Noise Reduction of Atmospheric Emitted Radiance Interferometer (AERI) Observations Using Principal Component Analysis

D. D. TURNER, R. O. KNUTESON, AND H. E. REVERCOMB

University of Wisconsin—Madison, Madison, Wisconsin

C. LO

Pacific Northwest National Laboratory, Richland, Washington

R. G. DEDECKER

University of Wisconsin—Madison, Madison, Wisconsin

(Manuscript received 31 August 2005, in final form 11 January 2006)

ABSTRACT

A principal component noise filter has been applied to ground-based high-spectral-resolution infrared radiance observations collected by the Atmospheric Emitted Radiance Interferometers (AERIs) deployed by the Atmospheric Radiation Measurement (ARM) program. The technique decomposes the radiance observations into their principal components, selects the ones that describe the most variance in the data, and reconstructs the data from these components. An empirical function developed for chemical analysis is utilized to determine the number of principal components to be used in the reconstruction of the data. Statistical analysis of the noise-filtered minus original radiance data, as well as side-by-side analysis of data from two AERI systems utilizing different temporal sampling, demonstrates the ability of the noise filter using this empirical function to retain most of the atmospheric signal above the AERI noise level in the filtered data. The noise filter is applied to data collected at ARM's tropical, midlatitude, and Arctic sites, demonstrating that the random variability in the data is reduced by 5% to over 450%, depending on the spectral element and location of the instrument. A seasonal analysis of the number of principal components required by the noise filter for each site shows a strong seasonal dependence in the atmospheric variability at the Arctic and midlatitude sites but not at the tropical site.

1. Introduction

High-spectral-resolution infrared remote sensors provide a wealth of data on the atmospheric state (e.g., Feltz et al. 2003; Susskind et al. 2003), surface properties (e.g., Nalli et al. 2001; Knuteson et al. 2004a), clouds (e.g., Smith et al. 1998; Turner 2005), aerosols (e.g., Sokolik et al. 1998), trace gases (e.g., He et al. 2001), and other important atmospheric parameters. However, because of the need to have high-temporal-resolution data in rapidly changing scenes resulting from aircraft/spacecraft flight speed or advection of the atmosphere over the instrument, temporal averaging may not be an option to improve the signal-to-noise ratio of the radiance observations. Spectral averaging will improve the signal-to-noise ratio at the expense of the spectral resolution, and thus may not be a satisfactory option either, depending on the application.

One approach that has been proposed is to utilize a principal component analysis (PCA) noise filter (Huang and Antonelli 2001; Aires et al. 2002; Goldberg et al. 2003; Antonelli et al. 2004). This approach takes advantage of the high correlation among the different spectral channels of the high-spectral-resolution observations and allows a significant fraction of the uncorrelated random error in the observations to be removed, thereby improving the signal-to-noise ratio of the radiance observations. However, to our knowledge, the PCA noise filter approach has only been applied to downward-looking instrumentation such as the Scan-

Corresponding author address: Dr. David D. Turner, Space Science and Engineering Center, University of Wisconsin—Madison, 1225 West Dayton Street, Madison, WI 53706.
E-mail: dturner@ssec.wisc.edu

ning High-Resolution Interferometer Sounder (S-HIS) and the Atmospheric Infrared Sounder (AIRS), not to ground-based upward-looking observations.

PCA has been used a variety of different ways by the atmospheric science community to analyze high-spectral-resolution observations. For example, several investigators have used PCA-based approaches to retrieve thermodynamic profiles or cloud properties from these datasets (e.g., Zhou et al. 2005). An early example demonstrating the value of PCA is the work by Smith and Woolf (1976). Goody et al. (1998) and Huang et al. (2002) demonstrate how PCA can also be used to compare high-spectral-resolution observations with models.

In this paper, we apply the PCA noise filter approach (hereafter referred to as the noise filter) to ground-based Atmospheric Emitted Radiance Interferometer (AERI) data collected as part of the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) program (Ackerman and Stokes 2003). We utilize an empirical function to determine the optimal number of principal components to use in the reconstruction of the data. We demonstrate that this objective approach removes random uncorrelated error from the data; a negligible amount of atmospheric signal is lost during the filtering process. This noise filter has been applied to 2 yr of data from each of the three ARM sites, demonstrating that the number of principal components required by the noise filter is a function of instrument, location, and time of the year.

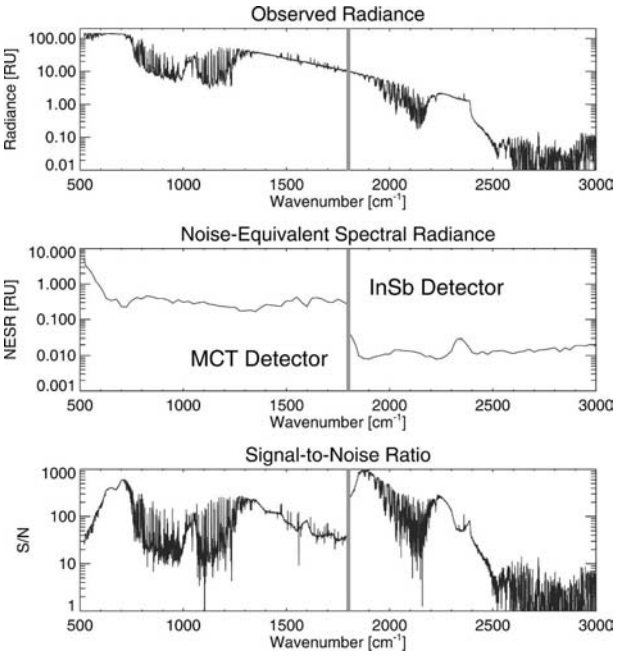


FIG. 1. Example of (top) a nighttime clear-sky radiance spectrum observed by the AERI at the SGP site, along with (middle) the noise-equivalent spectral radiance (NESR) uncertainty and (bottom) the signal-to-noise ratio. The data have been averaged for 3 min. The spectral bands of the two detectors utilized by the AERI are shown.

2. AERI

The AERI is an automated passive interferometer that measures downwelling infrared radiance with better than 1 cm^{-1} resolution. These instruments, which are built around a commercially available interferometer, utilize two well-characterized, high-emissivity blackbody targets running at 60°C and ambient temperature to provide accurate observations of the downwelling radiance. Accurate monitoring of the blackbody and structural temperatures, as well as correcting for instrument self-apodization and nonlinearity effects, result in an absolute radiometric calibration that is better than 1% (Knuteson et al. 2004c). The detector subsystem of the AERI contains both mercury cadmium telluride (MCT) and indium antimonide (InSb) detectors in a "sandwich" arrangement that provides sensitivity to both longwave (540–1800 cm^{-1}) and shortwave (1800–3000 cm^{-1}) radiation (henceforth referred to as the channel-1 and channel-2 bands, respectively). AERIs deployed at ARM's North Slope of Alaska (NSA) site have been modified to extend the sensitivity of the system to 400 cm^{-1} in order to make measurements in the rotational water vapor band, which becomes semitransparent at low water vapor amounts; however, the increased spectral range of the NSA AERI systems results in larger random error in the 800–1200 cm^{-1} window region relative to standard AERI systems. The radiometric uncertainty in the AERI measurements is estimated from the imaginary part of the calibration equation (Knuteson et al. 2004c). An example of an observed 3-min-averaged clear-sky spectrum, together with the associated single-sample noise estimate, is shown in Fig. 1. This noise estimate, which is derived from the sky scene, is slightly larger than the noise estimate derived from blackbody radiance spectra, as expected. Knuteson et al. (2004b) provide more details on the instrument's design and components, while additional calibration details are found in Knuteson et al. (2004c). ARM has AERIs deployed at its Southern Great Plains (SGP) site in north-central Oklahoma, at its tropical western Pacific (TWP) sites on Nauru Island and at Darwin, at the NSA site in Barrow, Alaska, and with its mobile facility.

ARM originally deployed AERIs to provide data that can be used to validate high-spectral-resolution radiative transfer models (Turner et al. 2004; Tobin et

al.1999) and from which temperature and water vapor profiles are retrieved (Feltz et al. 2003). For these applications, an approximate 8-min sampling strategy was used (3-min sky average with 5 min to view the two calibration blackbodies and for other housekeeping activities), which was a compromise between achieving a good signal-to-noise ratio and having enough temporal resolution to capture any changes in the *clear-sky* atmosphere (Knuteson et al. 2004b). However, the AERI data are also being used in retrievals of microphysical cloud properties (e.g., Mace et al. 1998; Comstock et al. 2002; Turner 2005; Turner and Holz 2005), and these microphysical properties can change quite rapidly (on the order of seconds) in the field of view of a vertically pointing narrow-beam instrument such as the AERI. An example of this phenomenon will be discussed in section 4a. However, the University of Wisconsin—Madison has recently demonstrated a "rapid sample" ("RS") capability by modifying the standard AERI control parameters so that each sky-view averaging period is significantly shorter and more sky views are performed before the blackbodies are viewed. This rapid-sample mode provides a sky spectrum approximately every 20 s instead of every 8 min, with periodic gaps when the instrument is viewing the calibration blackbodies. The sky averaging period for this prototype rapid-sample mode is approximately 12 s, as opposed to the 3-min sky averaging period used by ARM's standard AERI instruments. Three weeks of these prototype rapid-sample data were collected at the SGP site by a second AERI system from the University of Wisconsin—Madison that was deployed adjacent to the operational AERI in October through November 2003 in support of the Texas 2003 AIRS validation campaign. Rapid-sample data were also collected by two AERI systems deployed at the NSA site (the operational AERI at Barrow and the second unit deployed at Oliktok Point, Alaska) in support of the ARM Mixed-Phase Arctic Cloud Experiment (M-PACE) in September through October 2004 (Verlinde et al. 2005). These data, while extremely useful for cloud studies because of their higher temporal resolution, also have significantly more random noise resulting from the shorter averaging time used to view each scene. Therefore, it is desirable to reduce the random error in the rapid-sample data to facilitate more accurate retrievals of cloud microphysics from the high-spectral-resolution observations.

3. PCA methodology

As discussed in Antonelli et al. (2004), PCA can be used to both compress the high-spectral-resolution data and to improve the signal-to-noise ratio in the data by identifying and removing the principal components associated with uncorrelated random error. However, the reduction of the random error is a lossy compression procedure, and thus care must be taken to ensure that the principal components that are removed carry only random error and not real atmospheric signal. We present an empirical method that is used in the chemical analysis community to determine the "proper" number of principal components to be retained in section 3b, and demonstrate that a negligible amount of atmospheric signal is lost during this filtering process using this method.

a. Decomposition and reconstruction

PCA, which is also referred to as principal factor analysis, factor analysis, eigenanalysis, and empirical orthogonal function analysis by other practitioners, has seen widespread use in many different disciplines since the advent of the computer. PCA can be described as "a multivariate technique for reducing matrices of data to their lowest dimensionality by the use of orthogonal factor space and transformation that yield predictions and/or recognizable factors" (Malinowski 2002). The first step in this analysis is to decompose the data matrix into its eigenvectors (principal components) and eigenvalues. Typically, the data matrix used in the PCA is the covariance matrix $\mathbf{C}$ of the t-by-n (number of temporal samples by number of spectral elements) data matrix $\mathbf{M}$, where $\mathbf{C}$ is computed as $\mathbf{C} = \mathbf{M}^{\mathrm{T}}\,\mathbf{M}$ and T denotes the transpose. This is done because the eigenvectors of $\mathbf{C}$ are the same as the eigenvectors of $\mathbf{M}$ (Malinowski 1977a), and the covariance matrix $\mathbf{C}$ is a symmetric matrix and thus is easier to analyze. The magnitudes of the eigenvalues indicate the amount of variance explained in the data matrix by each of the associated eigenvectors, while the eigenvectors provide an abstract orthogonal basis for the vector space spanned by the data matrix. The eigenvector basis can be rotated and/or stretched (i.e., transformed) to provide physical meaning to each of the eigenvectors (Malinowski 2002). For example, Huang and Antonelli (2001) showed that the first eigenvector (i.e., the eigenvector associated with the largest eigenvalue) was well correlated with the window channels in the observed spectra.

The decomposition of the matrix $\mathbf{C}$ is typically performed using the singular value decomposition (SVD) approach, because the SVD approach yields eigenvalues and eigenvectors directly and is numerically stable and well behaved (Press et al. 1992). Applying SVD to $\mathbf{C}$ yields

$$\mathbf{C} = \mathbf{UDV}^{\mathrm{T}}, \tag{1}$$

where $\mathbf{D}$ is a diagonal matrix with the diagonal entries being the eigenvalues; that is, $\mathbf{D}_{ii} = \lambda_i$, and the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{C}$. Each λ_i represents a portion of the total variance in the data, so the eigenvectors associated with larger eigenvalues are oriented along axes in the vector space that have more variance, and $\Sigma_{i=1}^{n}\lambda_i$ is the total variance in the data. Because the covariance matrix $\mathbf{C}$ is real valued and symmetric, then $\mathbf{U} = \mathbf{V}$ and the eigenvectors are real valued. Following the SVD, the eigenvalues, and their associated eigenvectors in $\mathbf{U}$, are rearranged such that $\lambda_1 \geq \lambda_2 \geq, \ldots, \geq \lambda_n$, where the n is the dimension of $\mathbf{C}$. Note that $\mathbf{U}$, $\mathbf{D}$, and $\mathbf{V}$ are also square matrices with dimension n. An example of the typical eigenvalues derived from AERI observations is shown in the top panel of Fig. 2.

After performing the above decomposition, the dimensionality of the reconstructed matrix $\hat{\mathbf{C}}$ is reduced if k eigenvectors are used in the reconstruction and $k < n$. This reconstruction is achieved by only using the k eigenvectors associated with the k largest eigenvalues as

$$\hat{\mathbf{C}}_{n,n} = \mathbf{U}_{n,k}\mathbf{D}_{k,k}(\mathbf{V}_{n,k})^{\mathrm{T}}. \tag{2}$$

b. Deducing the number of factors

The total variability in the observed data can be described by the following three components:

(total variance) = (common variance) + (specific variance) + (error variance).

The true variance in the data is the sum of the "common" variance, which is found in all (or most) of the data, and "specific" variance, which refers to signals that are statistically significant but only seen in a portion of the data. The "error" variance is the uncorrelated random error that we desire the noise filter to remove. The key to the PCA noise filter is to properly determine k to reduce the dimensionality of the reconstructed covariance matrix $\hat{\mathbf{C}}$ (i.e., $k < n$) and include all of the eigenvectors that are associated with real variability, as opposed to random error, in the signal.

Unfortunately, there is currently no analytical way to determine the optimal value of k from real data that contain random noise that removes the maximum amount of uncorrelated noise while retaining all of the real variability in the data (Malinowski 2002). Various statistical and nonstatistical techniques have been used to estimate the value of k to be used. For example, Goldberg et al. (2003) have used the reconstruction score (RSc), which is defined as

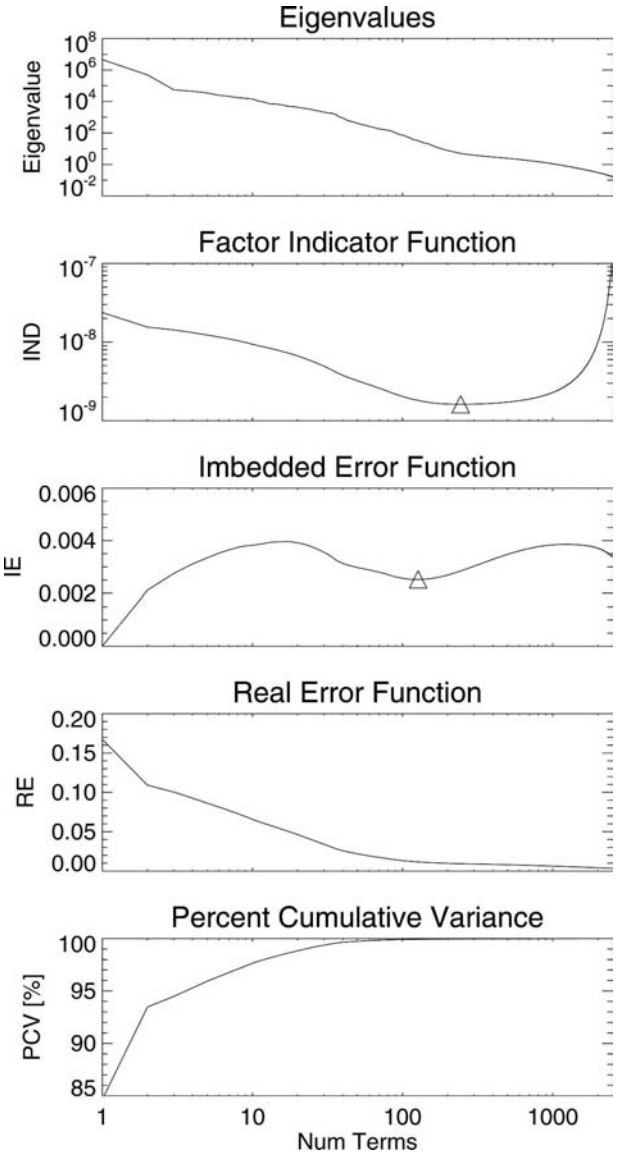


FIG. 2. The (sorted) eigenvalues, IND, IE, RE, and PCV functions, each computed for $k = 1$ to n, derived from the RS AERI data collected at the SGP from 24 to 28 Oct 2003. The triangles on the IND and IE panels indicate the optimal value of k that would have been selected from each.

$$\mathrm{RSc}(k) = \left[\frac{1}{n}\sum_{i=1}^{n}[I_i - R_i(k)]^2\right]^{1/2}, \tag{3}$$

where (as above) n is the number of spectral elements (channels) in the observations, and I_i and R_i are the noise-scaled observed and reconstructed radiances, respectively, where the latter have been reconstructed using k eigenvectors. In their analysis of AIRS data, the value of k used in the reconstruction is the lowest value of k (i.e., the smallest set of eigenvectors) needed such

that RSc < 1. There are two disadvantages to this technique. First, the data need to be reconstructed from the principal components multiple times for $k = 1$, $k = 2$, etc., until RSc(k) < 1. Second, RSc is a monotonically decreasing function that starts to lose sensitivity as the number of eigenvectors used in the reconstruction approaches k.

A second approach to determine the value for k is the percent cumulative variance (PCV) method

$$\mathrm{PCV}(k) = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}, \tag{4}$$

whereby k is determined such that the PCV is larger than some threshold (e.g., PCV > 0.999). While PCV is a useful quantity, the threshold is usually defined arbitrarily, which led Antonelli et al. (2004) to discard this method as a way to determine k. An example of the PCV computed from AERI observations is shown in Fig. 2.

The chemical analysis community has long used PCA techniques to reduce the dimensionality of their datasets, and a rich set of tools has been developed. Malinowski (1977a) has shown that all eigenvalues carry some component of the error, and should be separated into the following two groups:

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{k} \lambda_i^x + \sum_{i=k+1}^{n} \lambda_i^0, \tag{5}$$

where the λ^x are associated with the eigenvectors that contain a mixture of true signal and random error (i.e., the correlated error) and λ^0 with the eigenvectors consisting of the uncorrelated random error only. Malinowski (1977a) then showed that the real error (RE) is related to the imbedded error (IE) in the eigenvectors associated with λ^x and the extracted error (XE), which will be removed in the noise filtering process, as

$$(\mathrm{RE})^2 = (\mathrm{IE})^2 + (\mathrm{XE})^2. \tag{6}$$

All three of these error terms can be computed from the sorted eigenvalues as

$$\mathrm{RE}(k) = \left[\frac{\sum_{i=k+1}^{n} \lambda_i^0}{t(n-k)} \right]^{1/2}, \tag{7}$$

$$\mathrm{IE}(k) = \left[\frac{k \sum_{i=k+1}^{n} \lambda_i^0}{tn(n-k)} \right]^{1/2}, \tag{8}$$

$$\mathrm{XE}(k) = \left[\frac{\sum_{i=k+1}^{n} \lambda_i^0}{tn} \right]^{1/2}, \tag{9}$$

where, as above, t is the number of temporal samples, n is the number of spectral elements, and $t > n$. Malinowski (1977a, 2002) has shown that theoretically the RE is the difference between the pure data with no experimental error and the original data that contained the error, and that the XE is the difference between the reconstructed data and the original data (i.e., the amount of error that is removed by the PCA noise filter). The IE is the amount of correlated error that remained in the data after reconstruction. One method to determine the appropriate number of eigenvectors to use in the reconstruction is to find the minimum of IE(k), where IE is computed for each k from 1 to n. Examples of the RE and IE in an AERI dataset are shown in Fig. 2.

However, Malinokowski (1977b, 2002) pointed out that there are some cases where the IE does not have a well-defined minimum, because the PCA tends to exaggerate the nonuniformity of errors by placing too high of an emphasis on outlying data points. He has developed another empirical function, called the factor indicator function (IND), which can be used to deduce the dimensionality, and hence the optimal value of k, for the data matrix. The IND is computed as

$$\mathrm{IND}(k) = \frac{\mathrm{RE}(k)}{(n-k)^2}. \tag{10}$$

Malinowski (1977b, 2002) has demonstrated that when the error is uniform throughout the dataset and random (i.e., the noise is Gaussian distributed with a constant standard deviation for all spectral elements and times), the minimum of IND yields the correct value for k. An example of the IND for AERI data is also presented in Fig. 2.

c. *Applying the PCA noise filter to AERI observations*

We follow the seven steps outlined in Antonelli et al. (2004) to apply the noise filter to the AERI observations. These steps are summarized as follows: 1) normalize each observed spectrum by dividing it by the estimated noise-equivalent spectral radiance (NESR) derived from the imaginary component of the calibration equation; 2) generate the covariance spectrum $\mathbf{C}$ of the normalized observations; 3) derive the eigenvectors and eigenvalues using the SVD method; 4) determine the optimal number of eigenvectors to use in the reconstruction; 5) project each spectrum onto the vector

space spanned by this optimal number of eigenvectors; 6) reconstruct the data using these projection coefficients and the optimal number of eigenvectors, thus reducing the uncorrelated random error in the reconstructed observations; and 7) multiply each observation by the same noise spectrum used in step 1 to return the data back into radiance space.

As indicated in Antonelli et al. (2004), the normalization of the observed radiance by the NESR allows the PCA to fit the signal variability as opposed to the performance of the detectors, which manifests itself in the NESR spectra. Also, by applying the noise filter to the radiance normalized by the NESR (i.e., the signal-to-noise ratio), the random error in this ratio is assured to be uniform spectrally, and thus the IND method should yield the correct value of k. Nonetheless, because the AERI uses two different detectors to achieve the desired spectral range and each detector has its own unique noise characteristics (Fig. 1), we apply the noise filter independently to the channel-1 and -2 data.

Both the channel-1 and -2 data have approximately 2500 spectral elements. If the number of samples t is less than the number of spectral elements n, then the noise can affect the orientation of the principal components, which then affects the ability of the noise filter to remove the noise (Antonelli et al. 2004). Therefore, we require $t > 2n$ for when we apply the noise filter to AERI observations. Typically, we process 1 month of nominally sampled (i.e., 8-min resolution) AERI data at a time ($t \sim 5000$ samples).

Because the ARM AERIs are automated instruments, we do not know the range of atmospheric states that the AERI observed for any given time period. Thus, care has to be exercised to ensure that the noise filter does not remove cases of interest that are not well represented in the dataset. As pointed out by Antonelli et al. (2004), the use of a "dependent" PCA noise filter (i.e., the use of the data that we wish to noise filter for the generation of the eigenvectors) is critical to include observations that are not well represented in the set of observations. Therefore, we uniquely determine the set of eigenvectors directly from each dataset to which we are applying the noise filter. Additionally, because we desire to minimize the amount of atmospheric information loss, we conservatively select the number of eigenvectors k to use in the reconstruction. We have analyzed multiple years of data from the SGP, TWP, and NSA sites, and have found that the value of k suggested by the IE function is always less than the value suggested by the factor IND function. Also, in section 4, we will demonstrate that there is some correlation between different spectral elements in the noise-filtered (NF) minus original data residuals when k is determined from the IE function, which suggests atmospheric information is being removed, but not when k is determined from the IND function. Because the noise filter processing will be performed routinely in an automated fashion by the ARM program, our goal is to achieve the best noise reduction possible without introducing any artifacts in the data [such as the atmospheric information loss seen in Antonelli et al. (2004)]. Therefore, in our processing we reconstruct the data using the more conservative value of k suggested by the IND method.

4. Results

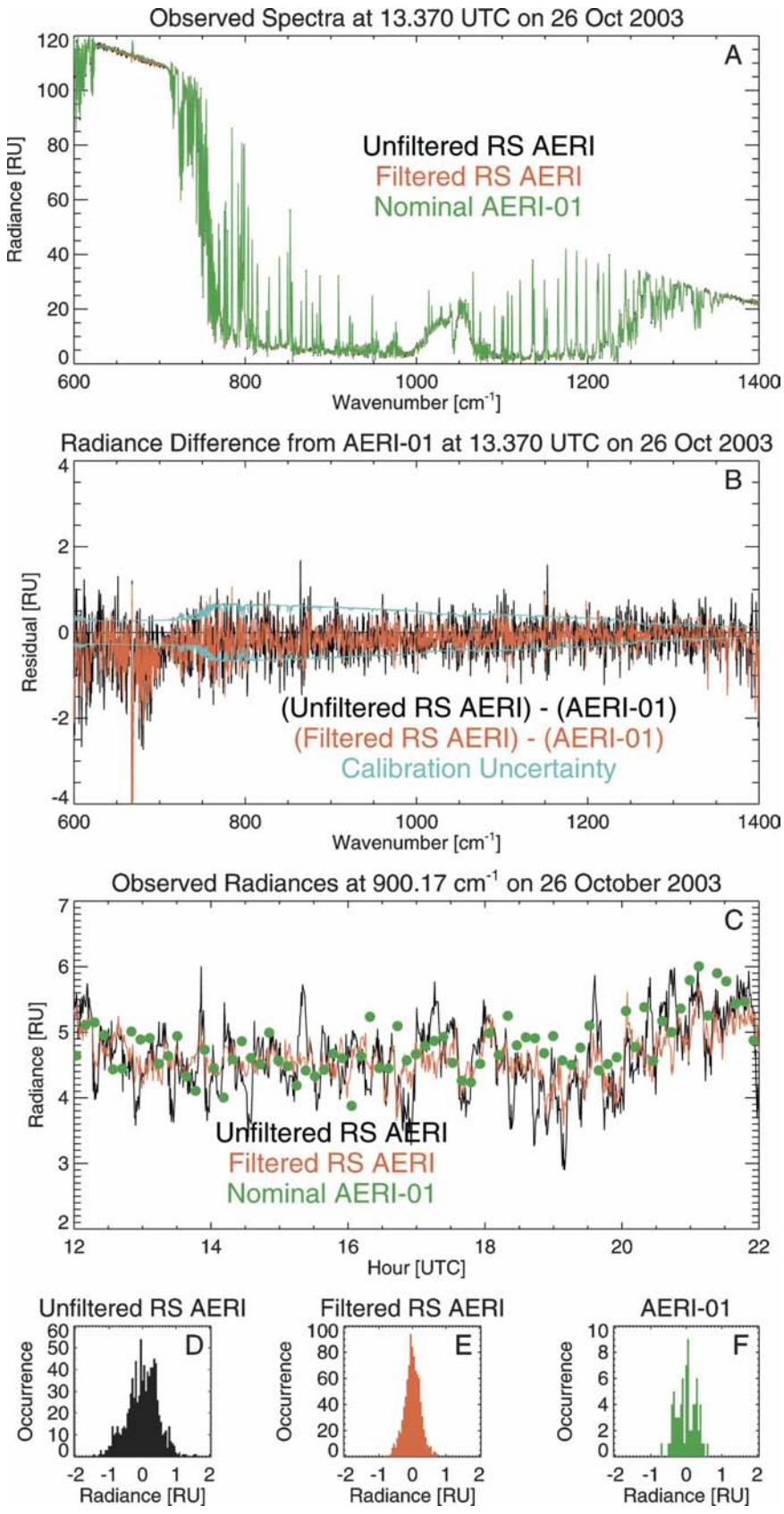
We have applied the PCA noise filter, using the dependent-set approach with k determined from the IND function, to AERI data collected at the SGP, NSA, and TWP sites. This section explores some of these results to indicate the performance of the noise filter.

a. Comparison using coincident observations

Evaluating the adequacy of the noise filter on real data, as opposed to simulated data where the "true" signal is known, is challenging with only a single instrument. In October–November 2003, the University of Wisconsin—Madison located their mobile AERI system at the ARM SGP site approximately 150 m from the operational ARM AERI-01 system. This mobile system was running in an RS mode, collecting 12-s averages of sky radiance every 20–25 s, as opposed to the 3-min averages of sky radiance collected every 8 min by the AERI-01. The mobile system was run in RS mode to demonstrate the utility of the RS observations in cloudy scenes and to collect a dataset that could be used to evaluate the noise filter algorithm relative to the observations of the operational AERI-01.

→

FIG. 3. (a) Observed spectra from the AERI-01 (unfiltered) and the unfiltered and noise-filtered RS AERI during a clear-sky period on 26 Oct 2003 at the SGP site. (b) Radiance residuals from the spectra in (a), along with the radiometric calibration uncertainty in the AERI observations. (c) The observed radiances from the AERI-01 and the unfiltered and filtered RS AERI at 900 cm^{-1} during the clear-sky period. Histograms of the detrended radiance observations from (c); the standard deviations for each of these distributions is (d) 0.461, (e) 0.243, and (f) 0.278 RU, respectively.


Observed Spectra at 13.370 UTC on 26 Oct 2003
A
Unfiltered RS AERI
Filtered RS AERI
Nominal AERI-01
Radiance [RU]
Wavenumber [cm-1]
Radiance Difference from AERI-01 at 13.370 UTC on 26 Oct 2003
B
(Unfiltered RS AERI) - (AERI-01)
(Filtered RS AERI) - (AERI-01)
Calibration Uncertainty
Residual [RU]
Wavenumber [cm-1]
Observed Radiances at 900.17 cm-1 on 26 October 2003
C
Unfiltered RS AERI
Filtered RS AERI
Nominal AERI-01
Radiance [RU]
Hour [UTC]
Unfiltered RS AERI
D
Filtered RS AERI
E
AERI-01
F
Occurrence
Radiance [RU]

The noise filter was applied to the RS data collected from 24 to 28 October 2003 (11 300 spectra). Figure 2 shows the eigenvalues as well as the IND, IE, RE, and PCV as a function of the number of terms used in the reconstruction. For this period, the IND function indicated that 250 eigenvectors should be used in the reconstruction of the channel-1 (5–19 μm) data.

To evaluate the adequacy of the noise filtering, comparisons of the unfiltered (UF) and filtered RS data were made with the AERI-01 observations in clear, stable conditions. Clear skies were selected to minimize the sampling differences between the two instruments; for example, we would expect to observe different radiance spectra using a 12-s sky average versus a 3-min sky average in a cloudy scene because of the inhomogenity that exists in all clouds. Figures 3a and 3b show an example of a single 12-s spectrum from channel 1, both unfiltered and noise filtered, from the RS AERI relative to the 3-min-averaged spectrum observed by the AERI-01 during a clear-sky period 26 October 2003. (The noise filter has not been applied to the AERI-01 data.) The spectral differences in Fig. 3b are within the radiometric uncertainties of the AERI observations, where these uncertainties are computed using the parameters in Table 3 of Knuteson et al. (2004c), with the exception of the CO_2 absorption band at 667 cm^{-1}; however, differences in surface air temperatures near the two instruments explain the residuals in this band. The spectral differences are also very similar in magnitude to comparisons of two collocated AERIs running with the same temporal resolution (e.g., Fig. 1 of Turner et al. 2004). The residuals in Fig. 3b demonstrate that the unfiltered RS data have larger differences from the AERI-01 observations than the noise-filtered RS data.

To quantify the reduction of the random error in the noise-filtered data, we examined the radiance at 900 cm^{-1} from the RS-unfiltered, RS-filtered, and AERI-01 data for a 10-h clear-sky period on 26 October 2003 (Fig. 3c). The data were detrended, where the trend was determined from a rolling 1-h average of the data, and histograms of the difference between the radiance observations at 900 cm^{-1} and the trend are shown for each (Figs. 3d–f). The difference in the standard deviations of the unfiltered and filtered RS radiance observations provide a measure of the amount of variability that the noise filter removed from the RS data. The standard deviation of the RS-unfiltered, RS-filtered, and AERI-01-detrended observations for this period was 0.461 RU, 0.243 RU, and 0.278 RU, respectively [radiance unit (RU), 1 RU = 1 mW $(m^2\ sr\ cm^{-1})^{-1}$]. Therefore, applying the noise filter reduced the variability in the RS data by approximately 50%, resulting in a slightly smaller amount of variability than was seen in the AERI-01 observations.

Clear-sky conditions will usually be well represented in any AERI dataset that spans from several days to months, and thus one would expect the reconstructed data to reproduce clear-sky scenes adequately. To evaluate how well scenes with clouds are reconstructed, we look at the following four different qualitative and quantitative measures: (a) that there are no discontinuities or general change in the character of the unfiltered-minus-filtered residuals (i.e., the extracted error) for any wavelength associated with clear/cloud boundaries; (b) that the reconstruction score given by Eq. (3) remains below approximately 1 for both clear and cloudy periods; (c) that the correlation between the extracted error from any two wavelengths is not statistically significant; and (d) that the cloud properties retrieved from the noise-filtered data agree well with the retrievals from the nominally processed data.

A wide range of cloud conditions were observed on 26 October 2003 at the SGP site—from overcast, opaque midlevel clouds from 0000 to 0130 UTC, to a thinning of these midlevel clouds from 0130 to 0300 UTC, to optically thin cirrus from 0300 to 0930 UTC, followed by clear skies for the rest of the day. A time series of brightness temperature at 900 cm^{-1} observed by the RS AERI is shown in Fig. 4, and the different cloud conditions can easily be identified in this plot. The extracted error at this wavelength is also shown in the figure, showing the residuals bouncing around zero in a random fashion; no obvious changes in the character of this residual is seen at the boundaries between different cloud types or clear/cloudy periods. The reconstruction score remains below unity for almost all of the samples during this period and, for the number of eigenvectors that were used in the reconstruction (k = 250), is insensitive to the presence or absence of clouds. The reconstruction score does increase slightly near the end of the day in the clear-sky conditions, but it always remains below 1.2. It should be noted that if a smaller number of eigenvectors was used in the reconstruction of the data for this day (e.g., k = 125 as suggested by the IE method), the reconstruction score becomes more variable (standard deviation is 0.150 for k = 125 as compared to 0.104 for k = 250), the mean value increases to 0.895 for k = 125 (compared to a mean value of 0.753 for k = 250), and there are numerous samples where the reconstruction score exceeds 1.2.

Because the noise filter is a lossy technique, it is very important to minimize the amount of information lost while preserving the best noise reduction possible. Antonelli et al. (2004) demonstrated that if too few eigenvectors are used in the reconstruction, then the ex-

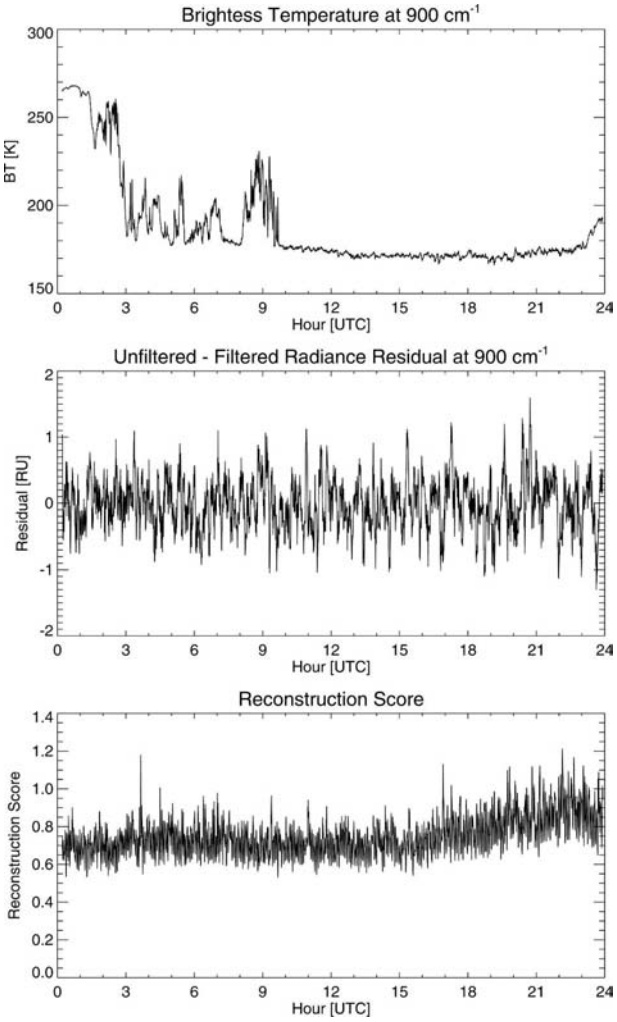


FIG. 4. (top) The observed brightness temperature at 900 cm^{-1} by the RS AERI on 26 Oct 2003 at the SGP site. (middle) The residual between the unfiltered and noise-filtered radiance observations at 900 cm^{-1}. (bottom) The reconstruction score of the data for $k = 250$.

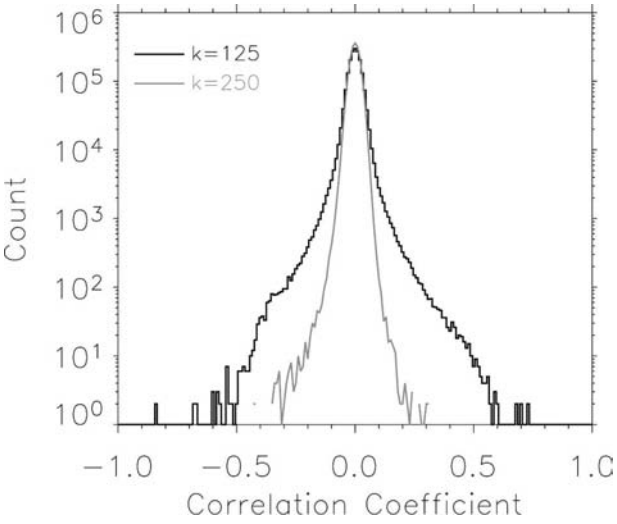


FIG. 5. Histogram of the correlation coefficients of the unfiltered minus noise-filtered residuals for all pairs of wavelengths from 520 to 1400 cm^{-1} for RS AERI data from 24 to 28 October 2003. The noise-filtered data were reconstructed twice: once using $k = 125$ as suggested by the IE function, and once with $k = 250$ as suggested by the IND function.

tracted error could be correlated with the observed radiance spectra (e.g., atmospheric absorption features may be present in the extracted error). However, the reason that the PCA noise filter works on high-spectral-resolution infrared radiance data is because many of the spectral channels are well correlated with other channels; thus, it stands to reason that if too few eigenvectors were used in the reconstruction and real atmospheric signals were being extracted, then there will exist a set of spectral elements in the extracted error spectra that have significant correlation. Figure 5 shows the correlation in the extracted error (i.e., the difference between the unfiltered and filtered residuals) for all pairs of spectral elements for wavelengths between 520 and 1450 cm^{-1} for RS data collected from 24 to 28 October 2003. If the reconstructed data used $k = 250$ eigenvectors as suggested by the IND function, then the absolute value of the correlation coefficient for almost all spectral pairs is less than 0.2. However, if the reconstruction uses $k = 125$ eigenvectors as suggested by the IE function, then there are hundreds of spectral pairs that exhibit significant ($|r| > 0.4$) correlations. This implies that if $k = 125$, then some real atmospheric signal is being removed by the noise filter, but if $k = 250$, a negligible amount of atmospheric signal is lost in the filtering.

The final "test" to ensure that the noise-filtering process preserves information on clouds is to utilize the noise-filtered data to retrieve cloud properties and compare the results to other coincident observations/retrievals. We have focused on the cirrus case from 0800 to 0945 UTC 26 October 2003, shown in Fig. 6. The reflectivity from the 35-GHz cloud radar (Clothiaux et al. 2000) and linear depolarization ratio from the Raman lidar (Turner et al. 2002) at the SGP site demonstrate significant variability in the cirrus cloud during this period. The cirrus optical depth and effective radius of the ice particles were retrieved from both the noise-filtered RS AERI and AERI-01 data using the mixed-phase cloud property retrieval algorithm (MIXCRA) of Turner (2005). The utility of the RS data during this period is obvious, because the coarser temporal resolution of the AERI-01 is unable to capture the significant changes in the optical depth in the cirrus. These large deviations are almost certainly

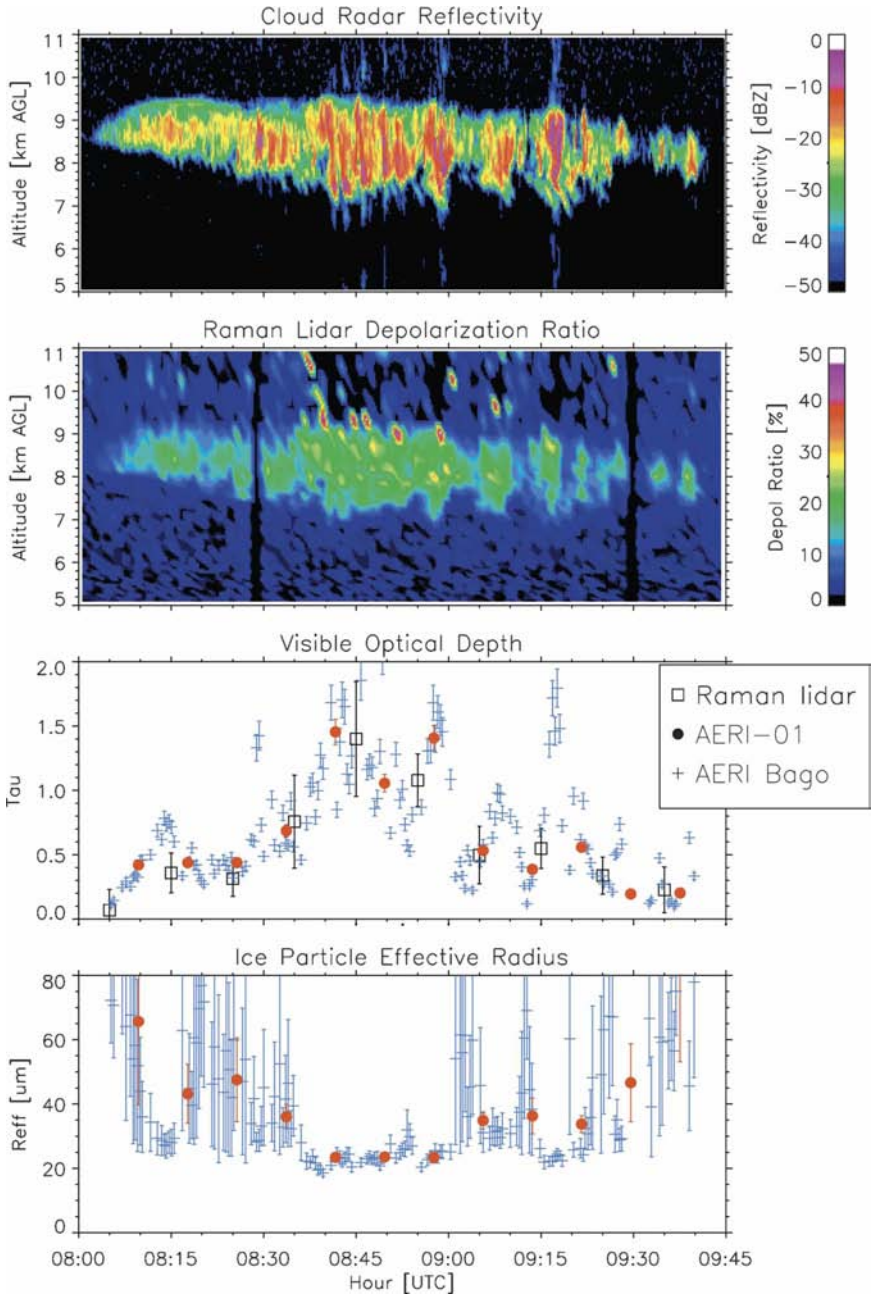


FIG. 6. Radar reflectivity and linear depolarization ratio from the cloud radar and Raman lidar, respectively, showing highly variable cirrus over the SGP site on 26 Oct 2003. The visible optical depth computed from 10-min averages of Raman lidar data and retrieved from the noise-filtered RS and unfiltered AERI-01 data show good qualitative agreement, although only the RS AERI is able to capture the high temporal variability in the optical depth field. The effective radius retrieved from the two AERI datasets is also in good qualitative agreement for the periods when the optical depth is above ~0.5, and thus the effective radius can be retrieved accurately.

real because they correlate very well with the magnitude of the reflectivity observations from the cloud radar. The retrieved optical depth from the noise-filtered RS data agrees very well with the optical depths retrieved from the AERI-01 and the 10-min-averaged Raman lidar data. For times when the optical depths are greater than approximately 0.5 and the effective radius can be retrieved (Turner 2005), the effective ra-

TABLE 1. Clear-sky (CS) results from the NF and UF channel-1 data in an opaque (675 cm^{-1}) and window (900 cm^{-1}) channels for several months from the NSA and SGP sites. The standard deviation (std dev) results are in mW (m^2 ster $cm^{-1})^{-1}$. The trends, computed with a 1-h moving average, have been removed from the window and opaque channels for both the NF and UF results before the statistics were computed; however, no trends were removed from the UF–NF data. The boldface numbers denote that NSA data from February and March 2004 were noise filtered simultaneously, because several weeks of data in February were missing due to instrument repair.

Site	Date	k	CS samples	UF window std dev	NF window std dev	UF – NF window std dev	UF – NF window skewness	UF opaque std dev	NF opaque std dev	UF – NF opaque std dev	UF – NF opaque skewness
NSA	Jan 2004	116	729	0.8372	0.1759	0.3708	−0.156	0.6921	0.3342	0.3002	0.628
NSA	Feb 2004	**92**	209	1.2048	0.1727	0.6314	0.780	0.9777	0.2998	0.4065	−0.925
NSA	Mar 2004	**92**	1024	0.9146	0.1578	0.4141	−0.240	0.7717	0.2958	0.3156	0.033
NSA	Apr 2004	126	1224	0.6997	0.1763	0.3203	0.007	0.5533	0.2144	0.2407	−0.041
NSA	May 2004	160	164	0.7409	0.2346	0.4689	0.095	0.6021	0.2207	0.2453	−0.307
NSA	Jun 2004	196	491	0.7463	0.2565	0.3673	0.142	0.4982	0.1920	0.1879	−0.110
SGP	Jan 2004	263	1212	0.2078	0.1189	0.0844	−0.110	0.2658	0.2221	0.0802	−0.099
SGP	Feb 2004	260	560	0.2015	0.1171	0.0792	0.094	0.3020	0.2763	0.0947	0.330
SGP	Mar 2004	298	728	0.2216	0.1199	0.0963	0.226	0.2622	0.2296	0.0735	−0.084
SGP	Apr 2004	309	226	0.2322	0.1255	0.1162	0.597	0.2623	0.2223	0.0895	−0.327
SGP	May 2004	326	213	0.2749	0.2133	0.0968	−0.154	0.2394	0.2014	0.0776	0.085
SGP	Jun 2004	328	94	0.3494	0.2638	0.1219	0.463	0.2625	0.2477	0.0787	0.487

dius retrieved from the two AERI systems agree well. It should be noted that if the RS data were not noise filtered, that the uncertainty in the retrieved optical depth and effective radius, which is provided by the optimal estimation technique used in MIXCRA, would be 30%–50% larger than the uncertainties from the noise-filtered RS values.

b. Magnitude of the noise reduction

The results from the RS AERI in section 4a suggested that the random variability in the signal, as indicated by the standard deviation of the detrended clear-sky data at 900 cm^{-1}, was reduced by a factor of 2. However, the following several questions can be asked. (a) Is the factor-of-2 reduction in the random variability what we should expect across the spectrum, in both opaque and window channels? (b) Are the results similar for both the channel 1 and channel 2? (c) Is there any site dependence?

To provide some insight into these questions, we analyzed data collected during stable clear-sky periods from January to June 2004 collected by the operational AERIs at the SGP and NSA sites. For each 1-month period, the noise filter was applied to the data for the entire month. We then identified the clear-sky periods by requiring that the observed brightness temperature at 900 cm^{-1} be below 155–210 K (depending on the month and site), and that the standard deviation of this brightness temperature in the 1-h window centered upon the sample of interest be less than 0.3–0.5 K (again, the actual value depended on site and month). We then selected periods where the clear-sky conditions persisted for at least eight consecutive hours.

From these clear-sky periods, we used a rolling 1-h mean to detrend the observations in a window and opaque channel in both channel 1 (900 and 675 cm^{-1}, respectively) and channel 2 (2500 and 2298 cm^{-1}, respectively), and then computed the standard deviation of the detrended data for each. These statistics are presented in Tables 1 and 2, and histograms of these data from April 2004 from the NSA site are shown in Fig. 7. The change in the standard deviation from the UF and NF data indicates the magnitude of the reduction of the random variability in each spectral channel. The magnitude of the reduction is both site and instrument specific. For example, at 900 cm^{-1}, the average reduction in the standard deviation of the detrended clear-sky data for this 6-month period resulting from the noise filter is 4.6 at NSA but only 1.6 at SGP, while at 675 cm^{-1} the average reduction in variance is 2.7 and 1.1 for the NSA and SGP sites, respectively. There is some monthly dependence in these statistics, however. For both sites, the reduction is larger in the window channels than in the opaque channels; this is true in both the channel-1 and -2 detectors. This phenomenon was expected, because Sromovsky (2003) has pointed out that the noise level in the atmospheric windows, especially in clear-sky scenes where the observed brightness temperatures are low, will be larger than opaque spectral regions because of the extrapolation applied by the calibration equation and the noise contributions from the calibration blackbodies.

The standard deviation of the detrended data represents real atmospheric variability in the detrended data as well as random error in the AERI observations. However, the differences between the unfiltered and

TABLE 2. Clear-sky (CS) results from the NF and UF channel-2 data in opaque (2298 cm^{-1}) and window (2500 cm^{-1}) channels for several months from the NSA and SGP sites. The standard deviation (std dev) results are in mW (m^2 ster $cm^{-1})^{-1}$. The trends, computed with a 1-h moving average, have been removed from the window and opaque channels for both the NF and UF results before the statistics were computed; however, no trends were removed from the UF–NF data. The boldface numbers denote that NSA data from February and March 2004 were noise filtered simultaneously, as several weeks of data in February was missing due to instrument repair.

Site	Date	k	CS samples	UF window std dev	NF window std dev	UF – NF window std dev	UF – NF window skewness	UF opaque std dev	NF opaque std dev	UF – NF opaque std dev	UF – NF opaque skewness
NSA	Jan 2004	118	729	0.0069	0.0016	0.0030	0.073	0.0093	0.0051	0.0033	0.135
NSA	Feb 2004	**121**	209	0.0064	0.0017	0.0025	0.177	0.0090	0.0036	0.0029	0.183
NSA	Mar 2004	**121**	1024	0.0068	0.0015	0.0026	0.240	0.0086	0.0030	0.0039	−0.118
NSA	Apr 2004	137	1224	0.0069	0.0018	0.0030	−0.047	0.0093	0.0038	0.0035	0.025
NSA	May 2004	146	164	0.0060	0.0015	0.0031	−0.022	0.0088	0.0040	0.0033	0.172
NSA	Jun 2004	157	491	0.0069	0.0022	0.0034	0.129	0.0087	0.0054	0.0034	0.012
SGP	Jan 2004	173	1261	0.0083	0.0021	0.0039	0.090	0.0124	0.0079	0.0052	−0.145
SGP	Feb 2004	173	560	0.0094	0.0023	0.0043	0.114	0.0130	0.0089	0.0061	0.099
SGP	Mar 2004	181	728	0.0094	0.0025	0.0043	0.360	0.0139	0.0102	0.0060	0.304
SGP	Apr 2004	189	226	0.0084	0.0022	0.0045	−0.079	0.0135	0.0111	0.0048	0.562
SGP	May 2004	194	213	0.0099	0.0027	0.0051	0.248	0.0143	0.0098	0.0060	−0.065
SGP	Jun 2004	193	94	0.0097	0.0030	0.0031	0.229	0.0157	0.0118	0.0069	−0.148

filtered data, provided that the filtered data have only had random error removed, provide a more direct way to look at the amount of random error that is removed by the noise filter each month. Tables 1 and 2 demonstrate that the standard deviation of the unfiltered-minus-filtered data is relatively constant for each month for each site. However, there are large differences between the standard deviations of the unfiltered-minus-filtered residuals for the channel-1 data from the SGP site relative to the NSA site, but the standard deviations for channel 2 are similar for the two sites. This suggests that the magnitude of the uncorrelated random error that is being removed from the NSA AERI channel-1 data is larger than that from the SGP system. This was anticipated, because the extension of the spectral range from a longwave limit of 540 cm^{-1} for the standard AERI to 400 cm^{-1} for the extended-range NSA AERIs results in a larger random error in the latter's channel-1 datasets (Knuteson et al. 2004b).

To investigate the sensitivity to the number of samples used in the determination of the principal component by the noise filter, we applied the noise filter algorithm to 3 months of data (March–May 2004) at NSA; the noise-filtered results were compared with the results when only data from April 2004 were used in the generation of the eigenvectors. For the 3-month dataset, the IND function indicated that the proper number eigenvectors needed for the reconstruction was $k = 140$, whereas for the 1-month dataset, the IND function indicated that $k = 126$. We attribute the slightly larger value of k for the 3-month dataset relative to the 1-month dataset to the need to have more eigenvectors to explain the atmospheric variability over the larger time window. However, the level of noise reduction, as gauged by statistics such as in Table 1, was nearly identical for the two datasets.

c. Dependence on environmental conditions

The number of eigenvectors required for reconstruction (k) by the IND function in Tables 1 and 2 suggest that there are significant differences between the SGP and NSA sites, and perhaps a seasonal dependence exists as well. To investigate this possible seasonal dependence of k, we applied the noise filter to nearly 2 yr of AERI data from the SGP, NSA, and TWP sites, where each month was processed individually. We required that each processing period have at least 5000 spectra; months with fewer samples than this (because of instrument downtime or other reasons) were grouped with an adjacent month before the noise filter algorithm was applied.

The results from the above analysis are presented in Fig. 8. The values of k determined from the IND method are significantly larger for the SGP and TWP sites relative to the NSA site. The smaller k values derived from the NSA data relative to the other two sites may indicate that the atmospheric variability observed by the AERI is smaller in the Arctic than in the other two sites, that the instrument noise in the NSA system is significantly larger than the TWP and SGP systems, or a combination of the two. The cloud fraction in the Arctic tends to be binary as either clear or overcast (i.e., broken cloud conditions do not often occur), and the clouds are typically low (Intrieri et al. 2002). However, cumulus clouds are frequently seen in

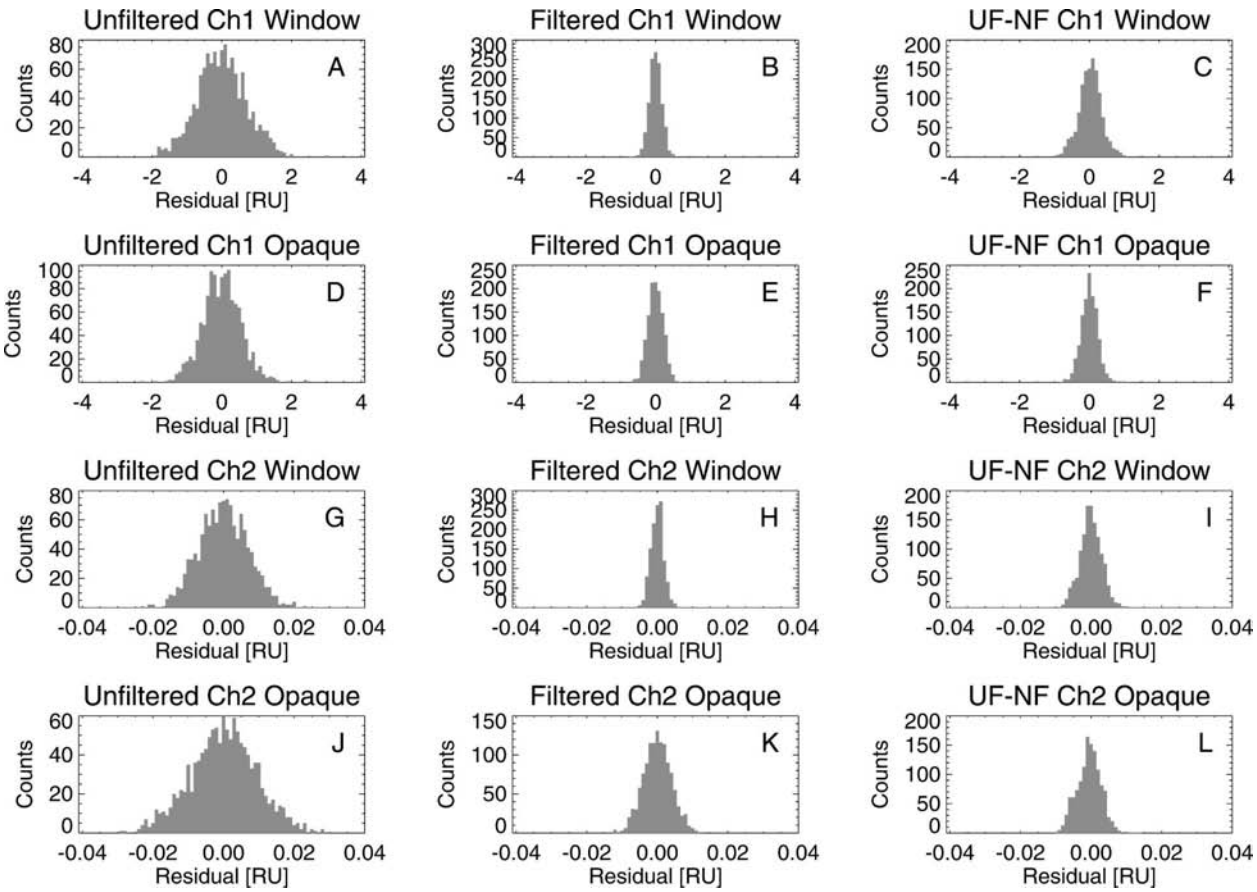


FIG. 7. Histograms of the (a), (d), (g), (j) unfiltered and (b), (e), (h), (k) noise-filtered clear-sky detrended AERI data collected at the NSA site in April 2004. See the text for details on how the dataset was selected. (c), (f), (i), (l) The differences between the UF and NF data are shown. This figure illustrates the impact of the noise filter for spectral elements in both window regions [(a)–(c) channel 1 and (g)–(i) channel 2] and opaque regions [(d)–(f) channel 1 and (j)–(l) channel 2). The exact wavenumbers for the window and opaque spectral elements are given in Tables 1 and 2.

the Tropics and midlatitudes, which can result in the AERI seeing a combination of clear and cloudy conditions during a sky-view period (e.g., Turner and Holz 2005), and there is a wide range of high-, mid-, and low-level clouds seen in both of these locations. Therefore, it is reasonable to expect that the atmospheric variability sampled by the AERI is smaller in the Arctic than at the SGP and TWP sites, which may partially explain the smaller values of k at the NSA site. However, Knuteson et al. (2004b) have indicated that the noise level is larger in the extended-range AERI systems that are used in the Arctic relative to the standard AERIs used at the other ARM sites. One way to compare the instrument noise level between different AERIs is to consider the magnitude of the variance of the radiance while looking at the hot blackbodies; such an analysis confirms that the noise level in the NSA AERI is 2–3 times larger in the 8–12-μm window than the SGP or TWP AERIs, which have similar instrument noise levels. Because the IND function is attempting to identify the number of eigenvectors that carry real information, if the noise level was increased across the spectrum, then the larger noise level would mask very small atmospheric signals in the higher-order eigenvectors. This results in a smaller number of eigenvectors being required to reconstruct the data.

This latter explanation (i.e., that a larger random error level results in a smaller value for k) is also supported by data collected during M-PACE. During M-PACE, the AERI at the NSA site was run in rapid-sample mode whereby each spectrum consisted of a 12-s average of sky radiance rather than the nominal 3-min average. This difference in the length of the averaging interval results in an increase in the random error of the rapid-sample data by roughly a factor of 4 relative to the nominally sampled data. When the noise filter was applied to the M-PACE data (the points that are circled in Fig. 8), the value of k returned by the IND

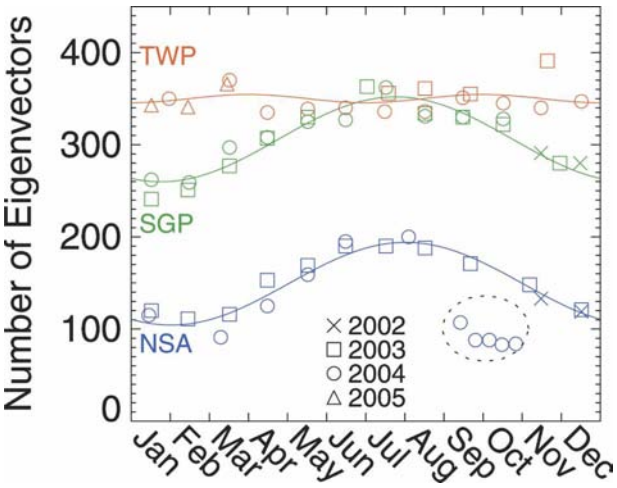


FIG. 8. The values of k derived by the IND function from nearly 2 yr of data, processed in 1-month periods, from each of the three ARM sites. A sinusoidal curve was fit to each of the datasets. The five NSA points circled with the dotted line were derived from 1-week long rapid-sample (12 s) data collected during M-PACE; all other data points were computed from data with nominal 3-min temporal resolution.

function was significantly lower than the values derived from the nominal data from previous years. Note that the M-PACE data were processed in 1-week intervals to maintain a relatively constant number of samples (~5000) per processing run to be consistent with the data processed from the nominally sampled AERI systems. The reduction in the temporal size of the sampling window (1 week versus 1 month) could be responsible for the change in k, because subsets of data require less eigenvectors to explain the variability during some arbitrary period of time (see section 4b). However, when all of the rapid-sample data collected during October were noise filtered as a single dataset, the derived value of k was almost exactly the mean of the k values derived from the weekly analysis.

The values of k determined from the data at the three ARM sites show a strong seasonal dependence in the SGP and NSA datasets, whereas the TWP data demonstrate negligible seasonal dependence. The peak variability, represented by the maximum in the sinusoidal curve fit to the k values, occurs approximately 10 days later at NSA than at the SGP (early August versus late July, respectively). Interestingly, the variability in the atmosphere captured by the spectral observations, represented by the value of k, during the summer at SGP is almost identical to the variability seen at the TWP site throughout its annual cycle. The seasonal variability of k observed in the SGP and NSA data is not explained by increased noise in the calibration due to the use of the hot/ambient calibration blackbody approach, because during the summer the warmer ambient conditions will decrease the temperature difference between the two blackbodies, which would increase the random error in the radiance observations, which would result in a smaller (not larger) value of k retrieved by the IND function. This is confirmed by considering the variance of the observed radiance when the instruments are viewing their hot blackbodies; these observations show no significant seasonal variability. Therefore, the seasonal variability seen in k appears to be due to real atmospheric variability and not an instrumental artifact; a more in-depth analysis of this phenomenon will be the subject of a future paper.

5. Summary

We have applied the principal component–based noise filter technique developed by Antonelli et al. (2004) to ground-based AERI observations of downwelling radiance. The technique decomposes a set of radiance observations that have a high correlation among different spectral elements into a set of principal components (eigenvectors), and then reconstructs the data using only a subset of the eigenvectors that explains the majority of the variance in the data. We have extended the Antonelli et al. (2004) technique by using an empirical function utilized by the chemical analysis community to determine the optimal number of eigenvectors to be used in the reconstruction of the data. We have demonstrated, via the use of reconstruction scores and by computing the correlation between spectral elements in the difference between the noise-filtered and original data, that this empirical technique provides a robust, objective way to determine the optimal number of eigenvectors to use in the reconstruction.

The ARM program has collected multiple years of high-spectral-resolution infrared radiance data from its AERI systems deployed at the SGP, TWP, and NSA sites. The data, which have been collected with a nominal sampling strategy of a 3-min sky average of data every 8 min, still have a significant random error component that can be reduced significantly (up to a factor of 4.5, depending on the spectral element in question as well as the location of the instrument); we plan to apply this noise filter to all of the AERI data in the ARM archive. Furthermore, ARM is moving into a new era with the AERI and will soon be collecting data in a rapid-sample mode whereby the sky averaging period will be decreased to approximately 12 s, allowing more frequent observations of sky radiance to be made. Data collected in rapid-sample mode is extremely useful in characterizing the variability of the microphysical cloud properties in cirrus and other optically thin clouds.

However, because of the decrease in the averaging time for the sky radiance, the rapid-sample data have a significantly larger random error component than the nominally sampled AERI data, and the noise filter algorithm described here will play an important role to reduce the random error in the rapid-sample observations.

Finally, the results from the noise filter algorithm also provide information on the variability of the atmosphere itself. By comparing the number of eigenvectors required for proper noise filtering, we have shown that the atmospheric variability, as observed by ground-based infrared radiometers, has a significant seasonal pattern at NSA and SGP sites, but there is no appreciable seasonal dependence at the TWP site. Furthermore, the peak in the seasonal variability is slightly later at the NSA site relative to the SGP site. These results are only the beginning: we plan on using the eigenvectors themselves to identify atmospheric conditions at the three ARM sites and to further quantify the differences seen between the sites.

Acknowledgments. We thank the staff in the high-spectral-resolution group at the University of Wisconsin—Madison for deploying the rapid-sample AERI at the SGP in 2003. We would also like to heartily thank the ARM operations staff, and especially the AERI instrument mentors, for maintaining the AERI and monitoring its data quality at all three sites. This research was supported by the U.S. Department of Energy's Office of Science (BER) as part of the Atmospheric Radiation Measurement program. Pacific Northwest National Laboratory is operated for the Department of Energy by Battelle under Contract DE-AC06-76RLO-1830.

REFERENCES

Ackerman, T. P., and G. M. Stokes, 2003: The Atmospheric Radiation Measurement Program. *Phys. Today,* **56,** 38–44.

Aires, F., W. B. Rossow, N. A. Scott, and A. Chedin, 2002: Remote sensing from the infrared atmospheric sounding interfereometer instrument 1. Compression, denoising, and first-guess retrieval algorithms. *J. Geophys. Res.,* **107,** 4619, doi:10.1029/2001JD000955.

Antonelli, P., and Coauthors, 2004: A principal component noise filter for high spectral resolution infrared measurements. *J. Geophys. Res.,* **109,** D23102, doi:10.1029/2004JD004862.

Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Moran, R. T. Marchand, M. A. Miller, and B. E. Martner, 2000: Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *J. Appl. Meteor.,* **39,** 645–665.

Comstock, J. M., T. P. Ackerman, and G. G. Mace, 2002: Ground-based lidar and radar remote sensing of tropical cirrus clouds at Nauru Island: Cloud statistics and radiative impacts. *J. Geophys. Res.,* **107,** 4714, doi:10.1029/2002JD002203.

Feltz, W. F., W. L. Smith, H. B. Howell, R. O. Knuteson, H. Woolf, and H. E. Revercomb, 2003: Near-continuous profiling of temperature, moisture, and atmospheric stability using the Atmospheric Emitted Radiance Interferometer (AERI). *J. Appl. Meteor.,* **42,** 584–597.

Goldberg, M. D., Y. Qu, L. M. McMillin, W. Wolf, L. Zhou, and M. Divakarla, 2003: AIRS near-real-time products and algorithms in support of operational numerical weather prediction. *IEEE Trans. Geosci. Remote Sens.,* **41,** 379–389.

Goody, R. M., J. Anderson, and G. North, 1998: Testing climate models: An approach. *Bull. Amer. Meteor. Soc.,* **79,** 2541–2549.

He, H., W. W. McMillan, R. O. Knuteson, and W. F. Feltz, 2001: Tropospheric carbon monoxide column density retrieval during the Pre-launch MOPITT Validation Exercise. *Atmos. Environ.,* **35,** 509–514.

Huang, H.-L., and P. Antonelli, 2001: Application of principal component analysis to high-resolution infrared measurement compression and retrieval. *J. Appl. Meteor.,* **40,** 365–388.

Huang, X., J. Farrara, S. S. Leroy, Y. L. Yung, and R. M. Goody, 2002: Cloud variability as revealed in outgoing infrared spectra: Comparing model to observation with spectral EOF analysis. *Geophys. Res. Lett.,* **29,** 1270, doi:10.1029/2001GL014176.

Intrieri, J. M., M. D. Shupe, T. Uttal, and B. J. McCarty, 2002: An annual cycle of Arctic cloud characteristics observed by radar and lidar at SHEBA. *J. Geophys. Res.,* **107,** 8029, doi:10.1029/2000JC000423.

Knuteson, R. O., F. A. Best, D. H. DeSlover, B. J. Osborne, H. E. Revercomb, and W. L. Smith Sr., 2004a: Infrared land surface remote sensing using high spectral resolution aircraft observations. *Adv. Space Res.,* **22,** 1114–1119.

——, and Coauthors, 2004b: Atmospheric Emitted Radiance Interferometer. Part I: Instrument design. *J. Atmos. Oceanic Technol.,* **21,** 1763–1776.

——, and Coauthors, 2004c: Atmospheric Emitted Radiance Interferometer. Part II: Instrument performance. *J. Atmos. Oceanic Technol.,* **21,** 1777–1789.

Mace, G. G., T. P. Ackerman, P. Minnis, and D. F. Young, 1998: Cirrus layer microphysical properties derived from surface-based millimeter radar and infrared interferometer data. *J. Geophys. Res.,* **103,** 23 207–23 216.

Malinowski, E. R., 1977a: Theory of error in factor analysis. *Anal. Chem.,* **49,** 606–612.

——, 1977b: Determination of the number of factors and the experimental error in a data matrix. *Anal. Chem.,* **49,** 612–617.

——, 2002: *Factor Analysis in Chemistry*. 3d ed. Wiley and Sons, 414 pp.

Nalli, N. R., W. L. Smith, and B. Huang, 2001: Quasi-specular model for calculating the reflection of atmospheric-emitted infrared radiation from a rough water surface. *Appl. Opt.,* **40,** 1343–1353.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in C: The Art of Scientific Computing*. 2d ed. Cambridge University Press, 994 pp.

Smith, W. L., and H. M. Woolf, 1976: The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiometer observations. *J. Atmos. Sci.,* **33,** 1127–1140.

——, S. A. Ackerman, H. E. Revercomb, H.-L. Huang, D. H. DeSlover, W. F. Feltz, L. Gumley, and A. Collard, 1998: In-

frared spectral absorption of nearly invisible cirrus clouds. *Geophys. Res. Lett.,* **25,** 1137–1140.

Sokolik, I. N., O. B. Toon, and R. W. Bergstrom, 1998: Modeling the radiative characteristics of mineral aerosols at infrared wavelengths. *J. Geophys. Res.,* **103,** 8813–8826.

Sromovsky, L. A., 2003: Radiometric errors in complex Fourier transform spectroscopy. *Appl. Opt.,* **42,** 1779–1787.

Susskind, J., C. D. Barnet, and J. M. Blaisdell, 2003: Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. *IEEE Trans. Geosci. Remote Sens.,* **41,** 390–409.

Tobin, D. C., and Coauthors, 1999: Downwelling spectral radiance observations at the SHEBA ice station: Water vapor continuum measurements from 17 to 26 μm. *J. Geophys. Res.,* **104,** 2081–2092.

Turner, D. D., 2005: Arctic mixed-phase cloud properties from AERI-lidar observations: Algorithm and results from SHEBA. *J. Appl. Meteor.,* **44,** 427–444.

——, and R. E. Holz, 2005: Retrieving cloud fraction in the field-of-view of a high-spectral resolution infrared radiometer. *Geosci. Remote Sens. Lett.,* **3,** 287–291.

——, R. A. Ferrare, L. A. Heilman Brasseur, W. F. Feltz, and T. P. Tooman, 2002: Automated retrievals of water vapor and aerosol profiles from an operational Raman lidar. *J. Atmos. Oceanic Technol.,* **19,** 37–50.

——, and Coauthors, 2004: The QME AERI LBLRTM: A closure experiment for downwelling high spectral resolution infrared radiance. *J. Atmos. Sci.,* **61,** 2657–2675.

Verlinde, J., and Coauthors, 2005: Overview of the Mixed-Phase Arctic Cloud Experiment (M-PACE). Preprints, *Eighth Conf. on Polar Meteorology and Oceanography*, San Diego, CA, Amer. Meteor. Soc., CD-ROM, P1.14.

Zhou, D. K., W. L. Smith, X. Liu, A. M. Larar, H.-L. Huang, J. Li, M. J. McGill, and S. A. Mango, 2005: Thermodynamic and cloud parameter retrieval using infrared spectral data. *Geophys. Res. Lett.,* **32,** L15805, doi:10.1029/2005GL023211.