When can Poisson random variables be approximated as Gaussian?

Matthew Hayman[(1)], Willem J. Marais[(2)], Hristo Chipilski[(3)], Robert A. Stillwell[(1)], Scott M. Spuler[(1)]

[(1)] *National Center for Atmospheric Research, Earth Observing Lab, Boulder, CO, USA, E-mail:mhayman@ucar.edu*

[(2)] *University of Wisconsin Space Science and Engineering Center, Madison, WI, USA, E-mail:willem.marais@ssec.wisc.edu*

[(3)] *National Center for Atmospheric Research, Advanced Study Program, Boulder, CO, USA, E-mail:hristoc@ucar.edu*

In photon counting lidar, photon detection is generally well described as a Poisson point process as long as photon arrival rates are sufficiently low to avoid significant nonlinear effects in the detection chain (e.g. deadtime). In general, it is straightforward to employ the Poisson Negative Log-Likelihood as a noise model in order to obtain a maximum likelihood estimate of photon counting lidar data. However data assimilation methods employing Kalman Filters (KF) and some processing methods such as Optimal Estimation (OE) were originally designed for observational data with Gaussian distributed noise, thus allowing for closed form solutions. A common assumption when applying these methods to photon counting lidar data is that, by the central limit theorem, the Poisson distributed photon counts may be *approximated* as Gaussian distributions. While this approximation is technically correct at high photon counts (assuming linear detection), the accuracy trade-off of the approximation is not always clear and is likely to depend on the particular estimation problem.

In addition, KF and OE are designed to operate on random variables of *known* or *constant* variance/covariance. However, when approximated as Gaussian, Poisson random numbers have variances that are a function of the estimated variable, the mean photons. In order to obtain separability between the noise and the estimated variable – as typically required for OE and KF – a second approximation is made: that the variance of an observation is approximated by the *observed* number of photons instead of the estimated variable.

In this work we aim to use simulations and observations from the MicroPulse DIAL to estimate when (1.) Poisson random variables may be reliably approximated as Gaussian and (2.) when the Gaussian approximated variance can be reliably approximated based on the observed photon counts. This will allow us to better specify when photon counting lidar data can be safely used in standard KF or OE frameworks.

We accomplish this by employing Poisson Total Variation as a maximum likelihood estimator while evaluating the two approximations against the Poisson noise model. This analysis will be conducted on both simulated data and high resolution lidar data.