HIPPARCOS MINOR PLANETS:

TOWARDS AN IMPROVEMENT OF THE MODEL ANALYSIS BY DETECTING INFLUENCE FACTORS.

M. L. BOUGEARD
Observatoire de Paris, DANOF/IERS, 75014 Paris and Université Lyon 1, Dep. Math., 69622 Villeurbanne, France
J.-F. BANGE
Bureau des Longitudes, URA CNRS 707
77 Avenue Denfert-Rochereau, Paris, France
M. MAHFOUZ
Université Lyon 1, Dep. Math. Statistique
AND
A. BEC-BORSENBERGER
Bureau des Longitudes, URA CNRS 707

Abstract. In order to evaluate a possible rotation between the Hipparcos and the dynamical reference frames, Hipparcos minor planets preliminary data are analysed. The resolution of the problem is very sensitive to correlations induced by the short length of the interval of observation. Several statistical methods are performed to appreciate the factors of bad conditioning. A procedure for variable selection and model building is given.

1. Hipparcos Minor Planets (HMP) model

For each of the 48 minor planets observed by the astrometric satellite Hipparcos, we denote by 'O' the observed abscissae on a Reference Great Circle (RGC) related to a Hipparcos satellite revolution (see Hestroffer et al, 1995, for the reduction procedure). Subtracting the corresponding position 'C' calculated through numerical integration of the motion equations leads to the following equation of condition, written with sufficient approximation as:

$$\mathbf{O} - \mathbf{C} = \mathbf{A}\Delta\overrightarrow{u^0} + \mathbf{B}\Delta\overrightarrow{u^0_T} + \mathbf{D}\,\overrightarrow{\theta} \qquad \text{(HMP)}$$

where **A** is a matrix depending on the minor planet's coordinates partial derivatives with respect to its initial elements; **B**, a matrix depending on the

S. Ferraz-Mello et al. (eds.), Dynamics, Ephemerides and Astrometry of the Solar System, 447–450.

earth's coordinates partial derivatives with respect to its initial elements and **D** a matrix linking the Hipparcos reference frame and the dynamical one. The unknowns are the 6-component vectors Δu^0 (corrections to the minor planet orbital elements), Δu_T^0 (corrections to the Earth elements) and θ (link between the dynamical and Hipparcos reference frame). The initial epoch is t_0=JD 2447800.5 .

2. Statistical exploration of the HMP-Model

Here we focus our attention to the estimation of the 6×1 vector $\boldsymbol{\theta}$ by using preliminary FAST data covering 37 months of the mission. The statistical approach is illustrated on a subset of 3 minor planets: 6-Hebe (91 observations on 20 RGC), 9-Metis (43 observations on 14 RGC) and 42-Isis (52 observations on 15 RGC). Performing a Least Squares regression by pooling the 186 equations of condition (parameters Δu_T^0 and θ are common) leads to a very large instability of the results that are not even meaningful! As a measure of possible multicollinearity phenomena, the **VIF** (Variance Inflation Factor) has been performed for each predictor in two cases: model 1 with full design matrix (A, B, D) and model 2 without the B-factors. We report some values for model 2 (Table 1).
Although both models fit the data quite well (coefficient of determination R2=0.97 and R2 adjusted=0.96) the very large **VIF** values assert that many factors are involved in the multicollinearity (Tassi, 1992). This concerns A-factors as well as D-factors: the overall model 1 including B-factors being worst (**VIF** values are larger). At this stage, the LS regression through any numerical approach (singular values or QR decomposition) is statistically misleading especially in terms of tests and error variance estimates (Bougeard and Michelot, 1991).

TABLE 1. Variance inflation factors (model 2)

	D_1	D_2	D_3	D_4	D_5	D_6	A_1	A_2	A_3	A_4	A_5	A_6
VIF	18	152	401	16	62	508	2997	6	23	203	5451	457

3. Procedures for multicollinearity detection

Consider the 42-Isis HMP data. Due to the short interval-length of observation several factors are measuring nearly similar physical phenomena. Although the correlation matrix is of full rank, three eigenvalues $\lambda_6 \leq \lambda_5 \leq \lambda_4$ appear as negligeable (Table 2).
Let T be the 6×3 matrix with columns V_6, V_5, V_4 the related eigenvectors. To detect the factors that may produce multicollinearity, the norm of each 3×3 sub-matrix inverse of T is computed. Those with minimal norm give indication on the influential factors (Table 3).

In conclusion, for obtaining a model meaningful from a statistical viewpoint, the variables/predictors to be dropped are: first A_1 and A_5, second A_4 or A_2 or A_3.

TABLE 2. 42-Isis data correlation matrix

	Correlation matrix						Eigenvalues	Cumulative
	A_1	A_2	A_3	A_4	A_5	A_6		proportion
A_1	1	0.04	-0.03	-0.98	-0.99	0.84	λ_1=3.135	0.619
A_2		1	-0.99	-0.12	0.03	-0.17	λ_2=2.0415	0.959
A_3			1	0.11	-0.04	0.18	λ_3=0.2316	0.998
A_4				1	0.95	-0.74	λ_4=0.0102	0.999
A_5					1	-0.91	λ_5=0.0030	1.0
A_6						1	λ_6=0.0001	1.0

4. Statistical building of the HMP model and Detection of influential observations

The above variable selection procedure was extended to 6-Hebe and 9-Metis predictors $A_1...A_6$ but also on predictors $B_1...B_6$ associated to the correction of the Earth orbit and $D_1...D_6$ associated to the link between reference frames (recall that large **VIF** values were associated to all these variables). Results obtained after a least squares regression on the selected model are given in Table 4. VIF and condition numbers are then acceptable. By reducing the effects of multicollinearity, estimates of θ with smaller mean squared error were produced. At this time, statistical tests to detect 'atypical' observations can be used. Here, we resort to the use of the Belsey-Kuh Welsch influence diagnostic statistics (already applied in Bougeard, 1987). Influential observations are given in table 5.

TABLE 3. Sub-matrices of rank 3

norm	2.301	2.305	2.498	2.500	3.004	3.005	3.357	3.361	1006.71
lines of T	1,2,5	1,3,5	2,4,5	3,4,5	1,2,4	1,3,4	1,2,6	1,3,6	2,3,6

5. Conclusion

In this paper, a statistical methodology for HMP model building was tested on a subset of three minor planets. Detection of multicollinearity was performed, factors of influence analysed on 42-Isis data (see also Bec-Borsenberger et al, 1994), a model built that provides estimate with smaller mean squared error. Then statistical influence tests have allowed the detection of influential observations.

TABLE 4. Results and collinearity diagnostics (arcsecond or milliarcsecond.JJ^{-1}: D_5, D_6)

predictor	parameter estimate	standard error	variance inflation	n^o	eigenvalue	condition index
D_2	-0.513	0.006	17.7	1	4.00	1.0
D_3	-0.005	0.012	34.1	2	2.37	1.3
D_5	-0.04	0.01	12.1	3	2.26	1.3
D_6	-0.08	0.02	29.1	4	1.84	1.5
...	...	...	...	...	...	...
$42 - A_2$	0.077	0.005	4.8	15	0.07	7.7
$42 - A_3$	-0.019	0.006	7.8	16	0.03	10.9
$42 - A_6$	-0.007	0.005	4.0	17	0.01	19.4

TABLE 5. Belsley-Kuh Welsch influence statistics (*: observation potentially suspicious)

atypical obs. ref.	minor planet	RGC	numb. of obs on this RGC	Rstudent	Dfitts	HATDIAG	comment
34	6	450	8	*			?
61	"	1404	3	*	*	*	suspicious
78	"	2003	2	*	*	*	suspicious
91	"	2641	2	*	*	*	suspicious
94	9	771	4	*	*		suspicious
102	"	947	4	*	*		suspicious
120	"	1414	1	*	*		suspicious
130	"	1533	1	*	*		suspicious
135	42	258	2	*	*		suspicious
140	"	411	4	*	*		suspicious
148	"	549	6	*			?

References

Bange, J-F.: 1993, *Mémoire de DEA*, Observatoire de Paris.
Bec-Borsenberger, A., Bange, J.-F., Bougeard, M. L.: 1995, *Astron. Astrophys.* **304** 176.
Bougeard, M. L.: 1987, *Astron. Astrophys.* **173** 191.
Bougeard, M. L., Michelot, C., 1991: *Journées Systèmes de référence Spatio-temporels*, Observatoire de Paris, p. 86.
Hestroffer, D. *et al* : 1995, *Astron. Astrophys.* **304** 168.
Tassi, P.: 1992, Méthodes Statistiques, Economie.