An Automated Cloud Mask Algorithm for the Micropulse Lidar

G. G. Mace
Department of Meteorology
Pennsylvania State University
University Park, Pennsylvania

To perform cloud base analysis, an automated technique for determination of cloud base height from the raw lidar returns is needed. Previously, a straightforward signal thresholding algorithm was used to determine cloud base height. The difficulty with this approach is illustrated in Figure 1a. The micropulse lidar operates at a wavelength of 523 nm, which is very near the maximum in the solar spectrum. As can be seen in Figure 1a, during the night (0130 - 1200 UTC), identification of significant lidar return using thresholding techniques is generally capable of identifying even thin, high-altitude cirrus above the electronic noise of the instrument. However, solar radiation introduces random noise that can be as large as the signal from many cirrus clouds (Spinhirne 1993), making identification of cirrus returns difficult using signal thresholds.

The cloud returns identifiable during the daylight hours in Figure 1a between 10 and 11.5 km stand out from the noise due to their spatial continuity in the time-height domain. In other words, the spatial variance in the cloud regions tend to be significantly less than that of non-cloudy regions. Our goal is to exploit this characteristic by identifying sets of observations that have identifiably and significantly different signal statistics. The algorithm we describe is modeled after the millimeter radar cloud mask described by Clothiaux et al. (1995), although significant modifications were necessary to account for the difference in lidar and radar signal characteristics. Without significant scatterers in a sampled volume, a radar receives no identifiable return and the recorded signal is just due to internal electronic noise. A lidar, on the other hand, receives identifiable signal return in clear air depending on the molecular backscattering cross section and the range to the sampled volume. Cloud and aerosol return are superimposed on the clear sky signal. Therefore to consider time-height signal statistics it is necessary to remove the range dependence of the clear sky profile. We accomplish this by normalizing the observed signal with a modeled clear sky signal using the standard lidar equation (Spinhirne 1993) and midlatitude standard atmospheres. After normalization, the observations form a swarm of points about unity with the scatter due to electronic noise, solar background and cloud and aerosol return.

In order to identify those regions in the data field whose statistics differ significantly from the statistics of regions known not to be cloudy, we calculate the mean and standard deviation of the normalized signal in a five-minute domain between a 30 and 40 km range. Each data point below 12 km within the five-minute temporal interval is then examined within the context of a 3™5 boxcar of data points (five minutes wide in the temporal domain and 900 meters in the vertical). The mean of the 15 normalized data points is then combined with the mean and standard deviation of the signal between 30 and 40 km and the probability that the mean signal in the boxcar is composed entirely of noise is calculated assuming the noise is Gaussian distributed. This probability is assigned to each of the points within the boxcar. The boxcar is moved vertically by one range gate and the process is repeated. At the top of the profile (12 km for our purposes), the temporal window is incremented by one minute and the procedure begins again from the first range gate. Eventually, each point in the time-height domain is assigned 15 separate probabilities and these 15 separate probabilities are multiplied to form an effective probability index at each data point. The result of this procedure can be seen in Figure 1b where the non-cloudy and cloudy regions have been effectively separated into two distinct regimes with the lower boundary between these two regimes representing cloud base. While the boundary between the cloudy and non-cloudy pixels is sharp there tends to be a transition in the regimes that spans several pixels. A threshold that correctly identifies the actual cloud base was determined empirically using Belfort ceilometer data while additional fine-tuning of the threshold was performed using whole sky images.

Acknowledgments

This research benefitted from many insightful discussions with Dr. James Mather and was supported by U.S. Department of Energy grant.

References

Clothiaux, E.E., M.A. Miller, B.A. Albrecht, T.P. Ackerman, J. Verlinde, D.M. Babb, R.M. Peters, and W.J. Syrett, 1995: An evaluation of a 94 GHz radar for remote sensing of cloud properties. *J. Atmos.Oceanic. Technol*. **12**:201-229.

Spinhirne, J.D., 1993: Micro pulse lidar. *IEEE Trans*. **31**:48-55.