9. DE-CENSORED MAGNITUDES OF FAINT STARS

The Tycho Catalogue is dominated by stars too faint to be detected every time they were crossing the star mapper slit systems. This means that some of their transits got lost in the detector noise. Simple median magnitudes of detections consequently have a large bias, of 0.2 and 0.8 mag at V_T = 10 and 11 mag. Therefore, the magnitudes of the faint stars have been derived not only from the detections, but taking into account the censored measurements. This procedure, called 'de-censoring', was used for computing the B_T and V_T magnitudes of about 1 018 000 stars. A comparison with ground-based photometric standard magnitudes showed that the bias was reduced to a value of 0.025 and 0.07 mag at V_T = 10 and 11 mag, which is less than the standard errors. The magnitudes have been corrected for this systematic error, and it has been verified that the quoted standard errors are fairly realistic external errors.

9.1. Introduction

The Tycho catalogue contains all the stars which received enough detections for the derivation of an astrometric solution with an acceptable accuracy. Among several conditions, a minimum number of 30 transits usable in astrometry was requested. Since this number is only about 25 per cent of the mean number of transits per star, the catalogue content is dominated by stars which are too faint to be detected each time they crossed the star mapper slit systems.

For these stars, the acquisition of a detection depends on two factors. The first is related to the circumstances of the transits; since the transits were detected only when the signal-to-noise ratio was larger than 1.5, the limit of detectability was shifted toward bright magnitudes when the background increased; moreover, the mean number of photons received from the star depends also on the slit system (vertical or inclined), the field (preceding or following), and so on. The second factor is a matter of chance; the actual number of photons counted when a star is crossing a slit system is approximately the mean value, but with statistical variations. When only the signals above a threshold are detected, the fainter signals are censored, and the remaining ones are on average brighter than the mean magnitude of the star. An estimation of this bias was derived from standard stars and is plotted in Figure 8.3(b) for the V_T magnitude. It appears from this figure that all stars fainter than V_T = 10.5 mag have a bias increasing with the true magnitudes in such a way that their measured magnitudes were in fact always about 10. Therefore, the bias cannot be determined from the mean measured magnitude alone.

The median magnitudes were kept for about 30 000 stars with $V_T < 8$ mag and $B_T < 8.5$ mag. The magnitude of the other stars were computed with a method taking into account the censored measurements. This method is presented hereafter. It is based on a statistical model that is described in Section 9.2. The verification of its validity and the derivation of some parameters is presented in Section 9.3. The complete procedure is explained in Section 9.4. The results and the correction of the remaining bias are presented in Section 9.5.

This chapter describes the de-censoring as it was actually carried out. The earlier brief descriptions in Section 1.6 and in Volume 1, Section 2.2 differ slightly.

9.2. The Statistical Model

Before retrieving information from non-detections, it is necessary to know how the measurements were done. The data acquisition was described previously in Chapters 4 and 8, but the process is too complicated to be taken completely into account. The statistical model is a simplified summary of the whole process, restricted to the main characteristics.

General Outline and Detection Censoring

In the absence of any star, background photons were counted in both the B_T and V_T channels. The intensity of the background was estimated in each channel, each time the transit of a star was predicted. When a star was crossing a slit system of the star mapper, the increase of the photon count was estimated within an interval of 6.7 samples (1 sample = 1/600 s); 6.7 is expressed as 1/0.15 hereafter.

The total number of photons contributing to the detection of the star was then: $N_{B_T+V_T} = n_{\text{back}_B} + n_{\text{back}_V} + n_B + n_V$, where $n_{\text{back}_B} + n_{\text{back}_V}$ photons were emitted by the background, and $n_B + n_V$ photons were emitted by the star in the two channels. $N_{B_T+V_T}$ obeys a Poisson distribution; this distribution depends only on one parameter, the mean value of $N_{B_T+V_T}$, denoted $\langle N_{B_T+V_T}\rangle$ hereafter:

$$\langle N_{B_T+V_T}\rangle = \langle n_{\text{back}_B}\rangle + \langle n_{\text{back}_V}\rangle + \langle n_B\rangle + \langle n_V\rangle \qquad [9.1]$$

$\langle n_{\text{back}_B}\rangle$ and $\langle n_{\text{back}_V}\rangle$ were derived from the intensities of the background expressed in counts per sample. $\langle n_B\rangle$ and $\langle n_V\rangle$ are related to the actual magnitudes of the star, B_T and V_T, in the following equations:

$$\langle n_B\rangle = 10^{0.4(B_{\text{cal}}-B_T)}/0.15 \qquad [9.2]$$

$$\langle n_V\rangle = 10^{0.4(V_{\text{cal}}-V_T)}/0.15 \qquad [9.3]$$

where B_{cal} and V_{cal} are calibration terms representing the instrument sensitivity. These terms were called m_{off} in Chapter 8; their values are given in Table 8.1.

The detection of a stellar transit depended on its signal-to-noise ratio, $SNR_{B_T+V_T}$. The signal-to-noise ratio of a signal S is generally defined as $SNR = S/\sigma_S$, and for a Poisson distribution with the parameter λ, $\sigma_\lambda = \sqrt{\lambda}$. $SNR_{B_T+V_T}$ is then defined as:

$$SNR_{B_T+V_T} = \frac{N_{B_T+V_T} - \langle n_{\mathrm{back}_B} \rangle - \langle n_{\mathrm{back}_V} \rangle}{\sqrt{N_{B_T+V_T}}} \qquad [9.4]$$

Thus, the statistical uncertainties of the two background contributions were neglected because they were formed as averages over long stretches of time. When $SNR_{B_T+V_T}$ was larger than 1.5, the detection was recorded and the magnitudes of the stars were searched for in both channels. When $SNR_{B_T+V_T}$ was below 1.5 the detection was censored, but, since the transit was predicted, the non-detection was recorded too; a non-detection consists essentially of the estimations of the background, and the calibration terms (or flags permitting their reconstruction).

The Spurious Non-Detections

The non-detections were not all generated by transits below the detection threshold. The assignment of a detection to a given star depended on the following successive conditions:

- the epoch of the detection was close to the predicted transit time. The limit in time corresponded to an offset in position of 3 arcsec, but, on the other hand, the predicted transit times used in detection were inaccurate, since they were based on a preliminary determination of the attitude of the satellite;
- after the updating of the prediction using the final attitude determination, the distance between the star and the position of the slit was less than 0.6 arcsec. This value was derived from an histogram of the distances, in order to discard the false detections generated by random variations of the background (this selection is less critical in astrometry, where a limit of 1 arcsec was applied).

When the two conditions above were not satisfied, the actual detections of the stars were lost, but the transits were recorded as non-detections. These lost detections are referred to as 'spurious non-detections'. The probability that a detection could be lost and changed into a spurious non-detection was included in the model as a constant parameter.

The Magnitude Censoring

When a detection was obtained, the photon counts in the two channels were considered separately in order to determine the B_T and V_T magnitudes. Using the background level estimated by the detection process (see Chapters 2 and 4), one obtains:

$$B_{\mathrm{meas}} = B_{\mathrm{cal}} - 2.5 \log(N_B - \langle n_{\mathrm{back}_B} \rangle) - 2.5 \log 0.15 \qquad [9.5]$$

$$V_{\mathrm{meas}} = V_{\mathrm{cal}} - 2.5 \log(N_V - \langle n_{\mathrm{back}_V} \rangle) - 2.5 \log 0.15 \qquad [9.6]$$

where B_{meas} and V_{meas} are the measurements of B_T and V_T. In practice, however, the magnitudes were derived from a slightly different method, and measurements were obtained only when it was possible to fit the slit response function to the photon counts. As a consequence, the faintest measurements were censored. This second censoring is called the 'magnitude censoring' hereafter. It took effect below a threshold that had to be determined. The B_T or V_T values which were not measured are flagged as such in the Tycho Epoch Photometry Annexes.

9.3. The Adaptation of the Model to the Actual Data

Since the model in the previous section is a simplification of the complexity of the processes involved in the photometric reduction, it was necessary to check that the data had the properties that were assumed. Moreover, two parameters still remained to be determined: the threshold of the magnitude censoring and the proportion of spurious non-detections.

In practice, several different versions of the model were used to compute de-censored magnitudes for about 1 018 000 stars of the Tycho catalogue. The parameters derived hereafter refer to the version used essentially for the 721 000 faintest stars. About 297 000 stars were treated with slightly different versions, but it was verified that the results were similar for the range of magnitudes where they were applied.

Correction of the Actual Measurements and Magnitude Censoring

The assumption that the cumulated photon counts obeyed a Poisson distribution with the parameter $\langle N_{B_T+V_T} \rangle$ derived in Equations 9.1 to 9.3 was the keystone of the model. This hypothesis had to be checked, especially for faint stars, when the censoring was important. On the other hand, the distribution of $N_{B_T+V_T}$ derived from the actual measurements cannot be directly compared with a Poissonian distribution, since it was biased as a result of the censoring. The following approach was then used to solve this problem: Only the B_T measurements of red standard stars were considered. Since these stars have B_T much fainter than V_T, their detections depended essentially on the photon counts in the V_T channel. Therefore, the statistics of the photon counts in B_T were not affected by the detection censoring, but only by the magnitude censoring (symmetrically, the V_T measurements of blue stars had the same property). The idea was then to check if the photon counts N_B obtained for a given $\langle N_B \rangle$ obeyed a Poisson distribution with parameter $\langle N_B \rangle$ as long as N_B is larger than a limit given by the magnitude censoring.

It appeared that the distribution of N_B contained an excess of measurements with large N_B. This was attributed to an overestimation of n_B, the number of photons received from the stars, or, in other words, to a non-linearity of the initial Tycho estimation process. Investigations were carried out in order to evaluate this bias. In practice, the amplitude of the signal, S_B, expressed in counts per sample, was considered instead of n_B; they are equivalent, since $S_B = 0.15 \times n_B$. For each transit, the mean signal $\langle S_B \rangle$ was derived from the real B_T magnitude and from the calibration terms according to Equation 9.2. At the same time, the estimated intensity of the background, $\mathrm{back_B}$, expressed in counts per sample, was used instead of $\langle n_{\mathrm{back_B}} \rangle$; again, $\mathrm{back_B} = 0.15 \times \langle \mathrm{n_{back_B}} \rangle$. The measurements of several standard stars in narrow intervals of $\mathrm{back_B}$ and of $\langle S_B \rangle$ were considered together, and the distributions of S_B actually obtained were compared to the theoretical distributions. An example of these comparisons is represented in Figure 9.1. Various thresholds in the signal-to-noise ratio in B_T were introduced, in order to take into account the magnitude censoring. It was found that the distribution of S_B is in agreement with the theoretical distribution when the following conditions are satisfied:

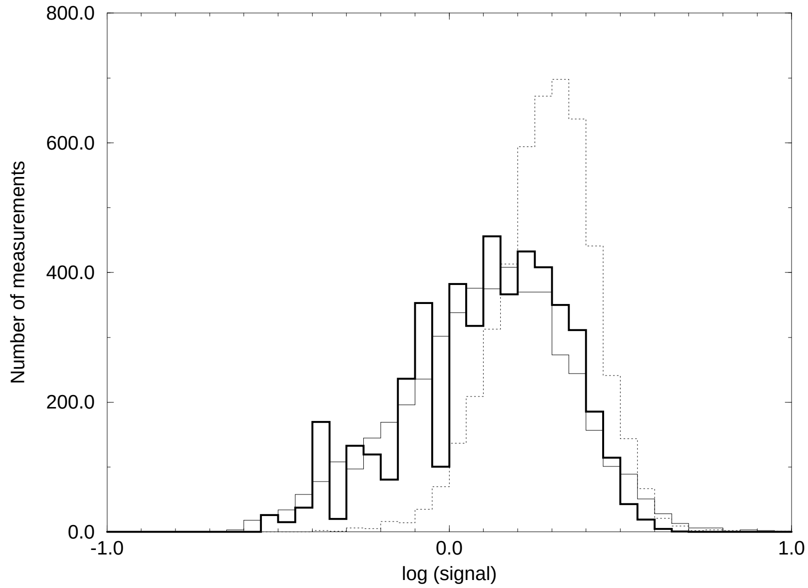

Figure 9.1. *The distribution of the signals in B_T, derived from the measurements of the red standard stars with $B_T - V_T > 0.7$ fulfilling the following conditions: transit of the inclined slit system, $\langle S_B \rangle$ between 0.5 and 2 counts per sample, and* $\mathrm{back_B}$ *between 1 and 2 counts per sample. In order to take into account the magnitude censoring, the measurements with a signal-to-noise ratio $SNR_B < 0.5$ are discarded. The dotted line refers to the signals estimated in the photometric reduction, and the thin solid line to the corrected signal. The distribution derived from the model is plotted as a thick line for comparison; it contains gaps and peaks because the total photon count in B_T, N_B, was assumed to be an integer number.*

- S_B was corrected using a relation depending on its initial estimation, on the background, and on the slit system of the star mapper. The corrected S_B are given in Table 9.1. This correction was important for B_T fainter than about 11 mag, but it became negligible for brighter measurements;
- the corrected measurements of B_T with signal-to-noise ratios $SNR_B < 0.5$ were discarded.

The measurements of the V_T magnitude should fit the model under the same conditions, because they were derived with the same algorithms as the measurements of B_T. The problems of the adequacy of the Poisson distribution and of the determination of the magnitude censoring were thus solved at the same time.

Verification of the Detection Threshold

It was assumed in the model that the detection condition was $SNR_{B_T+V_T} > 1.5$, with $SNR_{B_T+V_T}$ derived from Equation 9.4. Although the detection threshold $SNR = 1.5$ was included in the detection process (see Chapter 6), the validity of this hypothesis needed to be confirmed, because the signal estimations used in the detection process were derived from a different algorithm to the estimations used in photometry. Therefore, the detections were not actually based on $SNR_{B_T+V_T}$, but on an estimation which is

Table 9.1. Calculation of the corrected signal from the initial estimation and from the background, according to the slit systems; signals and backgrounds are expressed in counts per sample.

Slits			Vertical	Inclined
$S < 2$	and	back < 2	$S \times (0.8 \ln(1 + S) - 0.23)$	$S \times (0.8 \ln(1 + S) - 0.2)$
		back $\in [2, 3[$	$S \times (0.9(\ln(1 + S))^{1.1} - 0.45)$	$S \times (0.8(\ln(1 + S))^{1.3} - 0.35)$
		back $\in [3, 4[$	$S \times (0.9(\ln(1 + S))^{1.1} - 0.59)$	$S \times (0.8(\ln(1 + S))^{1.2} - 0.43)$
		back $\in [4, 5]$	$S \times (0.9(\ln(1 + S))^{1.1} - 0.638)$	$S \times (0.8(\ln(1 + S))^{1.2} - 0.505)$
$S > 2$	or	back > 5	$S - 3 \times \text{back}^2/(S + \text{back})^2$	$S - 2 \times \text{back}^2/(S + \text{back})^2$

called SNR_{det} hereafter. For this reason, it is necessary to check that, after corrections of the signals S_B and S_V, $SNR_{B_T+V_T}$ is equivalent to SNR_{det}.

A random set of detections with SNR_{det} close to 1.5, and with available measurements of both B_T and V_T was considered. The signals S_B and S_V were derived from the measured magnitudes B_T and V_T, and they were corrected by means of Table 9.1. $SNR_{B_T+V_T}$ were then calculated, and compared to SNR_{det}. It appeared that the median of $SNR_{B_T+V_T}$ was close to the median SNR_{det}, but the scatter was rather large: for 70 per cent of the detections, the ratio $SNR_{\text{det}}/SNR_{B_T+V_T}$ was between 0.8 and 1.6.

The detection condition $SNR_{B_T+V_T} > 1.5$ was finally considered as valid, although it is only an approximation of the actual statistical properties of the data. In reality, some transits were detected although $SNR_{B_T+V_T}$ was below the threshold, and some others were censored although they should have been detected.

The Proportion of Spurious Non-Detections

The proportion of spurious non-detections was derived from bright photometric standard stars: it was found that about 6 per cent of the transits of bright stars resulted in non-detections, which were in fact lost detections. For both faint and bright stars, this percentage is also the proportion of spurious non-detections among all the transits that would have been detected if they were searched in the right place.

9.4. The De-Censoring Procedure

As a matter of taxonomy, according to the definitions reported by Kendall & Stuart (1961), the censoring in Tycho is a ‘censoring on the left of Type I’, since the measurements smaller than a threshold were lost. Various techniques for solving statistical problems involving missing data were developed during the last decades (Little & Rubin 1987). Some of them were successfully applied in astronomy (Feigelson & Nelson 1985, Isobe *et al.* 1986). However, the censoring of the Tycho data was rather unusual, since it came from two different origins: the detection censoring was related to the photon counts in the B_T and in the V_T channels considered together, but the magnitude censoring depended only on the photon counts in one channel. Moreover, the rate of spurious non–detections should also be taken into account. Therefore, the iteration

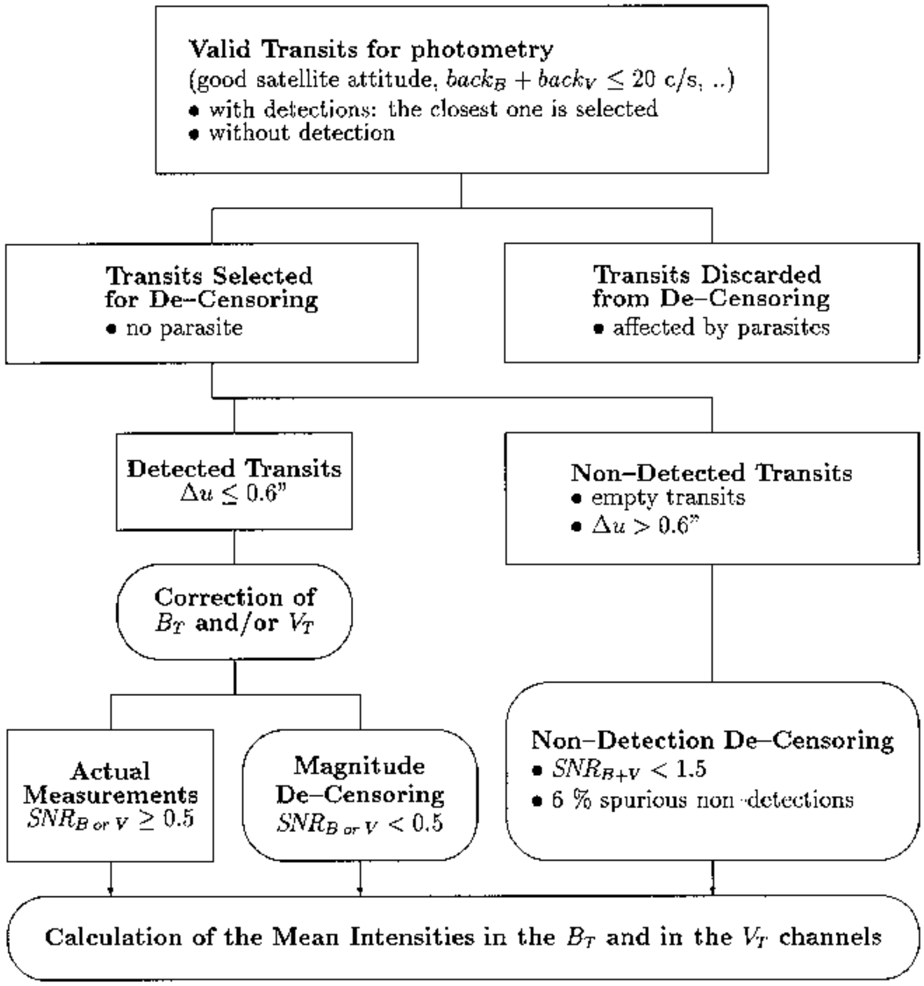

Figure 9.2. *The overall organisation of de-censoring. The data and the condition of selection are indicated in rectangular boxes; the oval boxes represent the processes.*

technique explained hereafter was preferred over the more sophisticated methods, such as 'survival analysis', presented in the papers mentioned above.

The basic principle of the calculation consisted of computing the mean intensity corresponding to each censored measurement, assuming the B_T and V_T magnitudes of the star were known. The mean intensities were related to the censored measurements, and new estimations of B_T and V_T were derived from all transits. The calculation was then repeated until it had converged. The different processes are explained hereafter, and the overall organisation is summarized in Figure 9.2.

The Input Data

Every time a star of the catalogue crossed a slit system of the star mapper, the following data were recorded in the complete photometric observation catalogue (they are also given in the Tycho Epoch Photometry Annex for a selection of stars):

- a parasite flag, indicating when another star was crossing a slit system at the same time;

- the background in both channels back_B and back_V, expressed in counts per sample;
- a set of flags determining B_{cal} and V_{cal}, the calibration terms defined in Equations 9.2 and 9.3.
- when the transit was detected, the photometric reduction attempted to derive the measurements B_{meas} and V_{meas}.

The transits affected by parasites or with the total background $\mathrm{back}_B + \mathrm{back}_V > 20$ counts per sample were not taken into account.

Each iteration was based on *a priori* values of the magnitudes, B_{iter} and V_{iter}. These values were assumed to be constant in time; as a consequence, the magnitudes derived from de-censoring are different from the median magnitudes when the stars are variable.

The Weights of the Transits

The transits of each star were obtained with backgrounds and instrumental sensitivities varying from one transit to another. In order to base the calculation on the most reliable transits, weights were derived from the transit parameters and assigned to the transits in each channel. The weights were defined as the squares of the signal-to-noise ratios, assuming the magnitudes B_{iter} and V_{iter}; for each channel, they were:

$$W_M = \frac{10^{0.8(M_{\mathrm{cal}} - M_{\mathrm{iter}})}}{0.15 \times (10^{0.4(M_{\mathrm{cal}} - M_{\mathrm{iter}})} + \mathrm{back}_M)} \qquad [9.7]$$

where M designates the B_T or the V_T channel. A couple of weights (W_B, W_V) was assigned to each transit used in the de-censoring, censored or not.

The Treatment of the Actual Measurements

When a measurement of the magnitude was recorded, it was used to reproduce the original signal:

$$S_M = 10^{0.4(M_{\mathrm{cal}} - M_{\mathrm{meas}})} \qquad [9.8]$$

The corrected signal, S'_M, was derived from the formulae in Table 9.1. The signal-to-noise ratio in the channel was then calculated from:

$$SNR_M = \frac{S'_M}{\sqrt{0.15 \times (S'_M + \mathrm{back}_M)}} \qquad [9.9]$$

When SNR_M was less than 0.5, the measurement of the magnitude was censored. Otherwise, the measured intensity of the star was derived by:

$$I_M = S'_M \, 10^{-0.4 M_{\mathrm{cal}}} \qquad [9.10]$$

I_M received the weight W_M calculated above and it was included in the calculation of the mean intensity of the star, $\langle I_M \rangle$, in Equation 9.18.

The Censored Magnitudes

A magnitude measurement was censored when a transit was detected, but the photometric reduction failed to derive a measurement, or when the corrected signal provided

$SNR_M < 0.5$. It was assumed that the censored magnitudes all satisfied this latter condition, and the upper limit of the total number of photons in the channel 'M', called $N_{M\,\mathrm{sup}}$ was derived from the equation:

$$\frac{N_{M\,\mathrm{sup}} - \mathrm{back}_M/0.15}{\sqrt{N_{M\,\mathrm{sup}}}} = 0.5 \qquad [9.11]$$

On the other hand, the actual number of photons, N_M, obeyed a Poisson distribution with the parameter $\langle N_M \rangle$:

$$\langle N_M \rangle = 0.15 \times (10^{0.4(M_{\mathrm{cal}} - M_{\mathrm{iter}})} + \mathrm{back}_M) \qquad [9.12]$$

The average photon count when $N_M < N_{M\,\mathrm{sup}}$ was thus derived. The mean intensity corresponding to the censored magnitude was then:

$$I_M = (0.15 \times \langle N_{M;N_M < N_{M\,\mathrm{sup}}} \rangle - \mathrm{back}_M) \times 10^{-0.4M_{\mathrm{cal}}} \qquad [9.13]$$

I_M then received the weight W_M, and it was taken into account in the calculation of the mean intensity of the star, as if it were derived from an actual measurement. This is the central idea of the de-censoring procedure.

The Censored Detections

The calculation of mean intensities had to take into account the contribution of 'censored detections' when a predicted transit was not detected.

The censored detections were due to a total photon count in both channels, $N_{B_T+V_T}$, below an upper limit related to the detection condition in signal-to-noise ratio. This limit, referred to as $N_{B_T+V_T\,\mathrm{sup}}$, was the solution of the equation:

$$\frac{N_{B_T+V_T\,\mathrm{sup}} - (\mathrm{back}_B + \mathrm{back}_V)/0.15}{\sqrt{N_{B_T+V_T\,\mathrm{sup}}}} = 1.5 \qquad [9.14]$$

with $N_{B_T+V_T} = N_B + N_V$, where N_B and N_V obeyed Poisson distributions with the parameters $\langle N_B \rangle$ and $\langle N_V \rangle$ derived from Equation 9.12. The mean intensities $I_{B\,\mathrm{cen}}$ and $I_{V\,\mathrm{cen}}$ were computed from simulations of pairs (N_B, N_V), taking into account only those such as $N_B + N_V < N_{B_T+V_T\,\mathrm{sup}}$ (in practice, only $N_B < N_{B_T+V_T\,\mathrm{sup}}$ and $N_V < N_{B_T+V_T\,\mathrm{sup}}$ were generated). The probability of getting a censored detection, P_{cen}, was calculated at the same time.

Spurious Non-Detections

Since spurious non-detections were lost detections, they also contributed to the censored data. However, the censoring is then in the opposite sense, it being assumed that the transits would have been detected if they had been searched for at the correct place. For a lost detection, the mean intensities $I_{B\,\mathrm{lost}}$ and $I_{V\,\mathrm{lost}}$ could in principle be calculated analogously to $I_{B\,\mathrm{cen}}$ and $I_{V\,\mathrm{cen}}$, but now assuming $N_B + N_V > N_{B_T+V_T\,\mathrm{sup}}$. In practice, however, $I_{B\,\mathrm{lost}}$ and $I_{V\,\mathrm{lost}}$ were derived from the intensities due to censoring and from the probability of censoring with the equation:

$$I_{M\,\mathrm{lost}} = \frac{I_{M\,\mathrm{iter}} - P_{\mathrm{cen}} I_{M\,\mathrm{cen}}}{1 - P_{\mathrm{cen}}} \qquad [9.15]$$

where $I_{M\,\mathrm{iter}}$ was derived from the input magnitude assumed in the present iteration step:

$$I_{M\,\mathrm{iter}} = 10^{-0.4M_{\mathrm{iter}}} \qquad [9.16]$$

Merging the Censored Detections and the Spurious Non-Detections

It was assumed that, when the photon count was above the detection threshold, it was possible that the detection was lost, with the probability $P_{\rm lost} = 6$ per cent. The intensities corresponding to censoring and to a lost detection were then combined. Expressing $I_{M\ \rm lost}$ as in Equation 9.15, the mean intensities for a transit without detection were finally:

$$I_M = \frac{P_{\rm lost} I_{M\ \rm iter} + P_{\rm cen}(1 - P_{\rm lost}) I_{M\ \rm cen}}{P_{\rm cen} + (1 - P_{\rm cen}) P_{\rm lost}} \qquad [9.17]$$

I_M then received the weight calculated in Equation 9.7.

The Mean B_T and V_T Magnitudes and the Convergence Criterion

After the calculations above, all the transits selected in de-censoring received intensities in the B_T and in the V_T channel. The actual measurements were tranformed to intensities by Equation 9.8, the correction in Table 9.1, and Equation 9.10. Mean intensities were related to the missing (censored) measurements censoring. The next process was then the derivation of the mean B_T and V_T magnitudes of the star for the iteration step.

In on-ground photometry, the mean magnitude of a star is computed as the average of the magnitude measurements. This method is used since the distribution of the logarithm of the photon counts obeys a Gaussian law, due to scintillation (Sterken & Manfroid 1992). This was not true for Tycho photometry, and the mean magnitudes were derived from the average intensities. For each channel, the average intensity of the star, $\langle I_M \rangle$, was computed from:

$$\langle I_M \rangle = \frac{\sum W_M I_M}{\sum W_M} \qquad [9.18]$$

The magnitude was then:

$$M_{\rm sol} = -2.5 \log \langle I_M \rangle \qquad [9.19]$$

The input magnitude of the next iteration step depended on the direction of the shift of the solution, $M_{\rm sol}$, from the input magnitude $M_{\rm iter}$. When the shift was in the same direction as in the previous step, the new input $M_{\rm iter+1}$ was even farther from $M_{\rm iter}$ than $M_{\rm sol}$. On the other hand, when $M_{\rm sol}$ was coming back toward the previous input magnitude, $M_{\rm iter-1}$, then $M_{\rm iter+1}$ was exactly $M_{\rm sol}$; moreover, a convergence flag was then turned on, when the shift was less than 0.05 mag.

New iterations were computed as long as the convergence was not obtained in both channels. When one of the convergence flags was on, it was revisited after each iteration, and it was turned off when the new solution was farther than 0.05 mag from the magnitude obtained when it was turned on. When B_T and V_T had both converged, an ultimate iteration was still done, including the computation of the percentiles and the estimation of the errors.

Computation of the Percentiles

The median and the 15th and 85th percentiles were computed for the measurements of V_T. This calculation was different from the calculation of the mean magnitudes in two aspects: no weights were assigned to the transits, and the censored or missing measurements were not changed in mean measurements, but in the distribution function of the magnitude. In practice, a histogram of the V_T magnitudes was calculated. The bin size of the histogram depended on the magnitude derived in the previous iteration. For stars fainter than 8.8 mag, it was:

$$\text{step} = \min(0.008 + 0.017 \times (V_T - 8.8)^2\ ,\ 0.10) \qquad [9.20]$$

Each actual measurement was simply counted in the corresponding bin. Each censored magnitude became a normalized distribution function which was integrated in the bins corresponding to magnitudes fainter than the measurement threshold. The treatment of the missing detections was more complicated, since the distributions were not restricted to magnitudes fainter than the detection threshold, due to the probability of missing detections. It was then necessary to calculate the distribution functions for the whole range of the histogram.

Each percentile was derived from the histogram of the magnitudes, using a linear interpolation on the bin where it was found.

The percentiles derived from de-censoring should be used with caution, since a basic hypothesis of de-censoring is that the stars are not variable. For this reason, they were included in the Tycho catalogue only when V_T was fainter than 9.5 mag. For stars brighter than this limit, the percentiles derived from the actual measurements were preferred, but a bias correction was applied.

Estimation of the Errors

The errors of the magnitudes B_T and V_T of the stars were derived from estimations of the variances of the mean intensities $\langle I_B \rangle$ and $\langle I_V \rangle$:

$$\text{var}(\langle I_M \rangle) = \frac{\sum W^2 \text{var}(I_M)}{(\sum W)^2} \qquad [9.21]$$

The problem was then to calculate the variance of the intensity for each transit, $\text{var}(I_M)$. The calculation was easy for the actual measurements since:

$$I_M = (0.15 N_M - \text{back}_M) \times 10^{-0.4 M_{\text{cal}}} \qquad [9.22]$$

and since the background was a constant determined for the transit, it appeared that:

$$\text{var}(I_M) = 0.15 \times (I_M 10^{-0.4 M_{\text{cal}}} + \text{back}_M 10^{-0.8 M_{\text{cal}}}) \qquad [9.23]$$

where I_M is the corrected intensity calculated by Equation 9.10.

The variance of the mean intensity derived for a censored magnitude or for a missing detection was much more difficult to calculate. It depended on the accuracy of the assumed magnitude (but this is the very result that was searched), and also on the accuracy of the model. The difference between $SNR_{B_T+V_T}$ and SNR_{det} was neglected in calculating I_B and I_V, but the effect of this simplification could not be ignored in the estimation of the variances. A problem this complicated cannot be solved analytically, but only with simulations; unfortunately, this technique would have been too time-consuming to treat one million stars within the time limit. On the other hand, the errors of the magnitudes of the photometric standard stars were derived from the actual measurements only, and it appeared that they were reliable estimates. Therefore, the censored detections and the censored magnitudes were finally ignored in calculating Equation 9.21.

The standard deviation of the intensity, $\sigma_{\langle I_M \rangle}$ was computed as $\sqrt{\mathrm{var}(\langle I_M \rangle)}$, and the error of the magnitude 'on the bright side', called σ_M^-, was defined as the offset in magnitude when the intensity was increased by $\sigma_{\langle I_M \rangle}$:

$$\sigma_M^- = 2.5 \log \left(1 + \frac{\sigma_{\langle I_M \rangle}}{\langle I_M \rangle}\right) \qquad [9.24]$$

σ_M^- is called the 'error on the bright side' since the probability that the actual magnitude is brighter than $M - \sigma_M^-$ is 16 per cent. On the other hand, the 'error on the faint side' is the error corresponding to a 1σ decrease of the intensity; it may be derived from σ_M^- with Equation 9.25:

$$\sigma_M^+ = -2.5 \log(2 - 10^{0.4\sigma_M^-}) \qquad [9.25]$$

σ_M^+ is larger than σ_M^-, and it is impossible to derive when $\sigma_M^- > 0.75$ mag. For this reason, only $\sigma_{B_T}^-$ and $\sigma_{V_T}^-$ were given in the Tycho Catalogue.

9.5. Verification of the Results of De-Censoring

Bias of the De-censored Magnitudes

The differences between the de-censored magnitudes and those of standard stars are plotted in Figure 9.3 for the V_T magnitudes. The large bias toward bright magnitudes that was present in Figure 8.3(b) has disappeared, but a closer examination of this plot reveals that the de-censored magnitudes are about 0.06 mag too faint when $V_T >$ 10.6 mag. For the B_T channel, a 0.05 mag bias was also found around $B_T = 11.4$ mag. These biases were determined as functions of the actual on-ground magnitudes, and were next converted into functions of the de-censored magnitudes. Since the number of standard stars was very small for $V_T > 11$ mag and $B_T > 11.8$ mag, the corrections are accurate only below these limits. They were added to the de-censored magnitudes, resulting in the final V_T and B_T. In Figure 9.4, the thick line represents the proportion of photometric standard stars with Tycho V_T fainter than on-ground V minus 50 per cent; it is always close to 0 per cent, confirming that the bias was properly corrected.

Verification of the Error Estimates

If the computation of the errors was correct, the proportion of stars actually brighter than $M - \sigma_M^-$ should be 16 per cent (again, M refers to the magnitude B_T or V_T). On the other hand, σ_M^+ should provide a 16 per cent proportion of stars actually fainter than $M + \sigma_M^+$. This was verified from the standard stars. In practice, the errors of the on-ground magnitudes must also be taken into account, and the differences between the Tycho magnitudes and the on-ground magnitudes were compared to the quadratic sums of the Tycho errors and of the on-ground errors. The results for V_T are presented in Figure 9.4. For V_T brighter than 9.5 mag, the proportion of stars outside the error

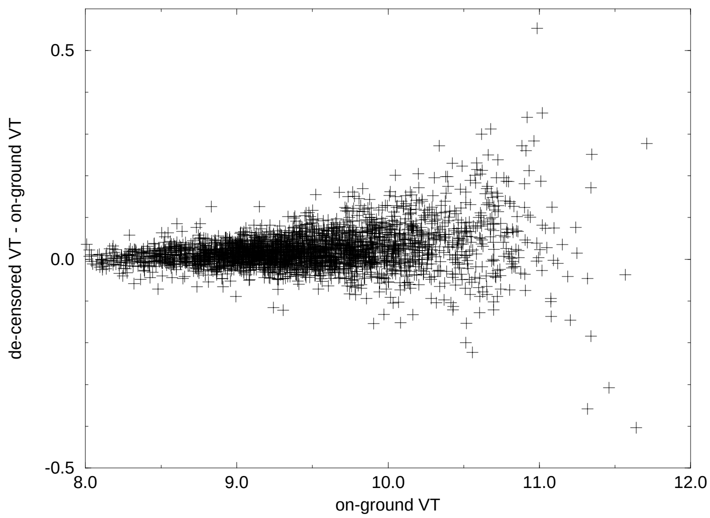

Figure 9.3. *The slight remaining bias of the de-censored V_T magnitudes versus the true magnitudes of standard stars. This bias was corrected for the final Tycho magnitudes (see Figure 9.4).*

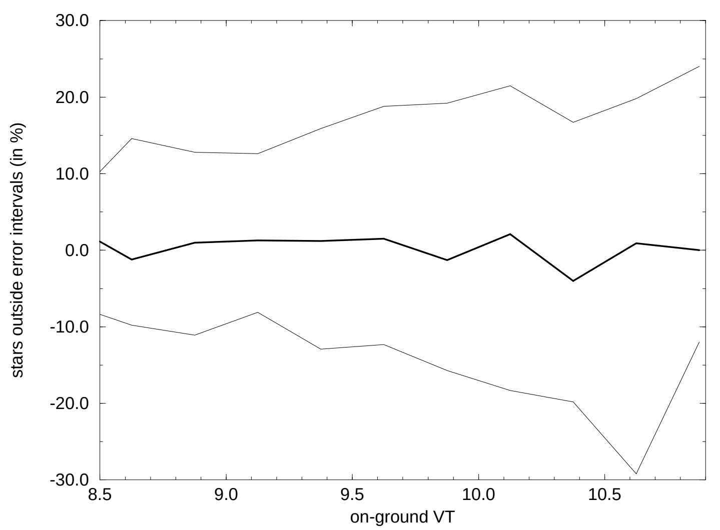

Figure 9.4. *The proportions of stars outside the error intervals for V_T. The upper line is the proportion of photometric standard stars with Tycho magnitudes too faint at the 1σ level when compared to on-ground magnitudes. The thick line is the excess of stars with Tycho magnitudes fainter than the on-ground magnitudes (see explanations in the text). The lower line is the proportion of photometric standard stars with Tycho magnitudes too bright at the 1σ level; for readability of the figure, it is given as negative. Note that the figure thus contains the faint stars at the top. The proportions of stars outside the error interval appear too small when V_T is brighter than 9.5 mag, but this is due to overestimations of the errors of the on-ground magnitudes.*

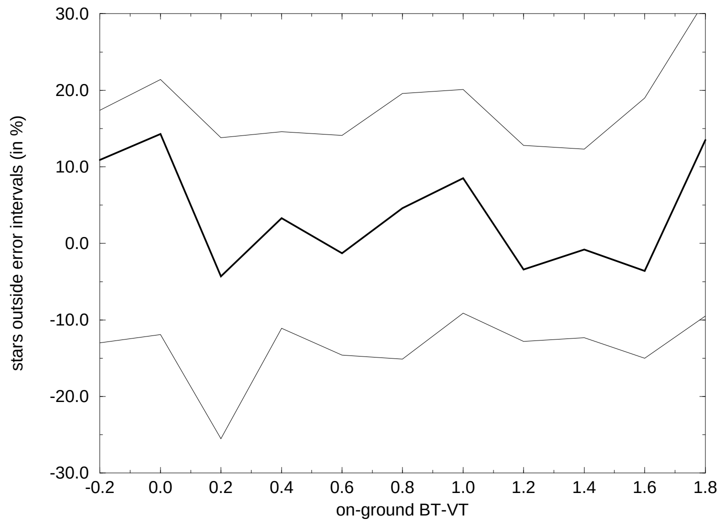

Figure 9.5. *The proportions of stars outside the error intervals for $B_T - V_T$ derived from the photometric standard stars fainter than $B_T = 9.5$ magnitudes. The upper line is the proportion of stars with Tycho $B_T - V_T$ too large (or too red) at the 1σ level. The thick line is the excess of stars with Tycho $B_T - V_T$ larger (or redder) than the corresponding on-ground value. The lower line is the proportion of stars with Tycho $B_T - V_T$ too small (too blue) at the 1σ level; for readability of the figure, it is given as negative. Note that the figure thus contains the red stars at the top.*

intervals are even smaller than expected, suggesting that the Tycho errors could be overestimated. In fact, for so bright magnitudes, the on-ground errors are on average 3 times larger than the Tycho errors, and the overestimation concerns obviously the former rather than the latter. For V_T around 11 mag, the ratio of the errors is the opposite, and the agreement is acceptable.

The estimation of the error of $B_T - V_T$ was also verified, as well as the absence of bias in the colour indices. Since a colour index corresponds to the ratio of two intensities, it is not straightforward to derive a standard deviation. The errors of $B_T - V_T$ were derived from the 16 per cent and 84 per cent percentiles, which were calculated by assuming that the errors of $\langle I_B \rangle$ and $\langle I_V \rangle$ obeyed a Gaussian distribution in a good approximation. These percentiles were then calculated using the formulae:

$$\sigma^-_{B_T - V_T} = 2.5 \log \frac{c + t}{c - 1/t} \qquad [9.26]$$

and

$$\sigma^+_{B_T - V_T} = 2.5 \log \frac{c + 1/t}{c - t} \qquad [9.27]$$

with c given by

$$c = \sqrt{\left(\frac{\langle I_B \rangle}{\sigma_{\langle I_B \rangle}}\right)^2 + \left(\frac{\langle I_V \rangle}{\sigma_{\langle I_V \rangle}}\right)^2 - 1} \qquad [9.28]$$

and t given by

$$t = \frac{\sigma_{\langle I_B \rangle} \langle I_V \rangle}{\langle I_B \rangle \sigma_{\langle I_V \rangle}} \qquad [9.29]$$

$\sigma^-_{B_T - V_T}$ is the error on the 'blue side' since the probability is 16 per cent that the actual $B_T - V_T$ is less than the measured one minus $\sigma^-_{B_T - V_T}$. On the other hand, $\sigma^+_{B_T - V_T}$ is the error on the 'red side'. For simplicity, only the largest of the two was given in the Tycho

Catalogue. The proportions of stars outside the error interval given by Equations 9.26 and 9.27 were derived for both sides, from the photometric standard stars. Again, the errors of the on-ground magnitudes were taken into account. The proportion of stars with Tycho $B_T - V_T$ larger than the corresponding on-ground value was also derived, in order to check that the Tycho colour indices were not biased. The results are shown in Figure 9.5. Apart from random variations due to the small number of standard stars (270 per bin on average, but only a few dozens at the extremities of the plot) the errors and the median look satisfactory.

J.-L. Halbwachs