Preliminary Diagnostic Testing of LDA Data-Sets

by

Hans R.E. van Maanen (S.E.P.T.A.R., dept. EPT-ASP)
(P.O. Box 60, 2280 AB Rijswijk, Netherlands)
(email: H.R.E.vanMaanen@siep.shell.com)

ABSTRACT

Determination of the quality of an LDA data-set is important for both the experimenter to decide on repetition of the experiment and the scientific community to know that the results are valid within certain limitations. Objective quality determination is also of importance for optimisation of LDA systems and for the novice user to get feedback on the quality of his/her experimental data.

The time interval distribution shows to be very useful for this purpose. A number of distinct deviations from the ideal exponential distribution occur which are indicative for incorrect settings of the burst validation criteria, dead time of the processor, velocity or particle rate bias, round-off errors of the clock and/or the data-file, arrival time noise and multiple validation. These aspects are of importance for different error sources in LDA. Multiple validation, which means that more than one velocity estimator is derived from a single Doppler signal, showed to be detrimental to the results and should therefore be avoided (Van Maanen (1999)). The time interval distribution is very effective for this purpose.

The raw velocity data (as a function of time) should behave in agreement with the flow under study. E.g. large excursions, which occur during a short period of time, are suspect. Especially in combination with the higher moments of the velocity probability distribution and the logarithmic velocity probability distribution, incorrect behaviour of the system can be revealed.

The auto correlation function, estimated using the slotting technique with local normalisation (Van Maanen (1999)), shows to be sensitive to multiple validation. The number of products in each slot is sensitive to velocity bias as well as round-off errors. Both properties are very useful for the determination of the quality of data-sets.

Although the diagnostic tools do not provide the final answer on the quality of the data-sets, they are very helpful in this respect. They can be used for intercomparison, but also for the training of novel users. As LDA still is a complicated technique, this is an important application. We have found it a very effective technique to teach novices in the field. One could also think that the results of the diagnostic testing should be made available for peer reviewers, so that they can be convinced of the quality of the underlying data, on which the conclusions are based.

1. Introduction.

Taking LDA measurements is a time consuming effort and high quality data are essential to obtain reliable conclusions, is important for optimisation of an LDA system and to make the decision to rerun the experiment or not. To determine the quality of the data, several tools have been developed, which will be discussed in the following sections.

2. The Time Interval Distribution.

The choices made to present time interval distributions (TID's) are to optimise these for diagnostics within the limits of variance, bias and resolution. The range is roughly $4t_0$ - $5t_0$, split into 100 bins, resulting in a resolution of $\approx 0.05t_0$. The vertical scale is logarithmic: the TID's should then be linear and deviations can easily be distinguished. To simplify the comparison between experiments with different data-rates, the TID's are "normalised" by the maximum value. The t_0 is estimated from a straight line, fitted through the TID, which should go through (0,1) in the ideal case (see fig. 1). More details can be found in Van Maanen (1999).

The homogeneous, but random, distribution of the tracer particles in space leads to the exponential TID between two successive Doppler signals (e.g. Erdmann and Gellert (1976)) and it can easily be confirmed by a Monte-Carlo simulation (fig. 1). Practical TID's usually show deviations from the ideal shape, which can be used to determine the quality of the experimental data.

In fig. 2 - 8 TID's from different experiments are presented and all show deviations from the theoretical one, the causes of which will be revealed later. The most important deviations are:

1. The absence of data for the shortest time intervals.
2. The shape of the TID for short intervals is convex.
3. The TID is concave in the vicinity of t_0.
4. The TID shows a crenellate structure.
5. A small number of bins do not contain data.
6. A large number of bins do not contain data.
7. The TID shows a doublet structure with two, clearly distinct, characteristic times.

In many cases the TID's show several of these deviations simultaneously. Monte-Carlo simulations have been run to find or verify the causes of these deviations. The ones that have been identified will be discussed in the next section.

3. Causes of deviations of TID's.

Every processor has a "dead time" in which it is processing a Doppler signal and it is not able to accept newly arriving ones. Including this into the TID removes the observations for the shortest time intervals (see fig. 9), but the boundary is sharp, whereas in practice it is more gradual. The dead time is therefore only a partial explanation. Yet it is an important parameter for the optimisation of the LDA system which can be estimated from the TID. It is important because the dead time determines the high-frequency limits of the data-set: using e.g. the slotting technique (Van Maanen and Tummers (1996) and Van Maanen (1999)): all slots below the dead time will remain empty, thus enforcing a minimum slot width and thus an upper frequency and resolution in the spectrum. Note that the dead time is also related to the transit time of the tracer particles.

Computers do their calculations and (mass) storage with a limited accuracy. This could lead to problems: the total duration of an experiment will mostly be long compared to the t_0. When the *arrival time* is stored in the data-file, the resolution can become of the same order of magnitude as the shorter time intervals: with single precision, the resolution after 1000 seconds is 100 μsec, which is 10% of the t_0 of fig. 2. This TID, however, has a much higher resolution. The round-off errors start to interfere with the bin-quantisation of the TID and are thus noticeable. Because the round-off errors increase with increasing time, from a certain moment on a number of bins will no longer collect data, which causes the crenellate structure. This is confirmed by simulation and the results are shown in fig. 10. Thus recording the *arrival time* can lead to an additional, but unnecessary, noise contribution to the velocity estimators due to arrival time error.

The limited resolution of the internal clock of a processor has a similar effect and it should be chosen sufficiently high to avoid unnecessary errors. This effect could be modelled very well, as is shown in fig. 11. Less extreme cases may go unnoticed, like the case in fig. 12, but it still occurs. The choices made for the presentation of the TID enable the detection of round-off errors of roughly $0.01t_0$ and larger, which should be sufficient for most cases.

The simulated TID of fig. 12 still shows some differences with that of fig. 6, notably the difference between the completely empty bins of the simulation and the partially filled bins of the experimental TID. The cause of this difference is the uncertainty in the individual arrival time estimators: one should define an "error band" for the velocity estimator, but also for the *arrival time estimator*, as has been discussed in Van Maanen and Nijenboer (1996) and in Van Maanen (1999). It is a similar source of noise contribution as the round-off errors. Introducing a uncertainty in the arrival time results in the TID of fig. 13, which resembles the structure of the TID of fig. 6. It also explains the convex structure of the TID at short time intervals, as is shown in fig. 14. This causes the fitted line to intersect the vertical axis above 1 (one) and can thus be used as a measure of this error source. Arrival time error can only be suppressed by using Doppler signals which are strong and have a high Signal-to-Noise ratio (SNR) in combination with a small measurement volume.

Velocity bias is a much debated phenomenon in LDA (see e.g. Absil (1995)) and is caused by a correlation between the (measured) velocity component and the data-rate. In a simulation a 1-D flow system has been analyzed by adding an average velocity to simulated turbulence (Van Maanen and Tummers (1996) and Van Maanen (1999)) and to take the data-rate proportional to the instantaneous velocity. The results for a turbulence intensity of 30% are presented in fig. 15 and for a turbulence intensity of 100% in fig. 16. These TID's show a concave shape, which increases with increasing turbulence intensity. This concave shape can be seen in several of the TID's, presented in fig. 2 - 8.

Another cause of bias in LDA is the confluence of fluids with different tracer particle concentrations or changes in the tracer particle concentration by e.g. centrifugal effects. In the simulation, the fluids had a concentration difference of a factor 4 (fig. 17) and a factor of 2 (fig. 18). Each fluid spent 50 % of the time in the measurement volume under the assumption that these times are far longer than the t_0 of the TID. The results show that this kind of bias introduces similar deviations as velocity bias, but to a smaller extent. It is therefore only useful for extreme cases.

Narrow band noise leads to an amplitude modulation of the Doppler signal, which can give rise to "multiple validation" of the Doppler signal: the detection and validation system concluded that the signal consisted of the Doppler signals of at least two tracer particles (Van Maanen (1999)). When this occurs, a large number of observations with short interval times (with a maximum of the transit time of the tracer particles!) are detected and the TID's show a kind of double structure in the decay, one t_0 corresponding to the transit time and the other to the real interval time, as is clear in fig. 19 (compare with fig. 8). Multiple validation leads to an intersection of the fitted line with the vertical axis below 1 (one), which can be used as a measure of this effect. As this drop is counteracted by e.g. the dead time, this usually points at serious problems. Multiple validation leads to deteriorating effects on the velocity estimators as is illustrated in fig. 20 and 21: fig. 20 shows the velocity without multiple validation and in fig. 21 with it. It happens with all processors (see fig. 22 and 23) and should simply be avoided.

Avoiding multiple validation often means an increase of the dead time of the processor, which, however, never needs to be longer than the transit time of the tracer particles. Thus by choosing a small measurement volume, it can be reduced as much as possible. A small measurement volume also increases the spatial resolution, it increases the light intensity and it allows higher concentrations of tracer particles. But the transit time can be quite variable in highly turbulent flows, which complicates the correct settings for the validation system: when fast signals are accepted, the chance for multiple validation at low velocities increases, whereas suppressing the fast signals may result in the loss of Doppler signals from the high velocities. Dynamic validation criteria, which could adapt to the "instantaneous" flow conditions would be ideal. The TID is a tool that can be used to see the influence of the settings on the results and can thus be helpful in finding the optimum settings.

4. The actual velocity trace.

The velocity trace itself can indicate the quality of the data: the excursions, shown in fig. 21 and 24, are incorrect. The velocity trace can also indicate the amount of noise, present in the velocity signal. However, it is only a qualitative judgement that can be made and it can give no guarantee that the experiment is running optimally.

5. The Velocity Probability Distribution.

The velocity probability distribution of most turbulent flows usually has a more or less Gaussian shape, shifted over the average velocity of the component being measured (Bradshaw (1971), Hinze (1975), Tennekens and Lumley (1983)). An example is shown in fig. 25. Skewed probability distributions are dominant and in oscillating flows bimodal distributions occur. Deviations from what can be expected may indicate errors. The distribution should be a continuous function and not have a "comb-like" structure as is shown in fig. 26, which points at quantisation errors. Fig. 27 shows a distribution which contains a small number of observations at elevated velocities, which are hard to distinguish on a *linear* scale. Therefore it is attractive to use a *logarithmic* scale, as in fig. 28. The small contributions at the higher velocities can now be distinguished and those above 0.2 m/s are obviously suspect. Calculation of the higher moments of the probability distribution can reveal whether this suspicion is correct or not.

6. The Auto Correlation Function estimated using the Local Normalisation.

The Local Normalisation has shown to be a good estimator of the auto correlation function (ACF) (Van Maanen and Tummers (1996) and Van Maanen (1999)), *provided the raw data are reliable*. Therefore it can be a quality tester and it showed to be very sensitive for multiple validation: the estimated ACF becomes uninterpretable. Fig. 29 shows an ACF like one might expect with a limited variance near $\tau = 0$ and a clear noise peak at the origin. In contrast, fig. 30 shows one which has been effected by multiple validation. The large variance for all values of τ indicates the problems.

7. The number of products in the slotted ACF.

The number of products in each slot of the ACF is -on average- equal to (Van Maanen (1999)):

$$N_p = N_t \bullet \left(\frac{\Delta t}{t_0} \right) = T_m \bullet \left(\frac{\Delta t}{t_0^2} \right)$$

in which:
N_p = Average number of products in each slot
N_t = Total number of observations.
Δt = Slot width
t_0 = Characteristic time of TID.
T_m = Total measurement time.
Comparison of this number with the actual number of products in each slot can reveal problems, as is illustrated in fig. 31. The low number of products in the first slots indicate a low number of observations with short time intervals (see fig. 32) and the comb-like structure points at round-off errors in the data-file. Note that the latter is not so obvious from the TID, shown in fig. 32.

8. Conclusions.

Unfortunately, diagnostics has not received the attention with the development of the LDA measurement technique that it should have had. It is -in my view- very important that shortly after the experiment it should become clear whether the data are useful or not. The analyses that have been described above can help in that respect, without pretending that these will give the final answer. In the future additional diagnostic tools may be developed, which can increase the probability of detecting data which are not OK. Such tools should be used to eliminate data-sets on which incorrect conclusions might be based. These tools are also valuable for the training of novel users, because they give feedback to the experimenter. This training task is often neglected but in our Laboratory it proved to be very useful for novice users in the field. One could also provide the results of the diagnostic analysis to the peer reviewers, to convince them of the quality of the underlying data-sets on which the conclusions are based and thus reduce the chances that erroneous results are reported in literature.

References.

L.H.J. Absil, "Analysis of the Laser Doppler Measurement Technique for Applications in Turbulent Flows", Ph.D.Thesis, Delft University of Technology (Netherlands), 1995.

P. Bradshaw, "An Introduction to Turbulence and its Measurements", Pergamon Press, London (1971).

J.C. Erdmann and R.I. Gellert, "Particle Arrival Statistics in Laser Anemometry of Turbulent flow", Applied Physics Letters, Vol. 29 (1976), pp. 408 - 411.

J.O. Hinze, "Turbulence", Mc. Graw-Hill, New York, 2nd ed. (1975).

H.R.E. van Maanen and M.J. Tummers, "Estimation of the Auto Correlation Function of Turbulent Velocity Fluctuations using the Slotting Technique with Local Normalisation", Proc. of the 8th Int. Symposium on Applications of Laser Techniques to Fluid Mechanics, Lisbon (Portugal), 8 - 11 July 1996, paper no. 36.4.

H.R.E. van Maanen and F.J. Nijenboer, "Application of the Wavelet Transform to Laser-Doppler Signal Processors", Proc. of the 8th Int. Symposium on Applications of Laser Techniques to Fluid Mechanics, Lisbon (Portugal), July 8 - 11 1996, paper no. 31.4.

H.R.E. van Maanen, "Retrieval of Turbulence and Turbulence Properties from Randomly Sampled Laser-Doppler Anemometry Data With Noise", Ph.D. Thesis, Delft University of Technology (Netherlands), 1999.

H. Tennekes and J.L. Lumley, "A First Course in Turbulence", The MIT Press, Cambridge (1983).

H.R.E. van Maanen, "Retrieval of Turbulence and Turbulence Properties from Randomly Sampled Laser-Doppler Anemometry Data With Noise", Ph.D. Thesis, Delft University of Technology (Netherlands), 1999.

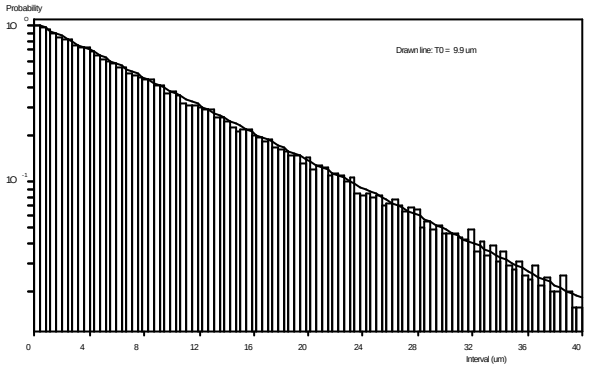

Figure 1: *Histogram of the positional distribution of homogeneously, but randomly, distributed point-like particles*.

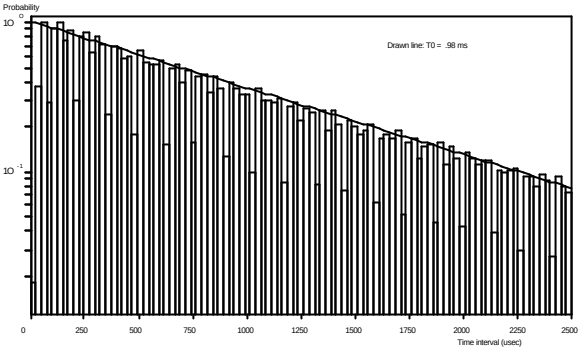

Figure 2: *Time interval distribution with an average data-rate of approximately 1 kHz.*

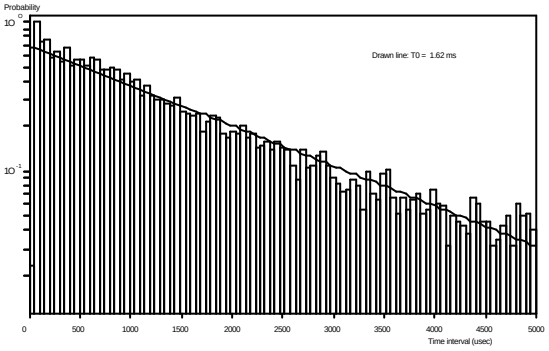

Figure 3: *Deviations: elevated below 150 msec, reduced between 150 and 500 msec, reduced around 2500 msec and elevated at 5000 msec.*

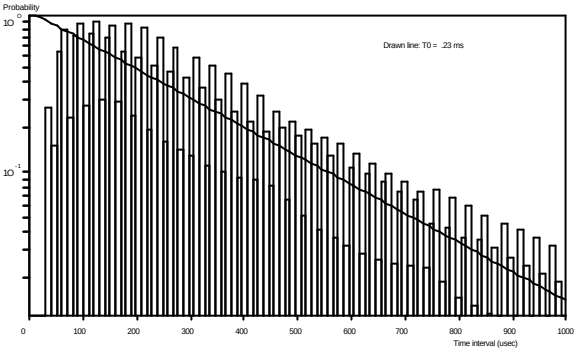

Figure 4: *Deviations: no observations below 30 msec, convex up to 200 msec, concave shape and crenel structure.*

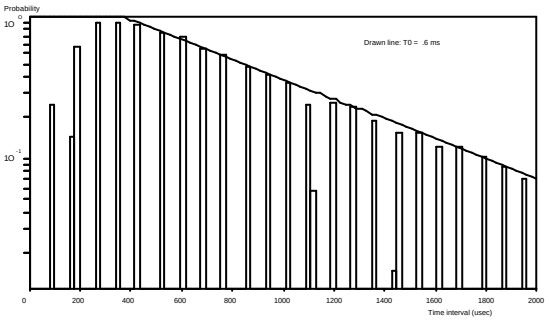

Figure 5: *Deviations: convex up to 400 msec., a large number of bins do not contain observations.*

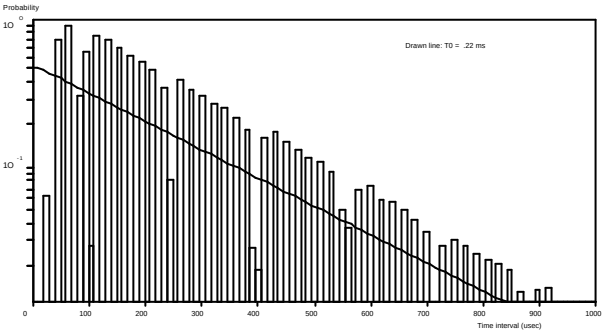

Figure 6: *Deviations: no observations below 30 msec., convex up to 70 msec. and a number of bins does not contain observations.*

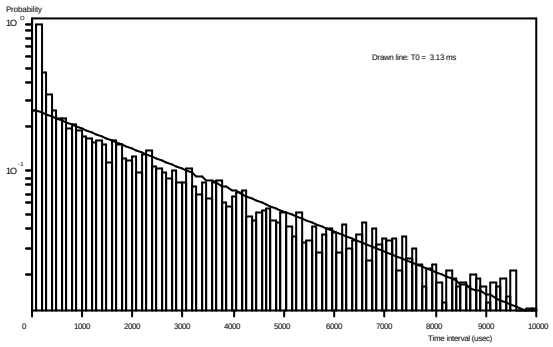

Figure 7: *Deviation: double t_0 structure.*

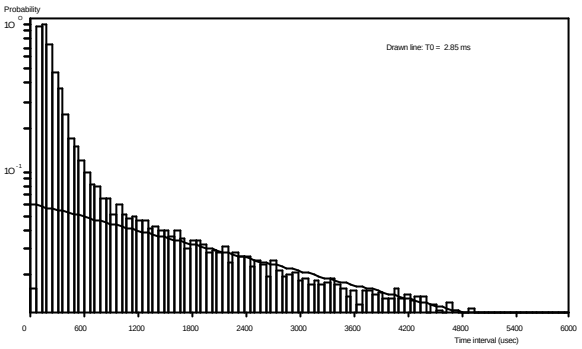

Figure 8: *Deviation: double t_0 structure.*

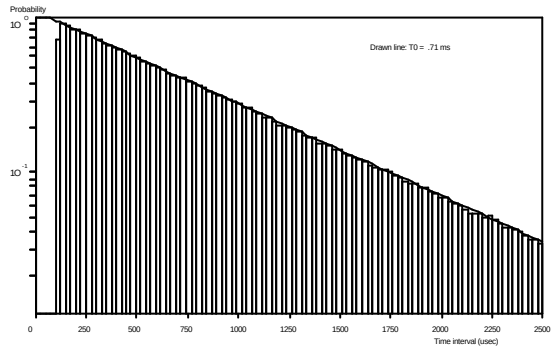

Figure 9: *A dead time explains the lack of observations for short time intervals. Note, however, the sharp boundary.*

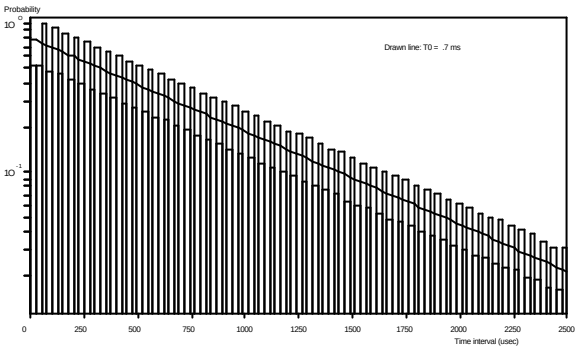

Figure 10: *When the accuracy to record the arrival time is too low, round-off errors occur, increasing with time.*

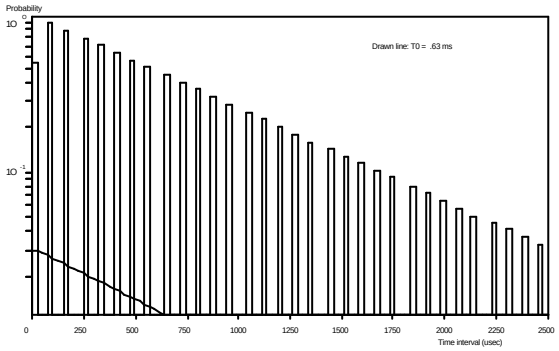

Figure 11: *Round-off errors of the internal clock of the processor lead to a distribution with empty bins.*

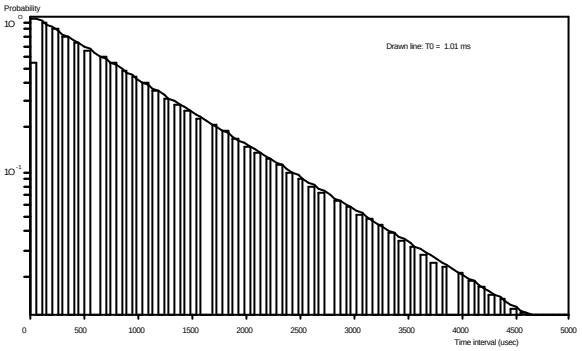

Figure 12: *The round-off errors of the clock do not explain the measured deviations completely.*

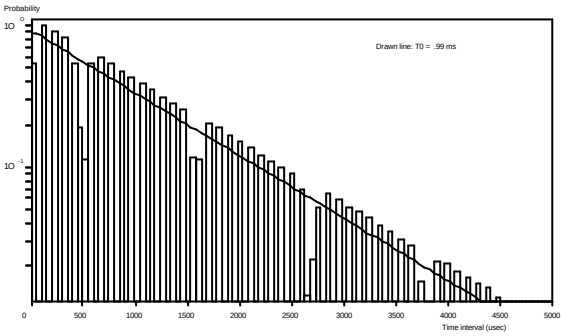

Figure 13: *Uncertainty in the arrival time and the round-off errors of the clock combined create the observed deviations.*

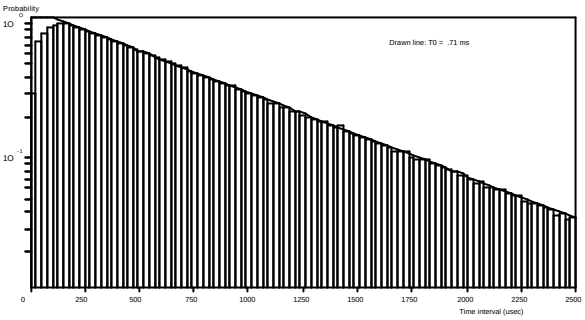

Figure 14: *Uncertainties in the arrival time estimators give rise to a convex shape at short time intervals.*

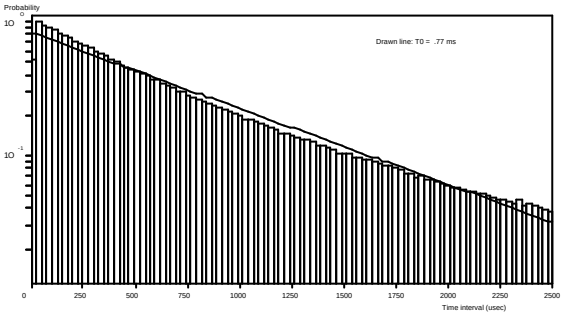

Figure 15: *A turbulence intensity of 30 % and 1-D velocity bias create a concave shape of the time interval distribution.*

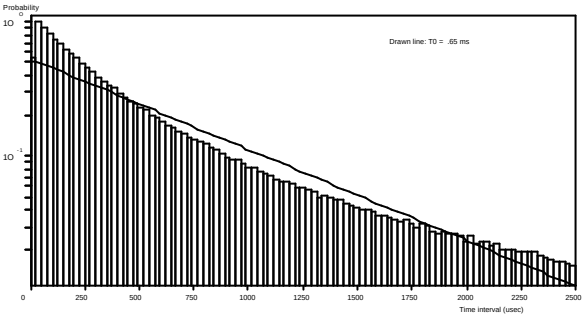

Figure 16: *A turbulence intensity of 100% and up creates a strong concave shape of the time interval distribution.*

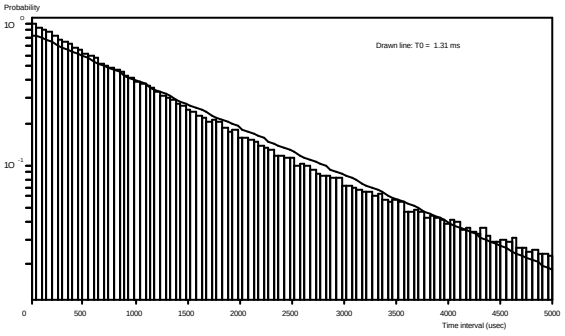

Figure 17: *The TID of mixing fluids. The tracer particle concentration ratio is 4, equal amounts of the fluids.*

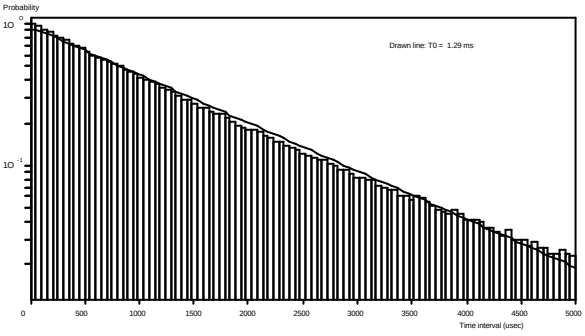

Figure 18: *The TID of mixing fluids. The tracer particle concentration ratio is 2, equal amounts of the fluids.*

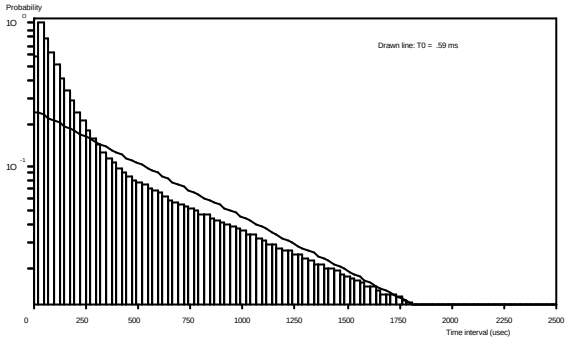

Figure 19: *Multiple validation leads to a TID with a double t_0 structure: the transit time and the interval time.*

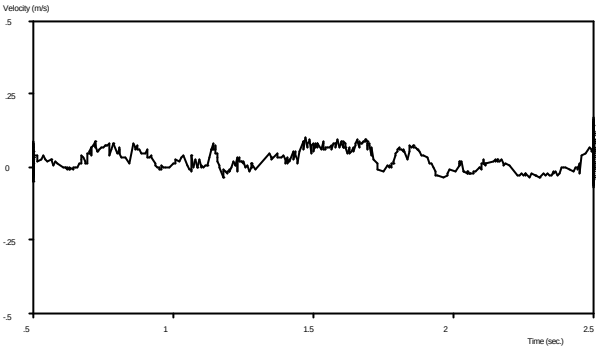

Figure 20: *Velocity fluctuations with correct settings so no multiple validation of the same Doppler signal can occur.*

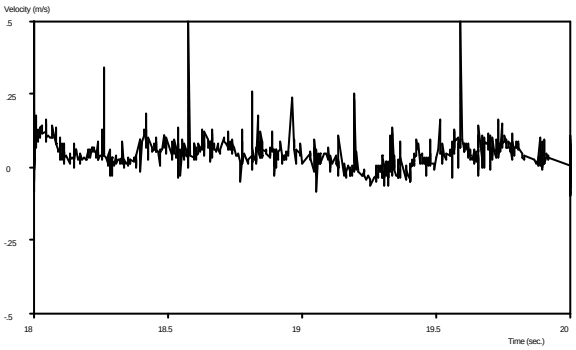

Figure 21: *Velocity fluctuations with incorrect settings so that multiple validation of the same Doppler signal can occur.*

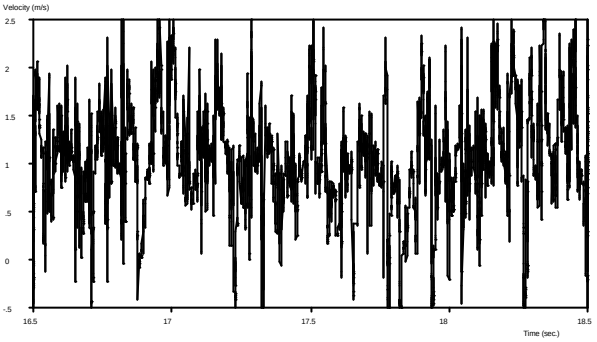

Figure 22: *Other processors show excursions too when multiple validations occur.*

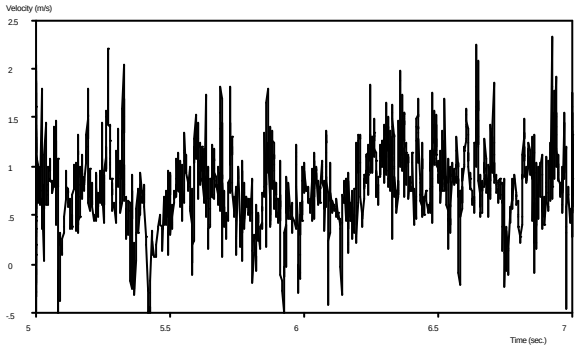

Figure 23: *The behaviour improves clearly when the settings are adjusted in so that multiple validation cannot occur.*

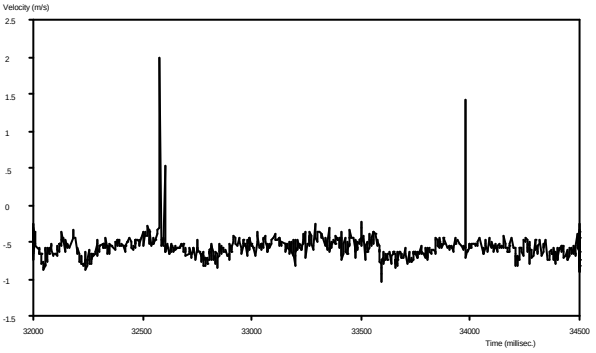

Figure 24: *Example of the raw velocity trace, which contains a few excursions, which indicate erroneous behaviour.*

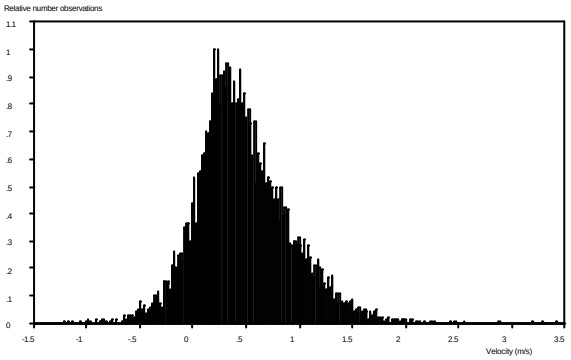

Figure 25: *An example of a velocity probability distribution.*

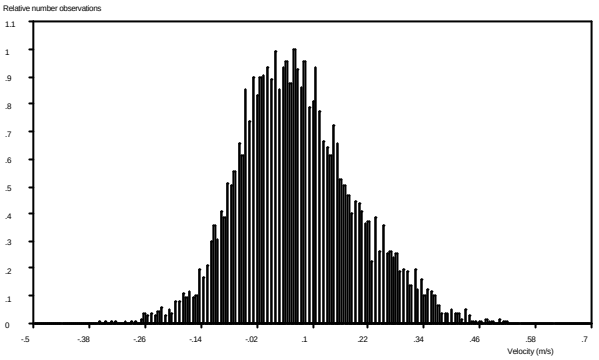

Figure 26: *An example of a velocity probability distribution with a "comb-like" structure.*

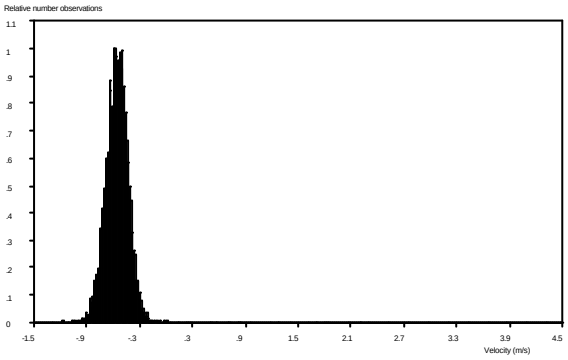

Figure 27: *An example of a velocity probability distribution with a small number of observations at elevated velocities.*

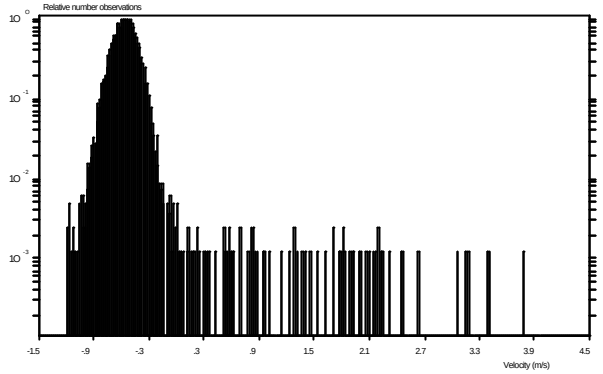

Figure 28: *The same velocity probability distribution as of fig. 27, but with a logarithmic scale on the vertical axis.*

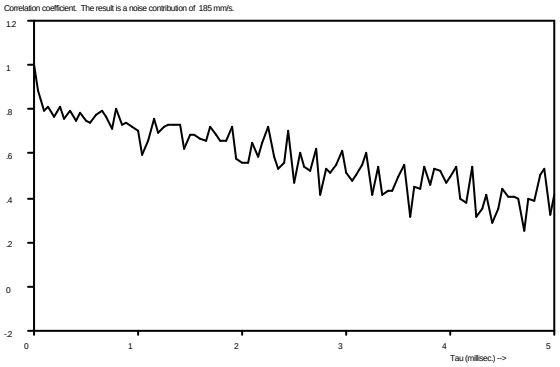

Figure 29: *The auto correlation function with local normalisation when the data are OK.*

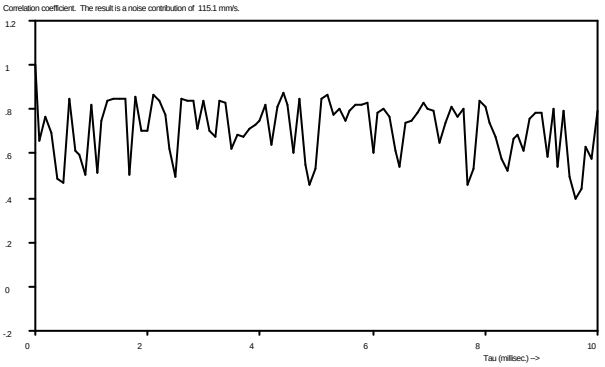

Figure 30: *The auto correlation function when multiple validation occurs. Note the high variance for all values of t.*

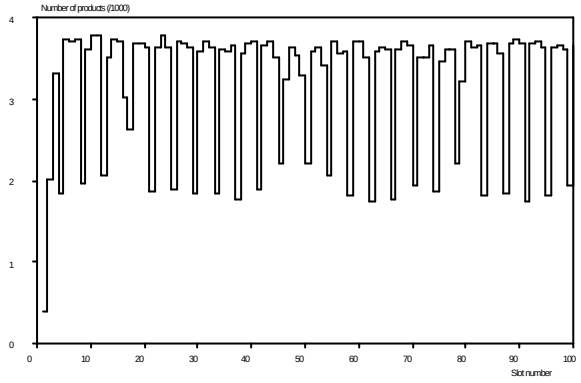

Figure 31: *Deviations of the number of products from the average can be used for diagnostics.*

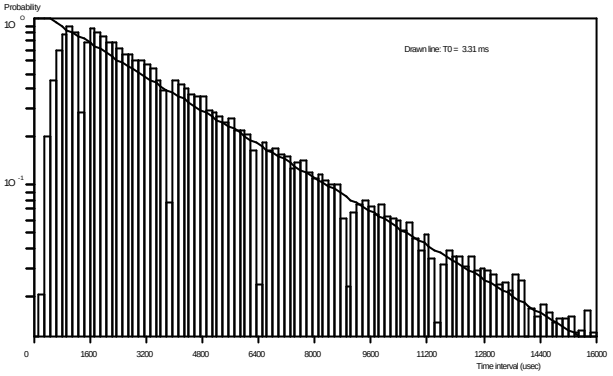

Figure 32: *TID with a lower number of observations for the short time intervals and round-off errors in the data-file.*