JOURNAL OF APPLIED AND SERVICE CLIMATOLOGY

Published Online September 5, 2025

An Assessment of Temperature from Multiple Gridded Climate Datasets in a Region with Strong Physiographic Gradients: Michigan, United States

Michael T. Kiefer[1], Jeffrey A. Andresen[1], and William J. Baule[1,2]

[1] *Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI*
[2] *Current Affiliation: Department of Atmospheric Sciences and Southern Regional Climate Center, Texas A&M University, College Station, TX*

Corresponding Author: Michael T. Kiefer, mtkiefer@msu.edu

In this study, three gridded climate datasets are examined for their ability to accurately reproduce temperature climatologies at observing sites across Michigan, United States during the period 1981-2018: Parameter-elevation Regressions on Independent Slopes Model (PRISM, versions D1 and D2), Gridded Surface Meteorological (gridMET), and Topography Weather (TopoWx). Michigan is the focus of this study since a comprehensive assessment of gridded climate datasets in a non-mountainous region with otherwise complex physiography is lacking. Two observational networks are utilized, one not assimilated into any of the gridded datasets (Enviroweather), and the other assimilated into all three datasets (Global Historical Climatology Network-Daily). Overall, TopoWx is found to exhibit the smallest deviation from observed daily temperatures, $PRISM_{D2}$ is found to exhibit the smallest deviations from observed annual extreme minimum temperatures, and gridMET is found to exhibit the largest estimate–observation differences across all timescales. Maps of contoured PRISM-version temperature differences reveal distinct spatiotemporal patterns. Generally speaking, differences are maximized along the lakeshores, in areas with low station density, and at airport station sites. Despite the challenges involved in assessing gridded climate datasets, results of such studies are critical sources of user information. Knowledge of dataset strengths and weaknesses can help inform users as to which datasets and variables are likely to be most accurate or appropriate for a given application and timescale. Study results are expected to be applicable to other regions where terrain gradients, complex coastlines, and land use patterns are important climate controls.

1. Introduction

Gridded climate datasets are used by researchers and practitioners in such varied sectors as agriculture, hydrology, risk management, and natural resource conservation (e.g., Daly et al. 2008; Rogers et al. 2017; Kiefer et al. 2019, 2022). Such spatiotemporally complete datasets are utilized in a multitude of ways, for example, as data sources for climatological analyses, climate model initial conditions, and crop model inputs. Examples of gridded climate datasets include the National Centers for Environmental Prediction (NCEP)/ National Corporation for Atmospheric Research (NCAR) Global Reanalysis (Kalnay et al. 1996), North American Regional Reanalysis (NARR; Mesinger et al. 2006), Climate Forecast System Reanalysis (CFSR; Saha et al. 2010), European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis 5th Generation (ERA5; Hersbach et al. 2020), nClimGrid-Daily (Durre et al. 2022), Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 2008), Gridded Surface Meteorological (gridMET; Abatzoglou 2013), and Topography Weather (TopoWx; Oyler et al. 2014, 2015, 2016). Spatiotemporal scales vary broadly across the datasets, from global to regional and annual to sub-daily, and variables contained in such datasets include, for example, temperature, precipitation, solar radiation, humidity, and wind speed. Although methodologies differ considerably between datasets, all gridded climate datasets are created by interpolating irregularly spaced observations to a regular grid. Interpolation techniques are used to nudge a background or first-guess grid toward observed values, with the background grid originating from either physics-based forecast model output (e.g., NARR), physiographic model output (e.g., PRISM), or some combination thereof (e.g., gridMET).

Regardless of methodology, all gridded climate datasets contain known and unknown errors and biases (e.g., Daly et al 2008; Oyler et al. 2014; Kiefer et al. 2019, 2022).

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

Although it is tempting to download a gridded dataset and begin using the data immediately, it is important to first evaluate dataset accuracy (e.g., Behnke et al. 2016; Walton and Hall 2018; Kiefer et al. 2019; Newman et al. 2019; Blankenau et al. 2020; Estes et al. 2022). Behnke et al. (2016) evaluated eight gridded datasets at weather stations across the continental United States (US), including PRISM, gridMET, and TopoWx, and found that the dataset with the best verification statistics varied regionally. They also found that weather variables with values closer to long-term average conditions were reproduced more accurately than extreme values near the distribution edges. They recommended that care be taken in selecting a gridded dataset, as no single dataset performed well across all regions, variables, and applications. Walton and Hall (2018) examined eight gridded datasets at weather stations across California and found that gridded datasets developed from physics-based forecast model background grids (e.g., NARR) exhibited systematic biases, whereas gridded datasets developed from physiographic model background grids (e.g., PRISM) exhibited biases in areas of strong physiographic gradients (e.g., coastlines, complex topography) and unphysical trends due to temporal inhomogeneities at individual weather stations.

Gridded climate datasets contain estimates of the current and past state of the atmosphere and it is prudent to treat them with some degree of skepticism. Ultimately, the user must decide how much estimate error is acceptable for their application. One complicating factor for gridded analyses as opposed to forecasts is that the gridded analysis and verification datasets are not necessarily independent of each other. Comparison of a gridded climate dataset to observations that were assimilated into that dataset may lead to an underestimation of gridded estimate error. Prior gridded dataset evaluations have either utilized exclusively non-independent observational datasets (e.g., Walton and Hall 2018) or have performed limited evaluation with independent observational datasets (e.g., Behnke et al. 2016). Additionally, it is important to keep in mind the uncertainty inherent in gridded climate datasets due to unresolved microclimatic variability.

Furthermore, as the values at each grid cell center (i.e., point) are area averages across the cell, an unknown portion of the differences between point observations and gridded estimates can be attributed to sub-grid variability or spatial representativeness error (Lorenc 1986; Daley 1991; Kitchen and Blackall 1992; Bosilovich 2006; Pinson and Hagedorn 2012; Janjić et al. 2018; Prakash et al. 2019). Quantifying sub-grid variability is made challenging by the need for dense networks of observing stations within one or more grid cells. In a study of spatial representativeness error in rural China, Liu et al. (2018), found that surface temperature measurements at station sites can be considered representative of the surrounding area within a radius of ~250-400 m, smallest for heterogeneous land use patterns. The small scale of spatial variability implied by these radii values, generally one or more orders of magnitude smaller than the grid spacing of gridded climate datasets, should be considered when interpreting the results of this and other gridded data assessment studies.

The objective of this study is to evaluate three gridded climate datasets for their ability to accurately reproduce daily and annual extreme temperature climatologies at observing sites across Michigan, US: PRISM, gridMET, and TopoWx. All three datasets are actively used to characterize spatiotemporal climate variability across the US; however, PRISM is the most widely used in research-related applications, with the primary PRISM reference (Daly et al. 2008) cited 1380 times between 2018 and 2022, compared to 134 and 902 citations for the primary TopoWx (Oyler et al. 2015) and gridMET (Abatzoglou 2013) references, respectively (source: Google Scholar, accessed 11 Aug 2023). Although PRISM is the primary focus of this assessment, the other two datasets are evaluated as they are comparable to PRISM in terms of spatiotemporal resolution but differ in methodology and input data sources (Section 2.2). Due to our stated focus on PRISM, and in the interest of keeping the assessment tractable, other gridded datasets examined by Behnke et al. (2016) and Walton and Hall (2018) are not considered here. Furthermore, an assessment of other climate variables is beyond the scope of this study.

This study focuses on Michigan for three reasons. First, although PRISM's origins lie in the complex topography of the western US (Daly et al. 1994, 1997), it is widely used in other US regions (Daly et al. 2008, 2015, 2021), including the Great Lakes (e.g., Duveneck et al. 2014; Seeley et al. 2019; Kiefer et al. 2022). Although PRISM has been evaluated previously (e.g., Daly et al. 2008), a comprehensive evaluation in a non-mountainous region with otherwise complex physiography is lacking. Michigan is characterized by complex coastlines and strong physiographic gradients of terrain, soils, vegetation, and climate, consistent with the broader Great Lakes region (Andresen and Winkler 2009). Second, PRISM is currently used in operational products in Enviroweather (EW), a web-based information system linking real-time weather data, forecasts, and biological and other process-based models for assistance in operational decision-making and risk management associated with Michigan's agriculture and natural resource industries (Andresen et al. 2012). Third, Michigan contains a mesoscale network (mesonet) of surface weather stations (EW) that are not assimilated into PRISM, gridMET, or TopoWx. Due to the relatively short

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

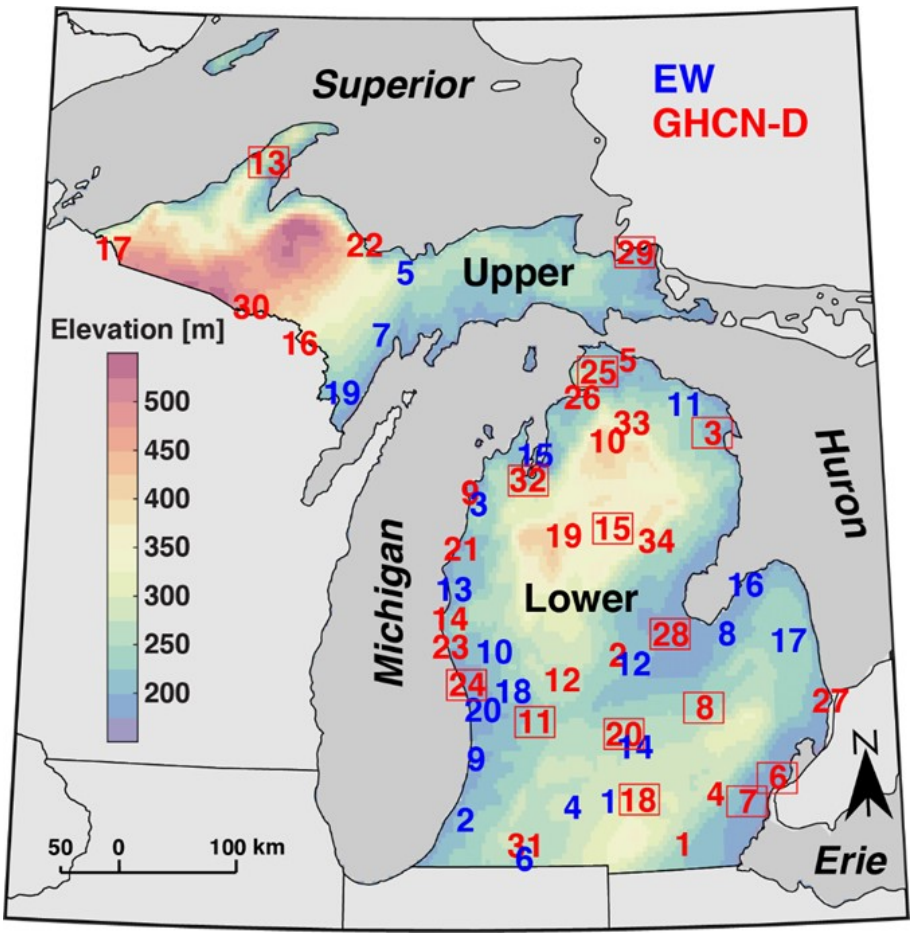

FIGURE 1: Michigan surface elevation [m] used in PRISM data products, with numbers denoting the locations of EW (blue) and GHCN-D (red) station sites; rectangle around number indicates airport site. See Tables 1-2 for station site metadata. The Upper and Lower Peninsulas and the four Great Lakes bordering Michigan are labeled.

period of record of the EW network (1996-present), and gaps in spatial coverage across Michigan, Global Historical Climatology Network-Daily (GHCN-D) station data is used to supplement the EW station analysis. GHCN-D is a global database of daily surface station climate summaries that have been subjected to quality assurance reviews (Menne et al. 2012). It is important to keep in mind that whereas EW observations are not assimilated into the gridded climate datasets examined in this study, GHCN-D observations are, rendering the gridded estimate and GHCN-D datasets non-independent (e.g., GHCN-D observations are assimilated into PRISM).

Although this study focuses on Michigan, we anticipate our results will be applicable to other areas of the US (e.g., Maine coastline) and other regions of the world [e.g., Black Sea (Eurasia)]. where terrain gradients, complex coastlines, and heterogeneous land use patterns are important climate controls. Thus, characterizing the ability of gridded climate datasets to reproduce daily and annual extreme temperature climatologies in Michigan has potential benefits beyond this specific study area.

2. Materials and Methods

2.1 Study Area

Michigan is located within the North American Great Lakes region (Fig. 1). This region encompasses the largest freshwater system in the world and is characterized by strong physiographic gradients of terrain, soils, vegetation, and climate (Curtis 1959; Anderson 2005; Andresen and Winkler 2009; Andresen 2012). During the autumn and winter (spring and summer) seasons, when lake surface temperatures are warmer (cooler) than the overlying air and adjacent land surfaces, cloud cover is enhanced (reduced) in Michigan relative to states upwind of the Great Lakes (Changnon and Jones 1972; Yu et al. 2014). Cloud cover effects combined with the direct moderating impact of the lakes on the overlying atmosphere yield higher (lower) mean temperatures downwind of the Great

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

Lakes in the autumn and winter (spring and summer) seasons, and overall higher mean temperatures annually. A review of the 1981-2010 PRISM climate normals (not shown; used in the production of PRISM daily temperatures, Section 2.2.1) reveals that January mean temperatures vary from higher than -4 °C in parts of the Lower Peninsula (hereafter, "Lower Michigan"), mainly along the southern Lake Michigan shore and the Detroit urban heat island, to lower than -10 °C across the western Upper Peninsula (hereafter, "Upper Michigan"), mainly in higher-elevation areas away from the lakeshores. July mean temperatures vary from higher than 22 °C across far southern Lower Michigan to lower than 18 °C in parts of Upper Michigan, including near the Lake Superior shore. Regarding diurnal variability, locations downwind of the Great Lakes experience a reduced diurnal range with higher daily minimum temperatures and lower daily maximum temperatures, relative to upwind locations (Changnon and Jones 1972; Scott and Huff 1996; Lofgren 1997; Notaro et al. 2013). Finally, the influence of the Great Lakes diminishes with distance inland, and is complicated by patterns of surface elevation, vegetation, and soil (Andresen and Winkler 2009).

2.2 Gridded Datasets

As stated previously, the objective of this study is an evaluation of three gridded climate datasets for their ability to accurately reproduce daily and annual extreme temperature climatologies at observing sites across Michigan, US: PRISM, gridMET, and TopoWx. Recall that although PRISM is the primary focus of this assessment, the other two datasets are evaluated as they have similar spatiotemporal resolutions to PRISM, but differing methodologies and input data sources. A brief description of the gridded datasets follows.

2.2.1 PRISM

PRISM is a gridded climate analysis system that interpolates individual station site observations to an 800-m grid covering the contiguous US (Daly et al. 1994, 1997, 2008), followed by interpolation to a 4-km grid for public distribution. PRISM datasets span a broad range of timescales, from 30-year normals, to annual, monthly, and daily time series, and include six daily climate elements: maximum and minimum temperature, mean dew point, maximum and minimum vapor pressure deficit, and precipitation. The interpolation process for annual, monthly, and daily time series begins with a background gridded dataset of 30-year climatological normals, developed in an earlier stage from long-term observations and a digital elevation model (DEM). Unique regression equations are developed for each grid point to produce estimates based on observations from nearby weather stations and the background value. Station weighting is based on five factors: distance, elevation, vertical layer, topographic facet, and coastal proximity (Daly et al. 1994, 1997, 2008). It is important to keep in mind that PRISM temperatures, as with the other gridded datasets examined in this study, are valid at 1.5 m above ground level, are interpolated from sheltered weather stations, and represent grid cell averages.

In this study, we evaluate two versions of the PRISM daily product: $PRISM_{D1}$ and $PRISM_{D2}$, released in June 2013 and October 2019, respectively. Changes to methodology and inputs implemented in $PRISM_{D2}$ include (Daly 2022): (1) an adjustment to correct for SNOw TELemetry (SNOTEL) Yellow Springs Instruments (YSI) Extended Range temperature measurements that were incorrectly transformed from voltage to temperature by the Natural Resources Conservation Service (NRCS); (2) an algorithm to day-shift National Weather Service (NWS) Cooperative Observer Program (COOP) temperature observations that were reported on the wrong day; (3) assimilation of gridded upper-air temperature and dew point data to improve accuracy in complex terrain; (4) modification of the interpolation procedure to better render strong temperature inversions. For this study, 4-km PRISM daily temperatures were downloaded for the period from 1 Jan 1981 to 31 Dec 2018 (see data availability statement).

2.2.2 gridMET

gridMET (Abatzoglou 2013) is a gridded dataset that combines the daily timescale of PRISM with the hourly timescale of the North American Land Data Assimilation System version 2 (NLDAS-2; Xia et al. 2012a, b) (in turn, derived from NARR). As a hybrid dataset, gridMET combines PRISM's high spatial resolution at daily, monthly, and longer timescales with NLDAS-2's high temporal resolution at sub-daily timescales. Primary gridMET variables include daily minimum and maximum temperature, downward surface shortwave radiation, wind velocity, maximum and minimum relative and specific humidity, and precipitation. To create gridMET daily temperature, the PRISM–NLDAS-2 monthly-mean temperature difference is added to the NLDAS-2 daily temperature. No data assimilation is performed beyond that which is performed in the production of PRISM and NLDAS-2. gridMET is available from 1979–present on a 4-km contiguous US grid, and was downloaded for the period 1981-2018 (see data availability statement).

2.2.3 TopoWx

TopoWx is a gridded dataset of daily minimum and maximum temperature based on station observations, DEM data, and remotely-sensed land skin temperature data

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

TABLE 1: Metadata for EW station sites: Station number, name, ID, latitude [deg], longitude [deg], elevation above mean sea level [m], and percent days missing [%]. The date range for the missing data check is 1 Jan 2004 – 31 Dec 2016.

Station Number	Name	ID	Latitude	Longitude	Elevation	% Missing
E1	Albion	alb	42.2619	-84.7741	289	0.63
E2	Bainbridge	bbc	42.1267	-86.2677	225	0.69
E3	Benzonia	bnz	44.5593	-86.1172	253	1.52
E4	Ceresco	cer	42.2206	-85.1526	289	3.01
E5	Chatham	cth	46.3405	-86.9273	270	7.12
E6	Constantine	cnt	41.8277	-85.6616	254	1.64
E7	Escanaba	esc	45.7452	-87.1844	221	1.22
E8	Fairgrove	fgv	43.5276	-83.4864	201	0.63
E9	Fennville	fev	42.5951	-86.1561	214	0.65
E10	Fremont	frm	43.4225	-85.9551	233	1.56
E11	Hawks	haw	45.2994	-83.8528	252	1.35
E12	Ithaca	ith	43.3154	-84.4884	219	2.53
E13	Ludington	ldt	43.9043	-86.3798	200	1.39
E14	East Lansing	msu	42.6734	-84.4870	264	2.25
E15	Old Mission	old	44.9320	-85.4966	219	1.01
E16	Pigeon	pig	43.8992	-83.2667	188	0.55
E17	Sandusky	sdk	43.4556	-82.8375	237	0.65
E18	Sparta	spo	43.1167	-85.7560	267	0.53
E19	Stephenson	stv	45.4082	-87.6096	208	6.49
E20	West Olive	weo	42.9715	-86.0765	194	0.93

(Oyler et al. 2014, 2015, 2016). Unlike PRISM or gridMET, TopoWx applies a temporal homogenization algorithm developed by Menne and Williams (2009) to correct for changes in observation practices, station siting, and instrumentation during the period of record. Missing values are filled by comparing with non-missing neighboring observations and applying spatial regression (Durre et al. 2010). Climate normals are computed using regression kriging, and daily anomalies are computed using moving window geographically weighted regression and inverse distance weighting. To help estimate climate normals in complex terrain and regions with low station density, TopoWx uses remotely sensed land skin temperature as an auxiliary predictor. Although no longer in production, TopoWx is available from 1948–2016 on an 800-m contiguous US grid, and was downloaded for the period 1981-2016 (see data availability statement).

2.3 Station Observations

The observation sites examined in this study are displayed in Fig. 1 and summarized in Tables 1-2. A total of 20 EW stations are examined, 17 in Lower Michigan and 3 in Upper Michigan, whereas a total of 34 GHCN-D stations are examined, 28 in Lower Michigan and 6 in Upper Michigan. Recall from Section 1 that EW is a web-based information system supporting Michigan's agriculture and natural resource industries, and GHCN-D is a global database of daily surface station climate summaries used in this study to supplement the EW station analysis. In contrast to the homogeneous EW network, the 34 GHCN-D stations consist of 14 airport sites and 20 COOP sites (Table 2). The EW and GHCN-D station sites were chosen to represent a range of latitudes, elevations, Great Lake proximities, and land-use characteristics. Of the 54 stations examined in this study, all but one [Hancock Houghton County AP (GHCN-D)] exhibit less than 10% missing data during the study period.

EW observations were subjected to a three-step quality assurance / quality control (QA/QC) procedure consisting of gross-range, step, and outlier checks similar to those

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

TABLE 2: Metadata for GHCN-D station sites: Station number, name, ID, latitude [deg], longitude [deg], elevation above mean sea level [m], and percent days missing [%]. An asterisk next to the GHCN-D site name denotes an airport (AP) site (14 stations). The date range for the missing data check is 1 Jan 1981 – 31 Dec 2016. For G16, Wastewater Treatment Plant is abbreviated WWTP.

Station Number	Name	ID	Latitude	Longitude	Elevation	% Missing
G1	Adrian 2 NNE	USC00200032	41.9163	-84.0158	232	2.68
G2	Alma	USC00200146	43.3869	-84.6482	224	1.50
G3	Alpena County Regional AP*	USW00094849	45.0716	-83.5644	209	0.02
G4	Ann Arbor Univ. Of Michigan	USC00200230	42.2981	-83.6639	274	0.05
G5	Cheboygan	USC00201492	45.6527	-84.4725	179	2.35
G6	Detroit City AP*	USW00014822	42.4092	-83.0100	191	3.38
G7	Detroit Metro AP*	USW00094847	42.2313	-83.3308	192	0.00
G8	Flint Bishop Intl. AP *	USW00014826	42.9666	-83.7494	235	0.01
G9	Frankfort 2 NE	USC00202984	44.648	-86.2100	289	5.44
G10	Gaylord	USC00203096	45.0332	-84.7113	412	3.60
G11	Grand Rapids Gerald R. Ford Intl. AP*	USW00094860	42.8825	-85.5239	237	0.00
G12	Greenville 2 NNE	USC00203429	43.2025	-85.2422	269	0.99
G13	Hancock Houghton County AP*	USW00014858	47.1686	-88.4889	334	14.25
G14	Hart 3 WSW	USC00203632	43.6747	-86.4238	235	8.82
G15	Houghton Lake Roscommon County AP*	USW00094814	44.3591	-84.6738	351	0.00
G16	Iron Mt. Kingsford WWTP	USC00204090	45.7858	-88.0841	326	1.36
G17	Ironwood	USC00204104	46.4656	-90.1892	436	1.06
G18	Jackson Reynolds Field*	USW00014833	42.2667	-84.4667	304	0.43
G19	Lake City Experimental Farm	USC00204502	44.3088	-85.2050	375	0.80
G20	Lansing Capital City AP*	USW00014836	42.7761	-84.5997	256	0.00
G21	Manistee 3 SE	USC00205065	44.2113	-86.2938	204	4.61
G22	Marquette	USW00014838	46.5458	-87.3795	203	2.62
G23	Montague 4 NW	USC00205567	43.4614	-86.4175	198	4.17
G24	Muskegon County AP*	USW00014840	43.1711	-86.2367	191	0.01
G25	Pellston Regional AP*	USW00014841	45.5644	-84.7927	215	0.08
G26	Petoskey	USC00206507	45.3725	-84.9766	183	2.55
G27	Port Huron	USC00206680	42.9750	-82.4194	180	2.53
G28	Saginaw MBS Intl. AP*	USW00014845	43.5331	-84.0797	201	0.00
G29	Sault Ste Marie Sanderson Field*	USW00014847	46.4794	-84.3572	220	0.02
G30	Stambaugh 2 SSE	USC00207812	46.0555	-88.6275	442	2.59
G31	Three Rivers	USC00208184	41.9299	-85.6385	247	2.94
G32	Traverse City Cherry Capital AP*	USW00014850	44.7408	-85.5825	188	0.13
G33	Vanderbilt 11 ENE	USC00208417	45.1702	-84.4397	276	2.67
G34	W Branch 3 SE	USC00208800	44.2541	-84.2011	270	1.53

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

applied across the Oklahoma mesonet (Shafer et al. 2000). GHCN-D observations were subjected to QA/QC procedures prior to being released to the public (Durre et al. 2010; Menne et al. 2012); no additional processing was applied in this study. The reader is cautioned that although COOP observation-day reporting inconsistencies are considered in the GHCN-D QA/QC system, and observations with QA/QC flags were masked in this study, the PRISM group identified inconsistent observing practices missed by the GHCN-D QA/QC system that have been rectified in $PRISM_{D2}$ using a day-shifting algorithm to reassign COOP temperature observations that were reported on the wrong day (Daly 2022). To the best of the authors' knowledge, the GHCN-D day-shifting algorithm implemented in $PRISM_{D2}$ is not publicly available, and thus was not implemented in this study. The GHCN-D observations were used to develop the gross-range, step, and outlier thresholds used for the EW data QA/QC. GHCN-D and EW data were downloaded for the periods 1981-2018 and 2004-2018, respectively, with the choice of time periods informed by the PRISM daily product period of record (GHCN-D) and availability of station records (EW) (see data availability statement).

Temperature was measured at the EW stations using Vaisala HMP 45C hygrothermometers, without mechanical aspiration; and at the GHCN-D stations using either Automated Surface Observing System (ASOS) 1088 hygrothermometers with mechanical aspiration (airport sites post-ASOS installation, about 1996 – present), HO-83 hygrothermometers with mechanical aspiration (airport sites pre-ASOS installation, before about 1996), or maximum-minimum temperature (MMTS) sensors without mechanical aspiration (COOP, all years). A comparison of gridded estimate–observation differences at airport sites before (1985-1994) and after (2005-2014) ASOS implementation was performed to quantify the potential impact of instrumentation changes on the gridded data assessment ($PRISM_{D2}$ only; Table S1, online supplemental material); results were similar for the other gridded datasets and are omitted. Note that the online supplemental material contains figures and tables, identified by a leading "S", that may be of interest to some readers, but are not essential for evaluating study methods, results, or conclusions. Across the eight station sites chosen for this exercise, gridded estimate–observation differences for daily minimum temperature were lower (more negative) during the ASOS decade compared to the pre-ASOS decade, whereas estimate–observation differences for daily maximum temperature differed little between the two decades. As ASOS hygrothermometers are known to report lower minimum temperatures than the pre-ASOS HO-83 hygrothermometers (National Research Council 2012), the more negative gridded estimate–observation differences for daily minimum temperature during the ASOS decade suggests factors other than instrumentation alone (e.g., changes to station siting).

2.4 Analysis Methodology

Assessment of the gridded datasets takes place in two stages. In the first stage, gridded estimates at the nearest grid point to each station site are compared with corresponding observations. Note that no attempt is made to interpolate the gridded datasets to a common grid; the reader is urged to keep in mind the difference in grid cell dimensions between PRISM and gridMET (4 km), and TopoWx (800 m). For a given station, the distance from grid-cell center to the station site may vary from one gridded dataset to another; such differences in distance contribute to an unknown degree to differences in comparison statistics between the gridded datasets (recall the discussion of sub-grid variability in Section 1). Maps are presented with markers denoting station sites color-coded by two statistical measures of estimate–observation differences: mean difference (MD) and root-mean-square difference (RMSD):

$$\boldsymbol{MD} = \frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}(\boldsymbol{E_i} - \boldsymbol{O_i}) \qquad (1)$$

$$\boldsymbol{RMSD} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}(\boldsymbol{E_i} - \boldsymbol{O_i})^{\mathbf{2}}} \qquad (2)$$

where O and E are the observed and estimated time series, respectively, and N is the number of values (e.g., Fig. 2). Corresponding tables containing network-median statistics are provided to assess overall differences across Michigan (e.g., Table 3). To help assess the statistical significance of differences in estimate–observation statistics between the gridded datasets, p values from a two-tailed t-test are included in the corresponding discussion. Both daily minimum and maximum temperatures (TMINd and TMAXd, respectively) and annual extreme minimum and maximum temperatures (TMINa and TMAXa, respectively) are examined. The maps of gridded estimate–observation differences include all months, whereas the tables depict full-year statistics and statistics for the four climatological seasons: winter (December-February), spring (March-May), summer (June-August), and autumn (September-November). In the second stage, maps of contoured differences between $PRISM_{D1}$ and $PRISM_{D2}$ are presented. The purpose of this second stage is to assess spatial patterns of PRISM version differences and to qualitatively examine relationships between spatial difference patterns and Mich-

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

igan's physiography.

Three time periods are considered in this study, 2004-2016, 1981-2016, and 1981-2018. The primary time period, 2004-2016, is utilized for comparison of all gridded datasets and observational networks, for three reasons: First, ASOS instrumentation was installed at the GHCN-D airport sites from about 1996 onward, introducing temporal inhomogeneities in the station time series; second, EW observations are less complete prior to 2004; third, TopoWx production ceased at the end of 2016. The secondary 1981-2016 time period is used for comparison of the four gridded datasets to GHCN-D observations alone (without EW). The tertiary 1981-2018 time period is used for assessment of the two PRISM versions and GHCN-D observations only (without gridMET, TopoWx, or EW observations), and is included because Kiefer et al. (2022) used this time period for their climatological study of extreme minimum temperatures in the US Great Lakes region. Lastly, the gridded dataset names are abbreviated in the figures and tables as follows: $PRISM_{D1}$ (P1), $PRISM_{D2}$ (P2), gridMET (G), and TopoWx (T).

2.5 Day Definitions

Before proceeding to results of the gridded climate dataset assessment, it is important to comment briefly on day definitions in the gridded datasets and observational networks. Although it is tempting to neglect this methodological detail and treat all day definitions as equivalent, it is important to limit (or in the least, be aware of) the degrees of freedom that complicate the gridded estimate–observation comparison. More generally, it is advisable that gridded dataset users be aware of how time periods are defined before using the data for a particular application. Regarding the datasets examined in this study, PRISM defines a day as the 24-hour period ending at 12:00 Coordinated Universal Time (UTC) (hereafter, PRISM day), and gridMET and TopoWx define a day as the 24-hour period ending at midnight local time (LT; hereafter, calendar day). Note that the term "calendar day" is used herein to refer to the day on which the maximum and minimum temperature occurred (i.e., 00:00:00 LT – 23:59:59 LT on day n), not the day the temperature was recorded (i.e., 00:00:00 LT day n report for the period 00:00:00 LT – 23:59:59 LT day n-1). Of the two observational networks, EW utilizes the same calendar day definition as gridMET and TopoWx, and GHCN-D utilizes multiple day definitions (e.g., calendar day, 24-hour period ending at 14:00 LT), that vary from station to station and, at some sites, change during the study period.

To investigate possible impacts on the gridded data assessment of day definition discrepancies between PRISM and the two observational networks, MD and RMSD were computed with and without a one-day time series lag (observation: day n, PRISM: day n+1) (TMAXd: Figs. S1-S2 and Tables S2-S4; TMINd: not shown). For both TMINd and TMAXd, MD was found to vary little between tests with and without the one-day PRISM lag, due to the cancellation of positive and negative estimate–observation differences. For TMAXd, however, RMSD was found to be substantially smaller with the one-day PRISM lag implemented, as the squaring of estimate–observation differences prevented cancellation; for TMINd, RMSD was found to vary little between tests. Generally speaking, TMAXd tends to occur during the mid to late afternoon hours (about 19:00-22:00 UTC in the central and eastern US), meaning that for a given calendar day, the PRISM TMAXd valid at 12:00 UTC on that day likely occurred during the prior afternoon. On the other hand, TMINd typically occurs around or shortly after sunrise (about 10:00-13:00 UTC in the central and eastern US), meaning that the calendar and PRISM day definitions are generally in sync. Based on these lag tests, a decision was made to partially account for day definition differences using a one-day lag for PRISM TMAXd (but not for PRISM TMINd or the other gridded datasets). The reader is urged to keep the simplicity of the one-day lag in mind when interpreting the results of this study: the day-definition differences between calendar-day observations and PRISM-day estimates are only partially mitigated by this step.

An additional test was performed at the 14 GHCN-D airport sites (Fig. 1; Table 2) using the National Centers for Environmental Information (NCEI) Integrated Surface Database (ISD; Smith et al. 2011) hourly observations ($PRISM_{D1}$ and $PRISM_{D2}$ only; Table S5). Generally speaking, replacing the calendar-day TMINd from GHCN-D with the PRISM-day TMINd computed from ISD hourly observations resulted in a change of MD sign from positive to negative and a two- to five-fold decrease in RMSD; for TMAXd, the switch to the hourly observation-based daily value resulted in a change of MD sign from negative to positive and an even larger reduction in RMSD than for TMINd. The use of observational networks and gridded datasets that use calendar day definitions precludes standardization of day definitions in this study. However, this exercise helps illustrate how the mixture of day definitions and observational data sources complicates the assessment of gridded climate datasets. This is an important limitation of our methodology that the reader is strongly urged to consider when interpreting the study results: the PRISM estimate–observation difference statistics presented in this study are likely exaggerated due to unmitigated day definition incongruities, particularly for RMSD, wherein cancellation of positive and negative values is prohibited by the squaring of differences.

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

3. Results and Discussion

3.1 Point Assessment

The assessment begins with maps of full-year gridded estimate–observation differences at the EW and GHCN-D station sites on the daily timescale (MD: Fig. 2; RMSD: Fig. 3; N: not shown), along with network-median statistics (EW: Table 3; GHCN-D: Table 4). Recall that the primary analysis period in this study is 2004-2016; for corresponding figures and tables covering the secondary analysis period, 1981-2016 (GHCN-D only), see Figs. S3-S4 and Table S6. Focusing on $PRISM_{D1}$ and $PRISM_{D2}$ in the top rows of Figs. 2-3, three aspects of estimate–observation differences become apparent. First, there is a general pattern of positive (negative) MD values for TMINd (TMAXd), indicating a tendency for both PRISM versions to underestimate the diurnal temperature range. Second, there is little indication in either the MD or RMSD plots for larger or smaller estimate–observation differences near the lakeshores than farther inland. Third, differences in the statistics between the two PRISM versions are generally small, a finding which is supported by the network-median statistics at the EW and GHCN-D station sites (Tables 3-4), and corresponding two-tailed t-tests at the EW stations sites: MD values differ between the two versions by 0.03 °C for TMINd (p values above .05 at 12/20 stations) and 0.12 °C for TMAXd (p values above .05 at 13/20 stations).

Broadening the analysis to all four gridded datasets, TopoWx exhibits the smallest deviation from observed TMINd and TMAXd, a result that is true for both systematic (MD) and random (RMSD) differences (Tables 3-4). A two-tailed t-test comparing MD between TopoWx and $PRISM_{D1}$ reveals p values below .05 at 16/20 (TMINd) and 18/20 (TMAXd) EW stations; thus, the improvement of TopoWx over PRISM on the daily timescale appears to be statistically significant. Possible reasons for TopoWx's superior performance include the integration of 800-m DEM data and 1-km land skin temperature data in the station interpolation process at the daily timescale, and the temporal homogenization of station data. However, the reader is reminded that the PRISM estimate–observation difference statistics presented in this study are likely exaggerated due to unmitigated day definition incongruities, particularly for RMSD, wherein cancellation of positive and negative values is prohibited by the squaring of differences. In contrast, gridMET exhibits the largest overall estimate–observation differences, particularly at the independent EW stations, wherein TMINd and TMAXd exhibit network-median MD of 0.32 and 0.24 °C, respectively [differences between gridMET and $PRISM_{D1}$ MD: p values below .05 at 16/20 (TMINd) and 18/20 (TMAXd) EW stations]. A possible source for this behavior is the propagation of NLDAS-2 temperature error (Behnke et al. 2016; Walton and Hall 2018) into gridMET. In other words, errors identified in gridMET may have their origin in errors in the parent dataset, NLDAS-2 (and, in turn, NARR) (Section 2.2.2). Overall, it is worth noting that MD and RMSD are generally smaller for GHCN-D than EW, consistent with the assimilation of GHCN-D observations into

TABLE 3: Summary statistics for 2004-2016 estimated TMINd and TMAXd [°C], with seasonal breakdown, at EW station sites: 20-station median number of days (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P_1 ($PRISM_{D1}$), P_2 ($PRISM_{D2}$), G (gridMET), and T (TopoWx). Note: for TMAXd, difference in day definition between EW and PRISM is partially accounted for by comparing day *n* observation to day *n+1* estimate (Section 2.5).

		TMINd				TMAXd			
	Months	P1	P2	G	T	P1	P2	G	T
N	**ALL**	**4688**	**4688**	**4688**	**4688**	**4671**	**4671**	**4671**	**4671**
	DJF	1146	1146	1146	1146	1138	1138	1138	1138
	MAM	1188	1188	1188	1188	1185	1185	1185	1185
	JJA	1191	1191	1191	1191	1188	1188	1188	1188
	SON	1174	1174	1174	1174	1170	1170	1170	1170
MD	**ALL**	**0.14**	**0.11**	**0.32**	**-0.10**	**-0.33**	**-0.45**	**0.24**	**-0.14**
	DJF	0.13	0.18	0.48	0.10	-0.06	-0.13	0.32	-0.16
	MAM	-0.20	-0.20	0.28	-0.19	-0.35	-0.40	0.06	-0.08
	JJA	0.43	0.33	0.41	-0.07	-0.21	-0.36	0.22	-0.14
	SON	0.10	0.10	0.25	0.05	-0.46	-0.51	0.13	-0.19
RMSD	**ALL**	**2.75**	**2.81**	**2.64**	**1.82**	**1.46**	**1.52**	**2.05**	**1.35**
	DJF	3.53	3.62	3.09	2.22	1.75	1.79	2.36	1.32
	MAM	2.68	2.78	2.50	1.82	1.58	1.61	2.12	1.48
	JJA	1.86	1.89	2.27	1.48	1.07	1.13	1.83	1.24
	SON	2.75	2.83	2.33	1.89	1.41	1.45	1.90	1.29

TABLE 4: Summary statistics for 2004-2016 estimated TMINd and TMAXd [°C], with seasonal breakdown, at GHCN-D station sites: 34-station median number of days (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P1 ($PRISM_{D1}$), P2 ($PRISM_{D2}$), G (gridMET), and T (TopoWx). Note: for TMAXd, difference in day definition between GHCN-D and PRISM is partially accounted for by comparing day *n* observation to day *n+1* estimate (Section 2.5).

		TMINd				TMAXd			
	Months	P1	P2	G	T	P1	P2	G	T
N	**ALL**	**4716**	**4716**	**4716**	**4716**	**4721**	**4721**	**4721**	**4721**
	DJF	1170	1170	1170	1170	1172	1172	1172	1172
	MAM	1193	1193	1193	1193	1194	1194	1194	1194
	JJA	1192	1192	1192	1192	1192	1192	1192	1192
	SON	1175	1175	1175	1175	1176	1176	1176	1176
MD	**ALL**	**-0.01**	**0.04**	**0.05**	**-0.15**	**-0.24**	**-0.32**	**0.09**	**0.05**
	DJF	-0.11	-0.15	0.14	-0.16	-0.20	-0.22	0.15	0.03
	MAM	-0.02	0.07	0.19	-0.12	-0.24	-0.31	0.19	0.17
	JJA	0.03	0.16	-0.07	-0.10	-0.23	-0.30	0.14	0.11
	SON	-0.04	0.05	-0.07	-0.09	-0.34	-0.39	0.00	-0.05
RMSD	**ALL**	**2.43**	**2.48**	**2.91**	**1.52**	**2.18**	**2.16**	**2.27**	**1.56**
	DJF	2.90	2.93	3.38	1.77	2.37	2.32	2.50	1.53
	MAM	2.44	2.41	3.06	1.60	2.50	2.48	2.58	1.92
	JJA	1.44	1.46	2.49	1.23	1.69	1.52	1.95	1.35
	SON	2.48	2.53	2.66	1.51	2.15	2.14	2.18	1.37

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

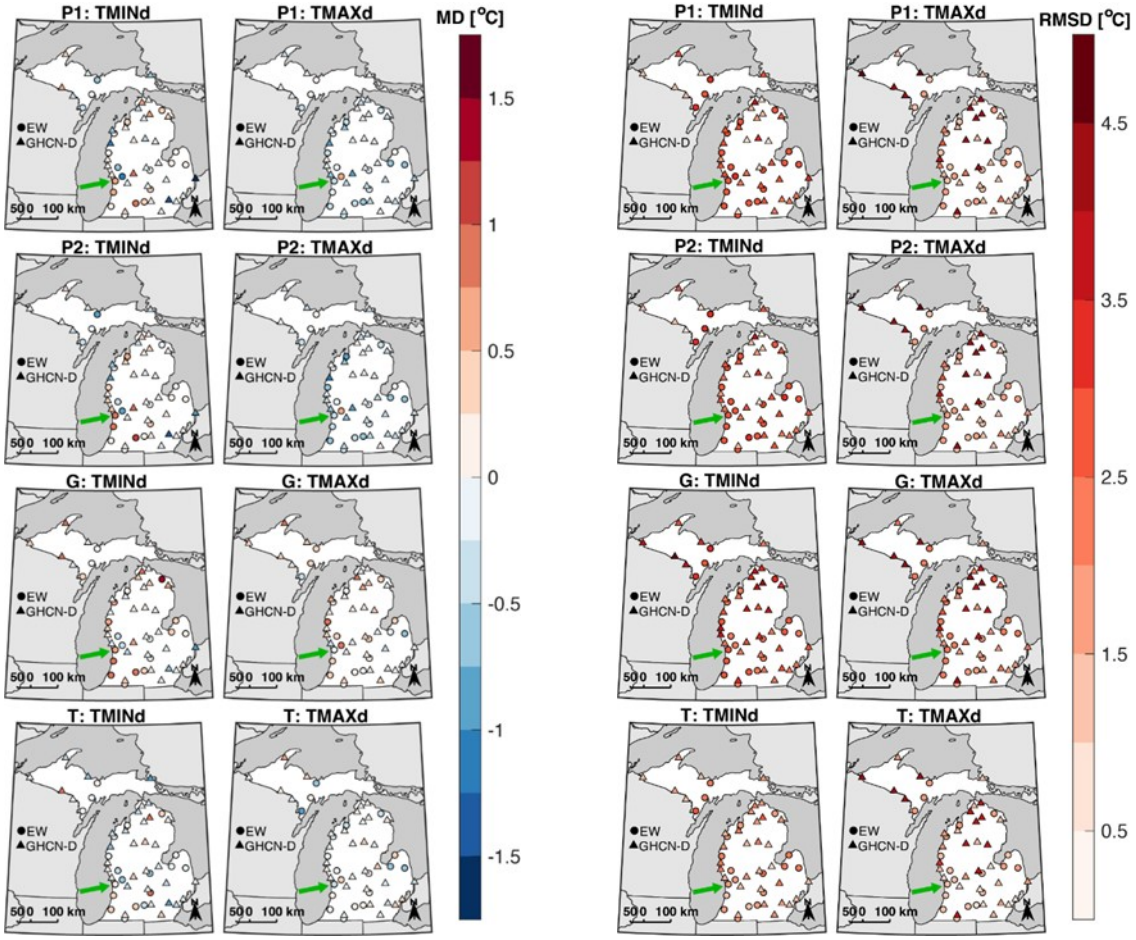

FIGURE 2 (LEFT) AND FIGURE 3 (RIGHT): EW (circle) and GHCN-D (triangle) station symbols, color-coded by daily temperature mean difference (MD) in Figure 2 and daily temperature root-mean-square difference (RMSD) in Figure 3, for period 2004-2016: columns: TMINd and TMAXd (w/ PRISM lag); rows: $PRISM_{D1}$ (P1), $PRISM_{D2}$ (P2), gridMET (G), and TopoWx (T). The minimum number of days (N) among both observational networks is 4190 for TMINd and 4223 for TMAXd; N at individual stations is not shown. Green arrow points to West Olive EW station site (E20 in Fig. 1, Table 1); next circle to northeast is the Sparta EW station site (E18 in Fig. 1, Table 1).

the gridded datasets. An important exception to this general statement is TMAXd RMSD: values are higher for GHCN-D, possibly due to unrectified day definition differences between the gridded estimates and station observations (Section 2.5). An assessment of gridded estimate–observation differences partitioned into the four climatological seasons [Tables 3-4 (all datasets); Figs. S5-S6, ($PRISM_{D2}$ only)] reveals no consistent seasonal cycle in MD values. However, a tendency for largest (smallest) RMSD in winter (summer) is evident, consistent with larger (smaller) day-to-day temperature variability associated with the equatorward (poleward) relative position of the jet stream.

To provide some context for the magnitude of MD in Tables 3-4, the full-year MD values in Table 3 (rows labeled "ALL") were added to observed TMAXd and TMINd at East Lansing (E14 in Fig. 1, Table 1) during the 1 Mar – 30 Sep 2010 growing season; growing degree days (GDD) base 10 °C were then computed with and without the MD added (Fig. S7). At the end of the growing season, cumulative GDD computed from daily temperatures with the gridded estimate–observation MD differed from the unadulterated cumulative GDD by approximately 20-45 °C, roughly equivalent to 1-2.5 calendar days with typical mid-summer GDD values. These cumulative GDD differences can lead to phenological dates differing by one or more days between datasets. Thus, gridded estimate–observation differences on the order of a few tenths of a degree can have a considerable impact when aggregated over weeks to months. More broadly, such differences may be relevant to thermal development applications that use GDD-based biofixes (e.g., codling moth models, due to aggregation of differences), and applications that consider temperature extremes (e.g., temperature-based insect mor-

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

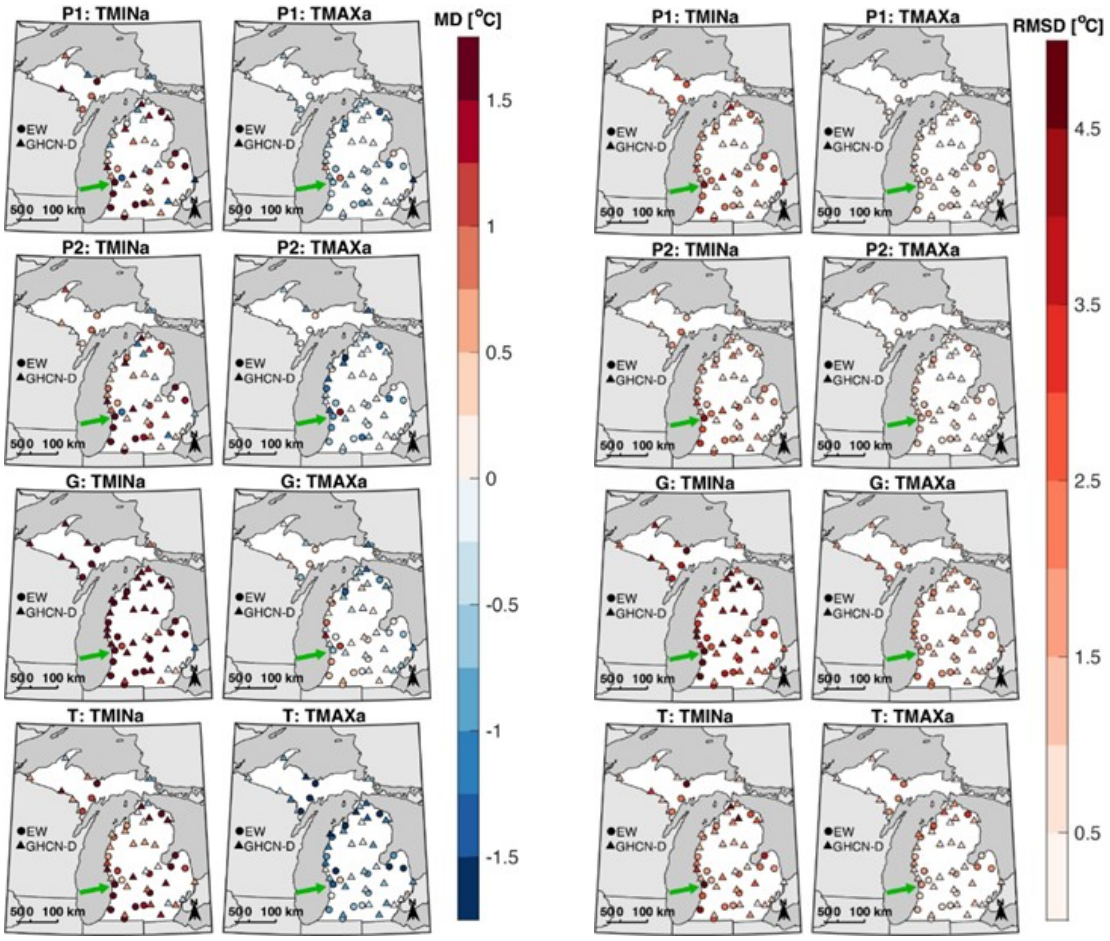

FIGURE 4 (LEFT) AND FIGURE 5 (RIGHT): EW (circle) and GHCN-D (triangle) station symbols, color-coded by annual extreme temperature mean difference (MD) in Figure 4 and annual extreme temperature root-mean-square difference (RMSD) in Figure 5, for period 2004-2016: columns: TMINa and TMAXa; rows: $PRISM_{D1}$ (P1), $PRISM_{D2}$ (P2), gridMET (G), and TopoWx (T). The minimum number of years (N) among both observational networks is 8 for TMINa and 9 for TMAXa; N at individual stations is not shown. Green arrow points to West Olive EW station site (E20 in Fig. 1, Table 1); next circle to northeast is the Sparta EW station site (E18 in Fig. 1, Table 1).

tality models, due to underestimation of spatiotemporal variability).

Before proceeding, it is worthwhile to pause the analysis of mean differences and comment briefly on individual-day differences between the gridded estimates and observations, and between the gridded datasets themselves. Extreme absolute differences are as small as 6.7 °C (TMAXd, $PRISM_{D2}$ vs. $PRISM_{D1}$) and as large as 30.2 °C [TMAXd, $PRISM_{D1}$ vs. GHCN-D (1981-2016)] (not shown). Such large differences occur infrequently: differences exceeding 10 °C (15 °C) occur on less than 5% (1%) of days across the study period of record. Possible reasons for the large differences include unresolved microclimatic variability, poorly resolved frontal zones, and unrectified day definition differences (Section 2.5).

Acknowledging the application of gridded climate datasets to studies of temperature extremes (e.g., Singh et al. 2016; Kiefer et al. 2022; Wang et al. 2021), an assessment of gridded estimate–observation differences at the tails of the climatological distribution is in order. Thus, TMINa and TMAXa are examined via maps of gridded estimate–observation differences at all EW and GHCN-D stations (MD: Fig. 4; RMSD: Fig. 5; N: not shown) and corresponding network-median statistics (Tables 5-6). For corresponding figures and tables covering the period 1981-2016 (GHCN-D only), see Figs. S8-S9 and Table S7. Beginning with $PRISM_{D1}$ and $PRISM_{D2}$ (top two rows in Figs. 4-5), two aspects of gridded estimate–observation differences are prominent. First, PRISM tends to overestimate (underestimate) TMINa (TMAXa), yielding a narrower annual extreme temperature range than what is observed. Second, there appears to be a tendency for estimate –observation differences to be largest at lakeshore stations. For TMINa, this is most prominent in southwest and northeast Lower Michigan, where MD values exceeding 1.5 °C and RMSD values exceeding 2.5 °C are common. The

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

TABLE 5: Summary statistics for 2004-2016 estimated TMINa and TMAXa [°C], at EW station sites: 20-station median number of years (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P1 (PRISMD1), P2 (PRISMD2), G (gridMET), and T (TopoWx).

	TMINa				TMAXa			
	P1	P2	G	T	P1	P2	G	T
N	12	12	12	12	13	13	13	13
MD	1.33	1.12	2.69	1.68	-0.44	-0.50	0.21	-1.01
RMSD	2.07	1.78	3.72	2.31	0.91	1.03	1.76	1.46

TABLE 6: Summary statistics for 2004-2016 estimated TMINa and TMAXa [°C], at GHCN-D station sites: 34-station median number of years (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P1 ($PRISM_{D1}$), P2 ($PRISM_{D2}$), G (gridMET), and T (TopoWx).

	TMINa				TMAXa			
	P1	P2	G	T	P1	P2	G	T
N	13	13	13	13	13	13	13	13
MD	0.30	0.31	1.98	0.67	-0.35	-0.36	-0.18	-0.72
RMSD	1.31	1.01	3.05	1.38	0.61	0.60	1.38	0.95

signal for a lakeshore proximity dependence of estimate–observation differences is generally weaker for TMAXa than TMINa, with the exception of $PRISM_{D2}$ MD (Fig. 4).

A comparison of network-median estimate–observation differences at the EW sites (Table 5) reveals that TMINa MD and RMSD are smaller for $PRISM_{D2}$, with the opposite true for TMAXa; corresponding statistics at the GHCN-D sites (Table 6) generally show smaller differences between the two PRISM versions. However, the small sample size (8-13 years; not shown) limits our ability to test the significance of the differences between the gridded datasets at the annual extreme timescale. One exception to the statement regarding smaller gridded estimate–observation differences at the GHCN-D sites is Pellston Regional Airport (G25 in Fig. 1, Table 2), where TMINa MD is 3.77 °C and 1.67 °C for $PRISM_{D1}$ and $PRISM_{D2}$, respectively. Pellston is situated in northern Lower Michigan where the local physiography produces a microclimate prone to cold air pooling and strong nocturnal temperature inversions. Changes to the PRISM system to better render strong temperature inversions (Daly 2022) appears to have had a beneficial impact on estimates at locations that, like Pellston, are sited in areas prone to this nocturnal phenomenon. The smaller magnitude of TMINa estimate–observation differences was the primary reason that $PRISM_{D2}$ was selected by Kiefer et al. (2022) for their climatological study of TMINa in the US Great Lakes region. Expanding the analysis to all four gridded datasets (Tables 5-6), the largest MD and RMSD are generally found with gridMET, with the exception of TMAXa MD. The positive sign of gridMET TMAXa MD indicates a counterintuitive tendency for the gridded dataset to overestimate TMAXa at the EW station sites; a comparison of MD at all timescales (Tables 3-6) points to a systematic overestimation of temperature by gridMET. A possible source for this behavior is the propagation of NLDAS-2 (and, in turn, NARR) temperature error (Behnke et al. 2016; Walton and Hall 2018) into gridMET (Section 2.2.2). Finally, estimate–observation differences are similar for PRISM and TopoWx, with the smallest overall differences found with PRISM.

We next examine scatterplots of estimated and observed daily temperatures at two EW sites chosen due to differences in station siting and contrasting estimate–observation differences (Figs. 6-7): Sparta and West Olive (E18 and E20 in Fig. 1, Table 1). West Olive is sited approximately 12 km inland of Lake Michigan at 194 m above mean sea level (AMSL), and as shown in Figs. 2 and 4, exhibits $PRISM_{D2}$ MD values of 0.84 °C (TMINd), -0.47 °C (TMAXd), 3.58 °C (TMINa), and -0.89 °C (TMAXa). Sparta is sited approximately 42 km inland of Lake Michigan, at 267 m AMSL, and exhibits corresponding MD values of -0.85 °C (TMINd), 0.72 °C (TMAXd), -1.23 °C (TMINa), and 1.29 °C (TMAXa). It is notable that the two sites, located approximately 31 km apart, exhibit opposite MD sign across all timescales. Careful examination of Figs. 2 and 4 reveals that Sparta, not West Olive, is the outlier among neighboring station sites. Notably, Sparta is located on western Lower Michigan's "fruit ridge", an approximately 410 km2 tree fruit production area of relatively higher topography with surface elevations exceeding 240 m AMSL and relatively fine-textured clay loam soils. Regular monitoring, maintenance, and QA/QC of station observations by EW staff, and the diligent monitoring of station observations by growers as part of their farming practices (Andresen et al. 2012), suggests that the outlier status of Sparta is likely a reflection of unresolved microclimatic variability, rather than the result of instrumentation error. For further information on unresolved microclimatic variability across Michigan, in the context of PRISM estimates of cold-season minimum temperatures, see Kiefer et al. (2023).

Careful examination of the scatterplots at Sparta and West Olive (Figs. 6-7) reveals characteristics of estimate–observation differences both general to gridded datasets and specific to the individual datasets and station sites. Generally speaking, the scatterplots exhibit larger clouds of points for TMINd than for TMAXd, indicative of the limited ability of gridded datasets to resolve spatiotemporal microclimatic gradients in the nocturnal boundary layer, especially those associated with surface inversions. Turbulent mixing in the daytime boundary layer yields weaker microclimatic gradients and smaller estimate–

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

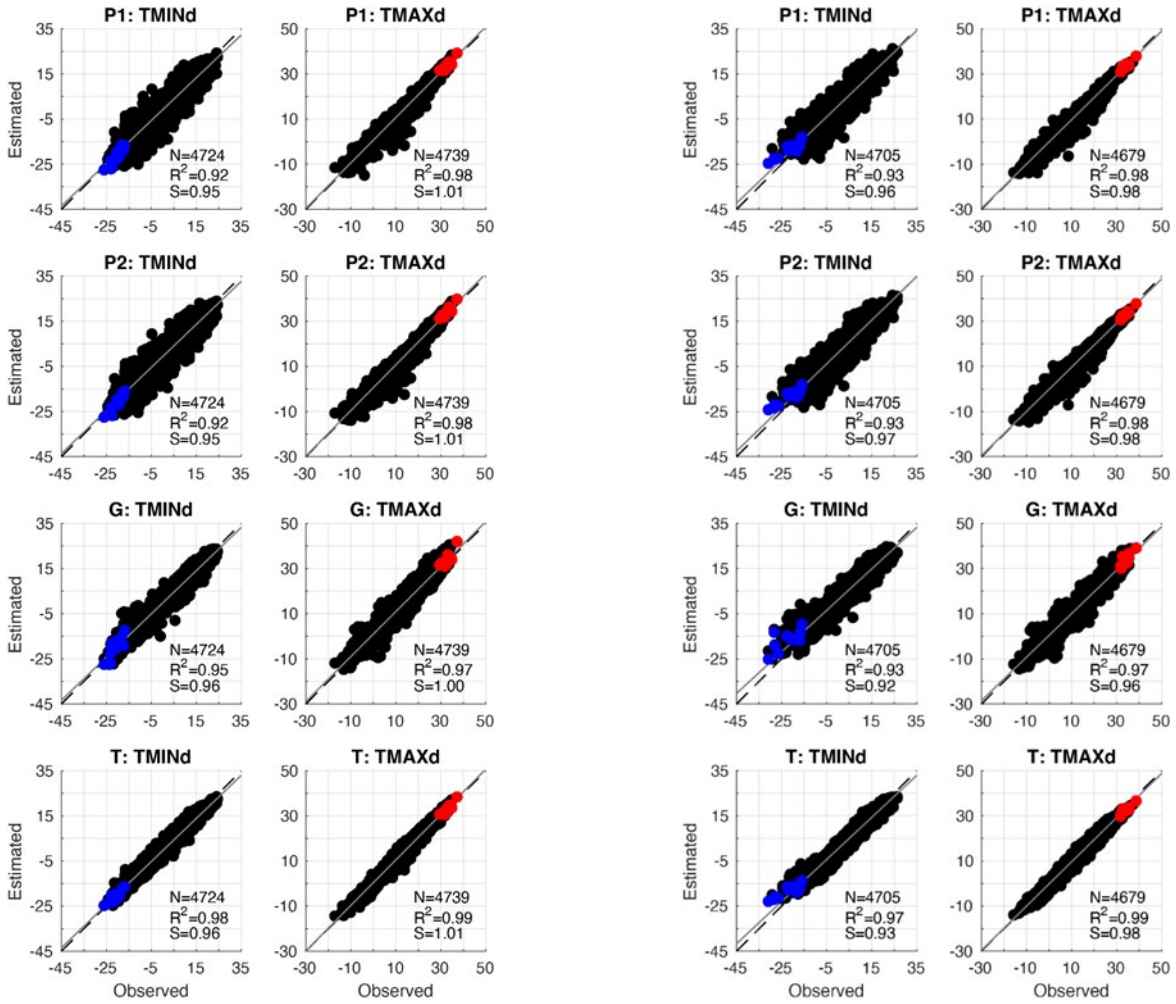

FIGURE 6 (LEFT): Scatterplots with Sparta EW station (E18 in Fig. 1, Table 1) and FIGURE 7 (RIGHT) with West Olive EW station (E20 in Fig. 1, Table 1) observed daily and annual extreme temperatures vs. four gridded estimates [°C]: columns: TMINd and TMAXd (w/ PRISM lag); rows: $PRISM_{D1}$ (P1), $PRISM_{D2}$ (P2), gridMET (G), and TopoWx (T). TMINa and TMAXa are indicated with blue and red fill colors, respectively, and dashed (solid) line indicates 1:1 (least-squares linear regression).

weaker microclimatic gradients and smaller estimate–observation differences for TMAXd. The larger spatial variability of TMINd was noted by Daly et al. (2008) in the context of their assessment of PRISM climatological normals: even in areas of gently varying terrain in the central and eastern US, spatial variability was found to be greater for monthly-mean TMINd than for monthly-mean TMAXd. Such a finding is broadly consistent with studies of spatiotemporal variability of atmospheric variables within the nocturnal boundary layer, wherein the magnitudes of inhomogeneity and non-stationarity increase with increasing static stability (e.g., Mahrt et al. 2013; Mahrt 2022). A second general point revealed by the scatterplot analysis is that a gridded climate dataset can perform comparatively well when averaged over all days (e.g., TMINd) but exhibit large estimate–observation differences at the distribution tails (e.g., TMINa). An example of this behavior is TopoWx: the overall tight fit of TopoWx estimated TMINd across a wide range of temperatures, at both Sparta and West Olive (Figs. 6-7), does not extend to the extreme low end of the temperature distribution at West Olive (Fig. 7). TopoWx exhibits a positive MD of 4.0 °C, considerably smaller than that of gridMET (5.78 °C) but larger than that of $PRISM_{D1}$ and $PRISM_{D2}$ (3.49 and 3.58 °C, respectively).

Returning to the opposing nature of MD at the two sites (Figs. 6-7), it is worthwhile to consider the counterintuitive nature of MD at Sparta. The process of interpolating point observations to a regular grid by weighting observations and background values is expected to generally yield higher (lower) minimum (maximum) temperatures in the gridded dataset, compared to observations (Daly et al. 2008).

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

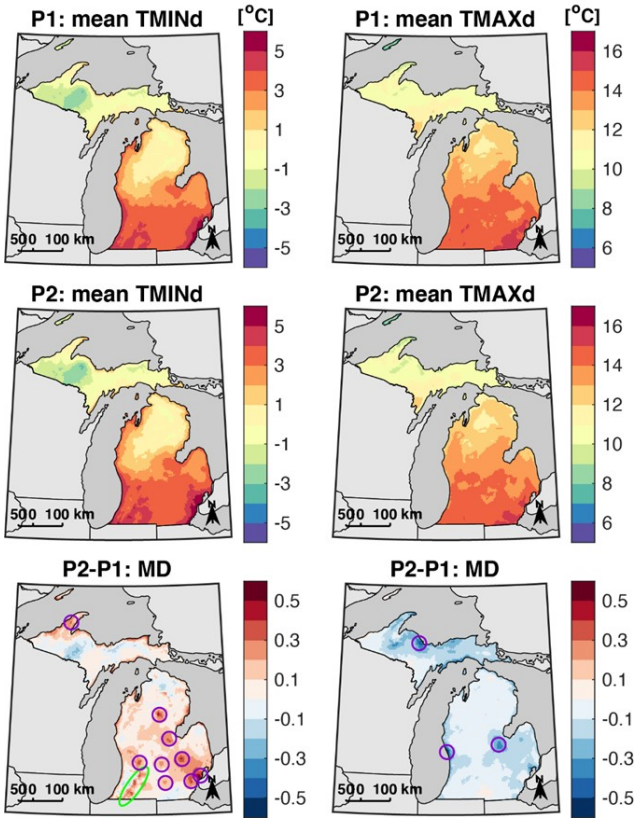

FIGURE 8: Maps of contoured $PRISM_{D1}$ and $PRISM_{D2}$ estimated daily temperatures: columns: TMINd and TMAXd; rows: $PRISM_{D1}$ (P1) 1981-2018 mean, $PRISM_{D2}$ (P2) 1981-2018 mean, and $PRISM_{D2–D1}$ (P2-P1) mean difference (MD). Green oval and purple circles denote features described in text.

This 'smoothing' effect of the interpolation process does in fact yield gridded values that generally overestimate (underestimate) observed TMINd and TMINa (TMAXd and TMAXa) (Figs. 2 and 4; Tables 3-6). However, the gridded estimate–observation differences at Sparta are the exact opposite: estimates of minimum (maximum) temperatures are mostly too low (high). Partitioning MD by season (Fig. S5) reveals that the underestimation (overestimation) of TMINd (TMAXd) at Sparta is greatest in spring (summer). A likely reason for the singular performance of the gridded estimates at Sparta is the inability of the gridded datasets to account for the complex physiographic gradients of soil, vegetation, and surface elevation in the vicinity of Sparta, and the resultingly complex spatial patterns of surface energy balance, nocturnal cold-air flows, and daytime boundary layer depths (not shown).

3.2 Spatial Analysis

Having characterized the gridded estimate–observation differences at point locations across Michigan, we now characterize spatial patterns of estimate–estimate differences. Given the differences in grids between PRISM, gridMET, and TopoWx, and the aforementioned utilization of PRISM in EW applications, this analysis focuses exclusively on the two PRISM versions. The preceding point analysis reveals that although gridded estimate–observation differences between the two PRISM versions are generally smaller than differences between PRISM and the other gridded datasets examined in this study, there are consistent differences between the two PRISM versions (Figs. 2-5; Tables 3-6). Overall, $PRISM_{D2}$ exhibits smaller (larger) MD than $PRISM_{D1}$ for TMINd and TMINa (TMAXd and TMAXa). To characterize spatial patterns of gridded estimate–estimate differences, Figure 8 depicts contoured 1981-2018 mean TMINd and TMAXd for the two PRISM versions, along with MD ($PRISM_{D2–D1}$). It is important to note that the complex nature of the PRISM climate analysis system (Section 2.2.1) makes attribution of features in the contoured plots difficult. Although we speculate as to the source of the spatial features, a comprehensive investigation is beyond the scope of this study.

Overall, the maps of contoured mean TMINd and TMAXd exhibit similar patterns for $PRISM_{D1}$ and $PRISM_{D2}$ (top two rows in Fig. 8), with lowest temperatures in western Upper Michigan and northern Lower Michigan, and highest temperatures in southwest and southeast Lower Michigan, and along the Lake Michigan shoreline. Despite similar temperature patterns, the maps of contoured TMINd and TMAXd MD (bottom row in Fig. 8) depict localized minima and maxima in the range of ±0.5 °C. The MD map for TMINd primarily exhibits positive values and contains three distinct difference patterns: lakeshore, southwest Lower Michigan, and airport site. First, positive values are found within one or two PRISM grid points of the shores of Lakes Michigan, Huron and Superior. The near-ubiquity of positive MD along the lakeshores suggests that this pattern is not related to the interpolation of individual station data; the introduction of gridded upper-air temperature data in $PRISM_{D2}$ is one possible source. Second, a string of MD maxima is present in southwest Lower Michigan (green oval in Fig. 8, bottom-left panel). It is noteworthy that the string of MD maxima appears to be collocated with a gap in weather station coverage (not shown). Recall from Section 2.2.1 that in the absence of surface station data, PRISM daily grids are weighted toward PRISM climatological normals, which are in turn a function of surface elevation. A careful examination of Fig. 1 reveals slightly higher surface elevation in the vicinity of the southwest Lower Michigan pattern than in surrounding areas (cf. Figs. 1 and 8), suggesting a tendency in PRISM for weaker nocturnal temperature inversions in southwest Lower Michigan associated with cold-air drainage to lower elevations. Third, the TMINd MD

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

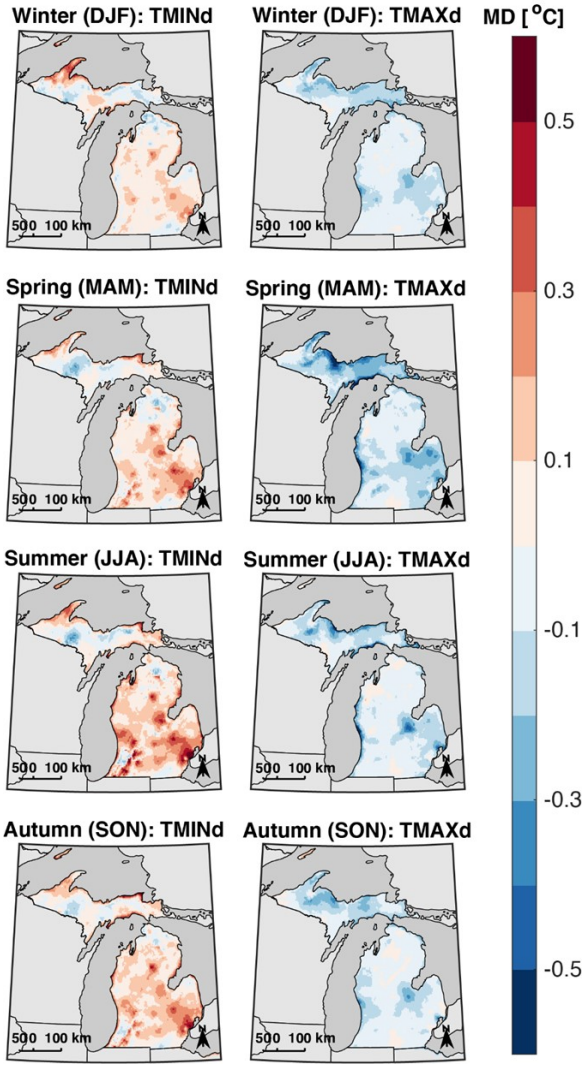

FIGURE 9: Maps of contoured $PRISM_{D1}$ and $PRISM_{D2}$ estimated daily temperatures: columns: TMINd and TMAXd; rows: $PRISM_{D2-D1}$ mean difference (MD), partitioned by season: winter (December-February), spring (March-May), summer (June-August), and autumn (September-November).

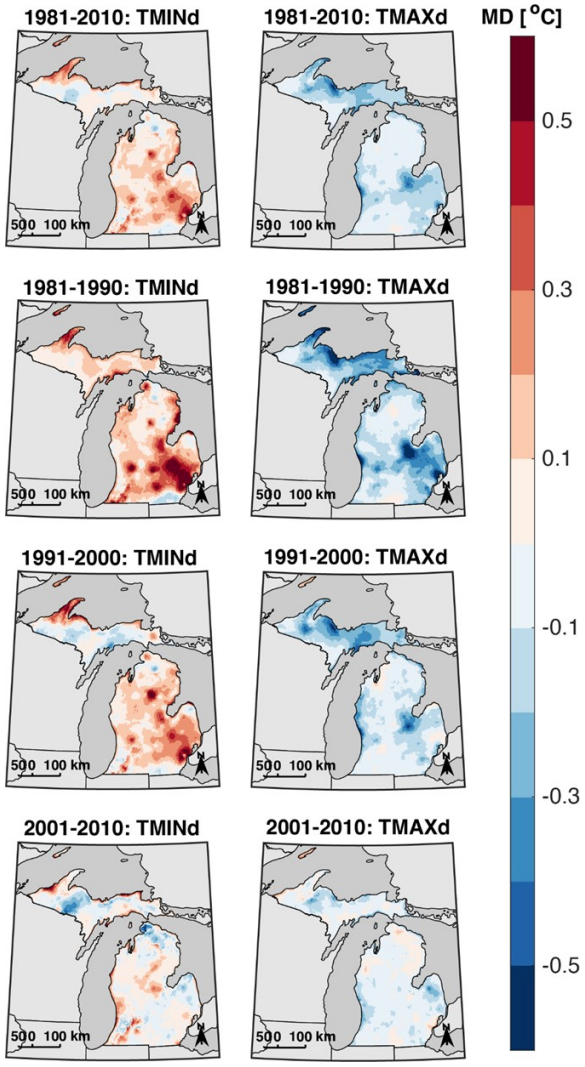

FIGURE 10: Maps of contoured $PRISM_{D1}$ and $PRISM_{D2}$ estimated daily temperatures: columns: TMINd and TMAXd; rows: $PRISM_{D2-D1}$ mean difference (MD), partitioned by decade: 1981-2010, 1981-1990, 1991-2000, and 2001-2010.

plot shows maxima collocated with numerous airport station sites (purple circles in Fig. 8, bottom-left panel). As will be discussed shortly, this pattern is related to differences in airport data assimilation between $PRISM_{D1}$ and $PRISM_{D2}$.

Proceeding to TMAXd (Fig. 8, right panels), the $PRISM_{D2-D1}$ MD map primarily exhibits negative MD and exhibits different patterns than the corresponding TMINd plot. There is a tendency for TMAXd MD to be most negative along the central shores of Lakes Michigan and Superior, and in an area southwest of Saginaw Bay (purple circles in Fig. 8, bottom-right panel). The approximate colocation of the MD minima and select GHCN-D station sites [clockwise from upper-left: Marquette (G22), Saginaw (G28), and Muskegon (G24); cf. Figs. 1 and 8, Table 2] suggests that differences in individual station assimilation between $PRISM_{D1}$ and $PRISM_{D2}$ may be the source of the phenomenon. Taken as a whole, the MD panels in Fig. 8 indicate a tendency for $PRISM_{D2}$ to exhibit a smaller diurnal range than $PRISM_{D1}$. Unlike the localized MD maxima and minima, the broad pattern of reduced diurnal range suggests a source not related to data assimilation at individual station sites. One possibility is the aforementioned inclusion of gridded upper-air temperature data in $PRISM_{D2}$, but not $PRISM_{D1}$ (further examination is left to future work).

Additionally, $PRISM_{D2-D1}$ differences exhibit cycles of seasonal variability, as shown in Fig. 9. TMINd MD is

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

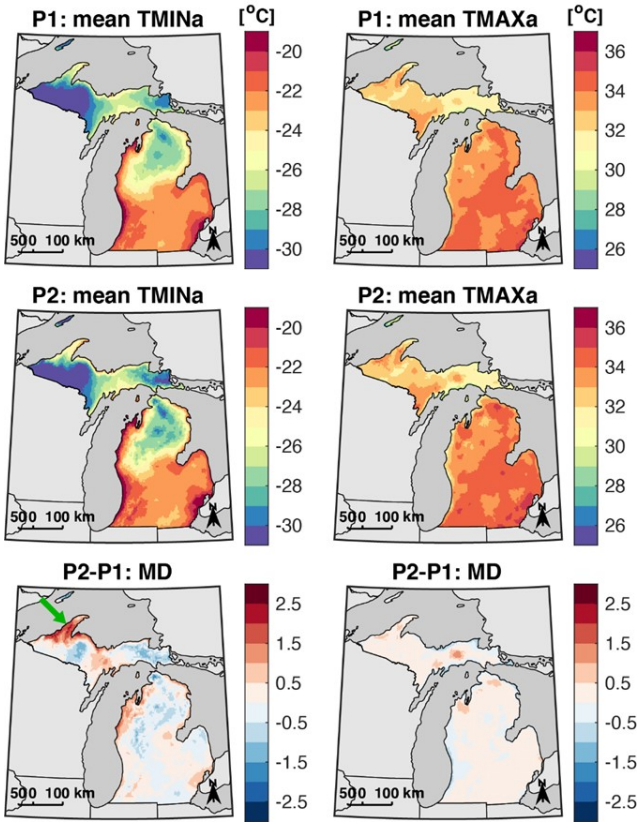

FIGURE 11: Maps of contoured $PRISM_{D1}$ and $PRISM_{D2}$ estimated annual extreme temperatures: columns: TMINa and TMAXa; rows: $PRISM_{D1}$ (P1) 1981-2018 mean, $PRISM_{D2}$ (P2) 1981-2018 mean, and $PRISM_{D2–D1}$ (P2-P1) mean difference (MD). Green arrow in bottom-left panel denotes the Keweenaw peninsula in western Upper Michigan.

maximized in summer (June-August), whereas TMAXd MD is maximized in spring (March-May). The prominence of the TMINd MD patterns seen in Fig. 8 during summer, a season with overall weaker wind speeds, suggests that station interpolation differences between the PRISM versions are most pronounced during light wind conditions. The $PRISM_{D2–D1}$ differences also exhibit cycles of decadal variability, as shown in Fig. 10. The local TMINd MD maxima at the airport station sites appear to be associated exclusively with the pre-ASOS period: features show up prominently during 1981-1990 (pre-ASOS installation) but are completely absent during 2001-2010 (post-ASOS installation); 1991-2000 is intermediate. Based on a review of station inventory files bundled with each PRISM file, the observations from airport sites (both human and automated) were not assimilated into $PRISM_{D1}$ until 2 Jul 1996, but they were assimilated into $PRISM_{D2}$ starting on 1 Jan 1981. This critical difference in input data sources between $PRISM_{D1}$ and $PRISM_{D2}$ is important to keep in mind when comparing PRISM daily variables (and quantities derived from daily variables) between versions.

Finally, maps of contoured $PRISM_{D1}$ and $PRISM_{D2}$ mean TMINa and TMAXa (Fig. 11) exhibit broadly similar patterns to that of TMINd and TMAXd, albeit with stronger minimum temperature gradients at the annual extreme timescale (cf. top two rows in Figs. 8 and 11). Mean TMINa values vary from below -30 °C in western Upper Michigan, away from Lake Superior, to above -20 °C in parts of Lower Michigan along the Lake Michigan shore and in the extreme southeast portion of the state. The MD patterns are noticeably different for TMINa and TMAXa than for TMINd and TMAXd (cf. bottom row in Figs. 8 and 11). Specifically, areas of positive MD are found in northwest Lower Michigan and parts of western Upper Michigan, in particular the Keweenaw peninsula (see green arrow in bottom-left panel). Areas of negative TMINa MD are found across interior Lower Michigan and parts of southwest and northeast Upper Michigan. Regarding the large positive MD in the Keweenaw peninsula, a likely source is the exclusion in $PRISM_{D2}$ of at least some portion of extreme minimum temperature observations at Hancock Houghton County airport (G13 in Fig. 1, Table 2; see the larger gridded estimate–observation differences for $PRISM_{D2}$ compared to $PRISM_{D1}$ in Fig. 4, left panels). Regarding TMAXa, MD values are of overall smaller magnitude than for TMINa, and with limited exceptions are of positive sign. The overall positive MD for TMAXa is of opposite sign to that of TMAXd (cf. Figs. 8 and 11, right panels), indicating that the changes to methodology instituted in $PRISM_{D2}$ yield lower daily maximum temperatures but slightly higher annual extreme maximum temperatures, compared to $PRISM_{D1}$. Further examination of the source of $PRISM_{D2–D1}$ differences is left to future work.

4. Summary and Conclusion

In this study, three gridded climate datasets were examined for their ability to accurately reproduce daily and annual extreme temperature climatologies at climate stations across Michigan: PRISM, gridMET, and TopoWx. Michigan was chosen for this study due to the limited prior assessment of gridded climate datasets in a non-mountainous region with otherwise complex physiography, the use of PRISM in Michigan's Enviroweather program, and the availability in Michigan of a network of weather stations that is not assimilated into any of the gridded datasets evaluated in this study. Although this study focused exclusively on Michigan, we expect the results to be applicable to other areas of the US (e.g., Maine coastline) and other regions of the world (e.g., Black Sea in Eurasia) with similar

physiography. The study took place in two phases; first, gridded estimates at the nearest grid point to each station site were compared with corresponding observations; second, maps of contoured PRISM-version differences were presented.

Overall, TopoWx was found to exhibit the smallest deviation from observed daily temperatures, $PRISM_{D2}$ was found to exhibit the smallest deviations from observed annual extreme minimum temperatures, and gridMET was found to exhibit the largest gridded estimate–observation differences across all timescales. TopoWx's superior performance at the daily timescale was attributed to the integration of DEM and satellite land-surface temperature data in the station interpolation process at the daily timescale, and implementation of a temporal homogenization algorithm. The overall poor performance of gridMET was attributed to the integration of NLDAS-2 temperature estimates, with propagation of error from NLDAS-2 to gridMET yielding broad temperature overestimation. For $PRISM_{D2}$, modifications to the PRISM system to accommodate strong temperature inversions appears to have had a beneficial impact on annual extreme minimum temperature estimates across Michigan. An in-depth assessment of gridded estimate–observation differences at two EW station sites pointed to the inability of the gridded datasets to fully account for the complex physiographic gradients of soil, vegetation, and surface elevation. Maps of contoured $PRISM_{D2–D1}$ temperature differences revealed distinct spatiotemporal patterns. Generally speaking, differences were maximized along the lakeshores, in areas with low station density, and at airport station sites.

Of the previous gridded data assessment studies cited in Section 1, only Behnke et al. (2016) and Walton and Hall (2018) considered all three temperature datasets examined herein. The finding of overall poorest performance for gridMET agrees with Behnke et al. (2016), and the finding that integration of DEM and satellite land-surface temperature data in TopoWx yields statistically-significant differences in dataset performance compared to PRISM and gridMET is broadly supported by Walton and Hall (2018). Blankenau et al. (2020) did not evaluate PRISM or TopoWx but did find that gridMET performed best among the datasets they evaluated in parts of the western and southeastern US, but not in the Great Lakes region or northeastern US in general.

This study has highlighted several challenges involved in evaluating gridded climate datasets. For example, day definition differences complicate the calculation of evaluation statistics, as some portion of the gridded estimate–observation differences may be due to observed and estimated values occurring at different times of day or on different days. Additionally, the assimilation of station observations into gridded datasets requires that caution be applied when using the very observations assimilated into the gridded dataset to evaluate it (e.g., PRISM and GHCN-D). However, even in a situation wherein the gridded and observational datasets use identical day definitions and are independent of each other (e.g., TopoWx and EW), one has to contend with differences stemming from sub-grid variability. Furthermore, differences in instrumentation between networks can complicate the assessment of gridded datasets using observations from multiple networks. Despite the challenges involved in assessing gridded climate datasets, studies such as this are critical sources of information for users. Knowledge of the strengths and weaknesses of gridded climate datasets can help inform users as to which datasets and which variables are likely to be most accurate for a given application and on a given timescale. Finally, users must also consider the production status of each gridded climate dataset: for example, TopoWx ended production in 2016, making it a suitable dataset for historical studies but not those that require data from more recent years.

Although this study makes an important contribution to our knowledge of gridded climate dataset strengths and weaknesses, much work remains. Caution is recommended when extrapolating the results of this study, since we examined only a small subset of available gridded climate datasets for a single variable within a single US state. Further investigation is warranted to assess whether the findings of this study (e.g., propagation of NLDAS-2 temperature error to gridMET) are applicable to other datasets. Future efforts will expand this assessment to other datasets, variables, and regions, will perform detailed process analysis to identify the sources of gridded estimate–observation and estimate–estimate differences, and will also explore in greater detail how gridded dataset errors may impact agricultural and ecological applications. This and future studies of gridded estimate–observation differences are likely to be of value to gridded climate dataset users across a range of disciplines.

Acknowledgments

Funding for this project was provided by the Michigan Department of Natural Resources (MDNR) Michigan Invasive Species Grant Program (MISGP) [grant IS16-3002]; the Great Lakes Integrated Sciences and Assessments Program (GLISA), a collaboration of the University of Michigan and Michigan State University (MSU) funded by the National Oceanic and Atmospheric Administration (NOAA) Oceanic and Atmospheric Research (OAR) office [NOAA-OAR-CPO grant NA15OAR4310148]; and Michigan State University AgBioResearch [Project MICL02548]. The color maps used in all figures except

Figs. 6, 7, and S7 were developed by Cynthia Brewer at the Pennsylvania State University (https://colorbrewer2.org/). Finally, comments and suggestions from two anonymous reviewers were helpful in revising the manuscript and are greatly appreciated.

Data Availability Statement

Data analyzed in this study are openly available as follows: $PRISM_{D1}$ and $PRISM_{D2}$ data were downloaded from https://ftp.prism.oregonstate.edu/data_archive/var/daily_D1_201910/yyyy (accessed 15 May 2019) and https://ftp.prism.oregonstate.edu/daily/var/yyyy (accessed 19 Oct 2019), respectively [note: before accessing link, replace "var" with the variable name (tmin or tmax) and "yyyy" with the four-digit year]; gridMET data were downloaded from https://www.northwestknowledge.net/metdata/data (accessed 31 Mar 2021); TopoWx data were downloaded from https://cida.usgs.gov/thredds/catalog.html?dataset=cida.usgs.gov/topowx (accessed 31 Mar 2021); GHCN-D data were downloaded from https://www.ncei.noaa.gov/data/global-historical-climatology-network-daily/access/ (accessed 9 Apr 2020); and EW data were downloaded from https://mawn.geo.msu.edu/station.asp?id=stn&rt=24 (accessed 22 Mar 2021) [note: before accessing link, replace "stn" with the three-letter EW station identifier; Table 1].

References

Abatzoglou, J.T., 2013: Development of gridded surface meteorological data for ecological applications and modeling. *Int. J. Climatol*., **33**, 121–131, https://doi.org/10.1002/joc.3413.

Anderson, B.J., 2005: The historical development of the tension zone concept in the Great Lakes region of North America. *The Michigan Botanist,* **44**, 127-138, http://hdl.handle.net/2027/spo.0497763.0044.304 [accessed 11 Aug 2023].

Andresen, J.A., 2012: Historical climate trends in Michigan and the Great Lakes region. *Climate Change in the Great Lakes Region*, T. Dietz and D. Bidwell, Eds., Michigan State University Press., 17-34.

Andresen, J., L. Olsen, T. Aichele, B. Bishop, J. Brown, J. Landis, S. Marquie, and A. Pollyea, 2012: Enviro-weather: a weather-based pest and crop management information system for Michigan. *Proc. of the 7th International Integrated Pest Management Symposium*, Memphis, TN, University of Illinois, Center for Innovation in Teaching & Learning, 32, https://ipmsymposium.org/2012/IPM_12_Proceedings_final.pdf [accessed 11 Aug 2023].

Andresen, J.A., and J.A. Winkler, 2009: Weather and climate. *Michigan Geography and Geology*, R.J. Schaetzl, J.T. Darden, and D. Brandt, Eds., Pearson Custom Publishing, 288–314.

Behnke, R., S. Vavrus, A. Allstadt, T. Albright, W.E. Thogmartin, and V.C. Radeloff, 2016: Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecol. Appl.*, **26**, 1338-1351, https://doi.org/10.1002/15-1061.

Blankenau, P.A., A. Kilic, and R. Allen, 2020: An evaluation of gridded weather data sets for the purpose of estimating reference evapotranspiration in the United States. *Agric. Water Manag.*, **242**, 106376, https://doi.org/10.1016/j.agwat.2020.106376.

Bosilovich, M.G., 2006: A comparison of MODIS land surface temperature with in situ observations. *Geophys. Res. Lett.,* **33**, L20112, https://doi.org/10.1029/2006GL027519.

Changnon, S.A., Jr., and D.M.A Jones, 1972: Review of the influences of the Great Lakes on weather. *Water Resour. Res*., **8**, 360–371, https://doi.org/10.1029/WR008i002p00360.

Curtis, J.T., 1959: *The Vegetation of Wisconsin: An Ordination of Plant Communities*. University of Wisconsin Press, 704 pp., https://muse.jhu.edu/book/16741 [accessed 11 Aug 2023].

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, Cambridge, 471 pp.

Daly, C., 2022: Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States. PRISM Doc., 28 pp, https://prism.oregonstate.edu/documents/PRISM_datasets.pdf [accessed 11 Aug 2023].

Daly, C., R.P. Neilson, and D.L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140-158, https://doi.org/10.1175/1520-0450(1994)033%3C0140:ASTMFM%3E2.0.CO;2.

Daly, C., G. Taylor, and W. Gibson, 1997: The PRISM approach to mapping precipitation and temperature. Proceedings on 10th Conference on Applied Climatology, Reno, NV, USA. American Meteorological Society, 10-12, https://prism.oregonstate.edu/documents/pubs/1997appclim_PRISMapproach_daly.pdf [accessed 11 Aug 2023]

Daly, C., M. Halbleib, J.I. Smith, W.P. Gibson, M.K. Doggett, G.H. Taylor, J. Curtis, and P.P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol*., **28**, 2031-2064. https://doi.org/10.1002/joc.1688.

Daly, C., J.I. Smith, and K.V. Olson, 2015: Mapping atmospheric moisture climatologies across the contermi-

nous United States. *PLoS ONE*, **10**, e0141140, https://doi.org/10.1371/journal.pone.0141140.

Daly, C., M.K. Doggett, J.I. Smith, K.V. Olson, M.D. Halbleib, Z. Dimcovic, D. Keon, R.A. Loiselle, B. Steinberg, A.D. Ryan, C.M. Pancake, and E.M. Kaspar, 2021: Challenges in observation-based mapping of daily precipitation across the conterminous United States. *J. Atmos. Oceanic Technol.,* **38**, 1979-1992, https://doi.org/10.1175/JTECH-D-21-0054.1.

Durre I., M. J. B.E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *J. Applied Meteor. and Climatol*., **49**, 1615-1633, https://doi.org/10.1175/2010JAMC2375.1.

Durre, I., A. Arguez, C.J. Schreck III, M.F. Squires, and R.S. Vose, 2022: Daily high-resolution temperature and precipitation fields for the Contiguous United States from 1951 to Present. *J. Atmos. Oceanic Technol.*, **39**, 1837–1855, https://doi.org/10.1175/JTECH-D-22-0024.1

Duveneck, M.J., R.M. Scheller, and M.A. White, 2014: Effects of alternative forest management on biomass and species diversity in the face of climate change in the northern Great Lakes region (USA). *Can. J. For. Res.*, **44**, 700-710, https://doi.org/10.1139/cjfr-2013-0391.

Estes Jr., M.G., T. Insaf, M.Z. Al-Hamdan, T. Adeyeye, and W. Crosson, 2022: Validation of North American land data assimilation system Phase 2 (NLDAS-2) air temperature forcing and downscaled data with New York State station observations. *Remote Sens. Appl.: Soc. Environ.*, **25**, 100670, https://doi.org/10.1016/j.rsase.2021.100670.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. Q. *J. R. Meteorol. Soc.*, **146**, 1999-2049, https://doi.org/10.1002/qj.3803.

Janjić, T, N. Bormann, M. Bocquet, J.A. Carton, S.E. Cohn, S.L. Dance, S.N. Losa, N.K. Nichols, R. Potthast, J.A. Waller, and P. Weston, 2018: On the representation error in data assimilation. *Q. J. R. Meteorol. Soc*., **144**, 1257-1278. https://doi.org/10.1002/qj.3130.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, https://doi.org/10.1175/1520-0477(1996)077%3C0437:TNYRP%3E2.0.CO;2.

Kiefer, M. T., J. A. Andresen, D. Doubler, and A. Pollyea, 2019: Development of a gridded reference evapotranspiration dataset for the Great Lakes region. *J. Hydrol.: Reg. Stud.*, **24**, 100606, https://doi.org/10.1016/j.ejrh.2019.100606.

Kiefer, M.T., J.A. Andresen, D.G. McCullough, W.J. Baule, and M. Notaro, 2022: Extreme minimum temperatures in the Great Lakes region of the United States: A climatology with implications for insect mortality. *Intl. J. Climatol.,* **42**, 3596-3615, https://doi.org/10.1002/joc.7434.

Kiefer, M.T., J.A. Andresen, D.G. McCullough, J.B. Wieferich, J. Keyzer, and S.A. Marquie, 2023: Microclimatic variability of cold-season minimum temperatures in Michigan, United States: A study with implications for insect mortality. *J. Appl. Meteorol. Clim.*, **62**,1187-1203, https://doi.org/10.1175/JAMC-D-23-0067.1.

Kitchen, M., and R.M. Blackall, 1992: Representativeness errors in comparisons between radar and gauge measurements of rainfall. *J. Hydrol*., **134**, 13-33, https://doi.org/10.1016/0022-1694(92)90026-R.

Liu, S., H. Su, J. Tian, and W. Wang, 2018: An analysis of spatial representativeness of air temperature monitoring stations. *Theor. Appl. Climatol.,* **132**, 857-865, https://doi.org/10.1007/s00704-017-2133-6.

Lofgren, B.M., 1997: Simulated effects of idealized Laurentian Great Lakes on regional and large-scale climate. *J. Clim.*, **10**, 2847-2858, https://doi.org/10.1175/1520-0442(1997)010<2847:SEOILG>2.0.CO;2.

Lorenc, A.C., 1986: Analysis methods for numerical weather prediction. *Q.J.R. Meteorol. Soc.*, **112**, 1177-1194, https://doi.org/10.1002/qj.49711247414.

Mahrt, L., C. Thomas, S. Richardson, N. Seaman, D. Stauffer, and M. Zeeman, 2013: Non-stationary generation of weak turbulence for very stable and weak-wind conditions. *Boundary-Layer Meteorol.*, **147**, 179-199, https://doi.org/10.1007/s10546-012-9782-x.

Mahrt, L., 2022: Horizontal variations of nocturnal temperature and turbulence over microtopography. *Boundary-Layer Meteorol.,* **184**, 401-422, https://doi.org/10.1007/s10546-022-00721-w.

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, https://doi.org/10.1175/JTECH-D-11-00103.1.

Menne, M. J., and C. N. Williams, 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700-1717, https://doi.org/10.1175/2008JCLI2263.1.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, https://doi.org/10.1175/BAMS-87-3-343.

National Research Council, 2012: The National Weather Service Modernization and Associated Restructuring: A Retrospective Assessment. *Appendix E: Automated*

Surface Observing System impact on the climate record. The National Academies Press, 99-100, https://doi.org/10.17226/13216.

Newman, A.J., M.P. Clark, R.J. Longman, and T.W. Giambelluca, 2019: Methodological intercomparisons of station-based gridded meteorological products: utility, limitations, and paths forward. *J. Hydrometeorol.*, **20**, 531-547, https://doi.org/10.1175/JHM-D-18-0114.1.

Notaro, M., K. Holman, A. Zarrin, S. Vavrus, and V. Bennington, 2013: Influence of the Laurentian Great Lakes on regional climate. *J. Climate*, **26**, 789–804, https://doi.org/10.1175/JCLI-D-12-00140.1.

Oyler, J.W., S.Z. Dobrowski, A.P. Ballantyne, A.E. Klene, and S.W. Running, 2014: Artificial amplification of warming trends across the mountains of the western United States. *Geophys. Res. Lett.*, **42**, 153-161, https://doi.org/10.1002/2014GL062803.

Oyler, J.W., A. Ballantyne, K. Jencso, M. Sweet, and S.W. Running, 2015: Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *Intl. J. Climatol.*, **35**, 2258-2279, https://doi.org/10.1002/joc.4127.

Oyler, J.W., S.Z. Dobrowski, Z.A. Holden, and S.W. Running, 2016: Remotely sensed land skin temperature as a spatial predictor of air temperature across the conterminous United States. *J. Appl. Met. Climatol.*, **55**, 1441-1457, https://doi.org/10.1175/JAMC-D-15-0276.1.

Pinson, P., and R. Hagedorn, 2012: Verification of the ECMWF ensemble forecasts of wind speed against analyses and observations. *Meteorol. Appl.*, **19**, 484-500, https://doi.org/10.1002/met.283.

Prakash, S., F. Shati, H. Norouzi, and R. Blake, 2019: Observed differences between near-surface air and skin temperatures using satellite and ground-based data. *Theor. Appl. Climatol.* **137**, 587–600, https://doi.org/10.1007/s00704-018-2623-1.

Rogers, B.M., P. Jantz, and S.J. Goetz, 2017: Vulnerability of eastern US tree species to climate change. *Glob. Change Biol.,* **23**, 3302-3320, https://doi.org/10.1111/gcb.13585.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Scott, R.W., and F.A. Huff, 1996: Impacts of the Great Lakes on regional climate conditions. *J. Great Lakes Res.*, **22**, 845-863, https://doi.org/10.1016/S0380-1330(96)71006-7.

Seeley, M., S. Goring, and J.W. Williams, 2019: Assessing the environmental and dispersal controls on *Fagus grandifolia* distributions in the Great Lakes region. *J. Biogeogr.*, **46**, 405-419, https://doi.org/10.1111/jbi.13491.

Shafer, M.A., C.A. Fiebrich, D.S. Arndt, S.E. Fredrickson, and T.W. Hughes, 2000: Quality assurance procedures in the Oklahoma Mesonetwork. *J. Atmos. Ocean. Tech.*, **17**, 474-494, https://doi.org/10.1175/1520-0426(2000)017%3C0474:QAPITO%3E2.0.CO;2.

Singh, D., D.L. Swain, J.S. Mankin, D.E. Horton, L.N. Thomas, B. Rajaratnam, and N.S. Diffenbaugh, 2016: Recent amplification of the North American winter temperature dipole. *J. Geophys. Res. Atmos.*, **121**, 9911–9928, https://doi.org/10.1002/2016JD025116.

Smith, A., N. Lott, and R. Vose, R., 2011: The integrated surface database: recent developments and partnerships. *Bull. Amer. Meteor. Soc*, **92**, 704–708, https://doi.org/10.1175/2011BAMS3015.1.

Walton, D., and A. Hall, 2018: An assessment of high-resolution gridded temperature datasets over California. *J. Climate*, **31**, 3789-3810, https://doi.org/10.1175/JCLI-D-17-0410.1.

Wang, Z.-H., C. Wang, and X. Yang, 2021: Dynamic synchronization of extreme heat in complex climate networks in the contiguous United States. *Urban Clim.*, **38**, 100909, https://doi.org/10.1016/j.uclim.2021.100909.

Xia, Y., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, D. Lettenmaier, V. Koren, Q. Duan, K. Mo, Y. Fan, and D. Mocko, 2012a: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.,* **117**, D03109. https://doi.org/10.1029/2011JD016048.

Xia, Y., K. Mitchell, M. Ek, B. Cosgrove, J. Sheffield, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, Q. Duan, and D. Lohmann, 2012b: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res.* **117**, D03110. https://doi.org/10.1029/2011JD016051.

Yu, L., S. Zhong, X. Bian, W.E. Heilman, and J.A. Andresen, 2014: Temporal and spatial variability of frost-free seasons in the Great Lakes region of the United States. *Int. J. Climatol.*, **34**, 3499-3514, https://doi.org/10.1002/joc.3923.

JOURNAL OF APPLIED AND SERVICE CLIMATOLOGY

Published Online September 5, 2025

Supplemental Online Material for

An Assessment of Temperature from Multiple Gridded Climate Datasets in a Region with Strong Physiographic Gradients: Michigan, United States

Michael T. Kiefer[1], Jeffrey A. Andresen[1], and William J. Baule[1,2]

[1] *Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI*

[2]*Current Affiliation: Department of Atmospheric Sciences and Southern Regional Climate Center, Texas A&M University, College Station, TX*

Corresponding Author: Michael T. Kiefer, mtkiefer@msu.edu

TABLE S1: Mean difference (MD: $PRISM_{D2}$ – station observation) for GHCN-D TMINd and TMAXd [°C], at eight airport (AP) sites for two 10-year periods, one pre-ASOS installation (1985-1994) and one post-ASOS installation (2005-2014). The eight airport sites included here exhibit sample size differences between the two decades of 3% or less; the remaining six airport sites exhibit sample size differences between the two decades of 20% or more and are omitted. See Table 2 for station metadata.

Station	TMINd		TMAXd	
	1985-1994	2005-2014	1985-1994	2005-2014
Alpena County Regional AP	-0.16	-0.38	0.08	0.14
Detroit Metro AP	-0.17	-0.32	0.08	0.15
Flint Bishop Intl. AP	-0.15	-0.36	0.03	0.09
Grand Rapids Gerald R Ford Intl. AP	-0.25	-0.38	0.11	-0.05
Lansing Capital City AP	-0.09	-0.24	0.05	0.13
Muskegon County AP	0.11	0.01	-0.09	-0.20
Sault Ste Marie Sanderson Field	-0.08	-0.57	0.14	0.10
Traverse City Cherry Capital AP	0.08	-0.01	-0.05	-0.03
Eight-station median	**-0.12**	**-0.34**	**0.06**	**0.09**

TABLE S2: Summary statistics for 2004-2018 estimated TMAXd [°C], at EW station sites: 20-station median number of days (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P1 ($PRISM_{D1}$) and P2 ($PRISM_{D2}$). In columns labeled, "Lag: no", difference in day definition between EW and PRISM is not accounted for: day n observation is compared to day n estimate. In columns labeled "Lag: yes", difference in day definition is accounted for: day n observation is compared to day n+1 estimate.

	TMAXd			
	P1		P2	
	Lag: no	Lag: yes	Lag: no	Lag: yes
N	5382	5381	5382	5381
MD	-0.31	-0.32	-0.41	-0.42
RMSD	3.97	1.46	4.01	1.51

TABLE S3: Summary statistics for 2004-2018 estimated TMAXd [°C], at GHCN-D station sites: 34-station median number of days (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P1 ($PRISM_{D1}$) and P2 ($PRISM_{D2}$). In columns labeled, "Lag: no", difference in day definition between EW and PRISM is not accounted for: day n observation is compared to day n estimate. In columns labeled "Lag: yes", difference in day definition is accounted for: day n observation is compared to day n+1 estimate.

	TMAXd			
	P1		P2	
	Lag: no	Lag: yes	Lag: no	Lag: yes
N	5449	5448	5449	5448
MD	-0.24	-0.24	-0.32	-0.32
RMSD	3.05	2.18	3.16	2.17

TABLE S4: Summary statistics for 1981-2018 estimated TMAXd [°C], at GHCN-D station sites: 34-station median number of days (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P1 ($PRISM_{D1}$) and P2 ($PRISM_{D2}$). In columns labeled, "Lag: no", difference in day definition between EW and PRISM is not accounted for: day n observation is compared to day n estimate. In columns labeled "Lag: yes", difference in day definition is accounted for: day n observation is compared to day n+1 estimate.

	TMAXd			
	P1		P2	
	Lag: no	Lag: yes	Lag: no	Lag: yes
N	13746	13745	13746	13745
MD	-0.14	-0.14	-0.27	-0.25
RMSD	2.91	2.33	3.24	2.29

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

TABLE S5: Summary statistics for (a) 2004-2018 and (b) 1981-2018 estimated TMINd and TMAXd [°C], at GHCN-D station sites (airports only): 14-station median number of days (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P1 ($PRISM_{D1}$) and P2 ($PRISM_{D2}$). In columns labeled, "GHCN-D", daily maxima and minima are obtained directly from the GHCN-D database and vary in day definition from station to station and across time. In columns labeled "ISD", hourly observations are obtained from the hourly Integrated Surface Database (ISD), and daily maxima and minima are computed as the highest and lowest values during the period from 12:00 UTC day n-1 – 12:00 UTC day n.

(a)	TMINd				TMAXd			
	P1		P2		P1		P2	
	GHCN-D	ISD	GHCN-D	ISD	GHCN-D	ISD	GHCN-D	ISD
N	5363	5363	5363	5363	5364	5364	5364	5364
MD	0.03	-0.31	0.09	-0.28	-0.38	0.12	-0.44	0.08
RMSD	2.60	0.82	2.62	0.56	3.92	0.48	3.95	0.41

(b)	TMINd				TMAXd			
	P1		P2		P1		P2	
	GHCN-D	ISD	GHCN-D	ISD	GHCN-D	ISD	GHCN-D	ISD
N	12862	12862	12862	12862	12863	12863	12863	12863
MD	-0.15	-0.45	0.15	-0.22	-0.22	0.27	-0.36	0.11
RMSD	2.70	1.30	2.67	0.48	3.88	0.91	4.05	0.42

TABLE S6: Summary statistics for 1981-2016 estimated TMINd and TMAXd [°C], with seasonal breakdown, at GHCN-D station sites: 34 -station median number of days (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P1 ($PRISM_{D1}$), P2 ($PRISM_{D2}$), G (gridMET), and T (TopoWx). Note: for TMAXd, difference in day definition between GHCN-D and PRISM is partially accounted for by comparing day n observation to day n+1 estimate (Section 2.5).

		TMINd				TMAXd			
	Months	P1	P2	G	T	P1	P2	G	T
N	**ALL**	**12961**	**12961**	**12961**	**12961**	**13038**	**13038**	**13038**	**13038**
	DJF	3206	3206	3206	3206	3238	3238	3238	3238
	MAM	3268	3268	3268	3268	3286	3286	3286	3286
	JJA	3276	3276	3276	3276	3290	3290	3290	3290
	SON	3234	3234	3234	3234	3250	3250	3250	3250
MD	**ALL**	**-0.10**	**0.04**	**0.05**	**-0.08**	**-0.13**	**-0.26**	**0.04**	**0.00**
	DJF	-0.18	-0.08	0.02	-0.17	-0.10	-0.26	0.00	-0.03
	MAM	-0.12	0.04	0.21	-0.10	-0.10	-0.28	0.19	0.07
	JJA	0.05	0.19	0.04	-0.10	-0.12	-0.23	0.07	0.07
	SON	-0.05	0.07	-0.10	-0.07	-0.27	-0.38	-0.17	-0.10
RMSD	**ALL**	**2.49**	**2.50**	**2.72**	**1.64**	**2.32**	**2.30**	**2.23**	**1.61**
	DJF	2.91	3.06	3.04	1.93	2.41	2.32	2.19	1.47
	MAM	2.38	2.44	2.78	1.69	2.63	2.61	2.57	1.96
	JJA	1.63	1.58	2.27	1.36	1.80	1.71	2.08	1.44
	SON	2.55	2.57	2.46	1.60	2.36	2.31	2.15	1.49

TABLE S7: Summary statistics for 1981-2016 estimated TMINa and TMAXa [°C], at GHCN-D station sites: 34-station median number of years (N), mean difference (MD), and root-mean-square difference (RMSD). Gridded datasets are labeled P1 ($PRISM_{D1}$), P2 ($PRISM_{D2}$), G (gridMET), and T (TopoWx).

	TMINa				TMAXa			
	P1	P2	G	T	P1	P2	G	T
N	36	36	36	36	36	36	36	36
MD	0.19	0.44	2.59	1.04	-0.38	-0.23	-0.26	-0.94
RMSD	1.59	1.34	3.48	1.87	0.77	0.61	1.29	1.22

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

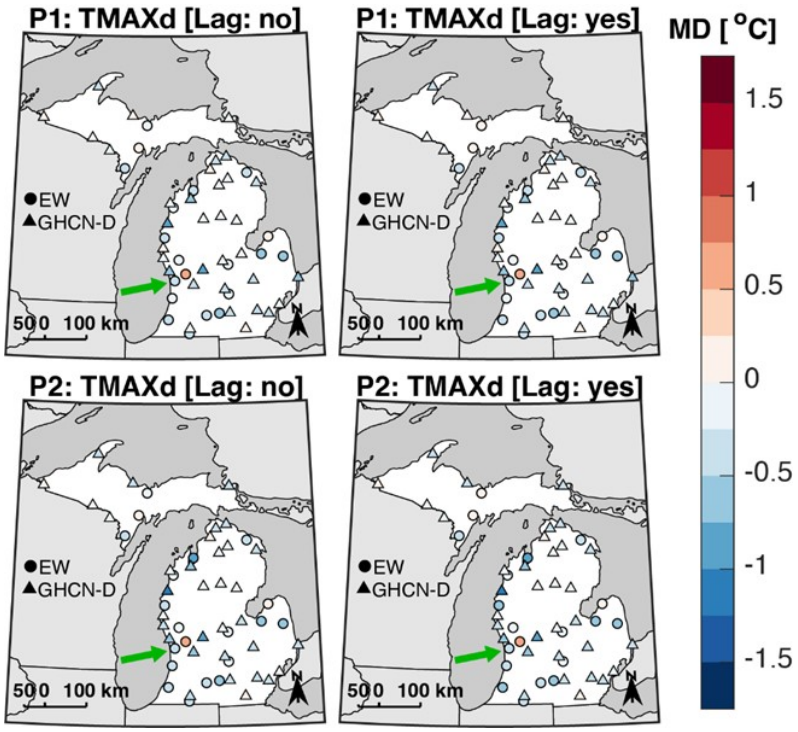

FIGURE S1: EW (circle) and GHCN-D (triangle) station symbols, color-coded by daily temperature mean difference (MD), for period 2004-2016: columns: TMAXd with and without PRISM lag; rows: $PRISM_{D1}$ (P1) and $PRISM_{D2}$ (P2). The minimum number of days (N) among both observational networks is 4223; N at individual stations is not shown. Green arrow points to West Olive EW station site (E20 in Fig. 1, Table 1); next circle to northeast is the Sparta EW station site (E18 in Fig. 1, Table 1).

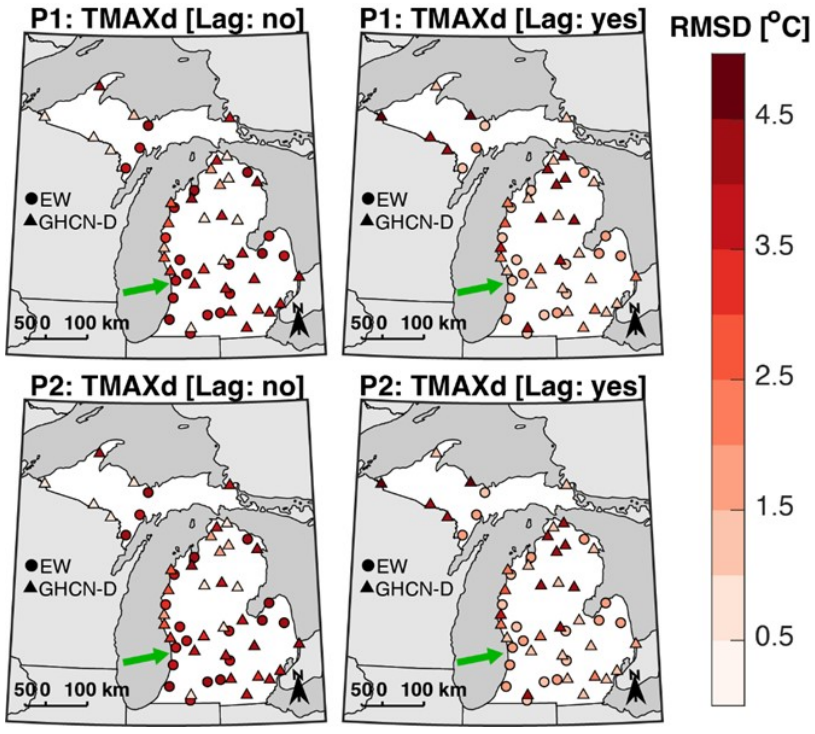

FIGURE S2: EW (circle) and GHCN-D (triangle) station symbols, color-coded by daily temperature root-mean-square difference (RMSD), for period 2004-2016: columns: TMAXd with and without PRISM lag; rows: $PRISM_{D1}$ (P1) and $PRISM_{D2}$ (P2). The minimum number of days (N) among both observational networks is 4223; N at individual stations is not shown. Green arrow points to West Olive EW station site (E20 in Fig. 1, Table 1); next circle to northeast is the Sparta EW station site (E18 in Fig. 1, Table 1).

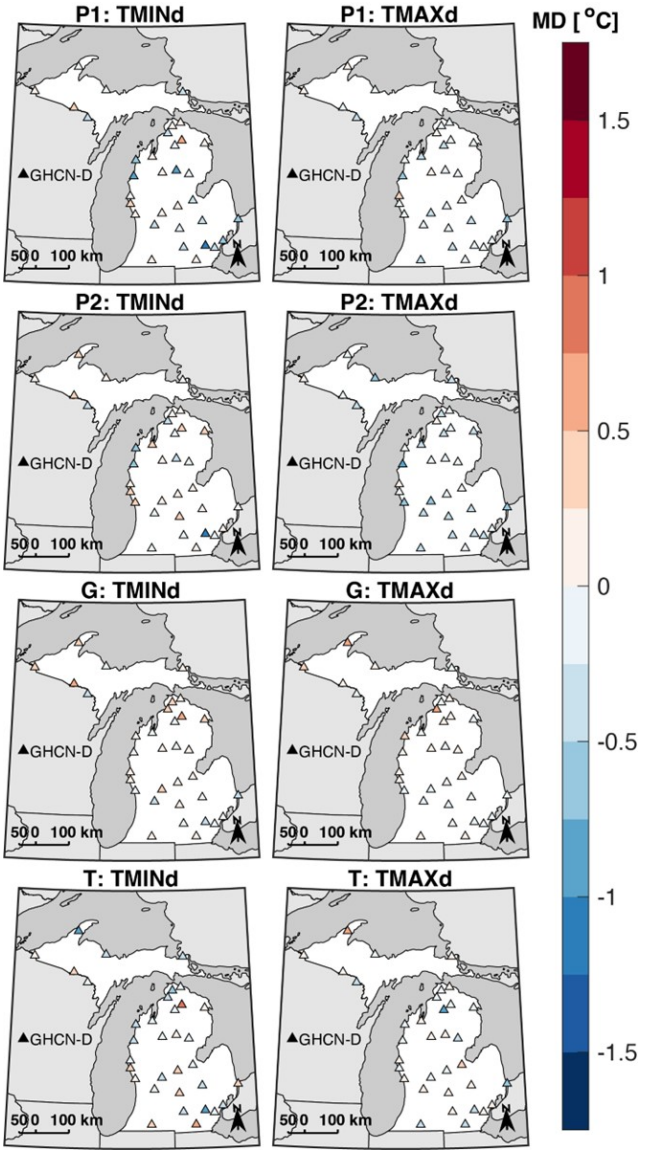

FIGURE S3: GHCN-D station symbols, color-coded by daily temperature mean difference (MD), for period 1981-2016: columns: TMINd and TMAXd (w/ PRISM lag); rows: $PRISM_{D1}$ (P1), $PRISM_{D2}$ (P2), gridMET (G), and TopoWx (T). The minimum number of days (N) is 11275 for both TMINd and TMAXd; N at individual stations is not shown. EW station sites are omitted as network period of record begins in 2004.

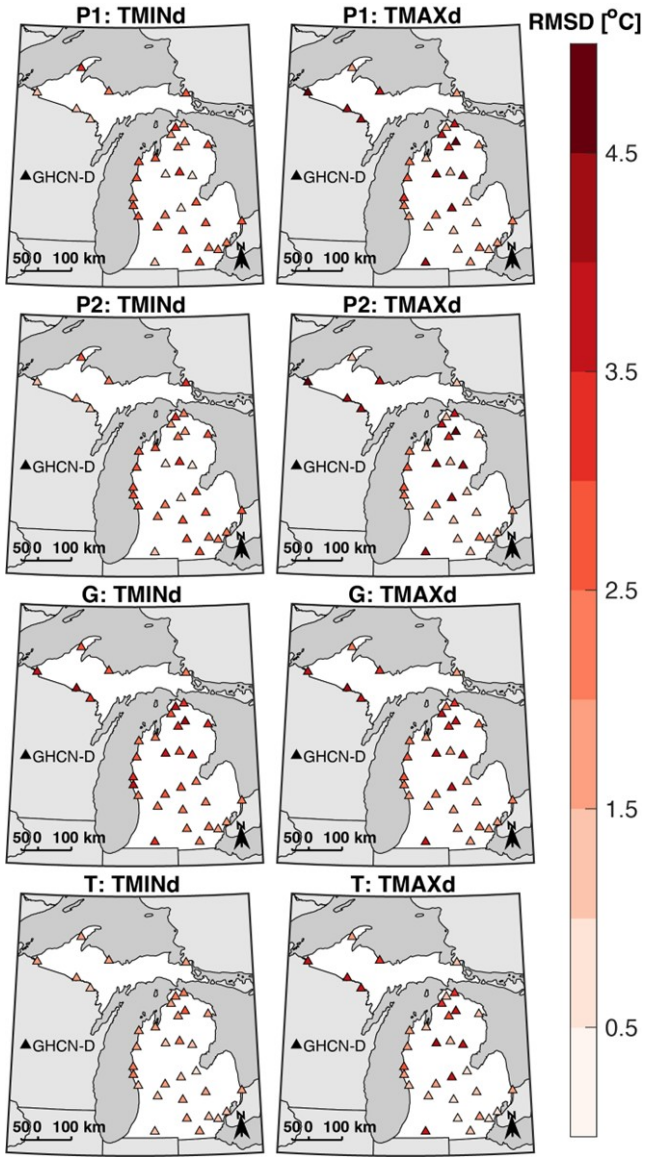

FIGURE S4: GHCN-D station symbols, color-coded by daily temperature root-mean-square difference (RMSD), for period 1981-2016: columns: TMINd and TMAXd (w/ PRISM lag); rows: $PRISM_{D1}$ (P1), $PRISM_{D2}$ (P2), gridMET (G), and TopoWx (T). The minimum number of days (N) is 11275 for both TMINd and TMAXd; N at individual stations is not shown. EW station sites are omitted as network period of record begins in 2004.

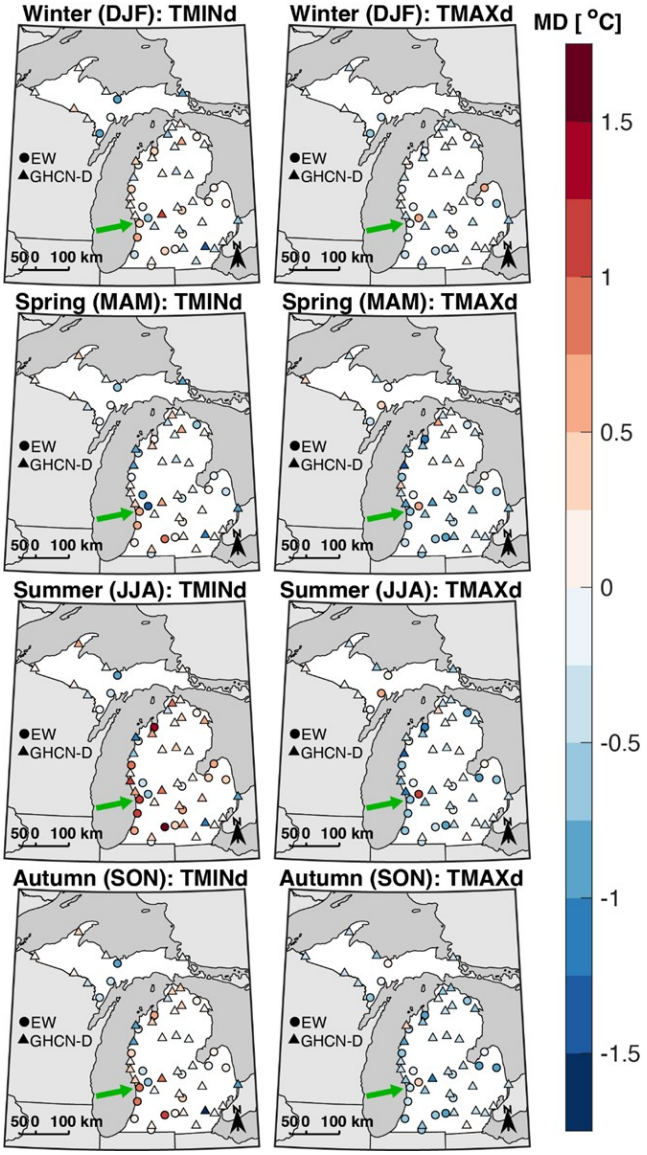

FIGURE S5: EW (circle) and GHCN-D (triangle) station symbols, color-coded by daily temperature mean difference (MD), for period 2004-2016: columns: TMINd and TMAXd (w/ PRISM lag); rows: winter (December-February), spring (March -May), summer (June-August), and autumn (September-November); for $PRISM_{D2}$ only. The minimum number of days (N) among both observational networks is 969 for TMINd and 970 for TMAXd; N at individual stations is not shown. Green arrow points to West Olive EW station site (E20 in Fig. 1, Table 1); next circle to northeast is the Sparta EW station site (E18 in Fig. 1, Table 1).

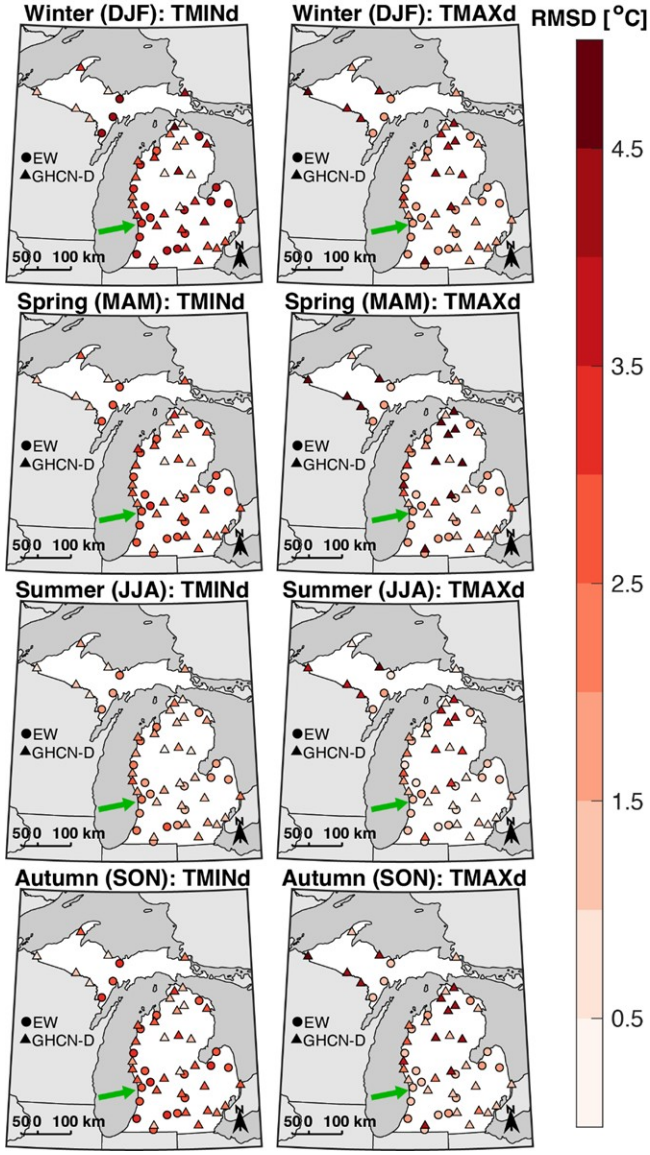

FIGURE S6: EW (circle) and GHCN-D (triangle) station symbols, color-coded by daily temperature root-mean-square difference (RMSD), for period 2004-2016: columns: TMINd and TMAXd (w/ PRISM lag); rows: winter (December-February), spring (March-May), summer (June-August), and autumn (September-November); for $PRISM_{D2}$ only. The minimum number of days (N) among both observational networks is 969 for TMINd and 970 for TMAXd; N at individual stations is not shown. Green arrow points to West Olive EW station site (E20 in Fig. 1, Table 1); next circle to northeast is the Sparta EW station site (E18 in Fig. 1, Table 1).

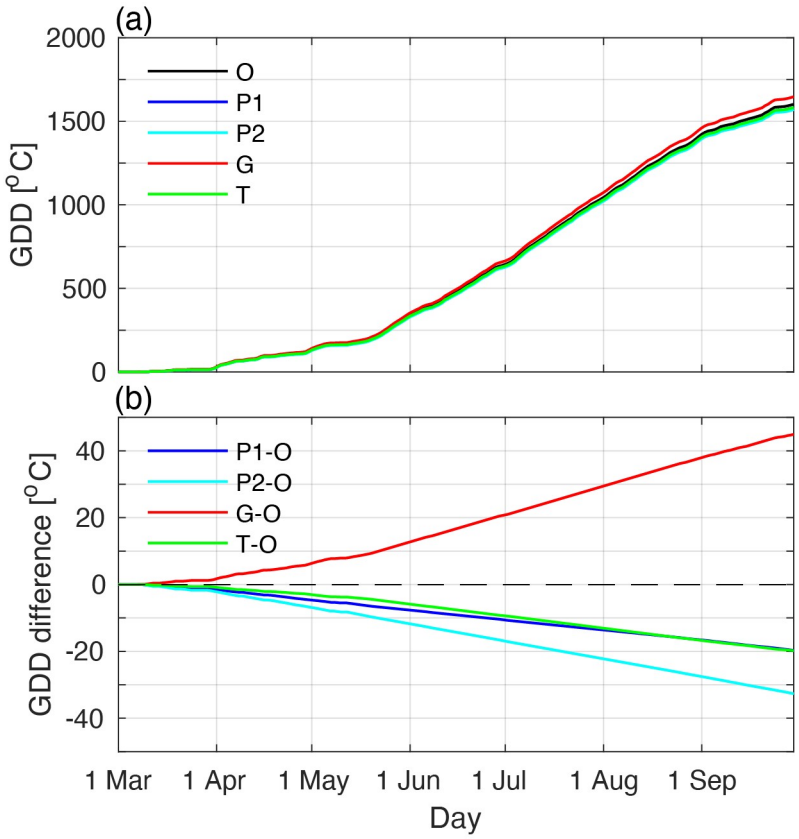

FIGURE S7: Time series of observed cumulative growing degree days (GDD) base 10 °C at East Lansing (E14 in Fig. 1, Table 1) between 1 Mar 2010 and 30 Sep 2010 (O), computed from observed TMINd and TMAXd, with additional time series constructed by adding the gridded estimate–observation mean difference (MD) in Table 3 to observed TMINd and TMAXd (P1: $PRISM_{D1}$; P2: $PRISM_{D2}$; G: gridMET; T: TopoWx). Panels (a) and (b) depict cumulative GDD and cumulative GDD relative to the observed (i.e., unadulterated) GDD, respectively.

Journal of Applied and Service Climatology, Volume 2025, Issue 001, DOI: https://doi.org/10.46275/JOASC.2025.09.001

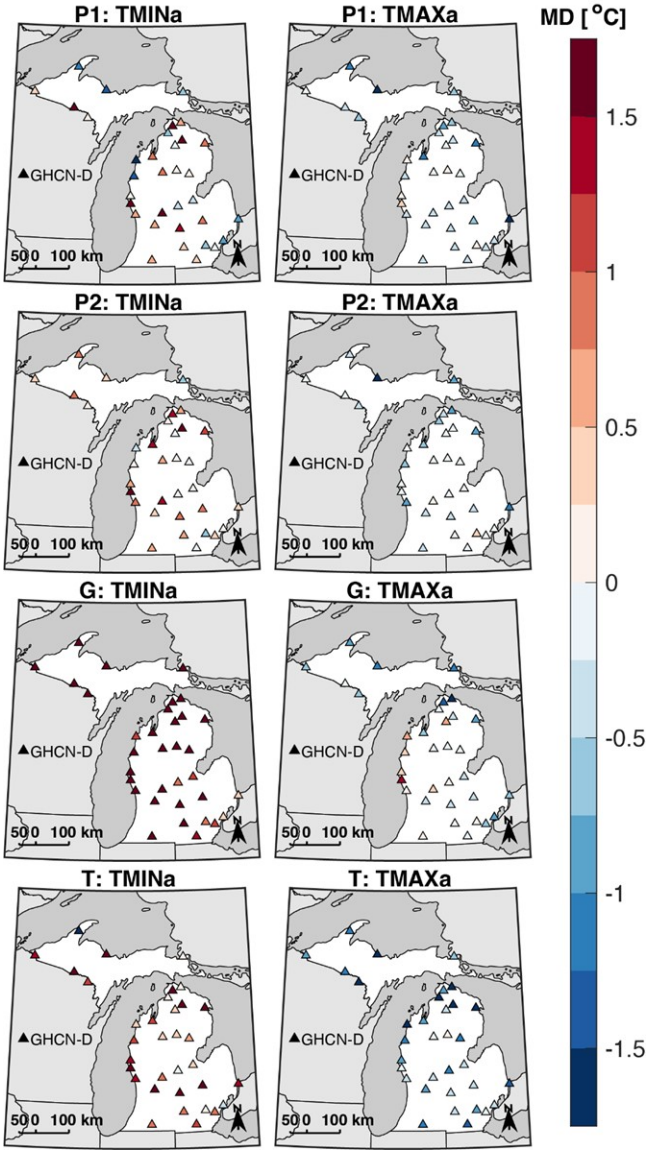

FIGURE S8: GHCN-D station symbols, color-coded by annual extreme temperature mean difference (MD), for period 1981-2016: columns: TMINa and TMAXa; rows: PRISMD1 (P1), PRISMD2 (P2), gridMET (G), and TopoWx (T). The minimum number of years (N) is 30 for both TMINa and TMAXa; N at individual stations is not shown. EW station sites are omitted as network period of record begins in 2004.

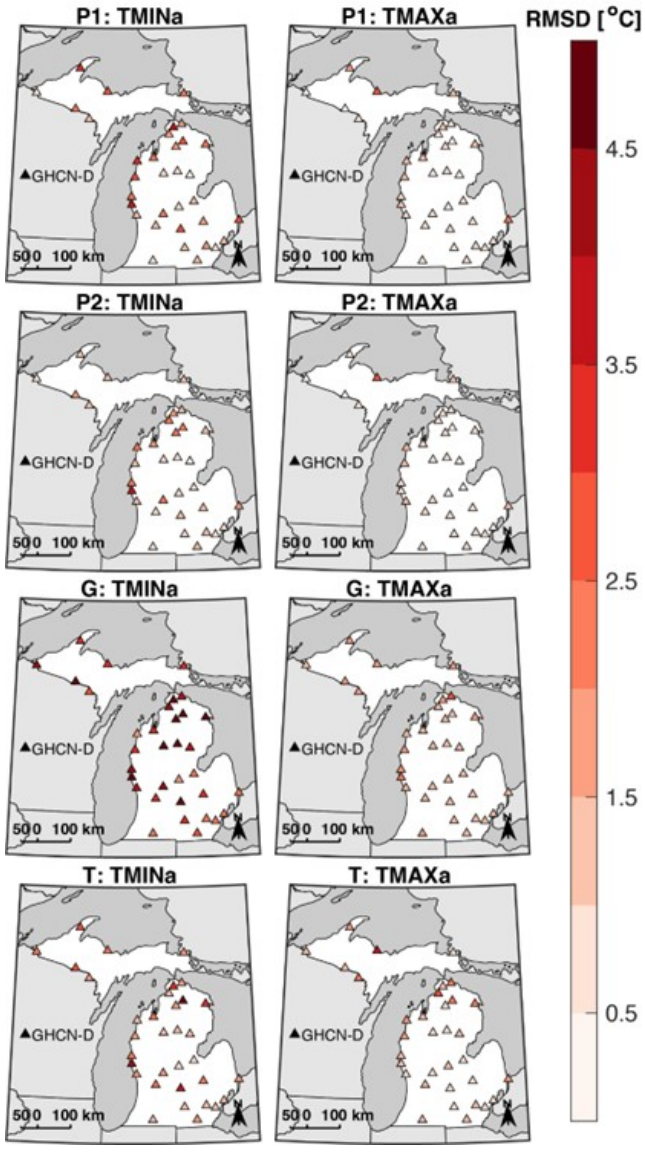

FIGURE S9: GHCN-D station symbols, color-coded by annual extreme temperature root-mean-square difference (RMSD), for period 1981-2016: columns: TMINa and TMAXa; rows: $PRISM_{D1}$ (P1), $PRISM_{D2}$ (P2), gridMET (G), and TopoWx (T). The minimum number of years (N) is 30 for both TMINa and TMAXa; N at individual stations is not shown. EW station sites are omitted as network period of record begins in 2004.