Spatiotemporal Climate Model Validation—Case Studies for MM5 over Northwestern Canada and Alaska

Ute C. Herzfeld*
Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado

Sheldon Drobot
Colorado Center for Astrodynamics Research, University of Colorado, Boulder, Colorado

Wanli Wu
National Center for Atmospheric Research, Boulder, Colorado

Charles Fowler and James Maslanik
Colorado Center for Astrodynamics Research, University of Colorado, Boulder, Colorado

Received 24 May 2006; accepted 2 March 2007

**ABSTRACT:** The Western Arctic Linkage Experiment (WALE) is aimed at understanding the role of high-latitude terrestrial ecosystems in the response of the Arctic system to global change through collection and comparison of climate datasets and model results. In this paper, a spatiotemporal approach is...

Objectives include a quantitative assessment of similarity between datasets and climate model fields, and identification of geographic areas and seasons that are problematic in modeling, with potential causes that may aid in model improvement. These are achieved by application of algebraic similarity mapping, a simple yet effective method for synoptic analysis of many (here, 45) different spatial datasets, maps, and models. Results indicate a dependence of model–data similarity on seasonality, on climate variable, and on geographic location. In summary, 1) similarity of data and models is better for temperature than for precipitation; and 2) modeling of summer precipitation fields, and to a lesser extent, temperature fields, appears more problematic than that of winter fields. The geographic distribution of areas with best and worst agreement shifts throughout the year, with generally better agreement between maps and models in the northeastern and northern inland areas than in topographically complex and near-coastal areas. The study contributes to an understanding of the geographic complexity of the Arctic system and modeling its diverse climate.

**KEYWORDS:** Climate modeling; Spatial data analysis; Algebraic map comparison

### 1. Introduction and objectives

The goal of the Western Arctic Linkage Experiment (WALE) is to investigate the role of high-latitude terrestrial ecosystems in the response of the Arctic system to global change. To further this goal, climate datasets and climate model results are compiled, collected, and compared for the WALE study region, which includes land areas in Alaska and northwestern Canada at 55°–70°N, 165°–110°W approximately [see McGuire et al. 2006, manuscript submitted to *Earth Interactions*; see also WALE project Web sites at http://wale.unh.edu (University of New Hampshire) and http://picea.sel.uaf.edu/projects/wale.html (University of Alaska Fairbanks)]. Through WALE, a suite of experiments was established. Climate modeling, data assimilation, and model validation play an important part in understanding the complexity and variability of the systems. This paper constitutes a contribution to WALE through several experiments, or case studies, which analyze WALE data and model results in the form of maps and grid models and hence facilitate an understanding of the geographic complexity of the Arctic system and may assist in modeling its diverse climate.

Central to the WALE project is the analysis and validation of the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5), which is a physical climate model developed with contributions from the climate research community (see www.mmm.ucar.edu). In the case studies presented in this paper, we focus on MM5 temperature and precipitation grids for the model years 1992–2000 (Wu et al. 2007) and compare them to several compiled or reanalyzed datasets that are frequently used in the climate research community.
These include temperature and precipitation datasets from 1) the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40; see Gibson et al. 1997); 2) University of Delaware climate datasets [UDEL (MW); see Willmott and Matsuura 1995]; 3) National Centers for Environmental Prediction (NCEP)–NCAR reanalyses (NCEP1 or NCEP or NCAR; see Kalnay et al. 1996). In addition, 4) temperature approximations derived from Advanced Very High Resolution Radiometer (AVHRR) Polar Pathfinder data (PATH or PCT or APP; see Fowler et al. 2002) are utilized (for all datasets, see http://wale.unh.edu/data.shtml).

For a previous comparison of temperature datasets from NCEP–NCAR, the 15-yr ECMWF Re-Analysis (ERA-15), and the Climatic Research Unit of the University of East Anglia (Jones et al. 2001), the reader is referred to the Arctic Climate Impact Assessment (ACIA 2005), for investigations of precipitation data from NCEP–NCAR and ERA reanalyses to Serreze and Hurst (Serreze and Hurst 2000).

The results of temporal analysis of these datasets (Drobot et al. 2006), which utilizes values averaged over the entire spatial domain and station values, lead to the question of geographical distribution of agreement and disagreement among datasets and between data and models. The geographic region comprising the WALE project domain encompasses a large range of topographic relief and considerable variability in local climatic conditions from the wet coastal ranges to the interior ranges, to regions in the shadow of westerly systems, and to a range of latitudes. Given this diversity within the study region, it is important to examine how models and data compare spatially, rather than through comparison of averaged fields alone. A visual comparison between results of model climate fields produced by MM5 and a couple of datasets is approached by Wu et al. (Wu et al. 2007). For an increasing number of maps, datasets, and models, as have been compiled under the WALE project, a quantitative and objective approach is needed.

The characteristic approach of this paper lies in a spatial analysis of MM5 model results and WALE datasets, and hence the analyses in Drobot et al. (Drobot et al. 2006) and Wu et al. (Wu et al. 2007) are taken a step further into a multidimensional spatiotemporal domain. To achieve this, we apply algebraic similarity mapping, a method for quantitative comparison of any number of input datasets, maps, or models, which was first developed for resource exploration (Herzfeld and Merriam 1990) and is adapted here for climate data analysis and the WALE experiment. Similarity mapping, or algebraic map comparison, utilizes a multidimensional algebraic algorithm to compare any number of input datasets, given as grids of a study region, and results in similarity maps. These maps show areas where the input datasets or model fields agree versus areas of mismatches. The similarity maps can be readily interpreted to answer the central questions of the paper. The similarity mapping method is also applied in Drobot et al. (Drobot et al. 2006) for a study of the most problematic month identified in the temporal data analysis (June). In this paper, we add another dimension, that of seasonality, by completing all case studies for summer (July) and winter (January). In our work as well as in the Drobot et al. (Drobot et al. 2006) and Wu et al. (Wu et al. 2007), the area of study is bordered by 55°–65°N, 160°–110°W to match the area of simulations with MM5 (Wu et al. 2007).

In summary, central questions in the spatiotemporal MM5 validation case studies in this paper are the following.
• How well and where do datasets and model fields match?
• How do differences between data and models vary as a function of geographical location?
• How large are seasonal effects, and how are they distributed regionally?
• What are the potential causes and remedies of discrepancies between data and models?

Hence the objectives of our study are 1) quantitative assessment of similarity between datasets and climate model fields, for temperature and precipitation, in a spatial domain; 2) identification of geographic areas that are problematic in modeling; 3) investigation of seasonal differences in model–data agreement and discrepancies; and 4) an attempt at attribution of discrepancies to aid in future modeling and data analysis.

2. Method: Algebraic similarity mapping

2.1. Overview

Algebraic similarity mapping facilitates spatial comparison of any number of geophysical, meteorological, geographic, or other spatial datasets and is the main method applied here to analyze temperature and precipitation datasets and models for the WALE project. Algebraic similarity mapping utilizes the algebraic map comparison method first derived in Herzfeld and Merriam (Herzfeld and Merriam 1990) as an aid in petroleum exploration based on several geophysical and geologic maps. Applications in exploration geology, basin analysis, oceanography, and marine geology are reported in Merriam et al. (Merriam et al. 1993), Hamann and Herzfeld (Hamann and Herzfeld 1991), and Herzfeld (Herzfeld 1992). Other methods of thematic map comparison are summarized in Merriam and Jewett (Merriam and Jewett 1988); however, the result of comparing two maps is usually a number, except for the case of a correlation map derived from two series of maps (Brower and Merriam 1990). An application of algebraic map comparison in physical modeling is described in Herzfeld (Herzfeld 1992).

In a nutshell, the map comparison algorithm utilizes a pointwise operator [the map comparison (MAPCOMP) operator], which calculates a combination of distances among standardized values in pairwise comparisons of any number of input maps. In simplified terms, one may envision that the MAPCOMP operator moves across the study area and derives a similarity value at each (grid) location from a stack of maps. An input map can be a gridded dataset for the study area, or the result of a model, for instance, MM 5-simulated 1992 temperatures at the 2-m level, or ERA-40 reanalysis precipitation data for an average of January 1992–2000 values.

More precisely, similarity mapping is based on an algebraic approach that proceeds by 1) standardizing input values in each map or spatial mode; 2) forming a functional of pairwise differences of standardized values; and 3) applying a seminorm to the functional in step 2, which results in a similarity value \( F(x) \) in each point \( x \) in a map area \( M \). The result is a spatial grid model of similarity values, which may be contoured, displayed as a three-dimensional model, or utilized as an input layer in a geographical information system (GIS) or in a physical model. The
next sections detail the mathematical principles as well as options for handling missing data values and integrating boundaries of geographic areas.

Standardization is necessary, wherever data from different sources or variables of different units are to be analyzed synoptically, as is the case in the WALE modeling and analysis project (http://wale.unh.edu/data.shtml). In our application, we use linear transformation of the range of data into the interval [0,1], and all calculations are performed inside a landmask outlining the study area. Similarity values close to 0 indicate good similarity, while higher values indicate poor correspondence among the input maps or models in a given location. The largest similarity value is 1. As a rule of thumb, similarity values of 0–0.2 indicate good similarity in comparisons with nine or more maps.

Advantages of the method are: 1) The spatial relationship of each value is preserved throughout the analysis, an essential point in all geographic analysis. 2) Any number of input maps can be compared simultaneously. 3) Several standardization methods are available to integrate different variables. 4) The results are spatial datasets and hence may be presented as similarity maps or comparison maps of the study area. 5) Missing data situations can be handled by a number of options, and landmasks can be used. 6) Application of the method is straightforward and does not require specific expertise or system components, as is the case for a geographical information system.

2.2. Standardization methods

A typical situation when studying a complex Earth system is that data of several different variables need to be investigated simultaneously. The problem of pre-analysis standardization arises as soon as more than one variable is involved in an analysis, or even when data on the same variable are reported in different units but have to be considered on a common scale. Because careless choices of the standardization method may distort the results and lead to uncertainties in the analysis, it is worth providing at least a couple of options for standardization.

Standardization is often understood as synonymous with Gaussian or so-called z-score standardization (transformation of observed data to a mean of zero and a standard deviation of one). The parameters mean and standard deviation, however, only make sense if the data show a Gaussian ("normal") distribution, an assumption that is often not met in geologic or environmental studies. We use the term "standardization" in the original meaning of a transformation that renders several different variables comparable. The map-comparison software offers several options for standardization. In this paper, the proportion-of-range standardization is applied, which is defined as follows.

Definition. Proportion-of-range standardization. If $y_{\min_j}$ and $y_{\max_j}$ are the minimal and maximal values observed on the variable $y_j$ (for $j = 1, \ldots, n$, where $n$ is the number of variables, e.g., temperature, precipitation, and humidity), and $y_{ij}$ are observations of $y_j$ (for $i = 1, \ldots, r_j$, where $r_j$ is the number of observations on $y_j$), then the proportion-of-range standardization $z_p$ is defined by

$$z_p(y_{ij}) = \frac{y_{ij} - y_{\min_j}}{y_{\max_j} - y_{\min_j}}.$$

(1)
We may also write \( y = z_p(y) \) for a standardized value. If one variable generally increases where most others decrease, then inverse proportion-of-range standardization \( z_{p-} \) is used:

\[
z_{p-}(y_{ij}) = \frac{y_{ij} - y_j}{y_{ij} - y_{\min}}.
\] (2)

Note that \( z_{p-}(y_{ij}) = 1 - z_p(y_{ij}) \). Both proportion-of-range standardizations map the observed data range into the interval \([0,1]\). The proportion-of-range transformation considers all data values equally important and is associated with the uniform distribution. The actual distribution of a variable depends on the selection of the study area and is usually unknown or very complex for spatial environmental variables or climate variables in topographically complex areas. In such cases, assumption of a uniform rather than a Gaussian distribution commonly leads to more robust results in the analysis (see Hamann and Herzfeld 1991). In consequence we utilize the proportion-of-range transformation in the application of similarity mapping to WALE data.

### 2.3. The MAPCOMP Algorithm

The algebraic map comparison algorithm is based on the calculation of weighted averages between standardized values of any pair of input maps, for each point of the map area (in practice, for each node of a common underlying grid). Let \( M \) denote a map area; \( n \) the number of input maps; and \( M_1, \ldots, M_n \) the input maps with \( m_{kt}(x) \) the standardized value of map \( M_k \) at location \( x \). For each grid point \( x \in M \) a difference matrix \( D(x) \in \mathbb{R}^n \) is formed with elements \( d_{st}(x) = m_{st}(x) - m_{kt}(x) \); note that different types of standardization may be used in each map, \( M_s, M_t, s, t = 1, \ldots, n \). As \( x \) varies over the study area, \( D(x) \) becomes a matrix functional. Now we define a norm in \( \mathbb{R}^n \) as

\[
F(x) = \frac{1}{k} \sum_{s < t} |d_{st}(x)|,
\] (3)

where \( k = n(n-1)/2 \) is the number of different pairwise comparisons among the \( n \) maps. Here, \( F(x) \) represents the average difference between maps at location \( x \), and hence \( F(x) \) is low if the maps are similar and high if they are different.

Use of Equation (3) assumes that all input maps are of equal importance. In a practical situation, however, one can imagine that one map may be considered more important, or that a model is to be compared with several different input maps. It may also be the case that a change in one variable has more drastic consequences for the environment than a change in another variable; for instance, imagine a 50% increase in wind speed versus a 50% increase in precipitation.

To meet similar requirements, we assign a weight \( w_i \) to the input map \( M_i \) for each \( i = 1, \ldots, n \) that captures the importance of map \( M_i \). (Note that all weights must be nonnegative and at least one positive.) The MAPCOMP equation for \( F \) including weighted input maps is given by
If a zero weight is used, the right-hand side of Equation (4) is actually only a seminorm.

Calculating $F(x)$ over the whole study area, that is, for every location $x \in M$, we build a grid model for the similarity map $F$. This grid model may be visualized as a map that shows relative values, giving the quality of agreement of the input maps. If a proportion-of-range standardization is used, then the values in the similarity map will also lie between 0 and 1, with low values indicating good correspondence among the input maps in a given location, and higher values indicating poor similarity. As a rule of thumb, a value below 0.2 indicates good similarity in comparisons with nine maps or more, and a value above 0.5 indicates very poor similarity.

In Equations (3) and (4), the 1-norm (absolute value) is employed to form $d_s(x)$, that is, the absolute value of pairwise differences of map values. One may also base the MAPCOMP equation on other norms, such as the Euclidean norm (2-norm) or the Mahalanobis norm. We have chosen the 1-norm for reasons of robustness in the ensuing data analysis.

A resultant value on the comparison map $F$ can stem from the fact that all maps are slightly different, or the fact that all but one are similar and one map differs from all the others. In that situation, a visual test may help to find the outlier. Using the MAPCOMP program, it is easy to compare only some of the $n$ maps by setting some weights to zero, and thus determine significant connections (Herzfeld and Sondergard 1988).

2.4. Missing-data handling and working with landmasks (e.g., study area for WALE)

Data may be missing at some of the grid nodes in the map area $M$, and this could happen at different locations of the individual maps. The program MAPCOMP offers two possibilities to handle such situations:

(i) $F$-algorithm: At location $x \in M$, only the maps that have data at $x$ are compared.

(ii) $G$-algorithm: At location $x \in M$, the comparison value is only calculated if all $n$ maps have values at location $x$.

In subareas where all maps have data, $F$ and $G$ coincide. If only one map has a value at a point $x$, no comparison is done. In places where 2 to $(n-1)$ maps have data, an $F$-map may be obtained, but not a $G$-map. The advantage of the $F$-map algorithm is that comparison is carried through in a larger part of the area. On the other hand, $G$-maps are easier to read, since all map areas are supported by the same number of input data fields.

In our analysis in the WALE project, the region is not a rectangle, and in addition, the data were required to be entered in a number of different reference
systems. MAPCOMPMASK, a new version of the program MAPCOMP, takes care of those requirements.

A landmask file can be created that is a matrix of a rectangular envelope of the study area in the projection that is to be used in the similarity mapping. Indicator values are used to identify grid nodes inside and outside the study area; notably the area need not be simply connected but may have any shape. In combination with the $F$- and $G$-algorithm options for missing data values, many different situations of data coverage can now be incorporated in a synoptic study.

In the ensuing analysis, we have taken the following approach. The area of $55^\circ$–$65^\circ$N, $160^\circ$–$110^\circ$W is determined by the simulations carried out with MM5 for the WALE project (Wu et al. 2007). Some WALE datasets have gaps in coverage within this area. In the following case studies, an $F$-comparison map has been calculated in case all datasets have data in the entire study area; otherwise, a different landmask is used, and calculations have been performed where all datasets contain data. In light of applications of the comparison maps, this means that 1) data for the entire area have been used wherever possible to provide a broader base for future studies; and 2) all comparison maps contain only values in locations where all input datasets have values, and therefore all maps are easy to read.

3. Geographic region

The area of the WALE project encompasses terrestrial regions in Alaska and northwestern Canada, approximately between $55^\circ$N and $70^\circ$N, $165^\circ$W and $110^\circ$W and rectangular in Equal Area Scalable Earth (EASE) grid (Armstong and Brodzik 1995; see section 4.2.). The study area is defined as the terrestrial regions between $55^\circ$–$65^\circ$N and $160^\circ$–$110^\circ$W (see Figure 1), and is the same as the area used by Drobot et al. (Drobot et al. 2006) and Wu et al. (Wu et al. 2007). The $65^\circ$N latitude circle runs just north of Fairbanks, Alaska; $160^\circ$W traverses the Alaska Peninsula between Sand Point and Chiguik; and $55^\circ$N passes between Ketchikan (N), Dawson Creek (N), Prince Rupert (S), and Prince George (S), containing most of southeastern Alaska in the WALE area. Parts of the Yukon and Northwest Territories of British Columbia and Alberta constitute the eastern half of the study area.

The study area includes parts of the lower Yukon and Kuskokwim drainages in the west, the Kilkuck Mountains (1000–1500 m), the Kaiyuh Mountains, the Nowitna River, the Kantishna, Tanana, and Delta Rivers and subsidiaries, the Alaska Range (Denali: 20 320 ft, approximately 6300 m), the Merrit Mountains (about 2000 m) between Fairbanks and the Yukon Territory, the Wrangell–St. Elias Mountains (reaching to 5920 m in Mt. Logan), the Chugach Mountains (up to 8504 ft in Hanagita Peak), the Coast Ranges (Mt. Waddington: 4016 m) along the Gulf of Alaska, the Mackenzie River in Canada, the Great Slave Lake in the Yukon Territory, and the Northern Rocky Mountains of Canada. The area is also known by its major highways, which are the Alaska Highway, the Stewart–Cassiar Highway, the Liard Highway, and the Klondike Highway (Whitehorse to Dawson City). In some datasets, coverage was lacking in some coastal areas and in a large area in southeast Alaska and in Canada east of the immediate coastal area and around the region of the Stewart–Cassiar Highway (see Figures 1a–d and compare Figure 2a with 2c). This area is topographically, ecologically, and climatically complex.
4. Datasets, model fields, and data processing

4.1. Data and models

All datasets compiled and/or contributed by researchers in the WALE project may be accessed on the University of New Hampshire Web site (http://www.unh.edu/data). The analyses in this paper are based on the following datasets and model fields.

1) MM5: The model MM5 was run for 10 model years, using NCEP-NCAR reanalysis data at the boundary of the WALE study area (Wu et al. 2007).

Figure 1. Study area: Geography and effects of coordinate transformation. (a) The study area, 55°–65°N, 160°–110°W, in northwestern Canada and Alaska with respect to EASE grid. (b) Topography of the WALE area (55°–70°N and 165°–110°W), approximated in a DEM. (c) A WALE data analysis map of the study area plotted with respect to EASE grid, for the example of a similarity map of nine datasets (1992–2000) of January ERA-40 reanalysis temperatures. (d) A WALE data analysis map of the study area plotted with respect to the geographic coordinate system, for same example as in (c).
Resultant model grids contain monthly averages of temperature and precipitation values for 1992–2000; these provide 12 (months) × 9 (years) × 2 (variables) input maps for similarity mapping and may also be considered as monthly or annual time series. The model fields may be treated in the same way as datasets.

2) **ERA-40**: The 40-yr European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA-40) reanalysis data are the results of data assimilations, following principles described in Gibson et al. (Gibson et al. 1997) for the 15-yr ECMWF Re-Analysis (ERA-15) reanalysis data. Several aspects of the ERA-40 project are covered in Simmons and Gibson (Simmons and Gibson 2000) and Simmons (Simmons 2001).

3) **UDEL (MW)**: UDEL climate data (see http://climate.geog.udel.edu/climate and http://www.cdc.noaa.gov/cdc/data.Udel-airT-precip.html) are derived from field observations, following interpolation methods described in Willmott and Robeson (Willmott and Robeson 1995) and Willmott et al. (Willmott et al. 1996). In this work, we utilize UDEL climate datasets in the form updated in June 2006 and referred to as UDEL-MW datasets in the WALE dataset compilation (wale.unh.edu/data.shtml); the original UDEL (MW) dataset is described in Willmott and Matsuura (Willmott and Matsuura 1995).

4) **NCEP, NCAR, NCEP1**: NCEP-NCAR reanalysis data are produced via data assimilation processes. The data used here are also referred to as NCEP1 and are an updated version of the dataset described in Kalnay et al. (Kalnay et al. 1996).

5) **PATH, PCT, APP**: Passive microwave data from the AVHRR Polar Pathfinder Mission of the Television Infrared Observation Satellite (TIROS) satellite are utilized as proxies for temperature data. Temperature data are calculated from satellite observations of pseudo-clear-sky temperatures following methods described by Fowler et al. (Fowler et al. 2002) and are stored with reference to EASE grid. Polar Pathfinder (PATH, APP) data are utilized here in the form identified as “PCT data,” which are derived using the latest revision of the data processing algorithms at the time of our study (May 2006).

All datasets are used in the form of monthly averages for 1992–2000 to match the MM5 model fields. The variables’ temperature and precipitation have been selected, as they are essential in any climate study, and were included in most WALE datasets at the time of this study.

### 4.2. Data processing

Data are accessed via the WALE data Web site (http://wale.unh.edu/data.shtml), transformed into EASE Grid (Armstrong and Brodzik 1995; see below for explanation) to facilitate comparison with datasets maintained at the University of Colorado’s Colorado Center for Astrodynamics Research (CCAR), analyzed with the similarity mapping software MAPCOMPMASK and MAPCOMPEASE5, and then transformed back into the geographic coordinate system using IDL. The effect of this coordinate transformation is illustrated in Figures 1c and 1d.

For properties and processing routines associated with compilation of the
WALE datasets, the reader is referred to the WALE data Web site, the WALE project Web sites [http://wale.unh.edu (University of New Hampshire) and http://picea.sel.uaf.edu/projects/wale.html (University of Alaska Fairbanks)], and other papers in this special theme. The datasets vary with respect to spatial resolution and the original interpolation method employed in their construction. For the comparative purpose here, all datasets were transformed into EASE Grid using bilinear interpolation or a similar interpolation method.

EASE Grid is a rectangular coordinate system based on the stereographic projection (see Snyder 1987) and intended for polar studies (see http://www.nsidc.org/..EASE grid by Ken Knowles and Mari-Jo Brodzik). The study area corresponds to EASE Grid rows 41-151/columns 41-151, a 111 x 111 submatrix in the upper-left corner of an EASE-Grid matrix, and is shown in Figure 1c.

MAPCOMPMASK takes care of the fact that the WALE region is not rectangular, and that different datasets may have different coverage within the study area. MAPCOMPEASE5 performs similar tasks, but specifically for EASE-Grid coordinate system. (For MAPCOMP, see http://www.iamg.org; for MAPCOMPMASK and MAPCOMPEASE5, write to the first author.)

5. Results

5.1. Interannual variability of temperature and precipitation fields

In the main result sections (5.2. for temperature and 5.3. for precipitation), the performance of MM5 as a predictor of climate fields for the study area is investigated, in order to answer the questions posed in the introduction. This will be achieved by calculating comparison maps between datasets and models. Prior to studying interdependencies of several variables over time, we investigate the interannual variability of each observational and model variable individually.

5.1.1. ERA-40

The similarity maps in Figure 2 show how much temperature and precipitation vary over the course of nine years (1992–2000), and in which places this interannual variability is highest or lowest. The temperature data (denoted by T2) are based on measurements at the 2-m level above the ground on a weather station tower.

Both the January (Figure 2a) and the July (Figure 2b) maps show very high similarity among the ERA-40 temperature datasets, as indicated by $F$-values (similarity values, or comparison values) below 0.15 almost everywhere in the study area. This means that interannual variability within the nine study years was low for January and July. The spatial distribution of lowest and highest values is almost reversed in January and July. For January, the highest values (0.1 < $F$ < 0.15) occur in an area over the coastal and adjoining inland mountains, with the exception of near-coastal areas, while the lowest values ($F$ < 0.05) are reached in the northeast inland part of the WALE area. For July, highest values ($F$ from 0.1 to just over 0.15) are reached in the northeast corner of the study area, in the northwest corner, and along the inland part of the southern area boundary, while the central areas have the least interannual variability and lowest $F$-values.

Figures 2c and 2d, in comparison with Figures 2a and 2b, show that January precipitation, according to ERA-40 reanalysis data, shows little variability in most
of the study area ($F < 0.5$), with the exception of coastal areas ($0.15 < F < 0.2$). Over the entire region, the similarity values for temperature compare to those for January precipitation. In contrast, summer (July) precipitation, according to ERA-40 data, exhibits a high interannual variability. Highest similarity values reach $0.3 - 0.35$, and values of $0.15 - 0.3$ dominate the map.

5.1.2. MM5

We created similarity maps for the two variables temperature at 2 m above the ground and precipitation for winter (January) and summer (July) (see Figure 3). Each similarity map is calculated from nine input maps, which are MM5 model output grids of monthly mean temperature and precipitation, for the years 1992–2000. The MM5-internal interannual variability of temperature is always low, as demonstrated by low $F$-values in both the similarity maps for January (Figure 3a) and July (Figure 3b). All $F$-values for January are below 0.2, with the largest part of the WALE study area showing values between 0.1 and 0.15 and maximum
values reached in the northwestern part of the study area. For July, internal variability is even lower: all $F$-values are below 0.15, and 0.1 is only exceeded in small parts of the study area.

In contrast, the similarity maps that indicate the internal interannual variability of MM5 precipitation predictions differ between summer and winter and have a lower spatial continuity than the maps of internal variability of temperature fields (Figures 3c and 3d). The areas of highest similarity values on the July map for precipitation variability (Figure 3d) approximately match those for lowest temperature variability; maximal $F$-values on the July precipitation variability map are between 0.25 and 0.3. Such high values occur only in coastal British Columbia for the January precipitation variability map (Figure 3c), and on the coast of the Gulf of Alaska opposite Kodiak Island. Throughout the interior of Alaska and Canada as covered by the WALE study area, $F$-values on the January precipitation variability map are below 0.1, which indicates a high similarity of the precipitation patterns in January of the nine test years, 1992–2000. The fact that temperature has a larger range of spatial continuity than precipitation is a property of the climate
variables temperature and precipitation, but we shall see in section 5.3. that MM5 does not adequately model the natural variability that occurs in summer, as captured in the interannual variability of temperature and precipitation fields derived from observations.

### 5.2. Spatiotemporal model validation for MM5 temperature fields, 1992–2000

In this section we address the objectives of 1) assessing spatial similarity between MM5 and four temperature climate fields, 2) identifying geographic areas that are problematic in modeling, 3) investigating seasonal differences by comparison of winter (January) and summer (July) maps, and 4) attributing any discrepancies to possible causes with the intention to suggest model improvements. The information to answer these questions is derived from comparison maps (similarity maps) calculated from 1992–2000 monthly temperature maps.

MM5 output grids of monthly mean temperature for the analysis region are compared to several observational and reanalysis datasets, one at a time:

1) ERA-40 versus MM5 (Figures 4a,b),  
2) UDEL(MW) versus MM5 (Figures 4c,d),  
3) PATH (PCT) versus MM5 (Figures 4e,f),  
4) NCEP1 (=NCAR) versus MM5 (Figures 4g,h), and finally  
5) for all sets synoptically (Figures 5a,b).

Each similarity map is based on 18 input maps (in 1–4) and on all 45 maps (in 5).

#### 5.2.1. ERA-40 versus MM5

Similarity maps of temperature for 1992–2000, ERA-40 versus MM5, are given in Figures 4a and 4b. The agreement of ERA-40 and MM5 is generally very good, as similarity values are 0.1–0.15 in most of the study area. Highest values are 0.15 in small areas, notably in an area in the east, which also stands out on the July map. Lowest values occur along the coast (0.1 and below) and in the northeast of the study area (0.0 < F < 0.5). A similarly distributed low along the coast has been derived for the interannual variability of MM5 (section 5.1.).

For July, the overall agreement between ERA-40 and MM5 is also good, but the spatial distribution of the similarity values is much more varied than for January, and also more varied than for just MM5, which points at regional differences between ERA-40 and MM5, considering the high spatial continuity in the map of interannual variability. Since in more than half of the study area similarity values are 0.1 or better (= lower), and in most of the study area 0.15 or better, MM5 and ERA-40 are found to match. The similarity mapping identifies a few localized problematic areas: in the eastern part of the WALE study area similarity values of $F > 0.5$ point at a discrepancy between MM5 and ERA-40. In addition, several small areas close to the coast and in the Coast Ranges have $F$-values of 0.2–0.25; here prediction is likely difficult due to topographic relief and resulting local variability of weather and climate variables.

#### 5.2.2. UDEL(MW) versus MM5

The geographic distribution of UDEL(MW) versus MM5 January temperature similarity values (Figure 4c) largely matches that of ERA-40 versus MM5, for the
Figure 4. Model validation for MM5 temperature prediction using algebraic similarity mapping with 18 input maps. Temperature at 2 m above ground. Note that UDEL(MW) data are not available for regions in the southwest of the study area (see www.unh.edu/data). (a) Similarity map ERA-40 vs MM5, January 1992–2000; (b) similarity map ERA-40 vs MM5, July 1992–2000; (c) similarity map UDEL(MW) vs MM5, January 1992–2000; (d) similarity map UDEL(MW) vs MM5, July 1992–2000; (e) similarity map NCEP1 (=NCAR) vs MM5, January 1992–2000; (f) similarity map NCEP1 (=NCAR) vs MM5, July 1992–2000; (g) similarity map APP (=PCT=PATH) vs MM5, January 1992–2000; and (h) similarity map APP (=PCT=PATH) vs MM5, July 1992–2000.
eastern half of the study area, but in the west and northwest higher values of 0.15–0.2 are found, indicating a lesser agreement between UDEL(MW) data and MM5. For July temperature fields (Figure 4d), the similarity map is quite different than for January—values are higher, the geographic areas of highest disagreement are almost complementary to those on the January map, and the spatial continuity has a shorter range. Values of 0.3–0.35 are reached at 60°–65°N, 130°W and values above 0.25 are reached in several Alaska locations. A high point with $F > 0.5$ exists at about 62°N, 115°W on every map, which suggests an exceptionally large error related to reporting or ingestion of station data at that location. Note that UDEL(MW) data are not available for a region in the southwest part of the study area (see Figures 4c and 4d).

5.2.3. NCEP1 (=NCAR) versus MM5

Similarity maps of temperature at the 2-m level, comparing NCEP1 (=NCAR) data fields with MM5 model outputs for January (Figure 4e), match the similarity maps ERA-40 versus MM5 (Figure 4a) and PATH (PCT) versus MM5 (Figure 4g), and show the large-range spatial continuity known from ERA-40 versus MM5—however, this observation does not hold for similarity maps for the month of July.

For July the similarity map NCEP1 (=NCAR) versus MM5 (Figure 4f) has high values of $F > 0.3$ over large regions, which reach $F > 0.4$ locally. Such high values also occur in the T2 comparison of UDEL(MW) versus MM5, but in different regions. The region with the best agreement is in the eastern part of the study area ($F < 0.1$), which coincides with the area of best agreement in the UDEL(MW) versus MM5 comparison. Similarity in the westernmost area is also good ($F < 0.15$). The highest values occur in the noted “bad spot” near the eastern

Figure 5. Comparison of MM5 with all four datasets (ERA-40, NCEP1 (=NCAR), APP (=PATH=PCT), and UDEL(MW)) simultaneously for temperature fields in years 1992–2000. Note that UDEL (MW) data are not available for regions in the southwest of the study area (see http://www.unh.edu/data). (a) January 1992–2000 (45 input maps) and (b) July 1992–2000 (45 input maps).
margin of the study area, but also near the southern margin of the study area. Mountainous areas about 100–300 km inland from the coast appear to be most problematic, a phenomenon that has also been observed in other comparisons and is attributed to the small-cell variability of natural processes in this area of extremely high relief.

5.2.4. PATH (=PCT=APP) versus MM5

At first scrutiny, the similarity maps of PCT (=PATH=APP) temperature versus MM5 temperature fields (Figures 4g and 4h) strike the observer as being a lot noisier than any other similarity maps in this section. This noisiness may be attributed to differences in spatial resolution or spatial continuity between the PCT data and the MM5 grids. Despite the noisiness, the large-scale spatial distribution of areas of similarity and dissimilarity matches those of similarity maps ERA-40 versus MM5 (Figures 4a and 4b), with slightly (up to 0.05) higher values in maps of PCT versus MM5. In January, the lowest values are reached in northern inland areas, and highest values with \( F \) between 0.15 and 0.2 in the Coast Ranges and in the western part of the study area.

In July, the lowest values with \( F \) of 0–0.05 occur in small areas, while coastal areas show higher values of \( F > 0.1 \). The overall similarity, that is, the agreement between model MM5 and map PATH (=PCT) data, is good, except for a known spot in the east-northeast of the study area, which is the notable bad spot location seen on almost every comparison map, and a small area on the southern coast where similarity is also larger than 0.4.

In summary of the pairwise comparisons of MM5 versus the four different datasets [ERA-40, UDEL(MW), NCEP1 (=NCAR), and PATH (PCT)] for the variable temperature at the 2-m level over the time range 1992–2000, we state the following:

Temperature at the 2-m level can be modeled satisfactorily with MM5 for January to match all four datasets. Modeling July temperature is far more problematic, which may be attributed to the fact that weather in the study area in July is characterized by high variability in both space and time, a fact that is reflected in the differences in the datasets for the month of July (see section 5.1.). This likely reflects localized convective processes and/or topographic effects. General agreement of MM5 with ERA-40 and PCT (=PATH) is still fairly good over large parts of the WALE area. For UDEL(MW)-versus-MM5 and NCEP1-versus-MM5 comparison maps, mountainous areas including the Coast Ranges and the interior ranges show dissimilarities, where \( F \)-values exceed 0.25 for large regions. This is unexpected as NCEP1 data are used to “drive” MM5. A bad spot is detected on all comparison maps in the same location, which indicates a mistake in MM5 in that location. (The location is obvious in all maps; see Figure 4.)

5.2.5. Comparison of MM5 with all four data types

Finally, we compare MM5 with all four datasets simultaneously, also for the nine model years of 1992–2000 (Figure 5). The purpose of this case study is to provide a synoptic representation of areas that are matched better in the modeling, compared to mapping from observations in the four datasets, and to detect areas
where the match is unsatisfactory, in light of potential improvement in modeling (see also section 5.3.). Based on Figure 5, the differences in modeling skill for January and July temperatures are now obvious—similarity is almost everywhere better than 0.15 for January, with slightly higher values in coastal areas (to 0.25), and the northwestern area is best ($F < 0.1$). In contrast, July temperatures are not predicted as consistently over the study area, indicated by a shorter range of continuity over the study area, and as accurately (values to 0.25 in several areas, and 0.5 in the bad spot).

5.3. Spatiotemporal model validation for MM5 precipitation fields, 1992–2000

A study similar to that for temperature at the 2-m level is carried out for precipitation, comparing data from ERA-40, UDEL(MW), and NCEP1 (NCAR) pairwise with MM5 for the 1992–2000 time series, and studying seasonal dependency.

5.3.1. ERA-40 versus MM5

ERA-40 January precipitation fields (see Figure 6a) are matched very well by MM5, with $F < 0.1$ for most of the study area. Higher values occur only at the northern coast of the Gulf of Alaska and on Seward Peninsula. In contrast, the similarity map for July precipitation (Figure 6b) is dominated by high values, indicating poor similarity, up to $0.35 < F < 0.45$, and values above 0.2 in more than half the map area, notably in central northern and eastern regions of the study area. A look at the maps of interannual variability of just ERA-40 (Figure 2d) and just MM5 (Figure 3d), which represent from the viewpoint of this study the internal variability of each set, indicates two notes: 1) the internal variability of each set may contribute to the overall higher values, but 2) a mismatch in the geographic distribution of precipitation fields may be the reason for the high dissimilarity.

5.3.2. UDEL(MW) versus MM5

Again, the January map (Figure 6c) of UDEL(MW)-versus-MM5-precipitation similarity exhibits a very good agreement, with $F < 0.1$ for most areas and high values near the Gulf of Alaska and in the eastern coastal location. The July map (Figure 6d) indicates that MM5 comes closest to UDEL(MW) of the three sets, ERA-40, UDEL(MW), and NCEP1, with good agreement ($F < 0.2$) in most areas and higher values south of 60°N.

5.3.3. NCEP1 (NCAR) versus MM5

NCEP1-versus-MM5 similarity maps (Figures 6e,f) are overall very similar to ERA-40-versus-MM5 comparison maps, with slightly higher $F$-values in most locations. The fact that the poorest agreement for July precipitation is seen between NCEP1 and MM5 is somewhat surprising given that the MM5 runs were forced with NCEP1. This may be indicative of a modeling problem.

The lowest values with $F < 0.05$ dominate over most inland areas, with the best agreement in the northeastern part of the study area, whereas the highest $F$-values occur at and near the southern coastal area ($F > 0.5$) with a steadily trending, but
not monotonous increase from inland to coastal areas. Overall similarity is good, as areas of good agreement are larger than areas of poor agreement.

The picture is different for July (Figure 6f), as overall agreement between NCEP1 and MM5 is poor, since high similarity values of $F > 0.3$ are derived for large parts of the study area. Most surprisingly, areas with the highest values are different than on any other similarity map, as they are located in interior regions but do not follow the main mountain ranges.
6. Summary and conclusions

Results of the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) runs for temperature and precipitation in northwestern Canada and Alaska at 55°N–65°N, −160°E–110°E, derived using National Centers for Environmental Prediction (NCEP)–NCAR reanalysis (NCAR, NCEP, NCEP1) reanalysis data at the boundary of the study area (Wu et al. see WALE Special Theme), have been evaluated through comparison with several climate datasets. These include temperature and precipitation datasets from 1) 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40; see Gibson et al. 1997), 2) University of Delaware climate datasets [UDEL(MW); see Willmott and Matsuura 1995]; 3) NCEP–NCAR reanalyses (NCEP1 or NCEP or NCAR; see Kalnay et al. 1996), and, for temperature only, 4) temperature approximations derived from Advanced Polar Pathfinder data (APP or PATH or PCT; see Fowler et al. 2002; for all datasets, see wale.unh.edu/data.shtml). The study is based on monthly datasets and model fields of the years 1992–2000.

The objectives are as follows:

1) quantitative assessment of similarity between datasets and climate model fields, for temperature and precipitation, in a spatial domain;
2) identification of geographic areas and seasons that are problematic in modeling;
3) investigation of seasonal differences in model–data agreement and discrepancies; and
4) an attempt at attribution of discrepancies to aid in future modeling and data analysis.

The analytical approach utilizes algebraic similarity mapping (Herzfeld and Merriam 1990) to compare various combinations of input datasets and model fields simultaneously and with reference to geographic location. As demonstrated in the WALE project, similarity mapping provides a useful tool for several aspects of climate modeling and data assimilation.

Results may be summarized as follows.

1) Spatiotemporal similarity between MM5 model fields and ERA-40, UDEL(MW), NCEP1, and PATH datasets depends on seasonality, climate variable, and geographic location.
2) Similarity of data and models is better for temperature than for precipitation.
3) Modeling of summer precipitation fields, and to a lesser extent, temperature fields, appears more problematic than that of winter fields.
4) The geographic distribution of areas with best and worst agreement shifts throughout the year for both temperature and precipitation, with generally better agreement between maps and model fields in the northeastern and northern inland areas than in topographically complex and near-coastal areas. However, for summer precipitation fields, areas indicative of modeling problems lie in the central and northern parts of the study area.
5) Seasonality effects are underrepresented in MM5, in particular the high variability of summer precipitation, and to a lesser extent, summer temperature.
Differentiating among the datasets, we draw the following conclusions for the time range 1992–2000:

6) MM5 successfully predicts January temperature fields that are very similar to ERA-40 reanalysis data, UDEL(MW) interpolated field data, NCEP1 reanalysis data, and temperature proxies derived from Pathfinder satellite data.

7) MM5 July temperature fields match ERA-40 reanalysis data and Pathfinder data, whereas similarity maps of UDEL(MW) versus MM5 and NCEP1 versus MM5 indicate various problematic areas.

8) January precipitation fields from UDEL(MW), ERA-40, and NCEP1 are matched very well by MM5 for all but the coastal regions. The poorest agreement is seen between MM5 and NCEP1.

9) Precipitation in July shows poor agreement for ERA-40 and NCEP1, and fairly good agreement for UDEL(MW) in the northern half of the study area (north of 60°N).

The poor agreement between MM5 and the NCEP–NCAR reanalysis data, which seems unexpected at first, may be explained by the fact the reanalysis data provide lateral boundary conditions and initial conditions for the MM5 model runs, but in subsequent model time steps the physical equations that govern the MM5 model starts to dominate, moving the MM5 model fields away from the reanalysis data fields.

In model validation, similarity maps are useful in the identification of areas that are problematic in a model. Investigation of the climate components and weather patterns that dominate in a problem area may aid in determining processes or variables that are not yet part of the model or are parameterized unrealistically, and thus may aid in improvement of physical models. Based on the similarity maps and their analysis, we conclude that the high interannual variability of summer precipitation is not adequately modeled. MM5 has a higher interannual variability for summer than for winter, but the actual climatic variability is even larger. Close investigation of the geographic distribution of areas with poor similarity values, which indicates mismatches between datasets and models, suggested the following possibilities for model improvement: 1) Seasonality effects are underrepresented in MM5; in particular the higher variability in summer months should be accounted for by the temperature variable and even more, the precipitation variable. 2) The energy of topographic relief may be parameterized in the model to generate a higher spatial variability in climate variables over mountain ranges. A better digital elevation model (DEM) may be helpful in this context, as the presently used DEM has maximum elevations of about 1200 m (rather than 6300 m) and represents the topographic roughness incorrectly. 3) Parameterization of proximity to the coast may be useful.

The present study investigates the similarity of spatial MM5 results with maps based on the current state of climate data fields. Datasets may also benefit from improvement in topographically complex areas where climate stations are not sufficiently dense to capture the high regional variability that exists in mountain weather and climate (see Barry 1992). Inclusion of elevation information from a digital elevation model in the interpolation and regionalization procedure (e.g., Phillips et al. 1992; Hudson and Wackernagel 1994; Goovaerts 2000) will only
lead to improved results if topographic information is actually used in the form in which it affects the variability in the climatic element within the study area. In most cases, climatologically relevant processes, such as orographic lifting of air masses, are influenced by morphological aspects and relative elevation differences. Approaches for interpolation of climate data with respect to topographic effects are presented in Benichou and Lebreton (Benichou and Lebreton 1987), Prudhomme and Reed (Prudhomme and Reed 1999), and Thomas and Herzfeld (Thomas and Herzfeld 2004).

In summary, the results of our study provide the scientific community with a better understanding of regional similarities and discrepancies in datasets and between datasets and MM5 model results, which may be used to assess the capabilities and limitations of present climate data and model fields for the Arctic region, contribute to developing improved climate data/model fields in the future, and expand on the original goals of the WALE project.

Acknowledgments. Work supported by NSF Award ARC 0095047: “Collaborative Research: Modeling the Role of High Latitude Terrestrial Ecosystems in the Arctic System: A Retrospective Analysis of Alaska as a Regional System.” Thanks are due to Dave McGuire (University of Alaska Fairbanks) for useful discussions, to Doug Young (National Snow and Ice Data Center), Steve Hart (CCAR), and Susanna Gross (CIRES) for system administration.

References


