Review on

Improvement of Vertical Velocity Statistics Measured by Doppler Lidar through Comparison with Sonic Anemometer Observations

In this investigation, vertical velocity variances derived from measurements by two different Doppler lidars are compared to in situ-measurements by sonic anemometers on a 300-m tower. The autocovariance technique of Lenschow et al. (2000) is used to correct for uncorrelated noise in the Doppler lidar measurements. The technique is also refined by analyzing in detail the optimal part of the autocovariance function (called lag time) to be used when applying the technique, based on theoretical considerations.

The need for thorough validation studies of remote-sensing measurements by a comparison to in-situ measurements is clear and therefore this investigation if fully justified. It is also true that different ways to use the autocovariance technique are in use and that its applicability / usefulness is not unambiguous. Thus, this study is useful to demonstrate the proper use of the autocovariance technique and the article can be recommended for publication.

Some points of this investigation have to be clarified and I would also suggest some minor reorganization of aims / sections.

Main comments:

(1)
In some cases in this investigation, the autocovariance technique results in a negative error. It is assumed that this is due to volume-averaging effects, i.e. caused by the limited spatial resolution of the lidar measurement, which can lead to too steep slopes in the turbulence spectra and, thus, to an underestimation of the variance. However, it is not clear to me if this assumption that a negative noise-error is related to this effect is physically based. Moreover, it is not clear from the shown spectra if the volume-averaging effect really has an impact on the analyzed measurements. In the results shown, it also does not clearly have a positive effect.
I would suggest either to define a lower threshold of 0 for the assumed noise error and not to discuss the volume-error in this way or to analyze possible occurring volume-errors more closely and to demonstrate the relationship between the latter and the ‘negative-noise’ error.

(2)
In the aims as well as in the conclusion, it is said that the “optimal parameters [i.e., lag time] that should be used when applying the autocovariance method” are determined.
I would have expected a comparison of variances against sonic-derived variances, as in Fig. 10, but using different lag times. Instead, the optimal lag time is determined by theoretical considerations mainly without demonstrating the validity of these considerations. Only in section 4.1, the ideal lag time determined in section 3 is shown for the sonic measurements, but only compared against a much larger lag time. What about a fixed lag times between 1 s and 10 s, e.g.? In their article, Lenschow et al. (2000) suggest to use “the first few lags” (p.

1333, last sentence in left column in their article).
I would suggest not to mix up the discussion of the temporal averaging of sonic measurements with the analysis of the lag times. Anyway, I don't understand why you analyze the noise error of the sonic measurements at all – the main conclusion from this is "that the sonic observations […] contain little noise" (clearly evident from the red line in Fig. 6). It would be better to do this for the lidar data.

Specific comments:

- p. 4, l. 1: "the first in-depth analysis": it is true that you perform a very detailed analysis, but I would suggest to formulate this more cautiously because you investigate a rather short period (2 days) with mainly neutral to stable conditions (no characteristic convective conditions for sure)

- p. 4, l. 28: "two-day period between 26 March and 28 March" → can you give the exact dates? also in local time so that the reader knows how many nighttime/daytime periods are used?

- p. 6, l. 25: 10-m wind speed instead of surface wind speed?

- p. 8, eq. 2 / l. 18: it is a bit confusing that you use *t* as the time lag in eq. 2 and then say that *w(t)* is a correlated variable, wherein *t* is again time

- p. 11, eq.: the argument in the integral should be *M*(t)* and not *M(t)*

- p. 11, l. 11-14: leave out the two sentences ("Using this method …. compared to the small values of sigma_w"); I would not refer to the results section at this point and leave out these sentences

- p. 11, l. 22-23: slow advection also means that your time series sample is shorter in spatial dimension and less turbulent structures are contained within which makes statistics less reliable – this is of course related to larger integral time scale, as you say. Nevertheless, could the increased sampling error also be of importance?
  See also main comment (1).

- p. 13, section 4.1: as said above (main comment 2), this subsection needs to be refined; first, describe only the averaging procedure for the sonic data. You can even consider to move this to a different section then, because it is more a method than a result. If you want to make a thorough analysis of the effect of averaging the sonic data, I would expect that you (1) show the averaging times which are used later in the study and (2) show the spectra. Did you also test a lowpass filter instead of averaging?
  I am also somewhat surprised that the averaging from 60 Hz to 1 Hz has more or less no effect on the variance while averaging to 0.1 Hz clearly eliminates relevant scales.

This can be much better understood when spectra are compared / shown.

- in Fig. 10, it is obvious that the bias between sonic-derived and lidar-derived variances is smaller for WC than for OU DL, but that the scatter is much larger; this raises the question how you selected the examples in Figs. 8 + 9? are the examples for WC the best matches? Actually, I would have expected a better match of zero-lag autocovariance ("raw" variance) for OU DL examples shown in Fig. 8 than for the WC examples in Fig. 9, because of the higher sampling frequency of OU DL and due to the measurement height: you argue that deviations of OU DL from sonic-derived variance "are likely due to differing measurement volumes". If the measurement volume is the main reason, the difference should actually be smaller at higher altitudes because you can expect the dominant scales to increase with altitude. Thus, deviations should be smaller for the example in Fig. 8 (300 m) than in Fig. 9 (100 m).

  Are differences between "fitted" and "raw" variances significant compared to the difference to sonic values?

- Fig. 9: can you mention values of Ri as in Fig. 8?

- p. 17, l. 20. / Fig. 8: "flattening of the spectra": it is very hard to see this; the spectra are very noisy. Did you consider windowing for calculation of the spectra?

- p. 22, l. 7-13: does this paragraph not belong to the previous subsection?

- p. 22. l. 14-15: it would be helpful to give the figure reference here (where can the slope of the best fit lines be seen?)

- Fig. 11: you could draw a vertical line at 0.25 Ri to indicate the critical Richardson number

- p. 22, l. 26 / l. 30: "When conditions are stable […], fitting is less clear" – "applying the extrapolation technique during stable conditions generally improves the estimates" → isn't that a contradiction?

- p. 24, l.9: You could add that especially unstable / convective conditions were not present