M., Essaifi N.,1992, *Astron. and Astroph.*,n°262,329
 Vondrak J., Ron C., 1992, this issue