6. UPDATING AND IDENTIFICATION OF TRANSIT DATA

The present chapter describes the prediction updating processes (Sections 6.1 and 6.2), and how their output was used to improve the identification of observed peaks in the photon counts with the transits of specific stars (Section 6.3). The role of these processes, and of transit identification, is to replace some or all of the imprecise inputs used in first prediction and transit detection (namely the Tycho Input Catalogue, the on-board satellite attitude and preliminary instrument geometry calibrations) by better data.

6.1. Prediction Updating-2

The second prediction step in the iterative Tycho data reduction scheme was called prediction updating-2. Its main role was to replace the imprecise on-board satellite attitude by an improved on-ground attitude for the purposes of the star recognition process (i.e. the Tycho Input Catalogue Revision) described in Chapter 5. The central problem of that process was to distinguish true stars from chance crossings of detection lines in its digital line maps. The ability to solve this problem crucially depended on the scatter of the transit lines about the true position of a star—and this scatter could not possibly be smaller than the uncertainty of the satellite attitude. Without prediction updating, the recognition process could not have used line map pixels as small as 0.5 arcsec. Instead it would have had to use pixels of about 3 arcsec. This would have created a 36 times higher rate of background detection line crossings per pixel, which in turn would have buried the true line crossings of all the fainter Tycho stars.

The basic task of prediction updating was much the same as that of the first prediction process, which was to provide a time-ordered list of expected group crossings of celestial objects, along with some auxiliary data. Despite this basic similarity, the actual process was completely different, and very much quicker. The first prediction process was controlled by the availability of on-board attitude and certain satellite status indicators. The most complex and time-consuming parts of the software were the selection of candidate stars for group crossings and the computation of satellitocentric apparent positions for these. Both could be avoided in prediction updating. This was made possible by a careful design of the Predicted Group Crossing (PGC) data stream.

Basic Algorithm

The prediction updating process was controlled by the Predicted Group Crossing (PGC) data stream, i.e. the output of the first prediction process. The original PGC tapes were sent from Heidelberg to Tübingen to be used in the detection process. Copies of them were also kept at Heidelberg to be used in prediction updating. They were processed in strict time sequence, with one predicted group crossing being treated at a time. The basic algorithm consisted of the following actions:

- read one predicted group crossing from the PGC data stream;
- check whether on-ground attitude is available for the appropriate instant of time;
- if so, use the apparent position of the object under consideration (which is included in the input PGC record), and the on-ground attitude to directly compute a new transit time. This is done exactly as in prediction, by linear interpolation between neighbouring pivotal points of the attitude (see Section 4.1);
- recompute some of the auxiliary data and flags (such as the instantaneous scan speed, the field coordinates at transit time, jet firings of the attitude control system etc.);
- compute the mean error of the updated transit time from the mean errors of the on-ground attitude alone (i.e. disregarding the uncertainty of the star position);
- create the output record for the PGC Updating-2 data stream and write it to tape.

No real updating could be done if no on-ground attitude was available for a particular predicted group crossing. But still a record was written to the PGC Updating-2 data stream in that case. It was flagged as non-updated, and furthermore marked by a large dummy value (corresponding to 1 arcsec on the sky) of the transit time mean error.

Note that this basic algorithm implied a number of approximations, all of which however, had negligible effects on the quality of its output. First, the apparent positions were not recomputed. This was possible because the original and updated predicted transit times differed by at most a few times 0.01 s, and because the apparent positions change only slowly (the biggest effect being the change of aberration due to the satellite's geocentric acceleration, at most 1 mas s^{-1}). Second, the search for transiting stars was not repeated. Due to the change in attitude it might have happened that a star had just missed the ends of the star mapper slits in the original prediction process, but now caused a transit (or vice versa). This possibility was avoided by using a star mapper model with slits that were somewhat longer than the real ones, both in first prediction and in prediction updating. Thus, no transits could be lost. Spurious transits produced in this way were easily recognized by their z coordinate being outside the actual slit boundaries after updating. Third, the originally predicted transit time might have fallen into the time interval between two pivotal points of the on-ground attitude, but slightly outside that interval after updating. In such cases the updating was performed as usual, which implied a very small extrapolation of the attitude (instead of the usual interpolation).

Input Attitude for Prediction Updating-2

The Tycho consortium was ready to use on-ground attitude from either NDAC or FAST, depending on availability. In practice, on-ground attitude from NDAC was available sooner (the first few months of attitude with sufficient precision of 0.2 arcsec were delivered by NDAC in December 1991). Thus, this was used for updating-2.

As explained in Section 4.1, it is important that an instrument geometry model consistent with the attitude determination is used in any computation of Tycho group crossings. This was automatically ensured by the presence of a special NDAC star mapper geometry file on each of the attitude tapes.

Output Data

The output of updating-2 consisted mainly of the PGC Updating-2 data stream, indicated in Figure 1.2, plus some protocol and log files. The output tapes were sent to Strasbourg, where they were used for the production of the Tycho Input Catalogue Revision, as described in Chapter 5.

Actual Processing

A series of test runs for prediction updating-2 was performed during 1991, using provisional on-ground attitude from both NDAC and FAST. This resulted in the discovery and removal of software errors, and also provided the opportunity to monitor the gradual progress in the quality of the on-ground attitude. The final production runs for prediction updating-2 were started on 21 February 1992, and completed 7 weeks later.

6.2. Prediction Updating-3

The role of prediction updating-3 was to replace both the originally used attitude and the star catalogue used for the first prediction. There were two reasons for using the Tycho Input Catalogue Revision instead of the Tycho Input Catalogue for a prediction updating. First, the number of stars was greatly reduced, from 3.15 million to 1.26 million. Second, the mean errors of the star positions were greatly reduced. Both of these differences resulted in a better agreement between predicted and actually observed transit times of stars, which was needed to improve the elimination of unrelated background transits and photon noise peaks in the final astrometry and photometry tasks.

Direct-Access Tycho Input Catalogue Revision Installation

While prediction updating-2 needed no star catalogue access at all, updating-3 needed an efficient access to the Tycho Input Catalogue Revision. The organisation and indexing of the Tycho Input Catalogue into a direct-access magnetic-disk file according to small regions on the sky was described in Chapter 4. That organisation was optimized for the needs of first prediction processing. To make it usable for prediction updating-3, the sequential record numbers of the Tycho Input Catalogue objects in that direct-access file were also included in the Predicted Group Crossing records. The Tycho Input Catalogue Revision itself was brought into the same sequence (with dummy records filling the 2 million gaps produced by the ‘unrecognized’ Tycho Input Catalogue stars). The resultant direct-access Tycho Input Catalogue Revision records did not include the improved positions for the Tycho Input Catalogue Revision objects, but the differences between the Tycho Input Catalogue Revision and Tycho Input Catalogue positions, represented by three-dimensional cartesian difference vectors. The dummy records were marked by a flag.

This organisation allowed the PGC-controlled retrieval of the Tycho Input Catalogue Revision data with a single magnetic-disk access, but only for those Tycho Input Catalogue entries having zero or one corresponding entry (‘companion’) in the Tycho Input Catalogue Revision. Those Tycho Input Catalogue entries yielding more than one companion in the star recognition process (Chapter 5) were managed in the following way. In the Tycho Input Catalogue Revision, the different companions to a particular Tycho Input Catalogue entry were distinguished by a running ‘companion number’. The ordered direct-access Tycho Input Catalogue Revision record described in the preceding paragraph was occupied by the companion with the highest companion number. A pointer at the end of the record contained the physical record number where the companion with the next-lower number could be found, and so on, until the companion with the companion number 1 was reached. The extra records for the additional companions were simply added at the end of the direct-access Tycho Input Catalogue Revision file.

In this way the PGC-controlled retrieval of the Tycho Input Catalogue Revision data with a single magnetic-disk access per object was achieved even in the case of multiple companions, as explained below.

Basic Algorithm

The basic algorithm of prediction updating-3 was similar to that of prediction updating-2, with a few additional actions caused by the change of the star list from the Tycho Input Catalogue to the Tycho Input Catalogue Revision. Again the selection of candidate stars for group crossings and the computation of satellitocentric apparent positions were avoided by the design of the input PGC data stream and the above-described special installation of the Tycho Input Catalogue Revision. The individual steps were as follows:

- read one predicted group crossing from the PGC data stream;
- check whether on-ground attitude is available for the appropriate instant of time;
- if so, use the direct-access record number (from the PGC record) to read Tycho Input Catalogue Revision data. Check whether a dummy Tycho Input Catalogue Revision record was found;
- if not, use the apparent position of the Tycho Input Catalogue object under consideration (from the PGC record), add the position difference (from the Tycho Input Catalogue Revision record) and directly compute a new transit time from the on-ground attitude, as in prediction updating-2;
- recompute some of the auxiliary data and flags (such as the instantaneous scan speed, the field coordinates at transit time, jet firings of the attitude control system etc.);
- compute the mean error of the updated transit time from the mean errors of the on-ground attitude alone (i.e. disregarding the uncertainty of the star position);
- create the output record for the PGC Updating-3 data stream, including flags and magnitudes from the Tycho Input Catalogue Revision, and write it to tape;

- if the component number under consideration is larger than 1, get the Tycho Input Catalogue Revision data for the next-lower component number, and repeat the preceding four steps.

This algorithm implies the same approximations as in prediction updating-2. In addition, there is an approximation by using the apparent position for the Tycho Input Catalogue object to get that for the Tycho Input Catalogue Revision companion(s). This is well justified, since the difference vector between a mean and an apparent position is a very slowly varying function of the location on the celestial sphere. The biggest effect is produced by the annual aberration. Its maximum possible derivative is about 0.1 mas per arcsec. So, for position updates (Tycho Input Catalogue Revision minus Tycho Input Catalogue) of a few arcsec, it is completely negligible in Tycho data reductions (larger offsets were taken care of in the Tycho reprocessing, Chapter 10).

A more important effect of this approximation is that implicitly the same proper motions and parallaxes were assumed for all companions of a Tycho Input Catalogue entry. This does not cause any problems, since the Tycho astrometric data are derived independently for all companions in the astrometry processing (Chapter 7).

Input Attitude for Prediction Updating-3

An NDAC-provided on-ground attitude was used throughout the mission. The particular set of 11 attitude tapes used for the final production runs of prediction updating-3 were all based on the 18-month NDAC Hipparcos sphere solution. That is, the attitude was the orientation of the satellite with respect to the celestial coordinate system defined by that preliminary Hipparcos catalogue. The typical rms errors of the attitude angles were 35 mas perpendicular to the scan direction, and a few mas along the scan direction.

Output Data

The output of updating-3 consisted mainly of the PGC Updating-3 data stream, indicated in Figure 1.2, plus some protocol and log files. While PGC Updating-2 contained an output record for every input PGC record, PGC Updating-3 was smaller by about a factor of 3, because there were no output records for the 2 million Tycho Input Catalogue entries missing in the Tycho Input Catalogue Revision. The PGC Updating-3 tapes were sent to Tübingen, where they were used for the identification of transits, as described in Section 6.3.

Actual Processing

A series of test runs for prediction updating-3 was performed during 1992, using provisional on-ground attitude from both NDAC and FAST, partly with a provisional Tycho Input Catalogue Revision and partly with the operational one.

The operational attitude (described above) was delivered to Heidelberg in small portions between May 1993 and August 1994. The final production for prediction updating-3 started in July 1993, and most of it was completed in February 1994. Updating-3 was an even quicker process than updating-2, due to the much smaller number of objects to be treated. It quickly caught up with the still running prediction process, then proceeded closely following the prediction process.

Prediction Updating-3 Redoing

The ‘prediction redoing’ process (see last paragraph of Section 4.1) produced a PGC data stream based on the final on-ground attitude, for the parts of the mission with bad real-time attitude. But it still used the Tycho Input Catalogue as input star catalogue. Therefore, a prediction updating step for redoing was still necessary in order to replace:

(a) the imprecise Tycho Input Catalogue positions by the more precise Tycho Input Catalogue Revision positions; and

(b) the object list and object numbering system of the Tycho Input Catalogue by that of the Tycho Input Catalogue Revision.

Thus, prediction updating-3 was run on the redoing PGC data stream (with slightly modified software to accomodate the changes in the data interfaces, see Section 4.1). The resulting special PGC Updating-3 tapes were also sent to Tübingen, to be used in the redoing transit identification.

6.3. Identification of Transits

The role of transit identification was to connect the prediction updating-3 data stream (Section 6.2) with the raw transit data stream produced in the detection process (Section 4.3). It had to perform two major tasks: assigning actual detections to the predicted group crossings given by updating-3 (and thus to stars on the sky), and to record potential disturbances caused by nearby other transits.

Assignment of Transits

The assignment of detections to the predictions given by updating-3 was carried out using the raw transit data stream (RT in Figure 1.2). No second analysis of the photon count data was done. The output of the process consisted of the identified transits data stream (IT in Figure 1.2). Each identified transit record contained most of the combined information of the corresponding updating-3 and raw transit data. Therefore it was 25 per cent bigger than a raw transit record. The complete amount of data however produced by the identification process was only about 30 per cent of the raw transit data stream. Due to the reduced error sources of updating-3 compared to the first prediction, the width of the time interval in which detections could be assigned to a certain prediction was only half of that used in the detection process. Furthermore, all raw transits which had been assigned to Tycho Input Catalogue objects not present in the Tycho Input Catalogue Revision were discarded. In that way a much cleaner input data stream for the following astrometric and photometric processing was produced.

In order to be assigned to a specific prediction, a detection from the raw transit data stream had to fulfill two conditions: (a) it had to lie inside the assignment interval; (b) the updated prediction (PGC Updating-3) had to be the update of the original (first) prediction to which the detection had been assigned in the detection process.

The second criterion was introduced to make sure that, for example, detections which had been assigned to a Predicted Group Crossing on the inclined slit group, and thus been treated by a specific geometric calibration parameter set in the estimation of their amplitude, would not turn into identified transits assigned to a Predicted Group Crossing Update on the vertical slit group, and vice versa. The same holds for the two fields of view and for the upper and lower slit branches.

As described in Sections 5.4 and 6.2, an original Tycho Input Catalogue object could be split into several Tycho Input Catalogue Revision companions, all carrying the original Tycho Input Catalogue identification number, distinguished only by a running companion number. In the terms of the second criterion, all companion predictions for the same Tycho Input Catalogue identification were treated as an update of the first prediction. So, a single detection could be assigned to all companion predictions if it was inside all assignment intervals.

For each identified transit an identity probability was computed, decreasing with increasing difference between predicted and actual transit time and with increasing difference between the Tycho Input Catalogue Revision magnitude and actual magnitude.

Parasite Recording

The second main task for the transit identification process was the so-called parasite recording. Parasites are transits (either detections or predicted group crossings; details are given in Section 7.2) above a limiting magnitude which are close enough to a given detection to produce a significant astrometric or photometric disturbance. The basic idea of their recording is to provide, for each potentially disturbed identified transit, the sum of the disturbing amplitudes. The decision whether to use the disturbed transit in the catalogue production is postponed to the astrometry and photometry processes.

A theoretical analysis of the relative additional positional scatter introduced by parasites of different brightness to detections of different brightness had shown that this scatter depends only on the brightness of the parasites (see Section 7.2). It had also been shown that recording all parasites brighter than the limiting magnitude of 10.5 was the best compromise between discarding too many detections and neglecting too many disturbances.

Parasitic amplitudes were recorded differently in the identified transit records, depending on their time distance to the detection under consideration. For distances greater than 9 samples, these amplitudes were taken directly from the raw transit data (i.e. from actual detections of other objects). For smaller distances one has to be aware that the parasite might be disturbed as well as the detection under consideration; for very small distances there may even be only one (combined) transit. So the parasitic disturbance in an interval of 18 samples centred on the detection under consideration was determined from the prediction data alone, recalculating only the parasitic amplitude from the Tycho Input Catalogue Revision magnitude given in the PGC Updating-3 data.

Basic Algorithm

The transit identification process was controlled by the PGC Updating-3 data stream. It was not possible, however, to treat the input data record by record, i.e. to read one predicted transit time, assign all appropriate raw transits, write the output data and step to the next prediction. The reason for this was that for the calculation of the parasitic amplitudes for a specific transit all detections and predictions within the parasite recording range had to be known, including those which followed the currently considered detection. Furthermore, although the prediction updating-3 process was controlled by the strictly time-ordered PGC data stream, it did not produce strictly time-ordered predictions (due to the treatment of the different Tycho Input Catalogue Revision companions described in Section 6.2). As the position of two companions of a Tycho Input Catalogue Revision star on the sky could differ by about 40 arcsec (see Section 5.4), detections and predictions of a time interval of about 150 samples in both directions had to be considered to be sure that no potential parasite was lost. But, on the other hand, in order to keep processing time as short as possible, all input records had to be read only once.

These complex and partly conflicting requirements made the transit identification software much more complicated than its basically simple task suggests. The working principle was to read the complete input data into a number of memory arrays, the 'record arrays'. Furthermore some additional arrays were used in which each element represents a time interval of one sample. These 'sample arrays' were filled during processing with amplitude data and with pointers to the corresponding PGC Updating-3 and raw transit record array elements. Internally they were split into three parts, the central part and the lower and upper buffer. The size of these buffers was set big enough to make sure that all records which could possibly contribute parasitic influences to any transit in the central region were always available.

The data processing consisted of the following actions:

- read PGC Updating-3 data until the end of the sample array is reached;
- read raw transit data for the same time interval. By filling up the raw transit sample array, identical transits can be identified. They are created in the transit detection process when a transit is assigned to more than one prediction (see Section 4.3). If two predictions belong to, for example, different slitgroups, different geometric calibration values are used in transit parameter estimation (see Section 4.4), so transit times of identical transits may differ slightly. Two transits were regarded as identical if their transit time difference was below 1 sample;
- treat the prediction records one by one in the sequence given by the PGC Updating-3 file;
- assign all appropriate raw transits for the actual prediction;
- calculate the parasitic amplitudes for all assigned transits, using data from the sample arrays;
- calculate the identity probability for each transit assigned to the actual prediction;
- write the data to the identified transit files;
- step to the next prediction;
- if the next prediction lies in the sample array range of the upper buffer, the contents of all sample arrays is shifted down by a certain amount of elements (or samples). Generally this amount is chosen in such a way that the next prediction to be treated is transported almost to the start of the central region. Thus all earlier possible parasites remain in the lower buffer;
- continue reading of PGC Updating-3 data.

Transit Identification Redoing

No differences existed in the two input data streams compared to the main processing. Thus the transit identification process for redoing was done exactly as for the main processing.

Actual Processing

During 1992, test data produced by prediction updating-3 were processed. The final production started in June 1993 and was essentially completed in March 1995, after some interruptions during which the Tübingen computers were busy with the reprocessing data (Chapter 10).

The output of the transit identification process consisted of the identified transit data stream, which was sent to Copenhagen for astrometry and stored in Tübingen for photometry. A total of 260 identified transit tapes was produced, corresponding to about 32 Gigabytes of data.

6.4. Verification Methods

Most of what was said in Section 4.5 about the verification of first prediction and detection equally applies (*mutatis mutandis*) for prediction updating and transit identification. In particular, it is true that the two processes can only be verified in combination, and that the accuracy of predicted group crossings can be checked in full depth (i.e. at the mas level) only in the detailed astrometric analysis of the identified transit data (see Chapter 7), using close-to-final star mapper calibrations and preliminary Hipparcos output catalogues.

The present section thus only shows an illustrative example of the improvements gained by prediction updating and transit identification. More details will be shown in Chapter 7 on astrometry processing.

Figure 6.1 displays a collection of 'cloud plots', showing the transit time differences 'detected minus predicted' versus the z coordinate along the slits, as explained in Section 4.5. The two plots at the top are repeated from Figure 4.10. They show the first results of prediction and detection on a stretch of data from the start of the mission, using as input the imprecise on-board attitude, a still unsatisfactory instrument geometry and the Tycho Input Catalogue (for more details see the description of Figure 4.10). The rms width of the 'clouds' is about 0.9 arcsec.

The two plots in the centre show results for a similar stretch of data, but after the application of a preliminary on-ground attitude. The rms width of the 'cloud' for the inclined slits (left) is 0.42 arcsec, dominated by the errors of the preliminary attitude. The attitude errors relevant for the vertical slits (right) are smaller by a factor of 20. Thus, the width of the clouds at right (0.26 arcsec rms) is entirely due to the errors of the star catalogue. Here, a small subset of the Tycho Input Catalogue, containing only stars with particularly good *a priori* positions, was used (this explains the small number

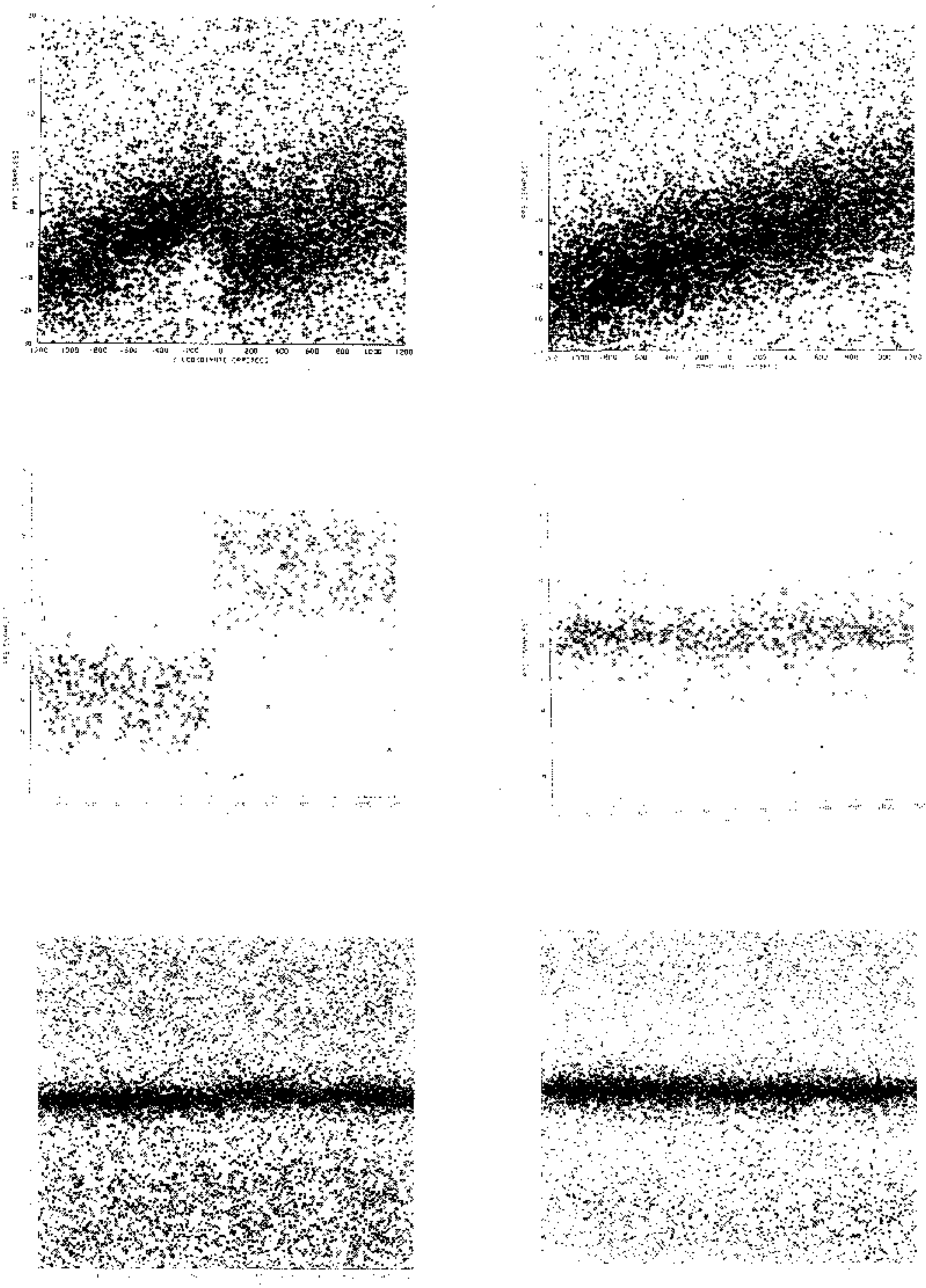

Figure 6.1. *Sample 'cloud plots' showing the improvements gained by successive stages of prediction updating and transit identification. The height of the two plots at the top corresponds to 12 arcsec on the sky, i.e. the width of the Predicted Group Crossing intervals. The other four plots have twice the vertical scale, i.e. their height corresponds to only 6 arcsec. The horizontal axis in all six plots spans the 40 arcmin length of the star mapper slits. Details are described in the text. It is evident that the discrimination between relevant transits and background transits is much improved from top to bottom (note the enlarged vertical scale).*

of points compared to the other plots). The instrument geometry has improved, but is still incorrect for the inclined slits.

The two plots at the bottom show the quality achieved by prediction updating-3 and transit identification, using as input the final on-ground attitude, a satisfactory instrument geometry, and the Tycho Input Catalogue Revision. The rms width of both clouds is 0.14 arcsec, which nicely agrees with the accuracy of the Tycho Input Catalogue Revision.

The cloud plots were produced only occasionally. Routine quality control during the mass production of transit identification was done using simple histogram plots showing the distribution of the differences between predicted and observed transit times, separately for both slit groups and fields of view. About one such plot per day of mission, covering 10 per cent of the data for that day, was checked for the occurrence of any unexpected features such as large offsets or unduly large scatter. Not a single serious alarm was raised; all (rare) cases of suspect histogram plots could be explained by disturbed data, mostly due to high radiation background.

U. Bastian, K. Wagner