

## NOTES AND CORRESPONDENCE

## A Numerical Method for Synthesizing Atmospheric Temperature and Humidity Profiles

MAIA S. TATARSKAIA\*

*Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado*

RICHARD J. LATAITIS AND B. BOBA STANKOV

*National Oceanic and Atmospheric Administration, Environmental Technology Laboratory, Boulder, Colorado*

VIATCHESLAV V. TATARSKII

*Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado*

9 January 1997 and 11 November 1997

## ABSTRACT

A numerical technique is described for synthesizing realistic atmospheric temperature and humidity profiles. The method uses an ensemble of radiosonde measurements collected at a site of interest. Erroneous profiles are removed by comparing their likelihood with prevailing meteorological conditions. The remaining profiles are decomposed using the method of empirical orthogonal functions. The corresponding eigenprofiles and the statistics of the expansion coefficients are used to numerically generate synthetic profiles that obey the same statistics (i.e., have the same mean, variability, and vertical correlation) as the initial dataset. The technique was applied to a set of approximately 1000 temperature and humidity soundings made in Denver, Colorado, during the winter months of 1991–95. This dataset was divided into four cloud classification categories and daytime and nighttime launches to better characterize typical profiles for the eight cases considered. It was found that 97% of the variance in the soundings could be accounted for by using only five eigenprofiles in the reconstructions. Ensembles of numerically generated profiles can be used to test the accuracy of various retrieval algorithms under controlled conditions not usually available in practice.

## 1. Introduction

Many indirect statistical techniques exist for remote retrieval of vertical temperature and humidity profiles in the atmosphere (see, e.g., Westwater 1993; Stankov et al. 1996; Tatarskii et al. 1996). Most of these methods require a knowledge of typical meteorological conditions at a particular location to initialize the retrieval. This site climatology is usually inferred from an ensemble of existing radiosonde measurements. Once the retrieval algorithm is initialized, the accuracy of the resulting retrieval can be evaluated by comparing it with some type of ground truth, typically additional radiosonde measurements.

Practical difficulties arise if the climatological data-

base is too small and, in particular, if it contains errors. Frequently, a sufficiently large database that accurately represents typical meteorological conditions at a particular site is not readily available. Retrieval algorithms initialized using these statistically incomplete datasets are limited in the range of meteorological conditions they can accurately represent. Furthermore, a thorough evaluation of the accuracy of a particular method requires many additional radiosonde measurements that may be difficult to obtain. A more serious problem involves errors in the radiosonde measurements themselves (Pratt 1985; Schwartz and Doswell 1991; Wade 1995). These errors, together with sparse datasets, make it difficult, if not impossible, to adequately evaluate and compare the accuracy of different profile inversion techniques.

This paper describes a numerical method for creating a synthetic database suitable for testing different indirect methods of retrieving humidity and temperature profiles. Ideally, this database should have the same statistical properties as a set of real input profiles but without any errors. In addition, it should be possible to arbitrarily extend this database to generate a more statistically com-

---

\* Current affiliation: National Oceanic and Atmospheric Administration, Environmental Technology Laboratory, Boulder, Colorado.

---

Corresponding author address: Dr. Maia S. Tatarskaia, NOAA/ETL, R/E ET4, 325 Broadway, Boulder, CO 80303.  
E-mail: mtatarskaia@etl.noaa.gov

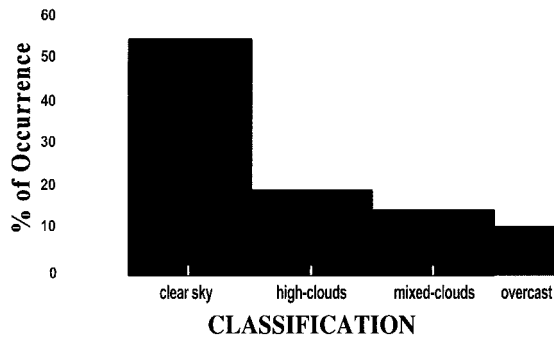


FIG. 1. Percent of occurrence of the four cloud classification categories for edited data (761 total profile sets).

plete set. The ability to generate such a synthetic dataset to serve as an error-free ground truth would be useful in evaluating the potential of various remote sensing techniques.

A set of radiosonde measurements for the site and meteorological conditions of interest is used to create such an ideal database. Erroneous profiles are removed by examining each set of temperature and humidity profiles from the point of view of the likelihood of the particular set, that is, how the profile set corresponds to the meteorological conditions at the time of launch,

how the peculiarities of the individual profiles correspond to each other and to the wind profile, and how natural they look. After removing the erroneous profiles using this synoptic approach, the remaining profiles are decomposed using the method of empirical orthogonal functions (EOFs) (e.g., Obukhov 1960). The statistics of the corresponding expansion coefficients are then used to numerically generate any number of synthetic profiles that obey the same statistics (i.e., profiles that have the same mean, variability, and vertical correlation) as the initial dataset.

The decomposition of radiosonde profiles into a set of EOFs is not a new idea. Tatarskaia (1974) used this approach in a retrieval algorithm to statistically characterize atmospheric temperature and humidity profiles near Dolgoprudnyy, Russia. A similar analysis was carried out by Liu et al. (1991), who examined the dominant humidity eigenprofiles (i.e., the dominant eigenfunctions or eigenvectors) for a variety of ocean locations and timescales.

An EOF approach was used to numerically generate stochastic temperature and humidity profiles representative of meteorological conditions at a particular site. The EOF procedure was applied to an ensemble of approximately 1000 sets of temperature and humidity profiles collected in Denver during the winter months of 1991–95. Because of the significant variability of the profiles, the dataset was divided into four subgroups corresponding to the four dominant cloud conditions observed during these periods (clear sky, high clouds,

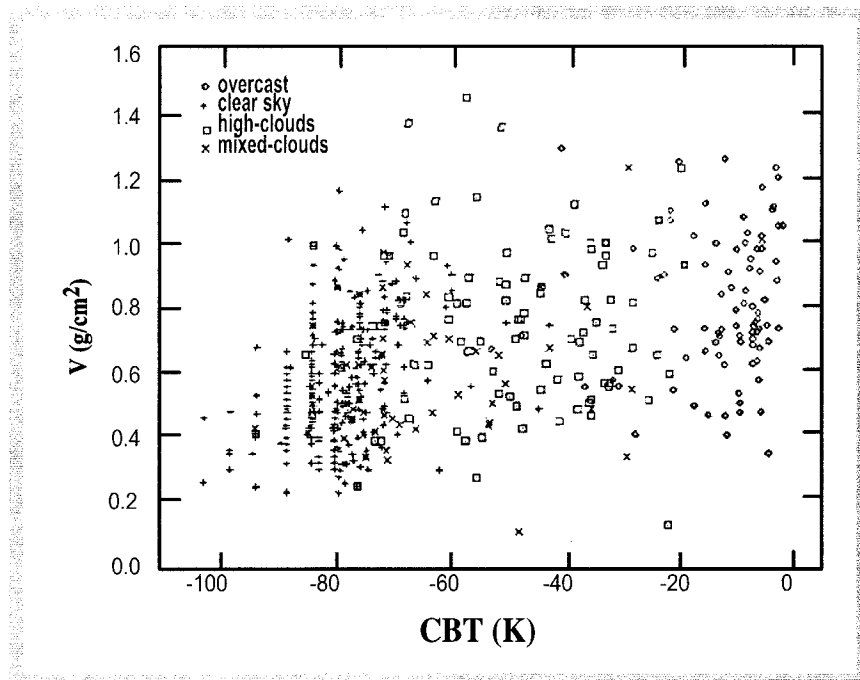


FIG. 2. Scatterplot of the radiosonde measurements as a function of the cloud-base temperature (CBT) and integrated water vapor ( $V$ ). Cloud-type/extent classifications are indicated.

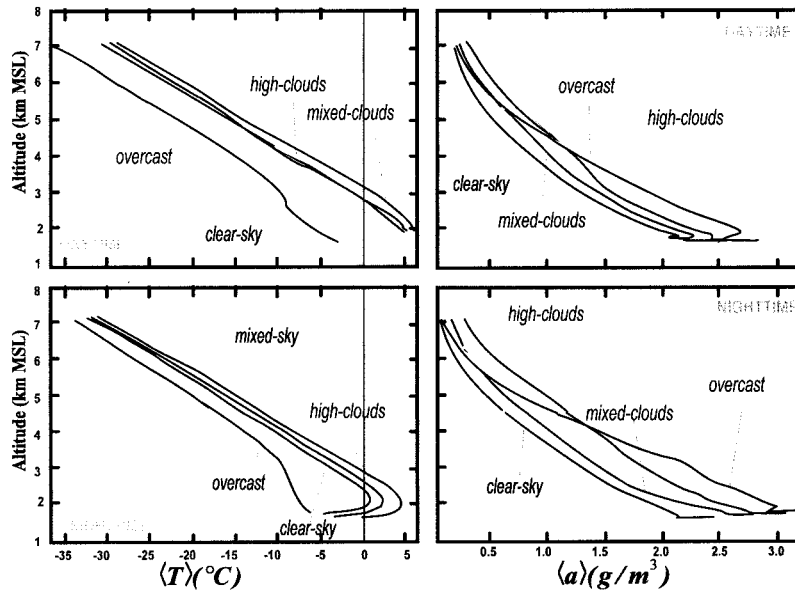


FIG. 3. Vertical profiles of mean temperature  $\langle T \rangle$  and absolute humidity  $\langle a \rangle$  for daytime and nighttime launches. The cloud-type/extent classifications are indicated. (MSL = above mean sea level.)

mixed clouds, and overcast). Each subset was further divided to reflect daytime–nighttime variability. This classification scheme described well the different types of profiles observed in this study. Of course, it is possible to use a classification scheme based on other parameterizations such as large-scale flow patterns deduced from synoptic maps, but the cloud-type classification appeared the most straightforward and only required observations made at a single station. Using a synoptic approach, erroneous profiles from each subset were removed, leaving 761 profile sets. The EOF technique was applied to each subgroup. The statistics of the original profile sets and examples of the synthetically generated profiles with corresponding statistics for selected cases are presented.

The ability to generate temperature and humidity profiles numerically whose statistics match a set of meteorological conditions for a particular site has two additional advantages. The first is the possibility of synthesizing an ensemble of profiles for a location that has no direct measurements by broadly matching meteorological and topographic conditions with a location for which the measurement statistics have been computed. Even though the meteorological and topographic conditions at two sites may not be identical, this offers a place to start in regions of sparse data. An example is the ocean's marine layer, where results from a single controlled experiment spanning a wide variety of meteorological conditions could be applied to similar ocean locations around the world. A second advantage is the

ability to study numerically the sensitivity of a particular profile retrieval technique to measurement errors. For example, a known level of noise could be introduced into the generated profiles and propagated through an inversion algorithm to determine the impact on the retrieval.

## 2. Origin of temperature and humidity errors

Temperature profiles measured by the National Weather Service (NWS) are usually accurate to within 1°C. On occasion, however, they contain errors at the lowest heights. These errors are often caused by the difference in the outdoor temperature and the temperature of the enclosure from which the radiosonde is released, and the relatively long time constant of the temperature sensor. It is nearly impossible to correct for this error because initial conditions at the time of launch (i.e., the radiosonde enclosure temperature) are rarely recorded independently. Humidity profiles are also subject to this error. In addition, it is not unusual for the humidity sensor to be damaged or incorrectly calibrated for low (>20%) and high (>80%) relative humidities. Consequently, humidity profiles very often contain significant errors. In contrast to temperature errors, humidity errors can occur at all altitudes. It is beyond the scope of this paper to outline all of the factors that affect the accuracy of radiosonde temperature and humidity measurements. The interested reader is referred to pa-

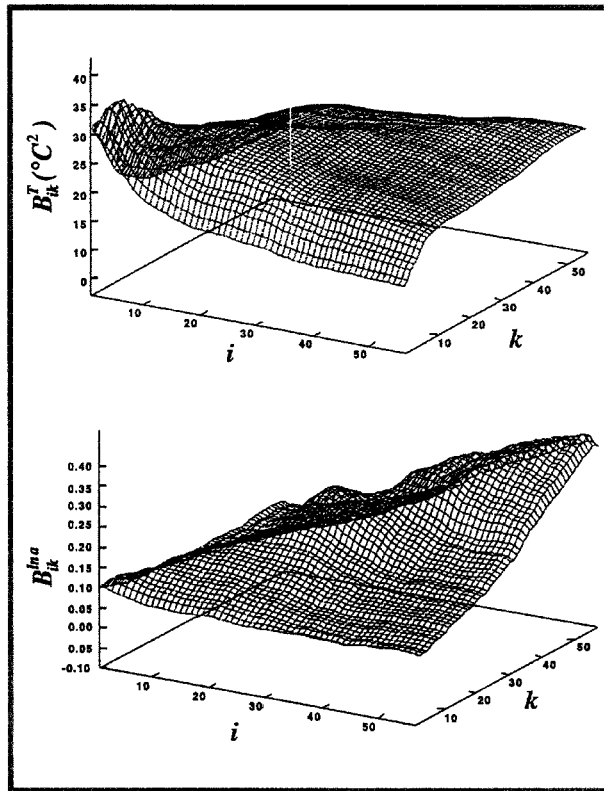


FIG. 4. Temperature  $T$  (top) and logarithmic humidity  $\ln a$  (bottom) correlation matrix  $B_{ik}$  as a function of indices  $i$  and  $k$ . Data shown are for nighttime launches and clear-sky conditions.

pers by Pratt (1985), Schwartz and Doswell (1991), and Wade (1995).

### 3. Use of synoptic data to identify erroneous profiles

Approximately 1000 pairs of temperature and humidity radiosonde profiles collected in Denver, Colorado, during the winter months (December–March) of 1991–95 were analyzed. The radiosondes were launched at 2300 UTC (1600 LT) and 1100 UTC (0400 LT), producing one daytime and one nighttime record. Clearly the profile statistics depend on meteorological conditions at the time of the launch. The problem was how to identify the primary meteorological conditions that best characterized the profiles in relatively broad terms. An examination of the data suggested that the type and extent of cloud cover were the best indicators of the profile statistics. Therefore, the following four classification categories were adopted:

- clear sky (<30% sky coverage),
- high clouds (>5-km altitude/cirrus or altocumulus/>30% sky coverage),
- mixed clouds (different cloud types at indeterminate levels/30%–50% sky coverage), and
- overcast (<5-km altitude/nimbostratus, cumulonimbus, cumulocongestus/>80% sky coverage).

Each profile set was placed into one of these categories. The data were also separated by time of launch (i.e., daytime or nighttime). We note that clouds of less than

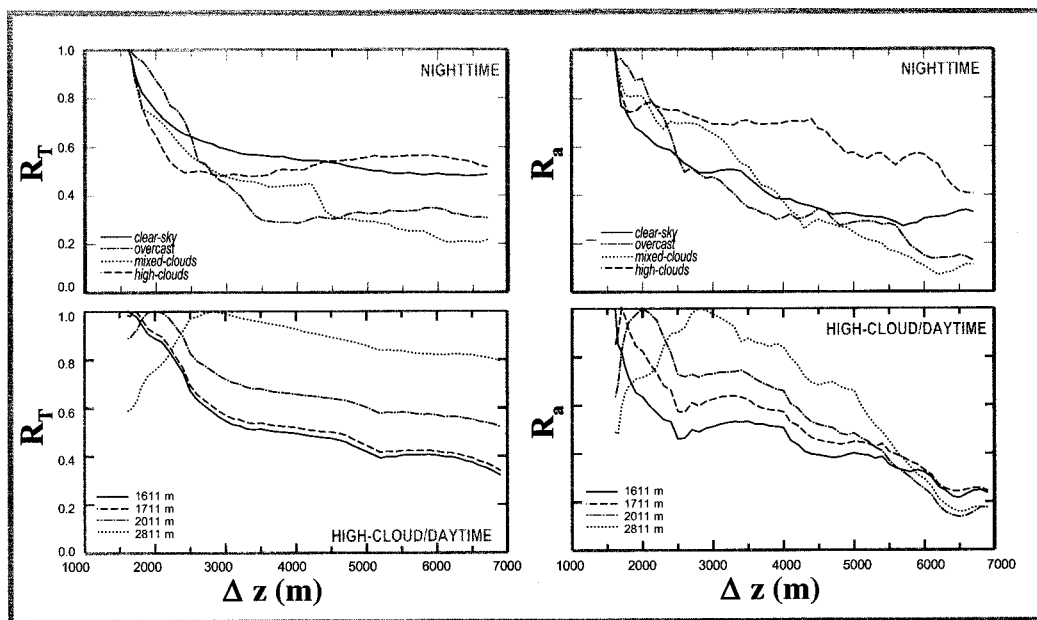


FIG. 5. Normalized vertical correlation function  $R$  for temperature  $T$  and humidity  $a$ . The top panels show the correlation between surface values and higher altitudes for the different cloud-type/extent classifications and nighttime launches. The bottom panels show the correlation as a function of altitude for the high-clouds case and daytime launches. The altitude about which the correlation was calculated is indicated.

5-km altitude and 50%–80% sky coverage are not included in the above classification because this situation was not encountered in this analysis.

Several criteria were used to eliminate erroneous profiles. For example, there were many instances when the lowest levels in the profiles were clearly in error because of the enclosure effect described earlier. These profiles were removed from the analysis. On the other hand, although calibration problems clearly affected some humidity profiles for both dry and saturated conditions, these profiles could not be eliminated a priori. Instead we considered how consistent profile shapes were relative to cloud type and extent. For example, if the radiosonde measurements were clearly inconsistent with the set of meteorological conditions indicated by the NWS synoptic maps and tables, the profile set was excluded from the analysis. The compatibility of temperature and humidity inversions, and their consistency with respect to the wind direction, were also considered. For example, westerly winds were typically observed at the 500-mb level when an inversion layer was present in the Denver area. A profile exhibiting a strong inversion in the absence of westerly winds would therefore be suspect. In this manner a total of 761 profile sets were obtained that were considered to be free of errors. A histogram showing the final distribution of the different cloud-type/extent classification occurrences is presented in Fig. 1.

Cloud-base temperature (CBT) and/or integrated water vapor (V) could also have been used to classify the data instead of synoptic observations. Both CBT and V were available from infrared and microwave radiometer measurements, respectively, made at the Denver site. A scatterplot of the radiosonde measurements versus CBT and V is presented in Fig. 2. The cloud-type/extent classification is identified. While there is little correlation between the radiosonde measurements and V, CBT does offer a way to distinguish between overcast and clear-sky conditions. CBT alone, however, will not allow a differentiation between the high- and mixed-cloud cases because both tend to vary over a wide range of CBTs. Therefore, only synoptic data (i.e., cloud type/extent) were used to classify the observations.

**4. The method of empirical orthogonal functions (EOFs)**

The method of EOFs was used to decompose the temperature and humidity profiles into the fewest number of significant eigenfunctions (eigenprofiles). The method uses a correlation matrix computed from radiosonde measurements to initialize an eigenfunction problem. The temperature correlation matrix is given by

$$\mathbf{B}_{ik}^T = \langle T'(z_i)T'(z_k) \rangle, \quad T'(z_i) = T(z_i) - \langle T(z_i) \rangle, \quad i, k = 1, \dots, N, \quad (1)$$

where the superscript  $T$  identifies temperature quantities, the angle brackets  $\langle . . . \rangle$  describe an average over the ensemble of radiosonde profiles in a particular classification,  $T'$  is the deviation of individual temperature profile from its mean value, and  $N$  is the number of levels. Note that  $\mathbf{B}_{ik}^T$  is an  $N \times N$  matrix. The corresponding eigenfunction problem is defined by

$$\sum_{k=1}^N \mathbf{B}_{ik}^T \varphi_\nu^T(z_k) = \lambda_\nu^T \varphi_\nu^T(z_i), \quad (2)$$

where  $\varphi_\nu^T(z_k)$ , ( $\nu = 1, \dots, N$ ) is the eigenprofile evaluated at an altitude,  $z_k$  (see, e.g., Tatarskaia 1974). These equations have nonzero solutions only for specific values of  $\lambda_\nu^T$  called eigenvalues. Equation (2) is a standard eigenvalue equation that can be solved using any one of a number of numerical techniques (e.g., Press et al. 1986; West and Harrison 1989). The temperature correlation matrix  $\mathbf{B}_{ik}^T$  uniquely determines  $\lambda_\nu^T$  and  $\varphi_\nu^T(z_k)$ .

The eigenprofiles  $\varphi_\nu^T(z_k)$  form a complete set of orthonormal basis functions that can be linearly combined to describe the vertical temperature structure of any profile in the initial dataset. The corresponding inversion equation is given by

$$T(z_k) = \langle T(z_k) \rangle + \sum_{\nu=1}^N \alpha_\nu^T \varphi_\nu^T(z_k), \quad (3)$$

where the expansion coefficients  $\alpha_\nu$  are random numbers defined in terms of any given temperature profile by the formula

$$\alpha_\nu^T = \sum_{k=1}^N T'(z_k) \varphi_\nu^T(z_k). \quad (4)$$

It follows from (4) that  $\langle \alpha_\nu^T \rangle = 0$  because  $\langle T'(z_k) \rangle = 0$ . In addition, because the eigenfunctions  $\varphi_\nu^T(z_k)$  are orthonormal [i.e.,  $\sum_{k=1}^N \varphi_\nu^T(z_k) \varphi_\mu^T(z_k) = \delta_{\nu,\mu}$ , where  $\delta_{\nu,\mu}$  is the Kronecker delta function], the expansion coefficients  $\alpha_\nu^T$  are statistically independent for different  $\nu$ ; that is,

$$\langle \alpha_\nu^T \alpha_\mu^T \rangle = 0 \quad \text{if } \nu \neq \mu. \quad (5)$$

Using (1), (2), and (4) it is also apparent that the variance of the coefficient  $\alpha_\nu^T$  is equal to the corresponding eigenvalue  $\lambda_\nu^T$ ; that is,

$$\langle (\alpha_\nu^T)^2 \rangle = \lambda_\nu^T. \quad (6)$$

Therefore, a typical value of  $\alpha_\nu^T$  is  $\sqrt{\lambda_\nu^T}$ , and the most important terms in (3) are the terms that correspond to the largest  $\lambda_\nu^T$ .

Once the eigenfunctions  $\varphi_\nu^T(z_k)$  and eigenvalues  $\lambda_\nu^T$  are computed, the correlation matrix can be reconstructed using the formula

$$\mathbf{B}_{ik}^T = \sum_{\nu=1}^N \lambda_\nu^T \varphi_\nu^T(z_i) \varphi_\nu^T(z_k). \quad (7)$$

For the dataset considered, it can be shown that if five of the dominant eigenfunctions are used (i.e., only those

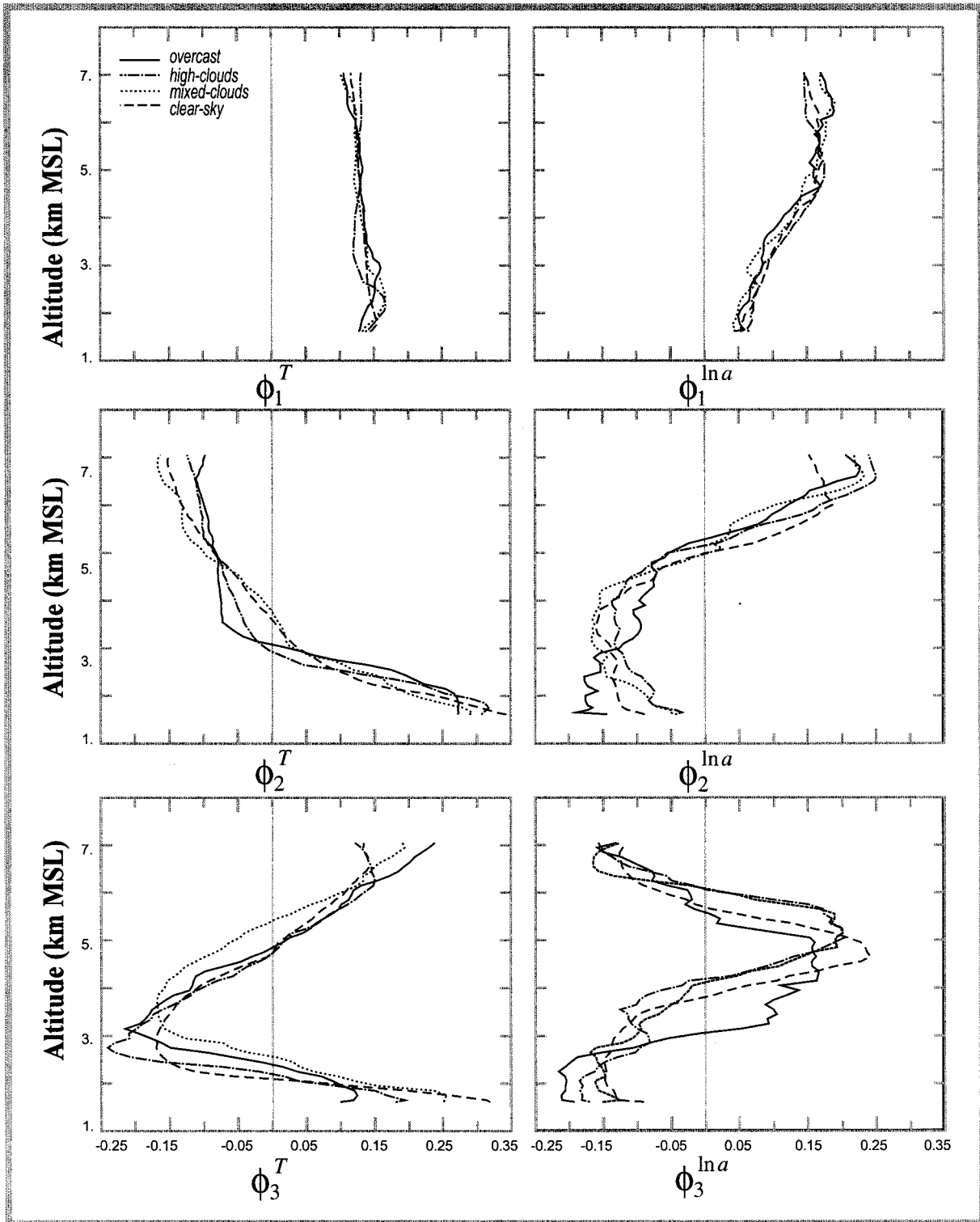


FIG. 6. The first three eigenprofiles of temperature  $\phi_v^T$  and logarithmic humidity  $\phi_v^{\ln a}$  for the different cloud-type/extent classifications and daytime launches.

TABLE 1. Eigenvalues and cumulative eigenvalue weights for the first ten eigenfunctions of the temperature and logarithmic humidity profiles (clear-sky conditions and nighttime launches).

$\nu$	$\lambda_\nu^T$ ( $^{\circ}\text{C}$ ) <sup>2</sup>	$P_\nu^T$	$\lambda_\nu^{\text{ln}a}$	$P_\nu^{\text{ln}a}$
1	1347	0.855	10	0.692
2	129	0.936	2.4	0.855
3	51	0.969	0.99	0.922
4	19	0.981	0.48	0.955
5	13	0.990	0.25	0.972
6	7.0	0.994	0.16	0.983
7	3.9	0.997	0.099	0.990
8	2.6	0.998	0.077	0.995
9	1.7	0.999	0.063	1.00
10	1.0	1.00	0.0038	1.00

eigenfunctions corresponding to the largest eigenvalues  $\lambda_\nu^T$ , the “filtered” correlation matrix differs by less than 3% from the original matrix. Therefore, we should be able to generate statistically accurate profiles using only the dominant eigenprofiles in the reconstructions. Use of the dominant terms in the expansion has the added advantage that the inversion is less sensitive to random errors, which primarily affect the higher-order eigenvalues.

To summarize, if we assume that the expansion coefficients  $\alpha_\nu^T$  are Gaussian random variables, we can numerically generate stochastic temperature profiles using the formula

$$T(z_k) = \langle T(z_k) \rangle + \sum_{\nu=1}^N \sqrt{\lambda_\nu^T} \xi_\nu \varphi_\nu^T(z_k), \quad (8)$$

where  $\xi_\nu$  are zero-mean Gaussian random numbers with unit variance. (The Gaussian approximation for the distribution of  $\xi_\nu$  is examined in more detail in section 5c.) The mean temperature profile  $\langle T \rangle$  and the eigenvalues and eigenfunctions  $\lambda_\nu^T$  and  $\varphi_\nu^T(z_k)$ , respectively, are computed from an ensemble of radiosonde measurements, the latter from the temperature correlation matrix and the corresponding eigenfunction equation given by (2). Humidity profiles can be generated in a similar way with one modification. In contrast to temperature, ex-

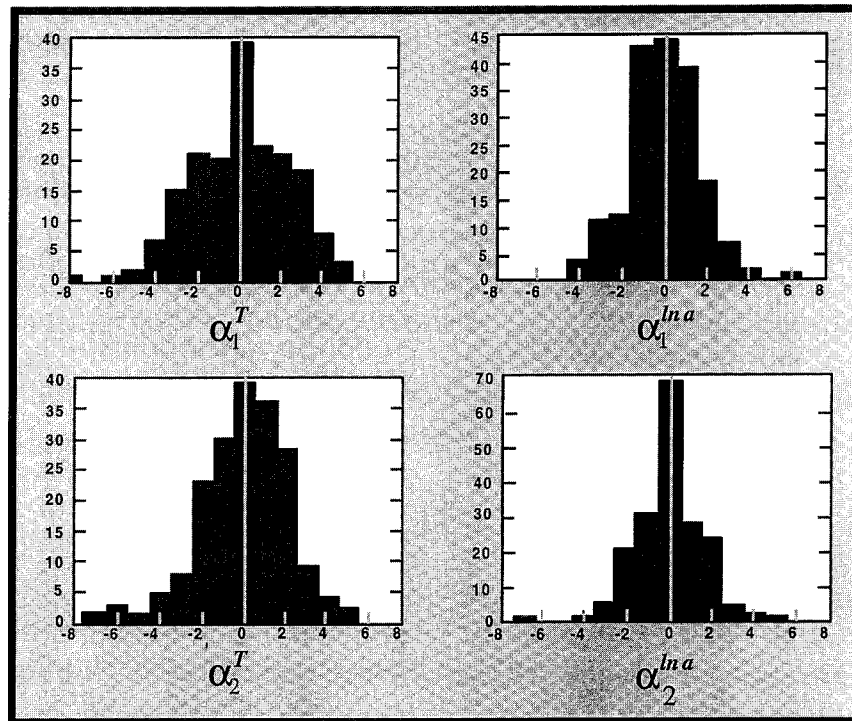


FIG. 7. Histograms of the first two temperature ( $\alpha_1^T$ ,  $\alpha_2^T$ ) and logarithmic humidity ( $\alpha_1^{\text{ln}a}$ ,  $\alpha_2^{\text{ln}a}$ ) expansion coefficients for clear-sky conditions and daytime launches.

cursions in the humidity can be very large, at times larger than the mean. This poses a problem because the humidity expansion equation [i.e., the equation corresponding to (3)] can produce negative values that are not realistic. To circumvent this problem we use the transformation

$$q \equiv \ln a, \tag{9}$$

where  $a$  ( $\text{g m}^{-3}$ ) is the absolute humidity normalized by  $1 \text{ g m}^{-3}$  (E. R. Westwater 1997, personal communication). The EOF decomposition procedure for humidity is then identical to that outlined above using  $q$  instead of  $a$ . At the final step we simply invert the results using (9); that is,

$$a(z_k) = \exp \left\{ \langle \ln a(z_k) \rangle + \sum_{\nu=1}^N \sqrt{\lambda_{\nu}^{\ln a}} \xi_{\nu} \varphi_{\nu}^{\ln a}(z_k) \right\}. \tag{10}$$

### 5. EOF analysis of temperature and humidity profiles

#### a. Mean profiles

Mean profiles of temperature and absolute humidity for daytime and nighttime launches are shown in Fig. 3 for the four cloud-type/extent classifications. The temperature profiles are very similar for all cloud types except overcast. In contrast, the humidity profiles vary significantly with cloud type. The discontinuity in the lowest level of the humidity profiles is due to the enclosure effect described in section 2.

The similarity in the mean temperature profiles for the clear-sky, mixed-clouds, and high-clouds classifications suggests that they can possibly be combined into a single category. To determine if these classifications (and all the others) are indeed statistically unique, the uncertainty in these mean profile estimates needs to be evaluated. The standard deviation  $\sigma$  in the mean profile estimates depicted in Fig. 3 is equal to the standard deviation of the process divided by the square root of the number of profiles in the classification set. Here  $\sigma$  can be estimated by splitting in half each set and calculating the mean profile for each subset. The deviation of the subset profiles about the original mean profile estimate provides a measure of  $\sigma$ . In all cases except one,  $\sigma$  was found to be less than the difference between any two mean profile estimates, suggesting that the proposed classification scheme is reasonable.

#### b. Vertical correlation of temperature and humidity

Vertical correlation matrices for temperature and logarithmic humidity were calculated for 57 levels. The altitude increment was chosen to be 50 m for the lowest levels ( $1611 \text{ m} < z < 1811 \text{ m}$ ) and 100 m for upper levels ( $1811 \text{ m} < z < 7011 \text{ m}$ ). The lowest level at 1611 m corresponds to the surface measurement. Figure

4 shows examples of the correlation matrices (nonnormalized) for  $T$  and  $\ln a$  for clear-sky and nighttime conditions.

Examples of the normalized vertical correlation function  $R$  for  $T$  and  $a$  (not  $\ln a$ ) are shown in Fig. 5. The top panels show the correlation between surface values and higher altitudes for the different cloud-type/extent classifications and nighttime launches. The vertical correlation length (e.g., the vertical separation over which the normalized correlation falls from 1 to 0.5) for temperature appears to be on the order of 1.5 km for all cases except clear sky, which has a larger correlation length ( $\sim 3 \text{ km}$ ). The larger correlation length for clear-sky conditions is due to the larger, more homogeneous air mass associated with the anticyclonic conditions that produce clearer skies. The shorter correlation lengths for cloudy conditions are the result of a more inhomogeneous air mass produced by the variable temperature stratification induced by the clouds. The humidity correlation length tends to be slightly larger than that for temperature for the high- and mixed-cloud cases, about the same for the overcast case, and slightly smaller for clear skies. In general, however, our experience indicates that humidity correlation lengths tend to be slightly larger than those for temperature. The reason is that temperature is a local quantity while absolute humidity is a conserved quantity and is more a property of the entraining air mass.

The bottom panels show the correlation as a function of altitude for the high-cloud case and daytime launches. The peak in the correlation functions identifies the altitude about which the correlation function was calculated (i.e., 1611, 1711, 2011, and 2811 m). The temperature and humidity correlation lengths both tend to increase with altitude, which reflects the diminished influence of topography at greater heights. The increase in the humidity correlation length with altitude is less pronounced than that for temperature because of its conservative nature.

#### c. Eigenprofiles, eigenvalues, and expansion coefficients for temperature and logarithmic humidity

Figure 6 depicts the dominant temperature eigenprofiles  $\varphi_{\nu}^T$  ( $\nu = 1, 2, 3$ ) and logarithmic humidity eigenprofiles  $\varphi_{\nu}^{\ln a}$  ( $\nu = 1, 2, 3$ ) for the different cloud-type/extent classifications and daytime launches. These first three eigenprofiles describe most of the large-scale meteorological variability in the temperature and humidity profiles. Higher-order eigenprofiles describe smaller-scale vertical variations.

Eigenvalues corresponding to the temperature and logarithmic humidity eigenprofiles are presented in Table 1 for clear-sky conditions and nighttime launches. To study the cumulative contribution of the eigenprofiles  $\varphi_{\mu}$  ( $\mu = 1, \dots, \nu$ ) to the total profile shape, a cumulative eigenvalue weight is defined as



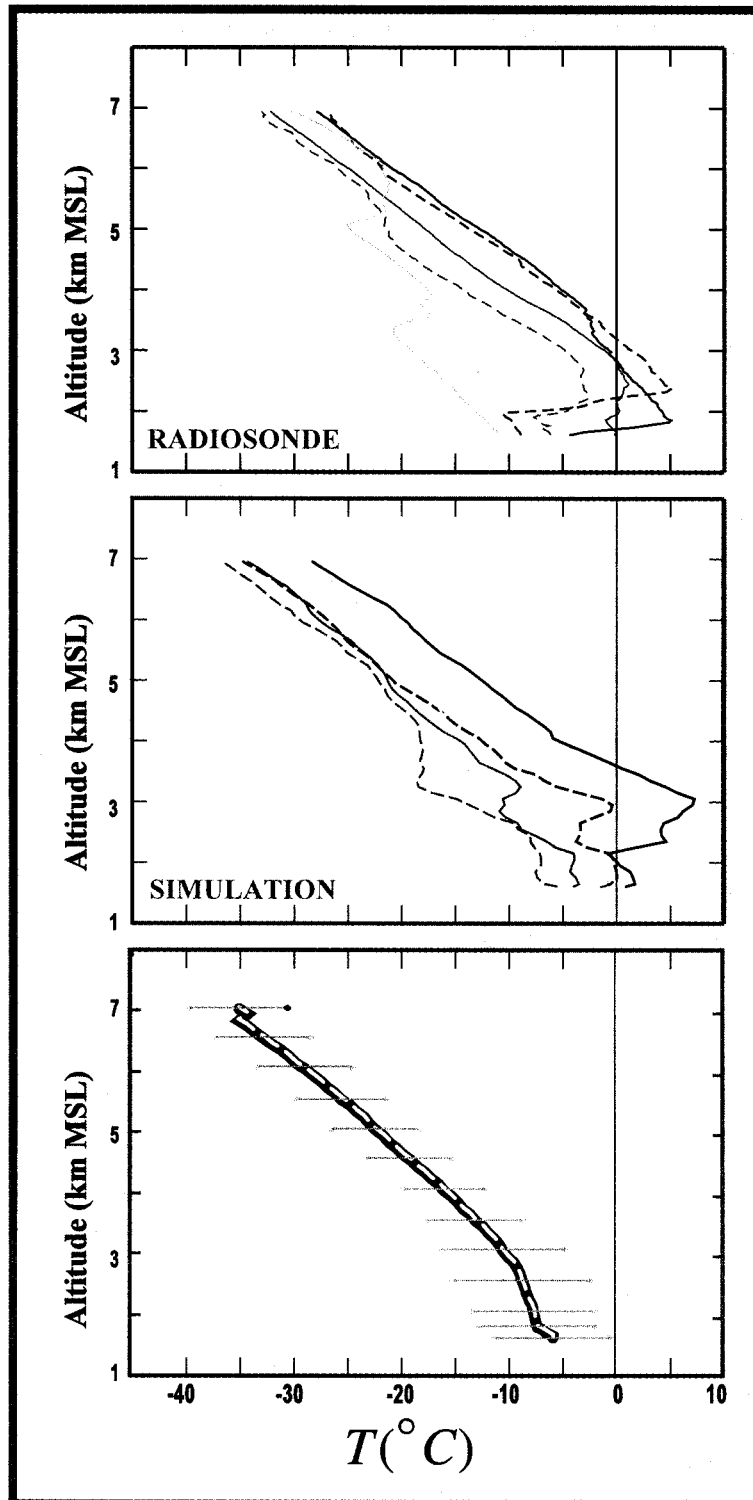


FIG. 8. Comparison of five radiosonde *temperature* records (top panel) and corresponding simulated profiles (middle panel) for overcast conditions and nighttime launches. The mean radiosonde profile (black solid line) and mean simulated profile (gray dashed line) are shown in the bottom panel for 10 000 simulations. The horizontal lines depict the standard deviation of temperature as a function of height for both the radiosonde measurements and simulations.

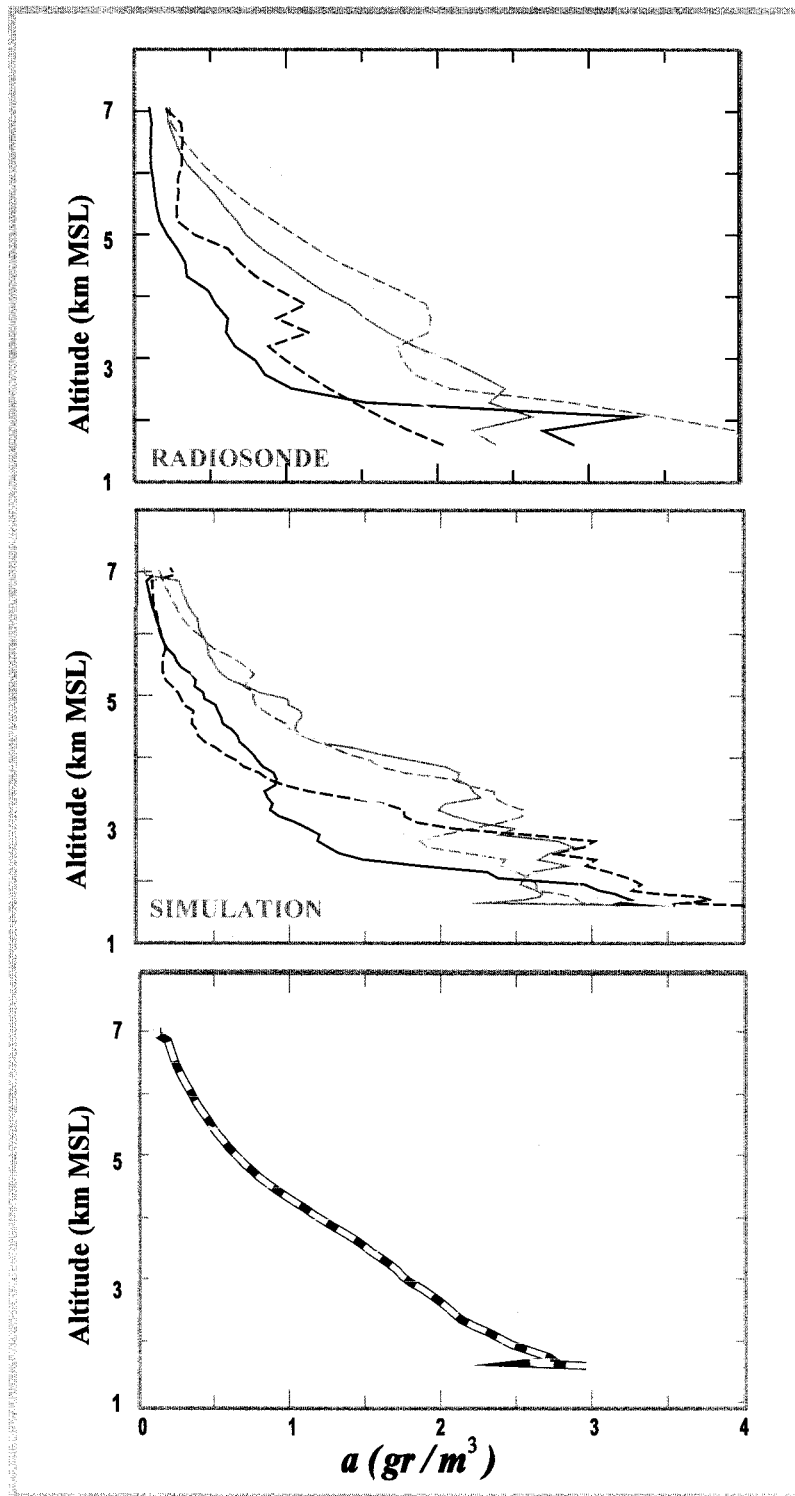


FIG. 9. Comparison of five radiosonde humidity records (top panel) and corresponding simulated profiles (middle panel) for overcast conditions and nighttime launches. The mean radiosonde profile (black solid line) and mean simulated profile (gray dashed line) are shown in the bottom panel for 10 000 simulations. The horizontal lines depict the standard deviation of humidity as a function of height for both the radiosonde measurements and simulations.

$$P_\nu = \frac{\sum_{\mu=1}^{\nu} \lambda_\mu}{\sum_{\mu=1}^N \lambda_\mu}, \quad (11)$$

where  $N = 10$  is the total number of eigenprofiles used in the reconstructions (out of a possible 57). Computed values for  $P_\nu$  are shown in Table 1. The cumulative weights for  $\nu = 10$  differed from unity by less than  $10^{-4}$  and were set equal to 1. The first three or four eigenfunctions describe most of the dominant structure in the profiles. Here  $P_\nu$  also describes the cumulative contribution to the total observed variance (i.e., the sum of the variance over all levels) of the radiosonde launches. Table 1 indicates that five (seven) eigenprofiles account for 99% of the variance observed in the temperature (humidity) profiles. Similar results were obtained for the other cloud classification categories.

Figure 7 depicts histograms of the expansion coefficients of the first two temperature eigenprofiles ( $\alpha_1^T$ ,  $\alpha_2^T$ ) and logarithmic humidity eigenprofiles ( $\alpha_1^{\ln a}$ ,  $\alpha_2^{\ln a}$ ) for the clear-sky case and daytime launches. The expansion coefficients approximate zero-mean Gaussian random variables. Although no goodness-of-fit test was applied, the observed distributions support the use of the Gaussian random numbers  $\xi_\nu$  that appear in the synthesis formulas (8) and (10).

## 6. Numerically generated temperature and humidity profiles

The initial set of 761 radiosonde measurements was used together with (8) and (10) to generate a corresponding set of synthetic profiles for the four cloud-type/extent classifications. The mean profiles, eigenprofiles, and corresponding eigenvalues for the dominant 10 terms in the expansion for the different cloud-type/extent classifications and daytime–nighttime conditions were calculated. These parameters completely characterize the profiles observed during the winter months of 1991–95 in Denver, Colorado. They are not presented here but are available upon request.

A comparison of the simulated and radiosonde profiles for temperature and humidity is shown in Figs. 8 and 9, respectively. The top panels contain five typical radiosonde records for overcast conditions and nighttime launches. Examples of the corresponding numerically generated profiles are shown in the middle panels. Individual profiles do not agree and are not expected to agree. Rather, it is the statistics of the profiles that are equivalent. The solid black line in the bottom panel shows the mean of all the radiosonde profiles in this category. The gray dashed line is the average of  $S = 10\,000$  simulations and falls directly on top of the mean radiosonde profile as expected. The horizontal lines depict the standard deviation of

the temperature and humidity for both the radiosonde measurements and simulations at the given altitude. The standard deviation of the simulated profiles is essentially identical to that for the radiosonde profiles. The degree of accuracy of the simulations is arbitrary and is proportional to  $1/\sqrt{S}$ . For 10 000 simulations we expect an accuracy of 1%, and there should be no visible difference between the mean radiosonde and simulation profiles and the corresponding standard deviations.

## 7. Summary and conclusions

We have described a numerical method for generating an ensemble of temperature and humidity profiles characteristic of a particular site and climatology. The method is based on an initial set of radiosonde measurements that has been edited to remove erroneous profiles. The remaining profiles are then decomposed using the method of empirical orthogonal functions, and the statistics of the expansion coefficients and corresponding eigenprofiles are used in a reconstruction algorithm to synthesize new profiles. We applied this technique to a multiyear dataset collected during the winter months of 1991–95 in Denver, Colorado.

This procedure differs fundamentally from the typical inversion problem in the sense that a single synthesized profile may differ significantly from the corresponding individual radiosonde measurements. Rather, it is the statistics of a set of generated profiles (i.e., the mean, variability, and vertical correlation) that correspond to the statistics of the initial dataset.

The statistics of the EOF decomposition completely characterize meteorological profiles for a particular site and climatology. In the future we plan to apply this technique to a multiyear set of radiosonde observations collected in the western Pacific during the Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (Webster and Lukas 1992). If successful, the computed eigenfunctions and corresponding statistics of the expansion coefficients could be used to generate a synthetic database appropriate for many tropical maritime ocean locations.

Numerically generated profiles can be used to test the accuracy of various indirect retrieval algorithms. One advantage of the numerical approach is the ability to generate any number of profiles to be used in a training and comparison set, overcoming the difficulty posed by an initial dataset having a limited number of profiles. A second advantage is the compactness of the approach. For the cases considered, only five eigenprofiles, and the mean and variance of the corresponding expansion coefficients, are all that are needed to fully characterize the profile statistics.

*Acknowledgments.* The authors thank the reviewers for the careful reading of the original manuscript and

their numerous suggestions, which significantly improved the paper.

## REFERENCES

- Liu, W. T., W. Tang, and P. N. Niler, 1991: Humidity profiles over the ocean. *J. Climate*, **4**, 1023–1034.
- Obukhov, A. M., 1960: Statistically orthogonal extensions of empirical functions. *Izv. Akad. Nauk SSSR, Ser. Geofiz.*, **3**, 432–439.
- Pratt, R. W., 1985: Review of radiosonde temperature and humidity problems. *J. Atmos. Oceanic Technol.*, **2**, 404–407.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1986: *Numerical Recipes*. Cambridge University Press, 818 pp.
- Schwartz, B. E., and C. A. Doswell, 1991: North American rawinsonde observations: Problems, concerns, and a call to action. *Bull. Amer. Meteor. Soc.*, **72**, 1885–1896.
- Stankov, B. B., E. R. Westwater, and E. E. Gossard, 1996: Use of wind profiler estimates of significant moisture gradients to improve humidity profile retrievals. *J. Atmos. Oceanic Technol.*, **13**, 1285–1290.
- Tatarskaia, M. S., 1974: Orthogonal expansions of the temperature and humidity fields of the lower troposphere. *Izv. Atmos. Ocean Phys.*, **10**, 290–296.
- Tatarskii, V., M. Tatarskaia, and E. R. Westwater, 1996: Statistical retrieval of humidity profiles from precipitable water vapor and surface measurements of humidity and temperature. *J. Atmos. Oceanic Technol.*, **13**, 165–174.
- Wade, C. G., 1995: An evaluation of problems affecting the measurement of low relative humidity on the United States radiosonde. *J. Atmos. Oceanic Technol.*, **11**, 687–700.
- Webster, P. J., and R. Lukas, 1992: TOGA COARE: The Coupled Ocean–Atmosphere Response Experiment. *Bull. Amer. Meteor. Soc.*, **73**, 1377–1416.
- West, M., and J. Harrison, 1989: *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, 704 pp.
- Westwater, E. R., 1993: Ground-based microwave remote sensing of meteorological variables. *Atmospheric Remote Sensing by Microwave Radiometry*, M. A. Janssen, Ed., J. Wiley and Sons, 145–213.