Memorandum on determination of the VCM of engineering tangent heights in MIPAS

Prepared by M.Ridolfi and B.Carli
22 July 1999

1. Introduction

The ORM code uses engineering pointing information in p,T retrieval. This information is included in the retrieval using the optimal estimation method, therefore also a variance-covariance matrix (VCM) is needed for a proper weighting of engineering data by the inversion algorithm. Since MIPAS pointing system is presently characterized only by very general specifications, some assumptions must be made and an algorithm must be set up to build a realistic VCM of the engineering tangent altitudes, starting from the specified performances. Based on the results of this algorithm, a suitable strategy must be studied for an efficient storage of MIPAS pointings VCMs in Level 2 framework.

In the present memorandum we elaborate an algorithm for deriving a realistic VCM of pointings to be used by p,T retrieval and subsequently propose a strategy for storing these VCMs in Level 2 framework.

2. MIPAS pointing performance

The VCM of pointings is built on the basis of some pieces of information provided by British Aerospace (BAe) which is responsible for the platform and for compiling the pointing budgets. BAe reports MIPAS pointing stability for 4.0 and 75 s time intervals for the three satellite axes (*x*-axis being the most critical for MIPAS pointing accuracy). *x*-axis stability, in terms of tangent altitude, is:

- 230 m for 4 s stability
- 660 m for 75 s stability

BAe provides also the total pointing accuracy:

- 2000 m is the total accuracy

The reported values have a confidence level of 95.4%, meaning that the above values are not exceeded in 95.4% of the cases. The errors are not purely statistical because they include e.g. linear drifts due to temporary unavailability of the stars used by the satellite star sensors. However, in order to exploit the formalism of the statistics, we will consider these errors as <u>gaussian</u> with standard deviation equal to **half** of the above reported figures (Note: we are assuming that the stability provided by BAe is an excursion from an average value, a quarter should be used if the provided value is a peak-to-peak excursion).

BAe tells also that for time intervals between 4 and 75 s no analyses have been made, however in these cases, the best approximation is to linearly interpolate between the above reported figures. This approximation will not be exploited in the proposed algorithm because it does not provide realistic stability figures for time intervals much less than 4s and much greater than 75s. In Sect.3 a more sophisticated interpolation scheme is proposed.

Another assumption we will use in the following is about the speed of MIPAS interferometer. We assume that MIPAS will be always operated at a 5 cm/s speed independently of the adopted spectral resolution. Furthermore we will assume the 'turn-around' time, i.e. the time required for speed inversion and positioning of the limb-scanning mirror, to be equal to 0.45 s. Scans with altitude step greater than 10 km characterized by a turn-around time greater than 0.45 s will not be considered here.

In this hypothesis the time Δt required for measuring a sweep with resolution identified by MPD is given by:

- $\Delta t(s) = \frac{MPD(cm)}{5cm/s} + 0.45s$

3. Algorithm

From the above figures, the total pointing error ($\sigma_{tot} = 1000$ m) can be intended as absolute error of the individual tangent heights, while the stability specifications can be exploited (as it will be explained) to derive the correlations between tangent heights.
Let's calculate explicitly the correlation $c_{i,k}$ between two generic tangent heights z_i and z_k, assuming that they have been measured at times t_i and t_k. The general expression of the correlation provides:

$$c_{i,k} = \frac{\sum_{j=1}^{N}(z_i(j) - \bar{z}_i)(z_k(j) - \bar{z}_k)}{\left[\left(\sum_{j=1}^{N}(z_i(j) - \bar{z}_i)^2\right) \cdot \left(\sum_{j=1}^{N}(z_k(j) - \bar{z}_k)^2\right)\right]^{\frac{1}{2}}} \tag{1}$$

where $\bar{z}_i$ is given by:

$$\bar{z}_i = \sum_{j=1}^{N} \frac{z_i(j)}{N} \tag{2}$$

and the index j ranges over an hypothetical set of N measurements of the tangent heights z_i and z_k with i,k =1, 2, …, N_{LS} (N_{LS} = number of sweeps in the considered limb-scanning (LS) sequence).
Let's indicate:

$$\varepsilon_i(j) = (z_i(j) - \bar{z}_i) \tag{3}$$

$\varepsilon_i(j)$ is the error on $z_i(j)$ in the sense that it is the deviation of $z_i(j)$ from its 'true' value which is represented by the average of equation (2).
If the two tangent heights z_i and z_k have been measured at times t_i and t_k such that $\Delta t = |t_i - t_k|$, their errors cannot differ too much due to the stability specifications of the pointing. In particular we will have:

$$\varepsilon_k(j) = \varepsilon_i(j) + \delta_{\Delta t}(j) \tag{4}$$

where $\delta_{\Delta t}(j)$ is a random term with standard deviation $\sigma_{\Delta t}$. In order to calculate $\sigma_{\Delta t}$ from the specified short- and long- term stability we will use the following function:

$$\sigma_{\Delta t} = \sigma_{tot} \cdot \left(1 - \exp\left(-\alpha \cdot \Delta t^{\beta}\right)\right) \tag{5}$$

with α and β constants determined imposing $\sigma_{\Delta t=4s} = \sigma_{4s}$ and $\sigma_{\Delta t=75s} = \sigma_{75s}$, where σ_{4s} = 115 m and σ_{75s} = 330 m are the standard deviations associated respectively to the 4s and to the 75 s specified stability. Please note that, as it is logically required, expression (5) provides $\sigma_{\Delta t} \Rightarrow 0$ for $\Delta t \Rightarrow 0$ and $\sigma_{\Delta t} \Rightarrow \sigma_{tot}$ for $\Delta t \Rightarrow \infty$. The behavior of $\sigma_{\Delta t}$ as a function of Δt is plotted in Fig.1.
The standard deviation of ε_k (equation (3)) can be expressed as:

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N}\varepsilon_k^2(j) = \frac{1}{N}\sum_{j=1}^{N}\left[\varepsilon_i(j)+\delta_{\Delta t}(j)\right]^2 = \\
&\frac{1}{N}\sum_{j=1}^{N}\varepsilon_i^2(j) + \frac{2}{N}\sum_{j=1}^{N}\varepsilon_i(j)\delta_{\Delta t}(j) + \frac{1}{N}\sum_{j=1}^{N}\delta_{\Delta t}^2(j) = \qquad (6)\\
&\sigma_{tot}^2 + \frac{2}{N}\sum_{j=1}^{N}\varepsilon_i(j)\delta_{\Delta t}(j) + \frac{1}{N}\sum_{j=1}^{N}\delta_{\Delta t}^2(j)
\end{aligned}$$

Now, since ε_k must have standard deviation equal to σ_{tot}, from equation (6) we get:

$$\sum_{j=1}^{N}\varepsilon_i(j)\delta_{\Delta t}(j) = -\frac{1}{2}\sum_{j=1}^{N}\delta_{\Delta t}^2(j) \qquad (7)$$

Substituting expressions (3) and (7) in (1) we obtain:

$$c_{i,k} = \frac{\sum_{j=1}^{N}\varepsilon_i(j)\left[\varepsilon_i(j)+\delta_{\Delta t}(j)\right]}{N\sigma_{tot}^2} = 1 - \frac{\sum_{j=1}^{N}\delta_{\Delta t}^2(j)}{2N\sigma_{tot}^2} \equiv 1 - \frac{\overline{\delta_{\Delta t}^2}}{2\sigma_{tot}^2} \qquad (8)$$

Where in the second step we have used:

$$\sigma_{tot}^2 = \frac{\sum_{j=1}^{N}\varepsilon_i^2(j)}{N} \qquad (9)$$

Considering that:

$$\overline{\delta_{\Delta t}^2} = \sigma_{\Delta t}^2 + \bar{\delta}_{\Delta t}^2 \qquad (10)$$

and that, from heuristic considerations, it should be:

$$\bar{\delta}^2 = 2\sigma_{tot}^2 - \sigma_{\Delta t}^2 - 2\sigma_{tot}\sqrt{\sigma_{tot}^2 - \sigma_{\Delta t}^2} \qquad (11)$$

equation (8) becomes:

$$c_{i,k} = \sqrt{1 - \frac{\sigma_{\Delta t}^2}{\sigma_{tot}^2}} \tag{12}$$

From a more qualitative point of view, this very simple result can also be justified as follows. The tangent height z_k has two error components: the first component (σ_1) is linked to the measurement of the neighboring tangent height z_i, the second component $\sigma_{\Delta t}$ does not depend on previous measurements. The two components must satisfy:

$$\sqrt{\sigma_{\Delta t}^2 + \sigma_1^2} = \sigma_{tot} \tag{13}$$

since the error associated to z_i is σ_{tot}, the correlation between tangent heights z_i and z_k is by definition:

$$c_{i,k} = \frac{\sigma_1}{\sigma_{tot}} = \frac{\sqrt{\sigma_{tot}^2 - \sigma_{\Delta t}^2}}{\sigma_{tot}} = \sqrt{1 - \frac{\sigma_{\Delta t}^2}{\sigma_{tot}^2}} \tag{14}$$

where the value for σ_1 has been extracted from equation (13).
Equation (12), together with expression (5) provides the tool for calculating the correlation between two generic tangent heights z_i and z_k. This tool can be exploited for computing the VCM **V** of the tangent heights whose elements $V_{i,k}$ are given by:

$$V_{i,k} = \sigma_{tot}^2 \cdot c_{i,k} \tag{15}$$

The VCM $\mathbf{V}_d$ relating to the differences between tangent heights (whose inverse is used by the ORM) can be obtained through the transformation:

$$\mathbf{V}_d = \mathbf{J}\mathbf{V}\mathbf{J}^t \tag{16}$$

where **J** is the jacobian matrix that represents the linear transformation leading from tangent heights to differences between tangent heights. If we indicate with $\Delta z_i \equiv z_{i+1} - z_i$, the jacobian **J** contains the derivatives:

$$J_{i,k} = \frac{\partial(\Delta z_i)}{\partial z_k} = \begin{cases} -1 & \text{if } i = k \\ 1 & \text{if } i = k-1 \\ 0 & \text{in the other cases} \end{cases} \tag{17}$$

with $i = 1, \ldots, N_{LS}-1$ and $k = 1, \ldots, N_{LS}$.

4. Software tool

A very simple software tool has been implemented which computes the VCMs $\mathbf{V}$ and $\mathbf{V}_d$ of MIPAS pointing system using the explained algorithm. Besides the parameters defining pointing performances described in Sect.2, the only inputs of this tool are the max. path difference and the number of sweeps of the LS sequence for which we want to calculate the VCM of pointings. The outputs of this program are:

- ✓ correlation matrix of tangent heights
- ✓ VCM of tangent heights
- ✓ correlation matrix of differences between tangent heights
- ✓ errors on differences between tangent heights
- ✓ VCM of differences between tangent heights
- ✓ inverse of VCM of differences between tangent heights

5. Results

In Fig.2 we report the correlations between different tangent heights for a scan of 16 sweeps and MPD = 20 cm as a function of the sweep index. In Fig.3 we report correlations of differences between tangent heights for the same scan of Fig.2 as a function of the index. The same quantities are reported in Fig's 4 and 5 respectively, for a scan of 16 sweeps and MPD = 5 cm (reduced resolution).
In the adopted approach, the absolute errors of both tangent heights and differences between tangent heights are constant with altitude. The errors on differences between tangent heights depend however on the selected MPD. The dependence of these errors on the MPD is shown in Fig. 6.
General comments are:

- the absolute error on tangent heights is a constant (does not depend on MPD)
- the correlation between tangent heights increases when decreasing the resolution (i.e. decreasing MPD)
- decreasing the resolution, in consequence of the increased correlations, the errors on the differences between tangent heights decrease (see Fig.6).

6. Recommended strategy for handling pointings VCM in Level 2 framework

The quantity required in input to the ORM is the inverse of the VCM of the differences between tangent heights. Given the invariance of the obtained results with respect to the sweep index we propose the following approach for storage / handling of pointing VCM in Level 2 framework.
The file 'PI_VCM.DAT' should contain VCMs of the tangent heights tabulated as a function of max. path difference. The tabulated VCMs should refer to a scan with a maximal number $N_{\max}$ of sweeps (e.g. $N_{\max}$ = 30 sweeps). Given a scan with N_{sw} (with $N_{sw} < N_{\max}$) sweeps to be analyzed and max. path difference MPD = xx, a block matrix of dimension N_{sw} x N_{sw} will be extracted from the VCM relating to OPD = xx. Rows and columns relating to corrupted sweeps will be then removed from this block. The remaining rows and columns will be transformed according to equation (16) and the resulting matrix will be inverted and provided in input to Level 2 processor. Given the particular symmetry of the VCMs of tangents heights, optimized strategies can be eventually considered for storage of these VCMs in Level 2 framework (note for example that the entire VCM can be easily reconstructed from its first row or column).

Despite the differences existing between the present algorithm and the algorithm developed during the first MIPAS pT retrieval study (ESTEC Purchase Order No: 142956 terminated in Sept.'95) for the calculation of pointing VCMs, the results of the two algorithms are consistent in the case of MPD = 20 cm. Presently it is not possible to use any longer the old algorithm due to the fact that in the old algorithm the spectral resolution was assumed constant. The present analytical expression provides a more simple and versatile calculation tool.

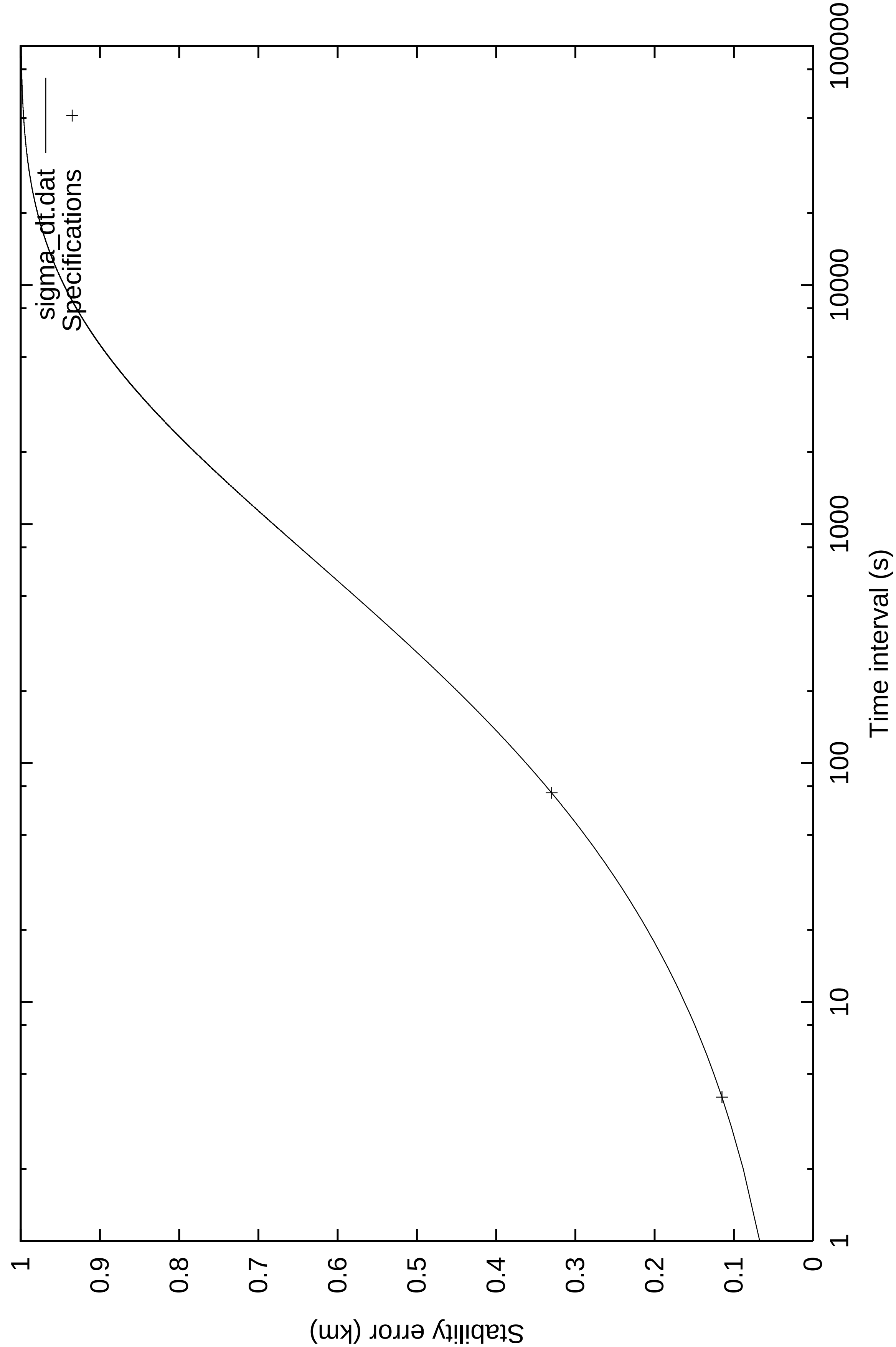
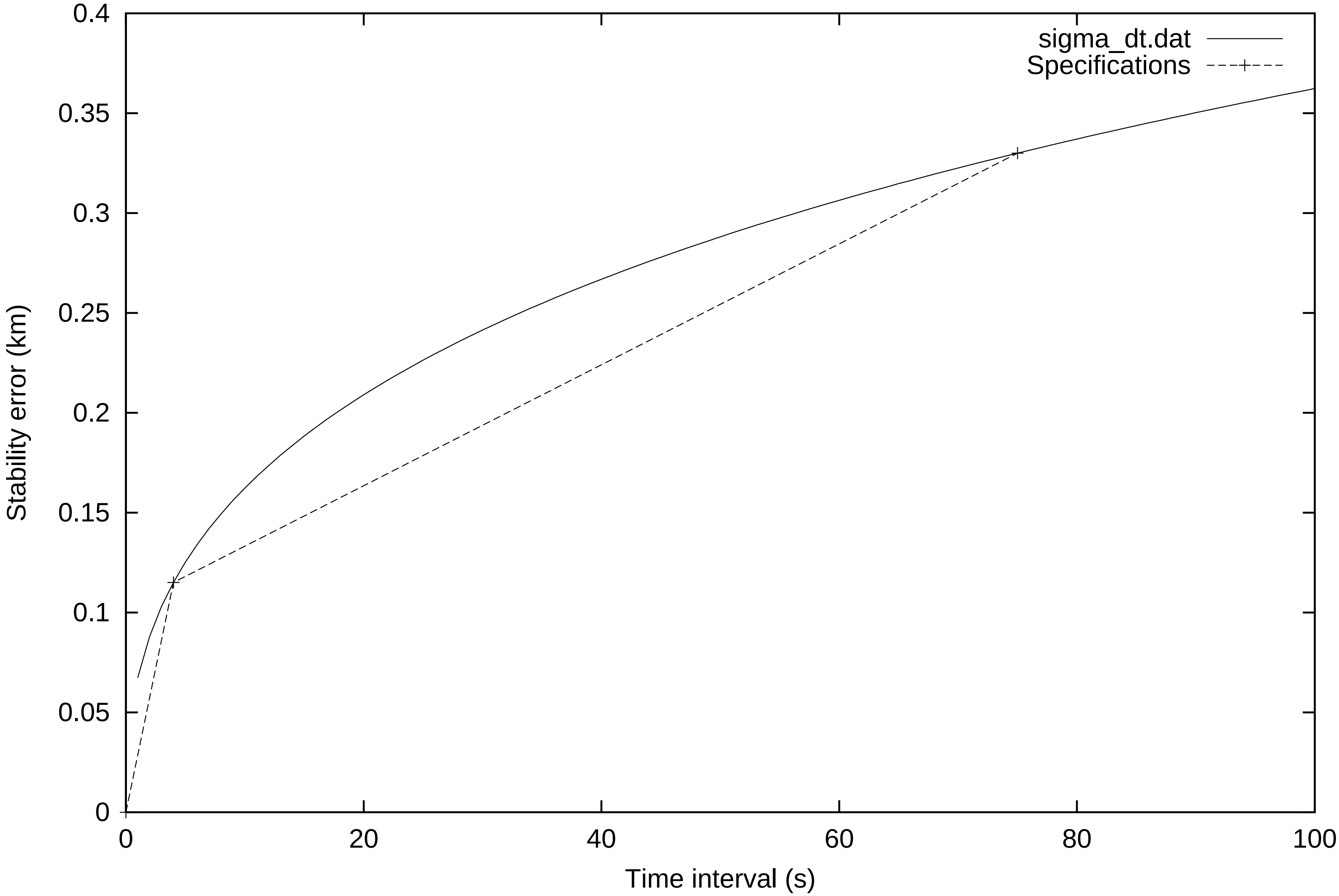


Fig.1a: Stability error as a function of time. Log-scale x

Fig.1b: Stability error as a function of time. Linear scale x



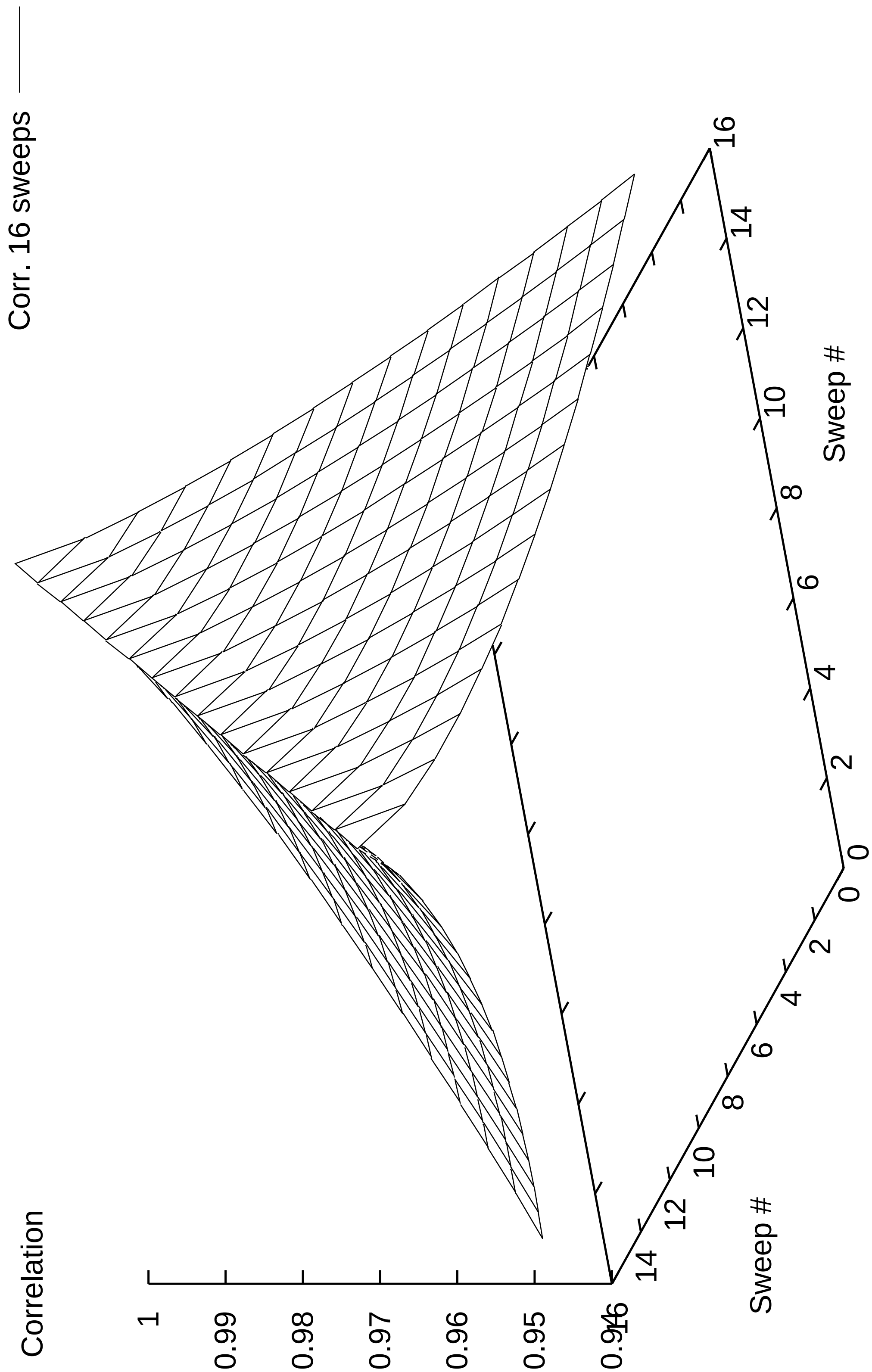


Fig.2: Correlations between Z, OPD = 20 cm, N. of sweeps = 16

Fig.3: Correlations between dZ, MPD = 20 cm, N. of sweeps = 16

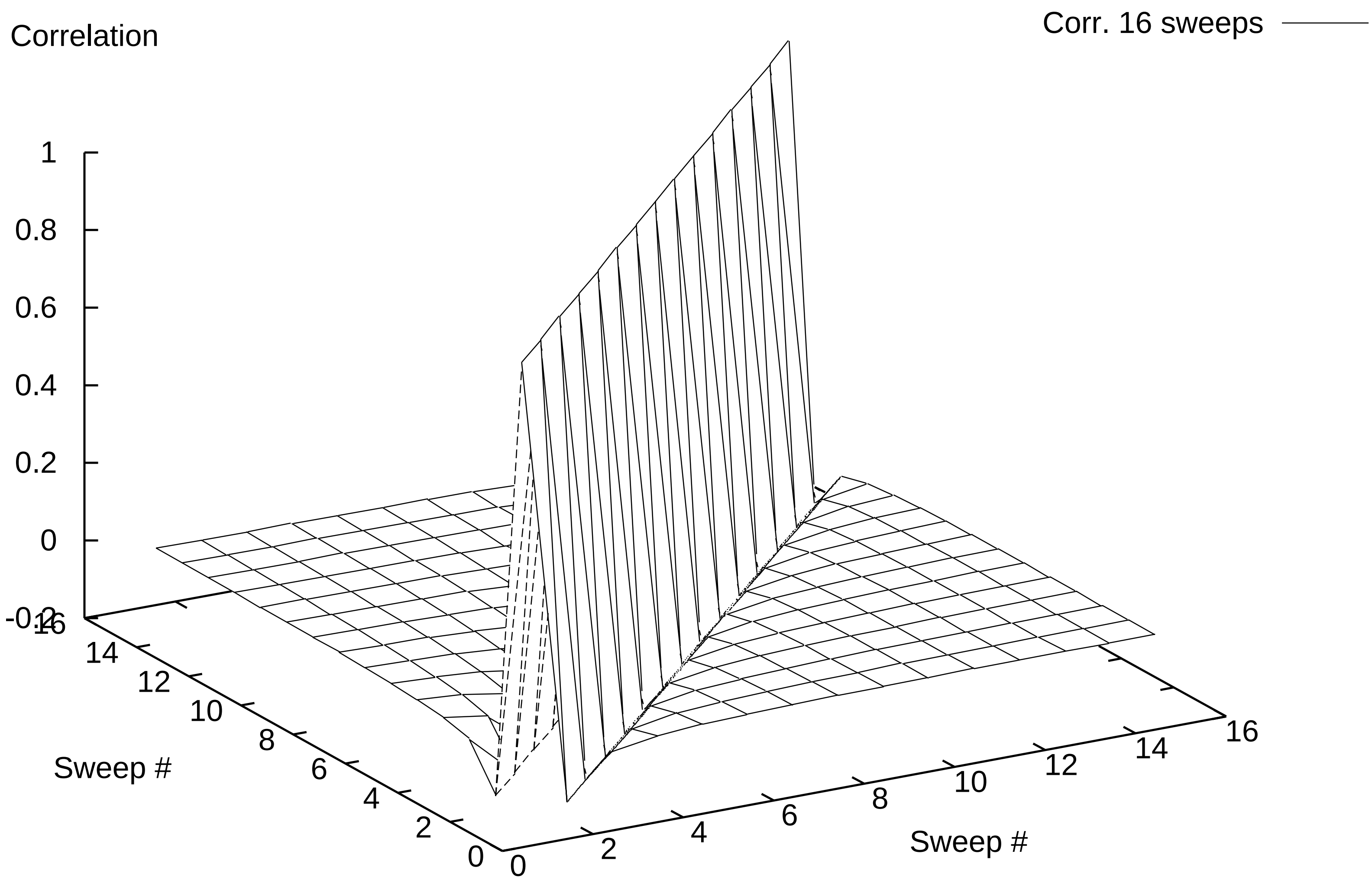


Fig.4: Correlations between Z, OPD = 5 cm, N. of sweeps = 16

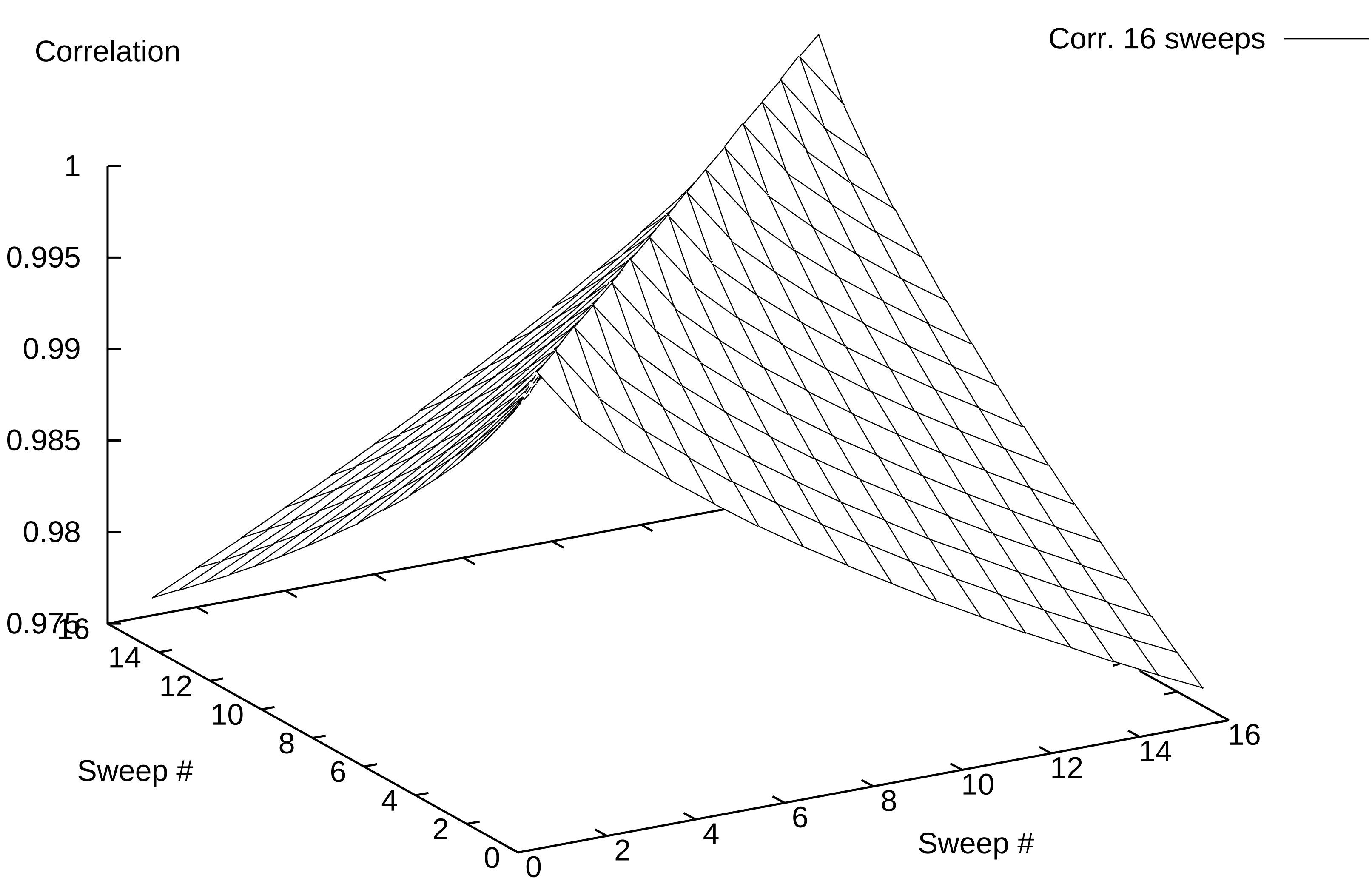


Fig.5: Correlations between dZ, MPD = 5 cm, N. of sweeps = 16

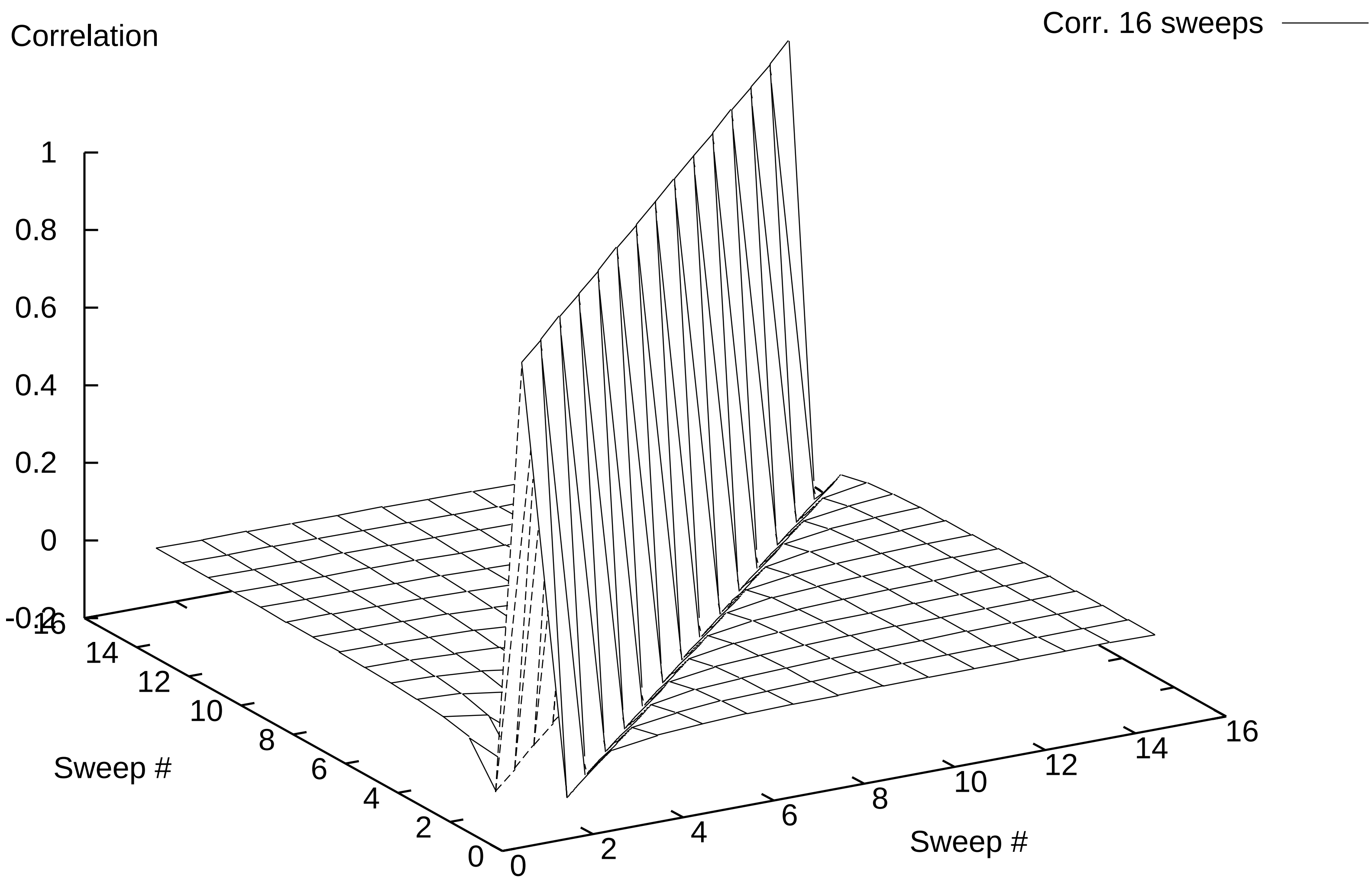


Fig.6: Errors on dZ for different MPDs. Errors do not depend on sweep.