

Monitoring Precipitation over the Arctic Terrestrial Drainage System: Data Requirements, Shortcomings, and Applications of Atmospheric Reanalysis*

MARK C. SERREZE AND MARTYN P. CLARK

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

DAVID H. BROMWICH

Byrd Polar Research Center, The Ohio State University, Columbus, Ohio

(Manuscript received 22 July 2002, in final form 17 October 2002)

ABSTRACT

An effort is under way aimed at historical analysis and monitoring of the pan-Arctic terrestrial drainage system. A key element is the provision of gridded precipitation time series that can be readily updated. This has proven to be a daunting task. Except for a few areas, the station network is sparse, with large measurement biases due to poor catch efficiency of solid precipitation. The variety of gauges used by different countries along with different reporting practices introduces further uncertainty. Since about 1990, there has been serious degradation of the monitoring network due to station closure and a trend toward automation in Canada.

Station data are used to compile monthly gridded time series for the 30-yr period 1960–89 at a cell resolution of 175 km. The station network is generally sufficient to estimate the mean and standard deviation of precipitation at this scale (hence the statistical distributions). However, as the interpolation procedures must typically draw from stations well outside of the grid box bounds, grid box time series are poorly represented. Accurately capturing time series requires typically four stations per 175-km cell, but only 38% of cells contain even a single station.

Precipitation updates at about a 1-month time lag can be obtained by using the observed precipitation distributions to rescale precipitation forecasts from the NCEP-1 reanalysis via a nonparametric probability transform. While recognizing inaccuracies in the observed time series, cross-validated correlation analyses indicate that the rescaled NCEP-1 forecasts have considerable skill in some parts of the Arctic drainage, but perform poorly over large regions. Treating climatology as a first guess with replacement by rescaled NCEP-1 values in areas of demonstrated skill yields a marginally useful monitoring product on the scale of large watersheds. Further improvements are realized by assimilating data from a limited array of station updates via a simple replacement strategy, and by including aerological estimates of precipitation less evapotranspiration ($P - ET$) within the initial rescaling procedure. Doing a better job requires better observations and an improved atmospheric model. The new ERA-40 reanalysis may fill the latter need.

1. Introduction

The hydroclimatology of the Arctic terrestrial drainage plays an important role in the climate system. The primary freshwater source to the Arctic Ocean is river discharge (Aagaard and Carmack 1989), the bulk contributed by the Ob, Yenisey, Lena, and Mackenzie. River discharge influences ocean salinity and sea ice conditions (McDonald et al. 1999; Steele and Boyd 1998), which can impact freshwater fluxes traveling through the Fram Strait and Greenland Sea into the North Atlantic. The degree of surface freshening in the North

Atlantic is thought to influence the global thermohaline circulation (Broecker 1997). Changes in the terrestrial hydrologic cycle may alter soil moisture, impacting plant communities. Arctic soils serve as potentially significant sources of carbon dioxide and methane. Fluxes appear to respond sensitively to altered soil moisture and temperature (Oechel et al. 1993).

Studies from numerous disciplines document significant change in the northern high-latitude environment (Serreze et al. 2000; Moritz et al. 2002). Pronounced winter and spring warming over Eurasia and northwest North America since about 1970 is partly compensated by cooling over eastern Canada and the northern North Atlantic. This has been attended by shifts in the atmospheric circulation characterized by dominance of the positive mode of the Arctic Oscillation–North Atlantic Oscillation (Thompson and Wallace 1998, 2000). Climate proxies (e.g., tree rings and varves), which are

* Byrd Polar Research Center Contribution Number 1275.

Corresponding author address: Dr. Mark C. Serreze, CIRES, University of Colorado, Campus Box 449, Boulder, CO 80309-0449.
E-mail: serreze@kryos.colorado.edu

primarily indicators of summer temperature, point to the twentieth-century Arctic as the warmest of the past 400 years. Negative snow cover anomalies have dominated the North American and Eurasian continents since the late 1980s and terrestrial precipitation has exhibited a general increase since 1900. Small Arctic glaciers have experienced generally negative mass balances. There is also evidence of increased plant growth and a longer growing season. Winter discharge from the major Siberian rivers appears to have increased (Serreze et al. 2003). Oceanic trends include a general downward tendency in sea ice extent and increased inflow of warm Atlantic-derived waters into the Arctic Ocean (Cavaleri et al. 1997; Dickson et al. 2000).

These recent changes point to the need to monitor the Arctic system and better understand interactions between system components. The terrestrial hydrologic budget is a high priority. Important tasks include assembling multidecade time series of budget components [e.g., precipitation (P), evapotranspiration (ET), $P - ET$, snow water equivalent, runoff] and related variables (temperature, permafrost active layer depth). A key element of monitoring is to provide time series updates and take advantage of new data streams, such as provided by National Aeronautics and Space Administration (NASA) Earth Observation System platforms.

There has been progress in achieving these goals. The University of New Hampshire (UNH) has released R-ArcticNET, an archive of historic monthly river discharge for the pan-Arctic drainage (Lammers et al. 2001). UNH is acquiring daily discharge updates for many rivers in Russia and Canada at time lags as short as several days. The National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR, hereafter referred to as NCEP) reanalysis system (Kalnay et al. 1996; Kistler et al. 2001) has emerged as a critical research tool. Reanalysis provides a modern picture of the atmospheric hydrologic budget through assessments of $P - ET$, calculated from wind and humidity profiles (Bromwich et al. 2000; Cullather et al. 2000; Rogers et al. 2001; Serreze et al. 2003). Efforts have been made to assemble quality precipitation datasets for the former Soviet Union (FSU) (Groisman et al. 1991) and for Canada (Groisman 1998; Mekis and Hogg 1999). Satellites are providing systematic coverage of snow extent, and efforts are on going to validate retrievals of snow water equivalent from passive microwave brightness temperatures (Tait 1998). Zhang et al. (1999) and Running et al. (1999) report on techniques to assess near-surface soil freeze–thaw status using passive microwave and scatterometer data.

A project known as Arctic-RIMS (Rapid Integrated Monitoring System) is bringing datasets and techniques together to provide readily accessible hydrologic products. Arctic-RIMS is a collaborative effort between University of Colorado, University of New Hampshire, The Ohio State University, and the NASA Jet Propulsion Laboratory. The project uses satellite data, the NCEP

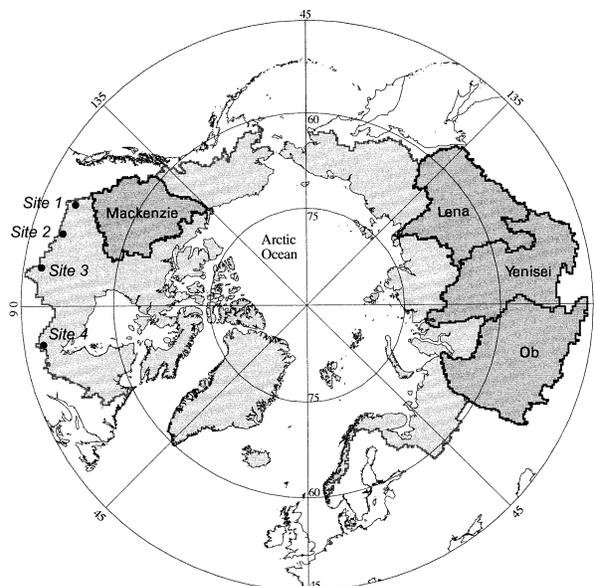


FIG. 1. The Arctic terrestrial drainage (all shaded areas) and its four major watersheds (the Ob, Yenisey, Lena, and Mackenzie). Also indicated are the locations of four well-instrumented sites used in Monte Carlo simulations.

reanalysis, in situ records (including observed discharge), and a permafrost/water balance model to compile fields of P , $P - ET$, ET , temperature, soil moisture, soil freeze–thaw state, active layer thickness, snow extent, and its water equivalent, soil water storage, and other variables. Historical time series are provided along with updates at a 1–2-month time lag. Gridded products are assembled over the complete Arctic terrestrial drainage (Fig. 1), defined as areas emptying into the Arctic Ocean as well as into Hudson Bay, James Bay, Hudson Strait, the Bering Strait, and northern Bering Sea.

The present paper discusses a core element of Arctic-RIMS—the provision of historic time series and updates of gridded monthly precipitation. This has proven to be a difficult task. The required station density to assemble quality historic time series at a spatial scale useful for input to hydrologic models exceeds what is available over most of the Arctic drainage. The problem is compounded by large errors in the measurement of solid precipitation and degradation of the station network since about 1990, the latter due to budget cuts in both the FSU and Canada. For example, the station coverage for the FSU in 1996 is about half of that available in the mid 1980s. Canada is also seeing a trend toward the replacement of manual observations by automated systems, providing data of suspect quality. Given this degradation, monitoring must rely more heavily on alternative data sources. We describe a monitoring approach that blends precipitation observations with output from the NCEP reanalysis. We show how statistical distributions of gridded station precipitation can be used to rescale monthly NCEP precipitation forecasts to remove

systematic biases, and how these reconstructions can be further improved with assimilation of available station updates and aerological estimates of $P - ET$. The overall performance of the approach, however, is limited by the need for both better historic time series and a better atmospheric model.

2. Station precipitation datasets

a. Primary archives

A readily available source of monthly station precipitation data for the pan-Arctic drainage is the Global Historical Climatological Network (GHCN) (Vose et al. 1992). There are several regional datasets. Groisman et al. (1991) assembled monthly time series for 622 stations in the FSU. Data are available through the early 1990s for most stations and through the late 1990s for a subset. The National Climatic Data Center holds dataset TD-9816 “Canadian Monthly Precipitation” (Groisman 1998). It contains nearly 7000 stations, although the vast majority are in the southern part of the nation. Mekis and Hogg (1999) describe a separate Canadian dataset of 495 stations.

As far as we are aware, the GHCN data contain no bias adjustments. As discussed shortly, bias adjustments are intended to account for gauge undercatch of precipitation and other measurement problems. The FSU dataset and TD-9816 are available as raw and adjusted monthly values. The Mekis and Hogg (1999) Canadian dataset provided to us contains adjustments. The Arctic Precipitation Data Archive in Offenbach, Germany, maintains station data from a number of different sources, and in both adjusted and raw form. Through Arctic-RIMS, we also obtained data for 105 stations (raw monthly totals) within the Ob, Yenisey, and Lena basins for the period 1966–90. These stations are not present in Groisman et al.’s FSU archive.

Efforts were made to assemble the most complete pan-Arctic dataset possible. We use the Groisman et al. FSU archive, TD-9816, and the additional 105 records just described. For Alaska, Scandinavia, Greenland, and areas in Europe not covered by regional sources, use is made of the GHCN archive. As justified below, only the raw data were used. Monthly values available from one archive were sometimes missing in another. Records were merged as necessary to obtain more complete time series. We use the regional datasets as they are presumably subjected to a higher level of quality control than possible for the global (GHCN) database.

The station records were assembled over the period 1960–89. There are two reasons for choosing this 30-yr period. First, as just mentioned, station data are much less abundant for later years, with Canada seeing a trend toward automation. Second, NCEP reanalysis output (which represents a basis of the monitoring strategy), is less reliable prior to 1960 due to the sparse assimilation database (see section 3). Hence, 1960–89 rep-

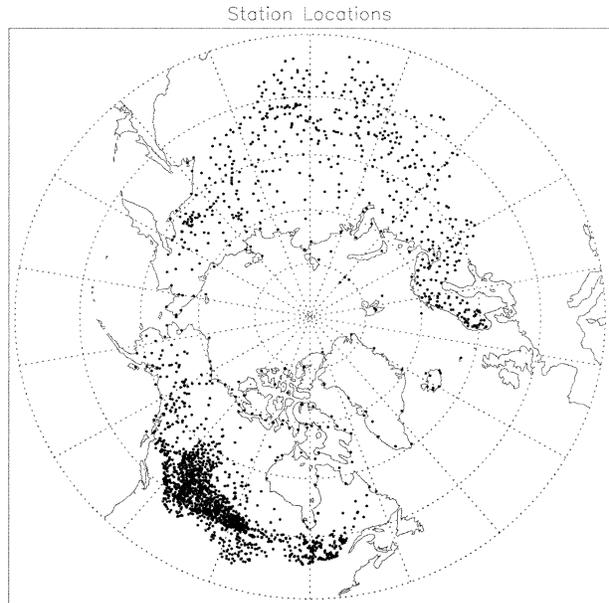


FIG. 2. Distribution of precipitation stations that are inside of or within 250 km of the boundaries of the Arctic terrestrial drainage having at least 10 yr of record length over the period 1960–89.

resents the 30-yr period for which reasonably complete time series of station precipitation are available along with “reliable” NCEP records. Figure 2 shows the location of stations with at least 10-yr of data over the period 1960–89 that are inside or lie within 250 km of the boundaries of the Arctic drainage.

b. Measurement problems and bias adjustments

The problem of biases in high-latitude precipitation data has long been recognized. However, what constitutes the most appropriate adjustment procedures is a lively area of debate and it will be some time before we have a “community dataset” representing a best-faith effort to deal with the problem. The primary issue is gauge undercatch of solid precipitation. This was examined through the World Meteorological Organization Solid Precipitation Intercomparison Project (Goodison et al. 1998; Yang et al. 2001). In summary, different countries use different gauge and wind shield combinations, which introduce variations in catch efficiency, especially for high wind speeds. Errors can reach 50%–100% in cold, windy environments. This has created artificial discontinuities in cold-region precipitation within countries and across international borders. Additional issues are wetting losses (the portion of precipitation that sticks to the walls of the gauge after it is emptied), evaporation losses, and treatment of trace precipitation amounts. The latter can be important as precipitation totals across much of the Arctic and subarctic are very low, especially in winter.

Bias adjustments in most archives (e.g., Legates and Willmott 1990; Groisman et al. 1991; Groisman 1998)

are climatological in representing constant multipliers to raw monthly precipitation totals. There have also been efforts to perform adjustments for individual days or months (Mekis and Hogg 1999; Yang 1999). Adjustments are site specific and require information on gauge type and changes through time, winds, site conditions, and measurement practices.

To illustrate a few of the problems, during the 1940s and 1950s, the FSU changed from use of Nipher-shielded gauges to improved Tretiyakov gauges. From parallel measurements, Groisman et al. (1991) adopted a correction to adjust precipitation measured with the Nipher gauge to be comparable with the Tretiyakov values. A wetting loss adjustment was also adopted. The wind corrections are a function of climatological wind speed, temperature, snowfall, and precipitation intensity at the gauge site. Techniques applied to the FSU data are inappropriate for Canada not just because of different gauges, but differences in the reporting of precipitation. The Canadian practice at most manually operated sites is to measure rainfall and snowfall separately. Rainfall is measured at gauges. A ruler is used to measure the depth of freshly fallen snow, which is converted into water equivalent using a 10:1 ratio. Starting in the early 1960s, some stations were equipped with Nipher-shielded elevated snow gauges that directly measure the water equivalent of snow.

Groisman (1998) adjusted the Canadian data by computing climatological ratios between the water equivalent measured at Nipher gauges (adjusted to account for undercatch) and from the manual ruler measurements. Ratios interpolated to the station locations were multiplied by the water equivalent at the stations as determined by the 10:1 ruler conversions. Before 1975, a wetting loss adjustment was performed, but information on the number of measurements per day that would allow for systematic corrections is not available at all stations. Wetting corrections were hence based on the mean number of days per month with rainfall or (if not available) a value interpolated from nearby stations. A small adjustment of rainfall was also included for estimated wind undercatch. After 1975 an improved “type B” gauge began to be used (Metcalf et al. 1997) and a wetting correction was considered unnecessary. There are no corrections for trace rainfall events. Mekis and Hogg (1999) apply broadly similar techniques to adjust the snowfall measurements and rain undercatch, but also include corrections for evaporation and trace precipitation amounts. Bias adjustment of data from automated systems is only beginning to be addressed.

We are reluctant to use bias adjusted data. First, we are concerned at the present lack of consensus regarding adjustment techniques for different regions. Second, from a practical viewpoint, the regional bias-adjusted datasets just described offer no coverage over Alaska, Scandinavia, and Greenland—regions for which there is no recourse but to use raw GHCN data. Third, we are concerned at potentially long delays in the posting

of bias-adjusted updates from regional sources. Hence we use raw data, leaving the option of adopting adjustments in a post-processing step, for example, using the gridded climatological adjustments of Legates and Willmott (1990).

3. Gridding the station data

a. Overview

As outlined in section 4, the monitoring approach makes use of the statistical distributions of monthly precipitation within grid cells using data over the period 1960–89. This requires the assembly of gridded time series. Here, we outline development of the gridded time series and the difficulties encountered. It is shown that the sparse station network limits the ability to preserve the “true” time series structure of grid cell precipitation, but is generally sufficient to define the statistical distributions. This finding bears directly on the monitoring strategy.

b. Data requirements

The problem of estimating grid cell precipitation from point data has been addressed in a number of studies (e.g., Rodriguez-Iturbe and Mejia 1974; Bras and Rodriguez-Iturbe 1976a,b). Our objective is to estimate the true precipitation at the spatial resolution selected. This implies a spatial average at the scale of the grid cell. Put differently, the estimated quantities for the grid cell should have similar statistics to the true quantities. In practice, the basic problem is that while precipitation statistics depend on spatial scale, the scale that can be adequately resolved depends on the density of the station network.

For example, consider monthly precipitation statistics for a 200-km grid cell in comparison to those for 20-km cells embedded within the larger grid cell. Within the larger grid cell, there will be variations in precipitation associated with the location of convective systems, the intensity of rain bands within synoptic-scale systems, and topography. Because of spatial averaging, the standard deviation of the precipitation time series for the larger cell will tend to be smaller than those for the 20 km cells contained within it. By the same token, precipitation time series for the 200-km cell may not be well correlated with the time series for the 20-km cells. Time-mean precipitation amounts (e.g., over a 30-yr period) are likely to be more similar.

With a sufficiently dense station network, the best estimate of grid cell precipitation is a simple average of the station values in the cell (the “drop-in-the-bucket” approach). Only one or two stations per 20-km cell would likely yield good results. However, for most areas of the globe, data are insufficient to generate such a high-resolution product. Values for cells with no stations in the cell could be obtained from spatial inter-

polation with distance weights, but if one must draw from stations well outside grid cell boundaries, the statistics will no longer be those of the desired 20-km scale. While the apparent solution is to use larger grid cells, more stations will generally be needed to get good cell values. In practice, one must compromise between choosing grid cells small enough for the desired application (for Arctic-RIMS, use in a hydrologic model), yet large enough so that the cell statistics are reasonably representative of the chosen cell size.

Satellite data streams for Arctic-RIMS are assembled at a 25-km resolution, based on the National Snow and Ice Data Center North Polar equal-area scalable earth grid (Armstrong and Brodzik 1995). This resolution cannot be supported for precipitation. Below we address data requirements for obtaining grid cell averages using Monte Carlo simulations applied to January and July station data from four reasonably well-instrumented sites over in the southern part of the Arctic drainage over Canada (Fig. 1). The tests focus on grid cells of 175 km and 350 km for July and January.

c. Monte Carlo simulations

As a first step in the analysis, grid cells of 175 and 350 km were centered over each of the four test sites. A drop-in-the-bucket approach was used to obtain grid cell precipitation for every year from 1960 to 1989, using all of the N stations located within each cell. Means and standard deviations were then obtained from the time series. At 350 km, each test cell contains more than 30 stations, while the 175-km cells contain 6–14 stations. These cell time series and associated statistics are the best approximations that can be obtained of the true (but unknown) values. Sampling errors will be present, the magnitude of which will depend on station density and interstation variability. In recognition we refer to these time series and statistics as those from the “dense network.” Our approach is best considered as a comparison between grid cell statistics approximating the truth with statistics generated from degraded networks. That the “true” time series and associated statistics are not exactly defined does not invalidate this approach.

For each grid cell, 500 different random selection of $N - 1$ stations were drawn. Each random selection was used to generate new time series using a drop-in-the-bucket approach. Time means and standard deviations were computed for each of the 500 time series, along with the squared correlation between the dense network and degraded ($N - 1$ station) time series. Based on the 500 realizations for each cell, we calculated (a) the absolute mean standard error (MSE) of the 30-yr mean precipitation as a fraction of the dense network mean, (b) the MSE of the 30-yr standard deviation as a fraction of the dense network standard deviation, and (c) the mean-squared correlation between the dense network and degraded time series. The entire process was then

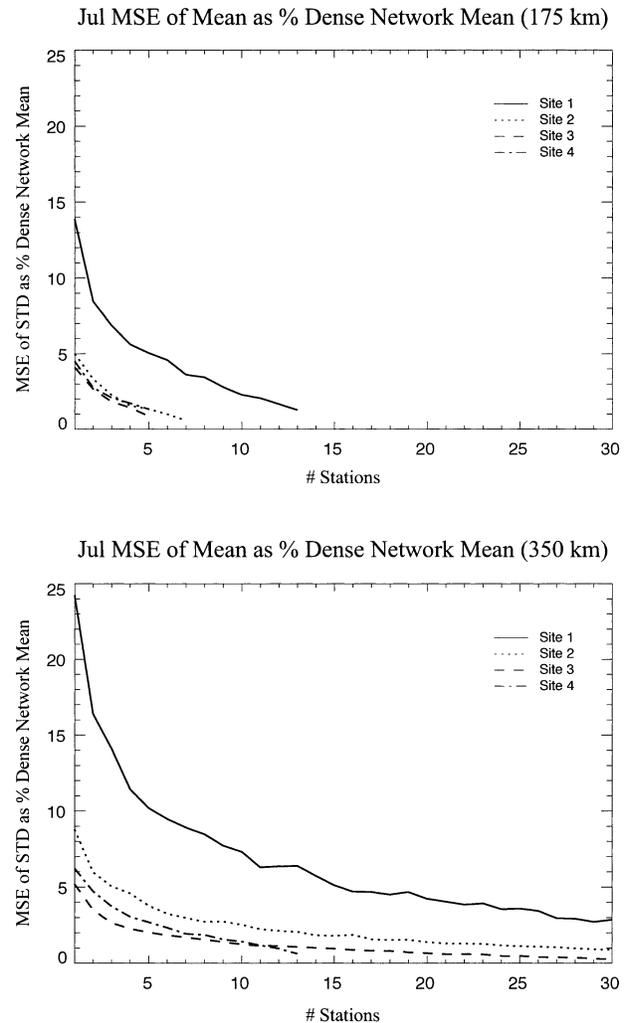


FIG. 3. Results from Monte Carlo simulations for four well-instrumented regions illustrating how well the dense-network mean precipitation (1960–89) based on all stations in cells of (top) 175 km and (bottom) 350 km can be estimated as stations are successively removed from the cell. Values are expressed as the absolute mean standard error (MSE) of the mean precipitation as a percentage of the dense network mean.

repeated for each cell using 500 random selections of $N - 2$ stations, $N - 3$ stations, and so on to $N - 1$ stations. Only those stations with at least 25 years of data were used. Missing station values complicate the analysis and were hence replaced with the station means. The use of 500 iterations obviously greatly oversamples the station network.

July results for the time series means, standard deviations and correlations are provided in Figs. 3–5, respectively. Results are shown for up to 30 stations. Sites 2–4 exhibit similar statistics, while site 1 is different. Site 1 is located near the foot of the Rocky Mountains (near Calgary) where topography results in strong local variability in precipitation. Some conclusions can nevertheless be drawn. For the 175-km cells, a reasonable

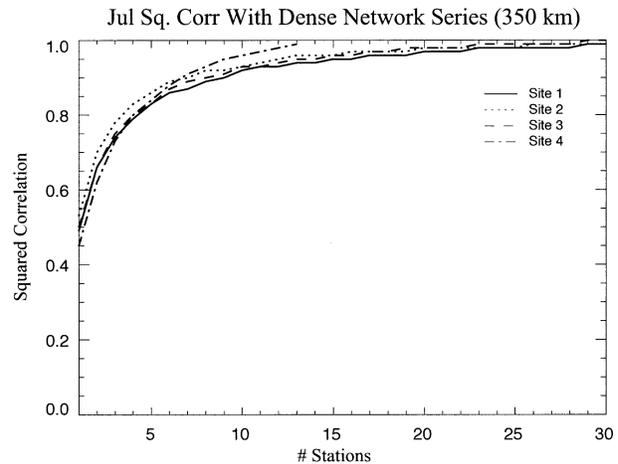
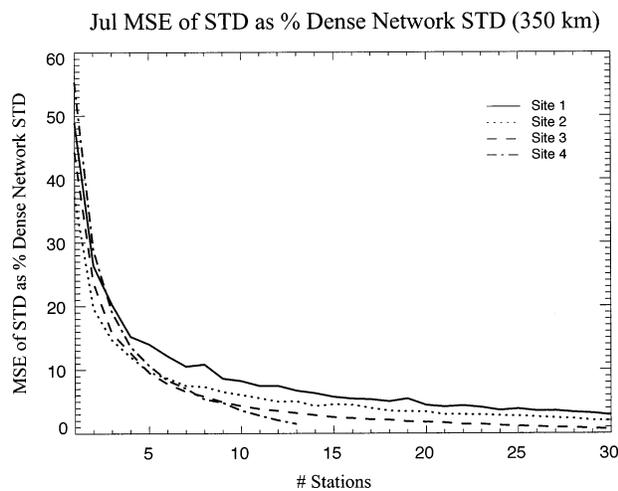
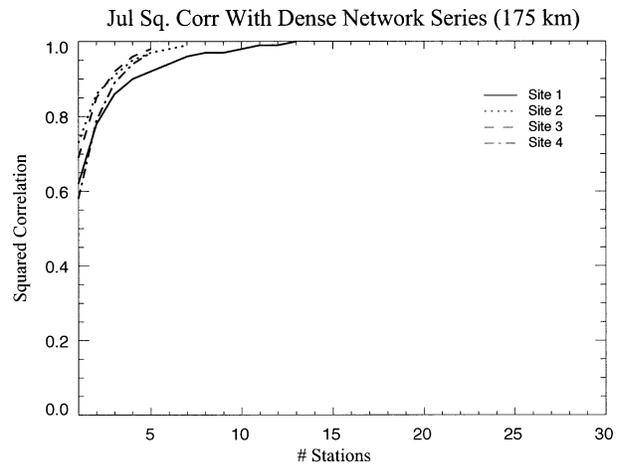
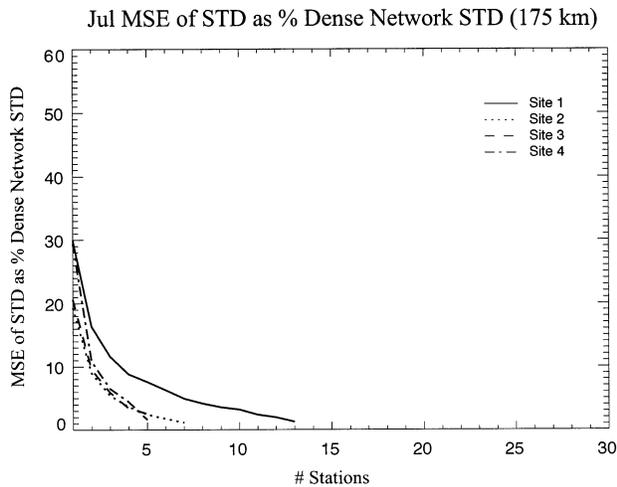


FIG. 4. Results from Monte Carlo simulations for four well-instrumented regions illustrating how well the dense network standard deviation of precipitation (1960–89) based on all stations in cells of (top) 175 km and (bottom) 350 km can be estimated as stations are successively removed from the cell. Values are expressed as the MSE of the std dev as a percentage of the dense network std dev.

FIG. 5. Results from Monte Carlo simulations for four well-instrumented regions illustrating the squared correlation between the dense network time series of precipitation (1960–89) based on all stations in cells of (top) 175 km and (bottom) 350 km and time series generated when stations are successively removed from the cell. Values are expressed as the mean-squared correlation with the dense network time series.

estimate of the dense network mean precipitation (within 15%) can be obtained with only one station in the cell. With the exception of site 1, this is also true for the 350-km cells. More stations are needed to estimate the dense network standard deviation. For the 175-km cells, having only one station yields a mean error from 20% to 30%. To get within 10%, two to four stations are required. More stations (four to seven) are understandably needed to get the same return for 350-km cells. Having only one station gives an error up to 55%. At 175 km, having only one station yields an average squared correlation with the dense network time series ranging from about 0.60 to 0.70. For the 350-km grids, another station is needed to get basically the same return.

Results for January (not shown) are broadly similar. There is some evidence that fewer stations are required to get the same return as July, consistent with the dominance of more widespread synoptic-scale precipitation in this month as compared to July, when some of the precipitation is associated with local convection. With respect to sampling the mean, the coefficient of deviation (standard deviation divided by the mean) is somewhat higher in January as compared to July for sites 1 and 3. Values for the two months are similar for sites 2 and 4.

The results in Figs. 3–5 can be rephrased in terms of the probability that the mean and standard deviation is within a given percentage of the corresponding dense-network statistic. For example, looking again at July, at

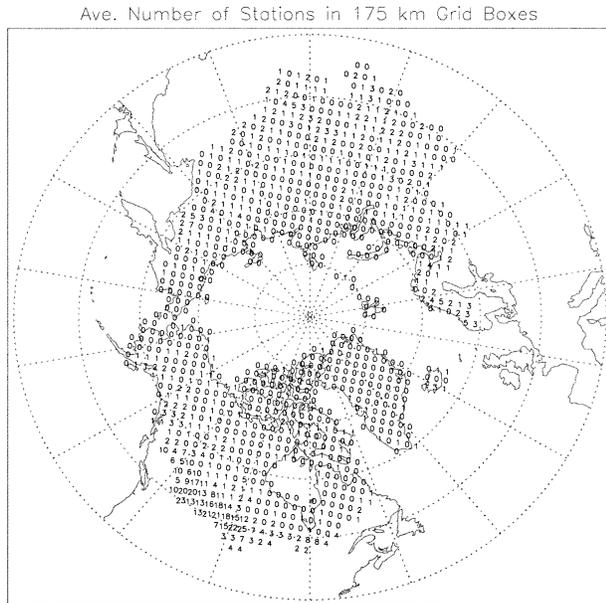


FIG. 6. Average number of stations within 175 km grid cells over the period 1960–89.

175 km with four stations there is a 54% (site 1) and 100% probability (the remaining sites) of getting the mean within 5% of the dense-network value. The probability of obtaining the standard deviation within 5% with four stations is lower, ranging from 32% (site 1) to 75% (site 3). Similarly, at 175 km in July with four stations, the probability of a squared correlation of at least 0.80 with the dense network time series is essentially 100% at all sites. By contrast, with only one station, the probability of a squared correlation of at least 0.80 falls to 28% at site 2 and 0% at the other locations.

A 175-km grid cell is small enough for use with the Arctic-RIMS hydrologic model. It is also on the same basic scale as the NCEP reanalysis output (2.5°) used in the monitoring approach. We conclude that preserving the dense network grid cell statistics (mean, standard deviation, and time series structure) at this scale requires at least four stations per box, and more in topographically complex regions such as site 1. However, Fig. 6 reveals that as averaged over the period 1960–89, only 38% of 175-km grid cells contain even one station. A similar map constructed for 350-km grids (not shown) naturally yields more grid cells with at least one station, but to get good statistics, more stations are needed per grid cell. A 350-km network is also very coarse, and of limited utility within Arctic-RIMS. In either case, the only recourse for obtaining full spatial coverage is spatial interpolation. For many cells the interpolation will have to draw on information from stations well outside of the grid box boundaries.

It is hence necessary to evaluate the effects of spatial interpolation on the precipitation statistics. To this end, the same basic Monte Carlo approach is applied to the

four test sites to see how well precipitation time series, means, and standard deviations from the dense network can be estimated from interpolation of stations that lie outside the grid cells. The procedure is applied to the 175-km test cells. Interpolation makes use of a modified version of the Shepard (1984) scheme. Willmott et al. (1985) provide a useful review. Interpolation weights are based on three categories of distance as follows:

$$\begin{aligned} \text{if } d \leq r/3, & \quad S = d^{-1} \\ \text{if } r/3 < d \leq r, & \quad S = 27/4r[(d/r) - 1]^2 \\ \text{if } d > r, & \quad S = 0, \end{aligned}$$

where d is the distance between the station and the center of the grid cell to be interpolated to, r is the search radius, and S is the interpolation weight. Values are defined for the maximum (MAX) and minimum (MIN) number of data points (station precipitation values) used in the interpolation. An initial search radius around each grid cell center is defined from the area of the spatial domain to be interpolated to and the number of available data points. If the number of data points within the search radius exceeds MAX, the closest data points up to MAX are used. If there are fewer than MIN data points in the search radius, the radius is expanded until at least MIN points are found. The interpolator uses spherical geometry to calculate distances, adjusts for uneven clustering of stations, and allows for extrapolation beyond the bounds of the station data based on local gradients.

The tests use station data in “rings” of (a) 87.5–175, (b) 87.5–350, and (c) 175–350 km from each of the grid cell centers. Tests (a) and (b) hence explore situations where no stations lie within the 175-km grid cell (true of 62% of all cells across the pan-Arctic) but stations are found within either one or two grid lengths distance, respectively. Experiment (c) assumes that stations are only found between one to two grid lengths from the cell center. For each test cell, 500 time series are compiled, each based on a random selection of a fixed number of stations within each data ring (i.e., MIN and MAX are set equal to each other). As before, results are expressed as the absolute MSE of the 30-yr mean and standard deviation as a fraction of the dense network values, and as the mean-squared correlation between the interpolated and dense network time series.

Results are shown in Table 1 based on a random selection of four stations. Four stations in the 87.5–175-km ring yields estimates of the mean and standard deviation comparable to those obtained from the drop-in-the-bucket approach using several stations (Figs. 3 and 4). The squared correlation with the dense network time series, however, is lower, ranging from 0.54 to 0.70. Using four stations in the 87.5–350-km ring also preserves the dense network mean and standard deviation reasonably well, while the squared correlations drop further to between 0.38 and 0.50. Using four stations in the 175–350-km ring further degrades the squared correlations (0.28–0.36), but even in the worst case (site

TABLE 1. Simulation results for four well-instrumented sites illustrating how well monthly precipitation time series, means, and standard deviations within 175-km grid cells from a dense network are represented through interpolation, using four stations located outside the cell bounds. See text for explanation.

Site	Data ring (km)	Squared correlation with dense network time series	MAE of mean as % dense network mean	MAE of std dev at % dense network std dev
Site 1	87.5–175	0.70	10.87	11.39
Site 2	87.5–175	0.57	8.10	10.37
Site 3	87.5–175	0.54	4.30	7.92
Site 4	87.5–175	0.54	3.18	7.72
Site 1	87.5–350	0.50	17.62	15.16
Site 2	87.5–350	0.43	9.50	11.34
Site 3	87.5–350	0.38	6.17	12.69
Site 4	87.5–350	0.38	5.97	11.10
Site 1	175–350	0.36	21.12	15.14
Site 2	175–350	0.28	10.67	10.42
Site 3	175–350	0.29	8.02	16.55
Site 4	175–350	0.30	7.97	14.98

1), the mean is estimated to be 21% of the dense network value.

If the objective is to calculate grid cell means and standard deviations at 175 km (hence statistical distributions), it seems that with spatial interpolation, one can get away with the fairly sparse pan-Arctic station network for most areas. By and large, however, the true time series structure is captured rather poorly. For the well-instrumented test sites, using more stations in the interpolation (e.g., 10 instead of 4) yields better characterization of the time series. However, using a minimum of 10 stations across the Arctic drainage typically degrades results, as it usually requires drawing from far outside the grid box bounds. The difficulty in capturing the dense network time series bears directly on how the NCEP data are applied in the monitoring approach.

d. The 1960–89 gridded time series

Based on the tests just described, we elected to compile the gridded time series at a resolution of 175 km. The time series was constructed as follows. If a grid cell contained at least four stations, precipitation was determined as the simple average of the stations. This provides the least-biased grid cell value (the Shepard interpolation by contrast provides center-weighted values). However, this takes care of only about 5% of grid cells (mostly in southern Canada and Scandinavia). The Shepard interpolation was used for the remainder of the cells.

The Shepard algorithm was modified to use only the four stations closest to each grid cell center. For most cells, this avoids drawing from distant stations. Most cells contain no stations, such that errors in the resulting precipitation statistics resemble those in Table 1 (precipitation distributions are preserved much better than the time series). For cells with 1–3 stations, results are obviously better than in Table 1. Further Monte Carlo tests show that having one station in a cell, and using the three closest stations outside the cell, yields mean squared correlations with the dense network time series

exceeding 0.75. Clearly, values obtained over central Greenland are based on distant coastal sites and are largely meaningless. In the future, we may adopt a special interpolation for Greenland, drawing from data over the ice sheet collected from automatic weather stations.

Figure 7 shows the resulting mean monthly fields of precipitation. Cold season precipitation (November–April) is below 20 mm and locally less than 10 mm over much of eastern Eurasia, northern Alaska, and northern Canada. Anticyclonic conditions prevail and precipitable water is low. High totals along the coasts of southeast Greenland, Scandinavia, and Alaska reflect frequent cyclone activity associated with the Icelandic and Aleutian lows and orographic uplift of moist air masses. The Atlantic and Pacific-side maxima weaken during summer. By contrast, summer is the season of peak precipitation over most land areas. This is due to increased cyclone activity over land, convective precipitation and development of a coastal baroclinic zone arising from differential heating between the cold Arctic Ocean and snow-free land. Serreze et al. (2001, 2003) provide further discussion.

4. Applications of the NCEP reanalysis

a. Overview

It is useful to consider four options for monitoring precipitation: 1) make do with gridding available updates of station data; 2) make direct use of gridded precipitation forecasts from numerical weather prediction (NWP) models like the NCEP reanalysis; 3) use the gridded observed precipitation time series and NWP output for 1960–89 (forecasted precipitation and other variables such as vertical motion) to develop linear regression models that can be applied to NWP updates (a form of statistical downscaling); 4) use nonparametric methods to constrain NWP output by the statistical distribution of the gridded observations. A common thread between options 2–4 is that output could be subse-

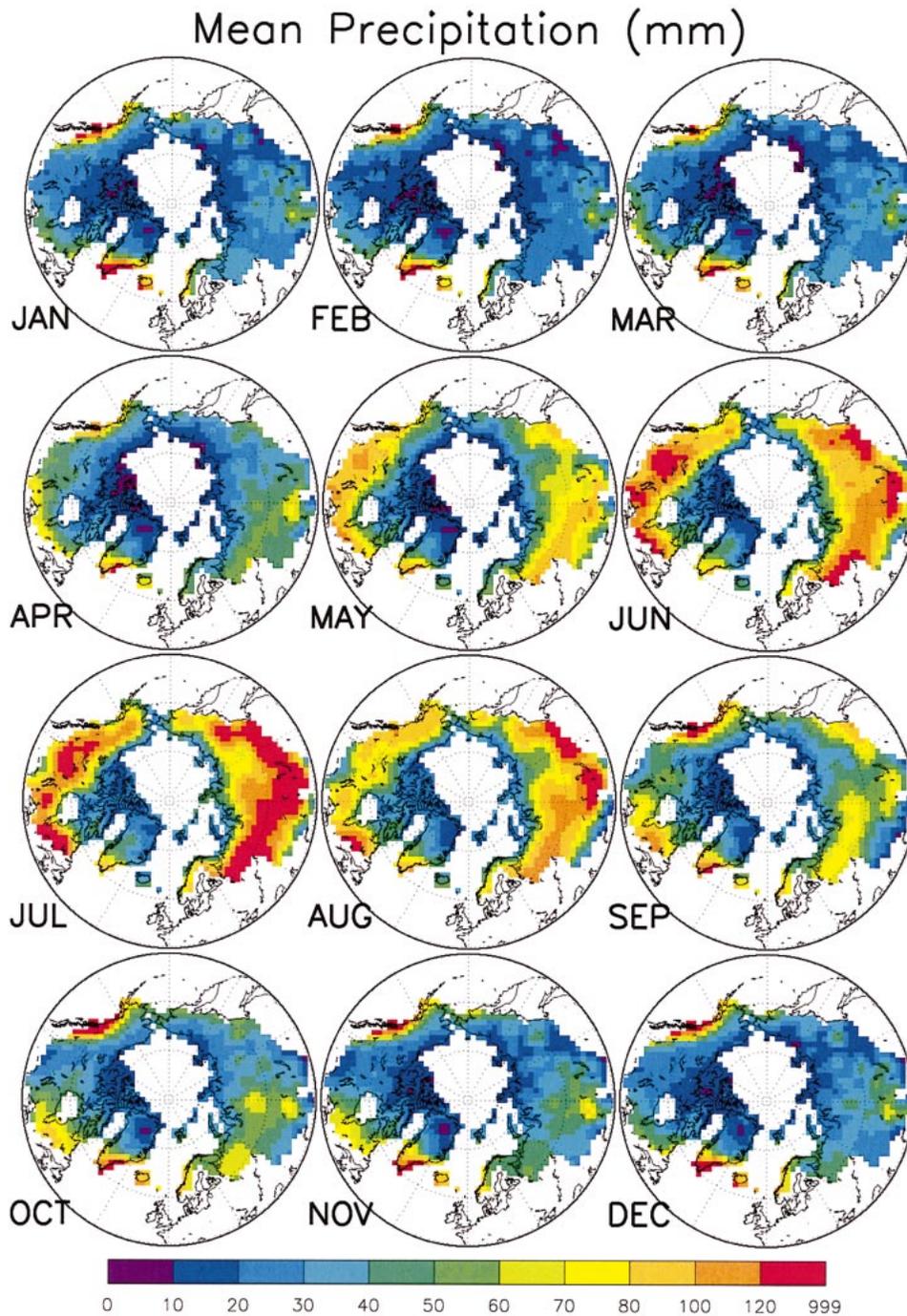


FIG. 7. Fields of observed monthly mean precipitation over the period 1960–89 (mm).

quently adjusted via assimilation of available station data updates.

The problem with option 1 is that station coverage since 1990 is much more sparse than for earlier decades and is insufficient by itself. Regarding option 2, NWP forecasts of precipitation typically contain large biases and cannot be used “as is” (Serreze and Hurst 2000). Option 3 (e.g., multiple linear regression) is clearly

problematic in that it requires faith in the observed gridded precipitation time series. As is evident from the results in section 3, the time series are of generally poor quality, meaning that one will be regressing against noise. The time series of individual stations represent truth (with due consideration of gauge undercatch and other biases). However, regression against station time series runs into problems of scale (relating point ob-

servations to relatively coarse-scale NWP output). Gridding the resulting station reconstructions also runs into the same problems discussed in section 3.

Option 4 emerges as the most viable. It recognizes that 1) biases in NWP precipitation forecasts are at least in part systematic; 2) systematic biases can be accounted for through rescaling procedures that require only the statistical distributions of observed precipitation rather than accurate representation of the gridded time series themselves; and 3) rescaling procedures can be applied to reconstruct precipitation from other variables, such as aerological estimates of $P - ET$, which can replace the rescaled NCEP precipitation forecasts if they are shown to provide better skill.

The following three sections review the NCEP reanalysis system used in the monitoring approach, characteristics of the NCEP precipitation forecasts, and their applications. Applications of $P - ET$ and data assimilation are reserved for section 5.

b. The NCEP reanalysis system

NWP models like the NCEP reanalysis start with a previous atmospheric forecast as a “first guess” of present atmospheric conditions. The first guess is adjusted through assimilation of observed atmospheric data (e.g., rawinsonde profiles, surface marine reports, aircraft observations of wind and temperature, synoptic reports of surface pressure, and satellite retrievals of temperature, humidity, and winds). This results in atmospheric analyses that are used to generate the next forecast. A general distinction can be made between variables in the NCEP archives (Kistler et al. 2001). Type “A” variables are strongly influenced by observations. Examples are geopotential height and u and v winds on pressure levels. Type “B” variables are those for which observed data directly affect the value of the variable, but where the model also has a strong influence. An example is humidity on pressure levels. Fields of type “A” and “B” variables are generally referred to as analyses. Type “C” variables are those for which no observations directly affect the variable (often termed forecasted or predicted variables). Examples are precipitation and radiation fluxes.

NCEP is a retroactive record of more than 50 years (continually updated) of global atmospheric analyses and forecasts. The effort involves recovery and assembly of numerous atmospheric datasets, which are then quality controlled and assimilated with a constant (“frozen”) data assimilation and forecast system. Output from operational systems (used for routine weather prediction) contain pseudoclimate signals (“jumps”) due to frequent changes in these systems (improvements in model physics and assimilation techniques). Reanalysis is intended to eliminate this problem. However, inhomogeneities are still present due to changes in the amount and quality of assimilation data (Kistler et al. 2001).

The NCEP reanalysis is performed with a T62 model with 28 vertical sigma levels. Outputs are provided every 6 h. Prior to 1958, rawinsonde coverage in the Arctic was very sparse, greatly degrading the reliability of the fields. Rawinsonde coverage increased after 1958, and again in the early 1970s. Satellite data began to be incorporated in the 1970s. Starting in 1979, drifting buoys began to provide regular reports of surface pressure over the Arctic Ocean. As discussed, we focus on the period 1960–89 for which reasonably complete time series of observed precipitation can be paired with NCEP records based on a reasonably robust assimilation database.

Two types of precipitation are computed in the NCEP model, convective and grid scale (dynamic). Convection is based on a simplified Arakawa–Schubert scheme (Pan and Wu 1994), which was found to result in improved forecasts of precipitation over the continental United States and Tropics as compared to the previous Kuo parameterization (Kalnay et al. 1996). Dynamic precipitation is parameterized by starting at the top layer of each model gridpoint column and checking for supersaturation. If supersaturated, latent heat is released to adjust the specific humidity and temperature to saturation, with the excess water falling to the next lower layer. If this next layer is supersaturated, then adjustment to saturation occurs again and the amount of precipitation is added to that from the higher layer. However, if the layer is unsaturated, some or all of the precipitation is evaporated. The process continues downward with all precipitation that penetrates to the bottom layer allowed to fall to the surface.

There are actually two NCEP data streams. The primary system described above, is hereafter referred to as NCEP-1. The second system, referred to as NCEP-2, is formally known as the NCEP–Department of Energy (DOE) Atmospheric Model Intercomparison Project (AMIP-2) Reanalysis. NCEP-2 addresses some of the known problems in NCEP-1 and incorporates improved physics. NCEP-2 is unfortunately only available from 1979 onward. Like NCEP-1 it is updated, but with a much slower turnaround. The assessments of reanalysis performance that follow include comparisons between NCEP-1 and NCEP-2.

c. Performance and biases

Monthly time series of precipitation (1960–89) from NCEP-1 (summed from daily values) were transformed to the 175-km grid array. This was accomplished using a Cressman (1959) interpolation with a 500-km radius of influence. For reasons discussed shortly, this large interpolation shell is needed to smooth the NCEP-1 output. Figure 8 shows the field of squared correlations between observed gridded precipitation and NCEP-1 forecasts based on the period 1960–89. Table 2 provides a monthly summary of the percent of grid cells with a squared correlation exceeding 0.50 (i.e., for which at least half the variance is shared). As evaluated in Table

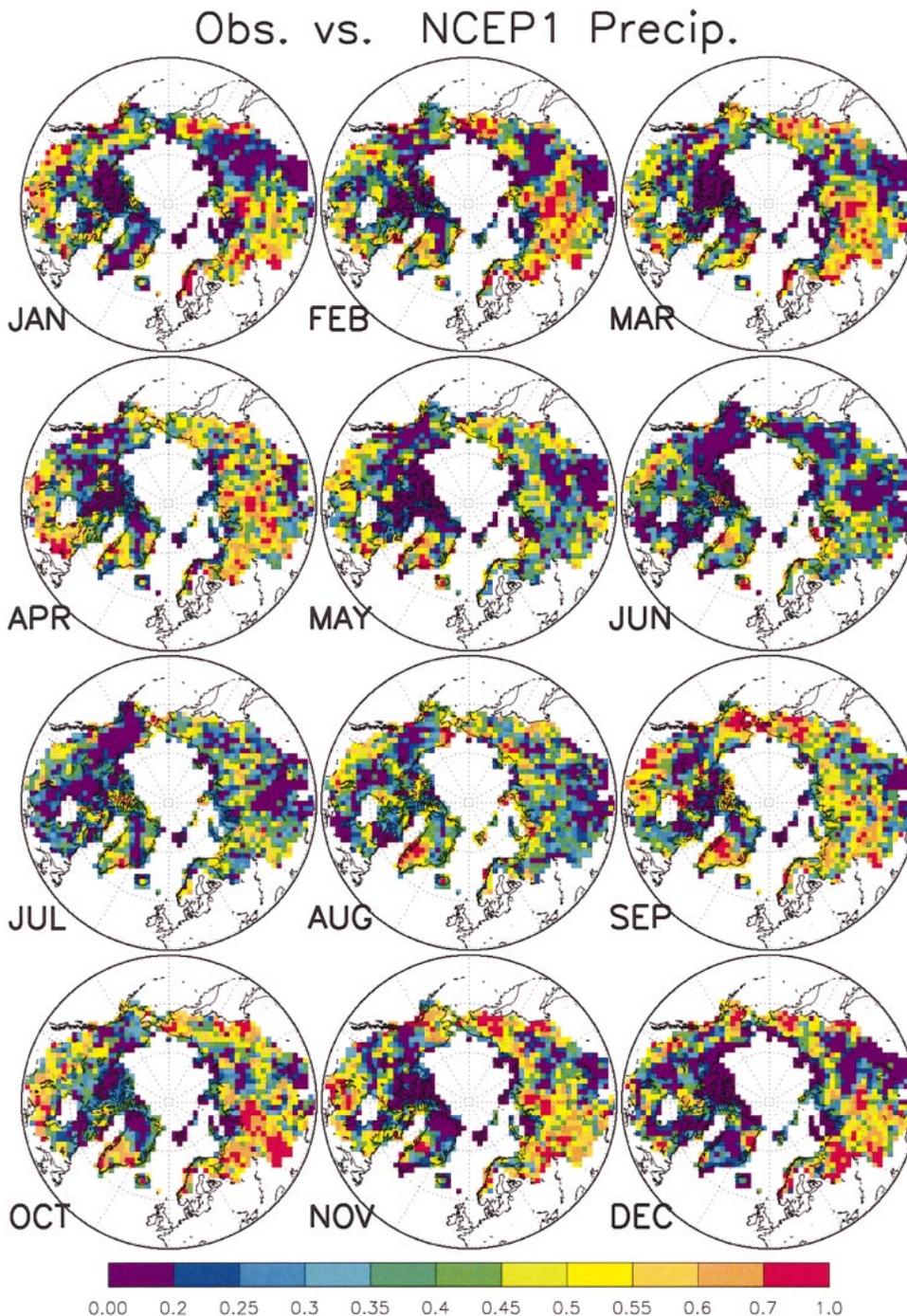


FIG. 8. Squared correlations between observed and NCEP-1 monthly precipitation over the period 1960–89.

2, overall performance is modest. Performance is best during the cold months, peaking in October when 35% of grid cells show a squared correlation exceeding 0.50. There is an obvious decline in performance during summer. As is evident from Fig. 8, performance varies regionally. Correlations are strong over western and cen-

tral Eurasia and southern Canada, but are quite weak over other parts of Eurasia and northern Canada.

As outlined at considerable length in section 3, the sparse station network means that the resulting gridded time series that are generated are typically not well representative of the 175-km scale. However, Table 2, Fig.

TABLE 2. Percent of grid cells by month for which the squared correlation between observed and NCEP-1 monthly precipitation exceeded 0.50 over the period 1960–89.

Month	Percent (%) of grids with squared correlation exceeding 0.5
Jan	27
Feb	31
Mar	30
Apr	34
May	20
Jun	14
Jul	15
Aug	21
Sep	36
Oct	35
Nov	35
Dec	28

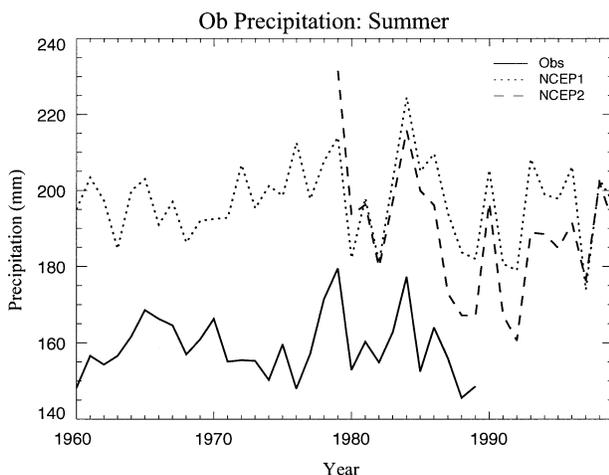
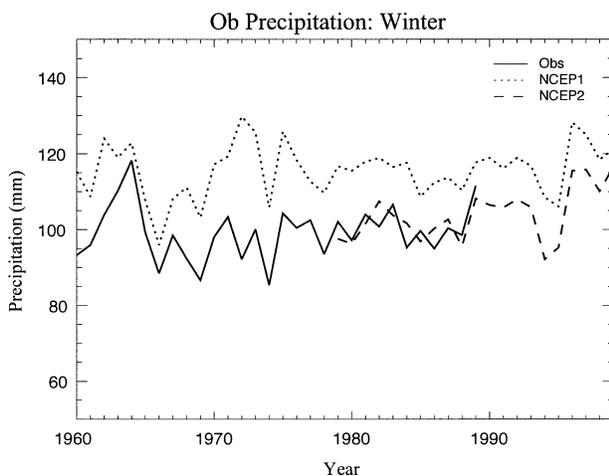


FIG. 9. Time series of basin-averaged precipitation for the Ob watershed based on observations, NCEP-1, and NCEP-2 for (top) winter and (bottom) summer.

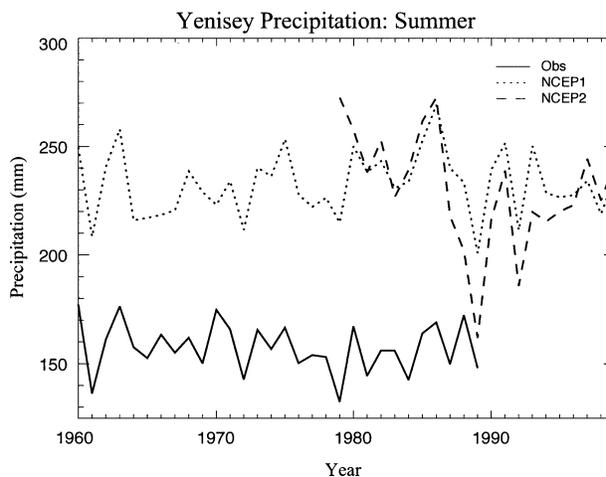
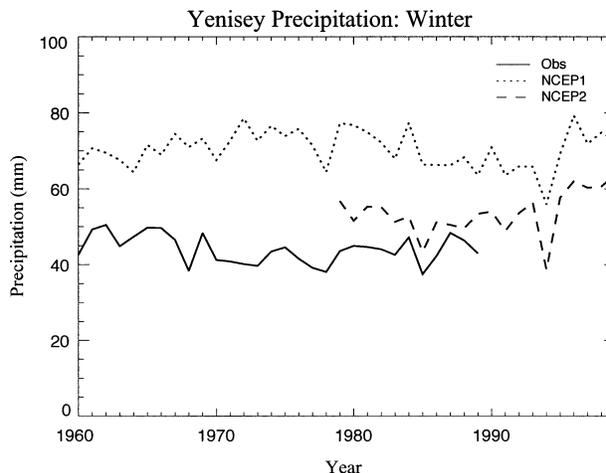


FIG. 10. Time series of basin-averaged precipitation for the Yenisey watershed based on observations, NCEP-1, and NCEP-2 for (top) winter and (bottom) summer.

8, and other assessments of skill that follow treat these time series as “truth.” This caveat must be kept in mind. For example, the consistently low correlations over northern Canada point to poor performance of NCEP-1. But this is also a region with a very low density of precipitation monitoring stations (Fig. 2). It is also a region where mean precipitation in all months is low (Fig. 7). Even small errors in precipitation measurement will hence degrade the observed time series. From both counts it is reasonable to speculate that the true correlations in this region are higher than what is indicated.

The NCEP-1 forecasts contain significant biases. This is illustrated by time series of winter and summer precipitation averaged for grid cells in the Ob, Yenisey, Lena, and Mackenzie basins, the four major Arctic-draining watersheds (Figs. 9–12). Results from NCEP-2 are shown for comparison. NCEP-1 totals are systematically too high, particularly during summer (note that

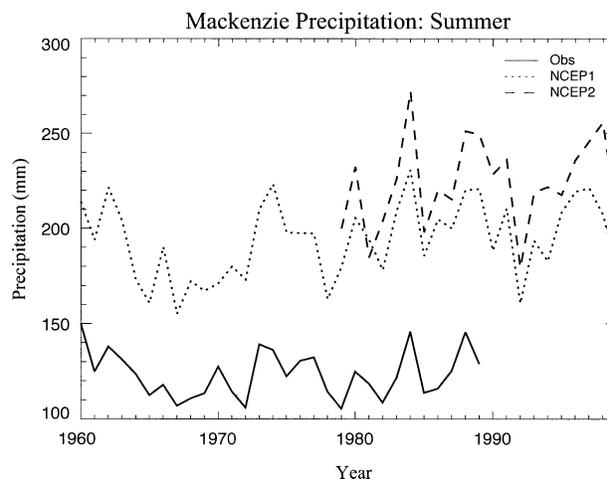
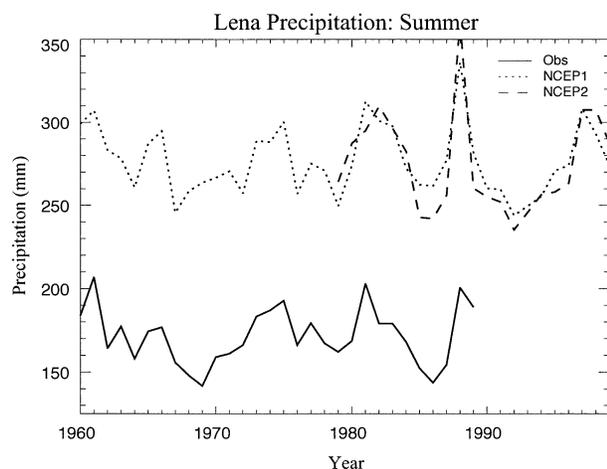
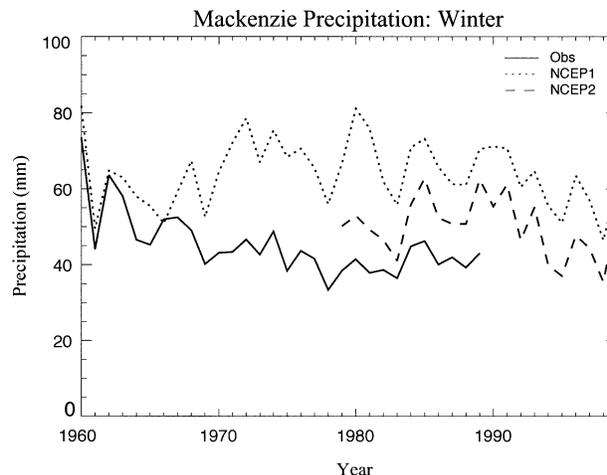
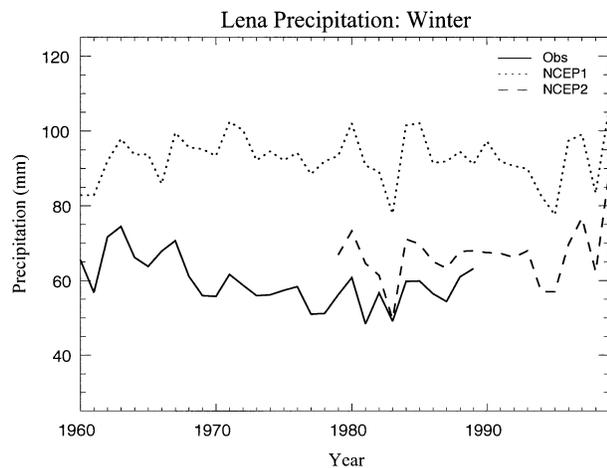


FIG. 11. Time series of basin-averaged precipitation for the Lena watershed based on observations, NCEP-1, and NCEP-2 data for (top) winter and (bottom) summer.

FIG. 12. Time series of basin-averaged precipitation for the Mackenzie watershed based on observations, NCEP-1, and NCEP-2 data for (top) winter and (bottom) summer.

gauge undercatch biases in observed precipitation are generally small in this season). The winter overestimates, at least in part, reflect the use of raw as opposed to bias-adjusted station precipitation data. Serreze and Hurst (2000) show that the summer bias in NCEP-1 is mostly due to excessive convective precipitation. Convection is particularly difficult to parameterize. The extent to which the problem reflects shortcomings in the convective scheme per se or is conflated with problems in other model components remains to be fully resolved. As shown by Serreze et al. (1998), during summer, downwelling solar fluxes in the model are too high due to insufficient cloud cover. This is consistent with the very high ET rates in the NCEP model, promoting excessive convection. High precipitation and ET are reinforced because soil moisture is updated by the modeled precipitation (Serreze and Hurst 2000). Note that the NCEP-2 reanalysis is closer to the observations dur-

ing winter, but still contains the strong summer biases of NCEP-1.

Summer is characterized by strong contributions from convective precipitation (Serreze and Etringer 2003, manuscript submitted to *Int. J. Climatol.*). As convective precipitation tends to occur at local scales, one cannot expect to obtain reliable summer-month grid cell time series with a sparse station network. This contributes to the lower summer correlations (Fig. 8). Furthermore, convective precipitation is very difficult to parameterize, and there is no reason to think that the summer biases in the NCEP-1 model just described are entirely systematic. The NCEP-1 archives include the convective precipitation component as a separate variable. The dynamic (large scale) component can be obtained as total precipitation minus convective precipitation. Fields of the squared correlation between the NCEP-1 dynamic precipitation fields and observations are very similar to

those in Fig. 8. It seems that including the convective component offers no improvement in skill.

Turning to other problems, NCEP-1 has a formulation of the horizontal moisture diffusion that causes moisture convergence, leading to unrealistic snowfall (“spectral snow”) over high-latitude valleys in winter (Kistler et al. 2001). Recent Arctic studies (Serreze and Hurst 2000; Cullather et al. 2000) show that the problem exhibits a definite association with topographic features and is present year-round. It has been corrected in the NCEP operational model as well as NCEP-2. NCEP also provides a precipitation dataset with a posteriori correction, but according to Cullather et al. (2000), these fields are still prone to problems and are overly dry. Smoothing the NCEP-1 data by the Cressman interpolation as we have done provides for a more realistic spatial pattern when compared to observations (Serreze and Hurst 2000).

The sensitivity of the squared correlation fields in Fig. 8 was examined with respect to different treatments of the NCEP-1 data. Tests included interpolating the monthly totals to the 175-km grid cell array using larger (750 km) and smaller (250 km) interpolation shells in the Cressman routine (as opposed to the 500-km interpolation shell used in the above results) and basing the monthly reanalysis totals on sums of daily precipitation exceeding different thresholds (e.g., 1 mm, 5 mm). No improvements over those shown in Fig. 8 were found.

Fields of monthly squared correlations between observed and NCEP-2 precipitation were computed for the 11-yr period 1979–89 and compared with results for the same period using NCEP-1. Although based on a short record, the NCEP-1 and NCEP-2 correlation fields have the same spatial structure. However, correlations over the 11-yr period from both reanalyses are stronger than shown in Fig. 8, which are based on 1960–89. There is an average increase of about 0.05 explained variance for most months. To examine this further, correlations between observed and NCEP-1 precipitation were examined for different record lengths, starting with 1960–89, then for 1961–89, 1962–89, etc., through 1979–89. A general increase in average correlations is observed when moving toward the more recent period. We suspect that this reflects improvements in the assimilation database through time.

d. Rescaling the NCEP-1 precipitation forecasts

If the regridded NCEP-1 precipitation forecasts are to be used in a monitoring strategy, the biases (Figs. 9–12) must be removed. The chosen approach is to rescale the forecasts via a probability transformation (Panofsky and Brier 1963). The procedure uses ranked (i.e., sorted) values of NCEP-1 and observed precipitation at each grid cell for 1960–89. The ranks are ascribed cumulative probabilities. Imagine that an update of NCEP-1 precipitation is desired for June 2004. We determine where the June 2004 NCEP-1 value falls in the 30-yr (1960–

89) NCEP-1 cumulative probability distribution. The June 2004 NCEP-1 value is rescaled by simply replacing it with the observed precipitation value at the same cumulative probability. Generally, linear interpolation is necessary because the NCEP-1 value to be rescaled lies between two of the ranked NCEP-1 values in the sample (1960–89) distribution. If the June 2004 NCEP-1 value is smaller than the smallest value in the NCEP-1 rankings, it is ascribed the smallest value in the observed 30-yr distribution. If the June 2004 NCEP-1 value is greater than the largest value in the NCEP-1 rankings, it is replaced with the largest observed value in the 30-yr distribution. The rescaling technique assures that any resulting rescaled NCEP-1 time series has the same mean and standard deviation as the corresponding observed time series.

To assess skill, the rescaling operation was evaluated via cross validation (Michaelson 1987) over the period 1960–89. For each year, the NCEP-1 and observed values for that year were withheld, and the rescaled NCEP-1 value was obtained from the distributions compiled from the remaining years. For example, the raw NCEP-1 and observed values for June 1960 are held out. A rescaled NCEP-1 value for June 1960 is determined from the NCEP-1 and observed distributions based on 1961–89 ($N = 29$). The raw NCEP-1 and observed June values for 1961 are then withheld. The rescaled value for June 1961 is determined from the NCEP-1 and observed distributions based on the years 1960 and 1962–89. The same procedure is followed for June of 1962, 1963, 1964, etc. The rescaled NCEP-1 value for any given year is hence independent of the observed and raw NCEP-1 values for that year. The skill measurement is the squared correlation between the cross-validated, rescaled NCEP-1 and observed time series ($N = 30$). The cross-validated monthly correlation fields (Fig. 13) have a spatial structure similar to that shown in Fig. 8. As expected, the squared correlations are lower. Depending on the month, from 5% to 12% fewer grid cells than shown in Table 2 have a squared correlation exceeding 0.50.

Results are further summarized in terms of field correlations. For each year, the field of observed precipitation over the pan-Arctic drainage was correlated against the field of cross-validated, rescaled NCEP-1 precipitation and also against observed climatology. A different climatology field was calculated for each year by withholding the data for that year. For example, the climatology field for June 1985 is based on grid cell means using all Junes except 1985. The mean field correlations based on the cross-validated, rescaled NCEP-1 data range from 0.80 in December to slightly more than 0.50 in July (Fig. 14). For every month, these are about 0.10 higher than those based on climatology. The improvement over climatology gained by the use of the rescaled NCEP-1 data is hence fairly modest.

An obvious issue with the rescaling approach is that with the short record lengths ($N = 29$ for the cross

Obs. vs. Rescaled NCEP1 Precip.

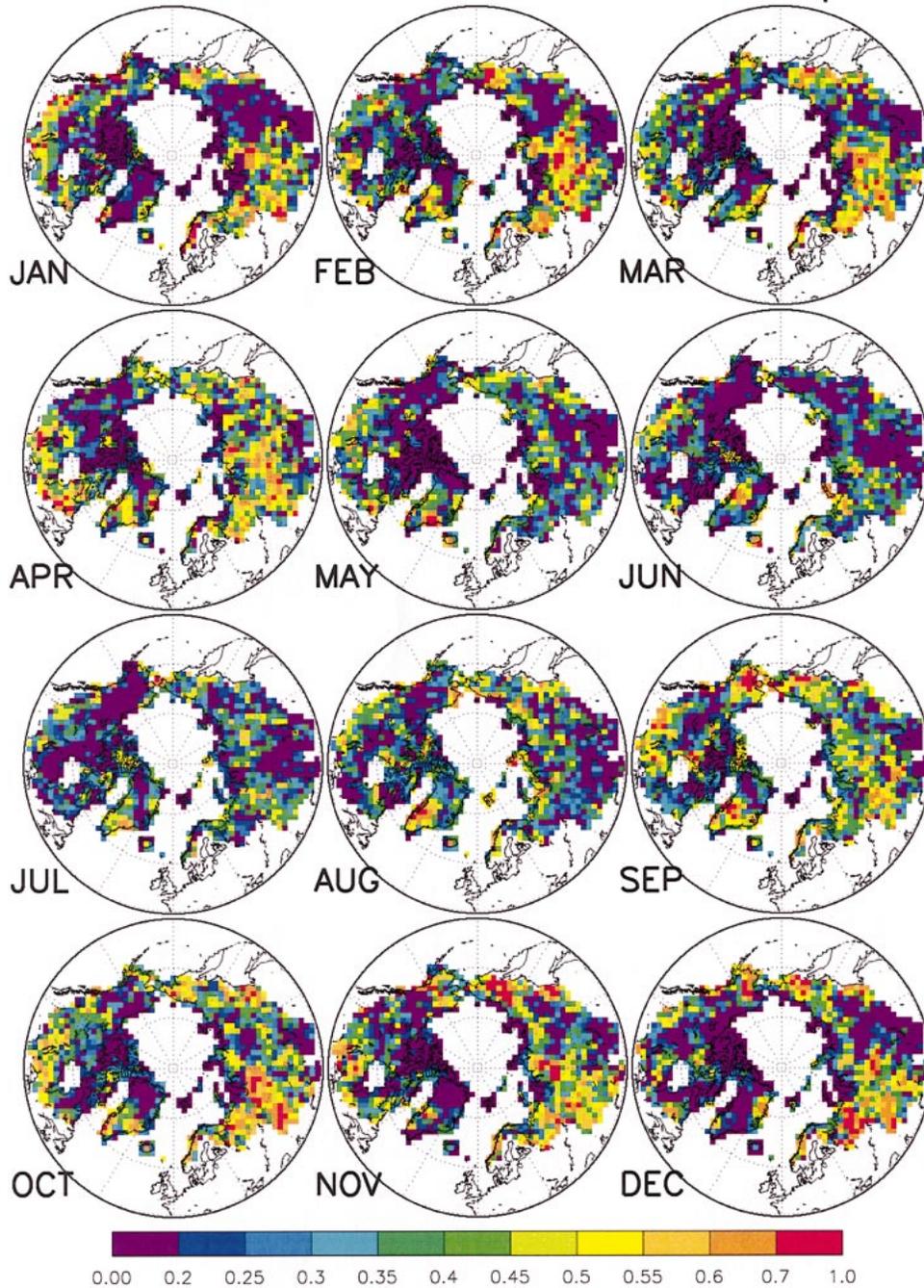


FIG. 13. Cross-validated squared correlations between observed monthly precipitation and rescaled NCEP-1 precipitation over the period 1960–89 based on a probability transformation.

validation) the observed and NCEP-1 precipitation distributions are inadequately defined. In particular, one expects that the extremes will not be well captured. In recognition, a rerun was conducted using observed and NCEP-1 precipitation for the longer period 1950–89. Results were worse than those just described, probably because of the very sparse assimilation database for the

NCEP-1 model in the 1950s. An alternative was also examined, adjusting the raw NCEP-1 values by: 1) expressing the NCEP-1 value as a z score (the anomaly divided by the standard deviation) with respect to the NCEP-1 mean and standard deviation; 2) rescaling by multiplying the z score by the standard deviation of observed precipitation and adding the mean observed

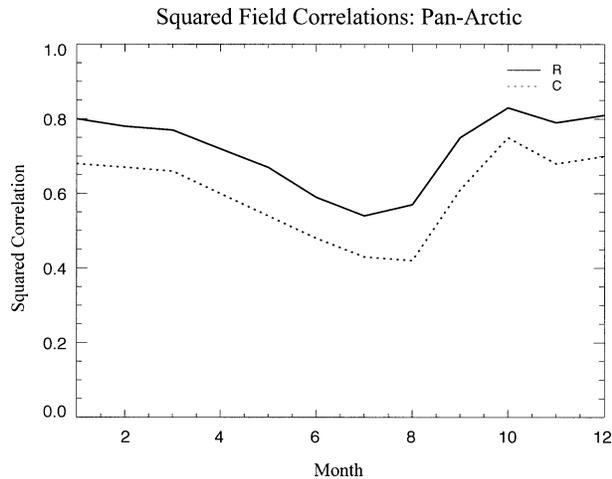


FIG. 14. Cross-validated squared field correlations for the Arctic drainage between observed and rescaled NCEP-1 precipitation (solid line) and between observed precipitation and the observed precipitation climatology (dotted line). Results are averaged over the 1960–89 period.

precipitation. When evaluated through cross validation, the squared correlations using this approach are slightly higher than in Fig. 13. However, the approach has the habit of yielding occasional negative precipitation values, which is avoided with the probability transformation.

5. Data assimilation and use of $P - ET$

a. Procedures

There are large areas where the cross-validated NCEP-1 product has little skill. With the aim of using NCEP-1 in a monitoring scheme, three experiments were performed. The experiments focus on the Eurasian sector of the Arctic drainage (20° – 180° E longitude) for which the Groisman FSU and GHCN datasets provide some station coverage in the 1990s:

Experiment 1 (R + C): Start with time series initialized to climatology, and replace climatology (C) by rescaled NCEP-1 precipitation time series (R) for grid cells where its performance beats climatology;

Experiment 2 (R + C + A): Start with time series initialized to climatology, replace climatology by rescaled NCEP-1 precipitation time series for grid cells where its performance beats climatology (as in experiment 1), but then further adjust the resulting time series through assimilating station observations, using a degraded network representative of what is expected to be available over coming decades;

Experiment 3: Modify the approach described in section 4 of rescaling of NCEP-1 precipitation forecasts by including a second variable (rescaled aerological estimates of $P - ET$). Specifically, replace

the rescaled NCEP-1 precipitation time series with precipitation time series based on rescaled $P - ET$ where the latter perform better. Assess the performance of this “enhanced” basic dataset against the original dataset based on rescaled NCEP-1 precipitation alone, then use the “enhanced” dataset in reruns of experiments 1 and 2.

The observed climatology used for initialization is the same as that used for Fig. 14: the climatology for a given grid cell, month, and year is based on withholding the data for that year, and calculating the mean based on all other years.

The data assimilation experiments make use of the Eurasian station coverage available for 1996. Within this sector, there were a total of 246 stations in 1996 with at least one month of data. This represents about 50% of the network available during the 1980s. All precipitation records over the period 1960–89 corresponding to the 1996 station network were extracted. The station data were then gridded in a similar fashion to that described in section 3: (a) if at least four stations were found in a grid cell (rarely the case), the cell value was taken as the simple average of the stations in the cell; (b) if there were less than four stations in the cell, a cell value was obtained via the Shepard interpolation using the four closest stations within two grid lengths (350 km) of the grid box center; (c) if there were less than four stations within two grid lengths, no value was obtained. This approach provided values at 57% of grid cells in the Eurasian sector. For all data assimilation experiments, a simple replacement strategy is used, where existing grid cell precipitation values based variously on climatology, rescaled NCEP-1 precipitation or rescaled $P - ET$ (depending on the experiment, see following discussion) are swapped out in favor of the interpolated station values.

For experiment 1, for each grid cell and month, the root-mean-square error (rmse) was calculated from the differences between the time series of observations and the cross-validated rescaled NCEP-1 precipitation. A corresponding rmse was then calculated between the observed precipitation time series and climatology. If the former rmse was smaller, the climatology time series for the grid cell was replaced by the rescaled NCEP-1 time series. Otherwise, the climatology time series was retained. Experiment 2 starts with the results from experiment 1. As just discussed, if an observed precipitation value is available, it replaces the existing grid cell value.

Regarding experiment 3, several studies (Bromwich et al. 2000; Cullather et al. 2000; Rogers et al. 2001; Serreze et al. 2003) have demonstrated the utility of reanalysis-derived estimates of $P - ET$ in studies of the Arctic hydrologic budget. Here $P - ET$ is calculated from the convergence of the vertically integrated moisture flux, adjusted by the time change in atmospheric precipitable water. Calculated $P - ET$ should not be

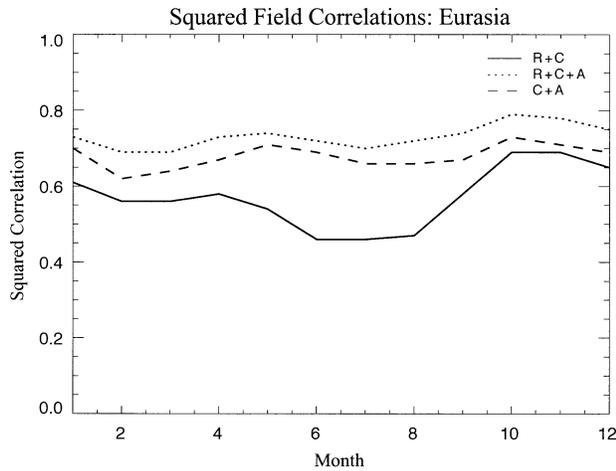


FIG. 15. Cross-validated squared field correlations for Eurasia between observed precipitation and (R + C, solid line) replacement of climatology by rescaled NCEP-1 precipitation; (R + C + A, dotted line) replacement of climatology by rescaled NCEP-1 precipitation and assimilation of a partial station database; (dashed line, C + A) climatology with assimilation of a partial station database.

confused with forecasted $P - ET$, based on the forecasts of each variable. Fields of the squared correlation between observed precipitation and calculated $P - ET$ have a spatial structure similar to Fig. 8. Time series of observed precipitation and calculated $P - ET$ are positively correlated (Serreze et al. 2003).

Monthly time series of calculated $P - ET$ for the period 1960–89 were translated to the 175-km grid array via a Cressman interpolation. As P and $P - ET$ are correlated, alternative estimates of P (termed here P1) can be obtained to complement those based on the NCEP-1 forecasts. The P1 values were obtained by rescaling grid cell $P - ET$ against the observed precipitation time series, using the same cross validation scheme as described earlier. At each grid cell and for each month, the rmse was calculated from the differences (1960–89) between observed precipitation and the P1 values. A corresponding rmse was then calculated from the differences between the observed precipitation and rescaled NCEP-1 precipitation. If the former rmse was smaller, the rescaled NCEP-1 value was replaced by the P1 value. As evaluated over the entire pan-Arctic domain, $P - ET$ performs better than NCEP-1 precipitation (as evaluated from the rmse or the squared correlation) for 20%–40% of grid cells. The improvement is most striking in the winter months.

b. Results

We focus first on results from experiments 1 and 2. Figure 15 gives mean field correlations over the Eurasian sector by month over the 1960–89 period based on the cross-validated rescaled NCEP-1 values with replacement by climatology (R + C, experiment 1), and with additional replacement by assimilation of gridded

TABLE 3. Squared correlations by season for the Ob, Yenisey, and Lena watersheds between time series (1960–89) of observed precipitation from rescaling (R); from rescaling and replacement of climatology (R + C); from rescaling, replacement of climatology, and data assimilation (R + C + A). Values in parentheses include $P - ET$ within the initial rescaling procedure.

Watershed	Ob	Yenisey	Lena
Winter			
R	0.36 (0.53)	0.00 (0.05)	0.07 (0.21)
R + C	0.43 (0.59)	0.01 (0.07)	0.12 (0.20)
R + C + A	0.52 (0.66)	0.04 (0.17)	0.25 (0.39)
Spring			
R	0.45 (0.51)	0.49 (0.62)	0.41 (0.54)
R + C	0.48 (0.55)	0.57 (0.69)	0.44 (0.58)
R + C + A	0.59 (0.65)	0.59 (0.72)	0.68 (0.79)
Summer			
R	0.27 (0.42)	0.35 (0.50)	0.67 (0.73)
R + C	0.24 (0.41)	0.34 (0.46)	0.78 (0.83)
R + C + A	0.37 (0.49)	0.65 (0.70)	0.94 (0.96)
Autumn			
R	0.43 (0.55)	0.55 (0.61)	0.38 (0.51)
R + C	0.47 (0.56)	0.62 (0.67)	0.44 (0.52)
R + C + A	0.60 (0.67)	0.69 (0.74)	0.61 (0.68)

station data (R + C + A, experiment 2). Also shown for comparisons are results from simply replacing climatology with the gridded station data (C + A, this can be considered as a subexperiment). As expected, the correlations based on R + C + A are substantially higher than those for R + C, that is, data assimilation has obvious benefits. The improvement is greatest in summer. While the approach shows promise, and might be improved with more elegant approaches to assimilating station precipitation data, note that almost the same return is obtained by simply replacing climatology with assimilation of data from the degraded network (C + A).

Table 3 lists cross-validated squared correlations between the observed time series and from rescaling the NCEP-1 precipitation (R, the “standard” technique outlined in section 4) and from R + C and R + C + A (experiments 1 and 2). The values in parentheses are from experiment 3, for which the rescaled NCEP-1 grid cell time series are first swapped out with the rescaled $P - ET$ time series (the P1 values) for grid cells where the latter performs better. All results are based on aggregating grid cells over the three major Eurasian watersheds (the Ob, Yenisey, and Lena). This watershed aggregation is useful in that one of the goals is to use the precipitation reconstructions within Arctic-RIMS to assess variability in river discharge to the Arctic Ocean.

Performance is quite mixed. For the Lena in summer, the simple rescaling approach (R) yields a squared correlation of 0.67, rising to 0.78 for R + C, and to 0.94 for R + C + A. We consider this to be quite good. By contrast, winter results for the Yenisey and Lena are extremely poor, as are the summer results for the Ob. Some of these points will be returned to shortly. Note,

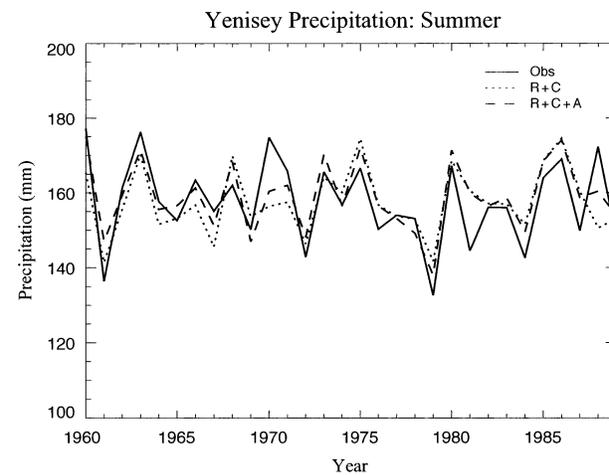
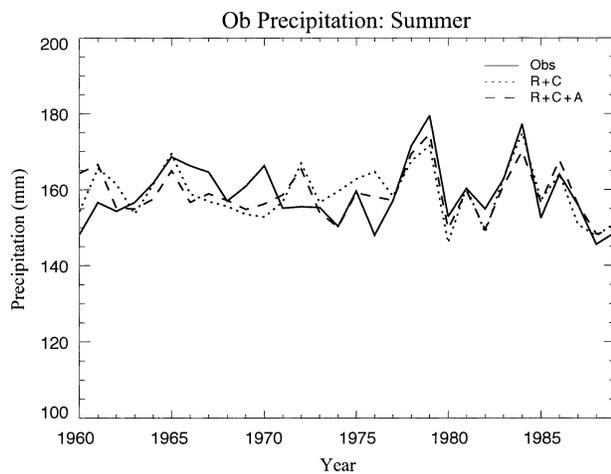
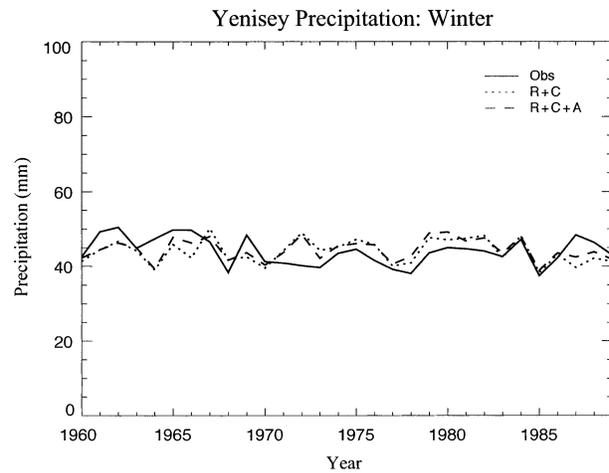
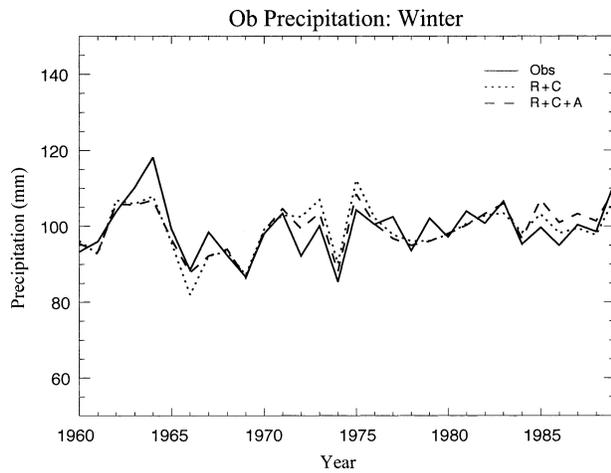


FIG. 16. Time series of basin-averaged precipitation for the Ob watershed for (top) winter and (bottom) summer based on: (solid line, obs) observations; (dotted line, R + C) replacement of climatology by rescaled NCEP-1 precipitation and $P - ET$; (dashed line, R + C + A) replacement of climatology by rescaled NCEP-1 precipitation, $P - ET$, and assimilation of a partial station database.

FIG. 17. Time series of basin-averaged precipitation for the Yenisey watershed for (top) winter and (bottom) summer based on: (solid line, obs) observations; (dotted line, R + C) replacement of climatology by rescaled NCEP-1 precipitation and $P - ET$; (dashed line, R + C + A) replacement of climatology by rescaled NCEP-1 precipitation, $P - ET$, and assimilation of a partial station database.

however, the improvement that is gained by including $P - ET$ in the initial rescaling (experiment 3, values in parentheses). While including $P - ET$ still yields poor winter results in the Yenisey, for other basins and seasons the increase in skill is striking. For example, for winter in the Ob, including $P - ET$ increases the squared correlation based on R by 0.17 and on R + C + A by 0.14.

Figures 16–18 provide winter and summer time series of precipitation averaged over the three drainage basins. These are based on results from experiment 3 that include replacement by $P - ET$ in the rescaling step. The R + C dataset captures the major aspects of the observed time series, but clearly has some problems, especially with regard to the extremes. The very good performance in the Lena basin for summer, even without data assimilation, is readily seen. As noted above, performance

over the Yenisey and Lena in winter is quite poor, even for R + C + A. It is important to point out, however, that the observed variability in winter precipitation for the Yenisey is itself rather low (the time series is quite flat, with extremes ranging between 40 and 50 mm), such that the squared correlations are especially sensitive to errors in both the observations and reanalysis output. The Lena is also not particularly variable during winter. There appears to be a trend in Lena basin winter precipitation, which is not captured in the reconstructions.

6. Summary and discussion

Our results paint a rather sobering picture of the ability to monitor precipitation over the Arctic drainage system and to provide historic time series. The station

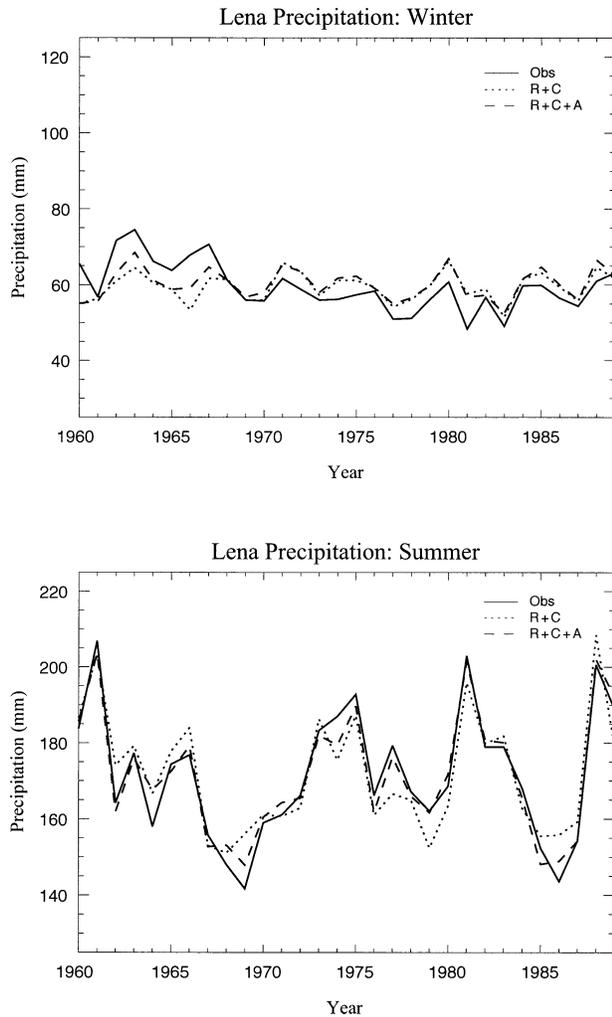


FIG. 18. Time series of basin-averaged precipitation for the Lena watershed for (top) winter and (bottom) summer based on: (solid line, obs) observations; (dotted line, R + C) replacement of climatology by rescaled NCEP-1 precipitation and $P - ET$; (dashed line, R + C + A) replacement of climatology by rescaled NCEP-1 precipitation, $P - ET$, and assimilation of a partial station database.

network is sparse. Except for a few areas, the station density required to provide accurate gridded time series is not met, even at a fairly coarse 175-km resolution. This is exacerbated by problems of gauge undercatch, and the lack of consensus regarding bias adjustment techniques. Particularly troublesome is degradation of the observing network since about 1990 due to station closure and, in Canada, a trend toward automation. It is this very degradation of the network that merits consideration of approaches such as presented here in developing a monitoring strategy.

Precipitation forecasts from the constantly updated NCEP-1 reanalysis, when rescaled to eliminate systematic biases, represent a useful starting point. If climatology is replaced by rescaled NCEP-1 values where the latter beats climatology, the basic spatial structure of

observed precipitation is captured. With notable exceptions (e.g., the Yenisey in winter), the basic time series structure can be reproduced at the large watershed scale. Assimilating data from a partial network of stations typical of that expected to be available in the coming decade (50% of the 1980s coverage) improves skill, as does replacing rescaled NCEP-1 precipitation forecasts with rescaled $P - ET$ where the latter performs better. These findings, however, must be tempered by recognition that the observed time series used to validate via correlation analysis have significant shortcomings.

An alternative approach has been examined in which precipitation is reconstructed via multiple linear regression (a downscaling approach), using as predictors the NCEP-1 monthly precipitation forecasts along with other reanalysis variables such as computed $P - ET$ from wind and humidity profiles, monthly sums of upward vertical velocity (ω) at 500 hPa, zonal and meridional moisture fluxes, sea level pressure, and a measure of lower-tropospheric stability. The output is then rescaled using the observed 1960–89 time series. The apparent skill is comparable to that based on the rescaling approach using NCEP-1 precipitation and $P - ET$. There are issues of colinearity between predictors. While there are methods to resolve these issues, the rescaling approach is on a better statistical footing in that, unlike regression, it does not assume that the observed time series are themselves accurate. Only the statistical distributions need be known. The rescaling approach also has the distinct advantage of simplicity and easier portability to next-generation reanalysis systems (discussed shortly). We have also used the rescaling approaches to reconstruct precipitation at the station locations, with subsequent interpolation of the reconstructed station values to the 175-km grid cell array. In general, the results are worse than those based on first interpolating the station data to the grid cell array.

One way to improve the results (and better assess performance) is to assemble a better historical station database. Our understanding is that a significant amount of data for Russia remains to be digitized. Data rescue is a planned activity under the National Science Foundation program for the Study of Environmental Arctic Change (SEARCH) (SEARCH SSC 2001). Another SEARCH activity is to resolve the problem of bias adjustments. Some rapid updates of high-latitude data (daily), which could be used for assimilation, are being provided through the Global Telecommunications System, but there are serious quality control issues.

However, as is evident from the Monte Carlo simulations, we will never have enough station data. The best avenue for improvement is a better atmospheric reanalysis. NCEP is planning a new global reanalysis, but will meanwhile provide updates with the existing system (Kistler et al. 2001). NCEP is also undertaking a North American Regional Reanalysis (NARR). Efforts will be needed to evaluate the high-latitude performance

of the NARR. The SEARCH program envisions undertaking a dedicated Arctic System Reanalysis (ASR). Initial efforts toward development of the ASR are already under way.

The best hope for the near future is the European Centre for Medium-Range Weather Forecasts reanalysis, known as ERA-40. Data will be made available at the National Center for Atmospheric Research in Boulder, Colorado. Current plans are to provide updates as is done with the NCEP system. We have had the opportunity to examine several years of Arctic precipitation data from an early run of ERA-40. The precipitation forecasts are not greatly improved over the earlier ERA-15 effort (1979–93), but are much better than those from NCEP-1.

Until ERA-40 data become available, reconstructions for Arctic-RIMS will use NCEP-1. The RIMS system as it presently stands uses the product based on statistical downscaling. It is updated on a monthly basis. The monitoring system will either migrate to the simpler (and better) rescaling approach (with assimilation of available station updates), or will generate both products in parallel. Presently, rescaling is performed using 1960–89 data. It might be better to use longer records to get a fuller distribution. For example, one could use observed precipitation distributions from 1950 to 1989 and reanalysis distributions for 1960 to 1999.

Reconstructed precipitation fields represent both a standard Arctic-RIMS product and input into a permafrost/water balance model. For use in the water balance model daily fields are desired. Temporal disaggregation also makes use of the NCEP-1 precipitation forecasts. Briefly, the daily reanalysis totals are interpolated to the 175-km grid and expressed as a fraction of the monthly reanalysis total at the grid. Adjusted daily totals are obtained by multiplying the reconstructed monthly totals by each of the daily NCEP-1 fractions. To avoid “drizzle” events, the NCEP-1 daily fractions are only computed on the basis of daily events exceeding a selected threshold. The same technique is used to obtain daily totals from historical monthly time series. The approach assumes that the daily reanalysis totals are suspect in terms of absolute magnitude but that the relative magnitudes of daily events are more correct. A problem is the spectral noise (“spectral snow”) in the NCEP-1 output. This is presently addressed through smoothing the NCEP-1 output. Again, better results are anticipated using ERA-40.

Acknowledgments. This study was supported by the National Science Foundation under the Arctic System Science program Grants OPP-9732461, OPP-9614297, OPP-9910315, OPP-9906906, OPP-9905381, OPP-9910334, OPP-0229769 and NASA Contracts NAG5-6820 and NAG5-9596. Sheng-Hung Wang is thanked for data processing. Nancy Geiger-Wooten is thanked for providing the graphics.

REFERENCES

- Aagaard, K., and E. C. Carmack, 1989: The role of sea ice and other fresh waters in the Arctic circulation. *J. Geophys. Res.*, **94** (C10), 14 485–14 498.
- Armstrong, R. L., and M. J. Brodzik, 1995: An earth-gridded SSM/I data set for cryospheric studies and global change monitoring. *Adv. Space Res.*, **16**, 155–163.
- Bras, R. L., and I. Rodriguez-Iturbe, 1976a: Evaluation of mean square error involved in approximating the areal average of a rainfall event by a discrete summation. *Water Resour. Res.*, **12**, 181–184.
- , and —, 1976b: Network design for the estimation of areal mean rainfall events. *Water Resour. Res.*, **12**, 1185–1195.
- Broecker, W. S., 1997: Thermohaline circulation, the Achilles heel of our climate system: Will man-made CO₂ upset the current balance? *Science*, **278**, 1582–1588.
- Bromwich, D. H., R. I. Cullather, and M. C. Serreze, 2000: Reanalyses depictions of the Arctic atmospheric moisture budget. *The Fresh Water Budget of the Arctic Ocean*, E. L. Lewis, Ed., Kluwer Academic, 163–196.
- Cavalieri, D. J., P. Gloersen, C. L. Parkinson, J. C. Comiso, and H. J. Zwally, 1997: Observed hemispheric asymmetry in global sea ice changes. *Science*, **278**, 1104–1106.
- Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374.
- Cullather, R. I., D. H. Bromwich, and M. C. Serreze, 2000: The atmospheric hydrologic cycle over the Arctic basin from reanalyses. Part I: Comparison with observations and previous studies. *J. Climate*, **13**, 923–937.
- Dickson, R. R., and Coauthors, 2000: The Arctic Ocean response to the North Atlantic Oscillation. *J. Climate*, **13**, 2671–2696.
- Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation measurement intercomparison. WMO Tech. Doc. 872, World Meteorological Organization, Geneva, Switzerland, 212 pp.
- Groisman, P. Y., 1998: National Climatic Data Center data documentation for TD-9816, Canadian monthly precipitation. National Climatic Data Center, Asheville, NC, 21 pp.
- , V. V. Koknaeva, T. A. Belokrylova, and T. R. Karl, 1991: Overcoming biases of precipitation: A history of the USSR experience. *Bull. Amer. Meteor. Soc.*, **72**, 1725–1733.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kistler, R., and Coauthors, 2001: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267.
- Lammers, R. B., A. I. Shiklomonov, C. J. Vorosmarty, B. M. Fekete, and B. J. Peterson, 2001: Assessment of contemporary Arctic river runoff based on observational records. *J. Geophys. Res.*, **106** (D4), 3321–3334.
- Legates, D. R., and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.*, **10**, 111–127.
- McDonald, R. W., E. C. Carmack, F. A. McLaughlin, K. K. Falkner, and J. H. Swift, 1999: Connections among ice, runoff and atmospheric forcing in the Beaufort Sea. *Geophys. Res. Lett.*, **26**, 2223–2226.
- Mekis, E., and W. D. Hogg, 1999: Rehabilitation and analysis of Canadian daily precipitation time series. *Atmos.–Ocean*, **37**, 53–85.
- Metcalfe, J. R., B. Routledge, and K. Devine, 1997: Rainfall measurement in Canada: Changing observational methods and archive adjustment procedures. *J. Climate*, **10**, 92–101.
- Michaelson, J., 1987: Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteor.*, **26**, 1589–1600.
- Moritz, R. E., C. M. Bitz, and E. J. Steig, 2002: Dynamics of recent climate change in the Arctic. *Science*, **297**, 1497–1502.
- Oechel, W. C., S. J. Hastings, G. Vourlitis, M. Jenkins, G. Richers, and N. Guilke, 1993: Recent change of Arctic tundra ecosystems

- from a net carbon dioxide sink to a source. *Nature*, **361**, 520–523.
- Pan, H.-L., and W.-S. Wu, 1994: Implementing a mass flux convection parameterization package for the NMC medium range forecast model. Preprints, *10th Conf. on Numerical Prediction*, Portland, OR, Amer. Meteor. Soc., 96–98.
- Panofsky, H. A., and G. W. Brier, 1963: *Some Applications of Statistics to Meteorology*. Mineral Industries Continuing Education, College of Mineral Industries, The Pennsylvania State University, 224 pp.
- Rodriguez-Iturbe, I., and J. M. Mejia, 1974: On the transformation of point rainfall to areal rainfall. *Water Resour. Res.*, **10**, 729–735.
- Rogers, A. N., D. H. Bromwich, E. N. Sinclair, and R. I. Cullather, 2001: The atmospheric hydrologic cycle over the Arctic Basin from reanalyses. Part II: Interannual variability. *J. Climate*, **14**, 2414–2429.
- Running, S. W., J. B. Way, K. McDonald, J. Kimball, S. Frolking, A. R. Keyser, and R. Zimmermann, 1999: Radar remote sensing proposed for monitoring freeze-thaw transitions in boreal regions. *Eos, Trans. Amer. Geophys. Union*, **80**, 213–221.
- SEARCH SSC, 2001: Study of environmental Arctic change, science plan. Polar Science Center, Applied Physics Laboratory, University of Washington, Seattle, WA, 91 pp.
- Serreze, M. C., and C. M. Hurst, 2000: Representation of mean Arctic precipitation from NCEP–NCAR and ERA reanalyses. *J. Climate*, **13**, 182–201.
- , J. R. Key, J. C. Box, and J. A. Maslanik, 1998: A new monthly climatology of global radiation for the Arctic and comparisons with NCEP/NCAR reanalysis and ISCCP-C2 fields. *J. Climate*, **11**, 121–136.
- , and Coauthors, 2000: Observational evidence of recent change in the northern high-latitude environment. *Climatic Change*, **46**, 159–207.
- , A. H. Lynch, and M. P. Clark, 2001: The Arctic frontal zone as seen in the NCEP/NCAR reanalysis. *J. Climate*, **14**, 1550–1567.
- , D. H. Bromwich, M. P. Clark, A. Etringer, T. Zhang, and R. Lammers, 2003: The large-scale hydro-climatology of the terrestrial Arctic drainage. *J. Geophys. Res.*, **108** (D2), 8160, doi: 10.1029/2001JD000919.
- Shepard, D., 1984: Computer mapping: The SYMAP interpolation algorithm. *Spatial Statistics and Models*, G. L. Gaile and C. J. Willmott, Eds., D. Reidel, 133–145.
- Steele, M., and T. Boyd, 1998: Retreat of the cold halocline layer in the Arctic Ocean. *J. Geophys. Res.*, **103**, 10 419–10 435.
- Tait, A. B., 1998: Estimation of snow water equivalent using passive microwave radiation data. *Remote Sens. Environ.*, **64**, 286–291.
- Thompson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–1300.
- , and ———, 2000: Annual modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000–1016.
- Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. Eischeid, 1992: The global historical climatology network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. Tech. Rep. ORNL/CDIAC-53, NDP-041, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 99 pp. plus appendixes.
- Willmott, C. J., C. M. Rowe, and W. D. Philpot, 1985: Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *Amer. Cartogr.*, **12**, 5–16.
- Yang, D., 1999: An improved precipitation climatology for the Arctic Ocean. *Geophys. Res. Lett.*, **26**, 1525–1628.
- , and Coauthors, 2001: Compatibility evaluation of national precipitation gauge measurements. *J. Geophys. Res.*, **106** (D2), 1481–1491.
- Zhang, T., R. L. Armstrong, and J. Smith, 1999: Detecting seasonally frozen soils over snow-free land surface using satellite passive microwave remote sensing data. Preprints, *Fifth Conf. on Polar Meteorology and Oceanography*, Dallas, TX, Amer. Meteor. Soc., 355–357.

