

## Improving Snowfall Forecasting by Diagnosing Snow Density

PAUL J. ROEBBER AND SARA L. BRUENING

*Atmospheric Science Group, Department of Mathematical Sciences, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin*

DAVID M. SCHULTZ\* AND JOHN V. CORTINAS JR.\*

*Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 9 April 2002, in final form 5 August 2002)

### ABSTRACT

Current prediction of snowfall amounts is accomplished either by using empirical techniques or by using a standard modification of liquid equivalent precipitation such as the 10-to-1 rule. This rule, which supposes that the depth of the snowfall is 10 times the liquid equivalent (a snow ratio of 10:1, reflecting an assumed snow density of  $100 \text{ kg m}^{-3}$ ), is a particularly popular technique with operational forecasters, although it dates from a limited nineteenth-century study. Unfortunately, measurements of freshly fallen snow indicate that the snow ratio can vary from on the order of 3:1 to (occasionally) 100:1. Improving quantitative snowfall forecasts requires, in addition to solutions to the significant challenge of forecasting liquid precipitation amounts, a more robust method for forecasting the density of snow. A review of the microphysical literature reveals that many factors may contribute to snow density, including in-cloud (crystal habit and size, the degree of riming and aggregation of the snowflake), subcloud (melting and sublimation), and surface processes (compaction and snowpack metamorphism). Despite this complexity, the paper explores the sufficiency of surface and radiosonde data for the classification of snowfall density. A principal component analysis isolates seven factors that influence the snow ratio: solar radiation (month), low- to midlevel temperature, mid- to upper-level temperature, low- to midlevel relative humidity, midlevel relative humidity, upper-level relative humidity, and external compaction (surface wind speed and liquid equivalent). A 10-member ensemble of artificial neural networks is employed to explore the capability of determining snow ratio in one of three classes: heavy ( $1:1 < \text{ratio} < 9:1$ ), average ( $9:1 \leq \text{ratio} \leq 15:1$ ), and light ( $\text{ratio} > 15:1$ ). The ensemble correctly diagnoses 60.4% of the cases, which is a substantial improvement over the 41.7% correct using the sample climatology, 45.0% correct using the 10-to-1 ratio, and 51.7% correct using the National Weather Service “new snowfall to estimated meltwater conversion” table. A key skill measure, the Heidke skill score, attains values of 0.34–0.42 using the ensemble technique, with increases of 75%–183% over the next most skillful approach. The critical success index shows that the ensemble technique provides the best information for all three snow-ratio classes. The most critical inputs to the ensemble are related to the month, temperature, and external compaction. Withholding relative humidity information from the neural networks leads to a loss of performance of at least 5% in percent correct, suggesting that these inputs are useful, if nonessential. Examples of pairs of cases highlight the influence that these factors have in determining snow ratio. Given the improvement over presently used techniques for diagnosing snow ratio, this study indicates that the neural network approach can lead to advances in forecasting snowfall depth.

### 1. Introduction

Forecasting snowfall depth is, at present, a two-step problem. First, an assessment must be made of the amount of liquid water that is to fall, the quantitative precipitation forecast (QPF) problem (e.g.,

Fritsch et al. 1998). Second, this liquid water must be converted into snow, hereafter the snow-density problem. Historically, very little research has addressed the snow-density problem, as discussed by Super and Holroyd (1997, p. 20) and Judson and Doesken (2000, section 1). As we argue below, current operational practice of forecasting snow density, and hence snowfall depth, is still largely a non-scientific endeavor. Consequently, even if predictions of QPF were accurate and precise, large errors in snowfall forecasts (by factors of 2–10) could still occur as a result of inaccurate predictions of snow density. In this paper, we present a method to assess the complex problem of snow-density forecasting.

---

\* Additional affiliation: NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma.

---

*Corresponding author address:* Paul J. Roebber, Dept. of Mathematical Sciences, University of Wisconsin—Milwaukee, 3200 N. Cramer Ave., Milwaukee, WI 53211.  
E-mail: roebber@uwm.edu

### a. *The QPF and snow-density problems*

The economic and social value of accurate QPFs are well recognized (e.g., Fritsch et al. 1998). QPFs, however, have historically posed a significant operational forecast challenge (e.g., Charba and Klein 1980) and, in recent years, have exhibited only very modest increases in skill (e.g., Olson et al. 1995; Roebber and Bosart 1998; Fritsch et al. 1998). QPF skill undergoes substantial seasonal fluctuations, with the lowest skills in the warm season in association with the increased potential for convective precipitation (e.g., Olson et al. 1995). Even in the cold season, however, skill remains modest, with mean threat scores for heavy precipitation ( $\geq 25$  mm) and heavy snow ( $\geq 100$  mm  $12$  h $^{-1}$ ) of 0.27 and 0.20, respectively (Olson et al. 1995).

In the cold season, forecasts of precipitation type are required as well, since large sectors of the economy (e.g., transportation, construction, agriculture, commerce) are affected by snow and freezing rain. The cold-season QPF problem requires (in addition to information concerning the myriad factors that govern precipitation generation, type, and amount) insight into the physical processes controlling the depth of the snowfall via the snow density. Snow-density forecasting is important not only for operational weather forecasts for snowfall, but also for avalanche forecasting (Perla 1970; LaChapelle 1980; Ferguson et al. 1990), snowmelt runoff forecasting, snowdrift forecasting, and as an input parameter in the snow accumulation algorithm for the Weather Surveillance Radar-1988 Doppler (WSR-88D) (Super and Holroyd 1997). Accurate forecasts of the depth of the snowfall are critical to many snow removal operations, since these activities are triggered by exceedances of specific snow-depth thresholds (Gray and Male 1981, 671–706; Minsk 1998, 123–132).

### b. *Forecasting snow density: The 10-to-1 rule and current operational practice*

Snow density is often assumed to conform to the 10-to-1 rule: the snow ratio, defined by the density of water ( $1000$  kg  $m^{-3}$ ) to the density of snow (assumed to be  $100$  kg  $m^{-3}$ ), is 10:1. For example, if a forecaster believes that 1 in. of liquid water will fall at a given point over a specified time, then the 10-to-1 rule implies a total snowfall of 10 in. As noted by Judson and Doesken (2000), the 10-to-1 rule appears to originate from the results of a nineteenth-century Canadian study. Potter (1965, p. 1) quotes from this study: “A long series of experiments conducted by General Sir H. Lefroy, formerly Director of the Toronto Observatory, led to the conclusion that this relation [10 to 1] is true on the average. It is not affirmed that it holds true in every case, as snow varies in density. . . .” The 10-to-1 rule has persisted, however, despite the almost immediate warnings concerning its accuracy (Abbe 1888, p. 386).

More comprehensive measurements of snow density

(e.g., Currie 1947; LaChapelle 1962, reproduced in Doesken and Judson 1997, p. 15; Power et al. 1964; Super and Holroyd 1997; Judson and Doesken 2000) have established that this rule is an inadequate characterization of the true range of snow densities. Snow densities can vary from  $10$  kg  $m^{-3}$  (a snow ratio of 100:1) to approximately  $350$  kg  $m^{-3}$  (2.9:1). In interviews with current National Weather Service (NWS) forecasters from around the country, we find that the 10-to-1 rule may be modified slightly to 12 to 1 or 20 to 1, depending on the mean or median climatological value of snow density at a particular station (e.g., Currie 1947; Holroyd and Super 1997, p. 27). The concept of uniformly applying a fixed snow ratio, though, is typically employed. Thus, the persistence of the 10-to-1 rule has served merely to exacerbate the snow-forecast problem by oversimplifying the determination of snow density.

Attempts to specify snow density, usually through reference to in-cloud or surface air temperatures, have shown only marginal diagnostic capability, even when using carefully collected measurements (e.g., Grant and Rhea 1974; Super and Holroyd 1997; Kyle and Wesley 1997; Judson and Doesken 2000). Super and Holroyd (1997, p. 27) found that the single best prediction method for snow density is persistence, but this still only explains about 30% of the variance. In the absence of explicit snow-density forecasts, some empirical techniques have evolved (e.g., Chaston 1989; Garcia 1994). [See the review in Wetzel and Martin (2001) for a complete list of such techniques.] As argued in Schultz et al. (2002), however, these techniques are inadequate and remain unverified for a large variety of events. To gain further understanding of the snowfall forecasting problem, we next review the physical processes that affect snow density.

### c. *Factors affecting snow density*

The density of snowfall is related to the ice-crystal structure by virtue of the relative proportion of the occupied volume of crystal composed of air. Snow density is regulated by (a) in-cloud processes that affect the shape and size of growing ice crystals, (b) subcloud processes that modify the ice crystal as it falls, and (c) ground-level compaction due to prevailing weather conditions and snowpack metamorphism. Understanding how these processes affect snow density is difficult because direct observations of cloud microphysical processes, thermodynamic profiles, and surface measurements are often unavailable.

Cloud microphysical research indicates that many factors contribute to the final structure of an ice crystal. One factor that has been shown to affect snow density is the habit (shape) of the ice crystal. In an environment supersaturated with respect to ice, incipient ice particles will grow to form one of many types of crystal habits (e.g., Magono and Lee 1966; Ryan et al. 1976; Fukuta and Takahashi 1999). The habit is determined by the

environment in which the ice crystal grows: specifically, the surrounding air temperature and degree of supersaturation with respect to ice and liquid water. For example, at temperatures of  $0^{\circ}$  to  $-4^{\circ}\text{C}$ , plates dominate; at  $-4^{\circ}$  to  $-10^{\circ}\text{C}$ , prisms, scrolls, sheaths, and needles dominate, with the specific habit at these temperatures controlled by the degree of supersaturation; at  $-10^{\circ}$  to  $-20^{\circ}\text{C}$ , thick plates, sectors, and dendrites are found, depending on supersaturation; and at temperatures lower than  $-20^{\circ}\text{C}$ , hollow columns and sheaths dominate. Power et al. (1964) showed that pure dendrites have the lowest density, although the variation in the density of dendritic aggregates is large [from approximately 5 to  $100\text{ kg m}^{-3}$ ; Magono and Nakamura (1965); Passarelli and Srivastava (1979)]. Although needles would be expected to be dense, they are the next lighter crystal habit since they tend to aggregate at the relatively high temperatures at which they form (Power et al. 1964). Finally, columns and plates were the next densest form. Snow, however, is not a homogeneous collection of ice crystals. As an ice crystal falls, it passes through varying thermodynamic and moisture conditions, with each new crystal habit being superimposed upon the previous structure, such that the final habit is a combination of various growth modes (e.g., Pruppacher and Klett 1997, p. 44).

Another factor affecting density is crystal size. Large dendritic crystals will occupy much empty air space, whereas smaller crystals will pack together into a denser assemblage. The ultimate size of crystals depends on factors that affect the growth rate, such as the residence time in the cloud and the degree of supersaturation with respect to ice. For a given degree of supersaturation with respect to ice, ice-crystal growth by vapor diffusion is dependent primarily on temperature, as well as secondary pressure effects (e.g., Byers 1965; Ryan et al. 1976; Takahashi et al. 1991; Fukuta and Takahashi 1999). At 1000 hPa, the growth rate maximizes near  $-14^{\circ}\text{C}$ , whereas at 500 hPa, the rate maximizes near  $-17^{\circ}\text{C}$  (e.g., Byers 1965, p. 123). The maximum growth rate is expected to occur near the level of maximum upward air motion within the cloud, where the greatest water vapor delivery occurs (Auer and White 1982). Through natural variations in ice-crystal size, some crystals will grow relative to their neighbors within the cloud and begin to fall, thus promoting the sweepout of smaller particles. If this sweepout occurs in a cloud of ice crystals, then aggregation leads to the formation of snowflakes and relatively low snow density. An ice crystal falling through a cloud of supercooled water droplets, on the other hand, will lead to rimed crystals (graupel) through accretional growth, and very high snow densities (e.g., Power et al. 1964; Judson and Doesken 2000).

After leaving the cloud, the snow density can also be affected by sublimation and melting. Sublimation occurs when the air surrounding the crystal is subsaturated with respect to ice, whereas melting occurs when the tem-

perature of the air surrounding the crystal is greater than  $0^{\circ}\text{C}$ . Substantial sublimation or melting of a snowflake can occur over relatively short vertical distances [ $\sim 500\text{ m}$ ; Houze (1993, p. 199)]. Consequently, the liquid water content and the density of a snowflake when it reaches the ground can be strongly dependent upon the low-level air temperature and relative humidity. Hence, the subcloud thermodynamic stratification through which an ice crystal falls also will be a critical factor in determining snowfall density.

Finally, once the snow lands on the ground, compaction of the ice crystals due to prevailing weather conditions and snowpack metamorphism may occur. Wind greater than approximately  $9\text{ m s}^{-1}$  can move ice crystals at the surface, fracturing the crystal during saltation and causing surficial compaction, increasing snow density (e.g., Gray and Male 1981, 345–350). The weight of the snowfall can further compress the snowpack. Within hours of falling, snow metamorphism can occur, in which water sublimates from the sharper edges of the ice crystals and deposits on the more rounded edges, making the snow crystals more rounded and dense (e.g., Gray and Male 1981, 277–285; Doesken and Judson 1997, 18–19). Such rounding is accelerated at temperatures approaching freezing.

#### *d. A new approach*

Although the complexity of the snow-density problem is considerable, there is some expectation that the bulk effects of these processes can be assessed through reference to the temperature and moisture profiles within and below the cloud. The existing observational network does not provide such profiles in great temporal or spatial detail (e.g., radiosonde measurements are taken only twice daily with approximate station separations of 300–400 km in the United States); however, the regional character of both air masses and measured snow densities (e.g., Potter 1965; Judson and Doesken 2000) suggests the possibility of isolating a useful signal from the existing, albeit deficient, observational record. Such a procedure would clearly have great operational utility. Hence, in this paper, the feasibility of the diagnosis of snow density using only routinely available measurements is addressed.

The difficulty of measuring snowfall and its implications for this study are discussed in section 2. The approach used to diagnose snow density, employing artificial neural networks (ANNs), is discussed in section 3. An interpretation and discussion of these results is offered in section 4. A summary of the findings of this paper is given in section 5.

## **2. Data**

The purpose of this study is to develop a methodology to diagnose snow density using operational data collected by the NWS over areas of the continental United

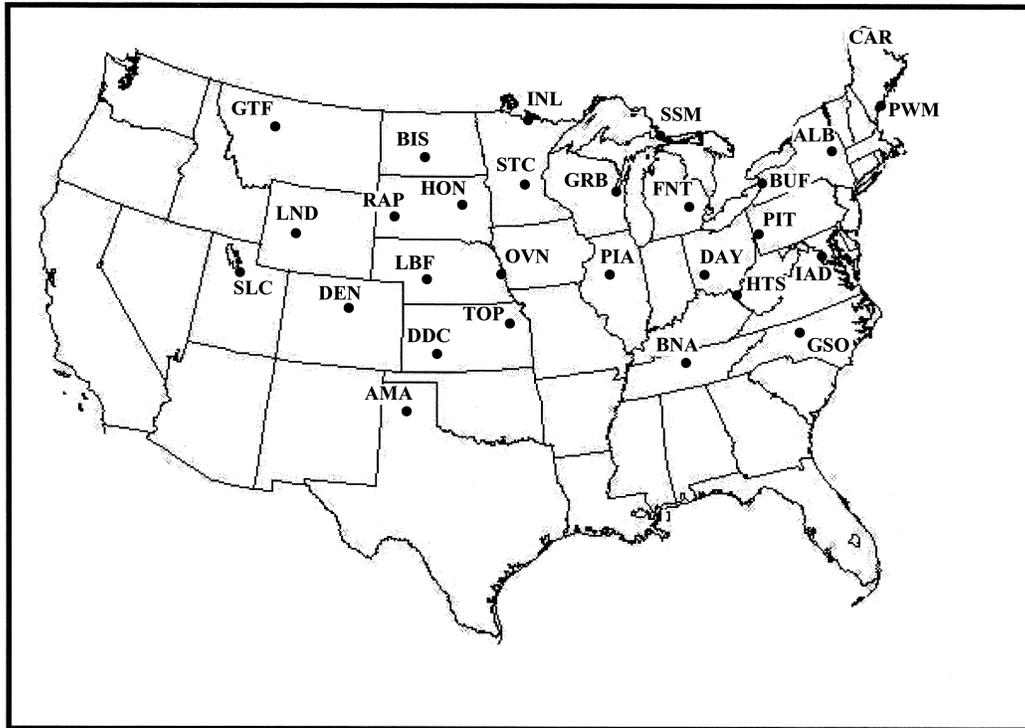


FIG. 1. Locations of the 28 radiosonde sites with collocated surface reports used in the study.

States receiving frequent, measurable snowfall. The review of factors affecting snow density in section 1 suggests that upper-air observations of temperature and humidity, and surface observations of temperature, wind speed, and precipitation, are needed. Thus, snowfall events for this study must be collocated with an upper-air observing site. Sounding data from these sites were extracted from the National Climatic Data Center/Forecast Systems Laboratory (NCDC/FSL) North American radiosonde dataset (Schwartz and Govett 1992; also see [http://raob.fsl.noaa.gov/Raob\\_Software.html](http://raob.fsl.noaa.gov/Raob_Software.html) on line).

The new-snow amounts (6-h values) come from the United States Air Force DATSAV2 Surface Climatic database, whereas the NCDC hourly precipitation dataset TD-3240 provided the liquid equivalent precipitation. Two different datasets were employed since the DATSAV2 reports generally included only 24-h precipitation measurements. NCDC summary of the day snow reports were examined to ensure consistency between datasets.

Based on potential snowfall events culled from these two datasets, quality control was accomplished in the following manner. First, since temporal variations in density can occur within an event over just a few hours (e.g., Super and Holroyd 1997, 25–27), the reported snow must have occurred within 6 h of the nominal time of a radiosonde launch (0000 or 1200 UTC). A snowfall that persists over a 12-h period, for example, will then be split into 6-h totals and associated with two successive soundings. Second, measuring the depth of new

snow is problematic due to metamorphism, compaction, drifting, the frequency of snow-depth measurement (e.g., Doesken and Judson 1997; Nouhan 1999; Doesken and Leffler 2000), and the type of gauge used (e.g., Goodison 1978; Groisman et al. 1991; Groisman and Legates 1994). To mitigate against errors and imprecision in measurement, the depth of new snow must have measured at least 50.8 mm (2 in.) and the liquid equivalent precipitation at least 2.8 mm (0.11 in.), this latter threshold being defined by a simple error analysis for snow ratio (e.g., Judson and Doesken 2000, p. 1584). Third, to reduce measurement errors associated with substantial snow drift, only events for which surface winds were less than or equal to  $9 \text{ m s}^{-1}$  at the time of the radiosonde report were considered (Gray and Male 1981, 345–350). This threshold is within the ranges for wet and dry snows reported by Li and Pomeroy (1997). The quality control measures produced 1650 snowfall events over the 22-yr period 1973–94 for the 28 stations shown in Fig. 1.

The values of the snow ratios for the 1650 events range from 1.9:1 to 46.8:1 (Fig. 2a) and the values for snow densities range from 21.4 to 526.3  $\text{kg m}^{-3}$  (Fig. 2b). Both snow ratio and snow density are provided for comparison with past studies. In so doing, an apparent paradox arises since snow ratio and snow density have a reciprocal relationship; that is, 10 units of ratio do not equate to a fixed number of density units (e.g., a snow-ratio bin of 6–8:1, equaling a density of 125–167  $\text{kg m}^{-3}$ , compared to a snow-ratio bin of 22–24:1, equaling

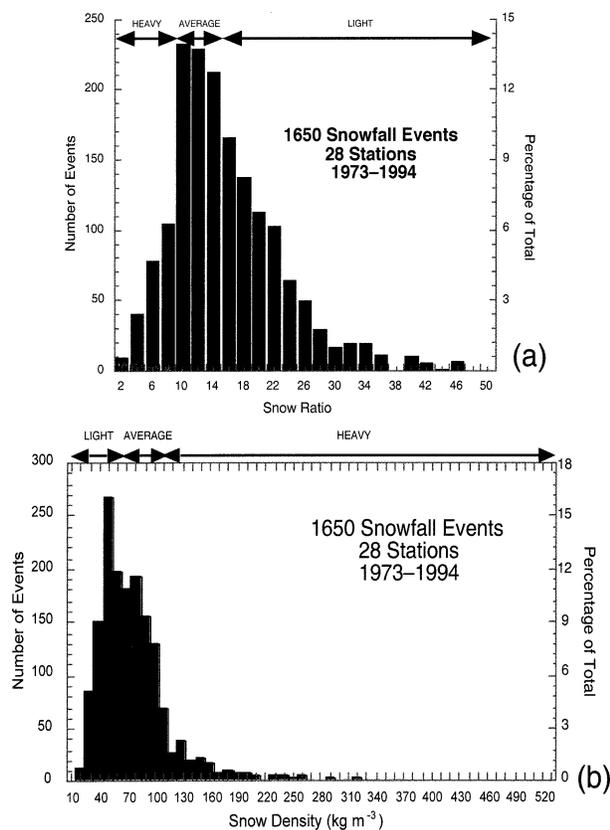


FIG. 2. Histograms for the dataset used in this paper (1650 snowfall events over 28 stations during 1973–94). The y axis is indicated with the number of events on the left side of the graph and the percentage of total events on the right side of the graph. Light, average, and heavy are defined in the text. (a) Snow ratio (in two-unit bins). (b) Snow density (in  $10 \text{ kg m}^{-3}$  bins).

a density of  $41.7\text{--}45.5 \text{ kg m}^{-3}$ ). The mean and median snow ratios are 15.6:1 and 14.1:1, respectively, whereas the mean and median snow densities are 80.9 and  $70.9 \text{ kg m}^{-3}$ , respectively. Figure 2a shows that the mode snow ratio is about 10:1 (14% of the events have snow ratios between 9:1 and 11:1), suggesting some weak indication for the validity of the 10-to-1 rule. [The mode snow density is between 50 and  $60 \text{ kg m}^{-3}$  (Fig. 2b).] These 14% are a subset of 41% classified as average (snow ratios between 9:1 and 15:1; snow densities between 67 and  $111 \text{ kg m}^{-3}$ ; see section 3). Another 14% of the events have snow ratios less than 9:1 and are classified as heavy. The remainder of the events, however, have snow ratios greater than average (45%) and are classified as light. The shape of the curve in Fig. 2b and its skewness toward high snow densities are consistent with previously published studies (e.g., LaChapelle 1962, reproduced in Doesken and Judson 1997, p. 15; Super and Holroyd 1997, p. 23; Judson and Doesken 2000).

The shape of the histograms of snow ratio or snow density from individual observing sites can differ substantially, even for locations in close proximity, as

shown by Super and Holroyd (1997) and Judson and Doesken (2000). The variability of the snow ratios from the individual sites in this paper is a subject of continuing research and will not be presented here.

### 3. Snow-density diagnosis

Despite the quality-control procedures described in section 2, it is expected that the computed snow ratios are subject to imprecision, given the difficulties in measuring both the depth and liquid equivalent of snowfalls. Consequently, for the purposes of diagnosis, all snow ratios are categorized according to three specified classes that reflect the distinct density characteristics of snowfalls (Fig. 2): heavy ( $1:1 < \text{ratio} < 9:1$ ), average ( $9:1 \leq \text{ratio} \leq 15:1$ ), and light ( $\text{ratio} > 15:1$ ). The average class is defined to include the mode of the snow-ratio distribution (Fig. 2a), but extending to higher ratios as suggested by Super and Holroyd (1997), while the heavy and light classes extend outward from either side of that class. Diagnosis of snow ratio then becomes a classification problem and the requirement is to define the boundaries in the input space (composed of the set of snow-ratio predictors) that define the separation between snow-ratio classes. In particular, a discriminant function is sought that evaluates every position in input space and produces a high value for the correct class and low values for the two others.

#### a. Artificial neural networks

ANNs are widely used in classification problems (e.g., Bishop 1996; Principe et al. 2000). A general review of applications of ANNs to problems in meteorology and oceanography can be found in Hsieh and Tang (1998). Further information about ANNs can be found in Marzban and Witt (2001), online (<ftp://ftp.sas.com/pub/neural/FAQ.html>), and in the appendix of this paper. An ANN can be defined as a network of many simple processors (processing elements), joined by communication channels (connections) that carry numeric data. The processing elements operate on the inputs they receive via the connections. The architecture of the ANN is defined by the connectivity. ANNs typically follow a training rule by which the weights of the connections are adjusted on the basis of examples provided by training data. In this way, ANNs are said to “learn” and can exhibit some capability for generalization beyond the training data (in other words, the network can produce reliable results for cases not represented exactly in the training data). It is important to recognize, however, that there are no methods for training ANNs that can create relationships not contained within the training data.

In this work, training of the ANNs is accomplished using supervised learning. The inputs (i.e., predictors for snow ratio) and the desired snow-ratio classes are given to the ANN during training so that the network

TABLE 1. Sigma ( $\sigma$ ) coordinate system used for sounding analysis:  $p$  is pressure and  $p_{\text{surf}}$  is surface pressure.

Sigma level	$\sigma = \frac{p}{p_{\text{surf}}}$
1	1.000
2	0.975
3	0.950
4	0.925
5	0.900
6	0.875
7	0.850
8	0.800
9	0.750
10	0.700
11	0.650
12	0.600
13	0.500
14	0.400

can adjust its connections to match the predicted and desired classes. This phase is distinct from testing, in which the trained network is only given the inputs, not the desired classes. Here, the term “population” is used to describe the set of all cases for which the system is designed, and “sample” (a subset of the population) for the set of cases that are actually available for training and testing. In order to accomplish training and testing, the sample is divided into a training set (here, 60% of the sample), a cross-validation set (20%), and a test set (20%). The training set is used to provide examples to the network to fit the weights and optimize the network snow-ratio class (accomplished by minimization of a cost function, in this case, the mean square error). Prior to training, data are reserved for a cross-validation early-stopping procedure, which provides a means to prevent overfitting of the network weights. The test set is used to assess independently the performance of the trained network.

The review of ANNs (see above and appendix) suggests the following salient points. First, the inputs should be selected based upon prior knowledge of the physical system. Second, the number of inputs should be minimized to constrain the dimensionality of the problem. Third, the training set should be comprehensive (i.e., representative of the population), such that the full range of scenarios to be classified are presented for training. Fourth, training should be accomplished

using early stopping to limit overfitting. Fifth, a variety of network architectures should be evaluated and combined into an ensemble to improve results. These criteria are all met in the study methodology as outlined below.

*b. Study procedure*

Given the relationship between thermodynamic profiles and snow density outlined in section 1, the classifier inputs were defined from measurements of temperature and relative humidity obtained from standard radiosonde launches. In order to account for on-the-ground compaction processes (Gray and Male 1981, 275–306), surface wind speed and liquid equivalent precipitation were also used as inputs. This procedure ensures that the inputs are based upon physical knowledge. Since surface elevation at the sites depicted in Fig. 1 varies from 1.6 km to near sea level, a sigma coordinate system was adopted to define a consistent set of 14 vertical levels for all soundings (Table 1). Further, to constrain the dimensionality of the problem, principal component analysis was used to extract a set of six factors (Table 2) from the original set of 29 inputs (temperature at all 14 levels, relative humidity at the first 13 levels, surface wind speed, liquid equivalent precipitation).<sup>1</sup> It is known that solar radiation is absorbed to some depth within a snowpack and that the grains become rounded during the melting process, leading to compaction (Gray and Male 1981, p. 326). A monthly index (January = +1, July = -1), meant to represent the effects of solar radiation, was added as a seventh input (Table 2).

The six factors produced by the factor analysis are associated with temperature (F1, F3), relative humidity (F2, F4, F5), and surface compaction (F6). The temperature and moisture factors are further stratified by

<sup>1</sup> The objective of factor analysis is to reduce a set of correlated variables to a smaller number of factors, which are linear combinations of the correlated variables. In this work, the extracted factors are constrained to be orthogonal (zero correlation between the factors). The factor extraction method used principal components analysis (Hotelling 1933). The initial factoring method, used to determine the number of factors, was a combination method, employing the larger of the 75% variance rule (Gorsuch 1983) and root curve analysis (Cattell 1966; Cattell and Jaspers 1967). A Varimax transformation method was employed (Kaiser 1958). See Gorsuch (1983) for more details on factor analysis.

TABLE 2. Set of variables extracted from sounding data and used as inputs into the artificial neural networks. The sigma levels of temperature ( $T$ ) and relative humidity (RH) measurements are indicated in parentheses, such that  $T(1)$  denotes the temperature at sigma level 1. The six orthogonal principal components are denoted F1–F6.

Name	Definition	Physical meaning
Month index	Jan = +1, Jul = -1, 0.33 per month	Solar radiation
F1	$T(1)-T(8)$	Low- to midlevel temperature
F2	$RH(1)-RH(7)$	Low- to midlevel relative humidity
F3	$T(6)-T(14)$	Mid- to upper-level temperature
F4	$RH(10)-RH(13)$	Upper-level relative humidity
F5	$RH(7)-RH(10)$	Midlevel relative humidity
F6	Surface wind, precipitation amount	External compaction

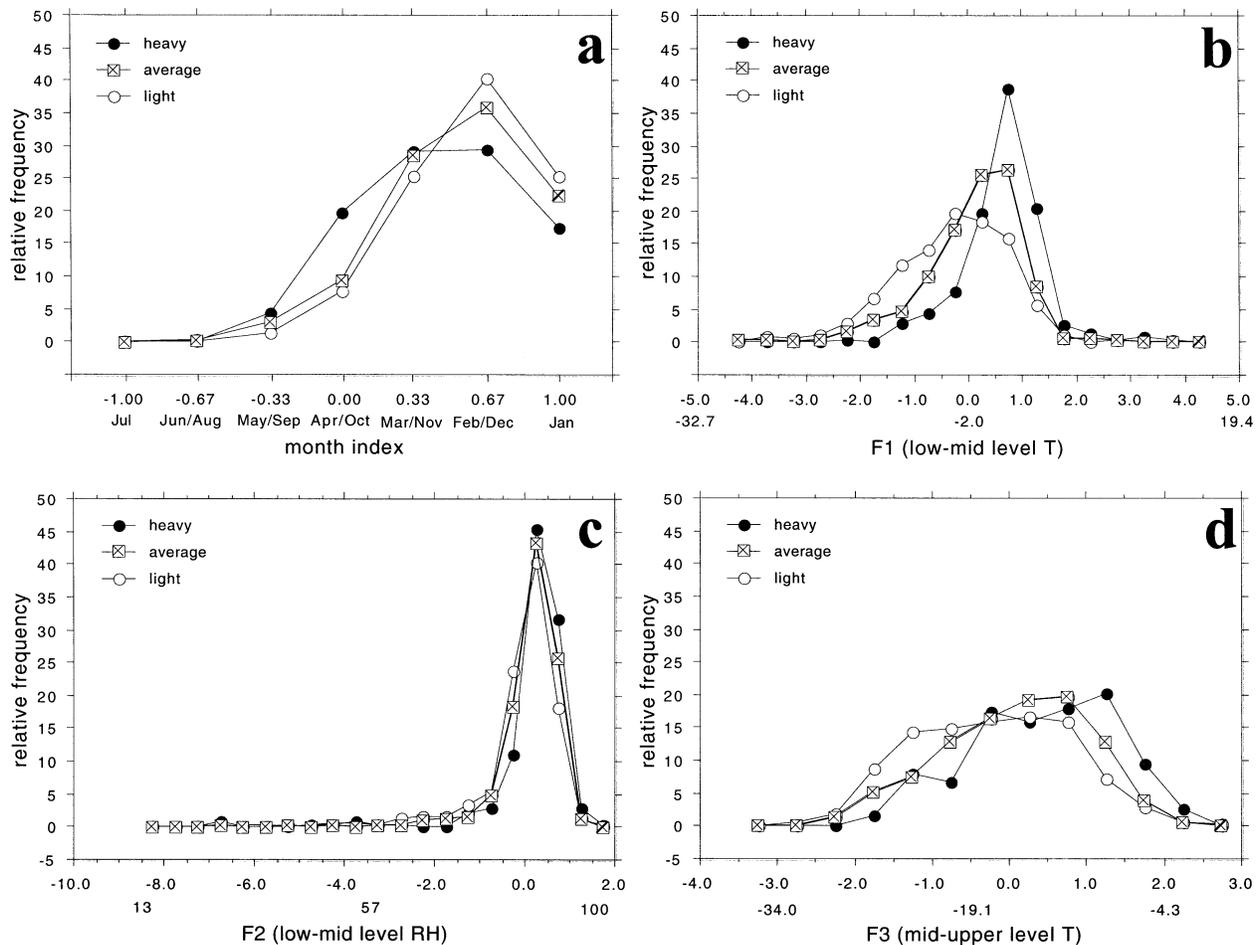


FIG. 3. Relative frequency distributions of the neural network inputs, stratified by observed snow-ratio class (heavy, solid circle; average, x-box; light, open circle). Shown are (a) month index, (b) first principal component (F1, a measure of low- to midlevel temperature), (c) second principal component (F2, a measure of low- to midlevel relative humidity), (d) third principal component (F3, a measure of mid- to upper-level temperature), (e) fourth principal component (F4, a measure of upper-level relative humidity), (f) fifth principal component (F5, a measure of midlevel relative humidity), and (g) sixth principal component (F6, a measure of surface compaction processes). For reference, approximate layer average quantities (as obtained from a linear regression) are also shown along the abscissa (e.g., for F1, the mean temperature for sigma levels 1–8). For F6, the wind ( $\text{m s}^{-1}$ ) and precipitation amounts (mm) are obtained under the null assumption for the other variable. All temperatures are degrees Celsius and relative humidities are percent.

level, representing the upper level (about 700–500 hPa), mid- to upper level (875–400 hPa), midlevel (850–700 hPa), and low to midlevel (below 850 hPa) (note that the exact pressure levels will vary for a given site, depending upon surface elevation; see Table 1).

The seventh input, the monthly index, should not be confused with climatological properties. For example, when the 1650 events are stratified by season (autumn, September–November; winter, December–February; spring, March–May) for sites within the high plains [Great Falls, Montana (GTF); Bismark, North Dakota (BIS); Rapid City, South Dakota (RAP); North Platte, Nebraska (LBF); Denver, Colorado (DEN); Dodge City, Kansas (DDC); and Amarillo, Texas (AMA)], an increased frequency of dense springtime snowfalls is apparent (autumn, 10.0%; winter, 6.4%; spring, 24.3%). This occurrence is associated with increased tempera-

tures and availability of moisture (mean surface temperature and relative humidity for autumn,  $-4.0^{\circ}\text{C}$  and 86.9%; winter:  $-6.9^{\circ}\text{C}$  and 87.4%, spring:  $-1.5^{\circ}\text{C}$  and 87.7%), properties that are directly accounted for by the sounding data. These variables are not associated with the month index, as determined by a low correlation coefficient of 0.30 and 0.04 between the month index and the low-level temperature (F1) and moisture (F2) factors, respectively.

An important physical aspect that is not accounted for by the sounding data, however, is solar radiation. For snowfall, this would most likely be in the form of diffuse solar radiation, which is associated with the time of year. Although explicit information on solar radiation is not available for the sounding sites, the relationship between the month index and solar radiation can be inferred using a data comparison between similar sound-

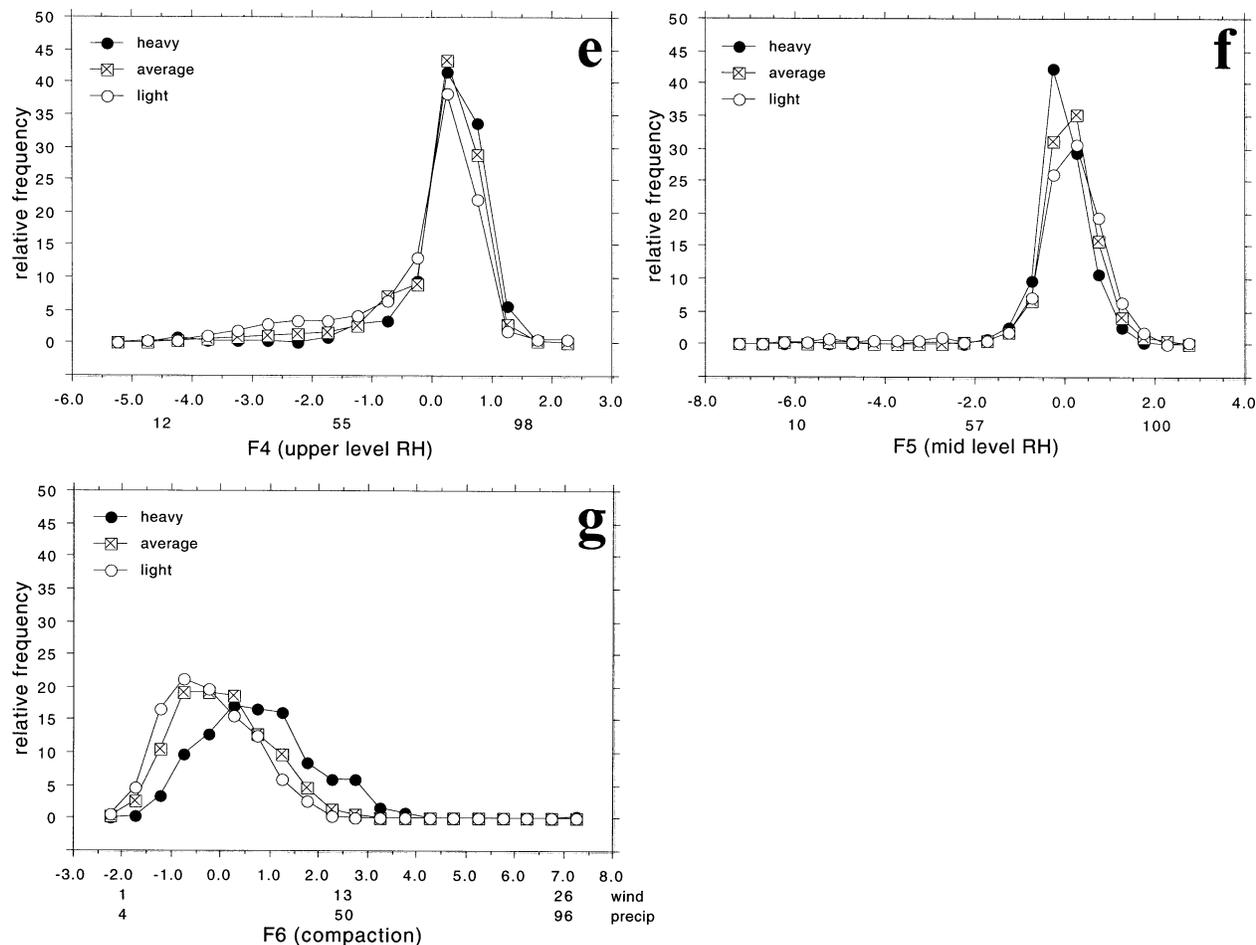


FIG. 3. (Continued)

ings. Three sets of matched sounding and surface data were constructed. Set A consists of 20 pairs of soundings–surface data that are closely matched (defined by the Euclidean distance between the sounding and surface data pairs), with one sounding taken in the autumn and the second in the spring. Set B consists of 64 pairs of closely matched soundings–surface data, in which one sounding was taken in either the autumn or spring and the second sounding was taken in the winter. Set C consists of 1000 randomly selected sounding–surface data pairs, which are used as a background measure of similarity.

Given sufficiently good matching in the sounding pairs and provided that solar radiation is an important factor in determining snow density, then set A should produce snow ratios between the matched pairs that are more similar than those in set B. The rms difference in sounding temperature and relative humidity (sigma-level data) for set A is 42% of that difference for the random pairs, compared to 22% for set B. For surface wind and liquid equivalent precipitation, the rms differences for sets A and B were 39% and 40%, respectively, of that of the random pairs. Hence, the sounding match is

somewhat better for set B. Despite this, the snow-ratio class agreement between sounding pairs is 60% for set A compared to only 42% for set B, reinforcing the importance of solar radiation.

Analysis of the frequency distributions of the inputs, stratified according to snow-ratio class, is a useful method for assessing their first-order diagnostic capability (Fig. 3). Based upon visual inspection, the greatest separations between the classes are revealed by the month index, F1 (low- to midlevel temperature), F3 (mid- to upper-level temperature), and F6 (compaction). However, these results do not necessarily imply a lack of utility of the remaining three factors, since nonlinear interactions between inputs may enhance diagnostic capability. This effect will be discussed in section 4a.

Since these seven inputs all have potential discriminatory capability, they were all used in the neural networks. Two types of network architectures were selected to define the ensemble, which was constructed from 10 individual networks (e.g., Opitz and Maclin 1999; see appendix). The training datasets for each individual member of the ensemble were distinct, having been obtained from the original training set using the bootstrap-

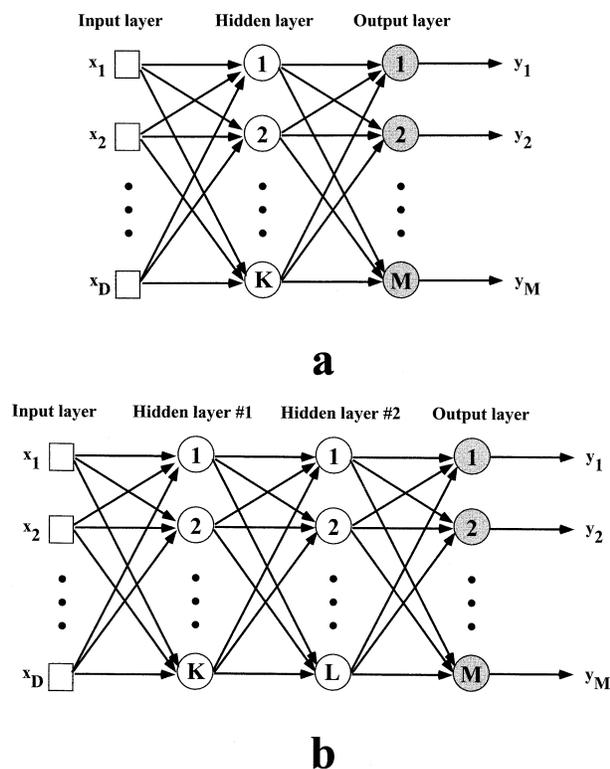


FIG. 4. Schematic diagrams of the MLP neural network architecture: (a) a single-hidden-layer MLP with  $D$  inputs,  $K$  hidden layer processing elements and  $M$  outputs and (b) a two-hidden-layer MLP with  $D$  inputs,  $K$  processing elements in hidden layer 1,  $L$  processing elements in hidden layer 2 and  $M$  outputs.

ping technique (Efron and Tibshirani 1993). This approach has been shown to be effective in producing unique networks (i.e., variable weights), which allows decorrelation of the errors of the individual networks and, hence, improved ensemble performance (Breiman 1996; Opitz and Maclin 1999).

The first network type, composing five members of the ensemble, is the multilayer perceptron (MLP) with a single hidden layer (Fig. 4a). MLP networks have been used extensively in classification problems (e.g., Bishop 1996; Principe et al. 2000). Each of the single-hidden-layer MLPs featured seven inputs, 40 processing elements in the hidden layer, and three outputs (the three snow-ratio classes). These architectural details were established through trial and error (see appendix), maximizing percent correct from the training dataset.

The second network type, also composing five members of the ensemble, is the MLP with two hidden layers (Fig. 4b). The two-hidden-layer MLPs, whose architectural details were also established through trial and error, featured seven inputs, seven processing elements in the first hidden layer, four processing elements in the second hidden layer, and three outputs. For all networks, activations were accomplished using the hyperbolic tangent function everywhere except in the output layer, where the softmax activation function

was used (so that the output values would lie between 0 and 1 and sum to 1).

When training neural networks, it is sometimes necessary to consider the frequency of occurrence of different classes in the training set. For the 1650 cases composing the entire 22-yr dataset, 14% are heavy, 41% are average, and 45% are light. For such unbalanced datasets where one class is substantially underrepresented compared to the other classes, the neural networks may simply treat that class as noise. Results from preliminary experiments with the networks for the snow-ratio problem exhibited this effect. Accordingly, a partial balancing criteria (270 heavy, 360 average, and 360 light cases in each training set) was applied to the bootstrap sampling, to improve training results across all three classes. As a result, the network outputs are not interpretable as posterior probabilities (i.e., a probability that takes into account the frequency of occurrence of a given class), without application of a renormalization procedure.

### c. Results

Following the training of the 10 individual networks, each of the networks was run on the independent test dataset. The individual network outputs were then combined into an ensemble using simple averaging, where the snow-ratio class with the highest ensemble value was compared with the observed class. The probability of detection (POD), false alarm ratio (FAR), bias, and critical success index (CSI) for each snow-ratio class were used as measures of performance. The POD is the fraction of observed events correctly forecast (range of 0–1, optimal performance is 1), the FAR is the fraction of event forecasts that do not verify (range of 0–1, optimal performance is 0), the bias is the ratio of forecast to observed events (<1 indicates underforecasts, >1 indicates overforecasts), and the CSI is a measure of forecast accuracy (range of 0–1, optimal performance is 1), defined by the ratio of successful event forecasts to the total number of event forecasts that were made or needed (Gilbert 1884; Schaefer 1990). Geometrically, the CSI can be thought of as the intersection of the forecast and observed events divided by the union of the forecast and observed events. As noted by Schaefer (1990), the CSI can be expressed in terms of POD and FAR as

$$\text{CSI} = [(\text{POD})^{-1} + (1 - \text{FAR})^{-1} - 1]^{-1}, \quad (1)$$

where  $\text{POD} = 0.00$  and/or  $\text{FAR} = 1.00$  yields a CSI of 0.00. Although the CSI is a biased score, since it depends upon the frequency of the event that is forecast (Schaefer 1990), it is useful in providing an indication of the relative worth of different forecast techniques applied to the same data. Two additional verification measures, applicable to the overall forecast results rather than specific snow-ratio classes, are the percent correct (PC; range of 0%–100%, optimal performance is 100%)

TABLE 3. Diagnosis results for sample of 333 independent test cases. Shown are (a) percent correct (PC) and Heidke skill score (HSS) performance measures, along with raw contingency table results and (b) specific class verification measures: probability of detection (POD), false-alarm ratio (FAR), bias, and critical success index (CSI). For comparison, all results are also shown for climatological forecasts, based upon the observed frequencies from the test sample (in parentheses) and forecasts based on the NWS Table 4-9 for surface temperatures (in brackets; note that the NWS Table 4-9 does not predict heavy snow ratios).

(a) Raw totals: PC, 60.4 (41.7) [51.7];  
HSS, 0.34 (0.00) [0.12]

		Predicted No. of cases		
		Heavy	Average	Light
Observed	Heavy	20 (3) [0]	12 (15) [20]	1 (15) [13]
No. of	Average	23 (15) [0]	86 (67.5) [70]	41 (67.5) [80]
cases	Light	6 (15) [0]	49 (67.5) [48]	95 (67.5) [102]

(b) Specific class verification measures

Score	Heavy	Average	Light
POD	0.61 (0.10) [0.00]	0.57 (0.45) [0.47]	0.63 (0.45) [0.68]
FAR	0.59 (0.90) [—]	0.42 (0.55) [0.49]	0.31 (0.55) [0.48]
Bias	1.49 (1.00) [—]	0.98 (1.00) [0.92]	0.91 (1.00) [1.30]
CSI	0.32 (0.05) [0.00]	0.41 (0.29) [0.32]	0.50 (0.29) [0.42]

and the Heidke skill score [HSS; Panofsky and Brier (1958); range of  $-1$  to  $+1$ , optimal skill is 1]. The PC, which is the ratio of all correct forecasts to the total number of forecasts, can be hedged by forecasting common events more frequently. The HSS accounts for this bias by comparing the PC to that obtained for no-skill random forecasts.

As expected, the ensemble approach resulted in improvement relative to the individual networks. Using the CSI as a measure of effectiveness, the ensemble outperformed the individual networks in 77% of the 30 possible situations [10 networks composing the ensemble times three CSI per network (one for each snow-ratio class)], with no individual network exceeding the

ensemble in all three snow-ratio classes. Overall, the ANN diagnosis achieves 60.4% correct and an HSS of 0.34 (Table 3a). The CSI shows that the highest forecast quality is generally associated with the light snow-ratio class, with a monotonic dropoff through the average to the heavy class, largely owing to increased false alarms (Table 3b). The results suggest, however, that the ANN method provides high-quality diagnoses for all snow-ratio classes.

The ANN diagnosis scores are compared with those obtained from two other methods (Table 3). The first method is a “climatology” derived from the observed frequencies of the snow-ratio classes in the test dataset [e.g., the observed frequency of the heavy class is 0.10,

TABLE 4. Diagnosis results for independent test cases, following data-flagging procedure (see text for details). Shown are (a) PC and HSS performance measures, along with raw contingency table results and (b) specific class verification measures: POD, FAR, bias, and CSI. For comparison, all results are also shown for climatological forecasts, based upon the observed frequencies from the test sample (in parentheses) and forecasts based on the NWS Table 4-9 for surface temperatures (in brackets; note that the NWS Table 4-9 does not predict heavy snow ratios).

(a) Raw totals: PC, 64.7 (41.9) [56.4];  
HSS, 0.42 (0.00) [0.24]

		Predicted No. of cases		
		Heavy	Average	Light
Observed	Heavy	20 (3) [0]	12 (17) [20]	0 (12) [12]
No. of	Average	23 (17) [0]	86 (78) [70]	41 (55) [80]
cases	Light	0 (12) [0]	26 (55) [14]	81 (40) [93]

(b) Specific class verification measures

Score	Heavy	Average	Light
POD	0.63 (0.09) [0.00]	0.57 (0.52) [0.47]	0.76 (0.37) [0.87]
FAR	0.54 (0.91) [—]	0.31 (0.48) [0.33]	0.34 (0.63) [0.50]
Bias	1.34 (1.00) [—]	0.83 (1.00) [0.69]	1.14 (1.00) [1.73]
CSI	0.36 (0.05) [0.00]	0.46 (0.35) [0.38]	0.55 (0.23) [0.47]

and the observed frequency of the average class is 0.45, so that the expected number of diagnosed heavy but observed average events is  $(0.10)(0.45)(333) = 15$ . The second method obtains the diagnoses from the NWS “new snowfall to estimated meltwater conversion” table [U.S. Department of Commerce (1996); see also section 4d of this paper; hereafter, NWS Table 4-9]. All of the class-specific performance measures (save bias) are substantially higher for the ANN ensemble. The NWS Table 4-9, which is based upon surface temperature alone, falls short of the ANN ensemble accuracy (as measured by CSI) for all snow-ratio classes. Overall, the ANN technique is the most skillful, with an HSS of 0.34, compared to 0.12 for NWS Table 4-9 and 0.00 for the climatological estimate. In the 1650-case sample, the correlation between F1 and surface temperature is 0.94; hence, NWS Table 4-9 is, in this respect, the functional equivalent of providing the ANN with F1 and withholding the six other inputs. A further exploration of the importance of the full set of inputs is provided in section 4.

#### d. Diagnosis failures

While the ANN results establish that the technique can provide improved information relative to existing approaches, it is of interest to understand the circumstances governing the failed diagnoses (hereafter, “hits” and “misses” for correct and incorrect ANN classifications, respectively). Analyses of the inputs (Table 2), using the Wald–Wolfowitz nonparametric test (Siegel 1956), show statistically significant differences in the distributions of the temperature inputs (F1 and F3;  $p$  values of 0.02 and  $<0.01$ , respectively) between the hits and misses for the light snow-ratio class. In particular, mean temperatures between sigma levels 1–8 are  $-4.1^{\circ}\text{C}$  for the misses, compared to  $-10.6^{\circ}\text{C}$  for the hits. Similarly, mean temperatures between sigma levels 6–14 are  $-13.0^{\circ}\text{C}$  for the misses, compared to  $-21.4^{\circ}\text{C}$  for the hits. Since sigma levels 1–8 extend over a surface-based layer to approximately 800 hPa, the relatively warm temperatures for the missed light snow-ratio cases are surprising, and warrant further study.

Since the snow data are collected within 6 h of a radiosonde launch, it is possible that some cases may have been adversely affected by rapidly changing weather conditions (consider, for example, the case where snow occurs along the cool side of a warm frontal boundary prior to a radiosonde launch; the warm front progresses northward, such that by the time the sounding is taken, the sampling occurs within the warm air). In the absence of temporally detailed vertical profiles, a first-order measure of such transitions was estimated using the variance of surface temperature and wind speed, calculated over the 12-h period centered on the radiosonde launch time (steady conditions would yield a small variance while a transition would produce a large variance). The analyses (based upon the two-sample  $t$

test with a 0.05 significance level) revealed no statistically significant difference in the variance between the hits and misses for the light snow-ratio class for either surface temperature or wind speed. Hence, we conclude that temporal data limitations are not substantially affecting the results.

A second uncontrolled error source is precipitation gauge undercatch. This effect, which can produce liquid equivalents that are 40%–70% less than snow collected and melted from snowboards (Peck 1972; Larson and Peck 1974; Goodison 1978; Gray and Male 1981, p. 204; Groisman and Legates 1994), will introduce a bias toward larger snow ratios. While it is not possible to directly assess the degree of undercatch from historical data, it is possible to screen the data for cases that seem questionable, based on existing data, known biases, and present physical understanding. The procedure employed was as follows. All observed light snow-ratio cases in the test sample (150 out of 333 events) were flagged if either of the temperature inputs (F1, F3) exceeded the mean of the light snow-ratio cases by more than one standard deviation. This procedure was motivated by the distribution test results noted above. An additional case, for which the observed snow-ratio class was heavy, but for which the temperature and compaction variables were inconsistent [the layer average low- to midlevel (mid- to upper-level) temperatures were 2.8 (1.8) standard deviations colder than the heavy class mean, while the precipitation was 0.9 standard deviations below the heavy class mean with a surface wind speed of  $5\text{ m s}^{-1}$ ], was also flagged. A total of 44 cases were flagged (representing 13.2% of the test dataset), including 14 hits.

Although we cannot conclude that these 44 cases are contaminated by measurement error, this is a plausible interpretation. Accordingly, we present the snow-ratio class diagnosis results, with these cases removed (Table 4), as a measure of the potential of this technique with cleaner data (note that this procedure does not screen for all questionable cases; see below). Screening the test data in this manner results in improvements in the HSS for both the ANN and NWS Table 4-9, with the largest gains in the average and light snow-ratio classes (as measured by CSI). The HSS suggests that the ANN technique can provide a 75% improvement in skill relative to the use of NWS Table 4-9.

Although the results of Table 4 are encouraging, 35.3% of the cases are still classified incorrectly. In real-world classification problems, ambiguity in the data is often unavoidable. This ambiguity arises from two sources: noisy data, as partially addressed above, and incomplete representation of the feature space. In the latter case, key inputs that are needed to precisely define classification boundaries are absent. As a result, the boundaries are represented in a lower-order dimension, and the classes overlap within the feature space. In section 5, we will provide some discussion of additional

TABLE 5. Cluster analysis of the 1650 snow cases, using a Kohonen self-organizing (feature) map (SOM) and the input factors F1, F3, and F6, representing temperature and surface compaction effects. Shown are cluster identifier (C1–C5), number of cases in the cluster ( $N$ ), layer average temperature from sigma levels 1–8 ( $T_{F1}$ , °C), layer average temperature from sigma levels 6–14 ( $T_{F3}$ , °C), surface wind speed (wind,  $\text{m s}^{-1}$ ), liquid equivalent precipitation amount (PP, mm), and percentage of cluster cases in the snow ratio classes (heavy, average, and light).

Cluster	$N$	$T_{F1}$	$T_{F3}$	Wind	PP	Heavy	Average	Light
C1	274	−14.0	−19.8	4.2	5.5	2.9	21.9	75.2
C2	336	−9.2	−17.9	5.8	7.3	4.2	44.9	50.9
C3	290	−4.0	−14.3	6.6	14.0	37.9	42.1	20.0
C4	319	−2.7	−13.6	4.4	8.7	21.3	49.5	29.2
C5	431	−4.7	−14.6	2.6	6.2	7.9	43.9	48.3

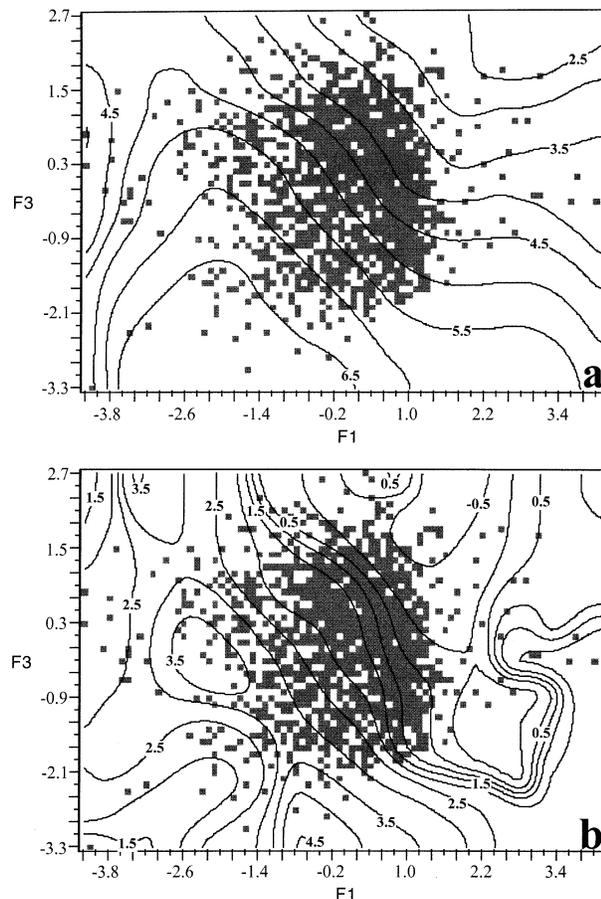


FIG. 5. The snow-ratio class for each of the 1650 events was projected in a three-dimensional volume with the axes represented by F1, F3, and F6. Snow-ratio class interfaces were then extracted from this volume and projected on the two-dimensional plane described by F1 and F3. Shown are the interfaces between (a) the heavy and average class and (b) the average and light class. Contour lines are the position of the interface as measured by F6 (contour interval is 0.5 units of F6). Locations with observed data are denoted by the gray squares (analyses in unsampled regions were generated using simple interpolation).

inputs that might lead to improved snow-ratio class separation.

As a means of understanding the degree to which this classification boundary issue affects the performance of the ANN, we have undertaken a cluster analysis of the input data, using all 1650 cases. Cluster analysis is concerned with organizing data into meaningful structures. This is distinct from the classification problem, since a single cluster defined by the inputs may overlap a classification boundary and multiple clusters may fall within a single class. A neural network technique known as the Kohonen self-organizing map (SOM) forms one basis for cluster analysis. The details as to how this network performs the cluster analysis are beyond the scope of this paper; interested readers should consult Kohonen (1995).

Experiments with the full set of inputs and other input combinations showed that the most optimal clusters (defined by distinct input values across the clusters and the relative separation of snow ratio classes; Table 5) were generated using three inputs (F1, F3, and F6) to an SOM. The results of the cluster analysis demonstrate the boundary issue, with only C1 showing an obvious separation into a specific snow-ratio class (light). Of particular note is the high percentage of average snow-ratio cases in each of the remaining four clusters, suggesting the existence of a complex boundary between the average and extreme classes. Visual inspection of the dataset using a three-dimensional projection, with the three axes represented by F1, F3, and F6, respectively, confirms this supposition (Fig. 5). Undulations, steep gradients, and local maxima and minima in the interfaces between the average and extreme classes are evident. Although the four remaining inputs (month index, F2, F4, and F5) provide additional discriminatory power, advances in diagnostic capability beyond that demonstrated in Tables 3 and 4 will likely require inputs related to physics that are presently neglected. This issue is discussed in section 5.

## 4. Discussion

### a. Interpretation of network diagnoses

An important but often overlooked aspect of the use of ANNs is the interpretation of the outputs. Since it is

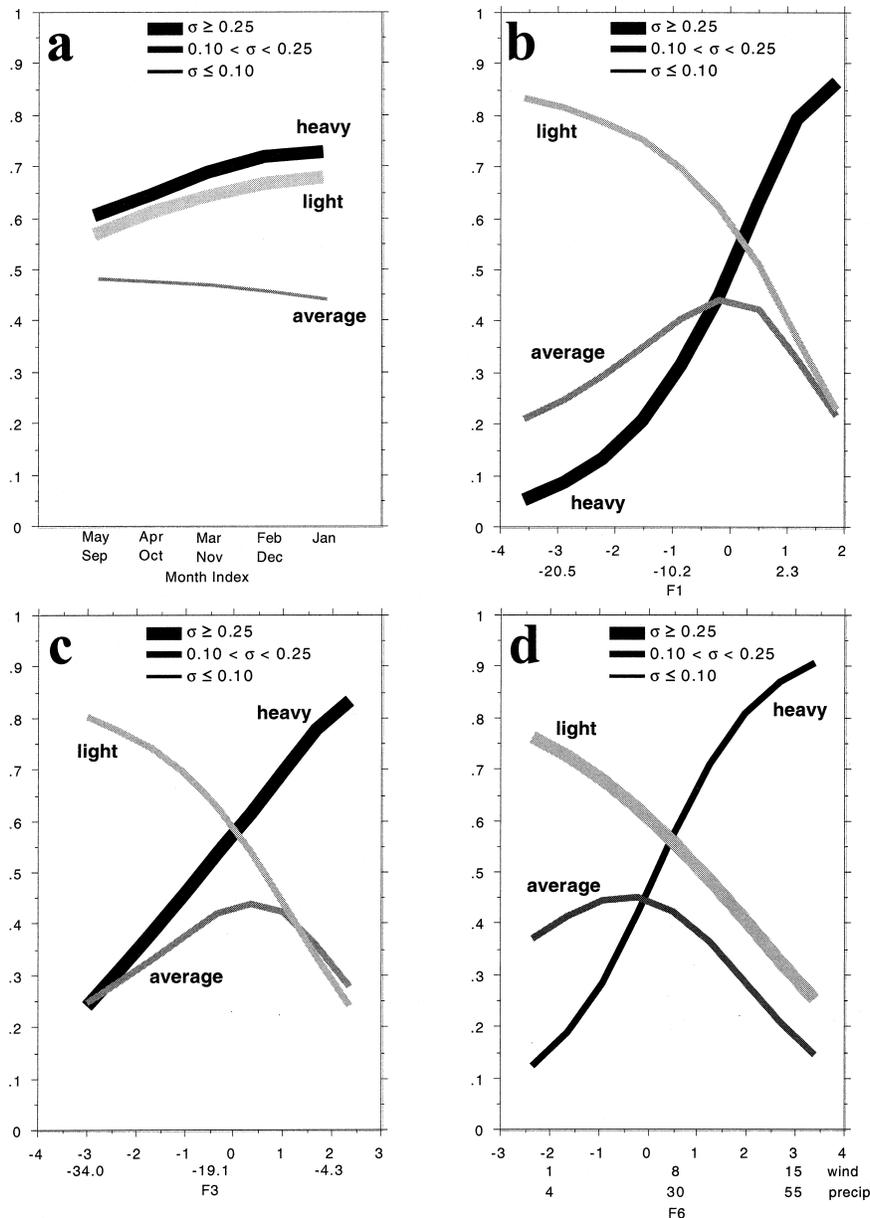


FIG. 6. Importance of inputs (see text for details) for heavy snow-ratio class (black), average snow-ratio class (dark gray), and light snow-ratio class (light gray). The thickness of the line indicates the variance of the response to the input, where  $\pm 1$  standard deviation  $\geq 0.25$  is a thick line, between 0.10 and 0.25 is a medium thickness line, and  $\leq 0.10$  is a thin line. Shown are (a) month index, (b) F1, (c) F3, and (d) F6. For reference, approximate layer average quantities (as obtained from a linear regression) are also shown along the abscissa (e.g., for F1, the mean temperature for sigma levels 1–8). For F6, the wind ( $\text{m s}^{-1}$ ) and precipitation amounts (mm) are obtained under the null assumption for the other variable. All temperatures are degrees Celsius.

not necessary to have prior knowledge of the precise relationship between the inputs and the desired outputs in an ANN, a trained network that generalizes well may yield new knowledge concerning these relationships.

The importance of inputs could be measured by several methods, including evaluation of the network weights (e.g., size, sums of products, or more elaborate

functions), or measuring the gradient of the output with respect to the input (i.e., a vector of partial derivatives that measures the local rate of change of the output with respect to the corresponding input, holding the other inputs fixed). Each of these approaches suffers from deficiencies that may lead to incorrect conclusions concerning importance. A thorough discussion of these is-

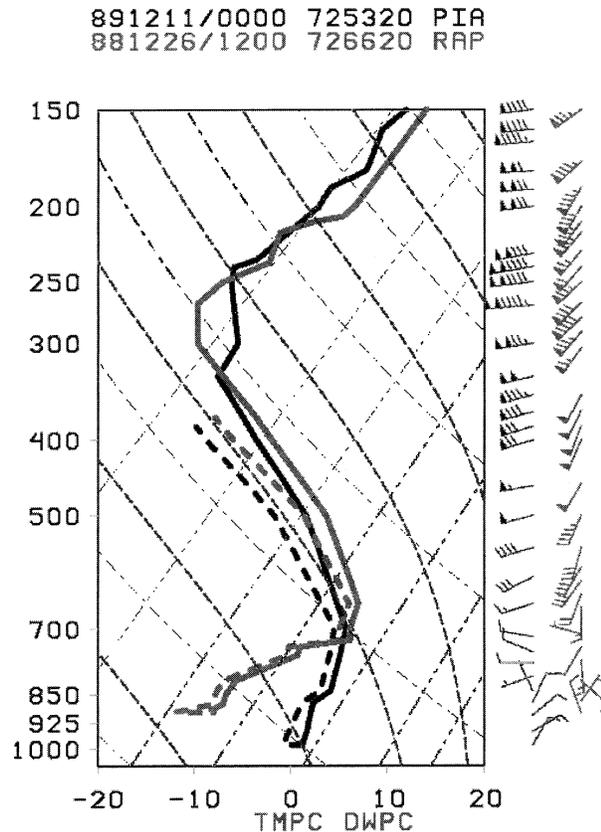


FIG. 7. Thermodynamic diagrams for 0000 UTC 11 Dec 1989 at PIA (black), and 1200 UTC 26 Dec 1988 at RAP (gray). The solid line is temperature and the dashed line is dewpoint temperature. Winds are plotted according to the standard meteorological convention.

sues can be found online (<ftp://ftp.sas.com/pub/neural/importance.html>).

A more robust method is to compute differences rather than derivatives, an approach that adheres to the strict definition of causal importance. This method is applied here by varying each of the seven inputs (Table 2) in turn across the range of values existing in the 201 correctly classified events of the test dataset (Table 3) while holding the other inputs to their observed values. The average value of an output snow-ratio class across all the cases (which can range from zero to one) is plotted as a function of the input value, as a means of defining the central tendency of the input importance. Since there may be substantial variations in the response under certain circumstances (consider, for example, cases in which no changes occur until a critical threshold is passed, such as the familiar example of precipitation), the standard deviation in the response across all cases as a function of the input value is also computed.

The results of this analysis are shown in Fig. 6 for the four most critical inputs (month index, F1, F3, and F6). The march of the seasons has the most substantial effect on the heavy and light snow-ratio classes, with

an increasing likelihood of either class in midwinter, albeit subject to considerable variation in the response (Fig. 6a). Increasing low- to midlevel temperatures have the effect of decreasing (increasing) the possibility of light (heavy) snow ratios, whereas the response of the heavy class is the most variable (Fig. 6b). Increasing temperatures aloft yields a similar response (Fig. 6c). Increasing winds and/or larger liquid equivalent precipitation amounts, both processes tending to produce greater surface compaction, also leads to decreasing (increasing) the likelihood of light (heavy) snow ratios, although in this case the light class response exhibits the greatest variability (Fig. 6d).

It is important to note that Fig. 6 shows the ANN response to an input rather than a prediction based upon that input, since a specific prediction will depend in a complex fashion on all of the inputs. The large variation in response ( $\geq 0.25$ ) for the heavy and light classes to month index (Fig. 6a), for the heavy class to low- to midlevel temperatures (Fig. 6b) and to mid- to upper-level temperatures (Fig. 6c), and for the light class to surface compaction (Fig. 6d), is an important indicator of the complexity of the problem. A specific comparison of two instances with a low- to midlevel layer average temperature of  $-11^{\circ}\text{C}$  demonstrates this complexity. In the first (second) example, the profile of the remaining inputs yields the following representative conditions: month April (March), mid- to upper-level layer average temperature  $\sim -20^{\circ}\text{C}$  ( $-7^{\circ}\text{C}$ ), low- to midlevel layer average relative humidity  $\sim 93\%$  (88%), midlevel layer average relative humidity  $\sim 91\%$  (90%), upper-level layer average relative humidity  $\sim 80\%$  (91%), and precipitation  $\sim 15$  mm (38 mm) with no wind. The primary difference between the two examples is in the mid- to upper-level layer average temperature and precipitation. Both of these factors act to produce higher densities in the second example, which results in a network probability of 0.77 for the heavy class. In contrast, the first example exhibits a network probability of 0.10 for that same class. Hence, although the general effect of an input can be isolated as depicted in Fig. 5, *the actual prediction in a specific case can be strongly modified by the conditions represented by the other inputs.*

Finally, the importance of the profiles of relative humidity (as represented by the principal component factors F2, F4, and F5) in the diagnosis of snow-ratio class was further examined. Experiments in which the networks were retrained while withholding the relative humidity inputs indicated an overall network performance loss of at least 5% (based upon percent correct), suggesting that these inputs are helpful, if nonessential, diagnostic quantities. A demonstration of this effect is provided by a pair of examples, in which all of the inputs except the low- to midlevel layer average relative humidities, are identical. In the first (second) example, these humidities are  $\sim 40\%$  (100%). The remaining conditions are the following: month (March), low- to midlevel layer average temperature  $\sim -8^{\circ}\text{C}$ , mid- to upper-

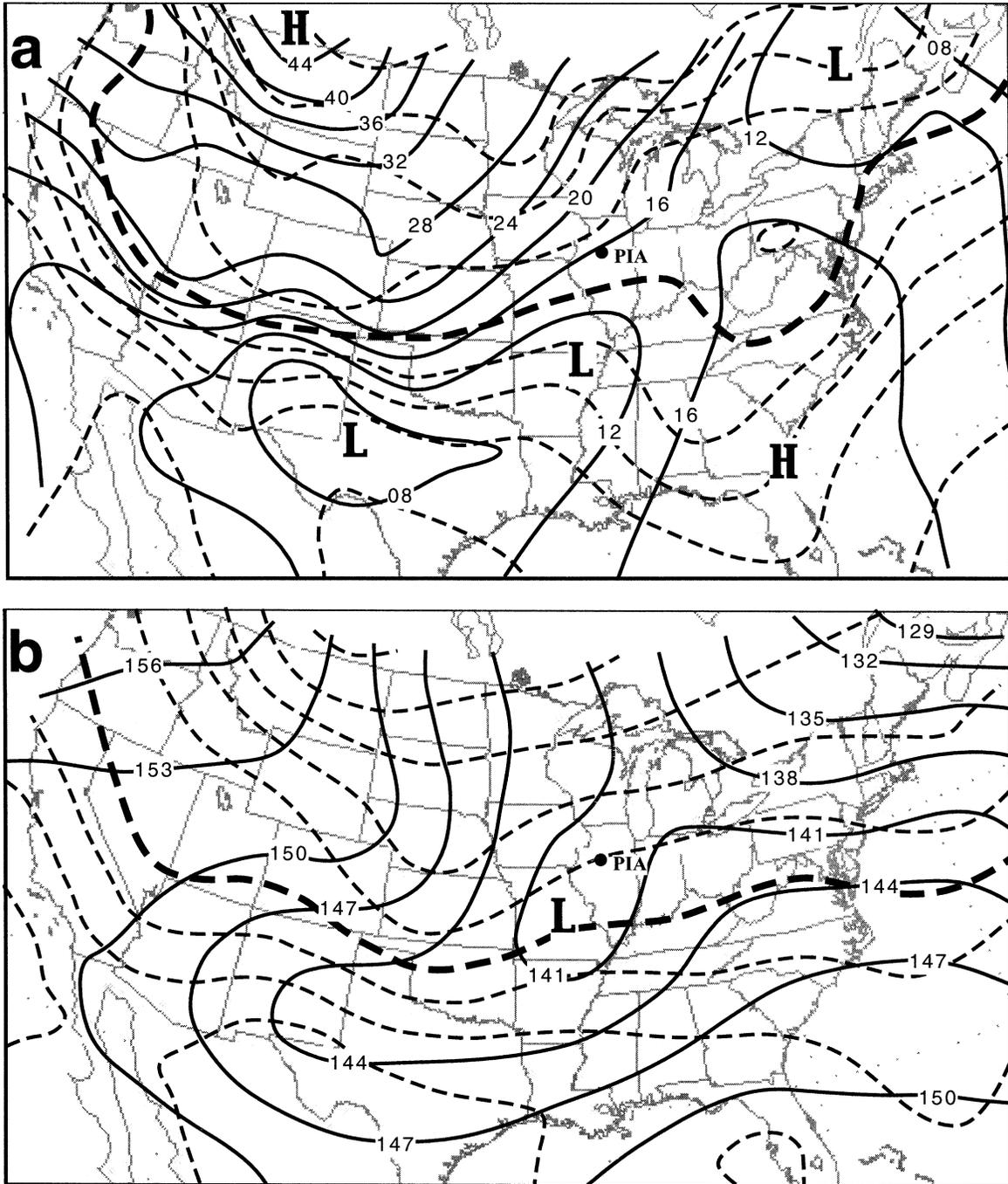


FIG. 8. (a) Surface analysis at 0000 UTC 11 Dec 1989, (b) 850-hPa analysis at 0000 UTC 11 Dec 1989, (c) surface analysis for 1200 UTC 26 Dec 1988, and (d) 850-hPa analysis at 1200 UTC 26 Dec 1988. The location of the case site is indicated by the black dot [(a), (b) PIA, (c), (d) RAP]. For surface analyses, solid (dashed) lines are isobars (isotherms), with contour interval of 4 hPa (5°C, thick line is 0°C). For 850-hPa analyses, solid (dashed) lines are geopotential height (isotherms), with contour interval of 3 dam (5°C, thick line is 0°C).

level layer average temperature  $\sim -5^{\circ}\text{C}$ , midlevel layer average relative humidity  $\sim 92\%$ , upper-level layer average relative humidity  $\sim 94\%$ , and precipitation  $\sim 24$  mm with no wind. For the first (second) example, the network probability for the heavy class is 0.60 (0.82).

In both examples, the relative warmth of the sounding and higher precipitation amounts skew the network toward the heavy class. In the second example, the elevated low- to midlevel layer average relative humidity indicates the availability of supercooled liquid water and

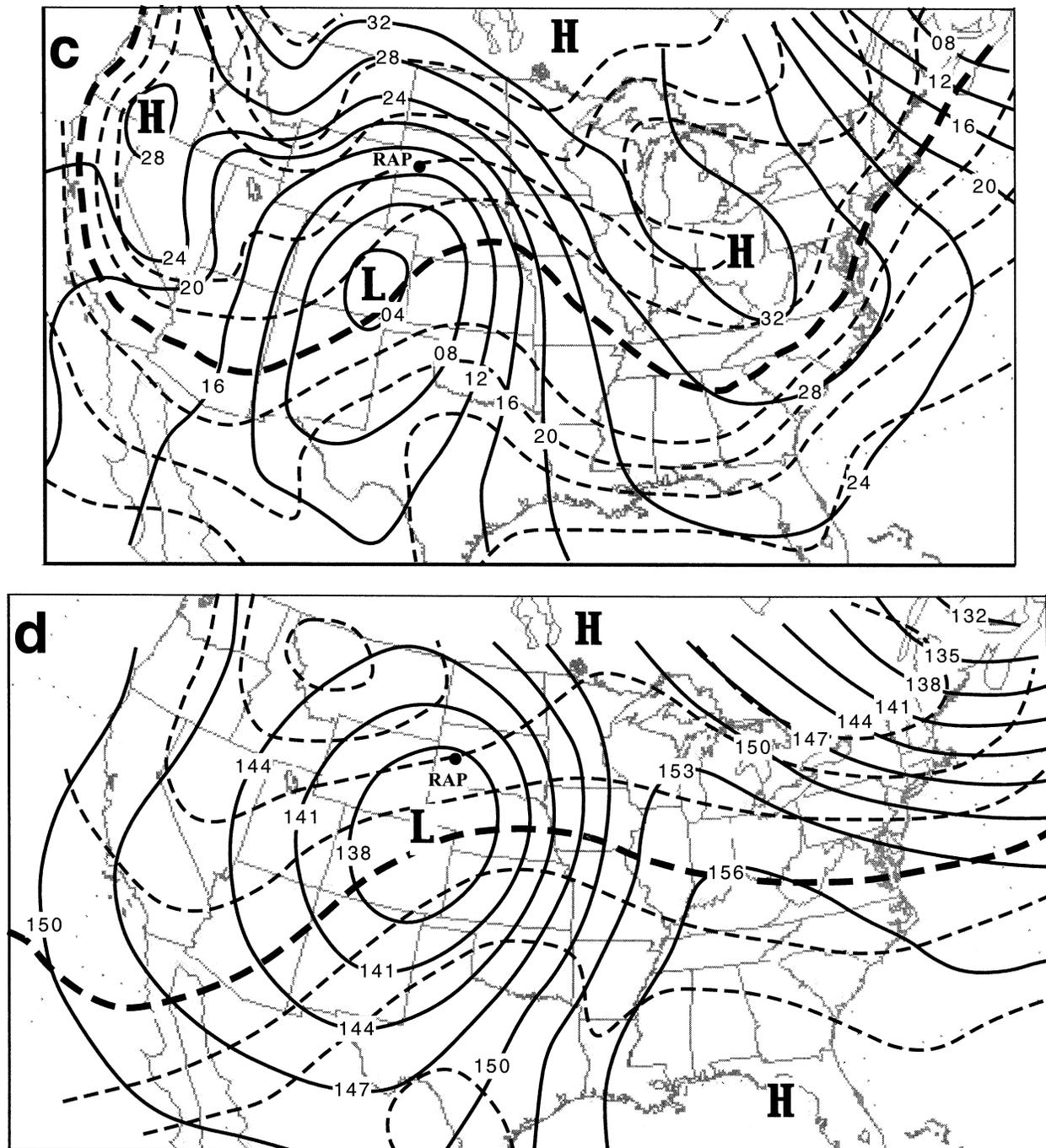


FIG. 8. (Continued)

the likelihood of riming, shifting the network prediction further toward the heavy class.

*b. Case examples*

In this section, a series of cases are presented that demonstrate the difficulties inherent in the diagnosis of

snowfall density. Of particular note is the degree to which case-to-case variations in snowfall density are dependent upon subtle features, not readily apparent from a synoptic perspective. A search was conducted for cases in which all but one of the seven input variables showed considerable agreement (as measured by standardizing the inputs to mean 0 and standard deviation

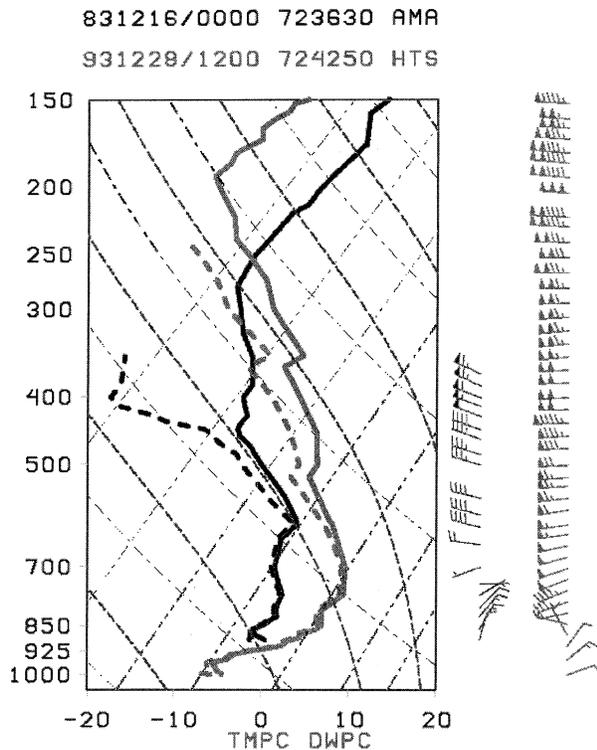


FIG. 9. Thermodynamic diagrams at 0000 UTC 16 Dec 1983 at AMA (black), and 1200 UTC 28 Dec 1993 at HTS (gray). The solid line is temperature and the dashed line is dewpoint temperature. Winds are plotted according to the standard meteorological convention.

1, and computing the Euclidean distances for all the inputs relative to the comparison case). Hence, the comparison cases will not necessarily be from the same site or date. This process was repeated for the two temperature inputs (F1 and F3) and the compaction input (F6).

Soundings from this analysis for low- to midlevel temperatures (F1) (Fig. 7) document the similarity in thermodynamic structure between RAP, and PIA, except for the strong surface-based inversion at RAP, consistent with that site being located on the cold side of a frontal boundary (Fig. 8c). For this event at RAP, the observed snow-ratio class was light (ratio of 18:1). PIA, with warmer low- to midlevel temperatures in association with a site location shifted farther toward the warm side of the low-level thermal gradient (Fig. 8a) and within the warm-core circulation at 850 hPa (Figs. 8b,d), developed heavy snow-ratio conditions (ratio of 7:1).

The comparison soundings for mid- to upper-level temperatures (F3) (Fig. 9) document the relative warmth at those levels at HTS, which led to a heavy snow ratio (8:1), while an average snow ratio (12:1) developed at AMA (note that this comparison demonstrates the need to normalize the raw sounding data to the sigma coordinate system described in section 3b). At the surface, both sites were positioned to the northeast of a cyclone and associated frontal boundary, with arctic air to the

north and west (Figs. 10a,c). However, the 500-hPa flow was nearly zonal across eastern North America in the HTS case (562-dam, 500-hPa height; Fig. 10d), whereas a trough was positioned across the southern and central United States in the AMA case (548-dam, 500-hPa height; Fig. 10b). In both of these cases, a diagnosis based upon surface temperatures alone (e.g., NWS Table 4-9) would have indicated light snow ratios (20:1).

Finally, the thermodynamic profiles at HON, and INL (Fig. 11), exhibit considerable similarity, consistent with the primary differences resulting from surface compaction processes, with stronger winds (8 versus  $5 \text{ m s}^{-1}$ ) and higher precipitation amounts (13.0 versus 2.8 mm) at HON (observed average snow ratio of 10:1) than at INL (observed light snow ratio of 22:1). Interestingly, low-level temperatures at both of these sites are fairly warm, which affirms the complexity in some instances in correctly diagnosing snow ratios. Examination of the synoptic charts reveals that both sites were positioned between a downstream anticyclone and an upstream cyclone (Figs. 12a,b), with HON located most directly within the pressure gradient ( $\sim 5 \text{ m s}^{-1}$  stronger geostrophic southerlies).

Each of these three cases suggest that subtle synoptic scale differences can be associated with a change in snow-ratio class. However, the results presented in Tables 3 and 4 indicate that considerable improvement beyond current operational practice can be obtained by more complete consideration of the available data.

### c. Deficiencies of NWS Table 4-9

The results obtained by application of NWS Table 4-9 to the test data (Tables 3 and 4) make clear that serious damage to the climatological record will ensue in the absence of careful direct measurements of snow depth and liquid equivalent. Unfortunately, current NWS practice is not promising in this regard. The U.S. Department of Commerce (1996, section 4.3.3.5.2) states, "Report the water equivalent of snow on the ground to the larger Weather Service Forecast Office; or the Smaller Weather Forecast Office at 1800 UTC if the average snow depth (to the nearest inch) is 2 inches or more. . . . Whenever the water equivalent of snow on the ground cannot be measured by melting or weighing of the core sample, enter an estimated water equivalent on the basis of a 1/10 ratio method unless a different ratio is more appropriate. . . . Table 4-9 (New Snowfall to Meltwater Conversion) can only be used as an observing aid in determining water equivalency of newly fallen snow."

Super and Holroyd (1997, p. 24) shed further light on this problem: "personal communication with Grant Godge of the National Climatic Data Center revealed this table [NWS Table 4-9] was developed as a guide for quality control checking of observations and was not intended to be a substitute for observations of [snow water equivalent]. The table's temperature dependence

of density is not based on actual measurements but rather on general impressions in the eastern Tennessee and western North Carolina areas. Hence, the reality of the apparent temperature dependence is uncertain." As was shown in the previous section, this temperature dependence is, in fact, inadequate [see also Kyle and Wesley (1997) for a specific example relevant to snowfalls in the high plains region].

## 5. Summary

Consideration of the cloud microphysical aspects of the production of snowfall combined with a principal component analysis of radiosonde and surface data led to the identification of seven factors that influence the diagnosis of snow ratio. They are solar radiation (month), low- to midlevel temperature, mid- to upper-level temperature, low- to midlevel relative humidity, midlevel relative humidity, upper-level relative humidity, and external compaction (surface wind speed and liquid equivalent precipitation amount). A 10-member ensemble of artificial neural networks was shown to improve the diagnosis of snow-ratio class compared to existing techniques (10-to-1 ratio, sample climatology, NWS new snowfall to estimated meltwater conversion table). The most critical factors in the neural network performance are related to month, temperature, and external compaction, with relative humidity information providing useful, but less essential input.

Until an integrated and well-verified scheme for forecasting snow is provided as a direct output from numerical weather prediction models, it will be necessary to convert liquid water to snow through the diagnosis of snow ratio. The method developed in this paper could be applied in the operational environment as follows: (a) QPF techniques, including the output from numerical models, are used to determine the amount of liquid equivalent precipitation; (b) forecast soundings and the QPF obtained from (a) are used within the neural network framework described in this study to diagnose the snow-ratio class; (c) the snowfall forecast is derived from the results of (a) and (b). For example, consider the simple case in which 0.5 in. of liquid equivalent precipitation is forecast and for which the snow-ratio class is diagnosed to be light. Then, the forecast snow depth would be approximately 10 in. (reflecting a median snow ratio of 20:1 in the light class). It should be noted that this forecast depth is twice that obtained by application of the 10-to-1 rule. Probabilities could be generated by using probabilistic QPFs, normalizing the network outputs by class frequency and accounting for the frequency distribution of snow ratios within each class. Such manipulations of forecast data could be accomplished through the "Smart Tool" portion of the Interactive Forecast Preparation System (IFPS), a component of the Advanced Weather Interactive Processing System (AWIPS) at NWS facilities. A pilot study of the

ANN technique will be implemented during winter 2002/03 in the Pacific and Yukon regions of Canada by the Meteorological Service of Canada. Given the large number of weights required to implement an ensemble of ANNs (e.g., a single hidden layer MLP with seven inputs, 40 processing elements, and three outputs requires 300 weights), operational implementation of the technique would be most readily accomplished by acquiring the training datasets used in this study and training the networks locally. Interested persons should contact the study authors.

Although the findings in this paper suggest that improvements in current snow forecasting practice are already obtainable, a key aspect, the role of in-cloud vertical motions (highlighted in the microphysical review of section 1), has been neglected in this study. It is expected that cloud-scale vertical motion input would further improve these results, perhaps substantially so. Empirical evidence for the importance of vertical motion is provided online [<http://www.nws.noaa.gov/er/hq/ssd/snowmicro>; see also Auer and White (1982)]. In this preliminary study, the intersection of areas of moderate vertical motion with temperatures favorable for dendritic crystal formation ( $-12^{\circ}$  to  $-18^{\circ}\text{C}$ ) within cloudy regions was found to be a useful indicator of snow advisory events. One means to obtain such vertical motions for a longitudinal study of this kind might be to use a mesoscale model and the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis (Kalnay et al. 1996) within a four-dimensional data assimilation framework. In this approach, the reanalysis data could be used to provide lateral boundary conditions in a nested mesoscale model configuration, with an interior domain of sufficient resolution to obtain the necessary cloud-scale vertical motions.

Other physical effects have also been neglected. In early and late season, ground temperature effects may lead to melting and compaction of new snowfalls. Strong low-level winds can lead to snow crystal fragmentation and density changes. Cloud aerosol effects may also contribute to the resulting density of snowfall. Borys et al. (2000) showed that snowfall rate decreased when clear-air equivalent sulfate concentration increased. They hypothesized that increased pollution results in more cloud condensation nuclei, more cloud droplets, and less efficient riming. Another factor that may negatively affect the performance of the network diagnosis is storm electrification. As reviewed in MacGorman and Rust (1998, 352–355), variations in the electrical state inside the cloud can produce significant changes in the likelihood of aggregation and fragmentation of ice crystals. The range of electric field over which these effects can be important are consistent with such measurements inside winter storms (e.g., Schultz et al. 2002). Such investigations could form the basis of future improvements in this technique.

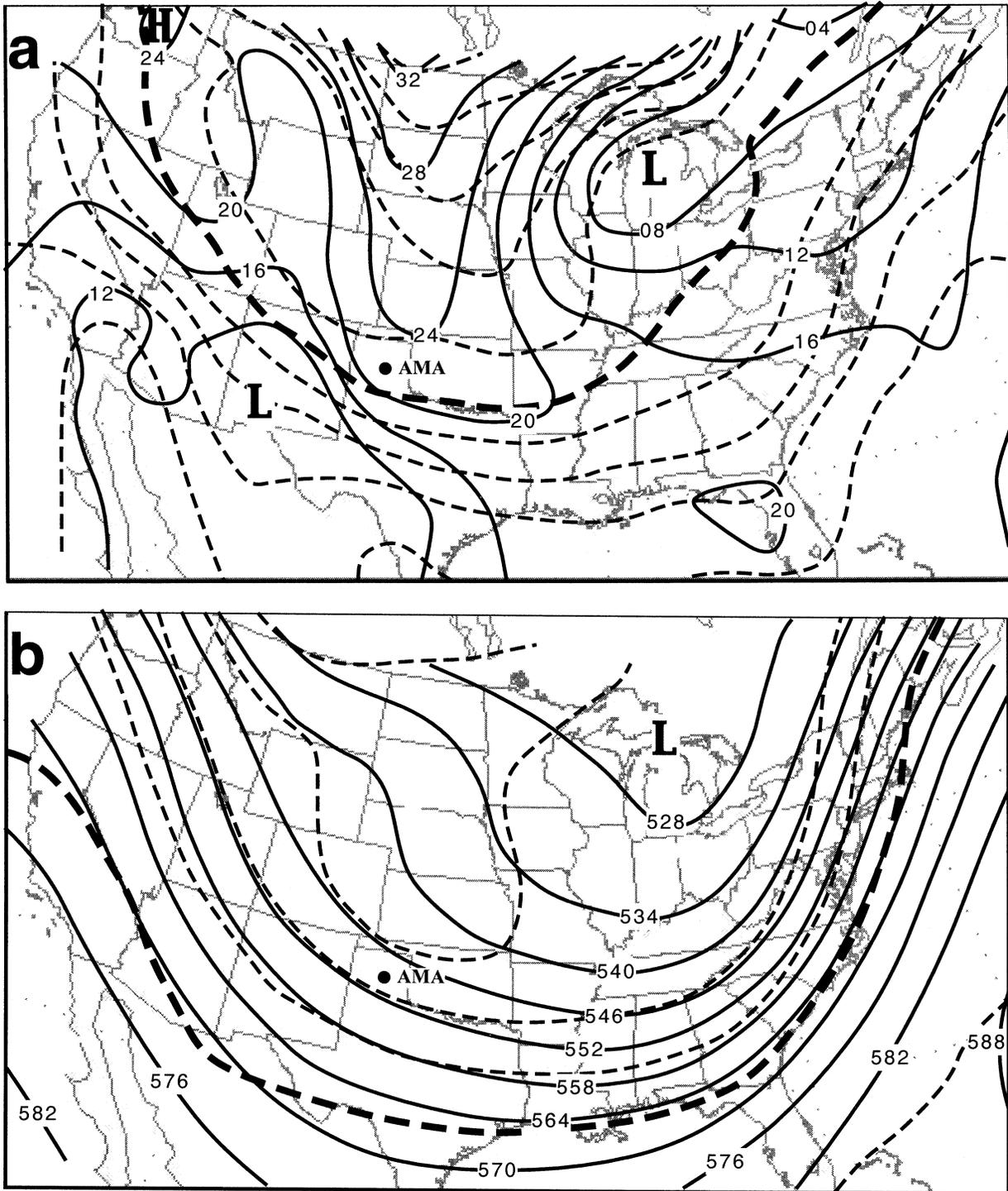


FIG. 10. (a) Surface analysis at 0000 UTC 16 Dec 1983, (b) 500-hPa analysis at 0000 UTC 16 Dec 1983, (c) surface analysis at 1200 UTC 28 Dec 1993, and (d) 500-hPa analysis at 1200 UTC 28 Dec 1993. The location of the case site is indicated by the black [(a), (b) AMA, (c), (d) HTS]. For surface analyses, solid (dashed) lines are isobars (isotherms), with contour interval of 4 hPa (5°C, thick line is 0°C). For 500-hPa analyses, solid (dashed) lines are geopotential height (isotherms), with contour interval of 6 dam (5°C, thick line is -15°C).

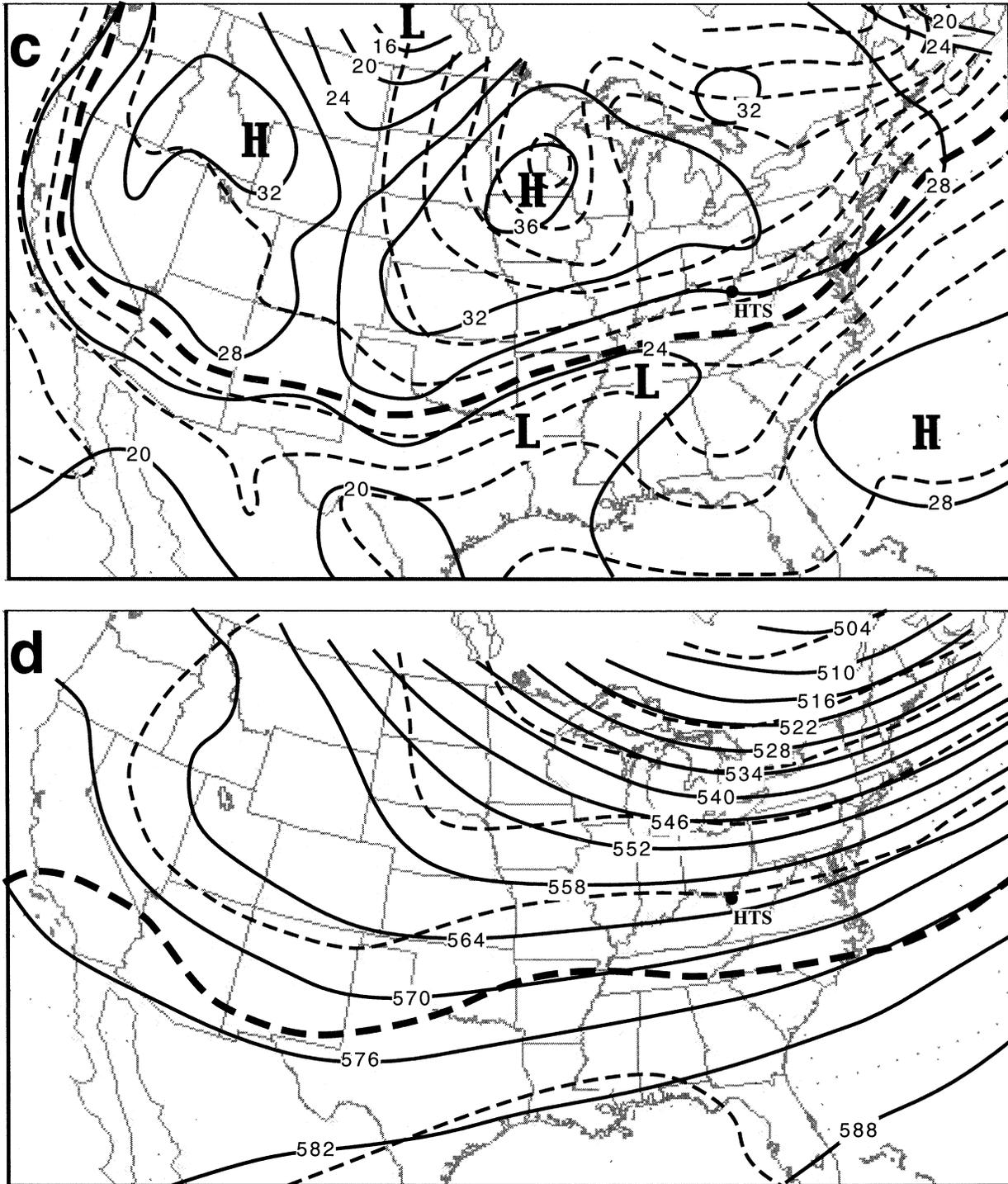


FIG. 10. (Continued)

*Acknowledgments.* The lead authors (P. Roebber and S. Bruening) thank the Department of Mathematical Sciences at the University of Wisconsin—Milwaukee (UWM) for providing some funding support for this work. Funding for D. Schultz and J. Cortinas was pro-

vided by NOAA/OAR/NSSL under NOAA–OU Cooperative Agreement NA17RJ1227. The use of the Storm Prediction Center microfilm collection is gratefully acknowledged. The comments of V. Larson of UWM helped to clarify aspects of the microphysical

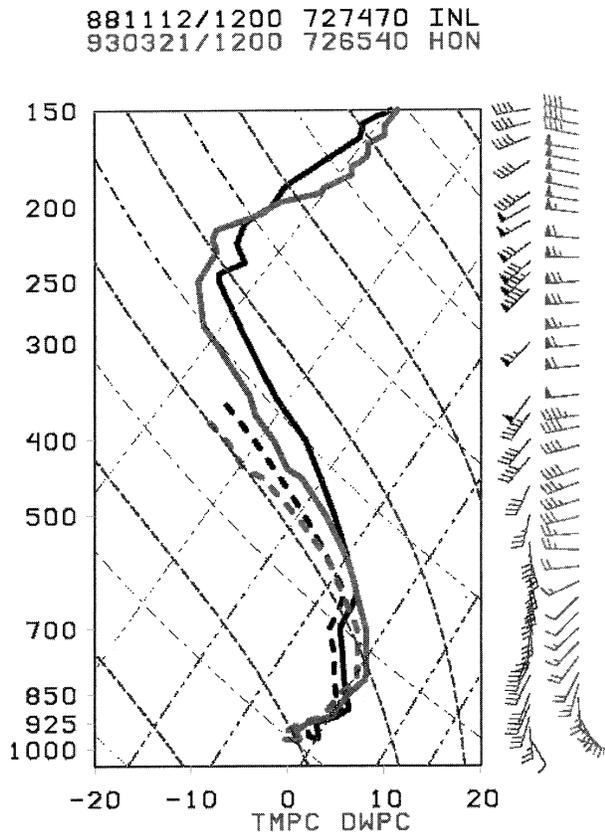


FIG. 11. Thermodynamic diagrams at 1200 UTC 12 Nov 1988 at INL (black) and 1200 UTC 21 Mar 1993 at HON (gray). The solid line is temperature and the dashed line is dewpoint temperature. Winds are plotted according to the standard meteorological convention.

discussion. Further comments by Larry Dunn, Ethan Greene, Steve Vasiloff, and the three anonymous reviewers were appreciated.

## APPENDIX

### Background on ANNs

This section provides additional detail concerning the construction and operation of ANNs, following the information provided online (<ftp://ftp.sas.com/pub/neural/FAQ.html>). An ANN builds discriminant functions from its processing elements and the architecture determines the number and shape of the discriminant functions. Since ANNs are sufficiently powerful to create arbitrary discriminant functions, ANNs can, in theory, achieve optimal classification. Further, since ANNs are adaptive systems, the parameters of the system under study need not be specified (in contrast, for example, to an expert system) but rather are extracted from the input data and the desired outputs (snow-ratio classes) by means of the training algorithm. In this way, if the relevant inputs are known but the precise relationship between the inputs and the snow-ratio classes is not known, diagnosis is

still possible. Additionally, postprocessing of the diagnoses may reveal new knowledge concerning these relationships (see section 4).

A major ANN architecture is the feed-forward type (i.e., the connections between processing elements do not form loops) known as the multilayer perceptron [MLP; Figs. 4a,b; Bishop (1996)]. This architecture can contain one or more hidden layers, that is, layers containing processing elements that are not directly connected to either the inputs or the desired outputs. It can be shown that an MLP with one hidden layer and an arbitrarily large number of processing elements can approximate any function (e.g., Bishop 1996). In practice, however, it is not possible to specify the number of processing elements that are required, and for more complicated functions, it is often more efficient to have several processing elements in a second hidden layer. Networks with too few hidden processing elements will generalize poorly as a result of underfitting (i.e., insufficient specification of the mapping between the inputs and the outputs), while networks with too many hidden processing elements will also generalize poorly, in this case due to overfitting (which produces a model of the statistical noise as well as the desired signal). Since there is no theoretical basis for defining the number of hidden processing elements, this aspect of the architecture is obtained through experimentation.

The connections between processing elements are constructed using combination and activation functions. In MLPs, each noninput processing element linearly combines values that are fed into it through connections with other processing elements, producing a single value called the net input. An activation function then nonlinearly transforms the net input (necessary to represent nonlinear discriminant functions), yielding an "activation," which is fed through connections to other processing elements. A common activation function, used in this study, is the hyperbolic tangent, which tends to produce fast training. The desired outputs, the snow-ratio classes, are defined using 1-of-C coding, such that three binary output variables (0/1) are specified, one for each snow-ratio class.

A requirement for a network to generalize well is its ability to represent every part of the multidimensional input space. In this way, it is possible to define the mapping from the inputs to the desired outputs. Since the data required to define the mapping is proportional to the hypervolume of the input space, networks with more inputs will require more training examples. Further, the training set must be a sufficiently large and representative sample of the population to promote interpolation (i.e., test cases are in the "neighborhood" of training cases) rather than extrapolation (i.e., test cases lie outside the range of the training cases or are within holes in the training input space). This also suggests that if irrelevant inputs are provided to the network, then performance will tend to be poor, since the network will squander resources in order to represent irrelevant

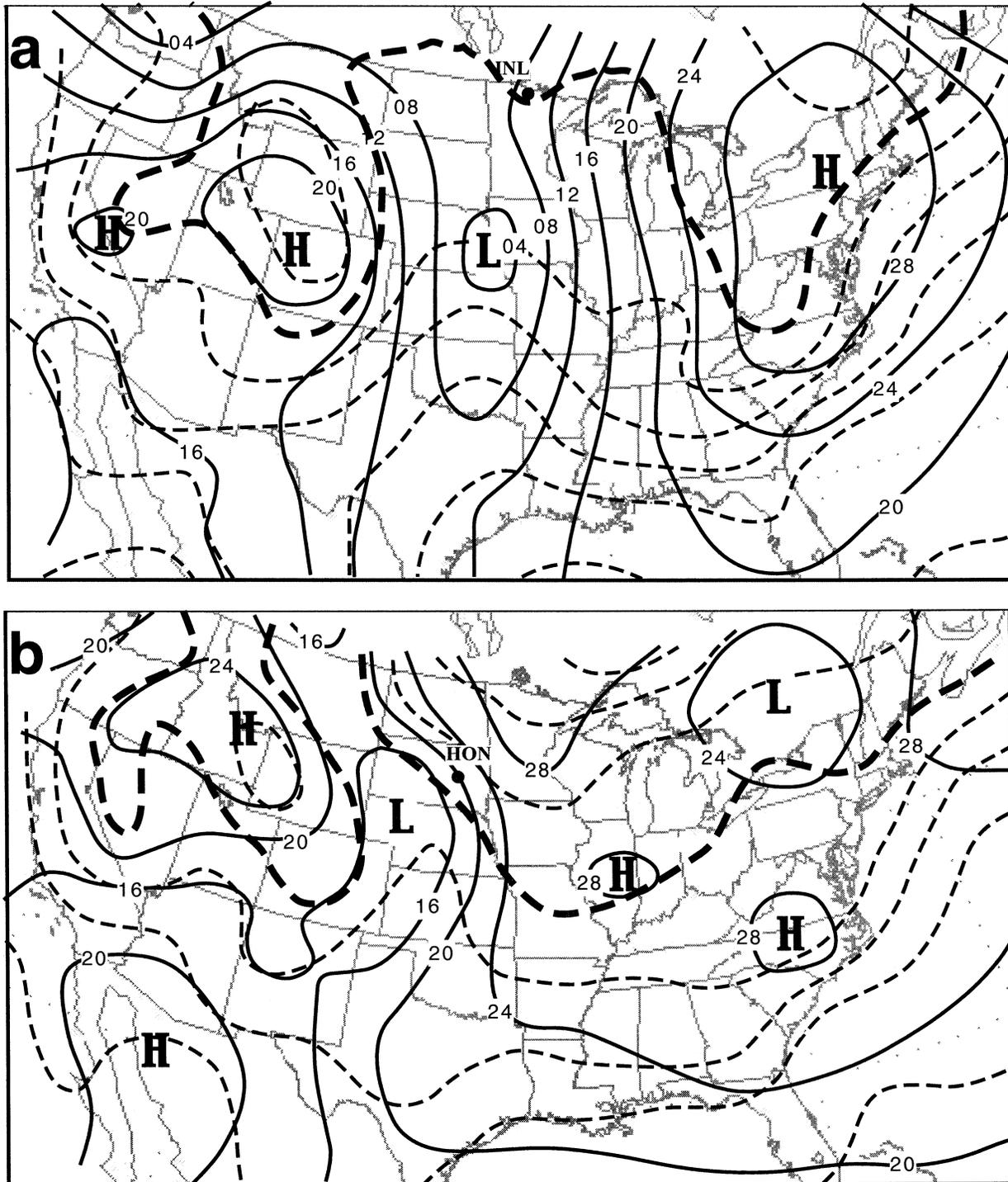


FIG. 12. (a) Surface analysis at 1200 UTC 12 Nov 1988, and (b) surface analysis at 1200 UTC 21 Mar 1993. The location of the case site is indicated by the black dot [(a) INL, (b) HON]. For surface analyses, solid (dashed) lines are isobars (isotherms), with contour interval of 4 hPa (5°C, thick line is 0°C).

portions of the mapping. Hence, careful selection of inputs is required to improve generalization. Another important generalization consideration relates to noise in the desired outputs. Noise increases the risk of ov-

erfitting (Moody 1992), since the network will attempt to approximate artificial boundaries in the training data rather than the desired signal. Problems associated with overfitting can be reduced by using large training sets,

applying early stopping, and combining networks (into a so-called committee of networks).

The principle of a committee of networks (hereafter, the ensemble network) is tied to the familiar concept of ensemble forecasting. For a set of  $M$  networks for which the errors are uncorrelated with each other and for which there is no bias, it can be shown that the error of the ensemble network will be  $M$  times smaller than the mean error of the individual networks (Principe et al. 2000). In practice, network errors will tend to be correlated such that the actual advantage of the ensemble is less than that suggested by the theory. By incorporating different network architectures into the ensemble, it is possible to reduce the occurrence of correlated errors and improve ensemble network performance (Opitz and Maclin 1999 and references therein). Further improvements are made possible by altering the training process, for example, by creating different training sets for each individual network through bootstrapping techniques (Efron and Tibshirani 1993; Breiman 1996; Opitz and Maclin 1999). Previous research suggests that ensembles of  $O(10)$  members will provide most of the advantage obtainable using this technique (e.g., Opitz and Maclin 1999).

#### REFERENCES

- Abbe, C., 1888: Appendix 46. 1887 Annual report of the Chief Signal Officer of the Army under the direction of Brigadier-General A. W. Greely, U.S. Govt. Printing Office, Washington, DC, 385–386.
- Auer, A. H., and J. M. White, 1982: The combined role of kinematics, thermodynamics, and cloud physics associated with heavy snowfall episodes. *J. Meteor. Soc. Japan*, **60**, 500–507.
- Bishop, C. M., 1996: *Neural Networks for Pattern Recognition*. Clarendon Press, 482 pp.
- Borys, R. D., D. H. Lowenthal, and D. L. Mitchell, 2000: The relationships among cloud microphysics, chemistry, and precipitation rate in cold mountain clouds. *Atmos. Environ.*, **34**, 2593–2602.
- Breiman, L., 1996: Stacked regressions. *Mach. Learn.*, **24**, 49–64.
- Byers, H. R., 1965: *Elements of Cloud Physics*. University of Chicago Press, 191 pp.
- Cattell, R. B., 1966: The Scree test for the number of factors. *Multivariate Behavioral Res.*, **1**, 140–161.
- , and J. A. Jaspers, 1967: *A General Plasmode (No. 30-10-5-2) for Factor Analytic Exercises and Research. Multivariate Behavioral Research Monographs*, Vol. 67, No. 3, Society of Multivariate Experimental Psychology, 212 pp.
- Charba, J. P., and W. H. Klein, 1980: Skill in precipitation forecasting in the National Weather Service. *Bull. Amer. Meteor. Soc.*, **61**, 1546–1555.
- Chaston, P., 1989: The magic chart for forecasting snow amounts. *Natl. Wea. Dig.*, **14** (2), 20–22.
- Currie, B. W., 1947: Water content of snow in cold climates. *Bull. Amer. Meteor. Soc.*, **28**, 150–151.
- Doesken, N. J., and A. Judson, 1997: *The Snow Booklet: A Guide to the Science, Climatology, and Measurement of Snow in the United States*. Dept. of Atmospheric Science, Colorado State University, 86 pp.
- , and R. J. Leffler, 2000: Snow foolin': Accurately measuring snow is an inexact but important science. *Weatherwise*, **53** (1), 30–37.
- Efron, B., and R. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, 436 pp.
- Ferguson, S. A., M. B. Moore, R. T. Marriott, and P. Speers-Hayes, 1990: Avalanche weather forecasting at the northwest avalanche center, Seattle, WA. *J. Glaciol.*, **36**, 57–66.
- Fritsch, J. M., and Coauthors, 1998: Quantitative precipitation forecasting: Report of the Eighth Prospectus Development Team, U.S. Weather Research Program. *Bull. Amer. Meteor. Soc.*, **79**, 285–298.
- Fukuta, N., and T. Takahashi, 1999: The growth of atmospheric ice crystals: A summary of findings in vertical supercooled cloud tunnel studies. *J. Atmos. Sci.*, **56**, 1963–1979.
- Garcia, C., 1994: Forecasting snowfall using mixing ratios on an isentropic surface—An empirical study. NOAA Tech. Memo. NWS CR-105, PB 94-188760, NOAA/NWS, 31 pp. [Available from NOAA/National Weather Service Central Region Headquarters, Kansas City, MO 64106-2897.]
- Gilbert, G. F., 1884: Finley's tornado predictions. *Amer. Meteor. J.*, **1**, 166–172.
- Goodison, B. E., 1978: Accuracy of Canadian snow-gage measurements. *J. Appl. Meteor.*, **17**, 1542–1548.
- Gorsuch, R. L., 1983: *Factor Analysis*. L. Erlbaum Associates, 425 pp.
- Grant, L. O., and J. O. Rhea, 1974: Elevation and meteorological controls of the density of snow. *Proc. Advanced Concepts Tech. Study Snow Ice Resources Interdisciplinary Symp.*, National Academy of Science, 169–181.
- Gray, D. M., and D. H. Male, 1981: *Handbook of Snow: Principles, Processes, Management and Use*. Pergamon Press, 776 pp.
- Groisman, P. Ya., and D. R. Legates, 1994: The accuracy of United States precipitation data. *Bull. Amer. Meteor. Soc.*, **75**, 215–227.
- , V. V. Koknaeva, T. A. Belokrylova, and T. R. Karl, 1991: Overcoming biases of precipitation measurement: A history of the USSR experience. *Bull. Amer. Meteor. Soc.*, **72**, 1725–1733.
- Hotelling, H., 1933: Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, **24**, 417–498.
- Houze, R. A., Jr., 1993: *Cloud Dynamics*. Academic Press, 573 pp.
- Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Amer. Meteor. Soc.*, **79**, 1855–1870.
- Judson, A., and N. Doesken, 2000: Density of freshly fallen snow in the central Rocky Mountains. *Bull. Amer. Meteor. Soc.*, **81**, 1577–1587.
- Kaiser, H. F., 1958: The Varimax criterion for Varimax rotation in factor analysis. *Psychometrika*, **23**, 187–200.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kohonen, T., 1995: *Self-Organizing Maps*. Springer-Verlag, 362 pp.
- Kyle, J. P., and D. A. Wesley, 1997: New conversion table for snowfall to estimated meltwater: Is it appropriate in the High Plains? NWS Central Region Applied Research Paper 18-04, NOAA/NWS Central Region, Scientific Services Division. [Available online at <http://www.crh.noaa.gov/techpapers/arp18/18-04.html>.]
- LaChapelle, E. R., 1962: The density distribution of new snow. Project F, Progress Rep. 2, USDA Forest Service, Wasatch National Forest, Alta Avalanche Study Center, Salt Lake City, UT, 13 pp.
- , 1980: The fundamental process in conventional avalanche forecasting. *J. Glaciol.*, **26**, 75–84.
- Larson, L., and E. L. Peck, 1974: Accuracy of precipitation measurements for hydrologic modeling. *Water Resour. Res.*, **10**, 857–862.
- Li, L., and J. W. Pomeroy, 1997: Estimates of threshold wind speeds for snow transport using meteorological data. *J. Appl. Meteor.*, **36**, 205–213.
- MacGorman, D. R., and W. D. Rust, 1998: *The Electrical Nature of Storms*. Oxford University Press, 422 pp.
- Magono, C., and T. Nakamura, 1965: Aerodynamic studies of falling snow flakes. *J. Meteor. Soc. Japan*, **43**, 139–147.
- , and C. W. Lee, 1966: Meteorological classification of natural snow crystals. *J. Fac. Sci., Hokkaido Univ., Ser. 7*, **11**, 321–335.

- Marzban, C., and A. Witt, 2001: A Bayesian neural network for severe-hail size prediction. *Wea. Forecasting*, **16**, 600–610.
- Minsk, L. D., 1998: *Snow and Ice Control Manual for Transportation Facilities*. McGraw-Hill, 289 pp.
- Moody, J. E., 1992: The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. *Adv. Neural Inf. Process. Syst.*, **4**, 847–854.
- Nouhan, V. J., 1999: Measuring snowfall, depth and liquid equivalent during and after extreme blizzard events. NWS Central Region Applied Research Paper 20-05, NOAA/NWS Central Region, Scientific Services Division. [Available online at <http://www.crh.noaa.gov/techpapers/arp20/20-05.html>.]
- Olson, D. A., N. W. Junker, and B. Korty, 1995: Evaluation of 33 years of quantitative precipitation forecasting at NMC. *Wea. Forecasting*, **10**, 498–511.
- Opitz, D., and R. Maclin, 1999: Popular ensemble methods: An empirical study. *J. Artif. Intell. Res.*, **11**, 169–198.
- Panofsky, H. A., and G. W. Brier, 1958: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University Press, 200 pp.
- Passarelli, R. E., Jr., and R. C. Srivastava, 1979: A new aspect of snowflake aggregation theory. *J. Atmos. Sci.*, **36**, 484–493.
- Peck, E. L., 1972: Snow measurement predicament. *Water Resour. Res.*, **8**, 244–248.
- Perla, R., 1970: On contributory factors in avalanche hazard evaluation. *Can. Geotech. J.*, **7**, 414–419.
- Potter, J. G., 1965: Water content of freshly fallen snow. CIR-4232, TEC-569, Meteorology Branch, Dept. of Transport, Toronto, ON, Canada, 12 pp. [Available from National Snow and Ice Data Center User Services, University of Colorado, Campus Box 449, Boulder, CO 80309-0449.]
- Power, B. A., P. W. Summers, and J. d'Avignon, 1964: Snow crystal forms and riming effects as related to snowfall density and general storm conditions. *J. Atmos. Sci.*, **21**, 300–305.
- Principe, J. C., N. R. Euliano, and W. C. Lefebvre, 2000: *Neural and Adaptive Systems: Fundamentals through Simulations*. John Wiley and Sons, 656 pp.
- Pruppacher, H. R., and J. D. Klett, 1997: *Microphysics of Clouds and Precipitation*. 2d ed. Kluwer Academic, 954 pp.
- Roebber, P. J., and L. F. Bosart, 1998: The sensitivity of precipitation to circulation details: Part I: An analysis of regional analogs. *Mon. Wea. Rev.*, **126**, 437–455.
- Ryan, B. F. D., E. R. Wishart, and D. E. Shaw, 1976: The growth rates and densities of ice crystals between  $-3^{\circ}\text{C}$  and  $-21^{\circ}\text{C}$ . *J. Atmos. Sci.*, **33**, 842–850.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575.
- Schultz, D. M., J. V. Cortinas Jr., and C. A. Doswell III, 2002: Comments on “An operational ingredients-based methodology for forecasting midlatitude winter season precipitation.” *Wea. Forecasting*, **17**, 160–167.
- , and Coauthors, 2002: Understanding Utah winter storms: The Intermountain Precipitation Experiment. *Bull. Amer. Meteor. Soc.*, **83**, 189–210.
- Schwartz, B. E., and M. Govett, 1992: A hydrostatically consistent North American radiosonde data base at the Forecast Systems Laboratory, 1946–present. NOAA Tech. Memo. ERL FSL-4, 81 pp. [Available from NOAA/FSL, 325 Broadway, Boulder, CO 80303.]
- Siegel, S., 1956: *Non-Parametric Statistics for the Behavioral Sciences*. McGraw-Hill, 312 pp.
- Super, A. B., and E. W. Holroyd III, 1997: Snow accumulation algorithm for the WSR-88D radar: Second annual report. Bureau Reclamation Tech. Rep. R-97-05, U.S. Dept. of Interior, Denver, CO, 77 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]
- Takahashi, T., T. Endoh, and G. Wakahoma, 1991: Vapor diffusional growth of free-falling snow crystals. *J. Meteor. Soc. Japan*, **69**, 15–30.
- U.S. Department of Commerce, 1996: Supplemental observations. Part IV, National Weather Service Observing Handbook No. 7: Surface Weather Observations and Reports, National Weather Service, Silver Spring, MD, 57 pp.
- Wetzel, S. W., and J. E. Martin, 2001: An operational ingredients-based methodology for forecasting midlatitude winter season precipitation. *Wea. Forecasting*, **16**, 156–167.