# Consideration of non-Poisson distributions for lidar applications

Andrew J. Gerrard, Timothy J. Kane, Jeffrey P. Thayer, Christopher S. Ruf, and Richard L. Collins

Poisson statistics are traditionally used to estimate the mean and standard deviation of the mean in time–range realizations of received photon counts from stationary processes in incoherent-detection lidar systems. However, this approach must be modified if the process under study is measurably nonstationary to account for any additional (and potentially unanticipated) variability. We demonstrate that the modified approach produces a different form for the estimated standard deviation of the mean for lidar return counts, which can also be applied to binning of higher-order data products. This modified technique also serves to determine optimum time–range integrations, diagnose system stability, and constrain operational modes. © 2001 Optical Society of America

*OCIS codes:* 280.3640, 000.4430, 000.5490.

## 1. Introduction

In many incoherent-detection lidar applications (as well as a number of coherent-detection systems), scattered photons from a given transmitted laser pulse are collected and counted by a receiver and sorted into range-resolved bins (or range realizations). Within a given range realization, the return photon counts from multiple transmitted pulses (transmitted within a small time period) are usually summed if the scattering medium is statistically stationary (in this paper we refer to stationary processes as processes that are at least wide sense stationary and nonstationary processes as those that are not stationary in the first sense). We call this initial summing and integration time period for a fixed range a nominal time–range realization and assume throughout this paper that the process is stationary within this realization. Depending on the experimental requirements, nominal time realizations can scale from pulse-to-pulse (or shot-per-shot) intervals to over 1-min integrations, whereas nominal range realizations usually are determined by the characteristics of the laser and receiver system. It is these data that are usually saved for future, off-line postprocessing. Poisson statistics (which govern many counting processes) state that the estimated mean of the measurement from a given nominal time–range realization $i$ corresponds to the total number of photons counted in the given bin $\hat{x}_i$ with an estimated standard deviation $\hat{\sigma}_i = \sqrt{\hat{x}_i}$ (where the circumflex denotes an estimated quantity). By collecting many nominal realizations, which are assumed to be uncorrelated in time and range, one obtains time–range-resolved data that contain information about the geophysical process under study. It is the ability to retrieve the geophysical information from the data accurately that requires proper representation of its associated error.

In postprocessing, $N$ consecutive nominal time realizations are often averaged to improve the signal-to-noise ratio (SNR) at the cost of temporal resolution (nominal range realizations can also be averaged together, but we consider only time integration at each of the various ranges throughout this paper without loss of generality). Again, from conventional Poisson statistics, it can be

A. J. Gerrard (gerraraj@morrisville.edu) is with the College of Agriculture and Technology at Morrisville, Galbreath Hall, State University of New York, Morrisville, New York 13408. T. J. Kane (tjk7@psu.edu) is with the Pennsylvania State University, 121 Electrical Engineering East, University Park, Pennsylvania 16802. J. P. Thayer (thayer@sri.com) is with the Radio Science and Engineering Division, SRI International, 333 Ravenswood Avenue, G-275, Menlo Park, California 94025. C. S. Ruf (cruf@umich.edu) is with the University of Michigan, 1521C Space Research Building, 2455 Hayward Street, Ann Arbor, Michigan 48109-2143. R. L. Collins (rlc@gi.alaska.edu) is with the Geophysical Institute and the Department of Electrical Engineering, University of Alaska, 903 Koyukuk Drive, Fairbanks, Alaska 99775.

Received 24 May 2000; revised manuscript received 5 December 2000.

0003-6935/01/091488-05$15.00/0

shown that the estimated mean (of the true mean, $\bar{\mu}$) can be expressed as

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \hat{x}_i, \tag{1}$$

and the estimated standard deviation of the mean is

$$\hat{\sigma}_{\hat{\mu}} = \sqrt{\hat{\mu}/N}, \tag{2}$$

both of which are analogous to the summing of return counts mentioned above. We refer to this method as the traditional Poisson approach to binning return counts from stationary processes (see Refs. 1–4 for an overview of Poisson distributions in counting statistics).

However, if the process under study is measurably nonstationary (for reasons discussed in Section 2), we note that the formulation of the estimated standard deviation of the mean given in Eq. (2) must be modified to effectively account for the additional variability in the return counts.[5] As a consequence of this modification, further information on the geophysical process under study, the stability of the lidar system, the optimal integration of time and range, etc. can be ascertained. Therefore the purpose of this paper is to promote awareness of such nonstationary return, as well as to note the importance and utilization of different error analysis methods applicable to many lidar applications.

## 2. Description of Nonstationarity and Subsequent Analysis

Returns from nonstationary processes, and hence non-Poisson lidar return count distributions, have been noted in a number of past papers (e.g., Ref. 5 pertains to return from a middle-atmospheric molecular and aerosol lidar, Ref. 6 from a tropospheric differential absorption lidar system). To understand why such distributions are measured over $N$ realizations, one needs to consider the conditional probability distribution functions (PDF's) from each nominal time–range realization, which we hereafter call child PDF's. If the statistics (namely, the mean and variance) of the child PDF's are the same over the $N$ realizations, then the process is considered stationary over the $N$ realizations, and Eqs. (1) and (2) are used to describe the estimated mean and the standard deviation of the mean. If the statistics of the child PDF's differ over the $N$ realizations, then the process is nonstationary over the $N$ realizations, and the above approach is not valid. The cause of such nonstationarity could be due to the nature of the geophysical process being studied, such as turbulence, clouds, gravity waves, or possibly an instrumental artifact such as system drift.

Nevertheless, if the characteristics of the child PDF's differ from realization to realization it is necessary to consider the joint PDF over the $N$ realizations, hereafter referred to as the parent PDF, which can be described as a combination of the various child PDF's (see Fig. 1). In atmospheric remote sensing

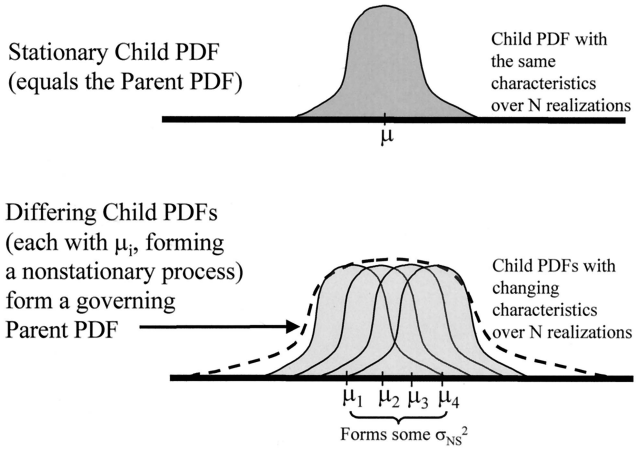


Fig. 1. Top, schematic depicting a stationary child PDF, which maintains its mean and variance over multiple realizations and is therefore equal to the overall parent PDF. Bottom, schematic depicting potential nonstationary return that is due to the different characteristics of multiple child PDF's. In this case, a new overall parent PDF is formed that includes the variance for all the children.

applications, this resultant parent PDF often takes on a wider, Gaussian-like, not Poisson-like, distribution. We note that the nonstationary component of the return is uncorrelated with the stationary Poisson process, although higher-order joint moments can exist. Although it can be shown that Eq. (1) still gives the estimated mean of the parent PDF over the $N$ realizations, Eq. (2) fails to describe the estimated standard deviation of the mean completely. Because the standard deviation of the mean is a fundamental quantity in any experimental investigation, two example approaches are presented below that can be used to calculate either this parameter or a variety of correlations between bins.

### A. Method 1

Because of the changing characteristics of the child PDF's, it can be shown that the total variance of the parent PDF ($\sigma_{\text{total}}^2$) is

$$\sigma_{\text{total}}^2 = \sigma_p^2 + \sigma_{\text{NS}}^2, \tag{3}$$

with the total variance found from the sample variance of an infinite number of data points

$$\sigma_{\text{total}}^2 = \langle (x_i - \bar{\mu})^2 \rangle, \tag{4}$$

(where $\langle\ \rangle$ denotes an expected value) that is itself composed of a stationary component that accounts for the variance of the Poisson nature ($p$) of photon counting,

$$\sigma_p^2 = \langle \sigma_i^2 \rangle = \bar{\mu}, \tag{5}$$

and a nonstationary (NS) component that accounts for any additional variance of the child PDF's (say each individual child PDF has a mean $\mu_i$),

$$\sigma_{\text{NS}}^2 = \langle (\mu_i - \bar{\mu})^2 \rangle. \tag{6}$$

Because the means of the individual child PDF's and the parent PDF are unknown, as well as the fact that there is a limited sample size, $\hat{\sigma}_{\mathrm{NS}}^2$ must be estimated[5] by

$$\hat{\sigma}_{\mathrm{NS}}^2 = \hat{\sigma}_{\mathrm{total}}^2 - \hat{\sigma}_p^2 \cong \left[\frac{1}{N-1}\sum_{i=1}^{N}(\hat{x}_i - \hat{\mu})^2\right] - \hat{\mu}. \quad (7)$$

The first term on the far right-hand side of Eq. (7) is the sample variance of the entire process, and the second term represents the bias that is due to the Poisson noise. With this parameter determined, it can be shown that the estimated (unbiased) standard deviation of the mean that incorporates the additional variability of the nonstationary process can be expressed as

$$\hat{\sigma}_{\hat{\mu}} \cong \left(\frac{\hat{\mu}}{N} + \hat{\sigma}_{\mathrm{NS}}^2\right)^{1/2}. \quad (8)$$

In the limit of large $N$, the modified standard deviation of the mean approaches the standard deviation of the nonstationary process over the range of combined bins, instead of zero produced by the Poisson error analysis of Eq. (2). In addition, if the process is stationary (i.e., $\mu_i = \bar{\mu}$), then $\sigma_{\mathrm{NS}}^2 = 0 = \hat{\sigma}_{\mathrm{NS}}^2$, and the traditional Poisson approach is maintained.

### B. Method 2

Similar results can also be obtained if one computes the estimated autocorrelation $\hat{R}(\tau)$ (where $\tau$ is the time lag) of the collected data.[7] The data over the $N$ bins are now considered as a noise-free, nonstationary time signal NS($t$), with an additional related white-noise process $p(t)$. Doing such, we can show that

$$R(\tau) = \langle[\mathrm{NS}(t) + p(t)][\mathrm{NS}(t+\tau) + p(t+\tau)]\rangle = \begin{cases} \hat{R}_{\mathrm{NS}}(\tau) & \text{for } \tau \neq 0 \\ \hat{\sigma}_{\mathrm{total}}^2 = \hat{R}_{\mathrm{NS}}(\tau = 0) + \hat{\sigma}_p^2 & \text{for } \tau = 0 \end{cases}. \quad (9)$$

At $\tau = 0$, the autocorrelation is equivalently the sample variance that is the addition of the nonstationary signal variance $\hat{R}_{\mathrm{NS}}(\tau = 0) = \hat{\sigma}_{\mathrm{NS}}^2$ and the noise variance $\hat{\sigma}_p^2$. As mentioned above, the variance of the nonstationary process can be expressed as

$$\hat{\sigma}_{\mathrm{NS}}^2 = \hat{R}_{\mathrm{NS}}(\tau = 0) \cong \hat{R}(\tau = 0) - \hat{\mu}. \quad (10)$$

Alternatively, one could also extrapolate from the lags of the autocorrelation function above 0 to find an estimate of $\hat{R}_{\mathrm{NS}}(\tau = 0)$, which can then be used to find an estimate of the noise variance as a consistency check of the procedure. It can be shown that the results of this method are similar to those of method 1. Note that this method also yields the correlation structure between bins, so that the resultant uncertainty of the binning process could be found by standard error propagation.

## 3. Higher-Order Data Binning

It is also important to note that the analyses presented above, as applied to return photon counts, can also be applied to the binning of higher-order lidar data products (i.e., lidar products that are derived from the nominal time–range raw photon counts before binning). For example, the range-resolved return counts in each nominal temporal realization from lidar measurements of middle-atmospheric parameters are often scaled to the (absolute or relative) atmospheric density (obtained from independent radiosonde measurement or from an atmospheric model) above the altitude where aerosols are usually expected (e.g., 30 km).[8,9] This scaling can remove signal fluctuation induced by changing conditions at lower altitudes (e.g., variable tropospheric and lower-stratospheric cloud and aerosol attenuation) and system drift, thus removing potential variability from these sources. The atmospheric density profiles are then averaged (i.e., binned) in time and range to increase the SNR. Because fluctuations in the densities can still be present (induced, for example, by gravity waves in the middle atmosphere), the subsequent uncertainty analysis would follow the techniques presented in method 1 or method 2 above, with the various estimated standard deviations now corresponding to the higher-order data products (hodp). For this particular example, Eq. (7) and relation (8) would be rewritten as

$$\hat{\sigma}_{\mathrm{NS}}^2 = \hat{\sigma}_{\mathrm{total}}^2 - \hat{\sigma}_{\mathrm{hodp}}^2 \cong \frac{1}{N-1}\sum_{i=1}^{N}(\hat{x}_i - \hat{\mu})^2 - \frac{1}{N}\sum_i \hat{\sigma}_i^2, \quad (11)$$

$$\sigma_{\hat{\mu}} \cong \left(\frac{\sum_i \hat{\sigma}_i^2}{N^2} + \hat{\sigma}_{\mathrm{NS}}^2\right)^{1/2}, \quad (12)$$

respectively, where $\hat{\sigma}_i$ is the estimated standard deviation of the estimated higher-order data product $\hat{x}_i$ of each bin $i$ (i.e., for the example below, $\hat{x}_i$ is the relative atmospheric density of bin $i$), and the mean is similar in form to Eq. (1). We note that the order of the data processing sequence is important, as one can account for and potentially remove certain types of variability in the data depending on the desired final data product.

## 4. Example of Analysis

To compare and contrast results between the traditional Poisson method and the new method (e.g., method 1), we performed numerical simulations with known nonstationary components using methods 1 and 2. The results lend strong credence to the above formulations and can be readily verified in most software packages. For brevity these numerical simulations are not included for further discussion.

However, we do provide an example with real lidar data in an effort to demonstrate the importance of the above concepts. A typical 6-h Rayleigh lidar data set taken on the night of 4 January 1997 with the Arctic Lidar Technology (ARCLITE) Facility Rayleigh lidar system is used to calculate various temperature profiles in altitude. Further discussion of this system and computation of these quantities are

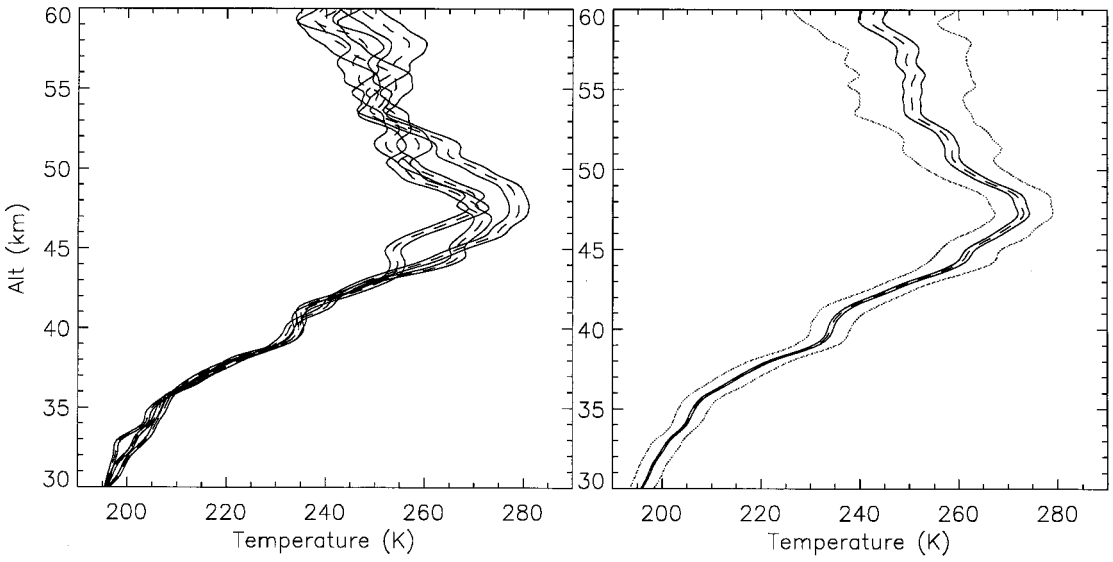


Fig. 2. Left, three consecutive middle-atmospheric temperature profiles from a 30- to a 60-km altitude, each calculated from 2 h of nominal photon count realizations from a 6-h data set. The three profiles are plotted together to show the variation of the mean temperatures (dashed curves) over the total 6-h period. The standard deviation bars (dark, bounding curves) are obtained from traditional Poisson error analysis. Right, plot showing the 6-h nightly mean-temperature profile (dashed curve) calculated from the nominal photon count realizations (the same mean is obtained from both traditional Poisson and the modified technique). The uncertainty bars with both the traditional (dark bounding curves) and modified analysis (light, bounding curves) are also plotted. We note that the new error analysis better acknowledges the overall variability of the process under study.

given in Ref. 5. We first convert the return photon counts to relative atmospheric density profiles (thus removing the above-mentioned lower-atmospheric attenuation and system drift); they are then binned into three consecutive, 2-h segments assuming the process was stationary in each of the three temporal segments [i.e., use of traditional Poisson analysis similar in form to that of Eqs. (1) and (2)]; and then they are converted into three individual mean-temperature profiles. These temperature profiles are displayed in Fig. 2 (left), along with the uncertainty bars for the first standard deviation of the mean. It is obvious that the individual means of the three consecutive 2-h temperature profiles vary greatly. These individual temperature profiles are then compared with the nightly mean-temperature profile that we obtained by binning all 6 h of density profiles using both stationary analysis [again with traditional Poisson analysis similar in form to Eqs. (1) and (2)] and the modified nonstationary analysis [i.e., use of modified method 1, Eq. (11) and relation (12)]. These nightly mean-temperature profiles and the various uncertainty bars from both methods are shown in Fig. 2 (right).

After inspecting the profiles of Fig. 2, one can see that both analysis techniques yield the same mean profile. In addition, it is also noted that the overall uncertainty bars from the traditional stationary analysis are $\sim\sqrt{3}$ times smaller than the individual uncertainty bars, as is also expected from Eq. (2). However, these overall uncertainty bars do not portray the wide range of temperatures that were observed over the course of the night. This result differs from the modified method 1 results, which yielded uncertainty bars that better reflect the overall variability of the process over the 6 h of data collection. In this sense, the uncertainty bars given by the modified technique include the standard deviation of the temperature variability throughout the night, and therefore provide more appropriate bounds of the measured temperatures.

## 5. Discussion and Conclusions

The topics covered in this paper are applicable to data obtained from many different types of lidar systems (e.g., tropospheric cloud-detection, ozone, mesospheric metal, middle-atmospheric Rayleigh). Hence the results should be useful to the lidar community in general, because there is a dearth of papers in the literature on the subject of non-Poisson (return) distributions. In addition, the topics in this study raise a number of related issues:

(1) A limiting factor in these modified methods is the accurate calculation of the estimated sample variance $\hat{\sigma}_{\text{total}}{}^2$ (or estimated autocorrelation, etc.) and the estimated Poisson variance $\hat{\mu}$ or $\hat{\sigma}_{\text{hodp}}{}^2$. These estimated quantities have standard deviations associated with them, and any detectable nonstationary return must have a variability greater than this combined error plus the estimated Poisson variance. In the above analysis we assumed that enough sample values existed such that these errors were negligible. However, such may not always be the case. Although sophisticated methods exist for estimating these variances with few data points,[3] the easiest and most practical way is to collect numerous data realizations or threshold the nonstationary variability.

With the relatively high speed of data-acquisition cards and low cost of memory, collection of many nominal time–range realizations allows for a good estimation of the sample variance, autocorrelation, Poisson variance, etc. from current lidar systems. Because we assumed that the process under study is stationary within the nominal time–range bin, shorter nominal integrations also further satisfy this requirement.

(2) The modified approach will work for any geophysical or instrumental variability present in the data. From a geophysical standpoint, the modified estimate of the standard deviation of the mean can be used to look at or compare fluctuations at various altitudes and times of the year, or between years, allowing for climatologies of such variability. From an instrumental standpoint, the modified technique can help detect system drift and aid in optimizing system performance and characterization. This is important because as data sets get longer and more numerous, there will be a move toward automated processing algorithms. Such algorithms could have difficulty detecting system drifts or glitches, and these methods would prove valuable in finding them.

(3) This modified procedure is valid for both time and range binning.

(4) It is important to realize that a lidar system can discern a change in the child PDF's only mean if the fractional variability of the geophysical process is greater than the 1/SNR value of the system for that particular bin. Hence the potential to discern nonstationary variability will depend on different ranges, scattering mechanisms, etc.

(5) The methods presented here can be used to determine the most effective integration periods (or ranges). To best minimize the uncertainty on the mean value and still retain high resolution, one would want to bin where there are no nonstationary components variance. Because the SNR varies with range, each range bin has an optimum binning length (in which case the bin length will have to be either a constant variable or a variable at each altitude, depending on the application). Such a technique was used in Ref. 5 to discern the optimal Wiener filter characteristics.

(6) It should be noted that longer data collection intervals tend to be associated with greater geophysical variability. This result is intuitively correct because the processes under study are not generally intrinsically stationary in time and range and could therefore vary over long integrations.

(7) Error propagation with other error reduction techniques (such as temporal or spatial filtering) should be used with care as the data are potentially correlated between time–range bins.

(8) If one compares data values obtained from traditional stationary techniques with similar results from other instruments, one might see seemingly significant statistical variations between results. However, the modified method allows for estimation of the degree of variability that is vital for radar, lidar, satellite, radiometer, etc. intercomparisons.

In conclusion, we have given an overview of modified analysis techniques that should be used to address and characterize potential nonstationarity in lidar return counts and potentially higher-order data products. If such nonstationarity exists in the data, the standard deviation of the mean calculated, for example, by Eq. (2), is not always appropriate, and one of the above methods should be used instead to describe this parameter. Such techniques provide a relatively simple way to better characterize the source of variability and nonstationarity, determine optimum time and height integrations, diagnose system stability, etc. Such data can also provide information on the analysis schemes, interpretation, and application of lidar-obtained products.

The authors thank Mark London, Christina Gerrard, and members of the Atmospheric Sensing and Lidar Laboratory for their assistance. A. J. Gerrard, T. J. Kane, and R. L. Collins were supported in part by National Science Foundation (NSF) Coupling, of Energetics and Dynamics of Atmospheric Regions program grant ATM 96-12870; J. P. Thayer was supported by the NSF under cooperative agreement ATM-9813556.

## References

1. P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 2nd ed. (McGraw-Hill, Boston, 1992).
2. J. R. Taylor, *An Introduction to Error Analysis; The Study of Uncertainties in Physical Measurements* (University Science, Mill Valley, Calif., 1982).
3. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. (McGraw-Hill, New York, 1991).
4. R. M. Measures, *Laser Remote Sensing: Fundamentals and Applications* (Krieger, Malabar, Fla., 1992).
5. J. P. Thayer, N. B. Nielsen, R. E. Warren, C. J. Heinselman, and J. Sohn, "Rayleigh lidar systems for middle atmosphere research in the Arctic," Opt. Eng. **36,** 2045–2061 (1997).
6. E. Durieux and L. Fiorani, "Measurement of the lidar signal fluctuation with a shot-per-shot instrument," App. Opt. **37,** 7128–7131 (1998).
7. C. S. Ruf and S. E. Beus, "Retrieval of tropospheric water vapor scale height from horizontal turbulence structure," IEEE Trans. Geosci. Remote Sens. **35,** 203–211 (1997).
8. C. S. Gardner, D. C. Senft, T. J. Beatty, R. E. Bills, and C. A. Hostetler, "Rayleigh and sodium lidar techniques for measuring middle atmosphere density, temperature, and wind perturbations and their spectra," in *World Ionosphere/Thermosphere Study Handbook*, S. H. Liu and B. Edwards, eds. (University of Illinois, Urbana-Champaign, Ill., 1989).
9. C. S. Gardner, "Sodium resonance fluorescence lidar applications in atmospheric science and astronomy," Proc. IEEE **77,** 408–418 (1989).