New Lidar Technique to Eliminate Noise Floor from Spectra: a Demonstration using Antarctic Observations and Modeling

Jackson Jandreau[(a)], Xinzhao Chu[(a)]
[(a)] *University of Colorado Boulder, CIRES, 216 UCB, Boulder, CO 80309, USA*
E-mail address: jackson.jandreau@colorado.edu

Abstract: Noise in lidar data greatly complicates the derivation of crucial second-order parameters (such as atmospheric wave energies, fluxes, and spectra), especially when the signal-to-noise ratio is low, leaving a bias in the derived parameters. Following studies exploring the use of covariance to eliminate the noise bias in the temporospatial domain, this study explores using the cross power spectral density to eliminate the noise floor in spectral data. It is found that the method is effective in most cases, and the study also explores the accuracy and precision of the approach.

1. Introduction

Lidar data capturing physical and dynamical processes inherently contains noise alongside the signal of interest due to photon noise in the detection process. This noise is often a limiting factor in using all of the data and often obscures features and trends. As such, it has always been of great interest to the field to minimize or eliminate such noise. For first-order products like atmospheric neutral density, temperatures, and winds measured by a lidar, minimization of the photon noise atop the signal is accomplished simply by averaging multiple samples together as the noise is zero-mean and incoherent between samples, and thus, drops towards zero with increasing sample incorporation (as in Eq. 1). Second-order products lose this mean-zero property, and thus, are left with cross terms and a bias term. While the cross terms still drop (after a sufficient number of samples are averaged) as they are uncorrelated, the bias term does not approach zero regardless of the number of samples incorporated (as seen in the right-hand side of Eq. 2a and 2b).

$$\overline{\rho'_{tot}} = \overline{\rho'} + \overline{\Delta\rho} = \overline{\rho'} \qquad (1)$$

$$\overline{{\rho'_{tot}}^2} = \overline{(\rho' + \Delta\rho)^2} = \overline{{\rho'}^2} + \overline{2\rho'\Delta\rho} + \overline{(\Delta\rho)^2} = \overline{{\rho'}^2} + \overline{(\Delta\rho)^2} \qquad (2a)$$

$$\overline{DFT(\rho'_{Total})DFT(\rho'_{Total})^*} = \overline{|DFT(\rho')|^2} + \overline{2DFT(\rho')DFT^*(\Delta\rho)} + \overline{|DFT(\Delta\rho)|^2} = \overline{|DFT(\rho')|^2} + \overline{|DFT(\Delta\rho)|^2} \qquad (2b)$$

Here ρ'_{tot} is total calculated atmospheric density perturbations, ρ' is the component of the perturbation due to real atmospheric disturbances, $\Delta\rho$ is the calculated perturbation due to noise in the signal. The overbar implies an averaging together of multiple measurements.

For useful atmospheric parameters like variances, energies, fluxes, and power spectral density (PSD), which are dependent on these second-order parameters, this requires the correction of this bias (in the case of variance) or noise floor (in the case of power spectra). In especially low signal-to-noise ratios (SNR), sometimes this correction can be nearly impossible to make using traditional methods.

As pointed out by Gardner and Chu [1], the photon noise bias can be eliminated if we compute these second-order parameters using two statistically independent measurements, and an ideal case is to employ two lidars to probe the same volume of the atmosphere at the same time. However, this is neither cost-effective nor practical; therefore, Gardner and Chu [1] proposed an interleaved data processing technique to use a single set of lidar data to create two samples of which the covariance will be zero. They also theorized that the interleaved approach should also be able to eliminate the noise floor in cross-power spectral density (Cross-PSD) and proposed an initial approach for that process. It is worth noting that the interleaved method shares a similar principle as the time-lagged method proposed in Gardner and Liu [2], i.e., separating data into two independent subsets so that noise from the two subset measurements is uncorrelated thus eliminated via averaging over many samples. Of course, the interleaved idea is more elegant than the time-lagged method on the aspect of minimizing the time or altitude shift between the two subset measurements thus achieving the highest possible accuracy. Jandreau and Chu [3,4] explore this method and compare it to previous approaches. In most cases, it appears to be an effective way to reliably remove noise-variance and noise floor, with a few caveats which are discussed briefly in Section 3.

Present day and future lidars are continuously improving their signal strength and resolution in order to measure weaker and higher-resolution phenomena such as turbulence. Under these conditions, the SNR generally also decreases which drives such biases even higher, thus methods such as these are invaluable. Improvements in processing techniques are especially valuable; not only do they strengthen the capability of future systems, but can be applied to pre-existing observations as well. This study explores the spectral application of the interleaved method through its use with lidar data, explores its performance using a model, and derives its accuracy and precision.

2. Application and Modeling

The raw lidar photon counts are split into even and odd bins, each separated by some δz for altitude interleaving or δt for time interleaving, and then processed identically. In this study, we process the photon counts into relative density perturbations ρ'_A and ρ'_B by subtracting the time mean and altitude mean, and then dividing by the time mean. These are then used to calculate the covariance, DFT, and Co-PSD where we define Co-PSD to mean the real portion of the complex Cross-PSD:

$$Cov[\rho'_{A,\mathrm{tot}}, \rho'_{B,\mathrm{tot}}] = \overline{[\rho'_A \rho'_B]} + \overline{[\rho'_A \Delta\rho_B]} + \overline{[\rho'_B \Delta\rho_A]} + \overline{[\Delta\rho_A \Delta\rho_B]} \quad (3a)$$

$$Co\text{-}PSD(\rho'_A, \rho'_B) = \frac{2\Delta z}{N_z} Re\left[\overline{DFT(\rho'_{A,tot})DFT^*(\rho'_{B,tot})}\right]$$

$$= \frac{2\Delta z}{N_z} Re\left[\overline{\begin{matrix} DFT(\rho'_A)DFT^*(\rho'_B) \\ +DFT(\rho'_A)DFT^*(\Delta\rho_B) \\ +DFT(\Delta\rho_A)DFT^*(\rho'_B) \\ +DFT(\Delta\rho_A)DFT^*(\Delta\rho_B) \end{matrix}}\right] \quad (3b)$$

Here, Δz (Δt) is the spatial resolution of the density perturbation series used to calculate the spatial (temporal) DFT, and N_z (N_t) is the length of the spatial series. When these two subsets are used and enough samples are averaged together, the cross terms will become negligible as before, yet this time, as the photon noise in sample A and sample B is incoherent and thus, the noise terms are uncorrelated, the noise floor term will also approach zero. This leaves a single term which approximates the PSD and statistically has no noise floor. Again, only the real portion of the Cross-PSD is of interest here, as this term contains the energy of the cross-correlation at a lag of zero and allows reduction of noise terms to better approximate the desired spectrum.

We can easily see this noise floor elimination by applying it to atmospheric neutral density observations taken with an Fe Boltzmann lidar over McMurdo station [5,6]. The data shown here is for vertical wavenumber spectra using the altitude interleaving approach, where the altitude spacing between samples A and B is 48 meters. The samples were subsequently binned from 1 minute and 48 m to 1 hour and 0.25 km in order to increase SNR. While 0.25 km is a higher resolution than is typical for this data, this will inflate the noise floor to emphasize the effect of the interleaved method for demonstration. Even with this higher resolution, the interleaved method is still able to derive a believable spectrum from what was otherwise definitively obscured by a noise floor.

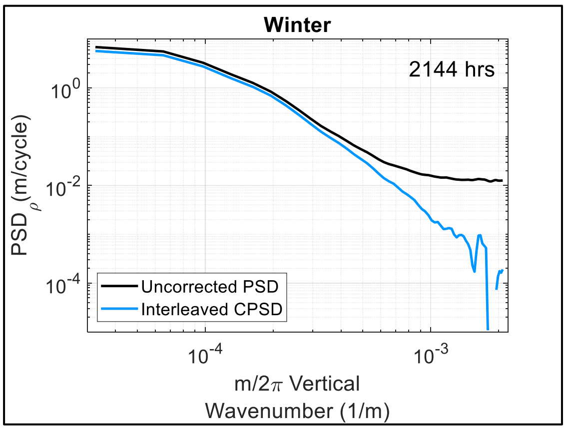

Figure 1. Interleaved method on real lidar data showing power spectral density versus vertical wavenumber.

A forward-model was also created to test the interleaved method. This model is a higher-fidelity model than that of Jandreau and Chu [4] but uses a similar concept to replicate observed spectra. As shown by Tsuda et al. [7], this model starts from a fitting of vertical wavenumber in the atmosphere, applies a random phase between 0 and 2π, and takes the inverse Fourier transform. This yields a unique spatial series of relative perturbations with the desired spectra. These perturbations are then scaled and added to a background density profile. The process of a lidar sensing these perturbations is then applied, scaling the "strength" of the signal to match that of the McMurdo systems and noise is added which matches the both the strength and statistical properties (to a 1st order approx.) seen in a given season of the lidar data (winter simulation is shown here). The process of generating counts based on "observed" density is based on the direct correlation between Rayleigh scattering and atmospheric density, and other processes which affect the observation are considered such as the inverse-square effect and subsequent correction process in order to reflect the actual measurement and processing of lidar data as accurately as possible.

Once the photon count is generated, the simulated data is processed identically to real lidar data using the interleaved method and Fig. 2 is generated. This time, we have the model spectra to which we can compare the effectiveness of the interleaved method. It is clear in Fig. 2 that the interleaved method removes the noise floor as in Fig. 1, and that the additional samples used here drive down the uncertainty some, but not entirely.

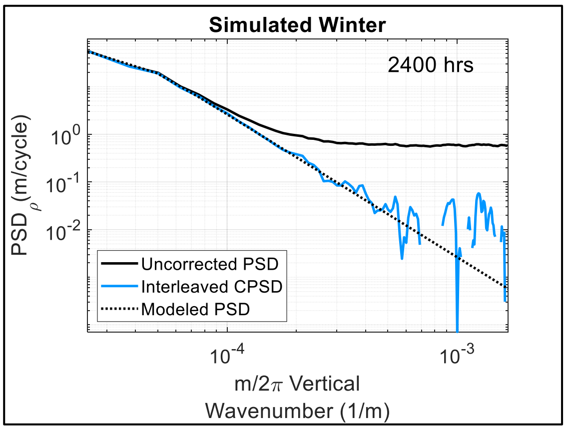

Figure 2. Interleaved method applied to model lidar data with modeled spectra shown.

3. Uncertainties

Calculating a Co-PSD which accurately reflects the atmospheric conditions relies on the creation of two data subsets which are very closely interleaved in time or in space. The interleaving can be done in either altitude or time, but the spatial or temporal difference between the two sets must be kept to a minimum in order to maximize coherency between the two subsets [1]. There will still be some difference between the theoretically ideal PSD and the Co-PSD calculated via the interleaved method, as samples A and B are measuring slightly different phases of the wave. This correction factor relates the spatial (or temporal) difference in the two subsamples to the spectra of the waves being measured, and expectedly scales with m to yield a larger relative correction as the spectral resolution approaches the size of the shift between samples. This correction factor is derived by Jandreau and Chu [4] as:

$$\frac{PSD(\omega) - CPSD(\omega)}{PSD(\omega)} = \frac{(\omega\delta t)^2}{2}$$

If the interleaving is done properly on data capable of supporting it, the correction factor will be nearly negligible at most spatial/temporal frequencies.

Additionally, it is key that we assess the uncertainty (precision) of the derived parameter. A downside to applying the interleaved method is that the subsampling process requires you to halve the number of photons used, which results in an uncertainty increase of $\sqrt{2}$ [1]. As such, applying the interleaved method to a small number of observations can sometimes result in a spectrum with so much uncertainty, the high accuracy of the interleaved method may no longer be worth it. Jandreau and Chu [4] explore this spectral uncertainty significantly, taking care to consider the different correlation times of the various components of the signal. This spectral uncertainty term for the interleaved approach is as follows, where τ is the correlation time of the gravity waves over McMurdo [1] and Δt is the resolution of the samples when averaged, where PSD(Δρ) is the portion of the PSD which is due to noise (can be found by the difference between the Co-PSD and PSD):

$$\Delta CPSD_{\rho'_A\rho'_B}(m) = \sqrt{\frac{2\tau}{T}\,Co\text{-}PSD^2_{\rho'_A\rho'_B} + \frac{\Delta t}{T}\left[\begin{matrix} 2\left(CPSD_{\rho'_A\rho'_B}\right)\mathrm{PSD}(\Delta\rho) \\ +\mathrm{PSD}(\Delta\rho)^2 \end{matrix}\right]}$$

4. Conclusions

With the increase in high-resolution lidar across many fields of atmospheric remote sensing, the field necessitates more complex methods to completely make use of these datasets. The interleaved method of variance has shown its capabilities in recent years, and this study has shown the capability of the interleaved method for performing spectral analysis. Additionally, we have determined the specifics of this processing approach and calculated precision and accuracy. Not only did this interleaved spectral processing technique reveal spectra absent of a noise floor, but it also allowed the resolvable resolution of this lidar to be confidently doubled from previous inspections of this dataset. This technique, alongside the interleaved method of variance calculation, is currently being used to characterize the McMurdo Fe Boltzmann lidar dataset. These methods have already revealed trends in this data that had not previously been seen, and we aim to publish these atmospheric characterizations soon.

5. Acknowledgements

This work was supported by the National Science Foundation (NSF) grants OPP-2110428 and AGS-2330168, and by NASA FINESST grant 80NSSC22K1854.

6. References

[1] C. S. Gardner and X. Ch, “Eliminating photon noise biases in the computation of second-order statistics of lidar temperature, wind, and species measurements,” App. Optics **59(27),** 8259–8271 (2020).

[2] C. S. Gardner and A. Z. Liu, “Measuring eddy heat, constituent, and momentum fluxes with high-resolution Na and Fe Doppler lidars: Measure eddy fluxes with lidars,” J. Geophys. Res.: Atmo. **119(17),** 10583–10603 (2014).

[3] J. Jandreau and X. Chu, “Comparison of Three Methodologies for Removal of Random-Noise-Induced Biases From Second-Order Statistical Parameters of Lidar and Radar Measurements,” Earth and Space Sci. **9,** e2021EA002073 (2022).

[4] J. Jandreau and X. Chu, “Bias-Eliminating Techniques in the Computation of Power Spectra for Characterizing Gravity Waves,” Earth and Space Sci, *under review*, (2024).

[5] X. Chu, W. Huang, W. Fong, Z. Yu, Z. Wang, J. A. Smith, and C. S. Gardner, “First lidar observations of polar mesospheric clouds and Fe temperatures at McMurdo (77.8°S, 166.7°E), Antarctica,” Geophys. Res. Lett. **38,** L16810 (2011).

[6] X. Chu, Z. Yu, C. S. Gardner, C. Chen, and W. Fong, “Lidar observations of neutral Fe layers and fast gravity waves in the thermosphere (110-155 km) at McMurdo (77.8°S, 166.7°E), Antarctica,” Geophys. Res. Lett. **38,** L23807 (2011).

[7] T. Tsuda, T. Inoue, D. C. Fritts, T. E. VanZandt, S. Kato, T. Sato, and S. Fukao, “MST Radar Observations of a Saturated Gravity Wave Spectrum,” J. of the Atmos. Sci. **46(15),** 2440–2447 (1989).