A Principal Component Analysis Noise Filter Value-Added Procedure to Remove Uncorrelated Noise from Atmospheric Emitted Radiance Interferometer (AERI) Observations

January 2006

C. Lo
Pacific Northwest National Laboratory
Richland, Washington

D. D. Turner and R. O. Knuteson
University of Wisconsin-Madison,
Madison, Wisconsin

Summary

This technical report provide a short description of the application of the principle component analysis techniques to remove uncorrelated random noise from ground-based high spectral resolution infrared radiance observations collected by the atmospheric emitted radiance interferometers (AERIs) deployed by the Atmospheric Radiation Measurement (ARM) Program. A general overview of the technique, the input, and output datastreams of the newly generated value-added product, and the data quality checks used are provided. A more complete discussion of the theory and results is given in Turner et al. (2006).

Contents

Table

Figures

A Principal Component Analysis
Noise Filter Value-Added Procedure to
Remove Uncorrelated Noise from
Atmospheric Emitted Radiance Interferometer (AERI)
Observations

1. Introduction

The atmospheric emitted radiance interferometer (AERI) is a ground-based passive sensor that measures downwelling emitted spectral radiance from 500-3000 cm^{-1} at 1 cm^{-1} resolution (Knuteson et al. 2004a, b). An AERI is located at each of the Atmospheric Radiation Measurement (ARM) Climate Research Facility (ACRF) locales: Southern Great Plains (SGP), North Slope of Alaska (NSA), and Tropical Western Pacific (TWP). The AERI radiance observations at these three locales have been used to validate radiative transfer models (e.g., Tobin et al. 1999; Turner et al. 2004), retrieve thermodynamic profiles (e.g., Feltz et al. 2003), and investigate the microphysical properties of clouds (e.g., Turner 2005). The original temporal sampling strategy was a three-minute average of sky radiance every eight minutes; this sampling strategy was selected for the validation of clear-sky radiative transfer models and for atmospheric profiling (Knuteson et al. 2004a). However, cloud microphysical properties can change on the order of seconds in the narrow field-of-view of the AERI. Therefore, the ARM Program is in the process of improving the temporal resolution of the AERI to collect a sky spectrum every 15-30 seconds. The increased temporal resolution results in less averaging performed by the instrument, and hence the larger component of random noise in the sky spectra. A new value-added procedure (VAP) has been developed that uses the high correlation in the observed radiance across the spectrum to reduce the uncorrelated random error in the data using principal component analysis (PCA). The VAP automatically determines the appropriate number of principal components to use in the reconstruction to eliminate as much random noise as possible. A significant reduction in the uncorrelated random error in the data has been proven for both regular temporal data and rapid-sample (RS) data.

The details of this algorithm have been published in Turner et al. (2006). We refer the readers to this article for further information about this VAP.

2. Input Data

The input files for this VAP are standard ARM AERI data files in NetCDF format. We have run this VAP on data from the three ACRF locales: NSA (Barrow, Alaska), SGP (north-central Oklahoma), and TWP (Nauru Island and Darwin, Australia). Table 1 lists the various ARM datastreams used in the VAP for input data, along with the specific variables in files that are used in processing.

Table 1. NSA Input Files and the Key Variables			
Datastream	**Variable Name**	**Variable Long Name**	**Units**
xxxaerich1FF.b1 **xxxaerich2FF.b1**	base_time	base time in epoch	seconds since 1970/01/01 00:00:00
	time_offset	time offset from base_time	seconds since base_time
	time	time offset from midnight	seconds since midnight
	wnum	wave number in reciprocal centimeters	cm^{-1}
	mean_rad	downwelling radiance interpolated to standard wave number scale	$mW/(m^2\ sr\ cm^{-1})$
xxxaerisummaryFF.b1	SkyNENCh1	AERI LW scene NESR spectral averages (Ch1)	$mW/(m^2\ sr\ cm^{-1})$
	SkyNENCh2	AERI SW scene NESR spectral averages (Ch2)	$mW/(m^2\ sr\ cm^{-1})$
	LWskyNEN	The noise equivalent radiance observed in the longwave channel during a sky view at 1000 cm^{-1}	$mW/(m^2\ sr\ cm^{-1})$

Atmospheric Emitted Radiance Interferometer:

The names of the input files are as follows:

> xxxaerich1FF.b1.YYYYMMDD.hhmmss – 8 minutes (regular) or 20-second (RS) data
> xxxaerich2FF.b1.YYYYMMDD.hhmmss – 8 minutes or 20-second data
> xxxaerisummaryFF.b1.YYYYMMDD.hhmmss – 8 minutes or 20-second data

Where:

xxx	–	the locations of the instrument (nsa, sgp, twp)
aeri	–	the main instrument name
ch1	–	Channel 1, sensitive to radiation from 5-25 μm
ch2	–	Channel 2, sensitive to radiation from 3-5 μm
summary	–	estimating the amount of noise across the spectrum (spectral resolution of 25 cm^{-1})
FF	–	facility (e.g., C1)
b1	–	data level, can be a1
YYYY	–	year, MM - month of the year, DD - day of the month, hh - hour of the day, mm - minute of the hour, ss - second of the minute of data start

   This VAP is defined as "pass-through," which means that all of the fields in the input files are propagated to the output files. The VAP will replace the primary data field "mean_rad," which contains the radiance data, with the noise-filtered radiance data in the output data files.

3. Output Data

This VAP is defined as "pass-through," which means the output data, including its file structure, is almost identical to the original input data file. However, there are two data fields subject to change. The primary data field from the AERI (the field "mean_rad") in the output file contains the noise-filtered radiance, and the qc_time field variable data from the input file were recalculated because of the time samples are synchronized among three input files and its attributes were rewritten to comply with the new quality control standards. All other fields in the output file contain the original values in the input file. Some global attributes are modified and/or deleted from the input files because of the inappropriate statement for ingest.

The output file does contain one additional field that is not in the input file. This field contains the eigenvalues that were derived from the data. In addition, several new global attributes were added to the output file to capture version information and other parameters that capture the state of this VAP.

The name of the output files is as follows:

xxxaerich1nf1turnFF.c1.YYYYMMDD.hhmmss
xxxaerich2nf1turnFF.c1.YYYYMMDD.hhmmss

– 8 minutes or 20-second data
One file is generated for each channel per day.

Where:

xxx	–	the location of the instrument (nsa, sgp, twp)
aeri	–	the main instrument name
nf1turn	–	identifies that this is Turner's version 1 AERI noise-filter VAP
FF	–	facility (e.g., C1)
YYYY	–	year, MM - month of the year, DD - day of the month, hh - hour of the day, mm - minute of the hour, ss - second of the minute of data start

This Value Added Procedure generates one quicklook plot; an example is shown in Figure 1. The name of the quicklook files is aerinf1turn_YYYYMMDD_quicklook.png

4. Principal Component Analysis and Determination of the Number of PCs

Principal Component Analysis (PCS) is a useful statistical technique that can be used to compress data via an orthogonal factor space. It has been widely used in the chemical analysis

and atmospheric science communities to reduce high-dimension data. For high-spectral-resolution infrared radiance data like that collected by the AERI where many of the spectral channels are well correlated with other channels, PCA methodology can be used to filter out uncorrelated random error without introducing any significant artifacts in the data such as the atmospheric information loss. PCA decomposes the data into a set of orthogonal vectors, or principal components, where each component explains some fraction of the total variance in the dataset. The VAP then determines the number of principal components to use in the reconstruction of the data; by neglecting to use all of the principal components, the reconstructed (noise-filtered) data will have a smaller random error component. The details of the noise-filtering process, including the determination of the number of principal components to use in the reconstruction, are presented in Turner et al. (2006).

The noise-filtered results from a mobile system running in rapid-sample mode at the SGP site have been compared with coincident observations with the operational AERI-01. These comparisons are also found in Turner et al. (2006).

5. Quality Control of the Input Data

To remove the uncorrelated random error and not the atmospheric signal, we require that the number of the time samples used in the filtering process be greater than two times the number of spectral elements (and preferably much larger). However, computer memory capacity limits how large of a window of AERI data can be noise-filtered in a single run. We have investigated the impact on the noise-filtered results to different sizes of the processing period (i.e., noise-filtering one month of nominally sampled data versus noise-filtering three months of nominally sampled data); we found that little sensitivity exists as long as the number of temporal samples is at least 5000 (because each channel has 2500 spectral elements). Therefore, we recommend that the nominally sampled data (i.e., the three-minute sky averages every eight minutes) be noise-filtered in one-month periods (i.e., run the VAP on one month of data at a time), and that the RS data be run on ten-day periods

To properly noise-filter the radiance data, we must account for the response of the detector. The AERI data files routinely provide a measured noise-equivalent spectral radiance (NESR) spectrum for each AERI sample; these spectra are stored in the "summary" data files. Therefore, the VAP reads in the Channel 1, Channel 2, and summary data files for each run. Because of how the AERI instrument operates, it is possible that these three data files will not have the same number of samples. Therefore, the VAP first performs time synchronization so that only the common time samples among those three types of files are kept.

The second step is to apply some simple quality control to the radiance spectra, as bad spectra can significantly skew the results of the PCA and thus the noise-filter data will be suboptimal. Two circumstances where the data are considered bad are as follows:

- spectra in which the mean observed radiance at 900 cm^{-1} is less than -5 mW/(m^2 ster cm^{-1}) or above 170 mW/(m^2 ster cm^{-1}) (the latter corresponds to a brightness temperature of approximately 55°C)

- samples in which the total longwave sky noise-equivalent radiance (captured by the field "LWskyNEN") is larger than 7 mW/(m^2 ster cm^{-1}) for nominally sampled data or 25 mW/(m^2 ster cm^{-1}) for RS data. If the sample is identified as bad, then it is removed from the dataset before the noise filter is applied.

6. Examples

We have applied the noise-filter to AERI data from the different ACRF locales for several years. Turner et al. (2006) presents examples drawn from this dataset. We provide additional examples in this report.

Figure 1 shows the time series of the original and noise-filtered radiance at 900 cm^{-1} of RS data for a (mostly) clear-sky day at the NSA locale. The noise-filtered radiance (blue) has less random fluctuation that indicates the removal of the random error. The spike at 18:55 Universal Time Coordinates (UTC) is associated with a real signal; it is not a noise spike.
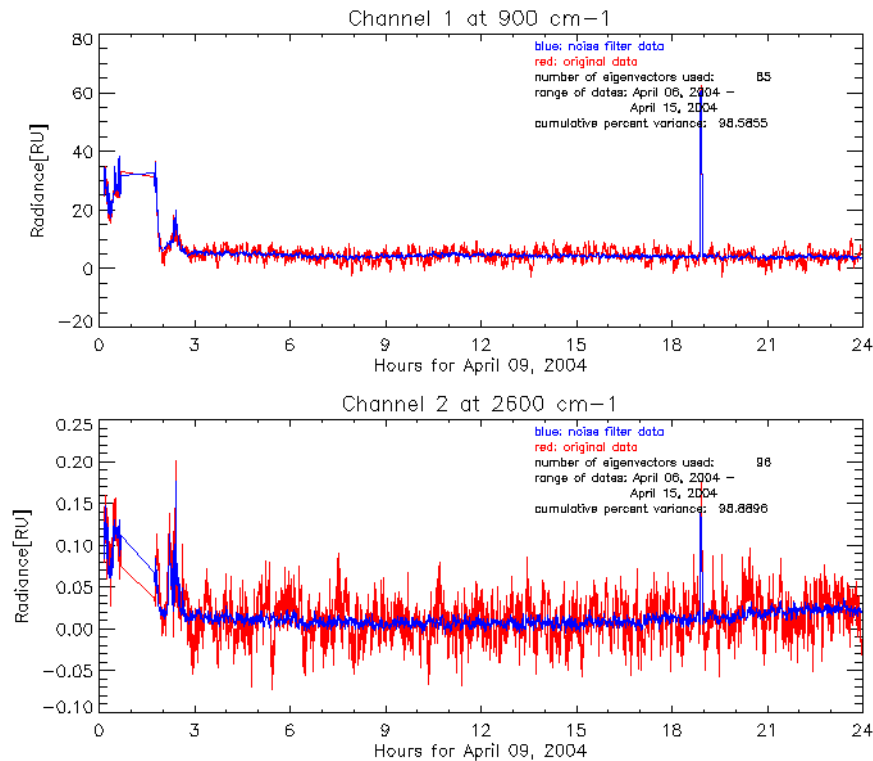
Figure 1. Quicklook image showing the time series of the radiance in window channels for Channel 1 (900 cm^{-1}) and Channel 2 (2600 cm^{-1}) in RS mode for NSA S01 data.

Figure 2 shows a Channel 1 spectrum from the nominally sampled AERI data on the first sample of January 1, 2004, at the NSA locale. In the middle panel, the residual between the noise-filtered (blue in the top panel) and original radiance spectrum (red in the top panel) shows only random differences; we observe no biases in any portion of the spectrum. In the bottom panel of Figure 2, the distribution of the differences between the noise-filtered and original radiances is Gaussian, suggesting that the differences between the original data and the noise-filtered data are due to the removal of normally distributed random error that was removed from the latter dataset.
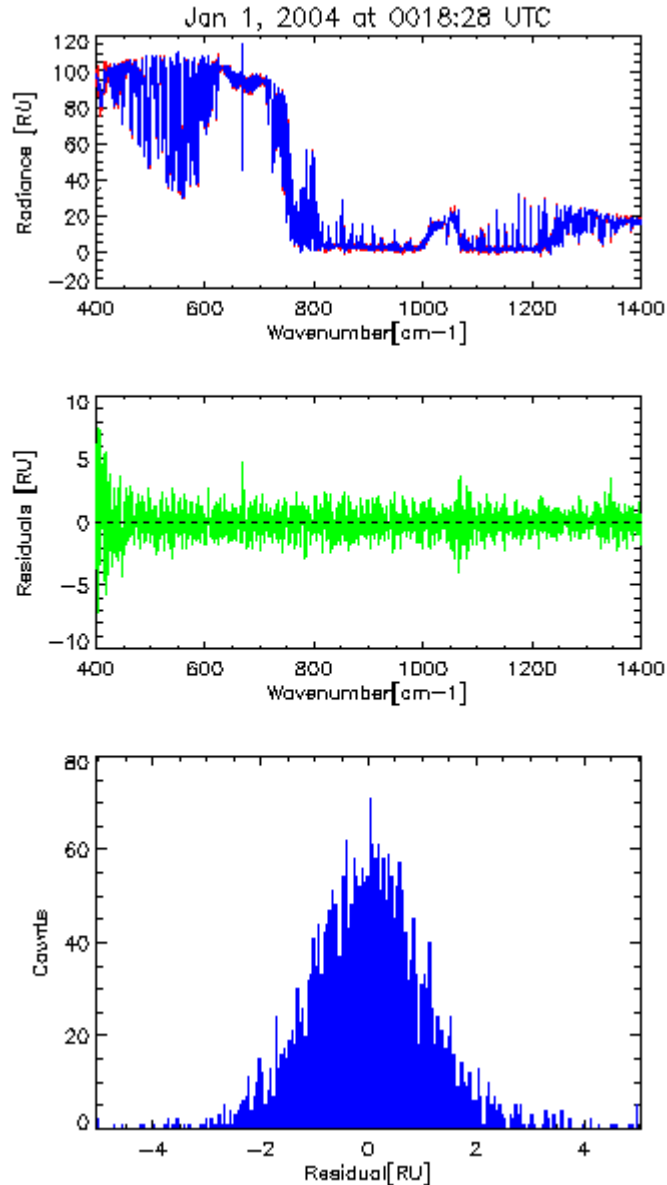
Figure 2. A nominally sampled spectrum for NSA C1 on January 1, 0018:28 UTC for filtered and unfiltered data and their difference (residuals).

6

7. Conclusions

The technique decomposes the radiance observations into its principal components, selects the ones that describe the most variance (with the largest eigenvalues) in the data, and reconstructs the data from these components. Principal components with small eigenvalues are typically associated with random error and are usually removed by the noise filter algorithm before the data are reconstructed. We use an empirical function to determine the proper number of principal components.

The results show significant reduction of the random error component of the radiance data for both RS and nominal samples, so the VAP will be applying this noise filter to all of the AERI data in the ARM Archive.

Further study is under way at the University of Wisconsin to use the calculated eigenvectors to explain the microphysical cloud properties.

8. References

Feltz, WF, WL Smith, HB Howell, RO Knuteson, H Woolf, and HE Revercomb. 2003. "Near-continuous profiling of temperature, moisture, and atmospheric stability using the Atmospheric Emitted Radiance Interferometer (AERI)." *Journal of Applied Meteorology* 42, 584-597.

Knuteson, RO, FA Best, NC Ciganovich, RG Dedecker, TP Dirkx, SC Ellington, WF Feltz, RK Garcia, HB Howell, WL Smith, JF Short, and DC Tobin. 2004a: "Atmospheric emitted radiance interferometer. Part I: Instrument design." *Journal of Atmospheric and Oceanic Technology* 21, 1763-1776.

Knuteson, RO, FA Best, NC Ciganovich, RG Dedecker, TP Dirkx, SC Ellington, WF Feltz, RK Garcia, HB Howell, WL Smith, JF Short, and DC Tobin. 2004b. "Atmospheric emitted radiance interferometer. Part II: Instrument performance." *Journal of Atmospheric and Oceanic Technology* 21, 1777-1789.

Tobin, DC, FA Best, PD Brown, SA Clough, RG Dedecker, RG Ellingson, RK Garcia, HB Howell, RO Knuteson, EJ Mlawer, HE Revercomb, JF Short, PFW van Delst, and VP Walden. 1999. "Downwelling spectral radiance observations at the SHEBA ice station: Water vapor continuum measurements from 17 to 26 μm." *Journal of Geophysical Research* 104, 2081-2092.

Turner, DD, DC Tobin, SA Clough, PD Brown, RG Ellingson, EJ Mlawer, RO Knuteson, HE Revercomb, TR Shippert, WL Smith, and MW Shephard. 2004. "The QME AERI LBLRTM: A closure experiment for downwelling high spectral resolution infrared radiance." *Journal of Atmospheric Science* 61, 2657-2675.

Turner, DD. 2005. "Arctic mixed-phase cloud properties from AERI-lidar observations: Algorithm and results from SHEBA." *Journal of Applied Meteorology* 44, 427-444.

Turner, DD, RO Knuteson, HE Revercomb, C Lo, and RG Dedecker. 2006. "Noise reduction of Atmospheric Emitted Radiance Interferometer (AERI) observations using principal component analysis." *Journal of Atmospheric and Oceanic Technology*, submitted.