Validation of TOVS Path-P data during SHEBA

Axel J. Schweiger,[1] Ronald W. Lindsay,[1] Jennifer A. Francis,[2] Jeff Key,[3]
Janet M. Intrieri,[4] and Matthew D. Shupe[5]

Received 23 May 2000; revised 24 April 2001; accepted 6 November 2001; published 28 September 2002.

[1] Products from the TIROS-N Operational Vertical Sounder (TOVS) Polar Pathfinder (Path-P) data set are compared with surface measurements and other satellite remote sensing retrievals during the Surface Heat Balance of the Arctic Ocean (SHEBA) field program (October 1997 to September 1998). The comparison provides estimates of Path-P retrieval uncertainties. Results are placed in the context of the natural variability and timescales of variability to allow potential users to judge the applicability of the data set for their purpose. Results show temperature profiles to be accurate within 3 K, total column precipitable water within 2 mm annually, and surface temperature within 3 K. Uncertainties in temperature retrieval are below "within-season" variability during all times of the year. Uncertainties in water vapor retrieval during winter and summer are slightly below observed variability in those seasons but are well below during spring. Uncertainty in retrieved cloud fraction is highly dependent on the timescale of observations. Cloud fractions from the surface and satellite are well correlated (correlation coefficient > 0.7) at timescales greater than 4 days but show weaker correlation at shorter timescales. Uncertainty in TOVS-retrieved cloud fraction is less than 20% for 5-day averages. In winter, TOVS-retrieved cloud fractions are higher than those reported in standard meteorological observations but match those derived from lidar data. This supports the notion that standard meteorological observations may underestimate cloudiness in winter. Cloud-top temperatures measured from the surface (lidar/radar) are significantly different from those estimated using TOVS and Advanced Very High Resolution Radiometer (AVHRR) radiances, which highlights the fundamental and inherent dissimilarity between these two measurement techniques. *INDEX TERMS:* 0320 Atmospheric Composition and Structure: Cloud physics and chemistry; 1640 Global Change: Remote sensing; 3349 Meteorology and Atmospheric Dynamics: Polar meteorology; 3339 Meteorology and Atmospheric Dynamics: Ocean/atmosphere interactions (0312, 4504); 4207 Oceanography: General: Arctic and Antarctic oceanography; *KEYWORDS:* TOVS, sea ice, validation, atmospheric profiles, cloud cover

Citation: Schweiger, A. J., R. W. Lindsay, J. A. Francis, J. Key, J. M. Intrieri, and M. D. Shupe, Validation of TOVS Path-P data during SHEBA, *J. Geophys. Res.*, *107*(C10), 8041, doi:10.1029/2000JC000453, 2002.

1. Introduction

[2] The TIROS-N Operational Vertical Sounder (TOVS) Polar Pathfinder (Path-P) data set is a collection of surface and atmospheric variables retrieved from the TOVS aboard NOAA's polar orbiting satellites. The data set covers the period from 1979 to 1998 and is available for the Arctic basin north of 60° latitude. It was designed to provide validation and forcing data for climate process models and information for the analysis of climate variability.

[3] The performance of general circulation models (GCMs) in the polar regions has been rather poor [cf. *Gates et al.*, 1999]. Similarly, satellite-derived global data sets have exhibited significant errors in the polar regions [cf. *Schweiger and Key*, 1992]. The Surface Heat Balance of the Arctic Ocean (SHEBA) experiment [*Perovich et al.*, 1999], which was conducted in the Beaufort Sea from October 1997 to September 1998, was largely motivated by these deficiencies. The research program is designed to construct detailed models of relevant local physical processes (process models) from SHEBA measurements, which will then be used to develop simpler and more realistic representations or parameterizations of those processes for implementation in GCMs. Improvements to high-latitude simulations by process models and GCMs hinges on the ability to evaluate their performance on a

[1]Applied Physics Laboratory/Polar Science Center, University of Washington, Seattle, Washington, USA.
[2]Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey, USA.
[3]Office of Research and Applications, National Environmental Satellite, Data, and Information Service, NOAA, University of Wisconsin, Madison, Wisconsin, USA.
[4]NOAA/Environmental Technology Laboratory, Boulder, Colorado, USA.
[5]Science and Technology Corporation, Hampden, Virginia, USA.

0148-0227/02/2000JC000453$09.00

basin-wide scale. SHEBA data provide unprecedented opportunities to validate satellite retrieval algorithms and products, which constitute the most promising sources of information for model validation. Satellite products can also be applied to evaluating basin-wide GCM output, forcing stand-alone coupled ice-ocean models, and studying variability of mechanisms governing climate processes.

[4] Temperature and humidity profiles have been retrieved operationally from TOVS radiances during the past 20 years by NOAA. Most development and validation efforts were focused on retrievals over "wet" oceans where few radiosonde stations exists and, consequently, where numerical weather forecast models must rely on the assimilation of data from the TOVS system. Over the past 10 years several efforts have focused on the improvement and validation of TOVS retrievals over ice-covered oceans [cf. *Claud et al.*, 1991; *Francis*, 1994; *Koepken et al.*, 1995]. Validation efforts have been limited to relatively short field experiments, which precluded the evaluation of TOVS products over an entire annual cycle. Remote retrieval of temperature and humidity profiles is severely hampered by clouds. Successful detection and removal of their effects must be achieved before profiles can be computed in cloudy areas. Retrieved cloud information has consequently been viewed more as a secondary product. Recently this perception has changed, and the potential for estimating cloud fraction and height from TOVS radiances has been recognized [cf. *Stubenrauch et al.*, 1999a, 1999b, 1999c; *Francis*, 1997]. The quality of TOVS-based cloud retrievals over Arctic sea ice was recently demonstrated through comparisons with meteorological observations from drifting ice stations [*Schweiger et al.*, 1999]. While these results are encouraging, we recognize that the quality of conventional cloud observations during winter may be questionable owing to the extended darkness. Data from upward-looking cloud radar and lidar during SHEBA, therefore, represent an excellent opportunity for more quantitative validation of satellite-derived cloud products.

[5] In this paper we compare surface and remote sensing data from the SHEBA experiment to variables from the TOVS Path-P data set for the same period (October 1997 to September 1998). We employ measurements obtained during SHEBA that span a wide variety of sources and extend over an entire annual cycle. Our goal is to provide estimates of the uncertainties in the Path-P products, allowing other researchers to judge whether a particular Path-P product is suitable for their particular application. Although we do provide possible explanations for discrepancies between validation and retrieval data, our focus in this paper is to provide uncertainty statistics. Research into specific causes and potential improvements will be left for future investigations.

2. Approach

[6] Validation of satellite data with surface measurements is inherently difficult. There is a mismatch in spatial and temporal resolution, as well as in the fundamental characteristics of the observations (e.g., viewing of clouds from above and below). It is important, therefore, to keep in mind that the "retrieval errors" presented in this paper are more properly referred to as "retrieval uncertainties." In our analyses we will use the term uncertainty. An uncertainty of X denotes a root mean square (RMS) difference between the Path-P retrieval and the observation. Assuming Gaussian distribution of the differences, the uncertainty represents a one-standard-deviation likelihood ($\sim$70%) that any randomly selected comparison of a satellite-derived value with an observation will be within X of the measurement. We also present mean errors to estimate possible biases in the data set. To quantify the accuracy with which the variability is captured (which is not evident from the RMS and mean errors) and to provide users with information to aid in judging whether a product is suitable for their application, we have included correlations with validation data at varying timescales. In some instances we compare expected variability in physical processes to the uncertainties at a particular timescale, to help users decide if their sought-after signal, e.g., a climate parameter or anomaly pattern, will exceed the noise level in the retrievals.

[7] To aid the reader, the observing system from which a measurement is obtained is defined in the text upon first mention, and a summary of data sources and their abbreviations is presented in Table 1.

3. Data Sources and Methodology

3.1. Path-P Data

3.1.1. Variables

[8] Table 2 lists the products contained in the Path-P data set, which are retrieved from infrared and microwave radiances measured by the TOVS instrument package. The TOVS has flown continuously on NOAA polar-orbiting satellites since 1978 and consists of three radiometer arrays: the high-resolution infrared radiation sounder (HIRS), microwave sounding unit (MSU), and stratospheric sounding unit (SSU). Data from the SSU are not used. We employ a modified version of the Improved Initialization Inversion ("3I") algorithm developed by the Atmospheric Radiation Analysis group at the Laboratoire de Météorologie Dynamique (LMD). The algorithm has evolved over time, including modifications to improve results over snow and ice surfaces [*Francis*, 1994]. Detailed descriptions of 3I can be found in the studies of *Chédin et al.* [1985], *Francis* [1994], *Stubenrauch et al.* [1996], and *Scott et al.* [1999]. Here we provide only a brief synopsis of the fundamental steps in the algorithm.

3.1.1.1. Calibration

[9] MSU and HIRS level 1b data are calibrated using the standard procedure outlined in the NOAA Polar Orbiter Users Guide [*Kidwell*, 1998]. Calibration coefficients provided in the level 1b data stream are applied to convert digital counts to brightness temperatures. Limb corrections to account for variations in viewing angle and the associated lifting of weighting functions and changes in surface emissivity are applied to MSU channels. MSU channel 1, which is particularly sensitive to surface emission over sea ice, is limb-corrected for atmospheric effects only. The view-angle dependence of surface emissivity in MSU channel 1 is particularly large over sea ice and is addressed in a later processing step. The limb-correction

Table 1. Frequently Used Abbreviations and Definitions

Variable Abbreviation	Description of Variable
3I	Improved Initialization Inversion algorithm
ARM	Atmospheric Radiation Measurement Program (DOE)
AVHRR	Advanced Very High Resolution Radiometer
CART	Cloud and Radiation Testbed (ARM program)
CASPR	Cloud and Surface Parameter Retrieval
C-LIDAR	Cloud fraction derived from LIDAR cloud height retrieval data set by time averaging over 1 day.
C-METOBS	Observation of cloud fraction by surface meteorological observer. Reported in eighths.
DABUL	Depolarization and Backscatter Unattended Lidar
DELTA	Differences (deltas) between modeled and observed TOVS brightness temperatures
ETL	Environmental Technology Laboratory (NOAA, Boulder)
GLAS-H	GPS/Loran Atmospheric Sonde Humidity
GLAS-T	GPS/Loran Atmospheric Sonde Temperature
GCM	General Circulation Model
HIRS	High resolution Infrared Sounder
ICP	Ice Crystal Precipitation
ISCCP	International Satellite Cloud Climatology Project
ITPP	International TOVS Processing Package
Level 1b	Satellite data provided in orbital swath with calibration and geolocation information attached.
Level 2	Geophysical retrievals in orbital swath format
Level 3	Geophysical retrievals gridded to regular grid. Path-P is gridded using a simple drop-in-the-bucket binning scheme.
LMD	Laboratoire Météorologie Dynamique
MMCR	Millimeter Cloud Radar
MSU	Microwave Sounding Unit
NOAA	National Oceanic and Atmospheric Administration
NP	North Pole drifting stations (former Soviet ice islands)
NWP	Numerical Weather Prediction
Path-P	TOVS Polar Pathfinder
SHEBA	Surface Heat Balance of the Arctic
SPO	SHEBA project office
RMS	Root Mean Square (Error)
TB(s)	Brightness Temperature(s)
Tc-CASPR	Cloud-top temperature retrieved from AVHRR using the CASPR system
TEMP-2m	Surface temperature measured at the SHEBA 20-m tower. Temperature measured at 2 m above surface.
TIGR	Thermodynamic Initial Guess Retrieval library of atmospheric profiles
TOA	Top of the Atmosphere
TOVS	TIROS-N Operational Vertical Sounder
WMO	World Meteorological Organization

procedure is adapted from the ITTP processing scheme [*Smith et al.*, 1985].

3.1.1.2. Interpolation of MSU Data and Grouping Into Retrieval Boxes

[10] The MSU and HIRS sensors have nadir footprint sizes of approximately 100 and 18 km, respectively. The 3I processing scheme interpolates surrounding MSU fields-of-view and groups HIRS pixels into 100-km retrieval boxes. Because of the change in footprint size along the scan line, a varying number of HIRS pixels are associated with each retrieval box (see Figure 1).

3.1.1.3. Air Mass Determination

[11] To reduce the number of situations involved in the first-guess search and to guide other algorithm branches,

Table 2. Path-P Variables

Parameter Name	Description	Units	~RMS Error
TEMP	temperature at 9 levels (50, 70, 100, 300, 400, 500, 600, 700, 850, 900 mb)	K	3
SKTEMP	surface skin temp	K	3
WVAPOR	precipitable water in 5 layers (300–400, 400–500, 500–700, 700–850, 850–surface)	mm	30%
FCLD	effective cloud fraction (εN)	%	30
HIRS_CLDY	% cloudy HIRS pixels	%	20
CLTEMP	cloud-top temperature	K	TBD
CLPRESS	cloud-top pressure	mbar	TBD
EMISS	surface emissivity @ 50 GHz		5%
PBLSTRAT	bulk PBL stratification	K	5
Cg	geostrophic drag coeff.		10%
ALPHA	wind turning angle	deg	20
ZANGLE	average view angle	deg	
ISICE	surface type		
PRESS	surface pressure (NCEP)	mbar	

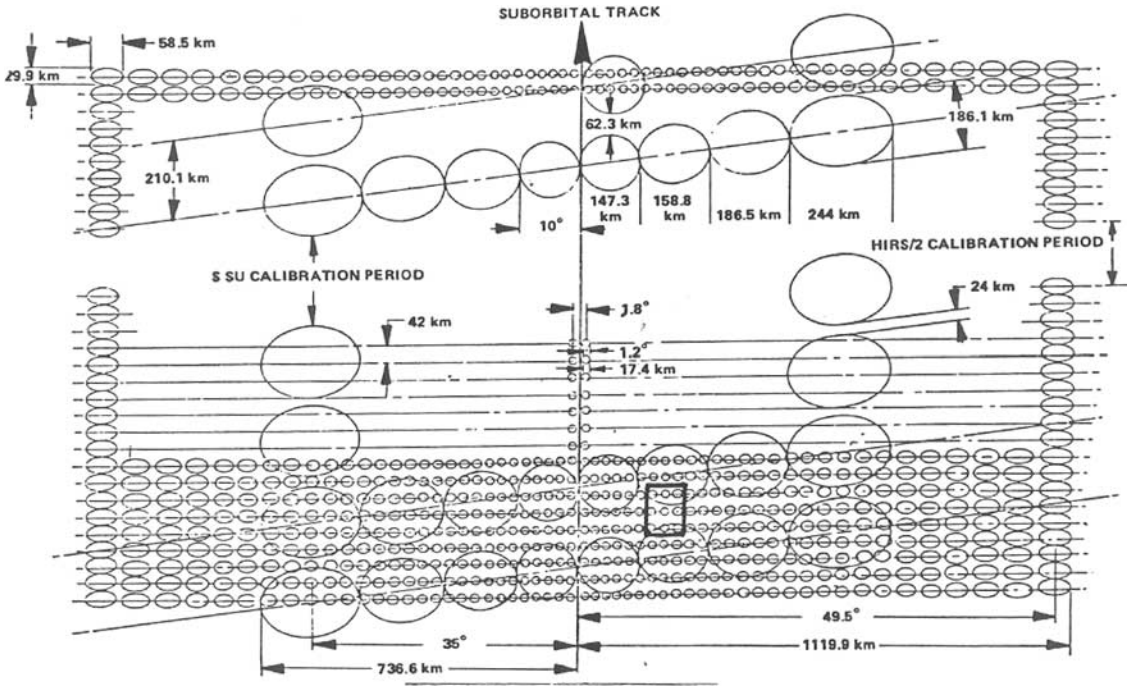


Figure 1. Grouping of MSU and HIRS footprints into 100-km retrieval boxes. From the work of *Chédin et al.* [1985]. Box (heavy lines) represents a sample grouping of HIRS and MSU fields-of-view into a 3I retrieval box.

each retrieval box is associated with a particular air mass, based on observed brightness temperatures in channels that are not affected by clouds (MSU 2–4 and HIRS channels with weighting function peaks above cloud tops).

3.1.1.4. Cloud Tests

[12] A series of cloud tests is performed to determine if individual retrieval boxes are clear, partially cloudy, or overcast. All cloud tests are spectral and involve operations on single or multiple channels in the visible, IR and microwave regions. Cloud tests have been modified since the original implementation of 3I to improve the cloud detection over polar surfaces [*Francis*, 1994]. Our current version employs the cloud tests detailed by *Stubenrauch et al.* [1999a] and *Scott et al.* [1999].

3.1.1.5. Temperature Profile

[13] Prior to temperature retrieval HIRS channels contaminated with clouds are ''cleared.'' Cloud clearing is based on regressions using combinations of microwave ''sister'' channels. The channels have similar weighting function as the corresponding HIRS channels (signal originates from the same portion of the atmosphere) but are not affected by clouds. Cloud clearing removes the effect of clouds in the HIRS channels and thus simulates radiances as if there were no clouds present. Cloud-cleared radiances are then used to search a library of profiles for a first-guess solution. This library, called the Thermodynamic Initial Guess Retrieval (TIGR) data set, contains 1761 unique temperature profiles selected from a global collection of over 150,000, and is searched to identify the 10 closest profiles using a proximity criterion based on the brightness temperature differences between TIGR profiles and observed TBs for a subset of channels. The average of these profiles is then used in combination with precomputed transmission functions and brightness temperature covariance matrices to retrieve the final temperature profile. The initial guess and covariances provide constraints for the least-squares solution to the underdetermined retrieval problem [cf. *Rodgers*, 1976].

3.1.1.6. Humidity

[14] Following the temperature retrieval, cloud parameters are determined as described below. For partially cloudy boxes (at least one clear HIRS pixel) retrievals are performed for the clear HIRS pixels. For overcast retrieval boxes, (all HIRS pixels cloudy) retrieved cloud parameters are used to ''clear'' channels sensitive to changes in atmospheric water vapor. Only retrieval boxes with effective cloud fractions of less than 90% are used in the subsequent retrieval of surface temperature and humidity. The procedure for the retrieval of humidity is analogous to the retrieval of temperature (least-squares solution constrained by a first guess). In the latest version of the ''native'' 3I system, the least-squares solution for water vapor retrieval was replaced with a neural network-based (NN) retrieval system [*Chaboureau et al.*, 1998]. Following some comparisons that showed that the NN solution had substantial biases in the polar regions, we retained the original retrieval method for generating Path-P products.

3.1.1.7. Surface Temperature

[15] Surface temperature is retrieved using a variety of methods depending on whether the retrieval box is classified as clear, partly cloudy, or overcast, and whether it is located over land, open ocean, or sea ice. For clear and partly cloudy retrieval boxes, a combination of HIRS window channels (channels 8 (11.1 μm), 18 (4.00 μm) and 19 (3.7 μm)) is used for the clear HIRS pixels within the box. Completely overcast boxes (all HIRS pixels have some cloud) are ''cloud-cleared,'' and a simultaneous retrieval of surface temperature and humidity is performed using the cloud-cleared

radiances. This method is only applied in boxes where the effective cloud fraction is less than 90%. The method based on MSU-2 correlations described by *Francis* [1994] is no longer used.

3.1.1.8. Cloud Fraction and Cloud Height

[16] The Path-P data set includes three variables describing cloud conditions. HIRS_CLDY is simply the fraction of HIRS pixels within a retrieval box that have been labeled cloudy through the application of 9 cloud tests. This variable is equivalent to the cloud fraction commonly retrieved from satellite imagers (e.g., AVHRR). Because of the relatively low resolution of TOVS sensors, another variable is commonly retrieved in the atmospheric sounding community—the ''effective cloud fraction'' (Path-P variable FCLD, but commonly abbreviated as $N\varepsilon$) is defined as:

$$N\varepsilon(p_k, \nu_i) = \frac{I_m(\nu_i, \theta) - I_{clr}(\nu_i, \theta)}{I_{cld}(p_k, \nu_i, \theta) - I_{clr}(\nu_i, \theta)} \quad (1)$$

where N is the cloud fraction, ε the cloud emissivity, and θ the viewing zenith angle. I_m is the measured radiance in channel i, I_{clr} the corresponding retrieved clear sky radiance and I_{cld} is a modeled radiance for a blackbody cloud with $\varepsilon = 1$ located at pressure level p_k. $N\varepsilon$ cannot be separated in this formulation. This method is a minimum residual or radiance difference technique because $N\varepsilon$ and cloud-top height p_k are found simultaneously by minimizing the difference between observed and calculated radiances in a number of channels. A weighted χ^2 technique which takes into account observed variances at different pressure levels and channels [*Stubenrauch et al.*, 1999b; *Eyre and Menzel*, 1989] is used to solve for $N\varepsilon$ and p_k. The weighted χ^2 removes the high biases in cloud-top height retrieval arising from the smaller differences between radiances from clear and cloudy situations in low-peaking HIRS channels. The technique appears particularly applicable in the Arctic where low clouds predominate.

3.1.2. Calibration Corrections (DELTAs)

[17] TOVS level 1b data are calibrated using the standard calibration procedure described in the NOAA Polar Orbiter Users Guides [*Kidwell*, 1998]. In addition, empirical corrections were applied to the calibrated brightness temperatures. These corrections (also known as DELTAs) are calculated from the differences between observed and modeled brightness temperatures for a large set of global radiosonde/TOVS match-ups. They are meant to reduce the effects of sensor drift and orbital decay, as well as to correct biases in the forward radiative transfer model used to generate the TIGR database. DELTAs used in Path-P were provided by the original algorithm developers at LMD, as these calculations require a large number of cloud-free radiosonde satellite match-ups. For this study DELTAs from NOAA-11 were applied to NOAA-14 because NOAA-14 DELTAs are not yet available. Any remaining calibration errors are contained in the uncertainties reported in our validation statistics. To avoid spatial discontinuities caused by satellite intercalibration differences (see below), data from only one satellite are used to generate Path-P products at any time (i.e., if two satellites are active, only one is used). Owing to orbit convergence at the poles, near complete coverage can be obtained using the 14 orbits by one satellite. Data from the NOAA-14 satellite were used for the SHEBA period.

3.1.3. Matching Surface Observations and Satellite Data

[18] The Path-P products used for validation are spatial and temporal averages of level 2 (geophysical data in orbital format) retrievals. Spatial averages are computed from all retrieval boxes whose centers are within a radius of 50 km from the SHEBA site These spatial averages (varying numbers per day) are then averaged over 24 hours to compute daily averages. The analysis was performed using level 2 data to minimize geolocation errors. However, because time and spatial averaging scales correspond to those of the gridded products, the uncertainties are generally applicable to the gridded Path-P products.

3.2. AVHRR Data

[19] The Advanced Very High Resolution Radiometer (AVHRR) data set is a product of the AVHRR Polar Pathfinder project [*Maslanik et al.*, 1997]. Global Area Coverage (GAC) data acquired from overpasses nearest 1400 local solar time (2400 UTC) were regridded to a 5-km pixel size. Cloud temperatures are computed for a 25 × 25 km area surrounding the SHEBA site.

[20] The cloud-top temperature [Tc-CASPR] is determined from the 11 μm (channel 4) brightness temperature, the clear-sky brightness temperature (not corrected for the atmosphere), and the visible cloud optical depth. The infrared optical depth is determined from the visible optical depth using a parameterization of cloud optical properties [*Key*, 2000]. If the infrared slant-path cloud optical depth is less than some threshold, the brightness temperature is assumed to be a function of both the cloud temperature and the upwelling radiation from the surface and atmosphere below the cloud. The threshold is chosen such that if a cloud's optical depth exceeds it, the transmittance is less than 1% (equivalent to an optical depth of 4.6) and the cloud is considered opaque. For opaque clouds the cloud-top temperature is simply the 11 μm brightness temperature. If the cloud is not opaque, the cloud temperature is determined by first computing the cloud transmittance t from the infrared optical depth, then calculating the cloud radiance that would be required to produce the observed radiance of the cloudy pixel given the cloud optical thickness and the observed clear-sky radiance (as an approximation of the surface radiance under the cloud):

$$L_c = \frac{L_4 - tL_{clear}}{1 - t} \quad (2)$$

where L_c is the cloud radiance, L_4 is the channel 4 radiance, and L_{clear} is the clear-sky radiance. If the estimated surface temperature is too low and the cloud is thin, then the cloud-top temperature will be too high and vice versa. If the adjusted cloud-top temperature is either lower than the tropopause temperature or higher than the maximum temperature from 950 mb to the tropopause, then it is reset to either the minimum or maximum temperature, as appropriate.

3.3. SHEBA Surface Data

3.3.1. Standard Meteorological Observations

3.3.1.1. Cloud Fraction

[21] Cloud fraction was estimated at the SHEBA field site by meteorological observers [C-METOBS]. Observations were made following WMO standards, which call for

reports in eighths of sky cover at 6-hour intervals near 0, 6, 12, and 18 UTC, unless observers were preoccupied by other tasks. Daily averages were computed from this data set.

3.3.1.2. Surface Temperature

[22] The measurements of 2-m air temperature data from the two SHEBA project office (SPO) 10-m towers [TEMP-2m] were used for comparison with TOVS-derived skin temperature (SKTEMP). A discussion regarding the differences between skin and 2-m temperatures is provided in section 4. The SHEBA data set is an average for the two towers after screening for points apparently contaminated by the influence of the ship.

3.3.1.3. Temperature (GLAS-T) and Humidity Profiles (GLAS-H)

[23] Temperature and humidity profiles were measured with the GPS/LORAN Atmospheric Sounding System (GLAS) during the entire SHEBA experiment. This system, developed by the National Center of Atmospheric Research, is based on a Vaisala RS 80-15 (GH or LH) sonde with a stated accuracy of 0.2 C for temperature and 2% for relative humidity. Actual accuracies for humidity, particularly at the low temperatures encountered in the Arctic, are likely to be lower. Measurements are available approximately twice daily (1115 and 2315 UTC) for the period 11 November 1997 to March 1998 and up to 4 times daily (0515, 1115, 1715, and 2315) for April to August 1998. GLAS radiosonde launches have a high sampling rate of 1 Hz, yielding measurements at levels within 0.5 mb of the standard Path-P levels. Because of the high sampling rate, we linearly interpolated radiosonde temperatures to the corresponding Path-P temperature level without incurring any significant error. Daily averages are then computed for those levels. Relative humidity profiles are also interpolated linearly to the Path-P layer boundaries, converted to specific humidity, and integrated over the Path-P layers to obtain layer-average precipitable water values.

3.3.2. Cloud Information From Lidar/Radar

3.3.2.1. The Depolarization and Backscatter Unattended Lidar (DABUL)

[24] The lidar system that was used to obtain the SHEBA cloud measurements is the Depolarization and Backscatter Unattended Lidar, which was developed at NOAA's Environmental Technology Laboratory. DABUL is an active remote sensing system that transmits pulses of laser light into the atmosphere. The energy scattered back to the system yields range-resolved information on the horizontal and vertical structure of clouds and aerosols, as well as the phase of clouds and precipitation throughout the troposphere. The DABUL range resolution is 30 m and real-time averaging of 5 s was employed during SHEBA to reduce the data quantity. For tropospheric cloud and aerosol information, the 5 s averaging times were more than sufficient for high signal-to-noise ratios. The lidar was located on the helicopter deck of the *CCGC Des Groselliers*. DABUL was operational from 1 November 1997 to 8 August 1998. For additional system information and optical engineering detail please refer to the studies of *Grund and Sandberg* [1996] and *Alvarez et al.* [1998].

[25] Measurements from the DABUL system are processed to extract information about Arctic cloud heights and phases. After the necessary system corrections (i.e., background, range, and overlap) are applied to the data, the intensity and/or depolarization fields are thresholded to determine cloud base and top heights for all cloud layers. Once the cloud boundaries are determined, the signal-weighted average values of the depolarization ratio for each layer are assigned. These ratios give an indication of the phase of each cloud layer. The cloud data are then averaged to create 10-min interval time series for the entire SHEBA data set. For additional detail on the determination of cloud statistics please refer to the work of *Intrieri et al.* [2002].

[26] For this study we translate the lidar cloud heights into a time series of a binary indicator for clear (0) versus cloudy (1) skies with a time resolution of 10 minutes. Daily cloud fractions are derived by computing time averages of the clear/cloudy flag. When comparing a cloud fraction derived from a temporal average (C-LIDAR) to one based on a spatial measurement (C-METOBS) or from satellite (HIRS_CLDY) one must assume that the spatial variability over an area translates to the temporal variability at a given point. In other words, one assumes there is a time-averaging scale that corresponds to a particular spatial scale that depends on the velocity of the clouds.

3.3.2.2. Cloud Radar

[27] NOAA/ETL also designed, built, and operated an 8 mm (35 GHz) Doppler cloud radar (MMCR, Millimeter Cloud Radar) at SHEBA. This radar was specifically designed to provide continuous observations of reflectivity, Doppler velocity, and Doppler spectral width of cloud hydrometeors for most cloud types. This system has the added feature of being sensitive to weakly precipitating clouds, unlike conventional weather radars. This radar is identical to those presently operating at the Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) program Cloud and Radiation Testbed (CART) facilities [*Stokes and Schwartz*, 1989], which are located at Lamont, Oklahoma; Barrow, Alaska; and on Manus Island near Papua New Guinea.

[28] Because the millimeter cloud radar was configured for obtaining long-term, continuous measurements in remote locations with minimum operator interface, it was designed to be vertically-pointing. It utilizes a high-gain antenna and a low peak-power transmitter with high duty-cycles for greater reliability. The MMCR cycles through 4 modes that use different combinations of pulse coding and range resolution. The modes are combined into a single product in post processing, which allows detection of clouds over a wide range of cloud types. The radar was housed in the DOE/ARM sea container, which was located near the lidar system on the helicopter deck of the *Des Groselliers*. Technical details concerning the MMCR can be found in *Moran et al.*'s [1997] study. Because lidar returns saturate in the presence of optically thick water clouds, cloud-top heights from the cloud radar and the DABUL lidar are combined by simply selecting the greater cloud-top height from either system.

3.3.3. Timescale Wavelet Analysis

[29] The timescales of the correlations between the surface-based estimates and the satellite-based estimates are investigated with the discrete wavelet transform, a technique well adapted to nonstationary time series and to compositing results from different stations or years. We used the maximum overlap wavelet decomposition [*Percival and Guttorp*, 1994; *Lindsay et al.*, 1996] to obtain the wavelet coefficients of each time series using the simplest wavelet

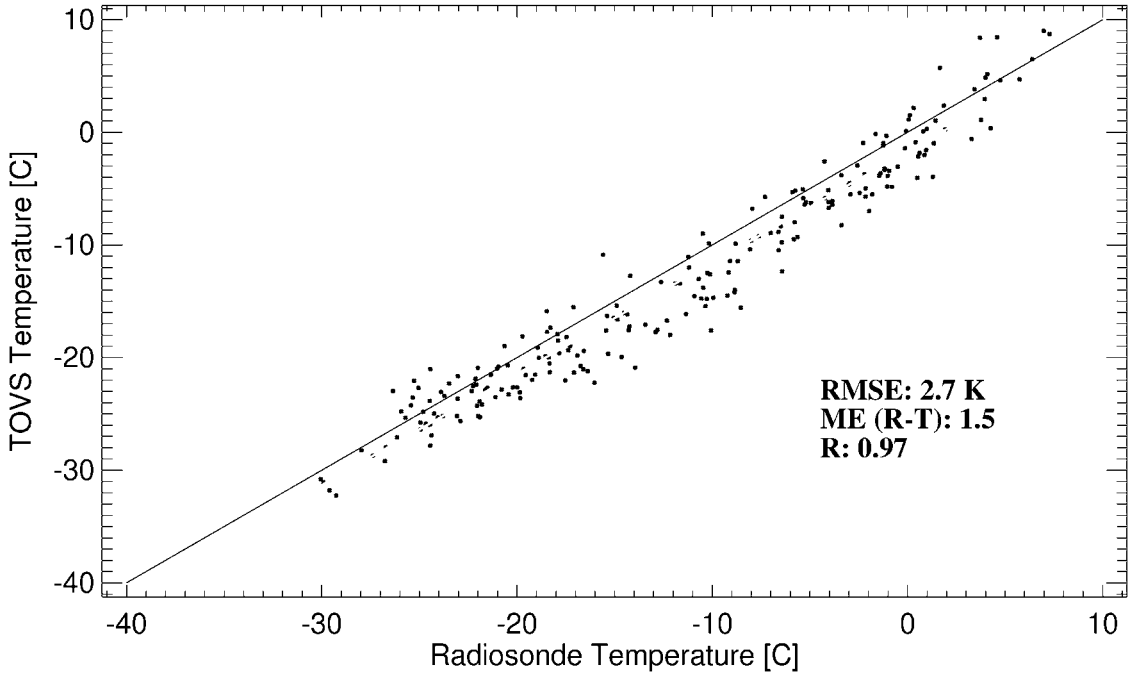


Figure 2. Comparison of retrieved Path-P temperature at the 850-mb level with radiosonde measurements during SHEBA. Lines of 1:1 correspondence and least-square fit are shown. Root Mean Square error: 2.7 K, Mean Error (Radiosonde-T): 1.5, and correlation coefficient: 0.97.

filter, the Haar. With this technique, the correlation at a particular scale is equivalent to finding the correlation between two band-pass-filtered time series obtained with a sequence of running-mean filters. Missing values were filled with linear interpolation. The wavelet coefficients are easily used to obtain the scale-dependent correlations of subsets of the data [cf. *Lindsay et al.*, 1996]. It is important to note the difference between statistics computed using *x*-day averages and those at a *x*-day timescale computed using the wavelet-based technique: The band pass operation that is at the root of the scale analysis effectively eliminates the high frequency, day-to-day, noise or errors in the time series (a low-pass filter) and isolates the intermediate-scale variability from the slow variations associated with the annual cycle (high-pass filter). Timescale analysis provides information on the variability at a specific timescale. Variability at shorter and longer timescales has been filtered out. In contrast, averaging is equivalent to a low-pass so that variations at longer timescales are still present in the data.

[30] We use timescale analysis here as a tool to judge whether a climate process occurring at a known timescale is observable using the data set. For example, a low correlation

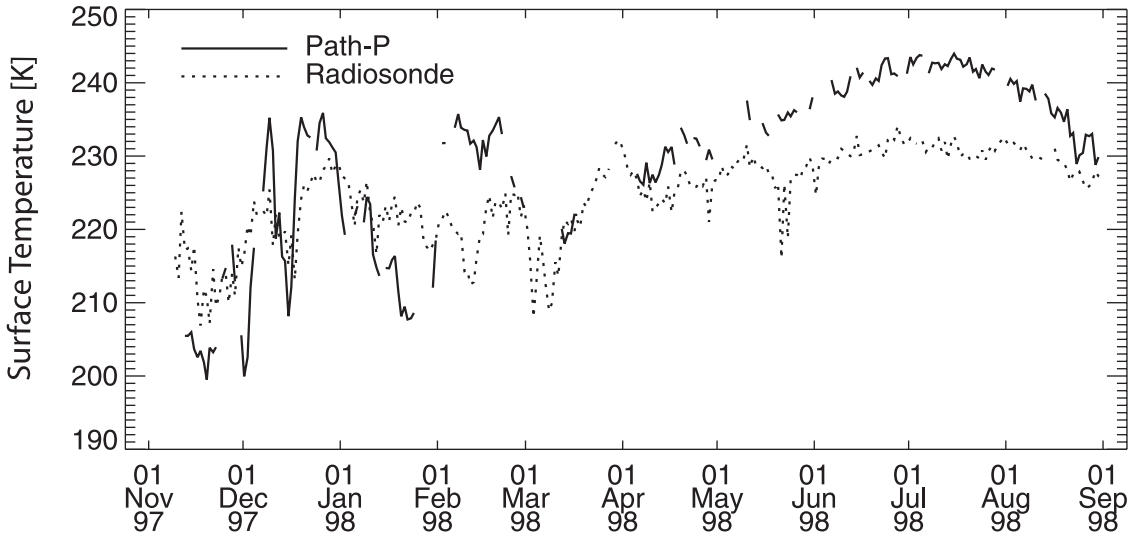


Figure 3. Time series of retrieved Path-P temperature (solid) at the 850-mb level and radiosonde measurements (dash) during the SHEBA field experiment.

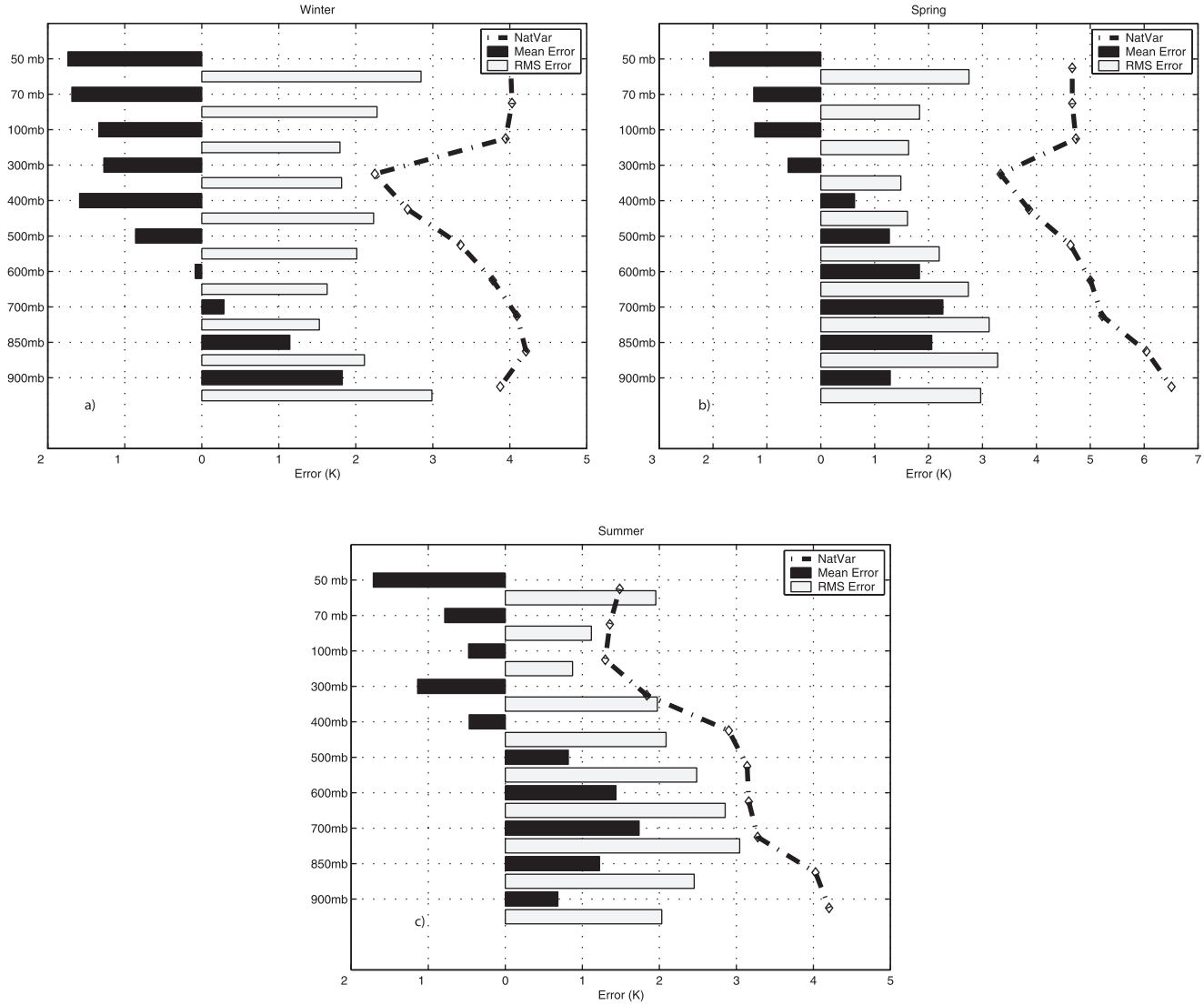


Figure 4. Comparison of uncertainties in TOVS Path-P temperature retrievals for SHEBA (a) winter (December, January, and February), (b) spring (March, April, and May), and (c) summer (June, July, and August) with variability (NatVar dash-dot line) during the same season. Too few points are available for the fall to present meaningful statistics.

at the 1-day timescale indicates that variations from one day to the next are not well captured (or observed differently in the comparison data set). High correlation at 3–5-day timescales would suggest that synoptic scale variability can be observed. If correlations are low below the monthly timescale, then one would conclude that only seasonal variability can be observed properly. Longer timescales are usually associated with higher correlations both because noise occurring at shorter timescales is removed from both time series, and there is commonly more variability at the longer timescales (annual cycle).

4. Results

4.1. Temperature Profiles

4.1.1. Uncertainties

[31] Atmospheric temperature profiles from Path-P (TEMP) are compared with radiosonde data (GLAS-T) at the SHEBA site for the entire year. Figures 2 and 3 show examples of comparisons for the 850 mb level temperatures. Uncertainties for daily values are on the order of 2.5 K with a small bias of 1.5 K (Path-P colder than measurements). Surface observations and Path-P retrievals are highly correlated with a correlation coefficient of 0.97. This result is particularly encouraging, as the lowest levels generally exhibit the largest uncertainties (except for near the tropopause). This is illustrated in Figure 4, which displays seasonal biases and RMS errors for each retrieval level. Too few data points are available for the fall—Our SHEBA data set begins in November and ended in August—so while no statistics are presented for this season, they are included in the annual averages. Uncertainties are smallest during spring when they barely exceed 2 K. Spring-time cloudiness is usually lower than in other seasons and is less likely to affect temperature retrievals. Uncertainties are generally greater at 850 mb and at 50 mb near the tropopause. These layers commonly contain temperature inversions. One should recognize that Path-P values represent layer-average values

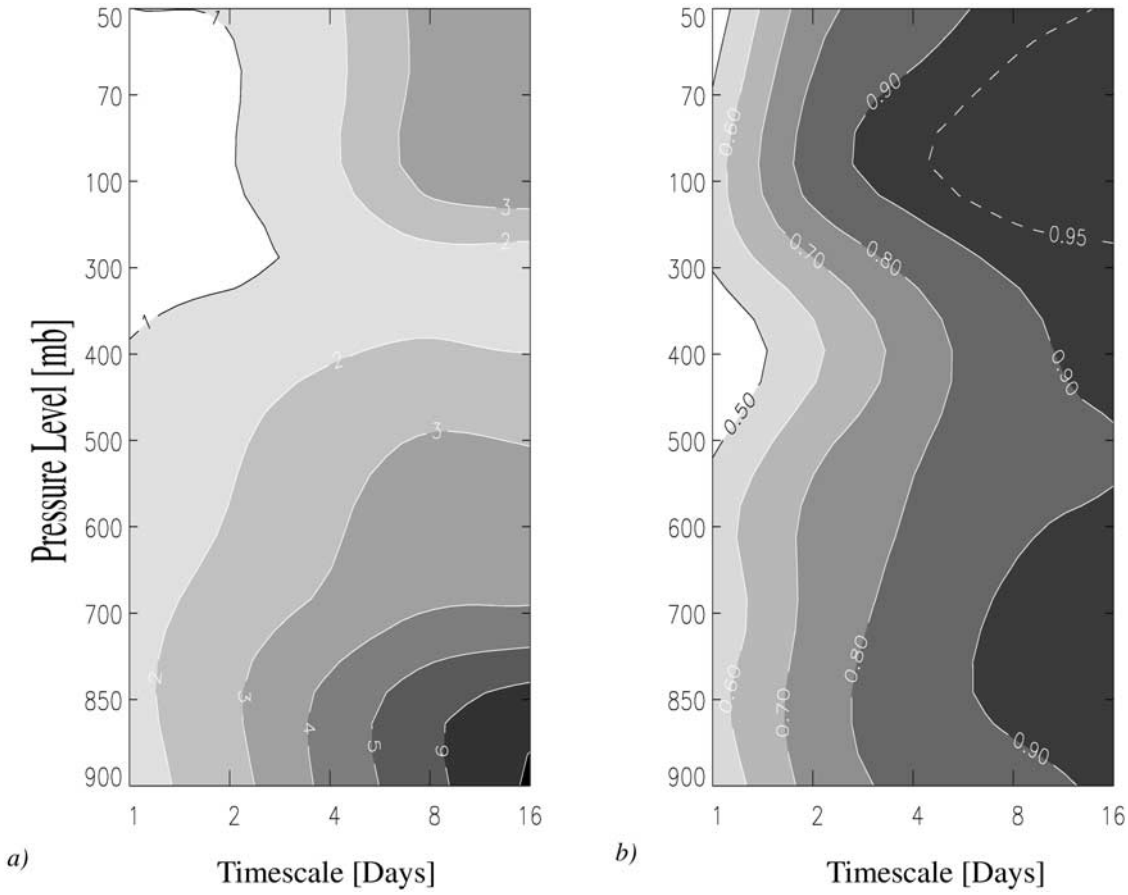


Figure 5. Temperature variance [C^2] of the radiosondes (a) and correlations (b) of TOVS-derived profiles with SHEBA radiosonde observations at different timescales.

interpolated to standard levels for a 100-km area, while the radiosonde values are for a point in vertical and horizontal space. The limited vertical resolution of the satellite retrievals is largely the reason for greater uncertainties in those atmospheric levels. Biases tend to be positive at lower levels and negative at higher levels.

[32] Uncertainties computed for data over the entire annual range maybe of limited value because observation variance has a large contribution from the annual cycle. We therefore ask the question how retrieval errors compare with intraseasonal temperature variability observed in radiosonde data. At all levels and in all seasons uncertainties are smaller than the intraseasonal variability (Figure 4). The relation of the RMS error to the observed natural variability is a measure of how well we can expect to observe variability within a given season. During spring, this relation looks best, as the natural variability far exceeds the retrieval uncertainties. This is largely because the spring season includes the transition from winter to summer and therefore a large part of the annual variability.

4.1.2. Timescales of Correlation

[33] Figure 5 shows the variance (a) of temperature profiles [GLAS-T] and the correlation (b) with Path-P profiles at timescales from 1 to 16 days. Variances are greater at longer timescales as well as near the surface and above the tropopause. Correlations at 1-day and 8-day timescales are on the order of 0.6 and 0.9. This result suggests that day-to-day variations in radiosonde data and Path-P retrievals are somewhat different, while week-to-week variations would be similar in either record. This in part reflects the greater variability at this timescale (greater than measurement error) and the reduction of measurement uncertainty by temporal averaging. Measurement uncertainty in this context includes differences in timing, sensor characteristics (e.g., vertical resolution), as well as spatial discontinuities. As discussed in section 3.3.3, it is important to note the differences in correlation computed using daily averages (see Figure 2) and those computed at a 1-day timescale. Correlations shown in Figure 5 do not include variability at longer timescales (e.g., the annual cycle).

4.2. Layer-Average Precipitable Water

4.2.1. Uncertainties

[34] SHEBA radiosonde data allow us to perform a detailed validation of the layer-average and total-column precipitable water amounts contained in the Path-P data set (Figure 6). Satellite-derived daily averages are highly correlated with radiosonde measurements (r = 0.88) and the RMS error computed from daily measurements over the entire year for total-column precipitable water is 2.2 mm. There is significant scatter in the measurements, however, leading us to question whether signals relating to climate processes would be above the retrieval noise level. To address this issue we compare the retrieval uncertainties during three seasons (winter, spring, summer) and compare them with the variability in the same season as obtained from radio-

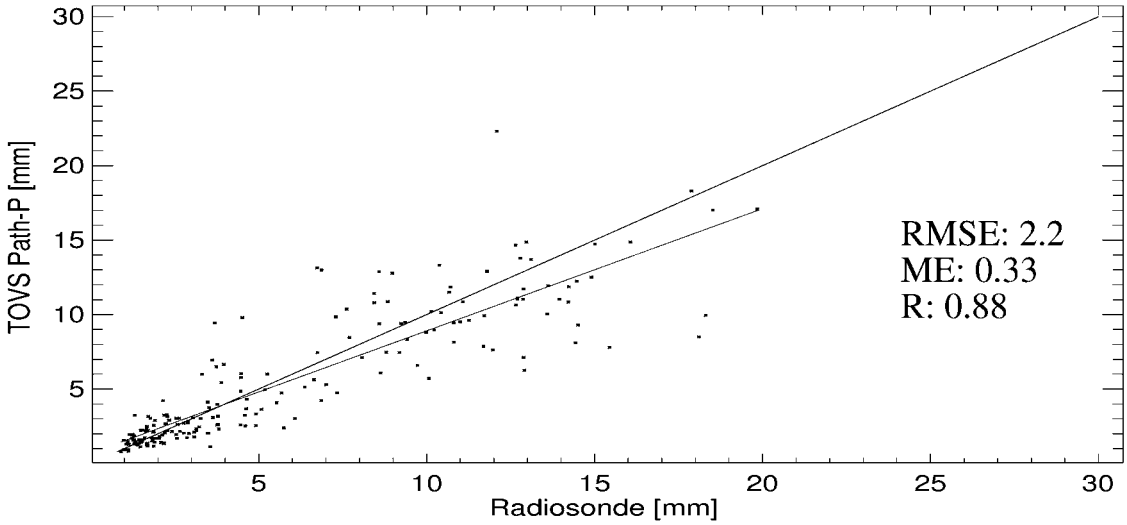


Figure 6. Comparison of daily averaged total-column precipitable water content (mm) from TOVS Path-P and from SHEBA radiosonde observations. Lines of 1:1 correspondence and least-square fit are shown. RMS error: 2.2 K, mean error: 0.33, and correlation coefficient: 0.88.

sondes (Figure 7). Residual errors from a linear fit are shown as well. Residual errors provide a measure of how much RMS errors might be reduced if systematic errors (biases, gains) were removed. To reduce the noise in both time series they are smoothed using a 5-day running mean prior to computing the statistics. In winter, when there is very little moisture in the Arctic atmosphere (<0.7 mm), retrieval uncertainties in most layers are smaller than the observed variability. We thus expect that during winter the variability in the distribution of moisture is captured in the Path-P data. As with temperature, the comparison of retrieval uncertainties to natural variability is most favorable in spring. RMS errors are well below natural variability. In summer, when thick water clouds dominate and complicate the retrieval of precipitable water, uncertainties are slightly smaller than the natural variability for total-column precipitable water. One might conclude from this that the distribution of water vapor in the Arctic atmosphere is not well observed in summer. However, it is important to note that these comparisons are between two very different data sets: point measurements (radiosonde ascents) and spatial/temporal averages (100 km/24 hours for Path-P products). Figure 8 shows a time series of total-column water vapor from Path-P and from radiosondes. Both time series are smoothed using a 5-day running mean. This time series shows that variability in measured moisture is well matched by satellite retrievals during all times, even in summer when retrieval uncertainties do not compare well with measured variability. Further, as shown in Figure 7, biases contribute significantly to the retrieval uncertainties during summer. In fact, residual errors from a linear fit are somewhat smaller than the uncertainties based on RMS errors. Removal of biases and gains would reduce the retrieval uncertainties, but this would require a larger data set for computing those statistics with confidence.

4.2.2. Timescales of Correlation

[35] Figure 9 presents precipitable water [GLAS-H] variance and correlations with Path-P retrievals [WVAPOR] at timescales of 1–16 days. Near the surface (850 mb-SRF) variances are similar at all timescales with a peak near 8 days. In higher layers, variances decrease with timescale. In the lowest layer the correlations between TOVS and radiosondes profiles are low for timescales less than 8 days, but the two data sets exhibit higher correlations at short timescales in upper layers. This result suggests that at short timescales the TOVS instrument and radiosonde balloons observe different phenomena at short timescales, but at longer timescales (>8 days) they capture the same atmospheric characteristics. A likely explanation is that both TOVS and radiosonde observations are rather noisy because of the low Arctic temperatures and the associated low moisture content, which yields lower correlations at short timescales. At longer timescales retrieval errors with inherent shorter timescales—such as differences related to location, timing, and fields-of-view—are less prevalent, leading to higher correlations.

4.3. Surface Temperature

[36] The Path-P products include an estimate of surface skin temperature (SKTEMP). Because a variety of methods is used to retrieve surface temperature, depending on cloud amount and surface type in the retrieval box, an unambiguous determination of whether the surface skin temperature represents a "clear sky" or an "all sky" value is difficult to make. In the Arctic, surface temperature can vary by several degrees depending on whether the area is clear or cloud-covered. The situation is further complicated by the fact that the response of surface temperature to a change in cloud cover varies depending on other conditions, such as wind speed. Because the surface warms more slowly after a cloud moves overhead than it cools after a cloud departs [*Walsh and Chapman*, 1998], the time during which a "hole in the clouds" is seen by a satellite remote sensing system will also influence the retrieved temperature. This time is likely to be relatively short for any clear pixel within a 100-km retrieval box as long as clouds are moving relative to the surface. Consequently, we do not expect a significant bias by inferring that SKTEMP represents an all-sky value. Surface

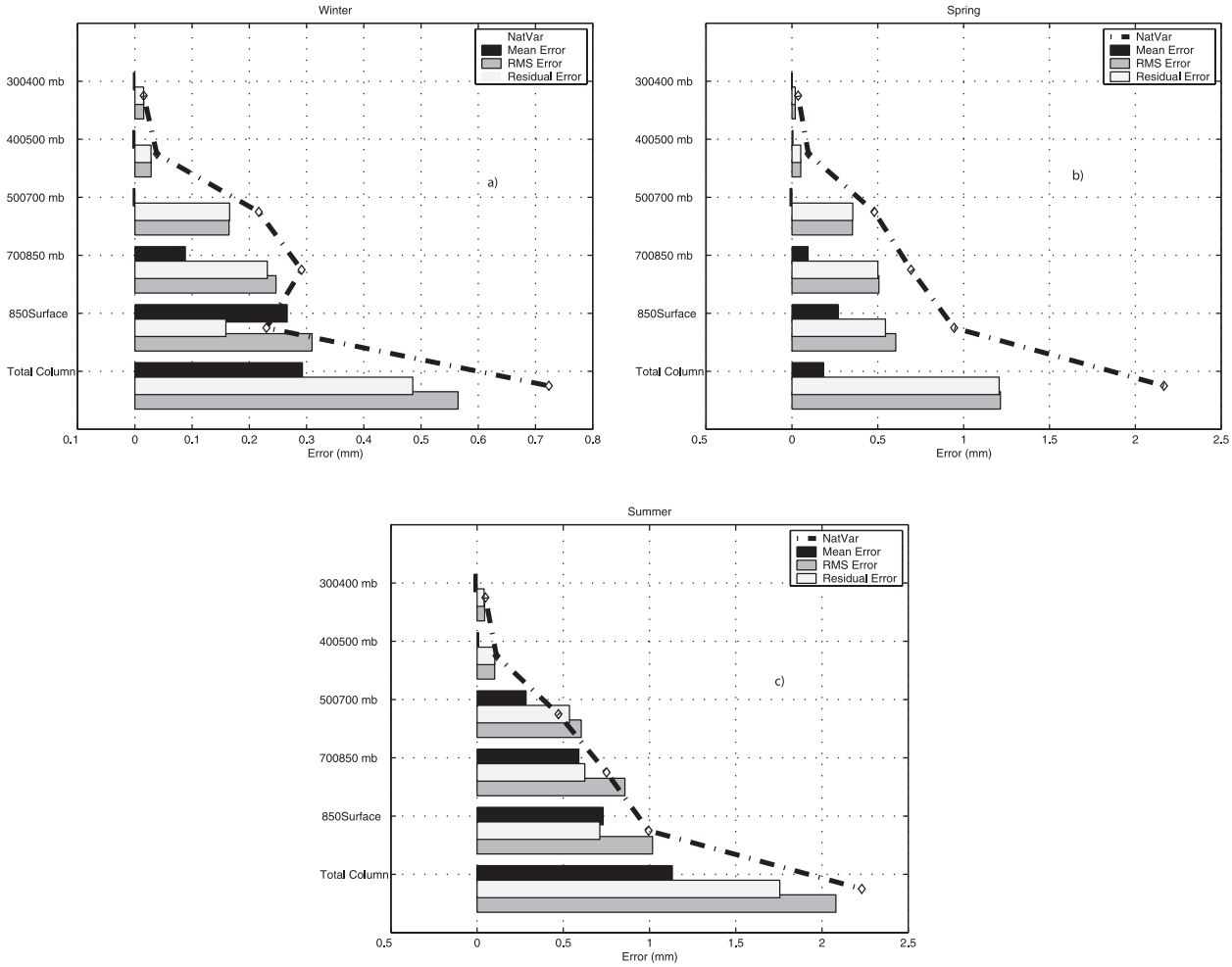


Figure 7. Comparison of uncertainties in TOVS Path-P precipitable water retrievals for (a) winter (December, January, and February), (b) spring (March, April, and May), and (c) summer (June, July, and August) with observed natural variability (NatVar) in the same season. Compared are mean errors, RMS errors, and residual errors from a linear fit, with the standard deviation of the radiosonde measurements as a measure of natural variability.

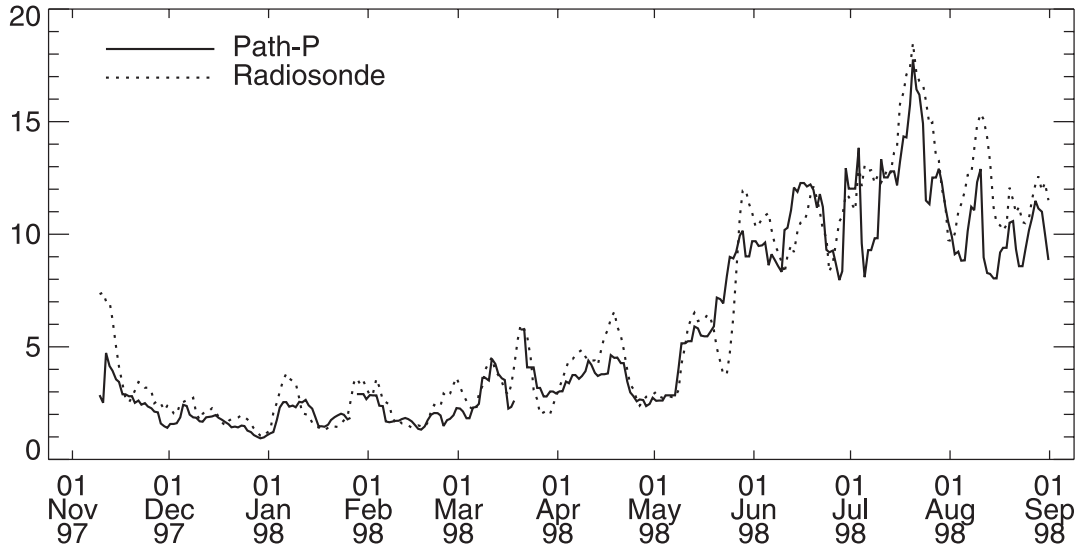


Figure 8. Time series of total-column precipitable water from TOVS Path-P (solid) and radiosonde measurements (dashed) during SHEBA. Both time series are smoothed using a 5-day running mean.

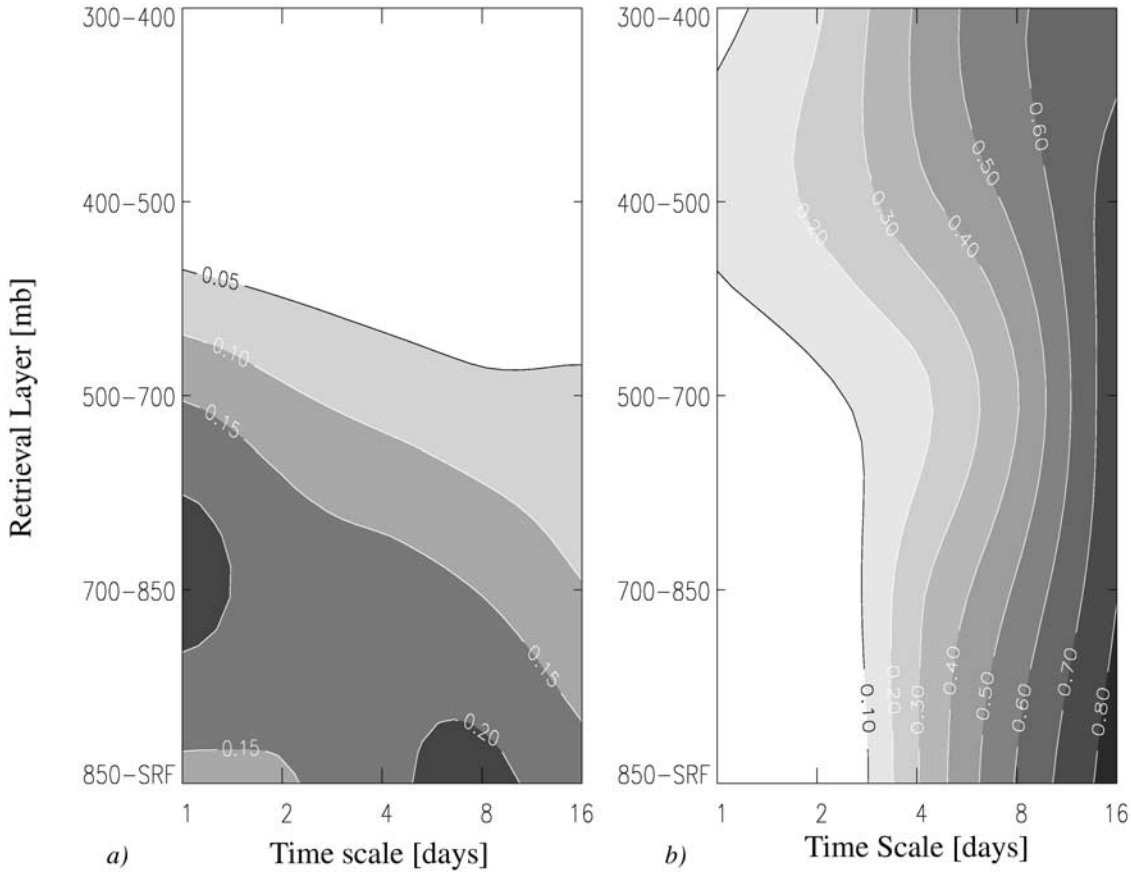


Figure 9. Precipitable water variance in [mm^2] (a) and correlations (b) of TOVS soundings with SHEBA radiosonde observations at varying timescales.

temperature is not retrieved by the 3I algorithm if the cloud cover exceeds 90%, however, thus we expect a slight clear-sky bias in the Path-P skin temperatures.

[37] Skin temperatures were not routinely measured at the SHEBA site. Instead we compare Path-P retrievals to air temperatures 2 m over the surface (Figure 10). Previous Arctic field programs in which both skin and 2-m air temperatures were measured show that the difference between these quantities is usually less than 1 K (colder surface) except during calm, clear days during winter, in which case the difference may be as large as 3 K [*CEAREX Drift Group*, 1990]. Our comparison with measured 2-m temperatures suggest that neither of these effects, which would yield a low bias, is apparent in the Path-P SKTEMP variable. In fact, Path-P SKTEMP appears to be a rather good proxy for ''all sky'' 2-m surface temperatures. Path-P values correspond closely to the measured 2-m temperature (RMSE = 3 K, r = 0.97). The lowest temperatures are overestimated slightly. This is contrary to our expectation of a ''clear-sky bias'' or ''skin-temperature'' bias. It is likely caused by low warm clouds missed by the cloud detection algorithm, which may actually overcompensate the ''clear-sky'' and ''skin'' biases. For higher temperatures (>−20 K) Path-P appears to underestimate surface temperature. This may in part be caused by limiting retrieved temperatures over ice-covered pixels to not exceed the melting point of ice. This attempt to keep retrievals ''physically plausible'' may introduce the small observed bias.

4.4. Cloud Fraction

[38] Cloud detection over Arctic sea ice has been problematic because of difficulties in defining unique spectral signatures for clouds and surfaces with existing satellite sensors. Earlier studies [e.g., *Schweiger and Key*, 1992] showed that Arctic cloud fractions contained in a widely used satellite-derived data set (International Satellite Cloud Climatology Project (ISCCP) C-series) exhibited characteristics that were significantly different from surface measurements, particularly in that the annual cycle was not captured. In a more recent study [*Schweiger et al.*, 1999], we compared cloud fractions from an updated ISCCP data set (D-Series) and from the TOVS Path-P data set with conventional surface-observed cloud fractions from the Russian North Pole (NP) meteorological stations adrift on the Arctic sea ice. We found that cloud fractions from the Path-P data set appear to capture the annual cycle of cloudiness, while the ISCCP data set still does not appear to capture the cycle well. Further, variability of cloud fraction in Path-P versus that from surface observations appears to be consistent at timescales of about 4 days, particularly during the summer. These results are encouraging, yet several significant questions remain:

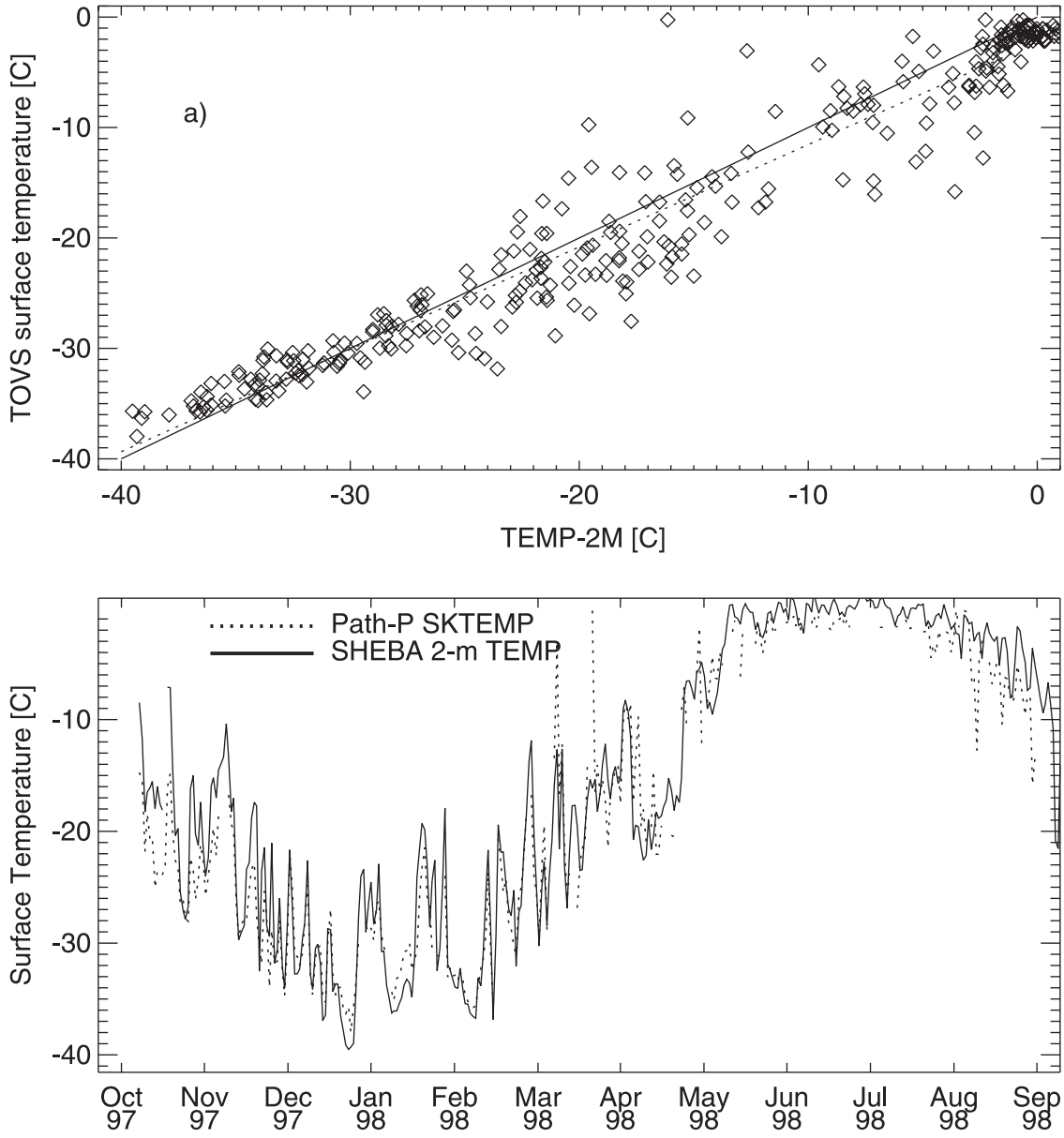


Figure 10. Daily averaged surface skin temperature (SKTEMP) from TOVS Path-P compared with 2-m air temperature from the SHEBA 10-m tower (TEMP-2m). The same data are shown as a scatterplot (a) and time series (b). RMS error is 3 C, correlation coefficient is 0.97. Lines of 1:1 correspondence and least-square fit are shown.

1. How reliable are "ground truth" data based on conventional cloud fractions estimated by human observers? Particularly, how does the absence of light during the Arctic winter affect these observations [cf. *Hahn et al.*, 1995]?

2. How does the presence of clear-air ice particle precipitation (ICP) (so-called diamond dust) affect cloud statistics obtain computed from satellite data? Can satellites detect it and are previously found discrepancies in winter time cloud fraction and in surface and satellite observations [*Schweiger and Key*, 1992; *Wilson et al.*, 1993] attributable to the presence of ICP?

3. Why do surface and satellite data correlate poorly at timescales of less than 4 days? Is this related to the limited spatial resolution of the satellite sensors compared to surface observations, or is it simply due to the level of noise in the retrieval system?

[39] Measurements made during the SHEBA period offer several avenues for attacking these problems. Standard meteorological observations of clouds can be paired with objective information obtained from a sophisticated cloud radar and lidar system that was deployed on the deck of the *Des Groselliers*. In addition several groups are working toward the validation and improvement of their satellite-based algorithms and are producing times series of their results for intercomparisons of satellite retrievals and surface measurements.

4.4.1. Subjective Versus Objective Cloud Fraction

[40] Cloud fraction estimated by meteorological observers is reported according to the WMO code as the fraction

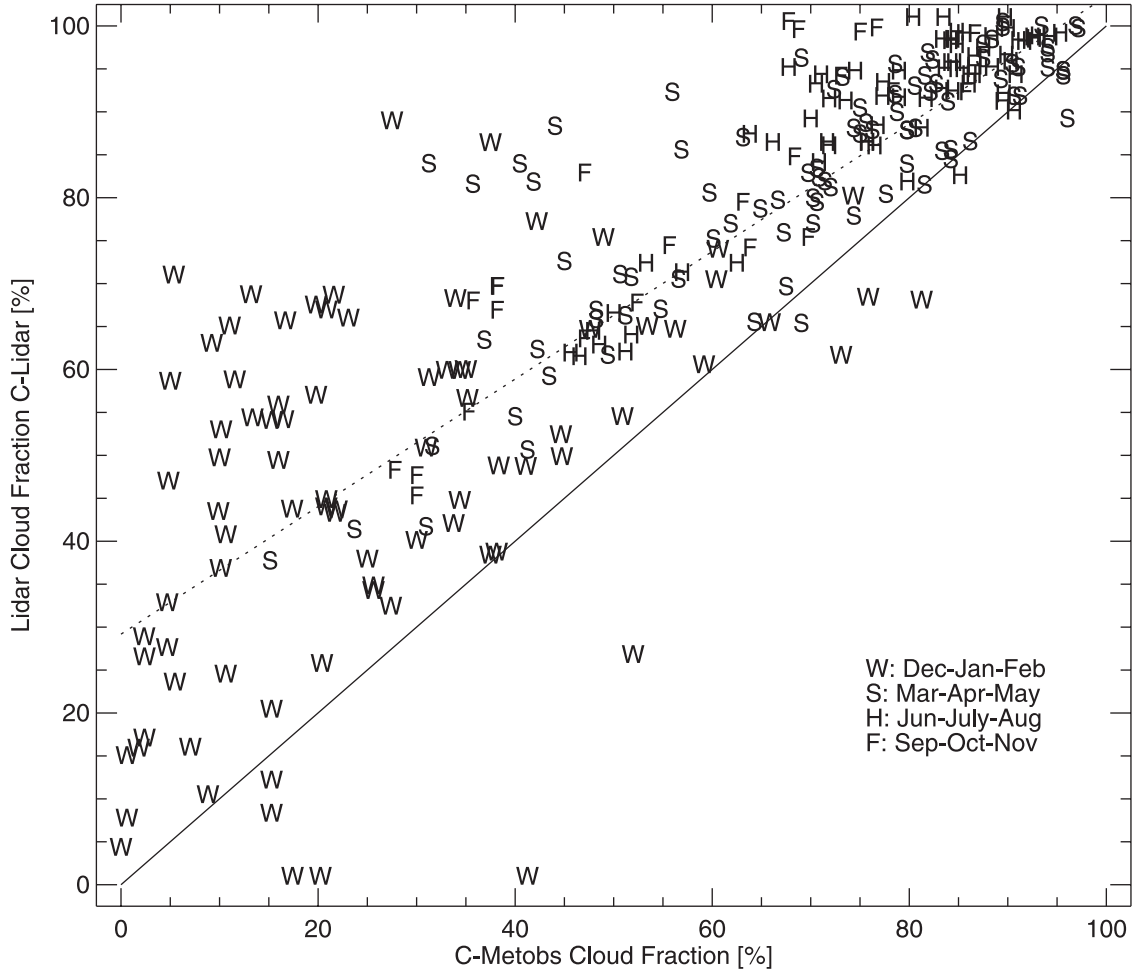
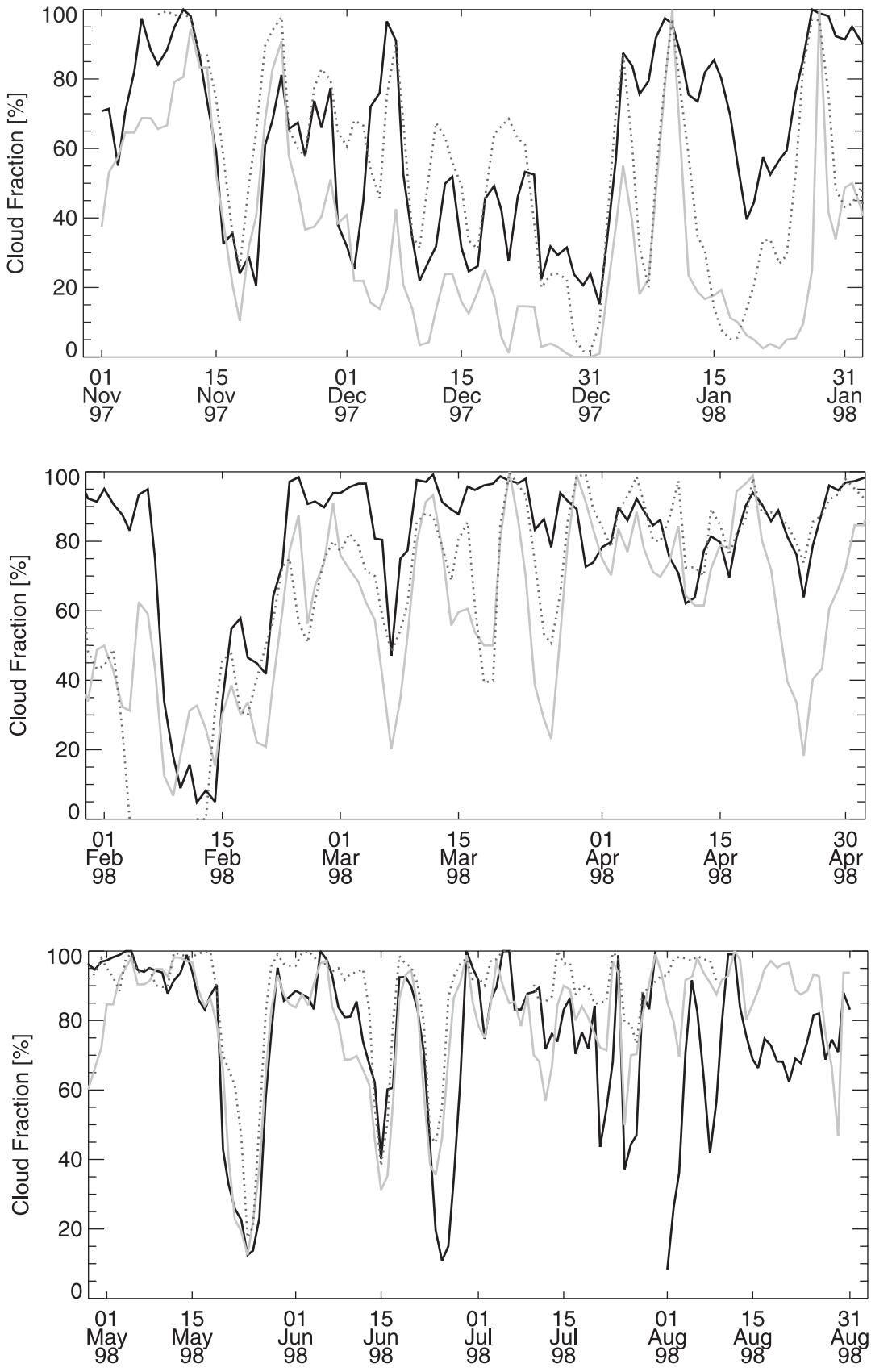


Figure 11. Comparison of cloud fraction from Lidar and from meteorological observations (C-METOBS) for the SHEBA period. Individual values represent 5-day averages. RMS error is 12 [% cloud cover], correlation coefficient is 0.86. Letter symbols W, S, H, and F denote seasons.

(in eighths) of the ''sky dome'' that is cloud-covered. The geographic area corresponding to an observation is dependent on factors such as visibility, cloud height, and cloud type. Lidar backscatter has been processed by the NOAA-ETL team into a time series of binary information (clear or cloudy) with a resolution of 10-minutes. To obtain ''cloud fractions'' that can be compared to values from the surface (C-METOBS) or satellite retrievals, lidar time series are averaged temporally. If we assume a uniform cloud field moving with a speed of 5 m s^{-1} over a stationary lidar system, then a 6-hour average will correspond roughly to a 100-km profile. The corresponding fields-of-view for a surface observer and a satellite sensor also depend on the cloud height. On clear days or days with high clouds, the area for which the surface observations are representative is much larger than it would be for low-cloud conditions. Moreover, satellites and meteorological observers only report a snapshot in time; and because the cloud field is not uniform, comparisons at small timescales are prone to disparity. Our study, consequently, will focus on comparisons at timescales of 1 day and longer. Figure 11 shows a comparison of 5-day averaged cloud fractions from lidar (C-LIDAR) and from C-METOBS. The two data sources are highly correlated (r = 0.86) at this timescale. C-LIDAR cloud fractions are generally higher than C-METOBS, particularly for reports with small cloud fractions, which tend to occur more frequently in winter. *Hahn et al.* [1995] have shown that polar cloud observations from observers are biased by the availability of moonlight. However, adding the their bias (<5%) to SHEBA observations is insufficient to account for the entire difference. A significant part of the difference is likely due to the fact that the Lidar is sensitive to ICP so that events of ICP are contained in the cloud statistics obtained from C-LIDAR. Satellite versus surface observations.

[41] Figure 12 shows a time series comparing cloud fractions (HIRS_CLDY) from the Path-P data set, C-LIDAR, and C-METOBS. Data are smoothed using a 3-day running mean. The three data sets exhibit surprisingly good agree-

Figure 12. (opposite) Time series of cloud fraction from TOVS Path-P (dark solid), the ETL cloud lidar (dotted), and meteorological observations (C-METOBS, light solid) made at the SHEBA camp. Data are smoothed using a 3-day running mean.


Cloud Fraction [%]
0
20
40
60
80
100
01 Nov 97
15 Nov 97
01 Dec 97
15 Dec 97
31 Dec 97
15 Jan 98
31 Jan 98
01 Feb 98
15 Feb 98
01 Mar 98
15 Mar 98
01 Apr 98
15 Apr 98
30 Apr 98
01 May 98
15 May 98
01 Jun 98
15 Jun 98
01 Jul 98
15 Jul 98
01 Aug 98
15 Aug 98
31 Aug 98
Path-P (N14)
Ship Obs: 3.0 day rmean
Lidar Obs: 3.0 day rmean

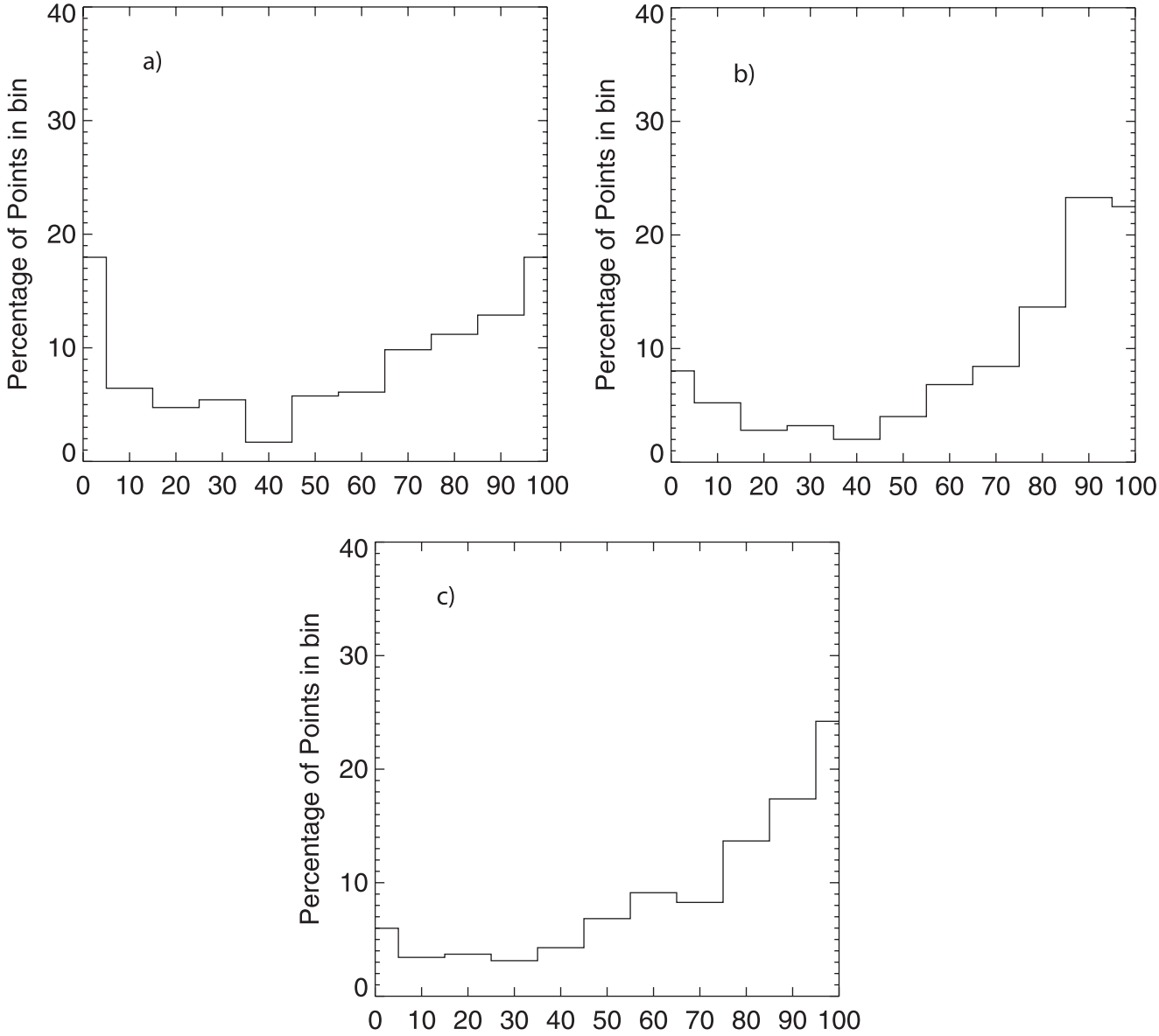


Figure 13. Distribution of cloud fraction during the SHEBA period derived from (a) meteorological observations (C-METOBS), (b) cloud lidar (C-LIDAR), and (c) TOVS Path-P (HIRS_CLDY).

ment given their vastly different perspectives, and the larger variations are captured well. Lower cloud fractions in C-METOBS are noticeable in both winter and spring, but agreement is better in summer. The Path-P cloud fractions during winter are generally closer to the C-LIDAR values than to the C-METOBS, especially when the cloud fraction is small. This suggests that the Path-P cloud detection algorithm properly detects clouds but includes ICP, which is not part of the cloud fraction reported by the surface observer. This observation would support of the explanation proposed by *Wilson et al.* [1993] for the discrepancy between winter time cloudiness in the ISCCP data set and climatologies based on surface observations. However, as shown by *Schweiger et al.* [1999], other factor are clearly responsible for the large differences in winter time cloudiness between the International Satellite Climatology Project (ISCCP) and surface-based cloud climatologies.

[42] Further work will be needed to characterize the specific cloud types and conditions that are responsible for differences between the satellite and surface cloud fractions, which will enable the data sets to be combined for generating long-term statistics and to study climate variability.

4.4.2. Principal Characteristics of Cloudiness

[43] One way to evaluate the quality of a data set is to examine if some of the fundamental characteristics of variability are captured. Arctic cloudiness has long been thought to have a primarily "bimodal" distribution: clear skies dominate in winter while completely cloud-covered skies dominate in summer. This distribution is apparent in the annual statistics computed from C-METOBS data during SHEBA (Figure 13a). The corresponding distribution derived from C-LIDAR exhibits a different shape. There are about 10% fewer completely clear cases, and more observations fall

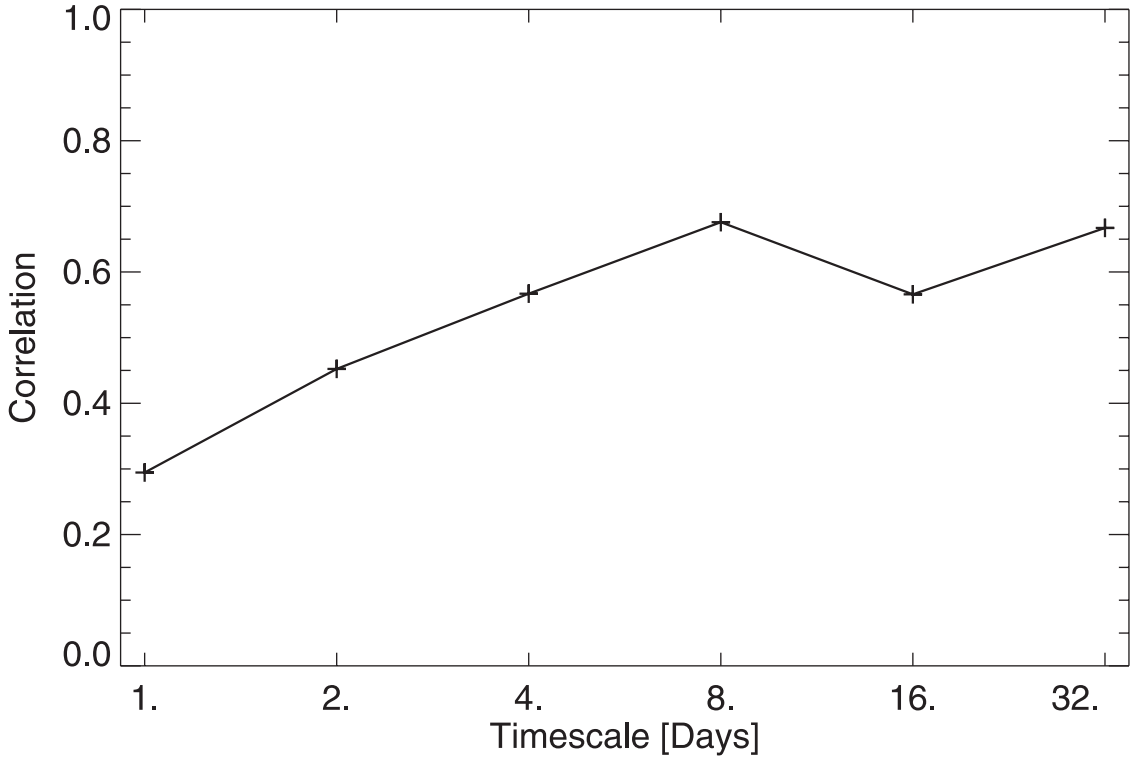


Figure 14. Correlation of cloud fraction from C-METOBS and TOVS Path-P at varying timescales.

into the 90–100% cloud fraction category. The cloud fraction (HIRS_CLDY) distribution from Path-P looks very similar to C-LIDAR. The similarity in the distribution of cloud fraction demonstrates that cloud fractions in the Path-P data properly capture the overall variability.

4.4.3. Timescales of Correlation

[44] We previously found [*Schweiger et al.*, 1999] that Path-P cloud fractions correlate well with surface observations from the NP station data set at timescales greater than 4 days. Here we repeat this analysis using data from SHEBA. Using a wavelet-based band pass filtering method [*Lindsay et al.*, 1996] we compute correlations at varying timescales. Results are very similar to those obtained from the NP station data base (Figure 14) with slightly higher correlations for most of the timescales. Correlations at 2-day timescales are on the order of 0.5 and reach a maximum of 0.7 (~50% variance explained) at the 8-day timescale. RMS differences between C-LIDAR and HIRS_CLDY at timescales greater than 4-days are less than 20%. These results confirm the quality of the observations from the NP station data and our previous results, but again pose the question why the correlations at shorter timescales are relatively poor. There are several possible explanations: Fundamental differences in the observations (e.g., viewed from below and above) and a temporal mismatch contribute to random noise in the comparison. At larger timescales this noise is reduced and a stronger signal becomes apparent. We considered another possibility: The relatively low resolution of the HIRS sensor masks variability at small spatial scales. It is likely that there is a correlation between spatial and temporal scale. For example, variations in cloudiness caused by changes in synoptic conditions (3-day timescale) may be more homogeneous spatially, and therefore more likely to be detected by the low resolution system. To investigate this issue we compared timescale correlations of surface data and cloud retrievals from AVHRR (not shown) which has a field of view of approximately 1-km. The time correlation structure is very similar to the one from TOVS so that resolution does not appear to be a significant factor. Further work is needed to clarify the source of this timescale dependence of the correlations.

4.5. Cloud-Top Temperature

[45] Mean monthly cloud-top temperatures from three sources are compared: CLTEMP from Path-P, CLTEMP-AVHRR from the CASPR algorithm, and values derived from a combination of ETL's surface-based cloud radar/lidar systems. Cloud-top height from the radar/lidar data is converted to cloud-top temperature using SHEBA radiosonde profiles. There is good agreement between the two satellite estimates (Figure 15). Although they use different instruments, resolutions, and algorithms, the satellite retrievals are similar in that they are based on emitted radiances. Retrievals from the radar/lidar systems, in contrast, are based on the interpretation of back scattered radiation. A large difference (over 10 K) between these two cloud-top estimates is notable. The likely explanation for this large difference is the presence of optically thin high clouds. These clouds emit little radiation and are relatively transparent to longwave emission from below (surface or lower clouds), yielding much higher "effective" cloud-top temperatures as seen from space. The active radar/lidar systems are sensitive to the high cloud particles, and consequently retrieve a physical boundary of the upper cloud, which is much higher (colder). Another factor is the possible mislabeling of clear versus cloudy scenes. These differences highlight the need for

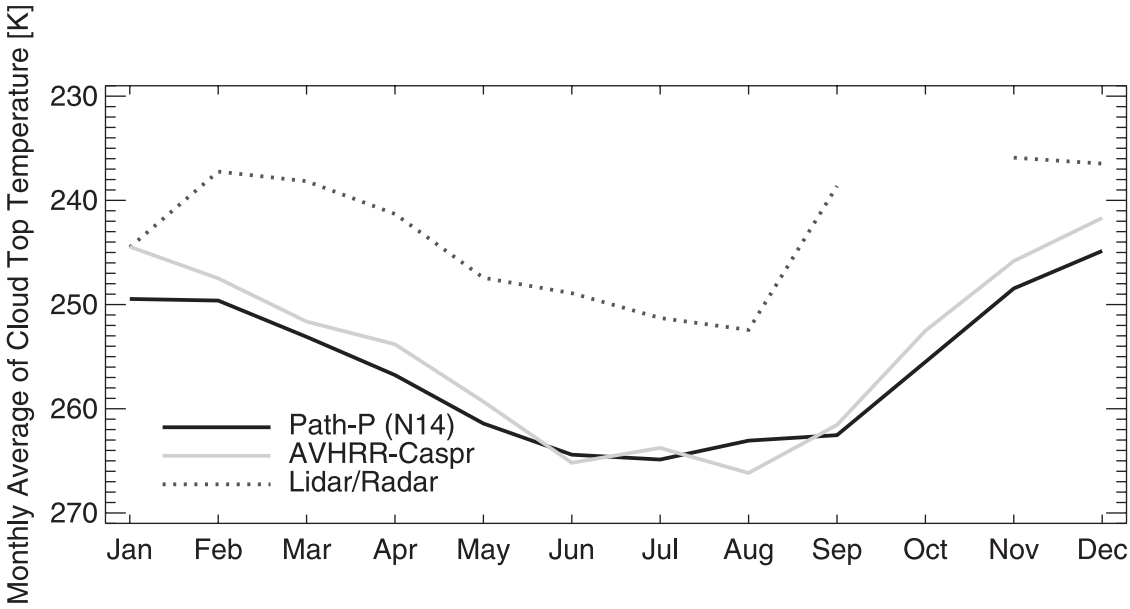


Figure 15. Comparison of cloud-top temperature derived from the combined lidar/radar system (dotted), the TOVS Path-P data set (dark solid), and the AVHRR-CASPR algorithm (light solid).

careful interpretation and application of measurements from differing sources. More detailed comparisons of these measurements, making use of SHEBA surface-based and aircraft-based information of cloud microphysics of the entire atmospheric column, should provide insights on how to best use and combine information from surface-based and satellite sources.

5. Discussion and Conclusions

[46] Validation of satellite-retrieved geophysical products with surface-based observations often involves accounting for and mitigating fundamental inconsistencies between measurements from disparate sources. Errors in a satellite retrieval may result from such a comparison, even though the space-based value may be an accurate representation of conditions in the footprint that includes the surface measurement site. We have attempted to mitigate these problems by analyzing characteristics of variability of the data sources and by averaging values over appropriate time and space scales. However, fundamental differences in the observing systems do contribute to uncertainties revealed in our analysis. We have also attempted to present uncertainties in the context of natural variability and at a range of timescales. We expect that by presenting the uncertainties in this manner, a potential user can determine whether a particular product is appropriate for his or her application.

[47] Using primarily data from the SHEBA experiment spanning an entire annual cycle, we have shown that TOVS Path-P products capture atmospheric and surface quantities in the Beaufort Sea region. We believe that the retrieval quality is comparable in other regions of the central Arctic where surface conditions are similar. Others have also demonstrated [e.g., *Claud et al.*, 1991], the high quality of 3I retrievals over unfrozen, high-latitude ocean, where the surface is more homogeneous, temperature inversions are less frequent, and cloud detection is more reliable than over a snow/ice-covered surface. Further work is required to validate Path-P products over polar land surfaces.

[48] Path-P retrievals of surface skin temperatures at the SHEBA experiment site display high correlation with measurements (r = 0.97). We reason that the response of the snow/ice surface temperature to changing cloud conditions, and attendant surface radiation fluxes, explains the lack of a clear-sky bias. Snow/ice skin temperature responds relatively quickly to increases in downwelling infrared radiation caused by the arrival of a cloud, but it cools more slowly when the cloud departs. The result is that skin temperature under a partially cloudy sky is probably close to that of the temperature under the clouds, as the surface would not have had sufficient time to cool to clear-sky values during breaks in the cloud cover. Our favorable comparison to SHEBA measurements (RMS = 3 K), therefore, gives us confidence that Path-P surface skin temperatures are representative of most atmospheric conditions and should be useful in a wide variety of applications in which high spatial and temporal resolution is not required.

[49] The Path-P retrievals of upper-level temperatures and moisture content offer opportunities to evaluate Arctic atmospheric models and NWP Reanalysis data sets, because the conventional measurements that constitute the primary input data to these models are sparse. Satellite-derived profiles cannot compete with radiosondes in vertical resolution, but the horizontal resolution, particularly in sparsely populated areas, is significantly higher for satellite retrievals. For this reason the Path-P data set may not be well-suited for detailed studies of cloud-radiation interactions or turbulent exchanges over varying ice types, but it may be useful for regional and basin-wide studies. With respect to accuracy, we are particularly encouraged by the comparisons of Path-P values to those from SHEBA radiosondes (RMS 2.5K, r 0.95), as the uncertainties are comparable to those from midlatitude ocean regions where retrieval algorithms are expected to perform well owing to the homoge-

neous background. Satellite-derived soundings over snow-covered and sea-ice-covered regions were, until recently, regarded as unusable owing to challenges introduced by extreme polar conditions, such as frequent temperature inversions, little cloud/surface contrast in temperature or albedo as seen from space, and inadequate sources of first-guess profiles for the retrieval algorithm. We have shown that temperature profile uncertainties are on the order of 3 K and are significantly below the natural variability. How does this retrieval uncertainty compare to a signal one may try to observe? This of course depends on the signal. If we use the intraseasonal variability (standard deviation within a season) as a proxy for signal, we conclude that temperature errors are well below this ''signal.'' Uncertainty in moisture profiles measured against such a ''signal'' compares less favorably. However, this disparity can be attributed to several sources other than retrieval error: fundamental differences in the physical characteristics observed by satellite and surface-based systems, errors in the ''ground truth'' validation data (radiosondes), and temporal/spatial differences in the observations. This is particularly so for water vapor, as radiosonde measurements tend to be noisy at low temperatures. Further, the distribution of water vapor as measured by radiosondes is highly variable with height affecting comparisons to a satellite retrievals with rather low vertical resolution. Spatial variability may similarly influence the comparison with a 100-km resolution system. To reduce measurement uncertainty below a different ''process signal,'' users may have to apply additional spatial or temporal averaging to the data. The removal of systematic errors through more extensive comparisons with validation data is an obvious approach to remedy the situation, but this would require a large, comprehensive validation data set. We believe that Path-P fields will be useful for characterizing spatial variability and deciphering interrelationships and feedbacks between processes involved in energy exchange between the atmosphere and the surface, and between low and high latitudes. Users should keep in mind, however, that residual errors caused by insufficient calibration and/or intersatellite differences may still exist, which may affect time series analyses of individual variables. Because we perform retrievals from a single satellite at any one time, spatial variability is not affected by intersatellite calibration differences. We therefore believe that the best use of the Path-P data set in the context of analyzing climate trends is to provide information on their manifestation (or lack thereof) in spatial variability.

[50] As we have discussed at length in section 4, comparing measurements of cloud properties (fraction and height) from surface-based observers or instruments to estimates from spaceborne sensors is far from straightforward. Clouds appear fundamentally different depending on one's perspective in space and time, as well as the device used to sense the existence of cloud particles (e.g., human eyes, bulk emission, lidar backscatter). We have shown that values obtained from these sources often disagree, but we have also explained that in many circumstances agreement would not be expected. Some of the disparities we observe, in fact, make perfect sense, as is the case for cloud-top height derived from satellite-measured infrared radiances and that derived from a radar/lidar system. Unless the cloud-top is very sharply defined, widespread, and optically thick, the satellite estimate will always be a lower height. Despite these issues, we have shown encouraging evidence that Path-P satellite retrievals of Arctic clouds properly capture the annual cycle, appear to include the presence of diamond dust in the cloud statistics, and are consistent with lidar/radar retrievals of cloud-top height. We believe the Path-P cloud products will be useful for characterizing spatial variability and untangling relationships between large-scale atmospheric dynamics and thermodynamics that may contribute to long-term change of the Arctic surface. The Path-P data set is the first of its kind to provide this information at a spatial and temporal resolution that matches the needs of modelers as well as investigators of basin-wide and regional short-term climate system behavior.

[51] Timescale analyses of Path-P variables with SHEBA validation data provide insights into what type of variations in the surface record we should expect to find in the satellite record. Processes operating at less than a 3-day timescales are generally less well represented. This is particularly true for the retrieval of clouds and precipitable water. However, differences in satellite and surface data at those timescales may be as much caused by differences in the observing system perspective as by measurement error.

[52] In summary, our validation of Path-P products with surface-based measurements from the SHEBA field program, as well as with radiosonde data from NP stations in other parts of the Arctic Ocean, demonstrates viability of atmospheric sounding over ice-covered oceans. The year-long SHEBA field program, with its nearly comprehensive suite of state-of-the-art measurements for the atmosphere, ice, and ocean, provides new opportunities for validating remote sensing products in the Arctic. While many issues still remain, it is clear that progress has been made, and that further efforts will result in a better understanding of processes that govern the Arctic climate system.

[53] **Acknowledgments.** This work was supported by NASA grants NAG5-4908, NAG5-7847, NAG5-4375, and NAG5-8625 and NSF grant OPP-9701514. We would like to thank the NOAA-Satellite Active Archive data center for providing TOVS data. The SHEBA project office is thanked for providing surface meteorological observations. Mark Ortmeyer was responsible for running the Path-P processing system at the University of Washington and is thanked for his heroic efforts. Andy Roach is thanked for preparing some of the figures. We also thank Dick Moritz and Steve Warren as well as two anonymous reviewers for valuable input.

References

Alvarez, R. J., II, W. L. Eberhard, J. M. Intrieri, S. P. Sandberg, and K. W. Koenig, Cloud backscatter and phase measurements in the Arctic using ETL's DABUL lidar, paper presented at the 4th Int. Symp. on Tropospheric Profiling, AMS, Snowmass, Colo., 20–25 September, 140–144, 1998.

Chaboureau, J. P., A. Chedin, and N. A. Scott, Remote sensing of the vertical-distribution of atmospheric water vapor from the TOVS observations: Method and validation, *J. Geophys. Res.*, *103*, 8743–8752, 1998.

Chedin, A., N. A. Scott, C. Wahiche, and P. Moulinier, The improved initialization inversion method: A high resolution physical method for temperature retrievals from satellites of the TIROS-N series, *J. Clim. Appl. Meteorol.*, *24*, 128–143, 1985.

Claud, C., N. A. Scott, A. Chedin, and J. C. Gascard, Assessment of the accuracy of atmospheric temperature profiles retrieved from TOVS observations by the 3I method in the European Arctic: Application for mesoscale weather analysis, *J. Geophys. Res.*, *96*, 2875–2887, 20 February 1991.

Coordinated Eastern Arctic Experiment (CEAREX) Drift Group, CEAREX drift experiment, *Eos Trans. AGU*, *71*, 1115–1118, 1990.

Eyre, J. R., and W. P. Menzel, Retrieval of cloud parameters from satellite sounder data: A simulation study, *J. Appl. Meteorol.*, *28*(4), 267–275, April 1989.

Francis, J. A., Improvements to TOVS retrievals over sea ice and applications to estimating Arctic energy fluxes, *J. Geophys. Res.*, *99*, 10,395–10,408, 1994.

Francis, J. A., A method to derive downwelling longwave fluxes at the Arctic surface from TOVS data, *J. Geophys. Res.*, *102*, 1795–1806, 1997.

Gates, W. L., et al., An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I), *Bull. Am. Meteorol. Soc.*, *80*, 29–55, 1999.

Grund, C. J., and S. P. Sandberg, Depolarization and backscatter lidar for unattended operation, paper presented at the *18th Int. Laser Radar Conf., OSA, Berlin, Germany, 22–26 July,* 1996.

Hahn, C. J., S. G. Warren, and J. London, The effect of moonlight on observations of cloud cover at night, and applications to cloud climatology, *J. Clim.*, *8*, 1429–1446, 1995.

Heinemann, G., TOVS retrievals obtained with the 3I-algorithm: A study of a meso-scale cyclone over the Barents Sea, *Tellus*, *47A*(3), 324–330, May 1995.

Heinemann, G., S. Noel, A. Chedin, N. Scott, and C. Claud, Sensitivity studies of TOVS retrievals with 3I and ITPP retrieval algorithms: Application to the resolution of meso-scale phenomena in the Antarctic, *Meteorol. Atmos. Phys.*, *55*(1–2), 87–100, 1995.

Intrieri, J. M., M. D. Shupe, B. J. McCarty, and T. Uttal, Annual cycle of cloud statistics from SHEBA, *J. Geophys. Res.*, *107*, 10.1029/2000JC000423, in press, 2002.

Key, J., *The Cloud and Surface Parameter Retrieval (CASPR) System for Polar AVHRR*, 59 pp., Coop. Inst. for Meteorol. Satell. Stud., Univ. of Wisc., Madison, 2000.

Kidwell, K., *NOAA Polar Orbiter Data User's Guide (TIROS-N, NOAA-6, NOAA-7, NOAA-8, NOAA-9, NOAA-10, NOAA-11, NOAA-12, NOAA-13 AND NOAA-14)*, Revision, November 1998.

Koepken, C., G. Heinemann, A. Chedin, C. Claud, and N. A. Scott, Assessment of the quality of TOVS retrievals obtained with the 3I algorithm for Antarctic conditions, *J. Geophys. Res.*, *100*, 5143–5158, 20 March 1995.

Lindsay, R. W., Temporal variability of the energy balance of thick Arctic pack ice, *J. Clim.*, *11*, 313–333, 1998.

Lindsay, R. W., D. B. Percival, and D. A. Rothrock, The discrete wavelet transform and the scale analysis of the surface properties of sea ice, *IEEE Trans. Geosci. Remote Sens.*, *34*, 771–787, 1996.

Maslanik, J., C. Fowler, J. Key, T. Scambos, T. Hutchinson, and W. Emery, AVHRR-based polar pathfinder products for modeling applications, *Ann. Glaciol.*, *35*, 171–182, 1997.

Moran, K. P., B. E. Martner, D. C. Welsh, D. A. Merritt, M. J. Post, T. Uttal, and R. G. Strauch, ARM's cloud profiling radar, paper presented at the 28th Conf. on Radar Meteorol., Austin, Tex., 1997.

Percival, D. B., and P. Guttorp, Long-memory processes, the Allan variance and wavelets, in *Wavelets in Geophysics*, edited by E. Foufoula-Georgiou and P. Kumar, pp. 325–343, Academic, San Diego, Calif., 1994.

Perovic, K. D., et al., The reference is "Year on ice gives climate insights", *Eos Trans. AGU*, *80*, pp. 481, 485–486, 1999.

Rodgers, C. D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Rev. Geophys. Space Phys.*, *14*, 609–624, 1976.

Schweiger, A. J., and J. Key, Comparison of ISCCP-C2 and Nimbus-7 satellite-derived cloud products with a surface-based cloud climatology in the Arctic, *J. Clim.*, *5*, 1514–1527, 1992.

Schweiger, A. J., R. W. Lindsay, J. R. Key, and J. A. Francis, Arctic clouds in multiyear satellite data sets, *Geophys. Res. Lett.*, *26*(13), 1845–1848, 1999.

Scott, N. A., A. Chedin, R. Armante, J. Francis, C. Stubenrauch, J.-P. Chaboureau, F. Chevallier, C. Claud, and F. Cheruy, Characteristics of the TOVS Pathfinder Path-B dataset, *Bull. Am. Meteorol. Soc.*, *80*, 2679–2701, 1999.

Smith, W. L., H. M. Woolf, C. M. Hayden, and A. J. Schreiner, The simultaneous retrieval export package, in *The Technical Proceedings of the Second International TOVS Study Conference, Igls, Austria, 18–22 February 1985*, edited by W. P. Menzel, pp. 224–254, CIMSS, Univ. of Wisc., Madison, 1985.

Stokes, G. M., and S. E. Schwartz, The Atmospheric Radiation Measurement (ARM) program: Programmatic background and design of the cloud and radiation test bed, *Bull. Am. Meteorol. Soc.*, *75*, 1201–1221, 1989.

Stubenrauch, C. J., N. A. Scott, and A. Chedin, Cloud field identification for Earth radiation budget studies, I, Cloud field classification using HIRS-MSU sounder measurements, *J. Appl. Meteorol.*, *35*(3), 416–427, March 1996.

Stubenrauch, C. J., W. B. Rossow, F. Cheruy, A. Chedin, and N. A. Scott, Clouds as seen by satellite sounders (3I) and imagers (ISCCP), part I, Evaluation of cloud parameters, *J. Clim.*, 2189–2213, 1999a.

Stubenrauch, C. J., A. Chédin, R. Armante, and N. A. Scott, Clouds as seen by satellite sounders (3I) and imagers (ISCCP), part II, A new approach for cloud parameter determination in the 3I algorithms, *J. Clim.*, *12*, 2214–2223, 1999b.

Stubenrauch, C. J., W. B. Rossow, N. A. Scott, and A. Chédin, Clouds as seen by satellite sounders (3I) and imagers (ISCCP), part III, Spatial heterogeneity and radiative effects, *J. Clim.*, *12*, 3419–3442, 1999c.

Walsh, J. E., and W. L. Chapman, *Arctic Cloud-Radiation-Temperature Associations in Observational Data and Atmospheric Reanalyses*, pp. 3030–3045, November 1998.

Wilson, L. D., J. A. Curry, and T. P. Ackerman, Satellite retrieval of lower-tropospheric ice crystal clouds in the polar regions, *J. Clim.*, *6*, 1467–1472, 1993.

J. Francis, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA.

J. Intrieri, NOAA/Environmental Technology Laboratory, Boulder, CO, USA.

J. Key, Office of Research and Applications, National Environmental Satellite, Data, and Information Service, NOAA, University of Wisconsin, Madison, WI, USA.

R. Lindsay and A. Schweiger, Applied Physics Laboratory/Polar Science Center, University of Washington, 1013 NE 40th Street, Seattle, WA 98109, USA. (axel@apl.washington.edu)

M. Shupe, Science and Technology Corporation, Hampden, VA, USA.