Uncertainty in Temperature and Precipitation Datasets over Terrestrial Regions of the Western Arctic

Sheldon Drobot,* James Maslanik, Ute Christina Herzfeld, and Charles Fowler

University of Colorado, Boulder, Colorado

Wanli Wu

National Center for Atmospheric Research, Boulder, Colorado

Received 9 January 2006; accepted 21 August 2006

ABSTRACT: A better understanding of the interannual variability in temperature and precipitation datasets used as forcing fields for hydrologic models will lead to a more complete description of hydrologic model uncertainty, in turn helping scientists study the larger goal of how the Arctic terrestrial system is responding to global change. Accordingly, this paper investigates temporal and spatial variability in monthly mean (1992–2000) temperature and precipitation datasets over the Western Arctic Linkage Experiment (WALE) study region. The six temperature datasets include 1) the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5); 2) the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40); 3) the Advanced Polar Pathfinder all-sky temperatures (APP); 4) National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalyses

* Corresponding author address: Dr. Sheldon Drobot, Colorado Center for Astrodynamics Research, University of Colorado, UCB 431, Boulder, CO 80309-0431. E-mail address: drobot@colorado.edu
(NCEP1); 5) the Climatic Research Unit/University of East Anglia CRUTEM2v (CRU); and 6) the Matsuura and Wilmott 0.5° × 0.5° Global Surface Air Temperature and Precipitation (MW). Comparisons of monthly precipitation are examined for MM5, ERA-40, NCEP1, CRU, and MW. Results of the temporal analyses indicate significant differences between at least two datasets (for either temperature or precipitation) in almost every month. The largest number of significant differences for temperature occurs in October, when there are five separate groupings; for precipitation, there are four significantly different groupings from March through June, and again in December. Spatial analyses of June temperatures indicate that the greatest dissimilarity is concentrated in the central portion of the study region, with the NCEP1 and APP datasets showing the greatest differences. In comparison, the spatial analysis of June precipitation datasets suggests that the largest dissimilarity is concentrated in the eastern portion of the study region. These results indicate that the choice of forcing datasets likely will have a significant effect on the output from hydrologic models, and several different datasets should be used for a robust hydrologic assessment.

**KEYWORDS:** Arctic; Temperature; Precipitation

1. **Introduction**

One of the central themes of the Western Arctic Linkage Experiment (WALE) is to investigate uncertainties in regional hydrology and carbon estimates with respect to variations in different driving datasets. This will provide scientists with a better understanding of how the different datasets influence the hydrology models, leading to a more complete description of model uncertainty. Such analyses are critical to understanding the larger WALE goal of determining how the Arctic terrestrial system is responding to global change.

Although there are numerous datasets that could be used to drive hydrologic models, very little research has focused on assessing the degree of similarity between these datasets. Among the notable exceptions is research from the Arctic Climate Impact Assessment (ACIA 2005), where some comparisons between temperature for the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalyses (NCEP1; Kalnay et al. 1996), 15-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analyses (ERA-15; Gibson et al. 1997), and the Climatic Research Unit/University of East Anglia CRUTEM2v (CRU; Jones et al. 2001) datasets were analyzed. The results indicated that temperature differences between the NCEP1 and CRU datasets were largest in winter and smallest in summer, with NCEP1 being warmer over North America; comparisons for NCEP1 and ERA-15 were similar, whereas ERA-15 was noticeably warmer than CRU in spring (ACIA 2005).

Further temporal and spatial analyses of Arctic temperature and precipitation datasets are needed to better inform the scientific community regarding uncertainty in these key data sources. Therefore, this paper quantitatively assesses temporal and spatial variations in mean monthly temperature and precipitation datasets over the WALE study region (Figure 1) from 1992 to 2000. The study region includes most land areas from 55° to 65°N and from −160° to −110°E.
Datasets used for temperature comparisons include the NCEP1, CRU, Advanced Polar Pathfinder all-sky temperatures (APP; Fowler et al. 2002), 40-yr ECMWF Re-Analysis (ERA-40), Matsuura and Wilmott 0.5° × 0.5° Global Surface Air Temperature and Precipitation (MW; http://rims.unh.edu/data/read_me.cgi?category=7&subject=4), and fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5; Wu et al. 2006, manuscript submitted to Earth Interactions, hereafter WDHML). For precipitation, datasets include NCEP1, CRU, ERA-40, MW, and MM5.

The following research questions drive this analysis.

- Are mean monthly temperature and precipitation fields from 1992 to 2000 over the WALE study region significantly different among the datasets? If so, which datasets are statistically different in which months?

Figure 1. The study region includes most land areas between 55° and 65°N, −160° and −110°E. This area is shaded in light gray.
Are mean monthly temperature and precipitation anomaly fields from 1992 to 2000 significantly correlated? If not, which datasets are statistically different?

Which regional differences contribute most to the statistical differences observed in the time series analysis?

Analysis methods used in this study include analysis of variance (ANOVA) with post hoc means comparisons to determine significant differences in the datasets, anomaly correlations to examine seasonal cycles, and similarity maps to highlight spatial regions where datasets differ.

2. Datasets

Each of the datasets analyzed here is used to validate or force hydrological models in the WALE project. The NCEP1 and ERA-40 are data-assimilated products, MW and CRU data are interpolated from station data, APP data are derived from satellite radiances, and MM5 output is based on model simulations forced by NCEP1 data at the boundary.

The NCEP1 data are on a 2.5° global grid and data were obtained online (http://www.cdc.noaa.gov/cdc/reanalysis/). They are generated by reanalyzing historical data with a state-of-the-art data assimilation model. Reanalysis provides a relatively consistent source for atmospheric data because the data are based on a static data assimilation scheme, the data analysis is global in extent, and many observations are used. Data quality has been discussed in numerous studies (Kalnay et al. 1996; Serreze et al. 1998; Serreze and Hurst 2000; Serreze et al. 2003).

In general, the reanalysis data capture the broad spatial patterns, but it overestimates annual total precipitation over land and has the seasonal precipitation maximum one month early over the Arctic (Serreze and Hurst 2000; Serreze et al. 2005).

Reanalyses also form the basis for the ERA-40 dataset. As with the NCEP1 fields, ERA-40 reanalyses provide a relatively consistent source for atmospheric data. For temperature, the ERA-40 system should be superior to those from NCEP1, and for precipitation, ERA-40 is also considered better than NCEP1 because it has smaller errors, captures large-scale patterns of precipitation more accurately, and is superior in depicting interannual variability (Serreze et al. 2005).

Based largely on station data, the MW dataset is compiled from numerous other datasets, including the Global Historical Climatology Network (Peterson et al. 1998), the Global Synoptic Climatology Network (National Climatic Data Center Dataset 9290c), and the Global Surface Summary of Day (http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=gsod.html). Monthly mean surface air temperature and precipitation were regridded onto 0.5° × 0.5° grids through the spherical version of Shepard’s algorithm, which employs an enhanced distance-weighting method (Shepard 1968; Willmott et al. 1985). Elevation influences were also incorporated as discussed in Willmott and Matsuura (Willmott and Matsuura 1995), and climatologically aided interpolation (CAI) was used to improve the spatial interpolation procedure (Willmott and Robeson 1995).

Similar to MW, CRU data are based on station data. The CRU data are obtained on a monthly, global 5° × 5° grid and data fields are based on variance-adjusted interpolation of land surface air temperature and sea surface temperature data. The
variance-adjusted interpolation damps artificial variance changes and has the greatest effect over data-sparse regions. More details are presented in Jones et al. (Jones et al. 2001).

The Advanced Very High Resolution Radiometer (AVHRR) Polar Pathfinder project all-sky skin temperatures used in this paper are from the 25-km² dataset described in Fowler et al. (Fowler et al. 2002) and Wang and Key (Wang and Key 2005) (as well as online at http://nsidc.org/data/nsidc-0066.html). This product begins initially as twice-daily gridded and calibrated satellite channel data on a 5-km² grid, based on common local solar times and adjusted for scan angle. The reduced-resolution data (25 km²) are derived from the 5 km² via subsampling.

Finally, the MM5 output is from release 3.6. The horizontal domain of the model in this study covers Alaska and western Canada with a grid size of 50 km, consisting of 50 north–south and 80 east–west grid points. The companion paper by WDHML outlines specific features of the MM5 datasets.

Because all datasets are not on common grids, the datasets were first converted to the Northern Hemisphere 25-km² Equal-Area Scalable Earth (EASE) Grid with bilinear interpolation (NCEP1, ERA-40, and CRU) or nearest neighbor (MM5); the APP data are native in EASE Grid. (More details on the EASE projection are available online at http://nsidc.org/data/ease/ease_grid.html.)

3. Analysis methods

3.1. Temporal analysis

To assess whether the observed differences among the different datasets are significant, we employ ANOVA with post hoc means comparisons. The ANOVA method provides a statistical test to determine whether the observed differences in the mean monthly values among all the datasets are significant; if the associated $p$ value of the ANOVA test is less than 0.05, then there is sufficient evidence to suggest that the differences among the observed group means are significant. Unlike most parametric tests, ANOVA is robust even for moderate departures from multivariate normality, particularly when groups are of equal sample size (as is this case here). However, the ANOVA procedure is sensitive to differences in variances of groups, especially for post hoc comparisons. Therefore, before applying the ANOVA tests, we first used Levene’s test of homogeneity of variance to test the ANOVA assumption that each dataset has the same variance. The Levene test is robust with respect to departures from normality, unlike the older Bartlett’s test of homogeneity of variance.

Although the ANOVA $p$ value indicates whether there is evidence to conclude that at least two years are statistically different, it does not provide details on which years are different. To ascertain this information, we employ techniques for multiple comparisons, which perform all pairwise combinations of different years at once. There are numerous multiple-comparisons tests, and there remains considerable debate over which test is most appropriate. If Levene’s test suggests that the datasets have equal variance, we employ Tukey’s Honestly-Significant-Difference (HSD) test, which is the most conservative of the post hoc tests, and therefore it will suggest dataset differences less often than other means comparison tests. If Levene’s test suggests that the assumption of homogeneity of variances is not
valid, we instead employ the Games–Howell (GH) test, which is formulated for unequal variances. Additional details on these tests are available in standard statistical textbooks.

Statistical tests to determine the strength of association among the datasets are based on correlating monthly anomalies. First, each of the datasets was converted to mean monthly anomalies based on individual 1992–2000 means. For example, the January anomalies for the CRU and MM5 datasets are based on the 1992–2000 January means for the CRU and MM5 datasets individually. This standardization removes both the systematic bias and the seasonal cycle and allows comparisons of the monthly anomalies alone.

3.2. Spatial analysis

To examine the spatial distribution of agreement and disagreement between temperature and precipitation datasets, similarity maps are calculated using the map comparison method developed by Herzfeld and Merriam (Herzfeld and Merriam 1990). Similarity maps may be calculated for any number of maps and datasets, and in this paper the method is used to compare temperature (or precipitation) datasets for nine years (1992–2000) simultaneously. The map comparison method is based on an algebraic approach that proceeds by 1) standardizing input values in each map or spatial model, 2) forming a functional of pairwise differences of standardized values, and 3) applying a seminorm to the functional in step 2, for each point in the 25-km$^2$ EASE Grid of the WALE region. The result is a spatial grid model of similarity values, which may be mapped to show areas of similarity versus areas of dissimilarity. In our application, we use a linear transformation to convert that data into the interval [0,1], and all calculations are performed inside a landmask outlining the study area (Figure 1). Similarity values close to 0 indicate good agreement, while higher values indicate poor correspondence among the datasets. A mathematical description of the method is given in the companion paper (Herzfeld et al. 2006), together with a demonstration of a variety of uses of algebraic similarity mapping in arctic system analysis. Applications of the algebraic map comparison method in oceanography, geology, geophysics, and mineral exploration are described in Herzfeld and Sondergard (Herzfeld and Sondergard 1988) and Merriam et al. (Merriam et al. 1993).

4. Results

4.1. Temperature

Are mean monthly temperature fields from 1992 to 2000 over the WALE study region significantly different among the datasets? If so, which datasets are statistically different in which months?

Boxplots of monthly mean temperatures indicate that there is considerable disparity among the datasets during most months (Figure 2). Interannual variability is greatest in winter months, where the boxplots encompass a larger range. In the summer months of June, July, and August, there is little interannual variability in the datasets, although there are systematic differences in the temperature values. Overall, the MW dataset is the warmest and the APP dataset is the coolest, with
the largest difference in mean monthly temperature between these two datasets exceeding 10°C in most months.

To test whether differences in monthly means are statistically significant, we employed the ANOVA with Tukey’s HSD because the Levene test indicated no violation for homogeneity of variance (not shown). Results suggest that differences in datasets vary seasonally and that only the ERA-40 and CRU datasets are not statistically different in at least one month (Table 1). From January through March, there are three separate groupings, with the APP data being significantly colder than the other datasets. The main reason for this is likely related to cloud masking and surface inversions. In winter, the snow-covered surface can appear similar to clouds (especially in the visible wavelengths), and cases where cloudy scenes are mistakenly classified as snow-covered ground often lead to cooler
temperatures. Additionally, the APP data are based on skin temperatures, rather than 2-m air temperatures, and they may be slightly colder due to near-surface inversions.

In March and April, both the APP and MW datasets are significantly different from the other datasets, with MW being warmer and APP still cooler. With the beginning of spring, and the associated melting of snow, the APP dataset is less affected by errors in cloud masking and the surface inversion is reduced, resulting in nonsignificant differences between APP and at least one other dataset throughout the summer. From June through October, the MW dataset remains significantly warmer than the other datasets, and the MM5 dataset is significantly colder than the other datasets. October is notable in that there are five distinct groupings, although several datasets are not significantly different from each other. With the return of snow, the APP dataset is again significantly different from the others. Finally, results for November and December are similar to January through March.

Overall, the CRU and ERA-40 datasets were not significantly different in all 12 months (Table 2), marking the highest number of nonsignificantly different values. In comparison, the APP, MW, CRU, and MW datasets were significantly different in all months. The latter is somewhat surprising given that they are both based on station data, but further research is needed to clarify this issue.

Are mean monthly temperature anomaly fields from 1992 to 2000 significantly correlated? If not, which datasets are statistically different in which months?

For some applications, understanding whether there are significant differences

---

**Table 1. Nonsignificant temperature ranges based on Tukey’s HSD. For each month, datasets that have the same letter are not significantly different.**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>MM5</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>c</td>
<td>c</td>
<td>d</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>NCEP1</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>ERA-40</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>CRU</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>APP</td>
<td>c</td>
<td>c</td>
<td>c</td>
<td>d</td>
<td>c</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>b</td>
</tr>
</tbody>
</table>

**Table 2. Counts of the number of months that temperature datasets were not significantly different.**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MM5</th>
<th>NCEP1</th>
<th>ERA-40</th>
<th>CRU</th>
<th>APP</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MM5</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>NCEP1</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA-40</td>
<td></td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRU</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
in the cycle of monthly anomalies of different datasets will be important. Figure 3 illustrates the monthly anomaly time series for each of the six datasets. There is obvious qualitative agreement, which correlation analyses quantitatively confirm (Table 3). All of the six datasets are significantly correlated with one another at $\alpha = 0.05$. The APP dataset shows the lowest correlations with the other datasets, ranging from a low of 0.70 in comparison with NCEP1 to a high of 0.82 with CRU. This is likely due to the influence of cloud masking and surface inversions in the APP data. Excluding the APP dataset, all other correlation coefficients are over 0.90. Therefore, while there are systematic differences in the datasets, the anomaly

![Figure 3. Temperature anomalies by month. All datasets are significantly correlated at $\alpha = 0.05$.](image)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MM5</th>
<th>NCEP1</th>
<th>ERA-40</th>
<th>CRU</th>
<th>APP</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>0.93</td>
<td>0.95</td>
<td>0.99</td>
<td>0.98</td>
<td>0.79</td>
</tr>
<tr>
<td>MM5</td>
<td>0.91</td>
<td>0.91</td>
<td>0.93</td>
<td>0.91</td>
<td>0.80</td>
</tr>
<tr>
<td>NCEP1</td>
<td></td>
<td>0.97</td>
<td>0.97</td>
<td>0.93</td>
<td>0.70</td>
</tr>
<tr>
<td>ERA-40</td>
<td></td>
<td></td>
<td>0.97</td>
<td>0.97</td>
<td>0.80</td>
</tr>
<tr>
<td>CRU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.82</td>
</tr>
</tbody>
</table>
correlations clearly indicate that each dataset closely resembles the others for temporal analyses of relative differences.

Which regional differences contribute most to the statistical differences observed in the time series analysis? What are probable causes of mismatches between datasets and models?

As shown in Figure 4, some of the largest differences in temperature occur in June. An investigation of spatial distributions of areas of disagreements, or mismatches, between datasets is facilitated by similarity mapping. Areas of high similarity values (above 0.3, as a rule of thumb) indicate problems in either observation or modeling, which may be attributed to geographic factors.

In the comparison of datasets with NCEP1, a maritime–continental antipodal effect appears to exist, with an axis near −130°E, trending southeast. In all cases, similarities are better east of the axis, mostly better than 0.15, whereas in the western part, the disagreement varies among the datasets, generally being highest in the center of a broad arch around the Gulf of Alaska (near 58°N and −140° to −130°E). For NCEP1–ERA-40 similarity values are always below 0.25; for NCEP–CRU and NCEP–MM5 they are below 0.3, except for small areas; but the two comparisons, NCEP1–MW and NCEP–APP, exhibit large regions of mismatches with values above 0.4 and 0.45. This may be due to either topographic effects of the coastal ranges and the Alaskan range (recall that the native spatial resolution of NCEP1 is much less than APP or MW), or to data correction (APP), or data extrapolation/interpolation routines that are used in presence of topographic relief. The pairwise similarity maps derived within the group of the three
datasets MM5, CRU, and ERA-40 generally show good agreement, with values below 0.1 in large areas and 0.15 in most areas.

Similarity maps involving APP data show areas of poor similarity along coastal areas in southeast Alaska, stretching north along the Fairweather and Saint Elias Ranges, with a maximum near Anchorage. The least mismatches exist between APP and MW. In the case of APP comparisons with NCEP, MM5, and ERA-40, and to a lesser degree with CRU, another problematic area lies in the northern-central part of the study area. Comparisons of MW with other data and model sets show smaller areas of poor similarity than the APP comparisons. Areas of dissimilarity are concentrated in the central and northern-central parts of the study area. The best match exists between the CRU and the MW datasets.

4.2. Precipitation

Are mean monthly precipitation fields from 1992 to 2000 over the WALE study significantly different among the datasets? If so, which datasets are statistically different in which months?

Similar to the temperature analyses, boxplots of monthly mean precipitation accumulation demonstrate substantial disagreement (Figure 5). In contrast to temperature, there is a clear seasonal pattern in the discrepancies between the datasets, with much larger differences in summer months. For example, the largest difference for precipitation is almost 70 mm (July), and the small differential is just over 27 mm (February). The summer results are related to differences in how the datasets capture convective precipitation, which is far more common in summer than winter. For example, Serreze and Hurst (Serreze and Hurst 2000) noted a deficiency in the NCEP1 precipitation fields due to excessive convective summer precipitation and enhanced soil water feedbacks, which leads to the higher precipitation values noted here. There also are differences in the maximum and minimum monthly peak precipitation among the datasets, which is further explored in the companion paper (WDHML).

Given the disagreement in the boxplots, it is not surprising that the results from Tukey’s HSD (January–June; August–December) or Games–Howell (July) suggest that there is little agreement among most of the datasets (Table 4). From January through March, no single dataset is significantly different from the others, although there are a large number of groupings. In April, both MW and CRU are significantly different from the other datasets, while in May, June, and July, NCEP1 is significantly larger than the other datasets, owing to the issues discussed in Serreze and Hurst (Serreze and Hurst 2000). In August, September, and November, there are only two separate groupings, while in October and December, CRU is significantly different from the other datasets. Overall, the MM5 and ERA-40 datasets are significantly different in only two months, and CRU and MW are significantly different in only three months (Table 5), which is remarkably different from the temperature analyses, where these datasets were not significantly different five and zero times, respectively. This example points out an interesting feature in the datasets: some datasets are similar in many months for one parameter and not the other (e.g., the above examples), some datasets are more similar than not for both parameters (e.g., ERA-40 and NCEP1), and some datasets are significantly different more often than not for both parameters (e.g., MM5 and CRU).
Figure 5. Boxplots of mean monthly precipitation for four datasets, 1992–2000. There is disagreement on the months of maximum and minimum peak precipitation values.

Table 4. Nonsignificant precipitation ranges based on Tukey’s HSD. For each month, datasets that have the same letter are not significantly different.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>b</td>
<td>c</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>c</td>
</tr>
<tr>
<td>MM5</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>NCEP1</td>
<td>b</td>
<td>b</td>
<td>c</td>
<td>b</td>
<td>c</td>
<td>a</td>
</tr>
<tr>
<td>ERA-40</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>CRU</td>
<td>c</td>
<td>c</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>c</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>MM5</td>
<td>c</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>NCEP1</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>ERA-40</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>CRU</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>c</td>
<td>b</td>
<td>d</td>
</tr>
</tbody>
</table>
Are mean monthly precipitation anomaly fields from 1992 to 2000 significantly correlated? If not, which datasets are statistically different in which months?

Although there is strong disagreement over the mean precipitation values, the anomaly correlations suggest that each of the datasets resembles one another over the seasonal cycle. The precipitation anomaly plot (Figure 6) is qualitatively similar, although not as strong as the case for temperature (Figure 3). Correlation analyses indicate that all datasets are significantly correlated, but that the strength of association for precipitation is much less than for temperature (Table 6). Physically, this is likely due to the spatial nature of precipitation and temperature, with higher spatial variability in precipitation leading to lower correlations. The lowest correlation is between CRU and MM5 (0.57), and the highest is between NCEP1 and ERA-40 (0.86).

Which regional differences contribute most to the statistical differences observed in the time series analysis? What are probable causes of mismatches between datasets and models?

---

**Table 5. Counts of the number of months that precipitation datasets were not significantly different.**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MM5</th>
<th>NCEP1</th>
<th>ERA-40</th>
<th>CRU</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>MM5</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>NCEP1</td>
<td></td>
<td>8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>ERA-40</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 6. Precipitation anomalies by month. All datasets are significantly correlated at \( \alpha = 0.05 \).**
Spatial algebraic map comparisons between pairs of variables, including nine maps of June precipitation input datasets for the years 1992–2000 (see Figure 7), were carried out analogously to the temperature similarity mapping (note that precipitation proxies cannot be derived from brightness temperatures, hence APP precipitation datasets do not exist). The similarity maps of June precipitation data for the years 1992–2000 visually fall into two groups:

1) comparisons involving MM5–CRU, MM5–MW, and CRU–MW;
2) comparisons involving NCEP or ERA-40.

Overall similarity is better in group 1 than in group 2. Comparisons inside the group MM5, MW, and CRU share similar patterns with good similarities in the coastal ranges, in the central north and in the northeast of the study area, in areas of dissimilarities up to around 0.35 extending from the northern part of the Gulf of Alaska northward, and in a second region between −120° and −130° longitude.

Similarity maps resultant from pairwise comparisons of NCEP1 and any other
dataset (MM5, CRU, MW, and ERA-40) have an overall worse similarity, with only small areas of similarity better than 0.1 and maximal values above 0.35 or 0.4. Areas of best agreement tend to occur over the southern coastal ranges in southeast Alaska, and in an area north of the Gulf of Alaska, whereas areas of poor similarity are found in the northern-central inland regions of the study area and along its northern limit; in particular, problematic areas in temperature and in precipitation data/models differ. In comparisons involving ERA-40, the general geographic distribution is similar, with somewhat lower values, indicating better agreement with other datasets, overall.

5. Conclusions

This paper analyzed temporal and spatial variability in six Arctic temperature and five precipitation datasets over the Western Arctic Linkages Experiment (WALE) study region from 1992 to 2000. The analysis methods included analysis of variance (ANOVA) with post hoc tests to assess whether there are statistically significant differences among datasets, anomaly correlations to determine whether the datasets showed significant association for the seasonal cycle, and algebraic similarity mapping to determine which geographic regions contributed most to the differences detected in the temporal analysis for the month of June. The following major conclusions resulted.

1) Interannual variability in temperature datasets is greatest in winter months, with the APP dataset typically being coldest, due to cloud masking and near-surface inversions, and the MW registering the warmest temperatures.

2) Differences in temperature datasets vary seasonally, and there is substantial evidence to conclude that most datasets are not statistically equivalent in all months.

3) While there are systematic differences in the datasets, anomaly correlations indicate that each dataset closely resembles the others for temporal analyses of relative differences of temperature.

4) Similarity mapping suggests that areas of poor temperature agreement lie mainly in the coastal areas and in the western central part of the study area.

5) For precipitation, the MM5 is higher in winter months and the NCEP1 dataset is higher in summer months, due to known errors in the NCEP1 data, and the largest discrepancies for precipitation occur in July.

6) Correlation analyses indicate that all precipitation datasets are significantly correlated, but that the strength of association for precipitation is much less than for temperature.

7) Similarity mapping for precipitation indicates that the largest areas of disagreement occur in the central and eastern portions of the study region.

As a final thought, these results indicate that the choice of forcing datasets likely will have a significant effect on the output from hydrologic models. As a result, several different datasets should be used for a robust hydrologic assessment.

Acknowledgments. This work was supported by NSF Grant 0095047. The CRU and ERA-40 data were acquired courtesy of Eugenie Euskirchin at the University of Alaska.
Fairbanks, and the APP data were obtained courtesy of Jeff Key at the University of Wisconsin. We thank the comments of two reviewers who pointed out many issues in the original manuscript, leading to a much better final product.

References


