

## RESEARCH ARTICLE

10.1002/2016JD025920

## Key Points:

- In situ testing of a popular gridded air temperature model in a semiarid watershed revealed high accuracy but also systematic biases
- Model accuracy increased by moving from 4 km to 800 m grid size, but further downscaling did not improve results
- Biases in daily temperature extremes are linked to siting differences between in situ and model source data, and topoclimatic mechanisms

## Correspondence to:

S. Strachan,  
scotty@dayhike.net;  
strachan@unr.edu

## Citation:

Strachan, S., and C. Daly (2017), Testing the daily PRISM air temperature model on semiarid mountain slopes, *J. Geophys. Res. Atmos.*, 122, 5697–5715, doi:10.1002/2016JD025920.

Received 8 SEP 2016

Accepted 21 MAY 2017

Accepted article online 26 MAY 2017

Published online 7 JUN 2017

## Testing the daily PRISM air temperature model on semiarid mountain slopes

Scotty Strachan<sup>1</sup>  and Christopher Daly<sup>2</sup> 

<sup>1</sup>Nevada Climate-ecohydrological Assessment Network, Department of Geography, University of Nevada, Reno, Nevada, USA, <sup>2</sup>PRISM Climate Group, Northwest Alliance for Computational Science and Engineering, Oregon State University, Corvallis, Oregon, USA

**Abstract** Studies in mountainous terrain related to ecology and hydrology often use interpolated climate products because of a lack of local observations. One data set frequently used to develop plot-to-watershed-scale climatologies is PRISM (Parameter-elevation Regression on Independent Slopes Model) temperature. Benefits of this approach include geographically weighted station observations and topographic positioning modifiers, which become important factors for predicting temperature in complex topography. Because of the paucity of long-term climate records in mountain environments, validation of PRISM algorithms across diverse regions remains challenging, with end users instead relying on atmospheric relationships derived in sometimes distant geographic settings. Presented here are results from testing observations of daily temperature maximum ( $T_{MAX}$ ) and minimum ( $T_{MIN}$ ) on 16 sites in the Walker Basin, California-Nevada, located on open woodland slopes ranging from 1967 to 3111 m in elevation. Individual site mean absolute error varied from 1.1 to 3.7°C with better performance observed during summertime as opposed to winter. We observed a consistent cool bias in  $T_{MIN}$  for all seasons across all sites, with cool bias in  $T_{MAX}$  varying with season. Model error for  $T_{MIN}$  was associated with elevation, whereas model error for  $T_{MAX}$  was associated with topographic radiative indices (solar exposure and heat loading). These results demonstrate that temperature conditions across mountain woodland slopes are more heterogeneous than interpolated models (such as PRISM) predict, that drivers of these differences are complex and localized in nature, and that scientific application of atmospheric/climate models in mountains requires additional attention to model assumptions and source data.

**Plain Language Summary** Knowledge of daily-to-seasonal climate (such as air temperature) in mountain areas is important for assessment of landscape conditions related to plants, animals, and resources such as water supply. Because few actual observations of climate processes exist in mountains, scientists have developed models to estimate parameters like temperature across landscapes. In this paper we test one commonly used spatial temperature model using observations and report the model error as well as influential factors. Our conclusions state that while for some science and management uses the model differences from observations are inconsequential, improper application of the model in other contexts without local verification or consideration of assumptions would lead to incorrect results. We also show that the location of long-term monitoring stations in mountain landscapes likely impacts model accuracy more than differences in network instrumentation practices. Therefore, scientists or managers seeking to leverage such models of temperature to make decisions need to be aware of both the representation of source data and assumptions made during the modeling process. This study underscores the need for additional long-term monitoring of climate processes in mountain areas given the importance of such regions to society in terms of resources and value.

### 1. Introduction

Atmospheric conditions in mountainous terrain at ecohydrologically important scales (e.g., subkilometer) remain challenging to both observe and estimate. Scientific disciplines ranging from snow hydrology to bioecology to palaeoclimatology struggle with obtaining accurate data with reasonable estimates of uncertainty on the most basic of climatic parameters in complex topography [Lundquist and Cayan, 2007; Fridley, 2009; Dobrowski, 2011; Graae et al., 2012; Stoklosa et al., 2014], which could have profound effects on interpretation of large-scale cause and local-scale effect [Lookingbill and Urban, 2005; Minder et al., 2010; Warren et al., 2014; Oyler et al., 2015]. The importance of research on mountain processes continues to be recognized as crucial

for resiliency of socio-ecological systems on a global scale [Messerli and Ives, 1997; Viviroli and Weingartner, 2004; Foley et al., 2005; Gurung et al., 2012], and therefore, scientific investigators and policymakers will use whichever data are made available in an information-poor environment.

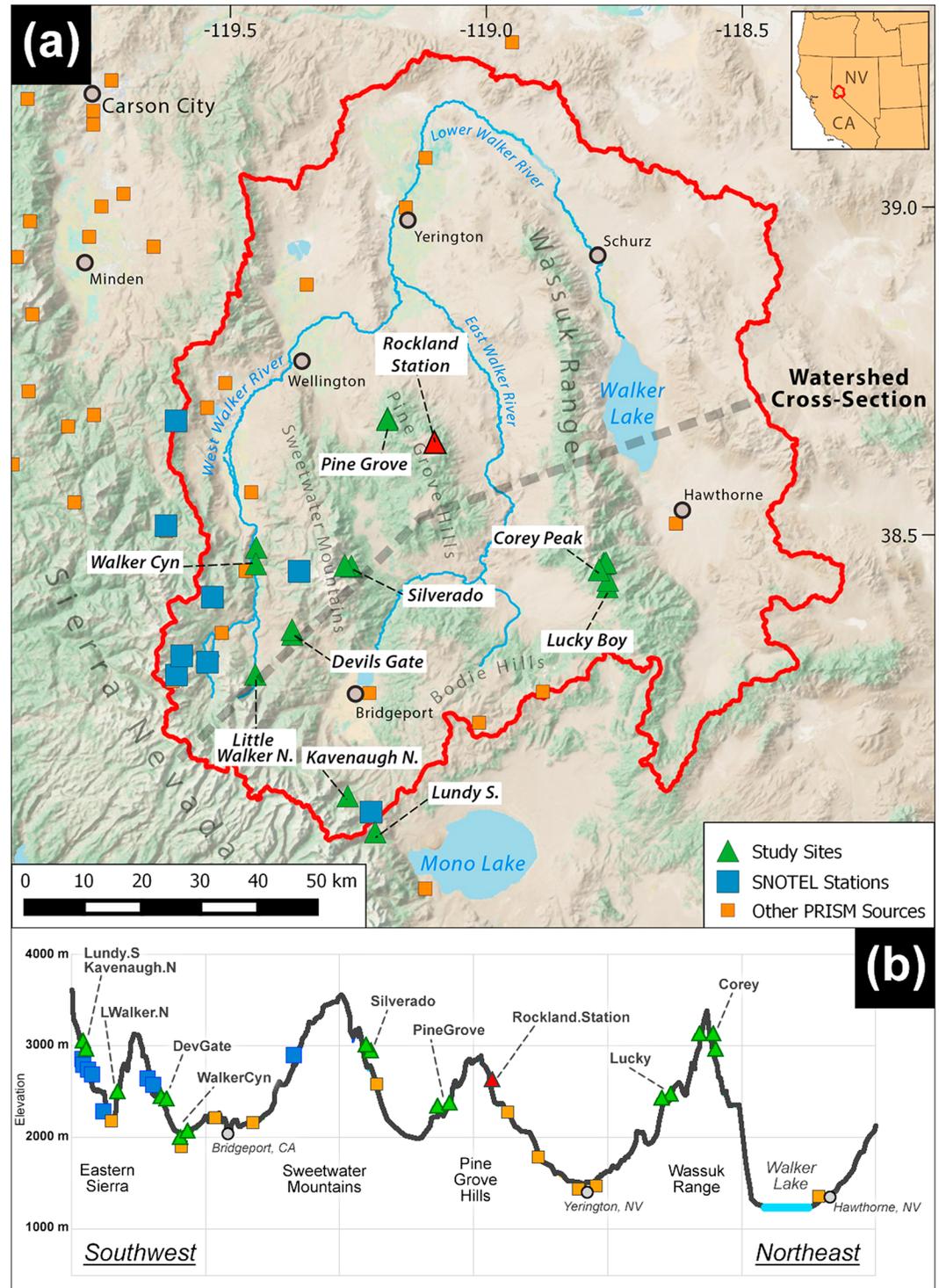
Gridded products extrapolating point observations of meteorological parameters to landscape scales have advanced in recent years [Brohan et al., 2006; Daly et al., 2008; Haylock et al., 2008; Thornton et al., 2012]. Regardless of model sophistication, the accuracy of these modeling efforts varies with density and quality of source data networks [Hamlet and Lettenmaier, 2005; Daly, 2006; Hofstra et al., 2009; McEvoy et al., 2014; Stoklosa et al., 2014; Oyler et al., 2015], and it is in mountainous topographic regions that observational data are the most scarce [Palecki and Groisman, 2011]. Source data for models are typically derived from valley-situated ground stations and ridgetop or upper air data, leaving the category of mountain slopes poorly represented in observational data sets. Temporal and spatial resolutions of climate models are increasing [e.g., Fowler et al., 2007; Skamarock et al., 2008; Feser et al., 2011], but newer ground networks that add calibration and verification data are generally coarse in spatial resolution [e.g., Leeper et al., 2015]. It is not uncommon to see analyses leveraging gridded climate data sets for process modeling or climate change impact prediction, and yet not including verification observations and resulting error estimates. In situ measurements of climate parameters such as air temperature are critical inputs to this process, as remote-sensing-derived estimates of near-surface (e.g., 2 m) air temperature still contain significant sources of error [Kalma et al., 2008; Hengl et al., 2012]. Validation of interpolated climate products within mountainous regions is therefore an important scientific activity, in order to inform users on model quality/accuracy as well as provide feedback to the modelers themselves on the performance of their products in varying geographical and seasonal settings [Minder et al., 2010; McGuire et al., 2012; Holden et al., 2015].

Here we present a case study of topographic temperature model testing in complex topography in the Great Basin region of North America, whereby the daily time step PRISM (Parameter-elevation Regression on Independent Slopes Model) [Daly et al., 2008] temperature product was evaluated for performance across a range of middle to high elevations on homogeneous topographic slopes at spatially distributed sites at the large watershed scale. Daily maximum ( $T_{MAX}$ ) and minimum ( $T_{MIN}$ ) air temperature estimations from PRISM at four spatial scales, i.e., 4 km grid, 30 arcsec (~800 m) grid, downscaled 3 arcsec (~80 m) grid, and point interpolations from surrounding grid cells, are compared to in situ observations made with air temperature sensors on 16 mountain sites. Results of the comparison are examined for possible instrumental, topographic, or mechanistic sources of differences. Impacts of model bias on example applications are also presented, with discussion on the challenges of modeling near-surface mountain meteorology.

## 2. Study Geography

Sixteen monitoring sites in mountainous topography are located within the Walker River Basin, a large (10,200 km<sup>2</sup>) semiarid watershed in the western United States. This watershed is considered to be on the climate transition zone between the Sierra Nevada and Great Basin Desert ecoregions of North America (Figure 1). Several contemporary scientific, conservation, and policy-related projects in the watershed integrate PRISM data as part of their efforts [Lopes and Allander, 2009; Millar and Westfall, 2010; Mejia et al., 2012; Knick et al., 2013; Saito et al., 2014; Hatchett et al., 2015].

The monitoring sites are associated with an ongoing palaeoclimate study using upper and lower treeline species (*Pinus flexilis* and *Pinus monophylla*), ranging from 1967 to 3111 m in elevation and generally co-located in opposite-aspect pairs. Sites are distributed across four mountain ranges, ranging from west to east (Figure 1), and represent a spatial gradient from Eastern Sierra to Great Basin ecosystems. Topographic positioning of the sites varies (Table 1), but, in general, the locations were designed to represent homogenous slope features rather than peaks, ridgetops, saddles, gullies, or canyon floors. In this way, the observations are specifically targeted at general air conditions that are free of influence from local cold-air pools and topographically enhanced wind velocity. Thus, the primary drivers of air temperature on the study sites are radiative processes, larger-scale airflow, and local lapse rates. While it is true that other processes such as cold-air pooling, windiness, and snow presence have dramatic and important local effects, proper estimation of general air conditions across topography is the first step to being able to correctly model air temperature behavior at the watershed scale. We have therefore optimized the locations of in



**Figure 1.** (a) Study area map with (b) elevation cross section. The distribution of paired study sites (green triangles) covered the major mountain ranges of the Walker Basin, starting with the Sierra front and progressing east. Stations providing source data for PRISM during the study period are shown (squares). A cross section of the watershed elevations (denoted by a dashed line in Figure 1a) provides topographic context for stations and study sites, which are plotted in Figure 1b even if they are located well away from the line itself. Most of the stations used by PRISM as source data are located in valley bottoms, with midelevation stations being primarily from the SNOW-TElemetry network (SNOTEL) and geographically skewed toward the western portion of the watershed.

**Table 1.** Study Site Characteristics and PRISM Error Statistics

Site Characteristics					PRISM <sub>800</sub> Error Statistics, Daily Values October 2013 to September 2015							
					$T_{MAX}$ (°C)				$T_{MIN}$ (°C)			
Site	Elev (m)	Slope (deg)	Aspect (deg)	DAH Index	Bias	MAE	SD	$r^2$ ( $p < 0.0001$ )	Bias	MAE	SD	$r^2$ ( $p < 0.0001$ )
Lucky.N	2480	21	325	-0.20	-1.37	1.83	1.78	0.97	-1.09	1.78	1.93	0.94
Lucky.S	2497	30	136	0.18	-3.45	3.49	1.54	0.97	-2.95	2.99	1.74	0.96
DevGate.N	2378	39	352	-0.52	0.70	1.15	1.30	0.98	-2.15	2.21	1.30	0.97
DevGate.S	2360	27	200	0.43	-3.47	3.49	1.31	0.97	-3.28	3.30	1.42	0.96
CoreyLow.N	2977	32	329	-0.31	0.78	1.68	1.96	0.96	-1.24	1.53	1.43	0.97
Corey.N	3104	34	307	-0.14	1.07	1.86	2.07	0.96	-0.27	1.22	1.49	0.97
Corey.S	3111	22	101	-0.07	0.95	1.42	1.46	0.97	-0.85	1.50	1.66	0.96
LWalker.N	2452	16	304	-0.04	-1.81	1.98	1.39	0.97	-0.91	1.14	1.30	0.96
Silverado.S	2897	26	157	0.30	-1.20	1.49	1.30	0.97	-2.67	2.80	1.83	0.94
Silverado.N	2937	31	270	0.20	-1.14	1.62	1.68	0.96	-1.99	2.13	1.69	0.94
PineGrove.N	2355	18	347	-0.24	-0.26	1.65	2.07	0.95	-2.58	2.65	1.68	0.95
PineGrove.S	2371	17	226	0.27	-3.19	3.32	1.86	0.96	-2.17	2.27	1.57	0.96
Kavanaugh.N	3000	37	310	-0.17	1.09	1.66	1.71	0.95	-1.29	2.07	2.17	0.91
Lundy.S	2911	33	212	0.51	-0.33	1.49	1.77	0.94	-1.27	1.86	2.00	0.92
WalkerCyn.N	2036	45	12	-0.67	1.69	1.82	1.68	0.97	-3.65	3.70	1.99	0.94
WalkerCyn.S	1967	38	188	0.57	-2.13	2.38	1.84	0.95	-2.88	2.93	1.72	0.95
Overall mean	2614	29	236	0.01	-0.75	2.02	1.67	0.96	-1.95	2.25	1.68	0.95

situ observations to focus on response of temperature to larger-scale topographic characteristics such as elevation, slope, and aspect, which represent the first order of variability in mountainous terrain.

### 3. Data

#### 3.1. PRISM Temperature Estimates

Grids of daily  $T_{MAX}$  and  $T_{MIN}$  estimates for the Walker River Basin were obtained from the PRISM AN81d data set [*Parameter-elevation Regression on Independent Slopes Model (PRISM) Climate Group*, 2016]. This data set spans 1981–present and covers the conterminous United States at 2.5 arcmin (~4 km) and 30 arcsec (~800 m) resolutions. Mapping of daily  $T_{MAX}$  and  $T_{MIN}$  was performed using the PRISM modeling system [Daly et al., 1994, 2002, 2003, 2008]. For each grid cell each day, PRISM calculated a local linear regression function between station temperature and a predictor grid (see below). Nearby stations entering the regression were assigned weights based primarily on the physiographic similarity of the station to the grid cell. In addition to distance and elevation differences between station and grid cell, physiographic factors accounted for included the level of the approximate wintertime inversion (if any) and a topographic index, which is measure of local topographic position (i.e., the station’s elevation relative to the surrounding terrain). Detailed descriptions of the PRISM model algorithms, structure, input grids, and operation are given in Daly et al. [2002, 2008]. Specific methods relevant to this study are described below.

The PRISM AN81d ([http://prism.oregonstate.edu/documents/PRISM\\_datasets.pdf](http://prism.oregonstate.edu/documents/PRISM_datasets.pdf)) temperature time series was developed using climatologically aided interpolation (CAI). CAI uses an existing climate grid to improve the interpolation of another climate element for which data may be sparse or intermittent in time [Willmott and Robeson, 1995; Funk et al., 2003; Hamlet and Lettenmaier, 2005; Daly, 2006; Daly et al., 2012, 2015]. This method relies on the assumption that local spatial patterns of the climate element being interpolated closely resemble those of the existing climate grid (called the predictor grid). The use of CAI in mapping daily temperature for the AN81d data set involved using existing PRISM 1981–2010 monthly  $T_{MAX}$  and  $T_{MIN}$  normals as the predictor grids [PRISM Climate Group, 2016]. For example, interpolating  $T_{MAX}$  for 5 January 2014 involved using the January 1981–2010 normal  $T_{MAX}$  grid as the independent variable in the local PRISM regression function for a grid cell and nearby station values of 5 January 2014  $T_{MAX}$  as the dependent variable. Details on normals mapping methods are available from Daly et al. [2008].

Station temperature data in and around the Walker River Basin used in the PRISM AN81d data sets (Figure 1) came primarily from four main networks: (1) U.S. Department of Agriculture (USDA) Natural Resources Conservation Service SNOW TElemetry (SNOTEL; <http://www.wcc.nrcs.usda.gov/snow/>), (2) National

Weather Service (NWS) Cooperative Observer Program (COOP; <http://cdo.ncdc.noaa.gov/CDO/cdo/>) and (3) Automated Surface Observing System (<ftp://ftp.ncdc.noaa.gov/pub/data/noaa/>), and (4) USDA Forest Service and Bureau of Land Management Remote Automatic Weather Stations (RAWS; <http://www.raws.dri.edu>). Station data were screened for adherence to a “PRISM day” criterion. A PRISM day is defined as 1200 UTC–1200 UTC (e.g., 4 A.M.–4 A.M. PST). Once-per-day observation times, such as those from NWS COOP stations, must fall within  $\pm 4$  h of the PRISM day to be included in the AN81d  $T_{\text{MAX}}$  and  $T_{\text{MIN}}$  data sets. The data set uses a day-ending naming convention; e.g., a day ending at 1200 UTC on 1 January is labeled 1 January. The PRISM day definition was established to match the “hydrologic day,” which is a standard in river modeling, and align with the observation times of most once-per-day observers (e.g., NWS COOP), which are in the morning.

### 3.2. In Situ Observations

At each site, observations of daily  $T_{\text{MAX}}$  and  $T_{\text{MIN}}$  (in  $^{\circ}\text{C}$ ) for 23 months (1 October 2013 to 1 September 2015) were made using Maxim iButton DS1922L thermochron dataloggers placed at approximately 1.7 m height above ground level in the vegetation canopy interspace. These data are completely independent of PRISM and were not used in model generation. The iButtons were configured using Maxim 1-Wire software to log their case temperature every 60 min, and the data were subsequently processed to extract the daily high and low temperature. Real-Time-Clock (RTC) drift of each iButton was evaluated at the end of the collection period to ensure that cumulative RTC error did not exceed 50% of the observation interval. In order to replicate standard weather station temperature measurements as closely as possible with the sensor deployments, the iButtons were placed inside six-plate Gill-type radiation shields using nonconductive mounting holders that mimicked typical temperature probe head positioning. During the 2013–2015 interval, winter-time bias of observed temperatures due to snow presence or interference with sensors at 1.7 m heights was minimized by record-setting low snowpack levels across the region [Swain, 2015], verified on the study sites with ground-level iButtons monitoring snow presence qualitatively. A subset of these observations is being continued through longer-term study.

Because this study focused on daily extremes, the time alignment of PRISM with the observation data in this study was checked using two different schemes: (1) PRISM Day, which consists of daily  $T_{\text{MAX}}$  and  $T_{\text{MIN}}$  taken from observations spanning the hours 1200<sub>DAY-1</sub>–1200<sub>DAY+0</sub> GMT (0400<sub>DAY-1</sub>–0400<sub>DAY+0</sub> PST local time), and (2) “Local Day,” which spans the hours 0800<sub>DAY+0</sub>–0800<sub>DAY+1</sub> UT (0000<sub>DAY+0</sub>–0000<sub>DAY+1</sub> PST local time) and lagging the PRISM day by one, such that the daily  $T_{\text{MAX}}$  and  $T_{\text{MIN}}$  from in situ observations from the interval 0000<sub>DAY+0</sub>–0000<sub>DAY+1</sub> PST were aligned with PRISM daily  $T_{\text{MAX}}$  and  $T_{\text{MIN}}$  from 0400<sub>DAY+0</sub>–0400<sub>DAY+1</sub> PST. It was found that statistical comparisons using the “Local Day” scheme were slightly improved, likely due to PRISM’s use of local midnight-to-midnight SNOTEL  $T_{\text{MAX}}$  and  $T_{\text{MIN}}$  data, which fall within the  $\pm 4$  h PRISM day window. Therefore, the results shown are derived from Local Day time alignment rather than PRISM Day.

## 4. Test and Analysis Methods

Our evaluation of PRISM temperatures in this context followed three general steps: (1) comparison of PRISM to observations, (2) investigation of potential error sources for both PRISM and the observations, and (3) evaluation of impacts of model departures on common hydrometeorological variables.

### 4.1. PRISM: Observation Comparison

Differences between model estimates and observations were initially calculated at the primary (i.e., 30 arcsec) scale of interest (PRISM<sub>800</sub>-Obs) for daily  $T_{\text{MAX}}$  and  $T_{\text{MIN}}$ , providing a first-order look at model performance.

### 4.2. Potential Error Sources

In order to better interpret the PRISM-Observations test, we investigated the likely sources of error present in the comparison: (1) observational methods, (2) model scale, and (3) bias in the model driven by topography and mountain meteorology given test site and source data locations.

#### 4.2.1. Accuracy of Deployed Sensors

For long-term testing/evaluation purposes of the observational method used in this study, an identical iButton/shield combination was placed on a scientific-grade weather station located at 2600 m elevation in the center of the study watershed. The baseline instruments included a Campbell Scientific CR3000 data logger, an HMP-60 temperature and relative humidity probe, and a Type-T thermocouple installed in

**Table 2.** Topographic Variables and Relationships to PRISM Error<sup>a</sup>

Variable	Description	$T_{MAX}$ Significance		$T_{MIN}$ Significance		Reference
		$r^2$	$p$	$r^2$	$p$	
Elev	Elevation of 10 m DEM	0.170	0.294	<b>0.872</b>	<b>0.001</b>	
Slope	Slope steepness	0.401	0.042	0.121	0.452	
Asp	Aspect, degrees	0.026	0.837	0.283	0.107	
TPI	Topographic position index	<b>0.535</b>	<b>0.002</b>	0.106	0.490	[Guisan et al., 1999]
TRI	Terrain ruggedness index	0.418	0.034	0.119	0.433	[Riley et al., 1999]
DAH	Diurnal anisotropic heating index	<b>0.783</b>	<b>0.001</b>	0.021	0.871	[Böhner and Selige, 2006; Böhner and Antonić, 2009]
StdH	Standardized height	0.168	0.300	<b>0.491</b>	<b>0.014</b>	
SlopH	Slope height	0.170	0.299	0.349	0.055	
NormH	Normalized height	0.303	0.103	0.192	0.230	
MSP	Midslope position	0.333	0.074	0.280	0.114	
VllyD	Valley depth	<b>0.522</b>	<b>0.010</b>	0.014	0.910	
TRASP	Topographic radiative aspect index	<b>0.818</b>	<b>0.001</b>	0.039	0.754	[Roberts and Cooper, 1989]

<sup>a</sup>Variables of primary (yellow) and secondary (green) significance are highlighted.

identical shielding and configured to Western Regional Climate Center and World Meteorological Organization specifications [World Meteorological Organization, 2008].

#### 4.2.2. Model Scaling Test

Because most of the sites are located on steep slopes containing rapid changes in elevation over short distances (Slope; Table 1), we tested the influence of model scale on error. Estimates of daily  $T_{MAX}$  and  $T_{MIN}$  (in °C) were obtained from PRISM at three additional spatial scales. Besides the native 30 arcsec grid (~800 m; PRISM<sub>800</sub>), an upscaled 2.5 arcmin (~4 km; PRISM<sub>4k</sub>) grid was used, as well as point-interpolated values (PRISM<sub>POINT</sub>) that used a two-dimensional spatial weighting of the surrounding PRISM<sub>800</sub> grid cells. In addition, the 30 arcsec PRISM model was downscaled to 3 arcsec (~80 m; PRISM<sub>80</sub>) for the study area. This downscaling was performed at each grid cell by: (1) calculating a local lapse rate between elevation and the 30 arcsec gridded values and (2) applying this lapse rate to a 3 arcsec digital elevation model (DEM).

#### 4.2.3. Topographic Relationships to Error

The PRISM temperature interpolation algorithm relied heavily on atmosphere-topographic relationships to make spatial predictions. We therefore explored the possibility that topographically driven meteorology was a factor in differences between PRISM and the observations. Several topographic characteristics for each site (Table 2) [Guisan et al., 1999; Riley et al., 1999; Böhner and Antonić, 2009] were computed separately using 1/3 arcsec (~10 m) DEMs with open-source Quantum Geographic Information System (GIS) software [Quantum Geographic Information System Development Team, 2015] running System for Automated Geoscientific Analysis processing algorithms [Böhner and Selige, 2006]. Elev is the elevation derived from the DEM for each site point location. Slope is the steepness of the land surface from horizontal in degrees of angle. Asp is the cardinal aspect of the site point location in degrees from true north. TPI is the topographic position index, a measure of local prominence with a default bandwidth setting of 75 m. TRI is the terrain ruggedness index, a measure of amount of elevation difference between adjacent grid cells. DAH, or diurnal anisotropic heating, is an index which represents incident radiative energy exposure. StdH is a measure of relative slope position within the catchment area, taking into account drainage minimums and summit heights. SlopeH is a measure of the total homogenous slope height associated with each point. NormH is the normalized altitude of the terrain, stretched between the summit and lowest point in the watershed. MSP, or midslope position, represents the fractional vertical position of each site on its associated slope feature. VllyD is the vertical distance to the nearest channel network base level, i.e., a local-scale valley/drainage depth from each site point location. In addition, a topographic radiative aspect index (TRASP) [Roberts and Cooper, 1989] was derived from aspect values using the formula:

$$TRASP = 0.5 - \frac{\cos\left[\left(\frac{\pi}{180}\right)(Asp - 30)\right]}{2},$$

where TRASP is a daily radiation loading index between 0 and 1 with zero being the coolest and one being the warmest slopes, and where Asp is the cardinal aspect of the slope in degrees east of north.

In an analysis similar to the topoclimatic work in *Bunn et al.* [2011], we investigated the influence of these nonindependent spatio-topographic variables on model departures from observations. A complimentary two-step process using independent cluster and ordination analyses was performed on the daily PRISM<sub>800</sub> bias time series. First, hierarchical clustering with the Ward2 minimal-variance implementation in multidimensional Euclidean space [Murtagh and Legendre, 2014] was used to group sites by similar error characteristics. Two distinct cluster groups existed for both daily  $T_{MAX}$  and  $T_{MIN}$  errors (each with different site subsets) based on evaluation of silhouette graphs generated using the same Euclidean distance calculations [Rousseeuw, 1987]. In the second step, we plotted each site's  $T_{MAX}$  and  $T_{MIN}$  error characteristics in two-dimensional ordination space by using Non-Metric Multidimensional Scaling (NMDS) [Oksanen et al., 2015].

NMDS is an ordination method which allows comparison of pairwise dissimilarities between variables in low-dimensional space using rank-based correlation (as opposed to linear correlation methods used in principal components analysis or principal coordinates analysis). It accommodates variables with unknown distributions and removes unit distances (losing the absolute magnitude of ordination distance but retaining relative positions of variables). In this case, topographic variables that are physiographically related but associated with different atmospheric mechanisms can still be evaluated separately and grouped for relative influence on PRISM departures in the same dimensional space. This aids in interpretation of topoclimatic mechanisms which may influence PRISM error, such as insolation or large-scale air convergence. NMDS is an iterative algorithm which requires successive ordinations (beginning with a randomized placement within  $n$  dimensions specified) to be compared to actual pairwise dissimilarities until the difference between the two ("stress") is minimized. Stress values of 0.1 or below are considered fair ordination fits.

Vectors of the topographic variables were fit to the  $T_{MAX}$  and  $T_{MIN}$  error ordinations [Oksanen, 2015], and significance assessed using permutation ( $n = 999$ ; Table 2) within the package *vegan* [Oksanen et al., 2015] in the open-source R software [R Development Core Team, 2015].

### 4.3. Impacts to Commonly Derived Hydrometeorologic Variables

In order to evaluate effects of model error to potential science questions, three use cases were examined. Variables were targeted that are often used in applied mountain climate impact studies and estimated using models like PRISM [e.g., *Horning et al.*, 2010; *Johnson et al.*, 2010; *West et al.*, 2015; *Copenhaver-Parry and Cannon*, 2016]. Because PRISM is leveraged for climate downscaling as well as plot-level studies, we thought that it is useful to briefly demonstrate three different scenarios of model application in mountain meteorology: (1) watershed-scale point-event prediction (such as first and last freezing days), (2) process modeling (precipitation as snow), and (3) cumulative temperature impacts (degree-day thermal sums).

#### 4.3.1. Frost-Free Season (2014)

We calculated the 2014 frost-free season (FF14) using both the observed data and the model outputs. The length of time between the first and last frost ( $T_{MIN} > 0^{\circ}\text{C}$ ) is considered an important metric of general growing season length for plants, and as a long-term climatic indicator has been increasing over recent decades in the western United States based on observations [Easterling, 2002; Kunkel et al., 2004]. The frost-free season at a given site is tied to synoptic mechanisms [Meehl et al., 2004] as well as local physiographic controls [Jordan and Smith, 1995], while regional trends in frost-free days are linked to global patterns [McCabe et al., 2015].

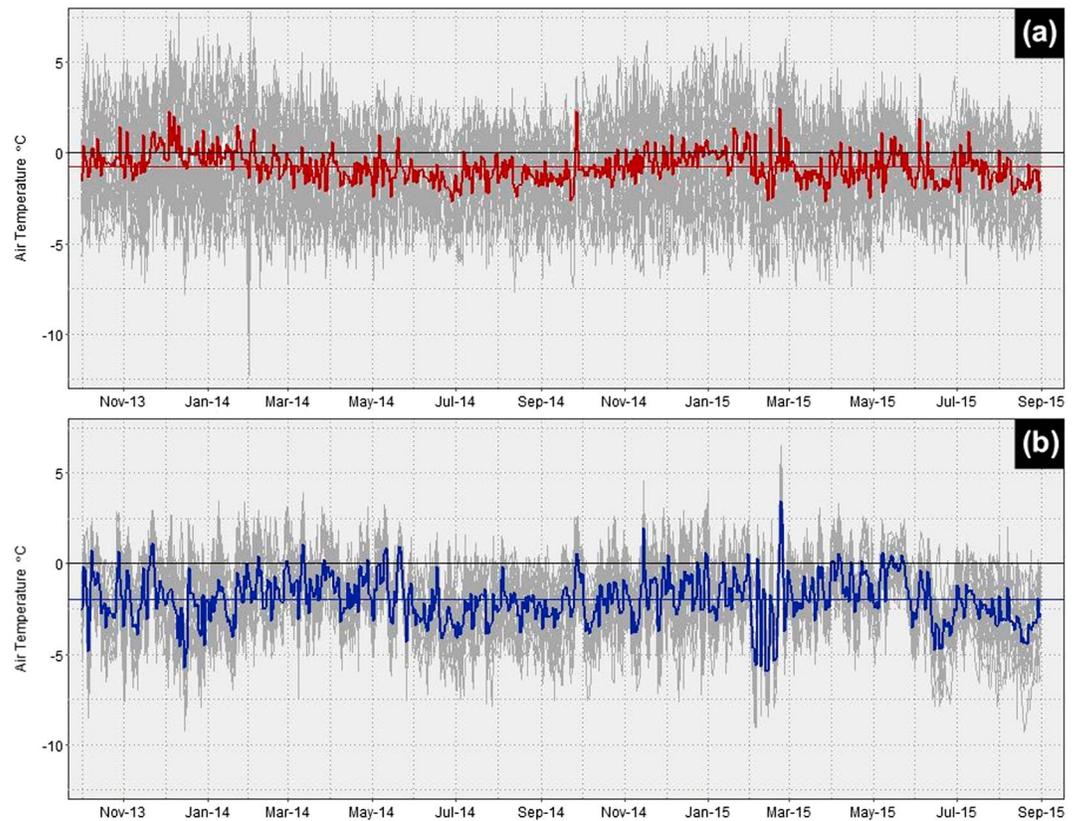
#### 4.3.2. Precipitation as Snow

We estimated precipitation as snow ( $P_{SNOW}$ ) using the following empirical model of frozen/liquid fractions developed in a previous study of high-resolution meteorological data across a mountain watershed [Daly et al., 2007]:

$$P_s = -0.1667T_m + 0.6667, 0 \leq P_s \leq 1,$$

where  $P_s$  is the proportion of daily precipitation as snowfall and  $T_m$  is the daily mean temperature in degrees Celsius. If  $T_m$  is  $-2.5^{\circ}\text{C}$  or less,  $P_s = 1$ , and if  $T_m$  is  $4^{\circ}\text{C}$  or above,  $P_s = 0$ .

Precipitation values from the daily PRISM 30 arcsec precipitation product were obtained for all 16 study sites and then partitioned according to this relationship with air temperature. Although this snow-fraction model was obtained in a different climatological region than the Walker Basin (western Oregon), given that this is strictly a test of relative estimates between observed temperatures and the PRISM model, we felt comfortable using it in this scenario.



**Figure 2.** PRISM 30 arcsec model daily departures from observations. All model departures (PRISM<sub>800</sub>-Observations; gray curves) are plotted for the interval October 2013 to September 2015. (a) Daily  $T_{MAX}$  departures are averaged between all 16 study sites (red curve), indicating seasonal variability and overall slight negative bias (horizontal red line). (b) Daily  $T_{MIN}$  averages for all study sites (blue curve) show stronger negative bias (horizontal blue line) and a less-pronounced seasonal variation in error. Error in daily  $T_{MAX}$  (Figure 2a, gray curves) has a much wider spread across sites than daily  $T_{MIN}$  (Figure 2b, gray curves).

#### 4.3.3. Heat Sums (2014 and 2015)

In a test of cumulative thermal sums, we estimated growing degree days (GDD) for the intervals 1 January to 1 September in both years from observations and model outputs using the following equation:

$$GDD = \left[ \frac{T_{MAX} + T_{MIN}}{2} \right] - T_{BASE},$$

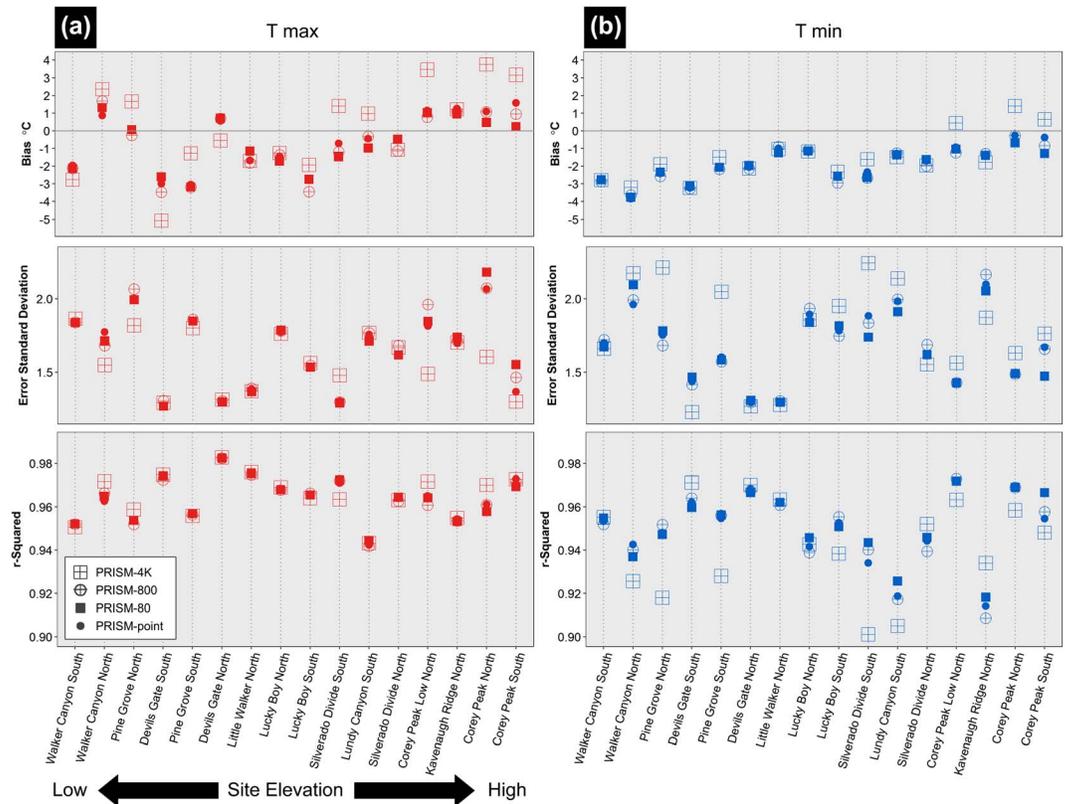
where  $T_{MAX}$  and  $T_{MIN}$  are daily maximum and minimum temperature and  $T_{BASE}$  is the base temperature (which is variable depending on the application). If  $[(T_{MAX} + T_{MIN})/2] < T_{BASE}$ , then  $[(T_{MAX} + T_{MIN})/2] = T_{BASE}$ .

We set  $T_{BASE} = 5^{\circ}C$ , as this represents a commonly used base temperature for climate-ecological studies [Crookston *et al.*, 2010; Thompson *et al.*, 2012, 2015; Bentz *et al.*, 2013], but such thermal sums are also a common feature in mountain hydrology models [Rango and Martinec, 1995; Bergström *et al.*, 2001; Hock, 2003]. Because PRISM does not contain hourly data, the true daily mean temperature was estimated using  $T_{MAX}$  and  $T_{MIN}$ . It is most common to use the mean of the two extremes as in the GDD equation above [McMaster and Wilhem, 1997].

## 5. Test and Analysis Results

### 5.1. PRISM: Observation Comparison

Systematic departures were clearly present between the model and observations, with a seasonal cool bias in PRISM daily  $T_{MAX}$  and consistent cool bias in PRISM daily  $T_{MIN}$  (Figure 2). Across all sites, daily  $T_{MAX}$  was underestimated on average ( $-0.75^{\circ}C$ ; Figure 2a), with mean bias varying by season. Mean absolute error (MAE) for



**Figure 3.** Model scaling test results of (a)  $T_{MAX}$  and (b)  $T_{MIN}$  over the entire 2 year study period. Tests of PRISM scale compared three error statistics (bias, standard deviation, and  $r^2$ ) from the standard 4 km (PRISM<sub>4K</sub>; square grid symbols) and 800 m (PRISM<sub>800</sub>; rounded grid symbols) grid products with point interpolations (PRISM<sub>POINT</sub>; solid dots) and downscaled 80 m grids (PRISM<sub>80</sub>; solid squares) across all 16 sites. The coarse-scale 4 km product has generally greater  $T_{MAX}$  (Figure 3a; red) bias than the fine-scale products, but does not typically diverge from the other scales for  $T_{MIN}$  (Figure 3b; blue) bias values. In most cases across all three error statistics, the finer-scale products display similar behavior, whereas the 4 km product frequently diverges, usually in a direction that denotes less accurate model performance.

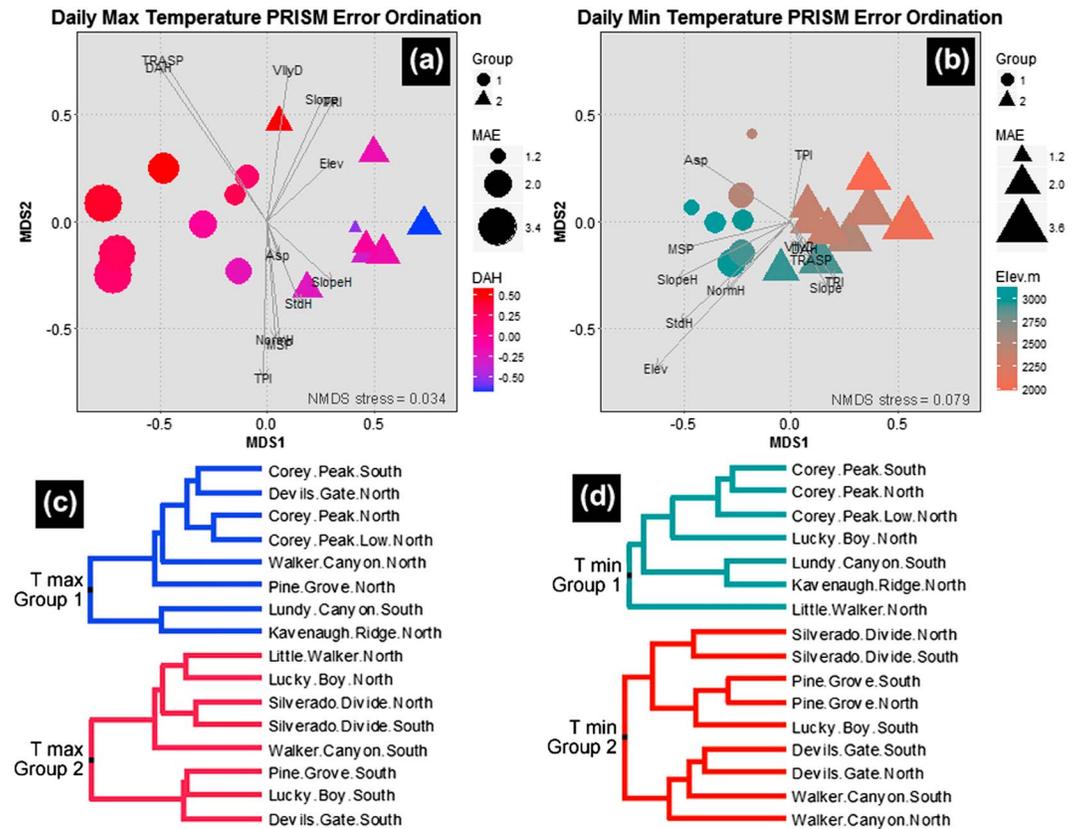
daily  $T_{MAX}$  ranged from 1.15 to 3.47°C, with standard deviations (SD) from 1.30 to 2.07,  $r^2$  values from 0.94 to 0.98 ( $p < 0.0001$ ), and overall bias from  $-3.47$  to  $1.69$ °C (Table 1). For daily  $T_{MIN}$ , negative bias was more pronounced ( $-1.95$ °C; Figure 2b). MAE ranged from 1.14 to 3.70°C, with SD spanning 1.30–2.17,  $r^2$  values from 0.91 to 0.97 ( $p < 0.0001$ ), and overall bias from  $-3.65$  to  $-0.27$ °C (Table 1). While some sites showed similar relative departures in MAE, SD, and  $r^2$  for both diurnal extremes, this was not true for all, indicating that underlying causes of the model departures were likely different for  $T_{MAX}$  as opposed to  $T_{MIN}$ .

## 5.2. Potential Error Sources

### 5.2.1. Accuracy of Deployed Sensors

Testing of the iButton/Gill shield deployment design revealed that this configuration logging at hourly intervals is capable of capturing the same daily  $T_{MAX}$  and  $T_{MIN}$  as the Campbell Scientific equipment. Over a continuous 93 week time frame, bias (calculated as *iButton-Campbell*) of the iButton daily  $T_{MAX}$  compared to the HMP-60 probe and Type-T thermocouple was 0.20°C and 0.45°C (standard deviation  $\sigma = 0.64$  and 0.66), respectively. Bias (*iButton - Campbell*) of the iButton daily  $T_{MIN}$  compared to the HMP-60 probe and Type-T thermocouple was 0.15°C and  $-0.41$ °C ( $\sigma = 0.40$  and 0.39), respectively. It should be pointed out that the two Campbell sensor types themselves diverged, with mean *HMP-60-Thermocouple* being 0.25°C for  $T_{MAX}$  and  $-0.56$ °C for  $T_{MIN}$  ( $\sigma = 0.31$  and 0.10, respectively).

Engineering specifications for Maxim DS1922L iButtons require that all data loggers be calibrated/validated against National Institute of Standards and Technology traceable reference devices, with a temperature accuracy of  $\pm 0.5$ °C when postprocessed with Maxim iButton software [Maxim Integrated, 2016]. Campbell

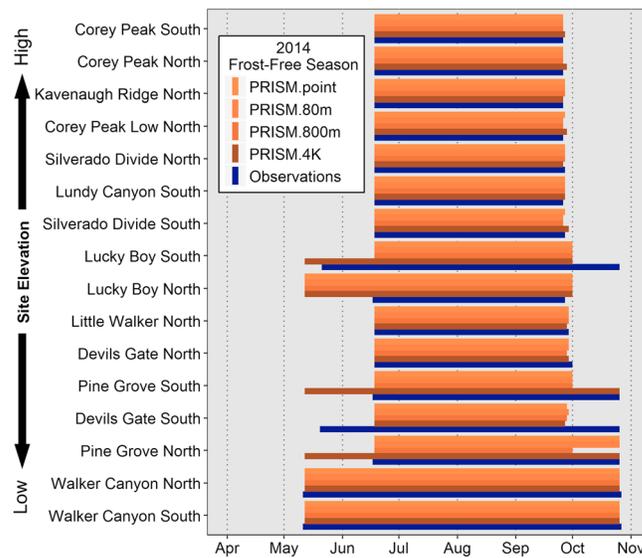


**Figure 4.** Ordination and cluster analysis results. Results from NMDS analysis are plotted for PRISM<sub>800</sub> (a and c)  $T_{MAX}$  and (b and d)  $T_{MIN}$ , indicating which site error characteristics are more related to each other. For  $T_{MAX}$ , cluster groups are partitioned by aspect/radiative exposure (DAH) as well as elevation. For  $T_{MIN}$ , sites are clustered by a combination of elevation and geographic location (not a test parameter). Ordination of PRISM error on each site from NMDS is plotted in two dimensions for  $T_{MAX}$  (Figure 4a) and  $T_{MIN}$  (Figure 4b), indicating site relationships to topographic variable vectors (arrows) and significance (arrow length), as well as site MAE (symbol size), cluster group (symbol type), and most significant variable (symbol color). Clearly, the topographic characteristics of the sites are not independent, and gradients of MAE and clusters do not align directly with any one variable. However, this serves as a visualization tool to aid interpretation of cluster partitioning as well as significance of DAH and elevation for  $T_{MAX}$  and  $T_{MIN}$ , respectively. The cluster trees (Figures 4c and 4d) show group breakdown by site, with more closely related PRISM-Obs error being plotted as neighbors. Comparison with the site map (Figure 1) aids interpretation of the results and appreciation for interacting factors. For instance, for  $T_{MAX}$  Group 1, we see that Lundy and Kavenaugh are most related, even though they are very different in DAH (Table 1); this is probably related instead to their spatial proximity to a SNOTEL (model source data) site. We observe similar spatial clustering for the Corey sites ( $T_{MAX}$  Group 1) and the Silverado pair ( $T_{MAX}$  Group 2); however, the rest of the cluster neighbors seem to fall out by aspect (north v. south). For  $T_{MIN}$ , we see that the groups are separated by a combination of spatial location and elevation (Group 1 is roughly all of the sites in the south/east/high portion of the study area, whereas Group 2 is the north/west/low set of sites).

Scientific specifications state that HMP60 probes within calibration are accurate to  $\pm 0.6^{\circ}\text{C}$  [Campbell Scientific, 2016]. With bias and standard deviation of the paired measurement differences equaling less than  $1^{\circ}\text{C}$  in all cases, we can safely state that the measurement methods using the iButtons for daily  $T_{MAX}$  and  $T_{MIN}$  are functionally equivalent to the research grade sensors in a noncanopied environment.

**5.2.2. Model Scaling Test**

Estimates of daily  $T_{MAX}$  and  $T_{MIN}$  were compared using the four different model scales (PRISM<sub>4km</sub>, PRISM<sub>800m</sub>, PRISM<sub>POINT</sub>, and PRISM<sub>80</sub>) over the entire 2 year study period. No systematic shifts were observed in model bias, error standard deviations, or correlation with observations by changing model scale below 800 m; however, we did observe greater departures in accuracy for many of the 4 km scale comparisons (Figure 3). This indicated that model departures from observations on aggregated time frames were not a matter of scale below the 800 m level even in very steep topography, but that application of the model at the 4 km



**Figure 5.** Frost-free 2014 season. Four spatial scales of PRISM daily data are compared to observations when calculating the frost-free season. Sites are sorted from low elevation (*bottom*) to high elevation (*top*). The PRISM temperature model performance in this test is quite good, with some disagreement between various scales of the model and the observations. Notably, PRISM<sub>4K</sub> missed the last springtime freeze on three sites when the finer-scale models did not.

( $r^2 = 0.818, p = 0.001$ ), which are closely related surface heat-loading factors. VlyD ( $r^2 = 0.522, p = 0.010$ ) and TPI ( $r^2 = 0.535, p = 0.002$ ) were significant factors for  $T_{MAX}$  as well. Variables associated with error characteristics for  $T_{MIN}$  were Elev ( $r^2 = 0.872, p = 0.001$ ) and *StdH* ( $r^2 = 0.491, p = 0.014$ ).

The error ordinations, significance of topographic factors, and cluster groups were visualized in two-dimensional plots (Figure 4). These results indicated that the mechanisms associated with PRISM departures were tied to real processes associated with geographic location and topography. The observation that significant topographic factors are not directly aligned with the ordination axes tells us that multiple (and probably interacting) processes are at work behind the nature of daily model bias.

### 5.3. Impacts to Commonly Derived Hydrometeorological Variables

#### 5.3.1. Frost-Free Season (2014)

PRISM<sub>800</sub> correctly approximated the last and first frosts of *FF14* at 11 of 16 sites. PRISM<sub>4K</sub> performed similarly but missed the last springtime freeze on three of the sites that the finer-scale models captured while correctly noting the first freeze on one site that the other scales did not (Figure 5). In four cases, PRISM<sub>800</sub> underestimated the length of *FF14*, and in one case the season length was overestimated. Underestimates were primarily on south facing slopes, and the overestimate was on a north facing slope. In only one case did the PRISM<sub>80</sub> and PRISM<sub>POINT</sub> downscaled data improve the *FF14* estimate. Differences among PRISM scales were limited to the lower elevation sites.

#### 5.3.2. Precipitation as Snow

PRISM's estimates of  $P_{SNOW}$  were similar to observation-driven estimates across most of the sites (Figure 6). However, we found that PRISM<sub>4K</sub> significantly underestimated  $P_{SNOW}$  at four upper sites and noticeably overestimated  $P_{SNOW}$  at three lower sites. Differences between the three finer-scale PRISM products were minimal and typically very close to  $P_{SNOW}$  estimates made using the observed temperature.

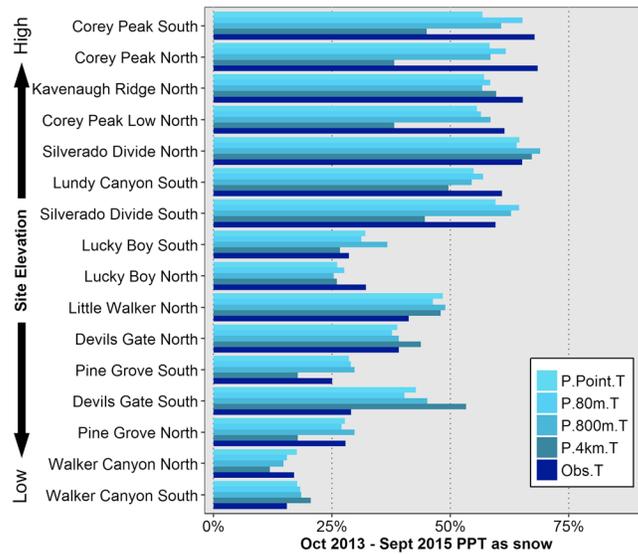
#### 5.3.3. Heat Sums (2014 and 2015)

As expected given the consistent cool bias in both  $T_{MAX}$  and  $T_{MIN}$ , PRISM generally underestimated GDD at all sites in both years (Figure 7). Once again, PRISM<sub>4K</sub> behaved differently than the finer model scales, in the cases of upper elevation significantly overestimating the thermal sums in both years. Most of the sites experienced between 200 and 400 more degree days in each year than estimated by the finer resolution

resolution could yield different results. Patterns in these differences also indicated that separate error mechanisms were in play for  $T_{MIN}$  as opposed to  $T_{MAX}$ , suggesting that local land-atmosphere interaction becomes increasingly important for modeling temperature in mountains below 4 km spatial resolution.

#### 5.2.3. Topographic Relationships to Error

Results from NMDS analysis of model error and associated topographic variables were as follows:  $T_{MAX}$  nonmetric  $R^2 = 0.999$ , stress = 0.034;  $T_{MIN}$  nonmetric  $R^2 = 0.994$ , stress = 0.079; indicating strong association with one or more topographic features for  $T_{MAX}$  and a fair association with one or more features for  $T_{MIN}$ . Variables of significance (measured using the squared correlation coefficient for the fit of the topographic vector and its scores in the ordination) for  $T_{MAX}$  were the indices DAH ( $r^2 = 0.783, p = 0.001$ ) and TRASP



**Figure 6.** Estimated precipitation as snow. Four spatial scales of PRISM daily data are compared to observations when calculating percent of precipitation as snow during the 2 year study period. Sites are sorted from low elevation (bottom) to high elevation (top). The PRISM temperature model performance in this test is generally excellent, with the exception of the PRISM<sub>4K</sub> product. It is evident that scale plays a role in application of modeled air temperature when estimating hydrometeorological parameters such as rain/snow boundaries. For example, the use of finer-scale PRISM estimates are likely to produce more accurate results when downscaling climate projections to mountainous landscapes.

PRISM products, with only the highest elevation sites showing minimal bias in these cases. These differences between scales and elevations could be significant for applications of the data in climate impacts modeling scenarios.

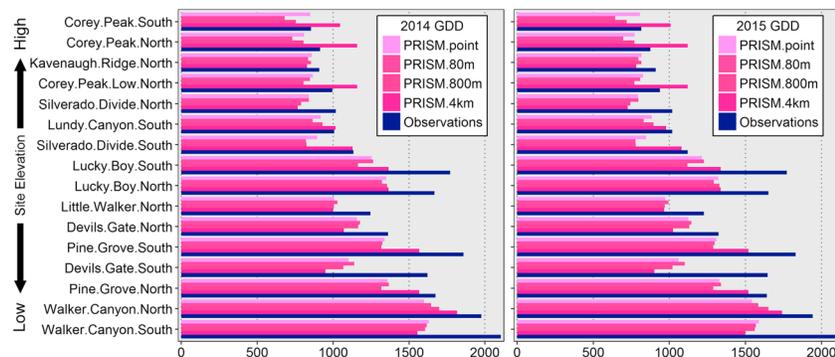
### 6. Discussion

Consistently high  $r^2$  values at all sites between observations and PRISM temperatures demonstrate that the model approximates daily temperature variations in open woodland environments well and that the response of the model to changing atmospheric conditions at the watershed scale is quite reasonable. Behavior of error statistics and derived products from the upscaled PRISM<sub>4K</sub> data set proved different than subkilometer-scaled PRISM outputs, indicating the importance of scale when estimating atmospheric conditions in complex terrain. The first-order error mode for PRISM<sub>800</sub> on these study sites appears to be a

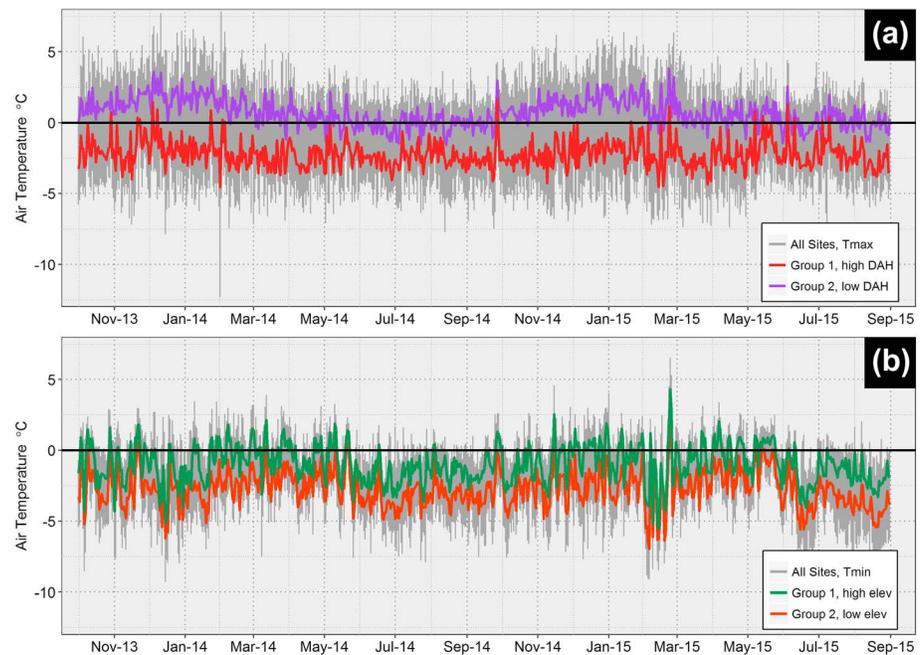
cool bias that underestimates daily  $T_{MAX}$  at 10 of 16 sites and  $T_{MIN}$  at all 16. Surprisingly, reducing PRISM<sub>800</sub> model scale by an order of magnitude had no consistent effect on results. Differences in error seasonality and relationships to site topography instead suggest mechanistic sources of model departures, tied to landscape/atmosphere interactions and how these are represented in source data or model processes.

#### 6.1. Topographic Mechanisms of Error

Diurnal radiation loading is associated strongly with modes of model error in estimating true daily  $T_{MAX}$  on open woodland mountain slopes, followed secondarily by relative topographic position (Table 2). While it



**Figure 7.** Growing degree days. Four spatial scales of PRISM daily data as well as observations are used to estimate thermal sums on the study sites. In this case, growing degree days (base = 5°C) are shown for January–September seasons in years 2014 and 2015. GDD5 or thermal sums of similar types are often incorporated in ecological prediction and snowmelt models tied to climate. Sites are sorted from low (bottom) to high (top) elevation. Finer scales of PRISM consistently underestimated GDD5 across the watershed, especially at lower and warmer elevations, while the PRISM<sub>4K</sub> product did not show the same consistency in bias.

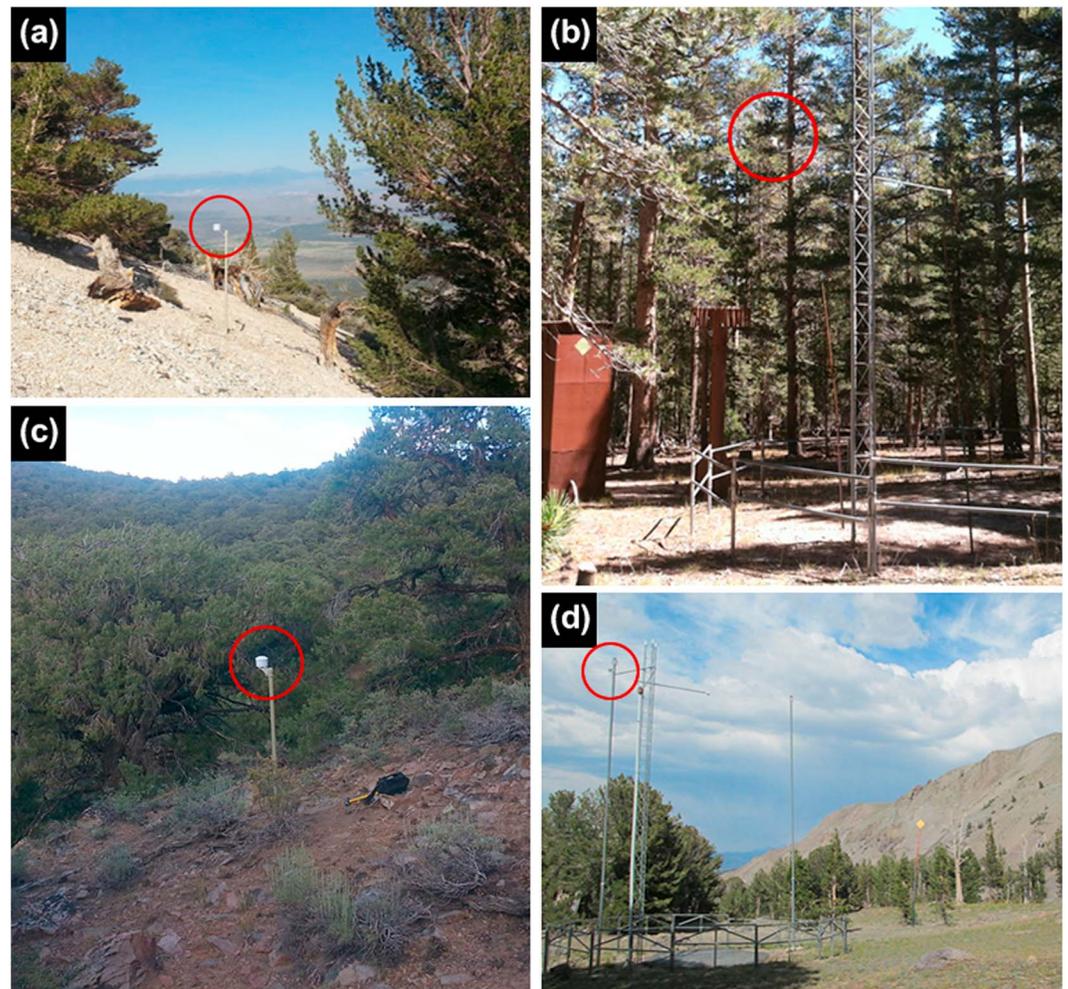


**Figure 8.** Error plots with cluster means. PRISM<sub>800</sub> daily errors (PRISM-Obs) are shown for all 16 sites (*gray*) with means for cluster groups (*colored lines*). Behavior of these clusters is visually quite different in terms of absolute and seasonal bias. (a) This result is strongly aligned with the association of radiative loading indices with  $T_{MAX}$  error (seasons with lower Sun angle result in greater relative departure, including wintertime overestimation of  $T_{MAX}$  at sites with less incident radiation). (b) Daily  $T_{MIN}$  error cluster groups do not show terribly different behavior and are differentiated primarily by a mean bias offset. This result would agree with a spatial or elevational bias in the model given the locations and types of source data stations in the watershed.

is also possible that this specific error is being confounded by incorrect assumptions of local lapse rates, the ordination results clearly point to a strong contribution of hillslope heat-loading error vectors as opposed to strictly elevational ones (Figure 4a). When we plot the mean PRISM<sub>800</sub> bias for the separate site clusters, we can immediately see that  $T_{MAX}$  for Group 1 was consistently underestimated by PRISM (Figure 8a), whereas Group 2 was overestimated in winter. These groups are primarily split between sites with high (Group 1) and low (Group 2) radiative loading (DAH; Table 1). The difference in how PRISM treats these two categories of sites approached 5°C at times during the low-Sun (winter) season, when slope and aspect had the greatest effect on contrasting solar exposures. The relationship between daily temperature extremes and incident solar radiation is well known [Bristow and Campbell, 1984; Thornton and Running, 1999], and the interactions of radiation with geography, land cover, slope, and aspect have also been explored in the context of complex terrain [McCutchan and Fox, 1986; Bolstad et al., 1998; McCune and Keon, 2002; Bennie et al., 2008]. These relationships have been applied in gridded models with local calibration data obtained at relatively high spatial resolution [e.g., Daly et al., 2007; Holden et al., 2015].

In the case of  $T_{MIN}$  departures (Figure 8b), there is no consistent seasonal or clustering pattern; rather, a systematic overprediction of drops in nighttime temperature occurred across all sites regardless of cluster group or season. That is, observed overnight lows were warmer across all instrumented sites relative to PRISM estimates. Because the most persistent negative bias in  $T_{MIN}$  was actually during summer, we do not suspect snow-on climatology bias in the model during these low-snow years, although snow can be a factor in lowering nighttime temperatures [Pepin et al., 2011]. There were no significant topographic factors associated with these errors besides relative elevation in the watershed (Table 2). This suggests a systematic bias in PRISM's station suite compared to the study sites.

Most of PRISM's source stations are located in flat terrain or valley bottoms (Figure 1), which are conducive to local cold air pooling and atmospheric decoupling, resulting in relatively low minimum temperatures. In contrast, the steeply sloping study locations experience much less cold air pooling and are likely to be more closely coupled to the free atmosphere (Figures 1 and 9). Valley-bottom decoupling from the upper air during



**Figure 9.** Examples of highly contrasting environments in both PRISM source data and test observations. Locations of temperature sensors are circled in red. (a) The Silverado South study site has sparse vegetation and high surface albedo. (b) The Virginia Lakes Ridge SNOTEL station is surrounded by tall vegetation. (c) The Pine Grove North study site is located in a medium-density woodland with neutral surface albedo. (d) The Leavitt Lake SNOTEL station is in a low-density woodland with surface albedo moderated by grasses. The height of the sensors above ground also differs, with SNOTEL sensors being located at times significantly higher than 2 m (Figure 9b photo source: USDA Natural Resources Conservation Service (NRCS), September 2014).

stable nighttime conditions is a notable (if understudied) feature of Great Basin topography, and cold-air pools at large scales have been documented as regularly occurring events in the region [Billings, 1954; Wells and Shields, 1964; Wells, 1983]. While PRISM's station-weighting functions account for topographic position for this very reason, the lack of midslope stations to draw from compromised its ability to replicate elevated  $T_{\text{MIN}}$  values at the study sites. Thus, PRISM's success in dealing with cold-air pooling and stable-air nighttime lapse rates is largely dependent on the availability of stations in a variety of topographic positions. Cold-air pooling, in general, remains a challenge for temperature modeling in mountain environments, given that the frequency and depth of pools is dependent on regional climatology, local airflow responses, and seasonal surface-atmospheric energy exchanges of individual watersheds [e.g., Whiteman, 1982; Bell and Bosart, 1988; Lundquist et al., 2008; Daly et al., 2010; Lareau and Horel, 2014; Holden et al., 2015].

## 6.2. Biases Related to Hardware and Station Micrositing

Both the PRISM source data and the test observations in this study are subject to a certain amount of bias forced by hardware (primarily radiative shielding of the sensor) and sensor/station microsite characteristics

(vegetative cover, snow presence/absence, and local surface albedo). While the test sensors were deployed in a uniform manner, the contributing network stations to PRISM consist of a wider variety of hardware and microsite factors. COOP stations in valley locations typically use the highly effective passively aspirated Stevenson screen sensor shelter [MacHattie, 1965], whereas the RAWS and SNOTEL networks use Gill-type passively aspirated plate shields similar to the test installations. COOP and RAWS stations generally have the sensors located at approximately 2 m in height above ground level (as do the test observations), whereas SNOTEL sensors can be much higher (Figure 9).

Measurements of  $T_{MAX}$  are primarily biased by shortwave radiation either entering the shield housing or being transferred to the air inside by the shield material itself. The influence of shortwave radiation as a bias source is directly related to the exchange between ambient air and the air inside the shield (i.e., wind speed [Richardson *et al.*, 1999]), as well as surrounding surface albedo [Huwald *et al.*, 2009]. The albedo of typical geology in the watershed (granites and hydrothermally altered rocks/soils) is high (Figure 9), and this study was conducted during one of the lowest-snow periods recorded for the region, which would minimize snowpack-driven wintertime increase in surface albedo, especially on south aspects. Increased albedo of north facing or heavily vegetated surfaces due to snow retention in winter would be offset by reduced incident radiation angle and increased shading. Because of these factors, we expect that seasonal biases in both source and test data in  $T_{MAX}$  due to snow presence are minimized during the study period. The strongest source of observation bias under these conditions is more likely to be the typical airflow across and through the shields during the warmest portion of the day. Surrounding vegetation significantly reduces average wind speeds at the sensor housing, with the consequence that Gill-type shields can experience  $T_{MAX}$  warm biases up to several degrees when wind speed is near zero and the sensor housing is not shaded [Nakamura and Mahrt, 2005; Huwald *et al.*, 2009].

Measurements of  $T_{MIN}$  are more influenced by microsite position and cool airflow during nighttime hours, rather than hardware factors. The presence of snow can certainly impact local air temperatures [Pepin *et al.*, 2011]; however, the abnormally low snow amounts during the study period would have minimized the duration of measurement bias during the snow season at these locations.

The question remains: how do these interacting factors impact source data bias versus test observation bias, and how do these conclusions relate to interpretation of PRISM model performance? A primary source of midelevation data for PRISM in the western U.S. is the SNOTEL network, a series of sites maintained for seasonal streamflow prediction that also includes basic meteorological sensors [Natural Resources Conservation Service, 2015]. The Walker Basin is no exception—there are eight SNOTEL stations within or near the watershed, and these dominate the local middle and upper elevation data contributions to the model. The topoclimatic siting characteristics of these stations therefore become an important factor in the performance of PRISM and other gridded models in the region. Because the SNOTEL network is designed to measure snow hydrology variables near the headwaters of major streams and rivers, the stations are frequently associated with upper elevation montane forests. They are situated such that they are not thermally representative of woodland slopes, but rather flats or even sink zones. Accordingly, it is likely that the SNOTEL stations in and around the Walker Basin experience air conditions with canopy-influenced reduction of radiation and increased cooling as well as more stable, frequent flow of cool air during nighttime conditions (Figure 9). These features make the SNOTEL sites a prime suspect in the PRISM departures in this study, due to local siting characteristics rather than systematic instrumental error.

Instruments and topographic position being equal, the siting characteristics of SNOTEL sensors would, in general, result in cooler  $T_{MAX}$  values than the test observations during times of low wind speed simply due to local shading and reduced local albedo. Elevated, more exposed sites in this region (such as our test locations) often experience persistent wind that would reduce  $T_{MAX}$  bias in the test data. For example, the mean wind velocity at the Rockland climate station (midelevation, in the center of the watershed, and open woodland ridge) during the study period was  $5.3 \text{ m s}^{-1}$ . Because of elevated wind speeds, it is possible that many of the testing sites would not experience significantly more  $T_{MAX}$  bias than forested SNOTEL sites, in spite of generally increased radiation factors. The differences in SNOTEL sensor height, however, could offset this on a case-by-case basis, depending on air turbulence at the sensor housing [Nakamura and Mahrt, 2005].

It has recently been pointed out that changes in SNOTEL temperature sensing hardware and calibration/processing practices could be introducing biases into the network [Oyler *et al.*, 2015]. Reported

shifts in temperature have been as much as  $\sim 1.1^\circ\text{C}$ , primarily on the warm side, but varying with temperature. At least some SNOTEL stations in the Walker were affected by this bias during the study period. A preliminary correction ( $\text{PRISM}_{800\text{corr}}$ ) using an empirically derived ninth-order function developed in collaboration with the National Park Service in Alaska (Daly et al., unpublished data) did not make significant improvements to the model performance in our study (mean adjusted  $\text{PRISM}_{800} T_{\text{MAX}}$  bias =  $-0.58^\circ\text{C}$  compared to  $0.75^\circ\text{C}$ , mean adjusted  $\text{PRISM}_{800} T_{\text{MIN}}$  bias =  $-2.67^\circ\text{C}$ , compared to  $-1.95^\circ\text{C}$ ). Thus, corrections to the model to compensate for SNOTEL bias appear to be overshadowed by other bias sources. We therefore interpret the primary mode of error in the overall test of the  $\text{PRISM}_{800}$  model to be a function of micro-siting differences between the PRISM source data and the study sites, and secondarily the larger topographic settings.

### 6.3. Applications and Use of PRISM in Complex Terrain

Our tests of the PRISM temperature product at different scales describe both strengths and weaknesses of gridded climate interpolation models in mountainous terrain, particularly on the proportionally large areas that are composed of elevated slopes. Our work also highlights several sometimes confounding sources of measurement bias, which vary across data networks. These regional and network-specific biases create challenges for both the model developer and the end user. We feel that this study's results are likely transferable across large portions of the Intermountain West where vegetation is limited to shrub and woodland cover on steep topography. The reasoning for this is tied to typical spatio-topographic distribution of source data for the model as well as the land cover types. Studies using PRISM (or other gridded products) directly or as part of downscaling efforts applied to temperature-related questions should examine the scale of application and cumulative effects of potential bias in both regional source data and representative topography. For example, studies estimating frost-free periods on steep mountain slopes may well be quite accurate in their conclusions, as first and last frost days in the intermountain west are most likely tied to transitional-season synoptic events which are interpolated well by PRISM regardless of source data bias. Predictive models which seek to partition precipitation or build/melt snowpack may be subject to greater error if coarse-scale grid products are used (Figures 6 and 7). If thermal sum calculations (such as *GDD5*) are used in snowmelt or climate impacts models, PRISM could introduce significant error (because of the cumulative nature of the statistic) if study sites do not share the topographic characteristics of PRISM's source data (Figure 7). Thus, researchers focused on processes which are sensitive to small changes in atmospheric temperature may discover that in order to arrive at accurate conclusions, local observations over full seasonal cycles for model calibration purposes would be wise, especially if model source data in the region are not topographically representative. This is becoming a more and more common practice, although the method of measurement, micro-siting factors, and radiative shielding practices introduce their own biases and must be understood when making comparisons with gridded data sets.

## 7. Conclusions for Climate Science in Mountain Regions

Our results indicate that PRISM temperature is reasonably representative of conditions on semiarid mountain slopes, but that absolute bias caused by a combination of factors can be significant and should be taken into account when applying PRISM data to science questions in similar settings or downscaling climate models over complex terrain. These findings highlight the importance of topoclimatic siting for near-surface observations in mountain science and how data processing (e.g., SNOTEL calibration) or prediction experiments (e.g., warming) that shift regional temperatures by a degree or two can be completely overshadowed or negated within interpolation models by source data bias at locally relevant scales.

We observed a significant improvement in model performance as grid resolution was increased from  $\sim 4$  km to  $\sim 1$  km, but no relative improvements were made by further downscaling to  $\sim 80$  m or using point interpolation. This suggests that at scales of 1 km or less, factors other than elevation, such as slope, aspect, and vegetation cover (which drive local airflow and radiative effects), have greater significance for site-level daily temperatures on mountain slopes. Therefore, as gridded models mature and incorporate diverse sources of data in pursuit of greater resolution and accuracy, it falls to the individual researcher to ensure that the application and use of modeled climate data are in fact appropriate to the science in question given model assumptions.

The relative amount of geographic area in mountain watersheds encompassed by elevated slopes is often quite large compared to ridgetop or canyon bottoms [Strahler, 1952], representing opportunity for significant spatial propagation of error if model assumptions are incorrect where slopes are concerned. The atmospheric

and mountain science communities need to make concerted efforts to rectify the current lack of representative data across the gradients of elevation, aspect, and vegetative cover. Better characterization of local station biases and increasing awareness of these in scientific application are also critical needs.

As future studies conduct in situ tests, the need for expansion of sensory networks that better capture a range of climate variables across diverse topographies will become abundantly clear. Conducting similar tests across a greater diversity of landscapes and topographic exposures both here in the Walker Basin and elsewhere in the semiarid western U.S., with the express intent of observations replicating “standard” 2 m weather station diurnal measurements that are used in interpolated products, is a lofty but obtainable goal in pursuit of improving the near-surface atmospheric modeling process and scientific applicability.

### Acknowledgments

The authors wish to thank the following collaborators who have contributed to field and data operations supporting this climate work in the Walker Basin: Constance Millar from the USDA Forest Service; Graham Kent from the Nevada Seismological Laboratory; Sergiu Dascalu and Fredrick Harris from the Nevada Research Data Center at the University of Nevada, Reno; Gregory McCurdy from the Western Regional Climate Center; Matthew Doggett with the PRISM Climate Group; the NRCS SNOTEL program; and Jehren Boehm from the Department of Geography at the University of Nevada. We also acknowledge the valuable inputs of Jessica Lundquist at the University of Washington and three anonymous reviewers who helped us refine our initial manuscript. At this time, the data used in this study may be obtained from the authors. Future archival will be at the Nevada Research Data Center (<https://sensor.nevada.edu>). Scotty Strachan was supported in part by National Science Foundation grants GSS-1230329 and EPSCoR IIA-1301726, as well as the University of Nevada, Reno, College of Science Dean's Office.

### References

- Bell, G. D., and L. F. Bosart (1988), Appalachian cold-air damming, *Mon. Weather Rev.*, *116*, 137–161, doi:10.1175/1520-0493(1988)116<0137:ACAD>2.0.CO;2.
- Bennie, J., B. Huntley, A. Wiltshire, M. O. Hill, and R. Baxter (2008), Slope, aspect and climate: Spatially explicit and implicit models of topographic microclimate in chalk grassland, *Ecol. Modell.*, *216*(1), 47–59, doi:10.1016/j.ecolmodel.2008.04.010.
- Bentz, B., J. Vandygriff, C. Jensen, T. Coleman, P. Maloney, S. Smith, A. Grady, and G. Schen-langenheim (2013), Mountain pine beetle voltinism and life history characteristics across latitudinal and elevational gradients in the western United States, *For. Sci.*, *60*(3), 434–449, doi:10.5849/forsci.13-056.
- Bergström, S., B. Carlsson, M. Gardelin, G. Lindström, A. Pettersson, and M. Rummukainen (2001), Climate change impacts on runoff in Sweden—Assessments by global climate models, dynamical downscaling and hydrological modelling, *Clim. Res.*, *16*(2), 101–112, doi:10.3354/cr016101.
- Billings, W. (1954), Temperature inversions in the pinyon-juniper zone of a Nevada mountain range, *Butler Univ. Botanical Stud.*, *11*(13), 112–118.
- Böhner, J., and O. Antonić (2009), Land-surface parameters specific to topo-climatology, *Dev. Soil Sci.*, *33*, 195–226, doi:10.1016/S0166-2481(08)00008-1.
- Böhner, J., and T. Selige (2006), Spatial prediction of soil attributes using terrain analysis and climate regionalisation, in *SAGA—Analyses and Modelling Applications: Göttinger Aeographische Abhandlungen*, vol. 115, edited by J. Böhner, K. R. McCloy, and J. Strobl, pp. 13–28, Verlag Erich Goltze GmbH, Göttingen, Germany.
- Bolstad, P. V., L. Swift, F. Collins, and J. Regniere (1998), Measured and predicted air temperature at basin to regional scales in the southern Appalachian mountains, *Agric. For. Meteorol.*, *91*, 161–176, doi:10.1016/S0168-1923(98)00076-8.
- Bristow, K. L., and G. S. Campbell (1984), On the relationship between incoming solar radiation and daily maximum and minimum temperature, *Agric. For. Meteorol.*, *31*(2), 159–166, doi:10.1016/0168-1923(84)90017-0.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.
- Bunn, A. G., M. K. Hughes, and M. W. Salzer (2011), Topographically modified tree-ring chronologies as a potential means to improve paleoclimate inference, *Clim. Change*, *105*(3), 627–634, doi:10.1007/s10584-010-0005-5.
- Campbell Scientific (2016), HMP60-L: Temperature and relative humidity probe. [Available at <https://www.campbellsci.com/hmp60> (Accessed 17 February 2016).]
- Copenhaver-Parry, P. E., and E. Cannon (2016), The relative influences of climate and competition on tree growth along montane ecotones in the Rocky Mountains, *Oecologia*, *182*(1), 13–25, doi:10.1007/s00442-016-3565-x.
- Crookston, N. L., G. E. Rehfeldt, G. E. Dixon, and A. R. Weiskittel (2010), Addressing climate change in the forest vegetation simulator to assess impacts on landscape forest dynamics, *For. Ecol. Manage.*, *260*(7), 1198–1211, doi:10.1016/j.foreco.2010.07.013.
- Daly, C. (2006), Guidelines for assessing the suitability of spatial climate data sets, *Int. J. Climatol.*, *26*(6), 707–721, doi:10.1002/joc.1322.
- Daly, C., D. R. Conklin, and M. H. Unsworth (2010), Local atmospheric decoupling in complex topography alters climate change impacts, *Int. J. Climatol.*, *30*(12), 1857–1864, doi:10.1002/joc.2007.
- Daly, C., W. P. Gibson, G. H. Taylor, G. L. Johnson, and P. Pasteris (2002), A knowledge-based approach to the statistical mapping of climate, *Clim. Res.*, *22*(2), 99–113, doi:10.3354/cr022099.
- Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. Pasteris (2008), Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, *Int. J. Climatol.*, *28*, 2031–2064, doi:10.1002/joc.1688.
- Daly, C., E. H. Helmer, and M. Quinones (2003), Mapping the climate of Puerto Rico, Vieques, and Culebra, *Int. J. Climatol.*, *23*(11), 1359–1381, doi:10.1002/joc.937.
- Daly, C., R. P. Neilson, and D. J. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*(2), 140–158, doi:10.1175/1520-0450(1994)033<0140:ASTMFM>3E2.0.CO;2.
- Daly, C., J. I. Smith, and K. V. Olson (2015), Mapping atmospheric moisture climatologies across the conterminous United States, *PLoS One*, *10*(10), doi:10.1371/journal.pone.0141140.
- Daly, C., J. W. Smith, J. I. Smith, and R. B. McKane (2007), High-resolution spatial modeling of daily weather elements for a catchment in the Oregon Cascade Mountains, United States, *J. Appl. Meteorol. Climatol.*, *46*(10), 1565–1586, doi:10.1175/JAM2548.1.
- Daly, C., M. P. Widrechner, M. D. Halbleib, J. I. Smith, and W. P. Gibson (2012), Development of a new USDA plant hardiness zone map for the United States, *J. Appl. Meteorol. Climatol.*, *51*(2), 242–264, doi:10.1175/2010JAMC2536.1.
- Dobrowski, S. Z. (2011), A climatic basis for microrefugia: The influence of terrain on climate, *Global Change Biol.*, *17*(2), 1022–1035, doi:10.1111/j.1365-2486.2010.02263.x.
- Easterling, D. R. (2002), Recent changes in frost days and the frost-free season in the United States, *Bull. Am. Meteorol. Soc.*, *83*(9), 1327–1332, doi:10.1175/1520-0477(2002)083<1327:RCIFDA>2.3.CO;2.
- Feser, F., B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn (2011), Regional climate models add value to global model data: A review and selected examples, *Bull. Am. Meteorol. Soc.*, *92*(9), 1181–1192, doi:10.1175/2011BAMS3061.1.
- Foley, J. A., et al. (2005), Global consequences of land use, *Science*, *309*(5734), 570–574, doi:10.1126/science.1111772.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldi (2007), Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling, *Int. J. Climatol.*, *27*(12), 1547–1578, doi:10.1002/joc.1556.

- Fridley, J. D. (2009), Downscaling climate over complex terrain: High finescale (<1000 m) spatial variation of near-ground temperatures in a montane forested landscape (Great Smoky Mountains), *J. Appl. Meteorol. Climatol.*, *48*(5), 1033–1049, doi:10.1175/2008JAMC2084.1.
- Funk, C., J. Michaelsen, J. Verdin, G. Artan, G. Husak, G. Senay, H. Gadain, and T. Magadzire (2003), The collaborative historical African rainfall model: Description and evaluation, *Int. J. Climatol.*, *23*(1), 47–66, doi:10.1002/joc.866.
- Graae, B. J., et al. (2012), On the use of weather data in ecological studies along altitudinal and latitudinal gradients, *Oikos*, *121*(1), 3–19, doi:10.1111/j.1600-0706.2011.19694.x.
- Guisan, A., S. B. Weiss, and A. D. Weiss (1999), GLM versus CCA spatial modeling of plant species distribution, *Plant Ecol.*, *143*(1), 107–122, doi:10.1023/A:1009841519580.
- Gurung, A. B., S. W. von Dach, M. F. Price, R. Aspinall, J. Balsiger, J. S. Baron, E. Sharma, G. Greenwood, and T. Kohler (2012), Global change and the world's mountains— research needs and emerging themes for sustainable development, *Mt. Res. Dev.*, *32*(S1), S47–S54, doi:10.1659/MRD-JOURNAL-D-11-00084.S1.
- Hamlet, A. F., and D. P. Lettenmaier (2005), Production of temporally consistent gridded precipitation and temperature fields for the continental United States, *J. Hydrometeorol.*, *6*(3), 330–336, doi:10.1175/JHM420.1.
- Hatchett, B. J., D. P. Boyle, A. E. Putnam, and S. D. Bassett (2015), Placing the 2012–2015 California-Nevada drought into a paleoclimatic context: Insights from Walker Lake, California-Nevada, USA, *Geophys. Res. Lett.*, *42*, 8632–8640, doi:10.1002/2015GL065841.
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, *113*, D20119, doi:10.1029/2008JD010201.
- Hengl, T., G. B. M. Heuvelink, M. P. Tadić, and E. J. Pebesma (2012), Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images, *Theor. Appl. Climatol.*, *107*(1–2), 265–277, doi:10.1007/s00704-011-0464-2.
- Hock, R. (2003), Temperature index melt modelling in mountain areas, *J. Hydrol.*, *282*(1–4), 104–115, doi:10.1016/S0022-1694(03)00257-9.
- Hofstra, N., M. Haylock, M. New, and P. D. Jones (2009), Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, *J. Geophys. Res.*, *114*, D21101, doi:10.1029/2009JD011799.
- Holden, Z. A., A. Swanson, A. E. Klene, J. T. Abatzoglou, S. Z. Dobrowski, S. A. Cushman, J. Squires, G. G. Moisen, and J. W. Oyler (2015), Development of high-resolution (250 m) historical daily gridded air temperature data using reanalysis and distributed sensor networks for the US Northern Rocky Mountains, *Int. J. Climatol.*, *33*(1), 121–131, doi:10.1002/joc.4580.
- Horning, M. E., T. R. McGovern, D. C. Darris, N. L. Mandel, and R. Johnson (2010), Genecology of *Holodiscus discolor* (Rosaceae) in the Pacific Northwest, U.S.A, *Restor. Ecol.*, *18*(2), 235–243, doi:10.1111/j.1526-100X.2008.00441.x.
- Huwald, H., C. W. Higgins, M.-O. Boldi, E. Bou-Zeid, M. Lehning, and M. B. Parlange (2009), Albedo effect on radiative errors in air temperature measurements, *Water Resour. Res.*, *45*, W08431, doi:10.1029/2008WR007600.
- Johnson, R. C., V. J. Erickson, N. L. Mandel, J. B. S. Clair, and K. W. Vance-Borland (2010), Mapping genetic variation and seed zones for *Bromus carinatus* in the Blue Mountains of eastern Oregon, USA, *Botany*, *88*(8), 725–736, doi:10.1139/b10-047.
- Jordan, D. N., and W. K. Smith (1995), Microclimate factors influencing the frequency and duration of growth season frost for subalpine plants, *Agric. For. Meteorol.*, *77*(1–2), 17–30, doi:10.1016/0168-1923(95)02233-N.
- Kalma, J. D., T. R. McVicar, and M. F. McCabe (2008), Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data, *Surv. Geophys.*, *29*(4–5), 421–469, doi:10.1007/s10712-008-9037-z.
- Knick, S. T., S. E. Hanser, and K. L. Preston (2013), Modeling ecological minimum requirements for distribution of greater sage-grouse leks: Implications for population connectivity across their western range, U.S.A, *Ecol. Evol.*, *3*(6), 1539–1551, doi:10.1002/ece3.557.
- Kunkel, K. E., D. R. Easterling, K. Hubbard, and K. Redmond (2004), Temporal variations in frost-free season in the United States: 1895–2000, *Geophys. Res. Lett.*, *31*, L0320, doi:10.1029/2003GL018624.
- Lareau, N. P., and J. D. Horel (2014), Dynamically induced displacements of a persistent cold-air pool, *Boundary Layer Meteorol.*, *154*(2), 291–316, doi:10.1007/s10546-014-9968-5.
- Leeper, R. D., J. Rennie, and M. A. Palecki (2015), Observational perspectives from US Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) network: Temperature and precipitation comparison, *J. Atmos. Oceanic Technol.*, *32*(4), 703–721, doi:10.1175/JTECH-D-14-00172.1.
- Lookingbill, T. R., and D. L. Urban (2005), Gradient analysis, the next generation: Towards more plant-relevant explanatory variables, *Can. J. For. Res.*, *35*(7), 1744–1753, doi:10.1139/x05-109.
- Lopes, T. J., and K. K. Allander (2009), Water budgets of the Walker River basin and Walker Lake, California and Nevada: U. S. Geol. Surv. Sci. Investigations Rep. 2009–5157.
- Lundquist, J. D., and D. R. Cayan (2007), Surface temperature patterns in complex terrain: Daily variations and long-term change in the central Sierra Nevada, California, *J. Geophys. Res.*, *112*, D11124, doi:10.1029/2006JD007561.
- Lundquist, J. D., N. Pepin, and C. Rochford (2008), Automated algorithm for mapping regions of cold-air pooling in complex terrain, *J. Geophys. Res.*, *113*, D22107, doi:10.1029/2008JD009879.
- MacHattie, L. B. (1965), Radiation screens for air temperature measurement, *Ecology*, *46*(4), 533–538, doi:10.2307/1934888.
- Maxim Integrated (2016), DS1922L: iButton temperature loggers with 8KB data-log memory. [Available at <https://www.maximintegrated.com/en/products/digital/data-loggers/DS1922L.html> (Accessed 17 February 2016).]
- McCabe, G. J., J. L. Betancourt, and S. Feng (2015), Variability in the start, end, and length of frost-free periods across the conterminous United States during the past century, *Int. J. Climatol.*, *35*(15), 4673–4680, doi:10.1002/joc.4315.
- McCune, B., and D. Keon (2002), Equations for potential annual direct incident radiation and heat load, *J. Veg. Sci.*, *13*(4), 603–606, doi:10.1111/j.1654-1103.2002.tb02087.x.
- McCutchan, M. H., and D. G. Fox (1986), Effect of elevation and aspect on wind, temperature and humidity, *J. Clim. Appl. Meteorol.*, *25*(12), 1996–2013, doi:10.1175/1520-0450(1986)025<1996:EOEAAO>2.0.CO;2.
- McEvoy, D. J., J. F. Mejia, and J. L. Huntington (2014), Use of an observation network in the great basin to evaluate gridded climate data, *J. Hydrometeorol.*, *15*(5), 1913–1931, doi:10.1175/JHM-D-14-0015.1.
- McGuire, C. R., C. R. Nufio, M. D. Bowers, and R. P. Guralnick (2012), Elevation-dependent temperature trends in the Rocky Mountain front range: Changes over a 56- and 20-year record, *PLoS One*, *7*(9), doi:10.1371/journal.pone.0044370.
- McMaster, G. S., and W. W. Wilhelm (1997), Growing degree days: One equation, two interpretations, *Agric. For. Meteorol.*, *87*(4), 291–300, doi:10.1016/S0168-1923(97)00027-0.
- Meehl, G. A., C. Tebaldi, and D. Nychka (2004), Changes in frost days in simulations of twenty-first century climate, *Clim. Dyn.*, *25*(5), 495–511, doi:10.1007/s00382-004-0442-9.
- Mejia, J. F., J. Huntington, B. Hatchett, D. Koracin, and R. G. Niswonger (2012), Linking global climate models to an integrated hydrologic model: Using an individual station downscaling approach, *J. Contemp. Water Res. Ed.*, *147*(1), 17–27, doi:10.1111/j.1936-704X.2012.03100.x.

- Messerli, B., and J. D. Ives (Eds.) (1997), *Mountains of the World: A Global Priority*, 495 pp., Parthenon Publishing Group, Carnforth.
- Millar, C. I., and R. D. Westfall (2010), Distribution and climatic relationships of the American Pika (*Ochotona princeps*) in the Sierra Nevada and western Great Basin, U.S.A.: Periglacial landforms as refugia in warming climates reply, *Arct. Antarct. Alp. Res.*, 42(4), 493–496, doi:10.1657/1938-4246-42.4.493.
- Minder, J. R., P. W. Mote, and J. D. Lundquist (2010), Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains, *J. Geophys. Res.*, 115, D14122, doi:10.1029/2009JD013493.
- Murtagg, F., and P. Legendre (2014), Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion?, *J. Classif.*, 31, 274–295, doi:10.1007/s00357-014-9161-z.
- Nakamura, R., and L. Mahrt (2005), Air temperature measurement errors in naturally ventilated radiation shields, *J. Atmos. Oceanic Technol.*, 22(7), 1046–1058, doi:10.1175/JTECH1762.1.
- Natural Resources Conservation Service (2015), SNOTEL sites, U.S.D.A. Natural Resources Conservation Service. [Available at <http://www.nv.nrcs.usda.gov/snow/snotel.html> (Accessed 1 July 2015).]
- Oksanen, J. (2015), *Vegan: An introduction to ordination*, pp. 1–10. [Available at <ftp://ftp3.ie.freebsd.org/pub/cran.r-project.org/web/packages/vegan/vignettes/intro-vegan.pdf>.]
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O. Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner (2015), Package “vegan”, *Community ecology package*, version 2.9. [Available at <https://github.com/vegandevs/vegan>.]
- Oyler, J. W., S. Z. Dobrowski, A. P. Ballantyne, A. E. Klene, and S. W. Running (2015), Artificial amplification of warming trends across the mountains of the western United States, *Geophys. Res. Lett.*, 42, 153–161, doi:10.1002/2014GL02803.
- Palecki, M. A., and P. Y. Groisman (2011), Observing climate at high elevations using United States Climate Reference Network approaches, *J. Hydrometeorol.*, 12(5), 1137–1143, doi:10.1175/2011JHM1335.1.
- Pepin, N., C. Daly, and J. D. Lundquist (2011), The influence of surface/free-air decoupling on temperature trend patterns in the western United States, *J. Geophys. Res.*, 116, D10109, doi:10.1029/2010JD014769.
- PRISM Climate Group (2016), Oregon State University [internet]. Descriptions of PRISM spatial climate datasets for the conterminous United States. [Available at [http://www.prism.oregonstate.edu/documents/PRISM\\_datasets.pdf](http://www.prism.oregonstate.edu/documents/PRISM_datasets.pdf), (Cited 25 July 2016).]
- Quantum Geographic Information System Development Team (2015), Quantum GIS Geographic Information System, Open Source Geospatial Foundation Project. [Available at <http://qgis.osgeo.org>.]
- R Development Core Team (2015), The R project for statistical computing. [Available at <http://www.r-project.org>.]
- Rango, A., and J. Martinec (1995), Revisiting the degree-day method for snowmelt computations, *Water Resour. Bull.*, 31(4), 657–669, doi:10.1111/j.1752-1688.1995.tb03392.x.
- Richardson, S. J., F. V. Brock, S. R. Semmer, and C. Jirak (1999), Minimizing errors associated with multiplate radiation shields, *J. Atmos. Oceanic Technol.*, 16(11), 1862–1872, doi:10.1175/1520-0426(1999)016<1862:MEAWMR>2.0.CO;2.
- Riley, S. J., S. D. Degloria, and R. Elliot (1999), A terrain ruggedness index that quantifies topographic heterogeneity, *Intermt. J. Sci.*, 5, 23–27.
- Roberts, D. W., and S. V. Cooper (1989), Concepts and techniques of vegetation mapping, Gen. techn. rep. INT-US Dep. of Agriculture, Forest Service, Intermountain Research Station (USA).
- Rousseeuw, P. J. (1987), Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, 20, 53–65, doi:10.1016/0377-0427(87)90125-7.
- Saito, L., F. Biondi, R. Devkota, J. Vittori, and J. D. Salas (2014), A water balance approach for reconstructing streamflow using tree-ring proxy records, *J. Hydrol.*, 529, 535–547, doi:10.1016/j.jhydrol.2014.11.022.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers (2008), A description of the Advanced Research WRF version 3, *NCAR Tech. Note NCAR/TN-475+STR*, 113 pp., doi:10.5065/D68S4MVH.
- Stoklosa, J., C. Daly, S. D. Foster, M. B. Ashcroft, and D. I. Warton (2014), A climate of uncertainty: Accounting for error in climate variables for species distribution models, *Methods Ecol. Evol.*, 6(4), 412–423, doi:10.1111/2041-210X.12217.
- Strahler, A. N. (1952), Hypsometric (area-altitude) analysis of erosional topography, *Geol. Soc. Am. Bull.*, 63(11), 1117–1142, doi:10.1130/0016-7606(1952)63.
- Swain, D. L. (2015), A tale of two California droughts: Lessons amidst record warmth and dryness in a region of complex physical and human geography, *Geophys. Res. Lett.*, 42, 999–1,003, doi:10.1002/2015GL066628.
- Thompson, R. S., K. H. Anderson, R. T. Pelltier, L. E. Strickland, P. J. Bartlein, and S. L. Shafer (2012), Quantitative estimation of climatic parameters from vegetation data in North America by the mutual climatic range technique, *Quat. Sci. Rev.*, 51, 18–39, doi:10.1016/j.quascirev.2012.07.003.
- Thompson, R. S., K. H. Anderson, R. T. Pelltier, L. E. Strickland, S. L. Shafer, P. J. Bartlein, and A. K. McFadden (2015), *Atlas of Relations Between Climatic Parameters and Distributions of Important Trees and Shrubs in North America: Revisions for All Taxa From the United States and Canada and New Taxa From the Western United States*, U.S. Geol. Surv. Prof. Pap., 1650–G, Reston, Va., doi:10.3133/pp1650G.
- Thornton, P. E., and S. W. Running (1999), An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation, *Agric. For. Meteorol.*, 93(4), 211–228, doi:10.1016/S0168-1923(98)00126-9.
- Thornton, P. E., M. M. Thornton, B. W. Mayer, N. Wilhelm, Y. Wei, R. Devarakonda, and R. Cook (2012), Daymet: Daily surface weather on a 1 km grid for North America, 1980–2008, Oak Ridge Natl. Lab. (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC), 1.
- Viviroli, D., and R. Weingartner (2004), The hydrological significance of mountains: From regional to global scale, *Hydrol. Earth Syst. Sci.*, 8(6), 1017–1030, doi:10.5194/hess-8-1017-2004.
- Warren, D. L., M. Cardillo, D. F. Rosauer, and D. I. Bolnick (2014), Mistaking geography for biology: Inferring processes from species distributions, *Trends Ecol. Evol.*, 29(10), 572–580, doi:10.1016/j.tree.2014.08.003.
- Wells, P. V., and L. M. Shields (1964), Distribution of *Arrea divaricata* in relation to a temperature inversion at Yucca Flat, southern Nevada, *Southwest Naturalist*, 9(2), 51–55, doi:10.2307/3668783.
- Wells, P. V. (1983), Paleobiogeography of Montane Islands in the Great Basin since the last Glaciopluvial, *Ecol. Monogr.*, 53(4), 341, doi:10.2307/1942644.
- West, A. M., S. Kumar, T. Wakie, C. S. Brown, T. J. Stohlgren, M. Laituri, and J. Bromberg (2015), Using high-resolution future climate scenarios to forecast *Bromus tectorum* invasion in Rocky Mountain National Park, *PLoS One*, 10(2), 1–15, doi:10.1371/journal.pone.0117893.
- Whiteman, C. D. (1982), Breakup of temperature inversions in deep mountain valleys: Part I. Observations, *J. Appl. Meteorol.*, 21(3), 270–289, doi:10.1175/1520-0450(1983)022<1314:COOTII>2.0.CO;2.
- Willmott, C. J., and S. M. Robeson (1995), Climatologically aided interpolation (CAI) of terrestrial air temperature, *Int. J. Climatol.*, 15(2), 221–229, doi:10.1002/joc.3370150207.
- World Meteorological Organization (2008), Guide to meteorological instruments and methods of observation, Geneva. [Available at <http://www.wmo.int/pages/prog/www/IMOP/CIMO-Guide.html>.]