

# Recent Results in Deriving Water Vapor Profiles from Remote Sensor Observations Using Kalman Filtering

Y. Han and E. R. Westwater  
CIRES, University of Colorado/  
NOAA, Environmental Technology Laboratory  
Boulder, Colorado 80303-3328

R. A. Ferrare  
Hughes-STX Corporation  
Lanham, Maryland

## Introduction

The remote sensing of water vapor profiles by a single instrument during all weather conditions remains an elusive goal. The best observations of vapor profiles are currently provided by Raman lidar, which measures mixing ratio with a vertical resolution of 75 m and a temporal resolution of 1 min up to a range of about 9 km at night. However, the Raman lidar is limited during daytime conditions and, like most optical and infrared instruments, does not penetrate most liquid-bearing clouds. At the present time, the most promising techniques for obtaining nearly continuous profiles of water vapor combines observations from several remote and *in situ* sensors.

At the Atmospheric Radiation Measurement (ARM) program's Cloud and Radiation Testbed (CART) central facility in Lamont, Oklahoma, several remote sensors are currently operated continuously. These instruments include a dual-channel microwave water vapor radiometer (WVR), a Fourier Transform Infrared Radiometer, several ceilometers, and Radio Acoustic Sounding Systems (RASS) at 915 and 50 MHz. Recently, Han and Westwater (1995) developed and applied a technique to WVR, RASS, and ceilometer data that achieved accurate results in cloudy conditions. This technique is being used operationally by ARM (Turner et al. 1997) to blend some of the ARM remote sensor data. Using Kalman filtering (KF), this mathematical technique can be extended to incorporate soundings from Raman lidar as well. During clear nighttime conditions, the soundings are little changed from those of the Raman; however, during cloudy and daytime conditions, the impact of the remaining sounders is substantial. In this investigation, we derive water vapor profiles by integrating data from a Raman lidar, WVR, RASS, and surface *in situ* instruments. The data collected during the FIRE II experiment, conducted in November and December 1991 in

Coffeyville, Kansas, are used for our preliminary test. The set of instruments was described by Melfi et al. (1989) and Han et al. (1994).

## Retrieval Algorithm

The measurements from which water vapor mixing profiles are derived consist of water vapor mixing ratio profiles from Raman lidar, surface temperature, pressure and humidity from *in situ* sensors, and integrated water vapor from the WVR. The measurements of RASS virtual temperature profiles are used to derive integrated water vapor and liquid and are also used to convert water vapor from mixing ratio to absolute humidity. Cloud base height, derived from a Raman or another lidar, is also used in retrievals. The vertical coordinate of the profile vector  $\mathbf{x}$  starts at the surface level. The second level is set at the lidar's first range gate. The remaining levels have adjacent intervals of 75 m, to be consistent with lidar's range gates. The top level is set at 10 km, a height above which the total amount of water vapor is negligible.

Under conditions when  $\mathbf{x}$  is partially measured by the lidar, surface instruments, and WVR, the retrieval of  $\mathbf{x}$  is an ill-posed mathematical problem. Additional information is required for such retrievals. In our retrieval algorithm, we introduce two such sources. One is the information contained in previous lidar measurements, and the other is the statistical information obtained from an *a priori* water vapor profile ensemble. The algorithm is a two-stage retrieval that is outlined in the following.

In the first stage, a KF technique (Gelb 1988) is applied to derive a state vector  $\mathbf{s}$ , which has the same vertical coordinates as  $\mathbf{x}$  but less range coverage. The measurement vector  $\mathbf{d}$ , on which the filter operates, comprises a vapor

mixing ratio at the surface and a lidar profile. The number of the elements in  $\mathbf{d}$  varies depending on the lidar measurements. The error of the surface vapor mixing ratio is estimated from the errors in the measurements of surface air temperature, pressure, and relative humidity. The error profile of lidar measurements is related to the statistics of the Raman water vapor return signal and is given with each lidar vapor mixing ratio profile. The measurement error covariance consists of these errors, and it is assumed that the off-diagonal elements are zero, which is equivalent to the assumption that the errors are uncorrelated with each other. The dimension of  $s$  depends on the following situation: when the maximum height of current lidar measurements is lower than those of all previous measurements, the dimension is the same as the previous; otherwise, the dimension is reset to that equal to the dimension of the new measurement vector, and the KF is re-initialized. Thus, the state vector  $s$  describes atmospheric water vapor from the surface to the maximum lidar range in the history of lidar measurements. In the first situation, the portion that is not measured by current observation is actually predicted from the past and the new measurements. The vector  $\mathbf{d}$  is linearly related to  $s$  (corrupted by errors) with coefficients contained in the so-called "observation matrix."

The evolution of  $s$  is assumed to be a first-order Markovian process and is characterized by a time-dependent transition matrix, which linearly relates the  $s$  vectors at two different times, and by the errors (assumed white noise) in the transition model. In our system, the time interval of the transition is 2 min. Since the variation of the atmospheric water vapor in 2 min is usually small in comparison with the estimated errors in the retrievals, we approximate the transition by advancing  $s$  without modification, which is equivalent to setting the transition matrix to unity. The transition errors may be estimated using a historic profile database. The Raman lidar itself may provide part of such a database if a large number of nighttime clear lidar measurements are collected. For this experiment, however, the collection of clear-sky lidar measurements was not statistically sufficient.

The database we used is a collection of 3-hourly radiosonde profiles from five observation stations at the central CART site during the Intensive Observation Period (IOP) conducted in April and May 1995. Over 700 radiosonde profiles are used for the error estimation. Assuming that the transition errors are time invariant and time uncorrelated, we estimated the errors by advancing each profile 3 h, calculating the mean-square difference between the advanced profiles and the profiles measured at that time, and then dividing the difference by the number of 2-min intervals in the advanced time period. In general, a certain time correlation between errors is

expected. Assuming a correlation time greater than 3 h, in the time intervals concerned, the correlations are likely to be positive, causing the estimate of the transition errors to be larger than those when the correlations are counted. Larger transition errors result in a larger error covariance of an *a priori* estimate, which is propagated from the last estimation. This effectively gives more weight to current measurements in the profile estimation. Existence of the correlation is also inconsistent with the assumption that the transition errors are white noise. However, our testing results have shown that our KF technique under this assumption works well for our system.

The recursive KF starts at an initial state and its error covariance. We use the first arrival of measurements and error covariance as the initial state. The state vector  $s$  and its covariance are propagated, according to the transition system, to the point when new measurements arrive. The propagated state vector and its covariance are seen as an *a priori* estimate. Using the error covariance matrices as weights, the *a priori* estimate is linearly combined with the new measurements that have been mapped to the  $s$  space by using the observation matrix. The error covariance of the new estimate is also calculated. The newest estimate and its covariance are used, on one hand, in the second stage of the retrieval process described in the following and, on the other hand, in the next process of KF estimation.

In the second stage, the estimate and error covariance of the profile obtained from KF, as well as integrated water vapor and its error from the microwave radiometer, comprise a new measurement vector and error covariance matrix. The measurement vector again is related linearly to the unknown profile by a new observation matrix. From an *a priori* ensemble of radiosonde data, a climatological mean of the water vapor profile and its covariance were derived. The climatological mean serves as an *a priori* estimate of the unknown profile. A covariance-weighted averaging is performed to combine the measurement vector and the statistical data. The result is our final estimate of the water vapor profile.

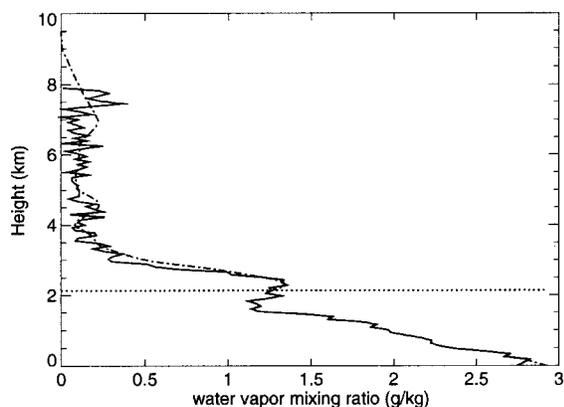
In the nighttime clear-sky cases in which the lidar covers the range close to the top of the retrieval vertical coordinate, the estimate from KF usually yields a much smaller error covariance in comparison with the statistical covariance. Hence, the averaging performed in the second stage is in favor of KF, resulting in an estimate differing little from that of KF. In other cases in which lidar measures only a portion of a profile, for the same reason as that in the clear-sky cases, the two-stage retrieval yields a profile with that portion similar to

the lidar measurements. The other part of the retrieved profile is a result of a balance among previous and current measurements and statistical data.

## Experiment Results

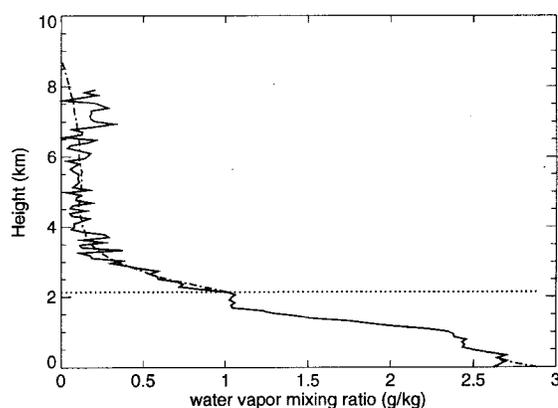
During the FIRE II experiment, the Raman lidar observed the atmosphere only during night. There were a total of 14 nights of observations, out of which there were 4 nights when low-level clouds with cloud base at about 2 or 3 km were observed. Unfortunately, during the cloudy periods when the lidar operated, there were no vapor profile observations by other instruments, such as radiosondes. Therefore, a direct comparison of retrievals with “ground truth” is not available. In order to demonstrate the system performance, we created “clouds” by extending existing cloud bases into clear periods. The Raman lidar soundings during these periods were then truncated at the artificial cloud bases and used as cloudy measurements. The original lidar soundings were used as ground truth.

We show here examples, which were obtained during the artificial cloudy period on November 24. Figure 1 shows a retrieval compared with ground truth, 16 min after the start



**Figure 1.** An example of retrieval using the technique described in the text. Dashed line, retrieval; solid line, ground truth (from lidar); dotted line, artificial cloud base. Data were collected during FIRE II. The cloud base is created by extending the existing cloud base into the clear period. The lidar sounding is truncated at the artificial cloud base and used as cloudy measurements. The original lidar sounding is used as ground truth.

of the cloudy period. As anticipated, the retrieved profile below the cloud base resembles the lidar measurement, and above the cloud base, the influence of the previous lidar measurements is obvious. Figure 2 shows an example in a situation when there are no historic lidar measurements available above cloud base height. This situation is likely to happen when clouds persist for a long period or during the daytime. The information for the profile above the cloud base is supplied by the statistical data and integrated water vapor, as well as the lidar measurement below the cloud through the correlations characterized by the error covariance matrix. In general, under such conditions, the profile above the lidar range is smoothed.



**Figure 2.** An example of retrieval when lidar provides no historic measurements for the portion above the cloud base. Other notations are explained in Figure 1.

## Summary

The Kalman filtering technique optimizes the use of information contained in past and current lidar measurements, surface *in situ* measurements, measurements of integrated water vapor, and statistical data. Under nighttime clear conditions, retrievals differ little from lidar measurements. Under cloudy or daytime conditions, the low portions of the retrieved profiles also differ little from the lidar measurements that cover those portions. The upper portions are constrained by the integrated water vapor measurements, and the profile structures are shaped by the previous lidar measurements and statistical data. This technique extends our capability to profile water vapor during cloudy and daytime conditions using data that are, or will be, taken at the ARM Southern Great Plains CART site.

## References

Gelb, A. (Ed.), 1988. *Applied Optimal Estimation*. The MIT Press, Cambridge, Massachusetts.

Han, Y., and E.R. Westwater, 1995. Remote sensing of tropospheric water vapor and cloud liquid water by integrated ground-based sensors. *J. Atmos. Ocean. Technol.* **12**, 1050-1059.

Han, Y., J.B. Snider, E.R. Westwater, S.H. Melfi, and R.A. Ferrare, 1994. Observations of water vapor by ground-based microwave radiometers and Raman lidar. *J. Geophys. Res.* **99**, 18695- 18702.

Melfi, S.H., D.N. Whiteman, and R.A. Ferrare, 1989. Observation of atmospheric fronts using Raman lidar moisture measurements. *J. Climate Appl. Meteor.* **28**, 789-806.

Turner, D., T. Shippert, J. Liljegren, Y. Han, and E. Westwater, 1997. Initial analysis of water vapor and temperature profiles retrieved from integrated ground-based remote sensors. Proceedings of the Sixth ARM Science Team Meeting, March 4-7, 1996, San Antonio, Texas (in press).