LDA Benchmark Generator II

Holger Nobach

November 23, 2000

Abstract

A program is described, which generates simulated, irregularly sampled LDA data sets suitable for testing processing algorithms or for comparing various estimation procedures of statistical quantities. It builds on a previous version designed to help evaluate spectral estimation algorithms and has been improved. The program is available on the internet at the address `http://www.nambis.de/benchmark`.

1 Introduction

The test of data processing algorithms for laser Doppler anemometry (LDA) requires data sets with known statistical characteristics. Except for some trivial cases, this is seldom possible via experiments and thus simulation routines can be particularly valuable. However the correct simulation of LDA data is not trivial due to many subtle physical effects relating to both the optical sampling of the flow velocity and the particular signal processing employed [1]. The experience of the last 8 years and especially the results of recent benchmark tests for the estimation of power spectra from LDA data [2] has led to a revised generation program with several new features, like different effects of the velocity bias, processor delay, arrival time uncertainty and periodic components in the velocity fluctuations. In its present form the program has been released to generate 66 different groups of LDA data sets with pre-defined statistical characteristics, each set of any group referenced with an index number. In this way users on different machines will be able to reproduce exactly the same data set. The data sets themselves are otherwise far too large to transfer directly. Each data set comprises the arrival time and the velocity value of each sample in ASCII format. The program `bm2` is available freely via internet at `http://www.nambis.de/benchmark`.

2 LDA Data Model

The basic simulation concept can be found in [1]. The equidistant sampled primary velocity time series (figure 1) is sampled non-regularly by the simulation of particle arrivals. The final step is the simulation of the processor effects, like time delay, arrival time uncertainty and noise. The simulation routines are described in detail in the next sections.

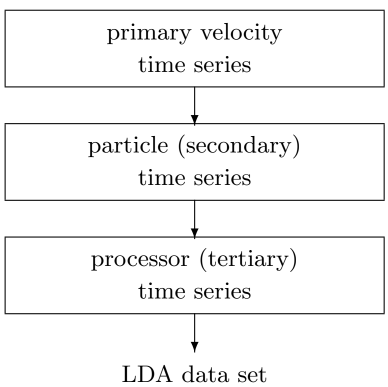

Figure 1: Signal flow of the LDA generation process.

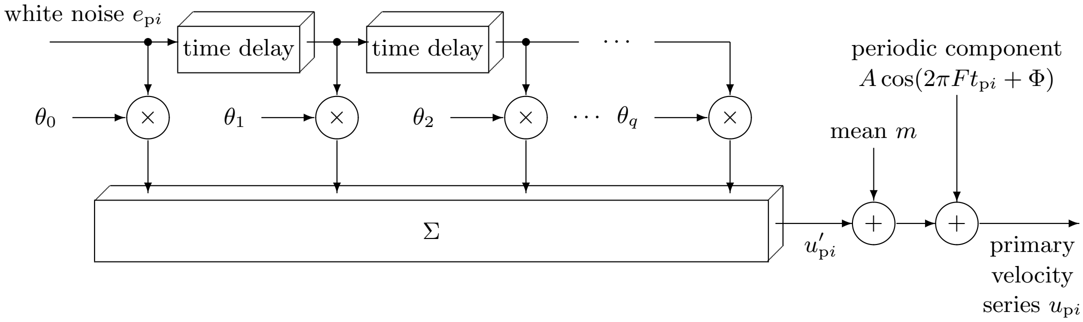

Figure 2: Generation filter for the primary velocity time series.

2.1 Primary Velocity Time Series

The primary velocity series is equidistant spaced with the sampling frequency of f_p. It is generated by an FIR filter of order q

$$u'_{\mathrm{p}i} = \sum_{j=0}^{q} \theta_j e_{\mathrm{p}(i-j)} \tag{1}$$

from an uncorrelated noise with e_p Gaussian distribution (figure 2). The filter coefficients θ_j are chosen in such a manner that the resulting series $u'_{\mathrm{p}i}$ has a given power spectral density (PSD) $S(f)$ and a given variance σ_u^2.

Three different spectral distributions are simulated, as illustrated in figure 3.

1. Band-limited random noise with Gaussian amplitude distribution and a very steep roll-off at 2 kHz.
2. A Pao-like spectrum, which decreases exponentially (in log coordinates) with increasing frequency

$$S(f) \sim \frac{e^{-1.5\alpha(f/f_d)^{4/3}}}{1 + \frac{(f/f_d)^{5/3}}{\alpha\pi}} \tag{2}$$

   where $\alpha = 0.1$ and f_d=30 Hz.
3. A spectrum with two distinct peaks at $f_1 = 10$ Hz and $f_2 = 100$ Hz

$$S(f) \sim \frac{1}{1 + (f/f_d)^4} + \frac{100}{\sqrt{2\pi}} e^{-200\,\mathrm{lg}^2(f/f_1)} + \frac{1}{\sqrt{2\pi}} e^{-200\,\mathrm{lg}^2(f/f_2)} \tag{3}$$

   where f_d=30 Hz.

In an additional step the final primary velocity series is found through

$$u_{\mathrm{p}i} = u'_{\mathrm{p}i} + m + A\cos(2\pi F t_{\mathrm{p}i} + \Phi) \tag{4}$$

with the mean velocity m and a periodic component with the amplitude A, the frequency F, the phase Φ (randomly chosen for each data set) and the sample times $t_{\mathrm{p}i}$. Examples of three different simulated time series are shown in figure 4 and a summary of the simulation parameters of the different spectral types is given in table 1.

2.2 Particle (Secondary) Time Series

The irregular sampling of the velocity time series is simulated through exponential distributed interarrival times or exponential distributed interarrival volume elements depending on whether a velocity bias is included or not. If the velocity bias is not included in the simulation the particle arrival times can be found directly through a summation of exponentially distributed interarrival times (lower line in figure 5). If the velocity bias is included in the simulation the volume passing through the measurement volume between two particle arrivals is distributed

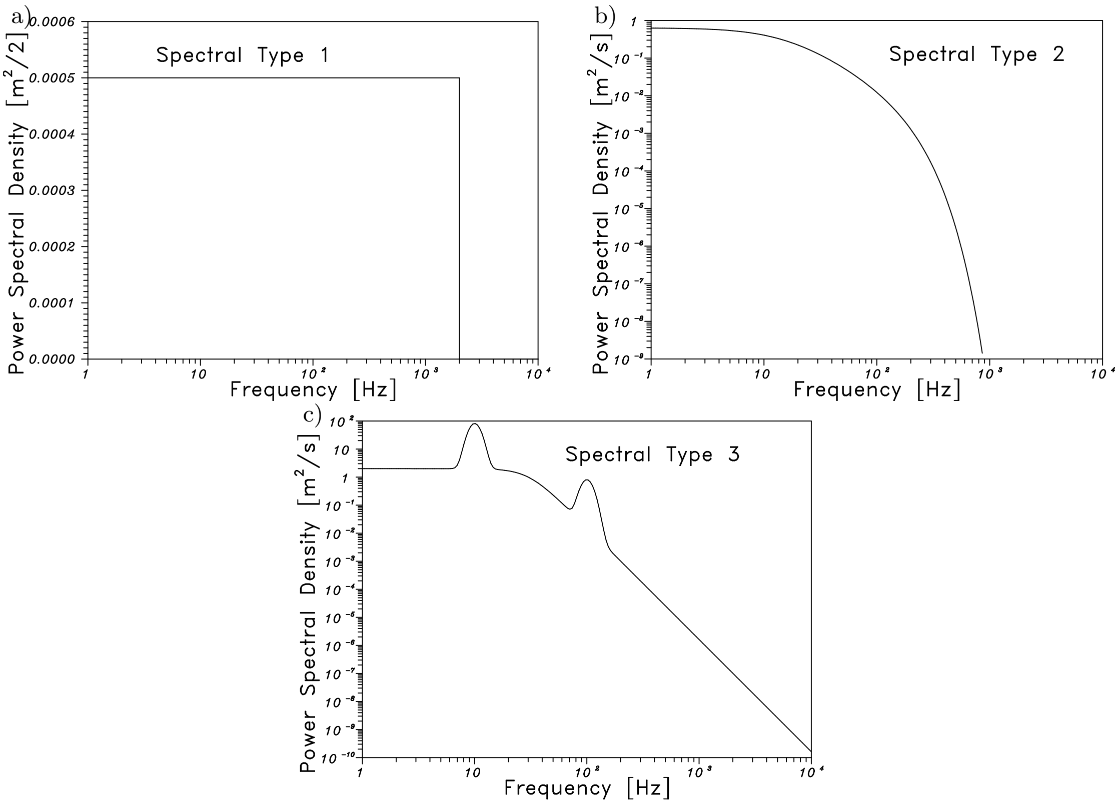

Figure 3: Spectral types: a) Band-limited random noise; b) Pao-like spectrum; c) Two-peak spectrum.

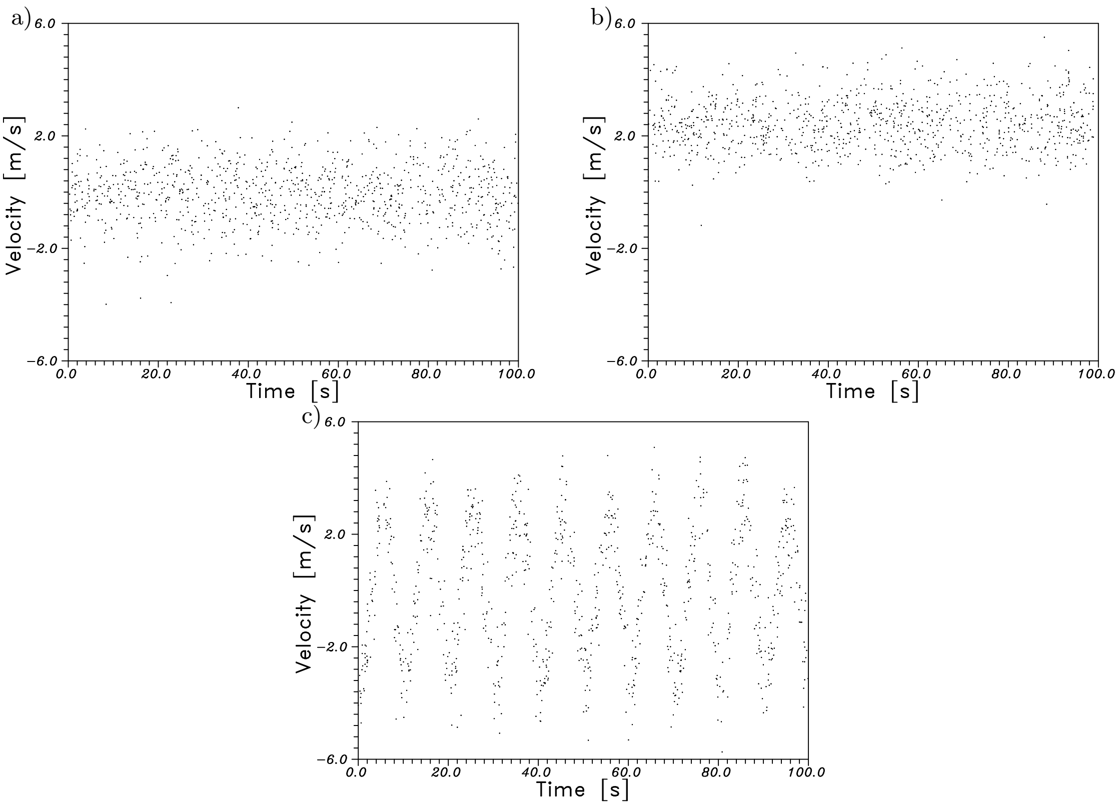

Figure 4: Velocity time series: a) without mean or periodic component; b) with a mean velocity and c) with a periodic component.

Type	σ_u^2 [m^2/s^2]	f_p [kHz]	q	m [m/s]	A [m/s]	F [Hz]	$\dot{n}$ [Hz]	v [m/s]	σ_n^2 [m^2/s^2]	t_d [ms]
1	1.0	20	128	–	–	–	1 000	–	–	–
2	1.0	4	256	$2.0^\times$	$3.0^\times$	0.1	10 or 100*	$1.0^\times$	$1.0^\times$	$1.0^\times$
3	1.0	2	2048	–	–	–	100	–	–	–

× if any
* low and high data rate

Table 1: Simulation parameters.

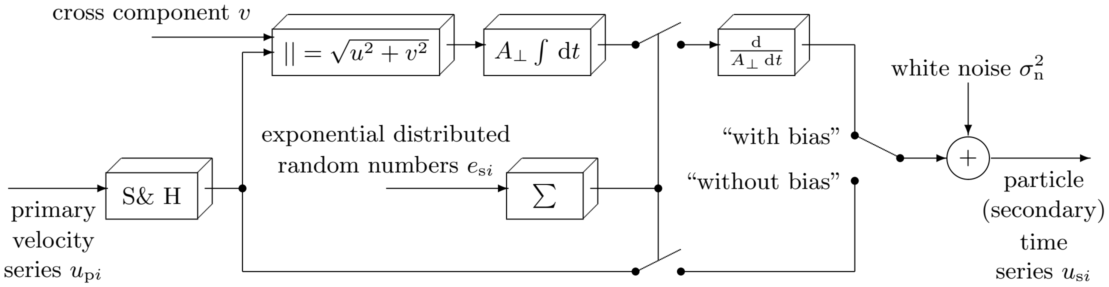

Figure 5: Irregularly sampling with different velocity bias implementations.

exponentially. It has to be integrated until the next particle arrives (upper line in figure 5) using the absolute magnitude of the velocity including a cross component v and the projection area of the measurement volume $A_\perp$ that is assumed to be constant.

Through different relations between the mean velocity m in the measurement direction and the cross component it is possible to generate data sets with different velocity bias conditions, e.g. a positive bias of variance estimation without a mean or cross velocity, a negative bias with a strong mean velocity and a non-proportional relationship between the data rate and the measured velocity component through the use of a cross component.

In any case, the arrival times of the particles will not coincide with the sampling times of the primary velocity series. To obtain the velocity at the particle arrival times a zeroth order interpolation of the primary series is used, i.e. the value of the previous sample is used, as illustrated in figure 6). The resulting continuous velocity function has a new PSD. It differs from the PSD of the primary series, e.g. it has no aliasing errors. Figure 7 shows the spectra of the three types after the sample-and-hold interpolation.

In a final step a white noise with a Gaussian distribution and a power of σ_n^2 is added to the velocity values, corresponding to the uncertainty of the frequency estimation in the signal processor.

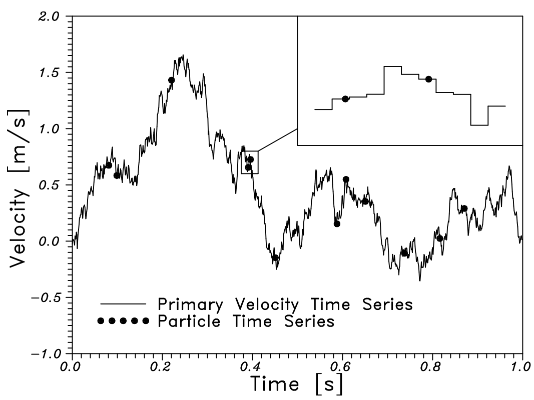

Figure 6: Interpolation of the primary time series at the computed particle arrival times.

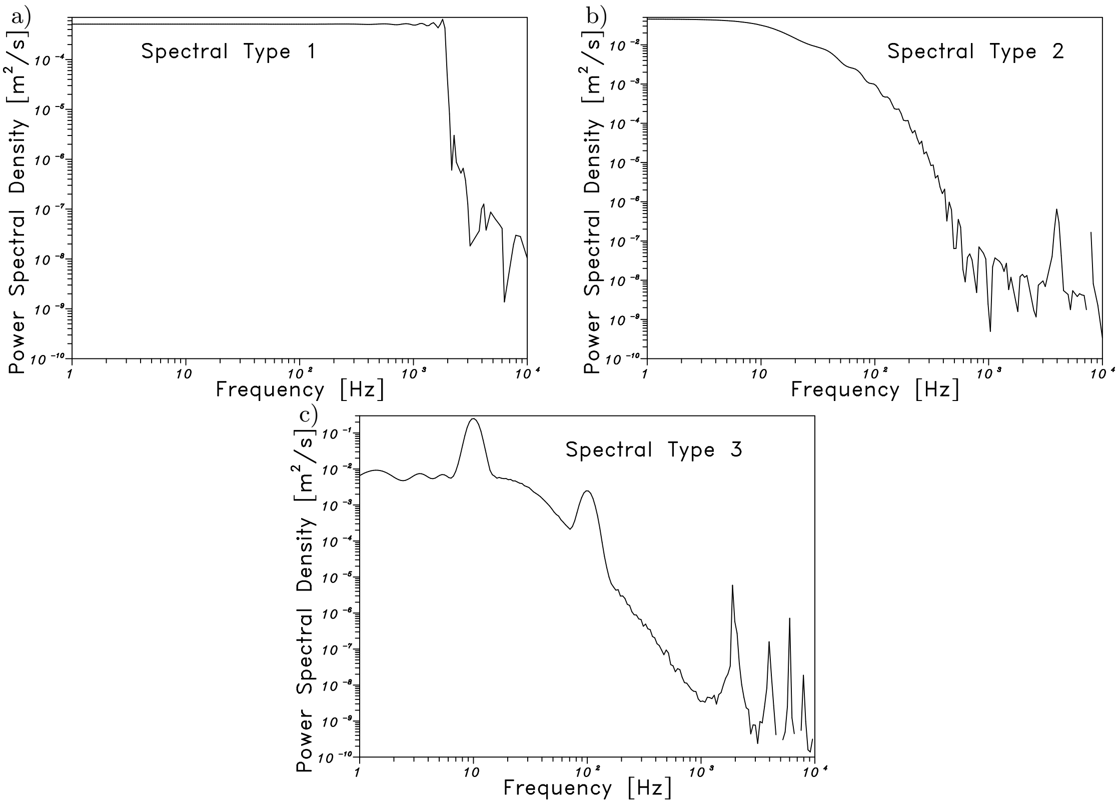

Figure 7: Spectral types after sample-and-hold interpolation: a) Band-limited random noise; b) Pao-like spectrum; c) Two-peak spectrum.

2.3 Processor (Tertiary) Time Series

The processor simulation assumes that only a single frequency measurement per burst is performed with a processor dead time of t_d. Thus, simulated velocity samples are only accepted in this third simulation step if the arrival times since the preceding particle is at least t_d. Furthermore, the actual arrival time t_{ti} is chosen randomly within the range $t_{si} \pm \frac{1}{2} t_d$, corresponding to the uncertainty of the arrival time estimation (t_{si} are the simulated particle arrival times from the particle time series).

3 Using the Programs and Program Parameters

The generation program `bm2` is available as a DOS-executable (`bm2dos.exe`) and as a Windows (32 bit) executable (`bm2win.exe`). It can be started through

```
bm2.exe name group number
```

or

```
bm2.exe name group number samples
```

where `bm2.exe` is either `bm2dos.exe` or `bm2win.exe`, `name` is the name of the output file (max. 8 characters) without an extension, `group` (1…66) defines the data characteristics (see table 2) and `number` (1…) is the index of the individual realisation. The upper limit of the index is given through the internal integer precision. Using different values for `number` leads to independent data sets, the same value leads to the same data set. This can be used to generate the same data sets on different machines to process them by different estimation routines. The parameter `samples` is optional. It defines the number of samples to be generated (default 100 000).

The program generates the output file `name.ts` containing the LDA data set as an ASCII text. Each row gives one sample of the format

Arrival Time Velocity

where the values are given in the floating point representation

[-]<digit>.<digits>E[+|-]<exponent>

and the values are separated by single spaces.

Beside this, the program generates a file `name.coe` containing the coefficients of the FIR filter for the generation of the primary velocity series. The additional program `co2.exe` calculates the theoretical spectrum and the autocorrelation function of the interpolated primary time series (see figure 7). It can be started through

```
co2.exe name f1 f2 Nf
```

where `name` is the name of the coefficients file without an extension, `f1` and `f2` (> 0) gives the frequency range of the spectrum and `Nf` is the number of spectral values. Because the theoretical spectrum of the interpolated velocity function is not sampled, it is possible to calculate the PSD for any frequency. The program `co2.exe` divides the spectrum equidistant on a logarithmic scale within the given frequency range. The result is written as ASCII data to the file `name.spc`. Each row gives one value of the spectrum

Frequency   Power Spectral Density

where the values are given in the floating point representation again. Furthermore, the program `co2.exe` calculates the theoretical autocorrelation function of the primary time series. The result is written as ASCII data to the file `name.acf`. Each row gives one value of the function

Time Lag   Correlation

where the values are given in the floating point representation again. This function is sampled equidistantly with the sampling frequency f_p of the primary velocity series. Nevertheless, the theoretical correlation function is continous. It can be reconstructed from the sampled function by a linear interpolation (the linear interpolation is exact for the used series generator). For a time lag $\geq q f_\mathrm{p}$ with the order q of the filter, the theoretical correlation is zero.

With the group parameter the individual characteristics of the LDA data set can be chosen. Table 2 gives an overview of all possible groups. The spectral types 1 and 3 are only represented by a standard parameter set. These spectral types are interesting to determine filtering effects of the signal processing algorithms. The spectral type 2 is used with various parameter combinations to determine the influence of several data set parameters on the results of the signal processing algorithms.

References

[1] W Fuchs, H Nobach, and C Tropea. Laser Doppler anemometry data simulation: Application to investigate the accuracy of statistical estimators. *AIAA Journal*, 32:1883–1889, 1994.

[2] L H Benedict, H Nobach, and C Tropea. Benchmark tests for the estimation of power spectra from LDA signals. In *Proc. 9th Int. Symp. on Appl. of Laser Techn. to Fluid Mechanics*, Lisbon, Portugal, 1998. paper 32.6.

Group	Spectral Type	Bias	Mean	Cross Velocity Component	Noise	Data Rate	Processor Dead Time	Periodic Component
1	1					↑		
2	2					↑		
3	2				×	↑		
4	2					↓		
5	2				×	↓		
6	2					↑	×	
7	2				×	↑	×	
8	2					↓	×	
9	2				×	↓	×	
10	2					↑		×
11	2				×	↑		×
12	2					↓		×
13	2				×	↓		×
14	2					↑	×	×
15	2				×	↑	×	×
16	2					↓	×	×
17	2				×	↓	×	×
18	2	+				↑		
19	2	+			×	↑		
20	2	+				↓		
21	2	+			×	↓		
22	2	+				↑	×	
23	2	+			×	↑	×	
24	2	+				↓	×	
25	2	+			×	↓	×	
26	2	+				↑		×
27	2	+			×	↑		×
28	2	+				↓		×
29	2	+			×	↓		×
30	2	+				↑	×	×
31	2	+			×	↑	×	×
32	2	+				↓	×	×
33	2	+			×	↓	×	×
34	2	−	×			↑		
35	2	−	×		×	↑		
36	2	−	×			↓		
37	2	−	×		×	↓		
38	2	−	×			↑	×	
39	2	−	×		×	↑	×	
40	2	−	×			↓	×	
41	2	−	×		×	↓	×	
42	2	−	×			↑		×
43	2	−	×		×	↑		×
44	2	−	×			↓		×
45	2	−	×		×	↓		×
46	2	−	×			↑	×	×
47	2	−	×		×	↑	×	×
48	2	−	×			↓	×	×
49	2	−	×		×	↓	×	×
50	2	∗		×		↑		
51	2	∗		×	×	↑		
52	2	∗		×		↓		
53	2	∗		×	×	↓		
54	2	∗		×		↑	×	
55	2	∗		×	×	↑	×	
56	2	∗		×		↓	×	
57	2	∗		×	×	↓	×	
58	2	∗		×		↑		×
59	2	∗		×	×	↑		×
60	2	∗		×		↓		×
61	2	∗		×	×	↓		×
62	2	∗		×		↑	×	×
63	2	∗		×	×	↑	×	×
64	2	∗		×		↓	×	×
65	2	∗		×	×	↓	×	×
66	3					↑		

× yes
\+ positive variance bias (large turbulence level)
− negative variance bias (small turbulence level)
∗ data rate $\not\sim$ measured velocity (cross component)
↑ high
↓ low

Table 2: Specification of the data characteristics.