JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 112, D20210, doi:10.1029/2006JD008300, 2007



# Using EOF analysis to qualitatively analyze, and identify inhomogeneities in, data from ground-based aerosol monitoring instruments

S. M. Gianelli,[1] B. E. Carlson,[2] and A. A. Lacis[2]

Received 1 December 2006; revised 17 July 2007; accepted 1 August 2007; published 30 October 2007.

[1] Empirical Orthogonal Function (EOF) analysis is performed on ground-based shadowband radiometer and Sun photometer data. The data come from Multifilter Rotating Shadowband Radiometer (MFRSR) instruments located at the central and extended facilities of the Southern Great Plains (SGP) research site of the Department of Energy's Atmospheric Radiation Measurement (ARM) program, as well as the Rotating Shadowband Spectroradiometer (RSS) 102 and 105 and the CIMEL Sun photometer CSPHOT located at the central facility at SGP. The EOFs show the variability in total aerosol optical depth and provide some qualitative information on the separation of the coarse and fine aerosol modes. In addition, as has been demonstrated previously with satellite data, EOF analysis also exposes several flaws and inconsistencies within the ground-based data sets. These inhomogeneities include optical depth anomalies in some MFRSR filters, wavelength shifts in the RSS, and problems with the data transmitter of the CIMEL instrument. Therefore EOF analysis is shown to be a quick and effective means not only of assessing the general aerosol behavior in the air above a particular monitoring instrument, but also of identifying both known and unanticipated influences on the data coming from within the instrument itself.

**Citation:** Gianelli, S. M., B. E. Carlson, and A. A. Lacis (2007), Using EOF analysis to qualitatively analyze, and identify inhomogeneities in, data from ground-based aerosol monitoring instruments, *J. Geophys. Res.*, *112*, D20210, doi:10.1029/2006JD008300.

## 1. Introduction

[2] The analysis of a large, multidimensional data set can often be greatly facilitated by separating the set into its constituent Empirical Orthogonal Functions, or EOFs [*Peixoto and Oort*, 1992]. Each EOF has a corresponding eigenvalue, which determines how much the EOF in question contributes to the total variance of the data set. Often, nearly all of the variance in a data set can be accounted for by its first few eigenvectors. This particular study examines data from different instruments that measure optical depth at different sets of wavelengths. It will be shown that EOF analysis not only can enable an objective comparison between optical depth data sets from different instruments, but it can also indicate where these data sets stray from their natural patterns and indicate flaws in the behavior of the instruments.

### 1.1. Examples of EOF Analysis in Atmospheric Science

[3] EOF analysis – sometimes referred to as eigenvector analysis, or principal component analysis – has been an important tool in atmospheric science since the fifties [*Lorenz*, 1956]. Generally it has been used to link temporal and spatial patterns in data sets. For example, *Chen et al.* [2002] used EOF analysis to isolate the temporal and spatial patterns corresponding to El Niño from a satellite data set of radiation fluxes, in order to focus on the upward trend in outgoing longwave flux and its corresponding effects. The first few eigenvectors in a typical set of EOFs reveal the primary natural variations in a given data set, but sometimes the later eigenvectors also show artificial, or instrument-related, variability. *Chelliah and Arkin* [1992] used EOFs to analyze satellite measurements of outgoing longwave radiation over the tropics. They found that the third and fourth eigenvectors of their set corresponded with data artifacts resulting from different equatorial crossing times and spectral windows utilized by the different satellites included in the data set.

[4] Studies pertinent to the use of data from ground-based aerosol monitoring instruments have employed EOF analysis previously to this one. *Box et al.* [1992] applied a combination of EOF analysis and Mellin transformation to retrieve full aerosol size distributions, both unimodal and bimodal, from real and synthetic optical depth data. Simplifying their mathematical approach somewhat, *Box et al.* [1996] used eigenvectors to estimate of the number of independent items of aerosol information that can be obtained from different combinations of wavelengths appli-

[1]Department of Applied Physics and Applied Mathematics, Columbia University, New York NASA Goddard Institute for Space Studies, New York City, New York, USA.
[2]NASA Goddard Institute for Space Studies, New York City, New York, USA.


0148-0227/07/2006JD008300$09.00

cable to multichannel radiometers, given a specified amount of relative error in the measurements. *Taha and Box* [1999] extended this work by applying it directly to MFRSR optical depth data, using the eigenvectors to separate the extinction of aerosols from that of ozone in the 610 nm channel. *Rabbette and Pilewskie* [2001] applied the technique to data from the solar spectral flux radiometer (SSFR), which measures incoming solar radiation at a continuum of wavelengths. They performed their EOF analysis directly on the irradiance spectrum instead of the optical depth data, and included cloudy days as well as clear ones in their analysis. They concluded that their data contained six consequential eigenvectors. When the EOFs were rotated in the analysis, the first four eigenvectors were dominated respectively by cloud scattering, water vapor absorption, Rayleigh scattering, and ozone absorption, while the fifth and sixth appeared to result from data artifacts. *Gianelli* [2004] evaluated several data sets from different ground-based aerosol monitoring instruments using EOF analysis, assessing the degree to which specific information concerning the aerosol size distribution can be retrieved with different combinations of wavelengths, and observing artificial sources of variability as well.

### 1.2. Calculating Empirical Orthogonal Functions

[5] Mathematically speaking, EOFs are the set of independent, orthonormal eigenvectors that most efficiently represent a given set of data. The objective is to have the first eigenvector account for as much of the variance as possible, with each subsequent eigenvector accounting for as much of the remaining variance as possible. For example, say that **F** is an M $\times$ N matrix, where M equals the number of mornings and afternoons for which RSS optical depth data exists, and N equals the number of channels in the RSS data. The element $F_{xy}$ in matrix **F**, then, contains the optical depth value measured for wavelength y on day x. Each row corresponds to the vector of the optical depth as a function of wavelength for a given morning or afternoon, and each column corresponds to the vector of the optical depth as a function of day for a given wavelength.

[6] Evaluating the total variance in the system first requires determining how the row vectors vary with each other. This done by calculating the covariance matrix **R**, a symmetric M $\times$ M matrix in which $\mathbf{R}_{xy}$ equals the covariance between row vectors x and y, or the degree to which the deviation of the components of row vector x from the vector's mean value coincides with the deviation from the mean of the corresponding elements of row vector y. In matrix notation, the calculation of **R** can be expressed by the simple formula $\mathbf{R} = (\mathbf{F}^*\mathbf{F}^T)/N$ [*Peixoto and Oort*, 1992]. The covariance matrix can then be diagonalized, and its eigenvalues and corresponding eigenvectors calculated. These eigenvectors are the empirical orthogonal functions. In EOF analysis, the largest eigenvalue is always the first, with the subsequent eigenvalues steadily decreasing in amount. The eigenvalues are ranked by magnitude because, as demonstrated by *von Storch and Zwiers* [1999] for example, the largest contribution to the total variance of the system comes from the eigenvector of the covariance matrix with the largest eigenvalue. The largest remaining contribution comes from the eigenvector with the second largest eigenvalue, and so forth. The sum of the eigenvalues represents the total variance within the data set, and the ratio of each individual eigenvalue to that sum indicates what fraction the corresponding eigenvector contributes to the total variance. The number of eigenvalues in an EOF analysis that contribute anything to the total variance (that is, have a value greater than zero) will never exceed N, the number of channels used in the analysis. Practically speaking, most of the variance can be accounted for in the first few eigenvalues, implying that only the first few of the M eigenvectors are needed to explain the significant behavior of the data set.

[7] The set of eigenvectors can be represented as an M $\times$ M square matrix **E**, and each eigenvector can be treated as a function of time, or more specifically, of the mornings and afternoons in the data set. In addition, each eigenvector has a corresponding set **C** of coefficients of projection, which are calculated using the formula $\mathbf{C} = \mathbf{E}^T*\mathbf{F}$. Each row of **C** can be treated as a function of wavelength, with the first row corresponding to the first eigenvector, the second row to the second eigenvector, and so on. The full data set **F**, then, can be expressed as the matrix product of **E** and **C**, which means algebraically that **F** is the sum of the matrix products of each eigenvector with its corresponding coefficients.

### 1.3. Data and Instrumentation

[8] A number of different ground-based instruments measure incoming solar radiation at different wavelengths, producing data that can then be used to provide information on aerosol amounts and properties. For example, tracking Sun photometers like the CIMEL Electronique 318A Spectral Radiometer, the instrument of choice for the AERONET network [*Holben et al.*, 1998], have detectors that track the sun across the sky to measure the direct beam. The CIMEL uses seven low-resolution filters with central wavelengths between 340 and 1020 nm. Other ground-based measuring instruments employ a rotating shadow band to separate the diffuse scattered solar radiation from the total incoming solar radiation, so that the direct beam can then be calculated from the difference between the two and the direct/diffuse ratio can be measured independently of calibration. The most common of these instruments is the Multifilter Rotating Shadowband Radiometer, or MFRSR [*Harrison et al.*, 1994], which consists of six low-resolution filters with central wavelengths ranging from 415 to 940 nm. A newer, high-resolution instrument called the Rotating Shadowband Spectroradiometer [(RSS), *Harrison et al.*, 1999] contains an array of over 1000 channels, with a wavelength range from 360 to 1100 nm. The RSS can separate the contributions of aerosols and gases to the total extinction at visible wavelengths more cleanly than the MFRSR can [*Gianelli et al.*, 2005]. In addition, data from the RSS can be easily converted into MFRSR and CIMEL "equivalents," where the wavelengths corresponding to channels in the MFRSR or CIMEL are isolated from the RSS data [*Gianelli*, 2004]. This enables objective, wavelength-for-wavelength comparisons between the RSS and the other instruments. The RSS 102 operated at the SGP facility from July 1999 to July 2000, and the RSS 105 has been operating there since May 2003. The data from the RSS 105 used in this EOF analysis cover the period from May 2003 to November 2005. The data from the CIMEL located at the central facility used in this analysis cover the years 1998 through 2000. As each MFRSR filter has distinct properties,

the effect of spectral leakage on MFRSR filters will vary from instrument to instrument. Therefore a number of different MFRSR instruments located in the same region were used in this study, including those run by the USDA UV-B Monitoring and Research Program [*Bigelow et al.*, 1998] and by ARM at the central facility C1 of SGP, plus instruments at the ARM extended SGP facilities E1, E3, E5, E6, E7, E8, E9, E10, E11, E12, E13, E15, E16, E19, E20, E22, and E25.

[9] For the RSS and MFRSR instruments, the optical depth data are obtained from morning and afternoon values of the slopes of Langley regression, after cloudy points have been removed. Rayleigh optical depths are subtracted out of the analysis before the EOFs are calculated, using the method developed by *Bodhaine et al.* [1999], but gaseous absorption due to water vapor, ozone, and nitrogen dioxide is maintained in the data, as is aerosol extinction. The CIMEL optical depth values used are the morning and afternoon means of Level 2.0 cloud-screened optical depth measurements obtained from the AERONET website.

### 1.4. The Use of EOF Analysis in This Study

[10] EOF analysis is well suited to the analysis of large sets of multiwavelength radiometer data. For one thing, it minimizes the number of significant independent variables. This is especially useful when analyzing data from a thousand-channel array like the RSS. Furthermore, it enables at least a partial separation of the contributions to the total extinction from different sources like aerosols, water vapor, ozone, and nitrogen dioxide. If all these quantities varied entirely independently from each other, their respective contributions to the total extinction would show up in the EOF analysis as separate eigenvectors. This is generally not observed, though, as aerosols and water vapor co-vary both in terms of their annual cycles and in terms of hygroscopic growth of aerosol particles. EOF analysis also allows for a rapid, qualitative examination of the data set as a complete entity, without the need to look at each wavelength individually. Finally, if the EOFs are calculated by first subtracting the mean optical depth values over the data set from each channel, any biases in measured optical depth values between different instruments vanish, allowing for a more objective comparison between different data sets [*Gianelli*, 2004]. Subtracting the means also removes the spectral signatures of ozone and nitrogen dioxide from the first two eigenvectors, leaving only the extinction due to water vapor and aerosols for analysis.

[11] Each eigenvector presented here is represented by two graphs. The top graph plots the components of the eigenvector itself versus day. The bottom graph plots the coefficients of projection corresponding to that particular eigenvector vs. wavelength. So the first EOF of a given data set consists of both the dominant temporal pattern and the dominant spectral pattern contained in the data. It was shown by *Gianelli* [2004] that the coefficients of projection of the first eigenvector in an optical depth data set are very nearly in exact proportion to the mean optical depth values for each channel over the course of the data set. *Box et al.* [1996], in their theoretical analysis, established that only two or three items of independent aerosol information are obtainable from optical depth data in the wavelength ranges of the MFRSR, RSS, and CIMEL instruments, assuming a relative error in optical depth measurements of 10%. With this in mind, it can be assumed that at most three EOFs contain significant information about the temporal and wavelength patterns of aerosol extinction. *Gianelli* [2004] observed that as aerosols partially correlate with water vapor, the second and third eigenvectors account for more variance when wavelengths sensitive to the absorption water vapor are included in the analysis. So if the first eigenvector represents the mean pattern in the wavelength dependence of the total extinction, then the next two eigenvectors (or the first two, if the EOFs are calculated after the mean values have been subtracted) represent the main sources of deviation from this pattern. Physically speaking, these deviations arise from variations in the aerosol size distribution. In addition to the natural variations in the optical depth values, unnatural patterns of variation can sometimes affect a data set as well. Such patterns may emerge in the EOF analysis when eigenvectors beyond the dominant two or three are examined.

## 2. Qualitative Bimodal Aerosol Information in EOFs

[12] Compounding the difficulty in precisely measuring physical aerosol properties with optical depth data from multiwavelength radiometers is the presence of multiple aerosol modes in the atmosphere. *Remer and Kaufman* [1998], for example, identify two sources each of coarse mode ($R_{EFF} > 1.0$ $\mu$m) and fine (accumulation) mode aerosol particles over the eastern United States. *Eck et al.* [1999] used CIMEL data to show that the complexity of the relationship between aerosol extinction and wavelength depends significantly on whether the aerosol size distribution is dominated by coarse mode or fine mode particles, while *Gianelli* [2004] demonstrated the necessity of including both a coarse and fine aerosol mode to obtain an accurate fit when retrieving aerosol information from RSS optical depth data. The results of that study suggest that the aerosol optical depth can, with reasonable accuracy, be separated into its coarse and fine mode components using the full range of CIMEL and RSS 102 wavelengths. The bimodal aerosol retrievals performed by *Gianelli* using the ''MFRSR equivalent'' variant of the RSS 102 data could not replicate the results of the retrievals exploiting the full wavelength range of the RSS, however, because the shorter wavelength range used in the retrieval caused the coarse mode optical depth to be systematically underestimated. Still, EOF analysis of the ''MFRSR equivalent'' version of the RSS 102 data set points to a way by which information about the coarse mode optical depth can be rapidly obtained using MFRSR data, even if just in a qualitative sense.

[13] Figures 1 and 2 show the first two eigenvectors of the ''MFRSR equivalent'' data set, with the mean values for each channel subtracted. Note that the spectral signature of ozone does not appear at all in either eigenvector, leaving only aerosols and water vapor. The first eigenvector shows the seasonal pattern in optical depth, relative to the mean. Both aerosols and water vapor have their maxima in summer and minima and winter, so this eigenvector reflects the general correlation between the extinction due to these quantities. The second eigenvector emphasizes the days that break from this pattern. The most obvious of these days

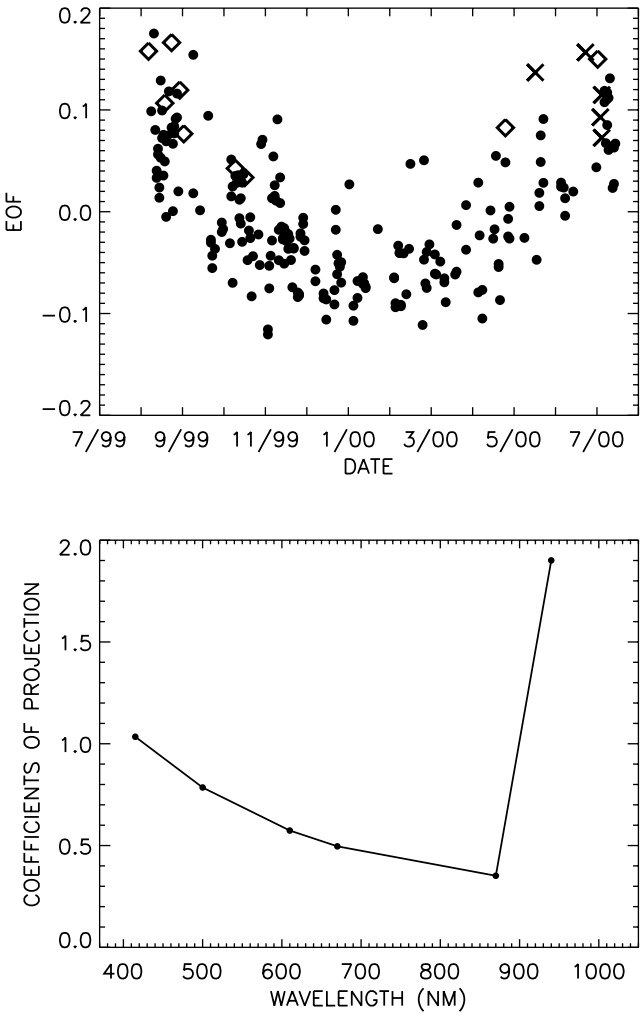


**Figure 1.** The first EOF of the "MFRSR equivalent" variant of the RSS 102 data, with the mean optical depth subtracted. This eigenvector accounts for 76% of the total variance. On the days marked with a diamond, the optical depth measured at 870 nm exceeded 0.1 and the retrieved fine mode optical depth at 550 nm exceed the retrieved coarse mode optical depth. On the days marked with an 'X', the optical depth measured at 870 nm exceeded 0.1 and the retrieved coarse mode optical depth at 550 nm exceed the retrieved fine mode optical depth.

occur in the high optical depth periods in August and October 1999. The coefficients of projection for both eigenvectors show curves that decrease with wavelength, outside of the 940 nm channel that is influenced by the strong water vapor absorption band. A typical aerosol curve, especially one dominated by the fine mode, will exhibit this kind of wavelength dependence. If the total aerosol extinction for a given day contains a significant coarse mode component, though, the decrease of optical depth with wavelength will be significantly mitigated, if not reversed. In the EOF analysis, such a result should manifest itself as a strongly negative value in at least one of the first two eigenvectors. Given that the first eigenvector reflects the general pattern and the second one reflects deviations over specific days or multiday stretches, it is more likely that a day with significant coarse mode optical depth will have a low or negative value for its component in the second eigenvector.

[14] The days that have been plotted with special symbols on the top graph are days where the measured optical depth at 870 nm, a wavelength where no gases absorb significantly, exceeds 0.1. If the value of the coarse mode optical depth at 550 nm, retrieved from the RSS data using the full wavelength range, is greater than the retrieved value of the fine mode optical depth on one of these special days, the day is marked with an X. If the fine mode optical depth is greater, the day is plotted with a diamond. In the second eigenvector, the days of high fine mode amounts and those of high coarse mode amounts separate very cleanly from each other. Only one high fine mode day produced an EOF value below zero, and even that value was higher than the value for any of the high coarse mode days.

[15] This suggests, then, a simple way to identify days of significant coarse mode optical depth in regular MFRSR

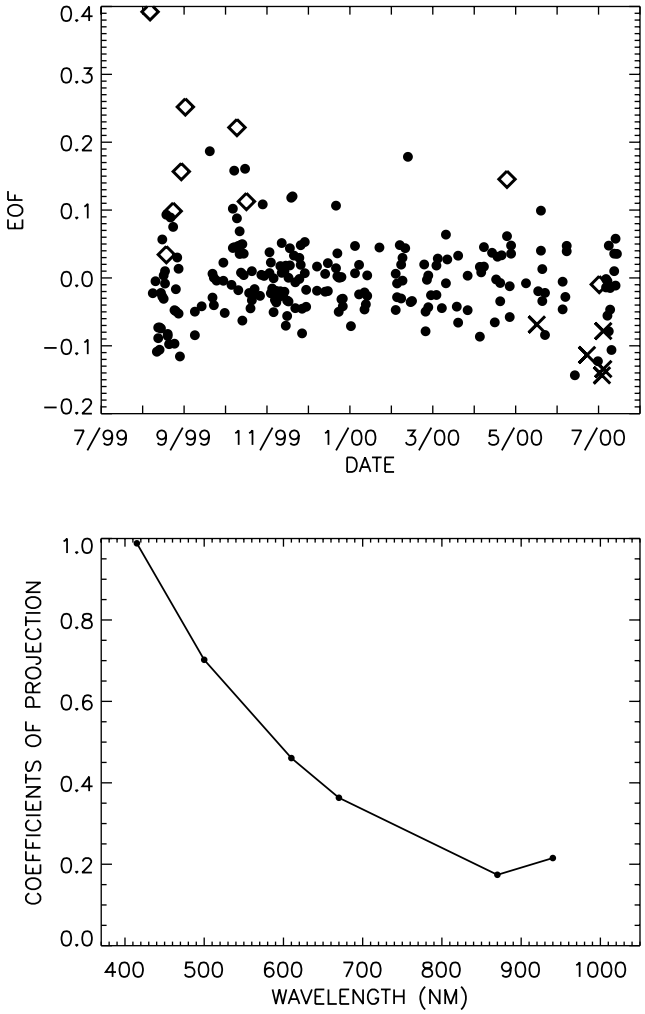


**Figure 2.** The second EOF of the "MFRSR equivalent" variant of the RSS 102 data, with the mean optical depth subtracted, accounts for 23% of the total variance. Here, the high coarse mode optical depth days are cleanly distinguished from the high fine mode optical depth days, enabling a fast and simple, if qualitative, means of determining the high coarse mode days for both RSS and MFRSR data.

**Table 1.** This Table Shows the High Measured Optical Depth Days, Categorized by Which Aerosol Mode Dominates According to the EOF Analysis, for the RSS 102 and the Co-located USDA and ARM MFRSR Instruments Over Days in the Data Set That All Three Instruments Have in Common[a]

| | RSS 102 | USDA MFRSR | ARM MFRSR |
|---|---|---|---|
| High fine mode days: | 10/6/99, 7/2/00 | 10/6/99, 4/18/00, 7/2/00 | 10/6/99, 4/18/00, 7/2/00 |
| High coarse mode days: | 6/23/00, 7/5/00 | 2/25/00, 6/23/00, 7/5/00 | 2/25/00, 6/23/00, 7/5/00 |

[a]The RSS 102 generally measured lower optical depth values than the MFRSR instruments did, explaining why the MFRSR's include one more day for each aerosol mode.

data. Days where the optical depth at 870 nm exceeds 0.1 were isolated from both the USDA and ARM instruments located at C1, and then separated depending on whether the value for the second EOF with the mean subtracted was positive for that day, indicating a strong fine mode, or negative, indicating a strong coarse mode. The results are presented in Table 1. Only days common to all three instruments are included. As *Gianelli* [2004] showed the USDA MFRSR had systematically higher optical depth values than the RSS 102, it was expected that fewer days would meet the cutoff optical depth of 0.1 in the RSS. For example, February 25, 2000 produced the highest retrieved coarse mode optical depth for the RSS 102 on the days where the optical depth measured at 870 nm did not reach 0.1, so its inclusion in the list of high coarse mode days for the other two instruments has more to do with higher optical depth measurements in the MFRSR instruments than a difference in the EOF calculations between the instruments. Otherwise, the agreement is good between the instruments.

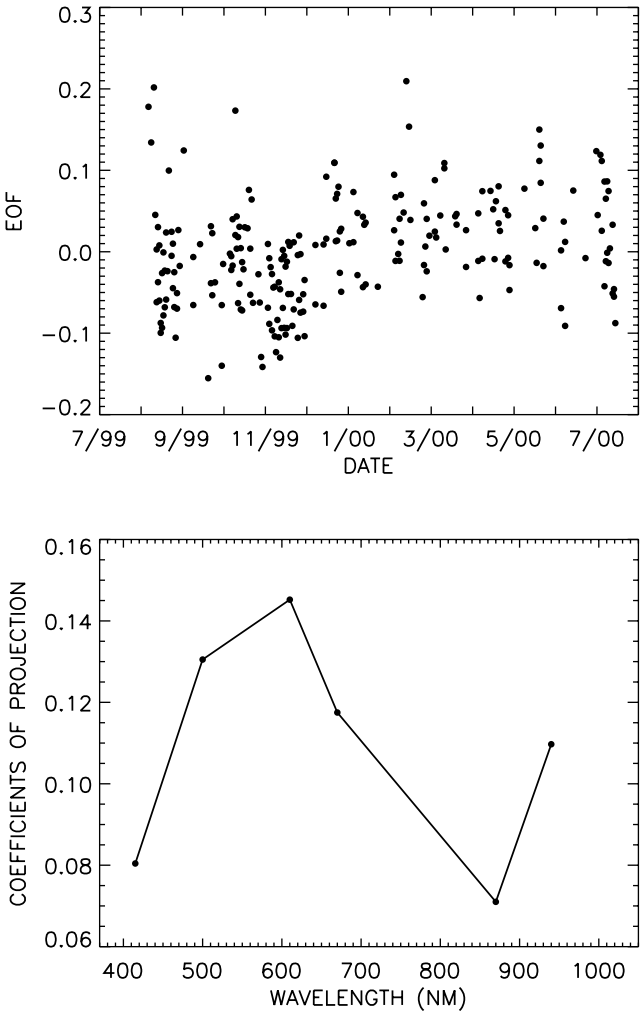


**Figure 3.** The third eigenvector of the "MFRSR equivalent" variant of the RSS 102 data set with the means subtracted contributes 0.2% to the total variance. The spectral signatures of both ozone and water vapor are present in this eigenvector.

## 3. Anomalous Optical Depth Measurements in MFRSR Filters

[16] EOF analysis also provides a very quick way of determining if a filter in a particular MFRSR instrument is producing optical depth values that are not fully consistent

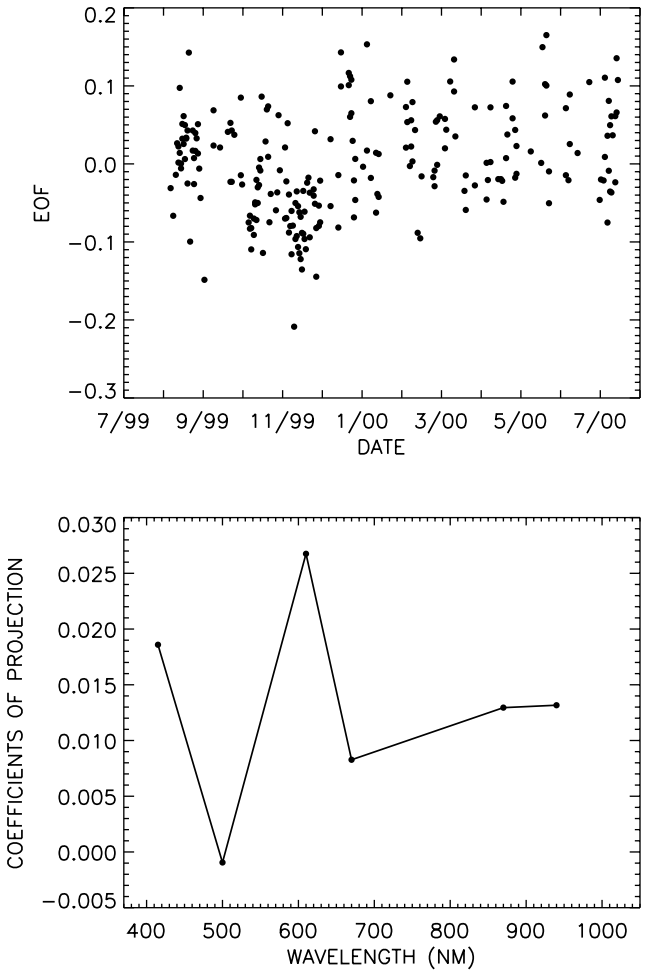


**Figure 4.** The fourth equivalent of the "MFRSR equivalent" data set contributes 0.02% to the total variance. The coefficients of projection follow a W-shaped curve.

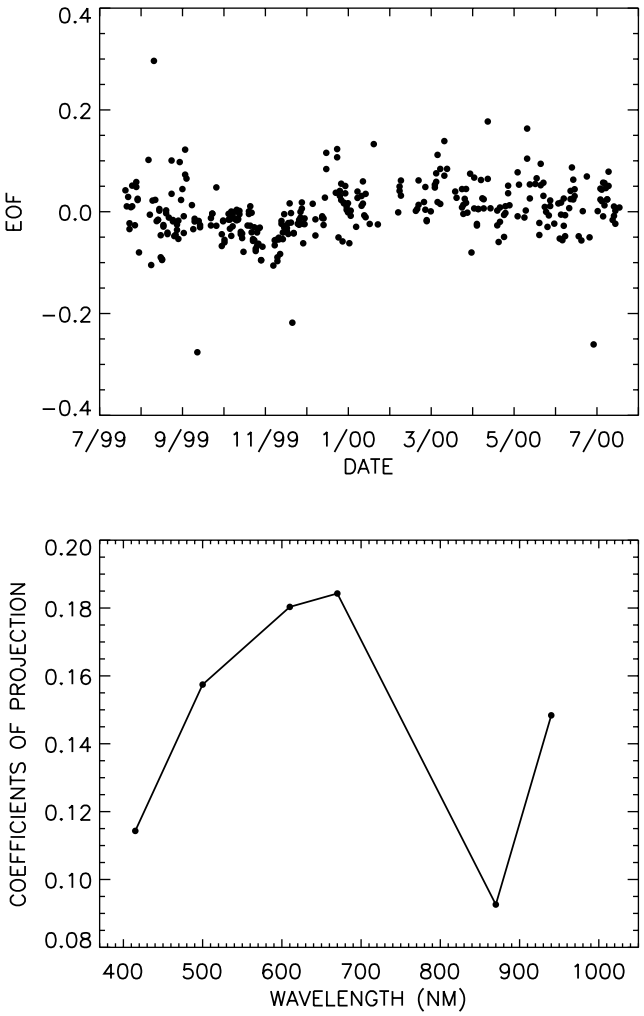


**Figure 5.** The third eigenvector for the MFRSR located at E5, calculated after the means have been subtracted, contributes 0.4% to the total extinction. In contrast to Figure 3, the coefficient of projection at 670 nm is higher than the one at Figure 3.

with the values seen in the other filters of the instrument. These anomalies might not be easy to find by other means. As the RSS 102 remained stable throughout its operation, the EOF analysis of the "MFRSR equivalent" should provide a good idea of what the EOFs of a stable MFRSR device will look like. Figures 3 and 4 show the third and fourth eigenvectors of the RSS 102 "MFRSR equivalent" data set, with the means subtracted. The third eigenvector contributes 0.2% to the total variance. The absorption spectrum of ozone is clearly present in the coefficients of projection, along with water vapor absorption at 940 nm. By contrast, the contribution of aerosols to this eigenvector appears to be minimal. The fourth eigenvector contributes 0.02% of the total variance, and its coefficients of projection follow a W-shaped curve. In neither eigenvector do any days produce an EOF component with a value noticeably outside the main cluster of values.

[17] When the EOFs are calculated for MFRSR data from the ARM and USDA sites in the SGP region, the spectral pattern of the coefficients of projection of the first two eigenvectors all very closely resemble what is seen in Figures 1 and 2. Gas absorption due to ozone does not appear, leaving two aerosol curves along with the signature of water vapor absorption at 940 nm. The magnitude of the coefficients at 940 nm may vary somewhat, depending on how closely the peak of the 940 nm filter's response function matches the peak of the water vapor absorption band, but the coefficients do not otherwise differ significantly from one instrument to the next. Differences do start to appear in the coefficients of the third eigenvector, though, at several of the sites.

[18] Figures 5 and 6 show the third and fourth eigenvectors at the E5 site, over the same time period of time that the RSS 102 operated. The coefficient of projection of the third eigenvector at 670 nm is actually a bit higher than the one at 610 nm, in contrast to what was observed with the RSS. Furthermore, one day has a noticeably high EOF value, and three days have values that are noticeably low. This eigenvector contributes 0.4% to the total variance, a bit more than what was observed in the RSS 102. The fourth eigenvector

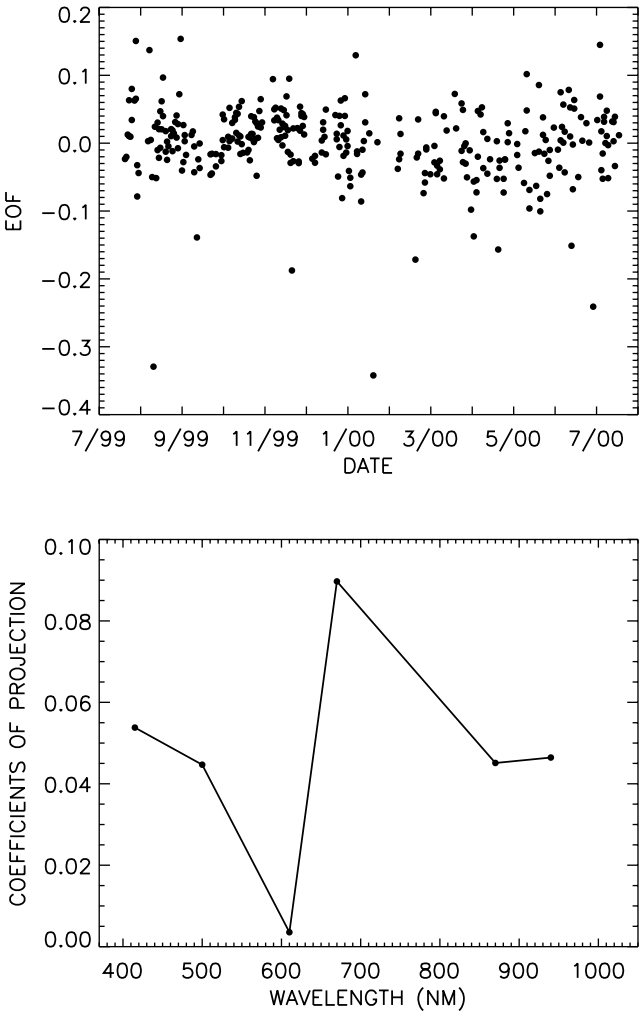


**Figure 6.** The fourth eigenvector of the MFRSR at E5 contributes 0.3% to the total variance, and does not resemble Figure 4 at all. Instead, the coefficients of projection feature a sharp spike at 670 nm superimposed over an inverted ozone spectrum.

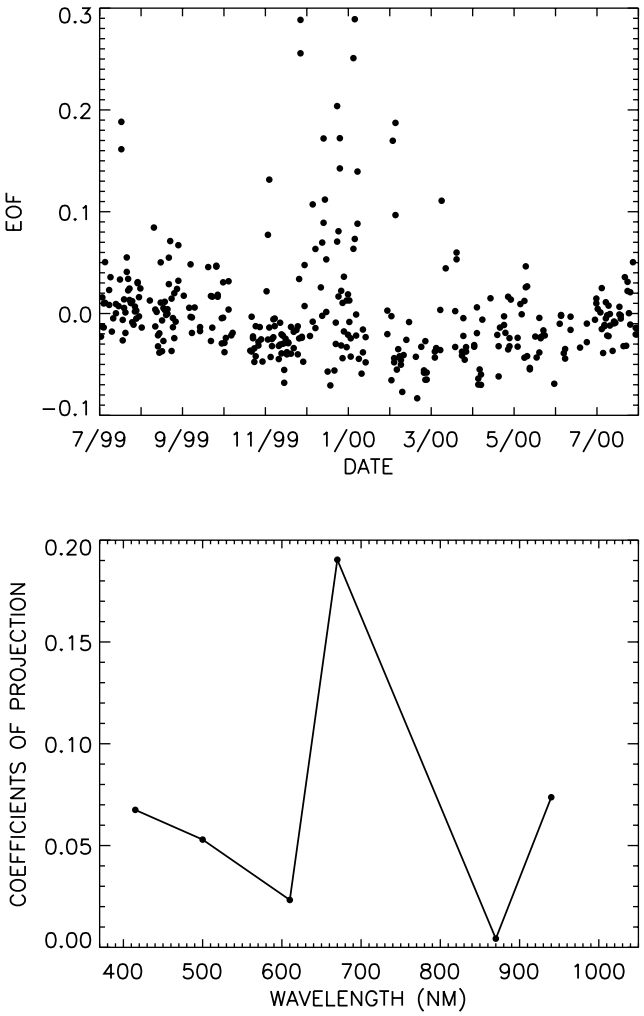


**Figure 7.** The third eigenvector of the MFRSR at E16 contributes 1.5% to the total variance. The EOF indicates many days, particularly during late fall and early winter, where the 670 nm filter had anomalous optical depth values.

looks nothing like the corresponding eigenvector for the RSS 102. It contributes 0.3% to the total variance – not much on the whole, but 15 times greater than the contribution of the fourth RSS eigenvector. Again, the value of the coefficient at 670 nm is noticeably high, while the remaining coefficients resemble an inverted ozone spectrum. These two figures suggest that the 670 nm filter produced anomalous optical depth values in the 670 nm channel on a handful of days. These days can be determined by the deviant values in the components of the eigenvector.

[19] The MFRSR instrument at E16 exhibited similar issues with the 670 nm filter, only with greater magnitude and frequency of occurrence. Figure 7 shows the third eigenvector, which contributes 1.5% to the total variance. The ozone signal is inverted, like Figure 6 instead of Figure 5. The EOF indicates that the 670 nm filter produced many anomalously high optical depth values from November 1999 through March 2000. The fourth eigenvector, shown in Figure 8, closely resembles the third eigenvectors of the data sets from RSS 102 and the other MFRSR devices in this study, although the slightly low value of the coefficient of projection at 670 nm offsets the high value in the third eigenvector on the days where the component of the third EOF was not positive and overly large. This eigenvector contributes 0.5% to the total variance.

[20] The cause of the anomalies in these two devices is not clear, but a possible hint comes from their filter response functions. Figure 9 shows the normalized response functions for the 670 nm filters for a number of MFRSR instruments in the ARM SGP network. The thinner lines represent the instruments at C1 (solid), E3 (dotted), E7 (dashed), and E10 (dot-dash). The 670 nm filters for the SGP site as a whole follow one of the two patterns exhibited here; either the response function actually peaks around 660 nm or 673 nm. The thicker lines, by contrast, represent the instrument at C5 (solid), and the two different heads that were used at E16 during the time span of this study. Head 373 (dashed) replaced head 372 (dotted) in January 2000. As Figures 7 and 8 show, the optical depth anomaly affected both heads. The curious aspect of the response functions of the filters at E5 and E16 is that all three filters in question

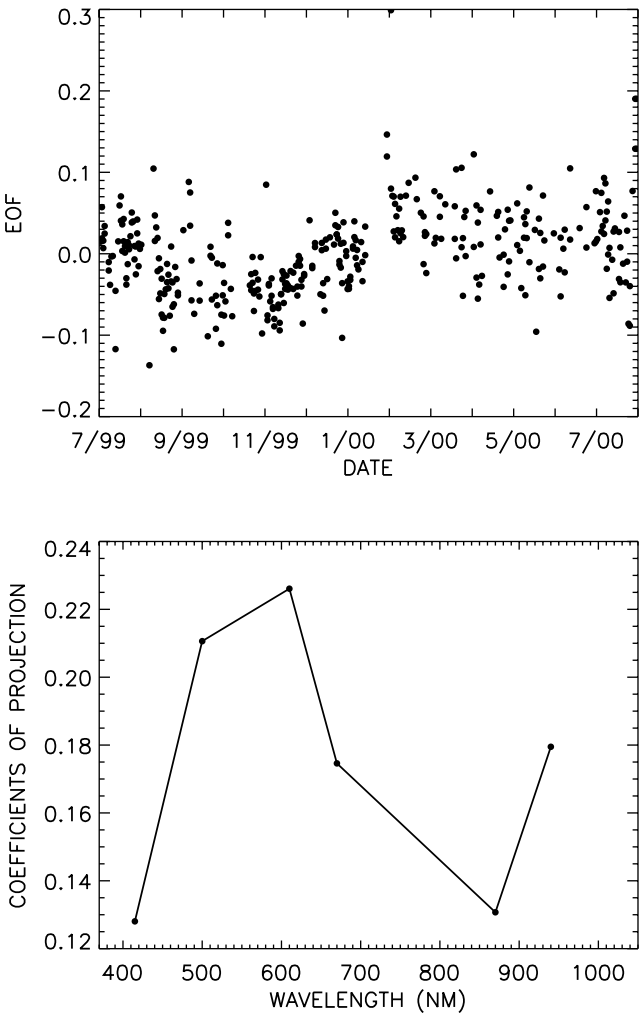


**Figure 8.** The fourth eigenvector of the MFRSR instrument at E16 contributes 0.5% to the total variance. The coefficients of projection resemble those of the third eigenvectors for the data sets of the RSS 102 and the MFRSR at E5.

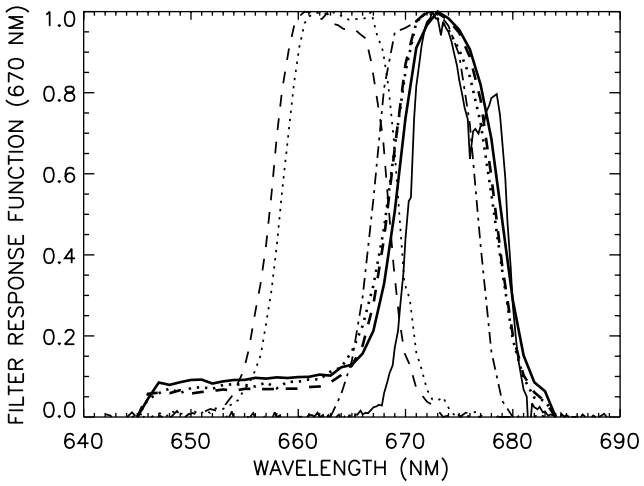


**Figure 9.** Some of the response functions for the 670 nm filters of the MFRSR instruments located within the SGP network. The thin lines represent C1 (solid), E3 (dotted), E7 (dashed), and E10 (dot-dash). The thick lines represent E5 (solid) and heads 372 (dotted) and 373 (dashed) that operated at E16 simultaneously with the RSS 102. Of all the MFRSR instruments at SGP, only the response functions for the filters used at E5 and E16 exhibited a tail extending to wavelengths shorter than 650 nm.

have a long tail extending down below 640 nm. Not one of the other MFRSR filters used at any of the extended facilities at SGP during the time period examined in this study exhibited this property.

[21] The fact that the optical depth anomaly occurs only on specific days indicates that the properties of these filters are not invariant over time. The peculiar filter response functions, in themselves, would not cause what is observed in Figure 9, but may indicate that these filters are particularly susceptible to changes in the response properties that could cause sideband leakage into the filters. Filter photometers have been observed elsewhere to allow in light from undesired wavelengths. For example, *Mitchell and Forgan* [2003] reported a bias between the optical depth measurements of filters centered at 868 nm in an SPO1A solar radiometer and those in a CIMEL instrument, and concluded that the higher SPO1A values resulted from sideband leakage in the SPO1A filter letting in additional light from a spectral region sensitive to the absorption of water vapor. Spectral leakage can cause aerosol extinction in a particular filter to be measured improperly, which in turn will compromise the accuracy of size distribution retrievals.

## 4. Wavelength Shifts in the RSS 102 and 105

[22] The RSS instruments have the advantage of high resolution over filter instruments like the MFRSR and CIMEL, but the central wavelengths of each pixel can shift over time, and these shifts will show up in the optical depth data. The pixel wavelengths for the RSS 102 were assumed to be constant, but the EOF analysis, as Figure 10 shows, indicates that the data set contained a couple of days where the pixel wavelengths deviated noticeably from their nominal values. The bottom graph plots optical depth values due to oxygen and water vs. wavelength on a low humidity day, as determined by a model simulation given typical pressure and humidity values for a winter day over the central facility at SGP. The coefficients of projection of the fifth eigenvector, plotted in the middle graph, resemble the first derivatives of the absorption spectrum of oxygen and water vapor. This behavior can be explained if the central wavelengths of the RSS 102 channels did not remain entirely steady, but rather fluctuated somewhat over time. Several subsequent eigenvectors also show first-derivative patterns, and collectively these indicate the magnitude and extent of the wavelength shifts, and on which days these shifts are significant.

[23] Figure 11 shows the fourth eigenvector of the RSS 105 data set, which indicates that the pixel wavelengths for the newer RSS shift to a more noticeable degree. The wavelengths used in the middle plot are the initial set of central wavelengths measured when the RSS 105 first

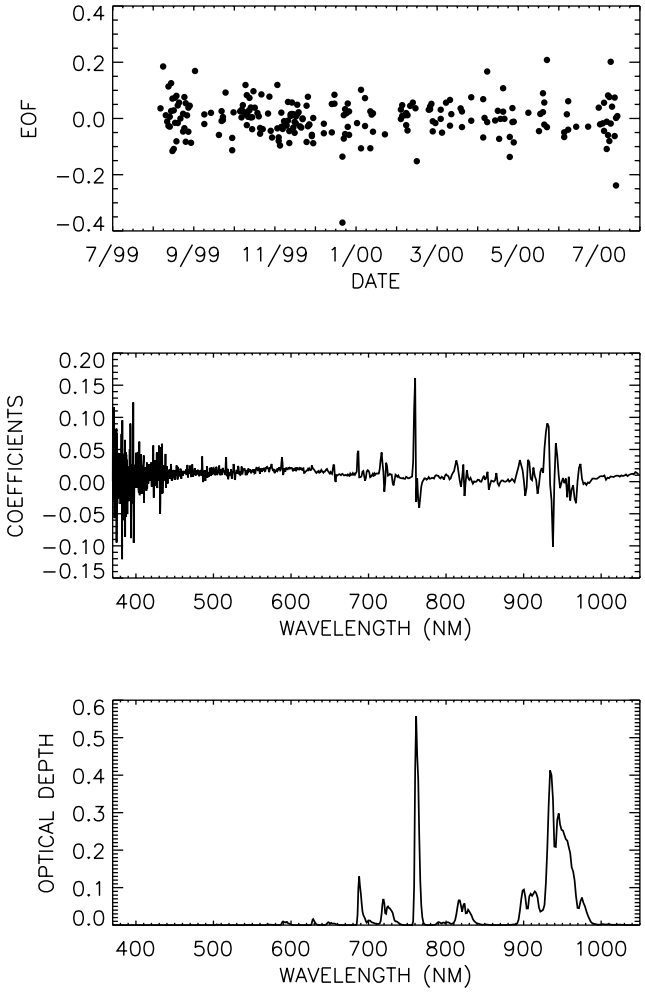


**Figure 10.** The fifth eigenvector for the RSS 102 data set at full resolution. The bottom graph shows the absorption spectrum of water vapor and oxygen. The first-derivative spectral pattern in the coefficients of projection indicates that the central wavelengths of the RSS pixels do not remain completely stable. The EOF shows that the pixel wavelength shifts for the RSS 102 remain centralized and, with one exception, manageably small.

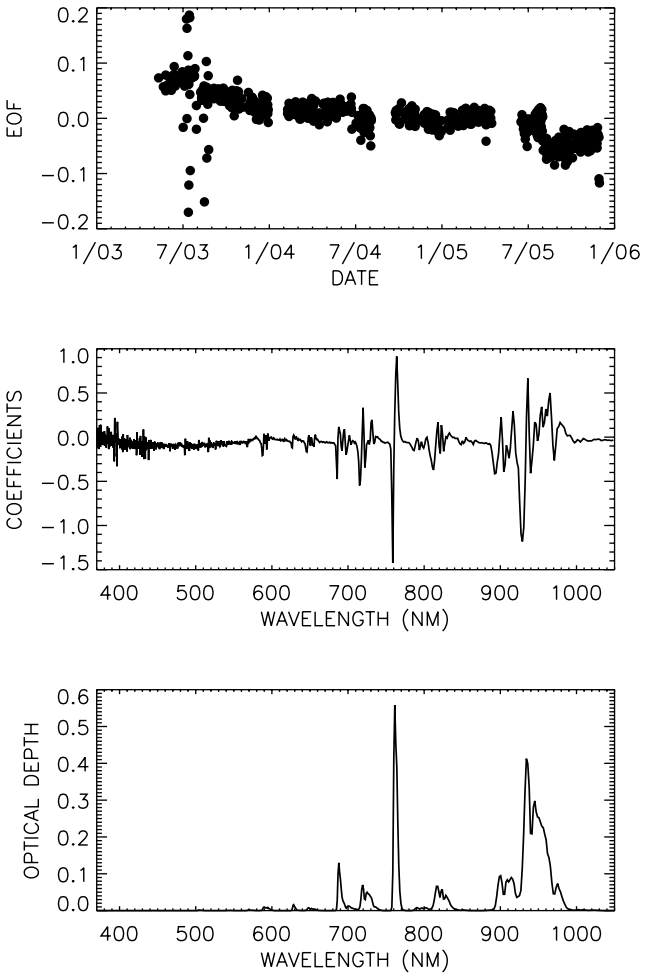


**Figure 11.** The fourth eigenvector for the RSS 105 data set at full resolution. By contrast, the EOF here shows that the pixel wavelength shifts for the RSS have followed a mostly steady trend, with a particularly sharp series of fluctuations in July and August of 2003.

started operating. Excepting a few days of large fluctuation in the summer of 2003, the eigenvector shows that the central wavelengths of the RSS 105 pixels have undergone a fairly steady shift over three years. This shift has been carefully monitored, though, and daily measurements of the pixel wavelengths are taken [Kiedron, private communication, 2005]. Anyone who wishes to measure gas amounts or monitor surface radiation fluxes at high resolution with the RSS 105 data needs to take into account the wavelength shifts by adjusting the pixel wavelengths, and the corresponding extinction coefficients and solar flux values, on a daily basis.

## 5. Problems With the Data Transmitter of the CIMEL at SGP

[24] The third eigenvector of the data from the CIMEL instrument located at SGP, presented in Figure 12, exhibits a fluctuation that adversely influences the entire data set. Because the fluctuation is not exactly annual (the time difference between the two primary peaks suggests a period of thirteen or fourteen months), and nothing resembling it appears in the EOF analysis from any of the other instrument, it must result from something instrument-related rather than from any natural cycle. According to the coefficients of projection, the two channels that appear to be most strongly affected by this are the first (at 340 nm) and the fifth (at 670 nm). In Figure 13, the third eigenvector is plotted after the EOFs have been re-calculated without these channels. The oscillation in the region between late 1998 and the spring of 2000 vanishes, with a steady upward trend in the values of the EOF taking its place. The negative values of the coefficients of projection mean that the EOF corresponds to a downward trending component within the measured optical depth. The cause of this trend was identified with the help of the AERONET principal investigator for this site [Wagener, private communication, 2004], who indicated that the particular instrument's data transmitter decayed over this time. While it was fixed on 14 April 2000, some problems persisted until 5 May. Figures 14 and 15 show the EOFs re-calculated again, first adding back the 340 nm channel without the 670 nm channel, and then including the 670 nm channel but

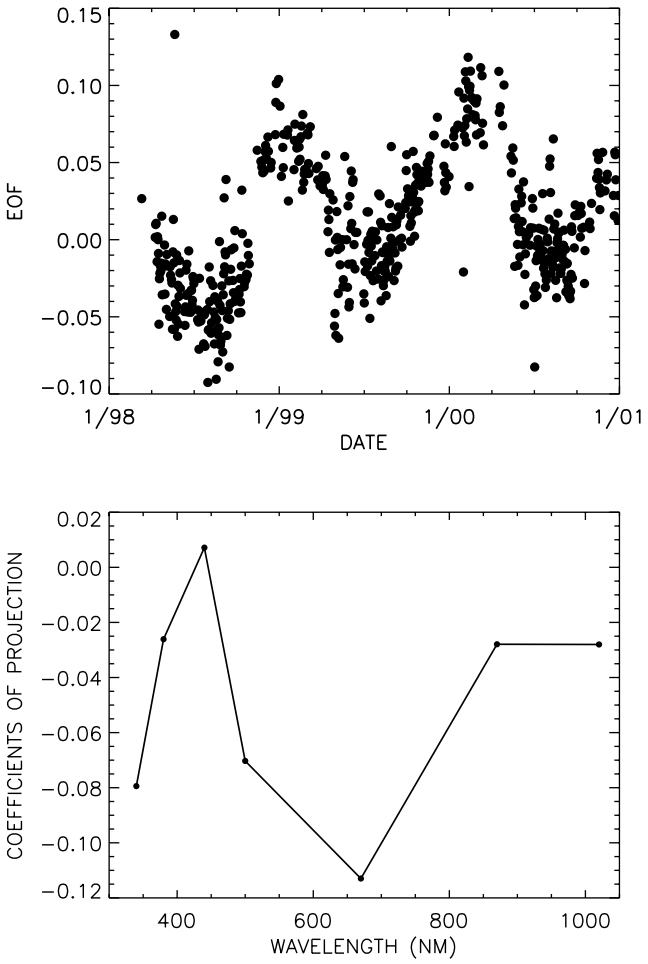


**Figure 12.** The third eigenvector for the data set of the CIMEL instrument located at the SGP central facility from 1998 through 2000. The eigenvector shows both an upward trend and an unnatural-looking fluctuation. From the coefficients of projection, the filters at 340 nm and 670 nm appear to be the most strongly affected.

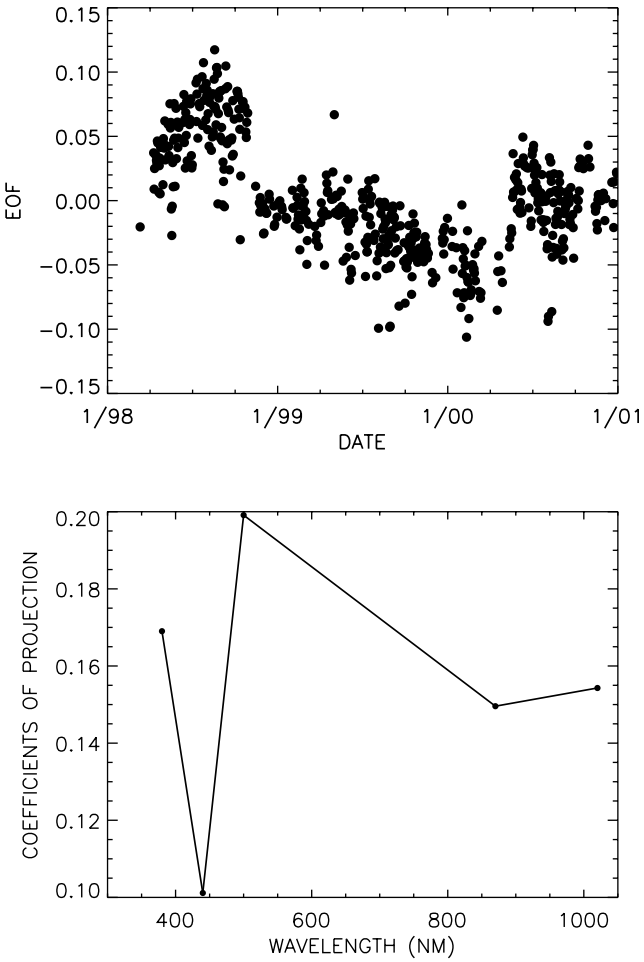


**Figure 13.** The third eigenvector of the CIMEL data set after the 340 nm and 670 nm data sets have been removed. A steady upward trend in the eigenvector exists between late 1998 and early 2000. From the coefficients, it can be seen that this trend causes a decrease in measured optical depth values. A problem with the data transmitter for this instrument caused this behavior, and was corrected in May 2000.

excluding the 340 nm channel. A dip in the EOF values in the middle of 1999, present in Figure 12, shows up in Figure 14 as well. The dip appears to have been caused by problems in the two ultraviolet channels, as the magnitudes of the coefficients of projection for the 340 nm and 380 nm channels are both large and opposite to each other. As the EOF values over this time period and the coefficient value for the 340 nm channel are both negative, it means that the measured optical depth was unnaturally high at 340 nm but low at 380 nm. Conversely, the low EOF values in the middle of 1998 and the high EOF values at the beginning of 2000 show up in Figure 15, indicating that data transmitter problems that caused the unnatural trend in all the filters during this time period affected the 670 nm filter the strongest.

## 6. Summary and Conclusions

[25] Generally speaking, EOF analysis of optical depth data from a particular instrument can uncover much information, both concerning the physical processes in the atmosphere above the instrument and also the performance of the instrument itself. The first two or three eigenvectors provide information regarding the general aerosol properties above the site. For example, days of large optical depth can be very simply categorized according to which aerosol mode dominated the extinction. EOF analysis can also quickly uncover specific days or periods of time in the set that deviate from the general patterns, though, and often the lesser eigenvectors reflect both known and unexpected instrument-specific issues. Eigenvectors beyond the first and second in a data set can shed light on potentially problematic issues with MFRSR filters, for example. In a high-resolution instrument like the RSS, EOFs can be used to monitor the wavelength stability of the pixels. Hardware problems affecting the data set can be observed as well, as was seen in the case with the CIMEL instrument at SGP and the issues with the data transmitter. Indeed, any issue that adversely affects the quality of optical depth data will show up somewhere in an EOF analysis. An EOF analysis of a particular instrument will only contain as many pertinent eigenvectors as there are channels used in the analysis, though. So for instruments like the MFRSR and CIMEL, an

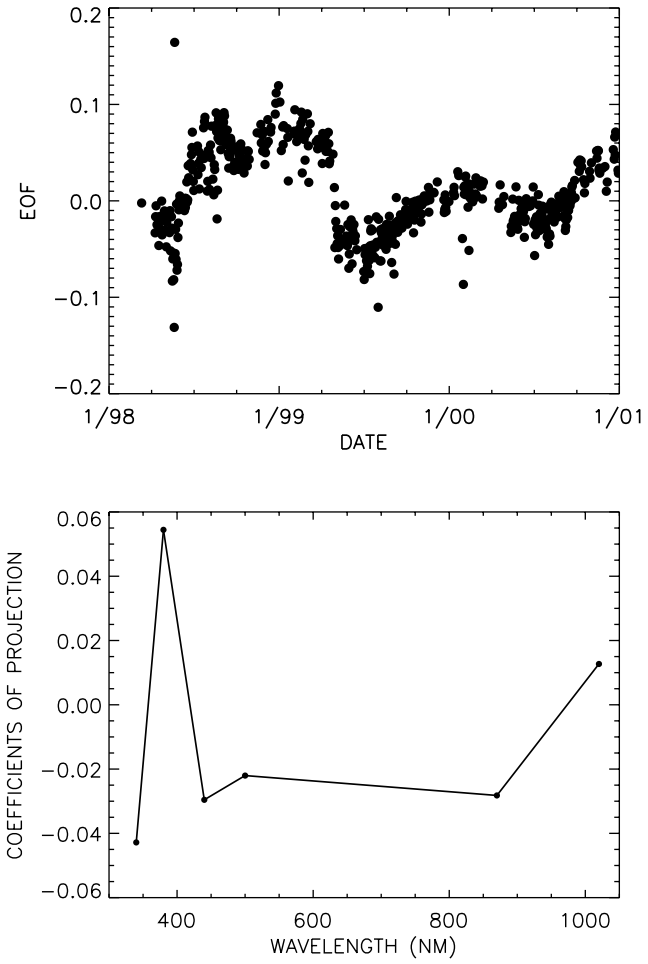


**Figure 14.** When data from the 340 nm channel are re-introduced, the dips in the value of the eigenvector observed in Figure 12 re-emerge. They appear to be the result of problems with the two ultraviolet channels.

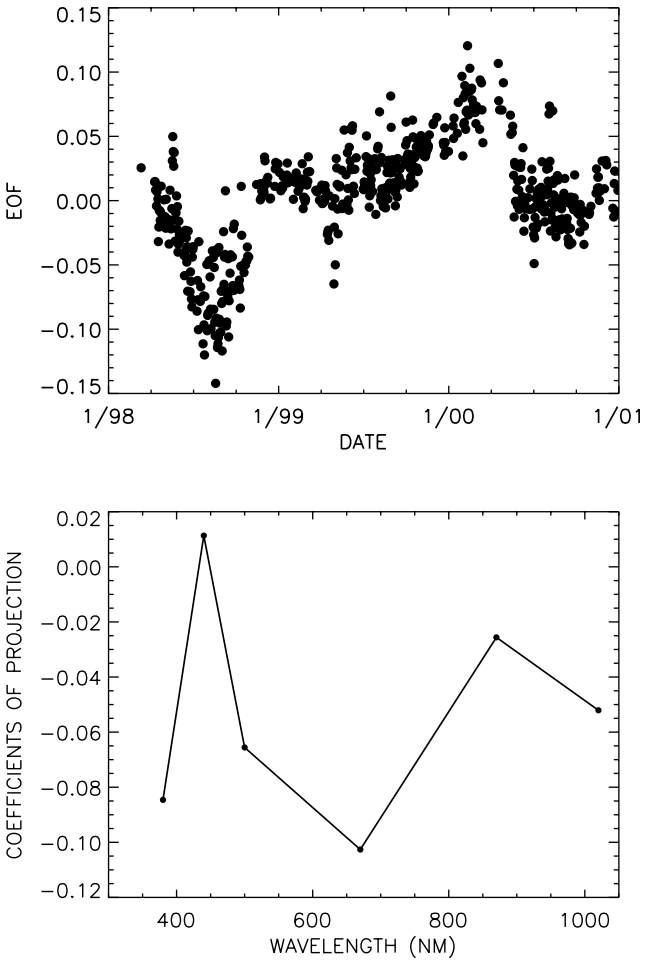


**Figure 15.** Re-introducing the 670 nm filter instead of the 340 nm filter brings back particularly low EOF values in the middle of 1998 and high values in early 2000, suggesting that the data transmitter problems affected the data from the 670 nm filter even more than the data from the other channels.

eigenvector could conceivably contain multiple instrument-related issues if several things went wrong over the course of the data set, and the less obvious ones might be hard to identify. An EOF analysis of RSS data, by contrast, would be more likely to give each source of inhomogeneity its own eigenvector.

[26] In each case examined for this study, information on the behavior of the optical depth measurements of a given instrument was obtained rapidly by performing an empirical orthogonal function analysis on the data set as a complete entity, without going through the more time-consuming process of examining each wavelength separately on each day. This reduces the time needed for data analysis significantly even for the MFRSR and CIMEL, but especially for the thousand-channel RSS.

## References

Bigelow, D. S., J. R. Slusser, A. F. Beaubien, and J. H. Gibson (1998), The USDA Ultraviolet Radiation Monitoring Program, *Bull. Am. Meteorol. Soc.*, *79*, 601–615.

Bodhaine, B. A., N. B. Wood, E. G. Dutton, and J. R. Slusser (1999), On Rayleigh optical depth calculations, *J. Atmos. Oceanic Technol.*, *16*, 1854–1861.

Box, G. P., K. M. Sealey, and M. A. Box (1992), Inversions of Mie extinction measurements using analytic eigenfunction theory, *J. Atmos. Sci.*, *49*, 2074–2081.

Box, G. P., M. A. Box, and J. Krücker (1996), Information content and wavelength selection for multispectral radiometers, *J. Geophys. Res.*, *101*, 19,211–19,214.

Chelliah, M., and P. Arkin (1992), Large-scale interannual variability of monthly outgoing longwave radiation anomalies over the global tropics, *J. Clim.*, *5*, 371–389.

Chen, J., B. E. Carlson, and A. D. Del Genio (2002), Evidence for strengthening of the tropical general circulation in the 1990s, *Science*, *295*, 838–841.

Eck, T. F., B. N. Holben, J. S. Reid, O. Dubovik, A. Smirnov, N. T. O'Neill, I. Slutsker, and S. Kinne (1999), Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, *J. Geophys. Res.*, *104*, 31,333–31,349.

Gianelli, S. M. (2004), Retrieving aerosols, ozone, and $NO_2$ using MFRSR, RSS, and CIMEL data, Ph.D. thesis, Columbia Univ., New York, New York.

Gianelli, S. M., B. E. Carlson, and A. A. Lacis (2005), Aerosol retrievals using rotating shadowband spectroradiometer data, *J. Geophys. Res.*, *110*, D05203, doi:10.1029/2004JD005329.

Harrison, L., J. Michalsky, and J. Berndt (1994), Automated multifilter rotating shadow-band radiometer: An instrument for optical depth and radiation measurements, *Appl. Opt.*, *33*, 5118–5125.

Harrison, L., M. Beauharnois, J. Berndt, P. Kiedron, J. Michalsky, and Q. Min (1999), The rotating shadowband spectroradiometer (RSS) at SGP, *Geophys. Res. Lett.*, *26*, 1715–1718.

Holben, B. N., T. F. Eck, I. Slutsker, D. Tanré, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov (1998), AERONET – A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, *66*, 1–16.

Lorenz, E. N. (1956) *Empirical Orthogonal Functions and Statistical Weather Prediction*, MIT Dept. of Meteorology, Science Report 1, 49 pp.

Mitchell, R. M., and B. W. Forgan (2003), Aerosol measurement in the Australian outback: Intercomparison of sun photometers, *J. Atmos. Ocean. Tech.*, *20*, 54–66.

Peixoto, J. P, and A. H. Oort (1992), *Physics of Climate*, American Institute of Physics, College Park, MD.

Rabbette, M., and P. Pilewskie (2001), Multivariate analysis of solar spectral irradiance measurements, *J. Geophys. Res.*, *106*, 9685–9696.

Remer, L. A., and Y. J. Kaufman (1998), Dynamic aerosol model: Urban/industrial aerosol, *J. Geophys. Res.*, *103*, 13,859–13,871.

Taha, G., and G. P. Box (1999), New method for inferring total ozone and aerosol optical thickness from multispectral extinction measurements using eigenvalue analysis, *Geophys. Res. Lett.*, *26*, 3085–3088.

von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge Univ. Press.

B. E. Carlson and A. A. Lacis, NASA Goddard Institute for Space Studies, New York City, NY, USA.

S. M. Gianelli, Department of Applied Physics and Applied Mathematics, Columbia University, New York NASA Goddard Institute for Space Studies, New York City, NY 10027, USA. (smg58@columbia.edu)