BAMS Article

Are Atmospheric Models Too Cold in the Mountains? The State of Science and Insights from the SAIL Field Campaign

William Rudisill,[a] Alan Rhoades,[a] Zexuan Xu,[a] and Daniel R. Feldman[a]

KEYWORDS: Mountain meteorology; Model evaluation/performance; Temperature; Climate

ABSTRACT: Mountains play an outsized role in water resource availability, and the amount and timing of water they provide depend strongly on temperature. To that end, we ask the question: How well are atmospheric models capturing mountain temperatures? We synthesize results showing that high-resolution, regionally relevant climate models produce 2-m air temperature (T2m) measurements colder than what is observed (a "cold bias"), particularly in snow-covered midlatitude mountain ranges during winter. We find common cold biases in 44 studies across global mountain ranges, including single-model and multimodel ensembles. We explore the factors driving these biases and examine the physical mechanisms, data limitations, and observational uncertainties behind T2m. Our analysis suggests that the biases are genuine and not due to observation sparsity or resolution mismatches. Cold biases occur primarily on mountain peaks and ridges, whereas valleys are often warm biased. Our literature review suggests that increasing model resolution does not clearly mitigate the bias. By analyzing data from the Surface Atmosphere Integrated Field Laboratory (SAIL) field campaign in the Colorado Rocky Mountains, we test various hypotheses related to cold biases and find that local wind circulations, longwave (LW) radiation, and surface-layer parameterizations contribute to the T2m biases in this particular location. We conclude by emphasizing the value of coordinated model evaluation and development efforts in heavily instrumented mountain locations for addressing the root cause(s) of T2m biases and improving predictive understanding of mountain climates.

SIGNIFICANCE STATEMENT: Mountain climates are rapidly changing, and along with them are the temperature-sensitive components of the water budget that societies have relied on. Yet atmospheric models, from those that predict the weather to those that predict the future climate, are several degrees too cold on average in these same mountain regions. This cold bias has not been systematically identified in the published literature yet, so we discuss evidence of its pervasiveness across models, its potential causes, and pathways to eliminate it using targeted models and observations. With community support, this bias can be uprooted, thereby enabling model projections that better project the climatic and water resource changes in these vital regions.

DOI: 10.1175/BAMS-D-23-0082.1
Corresponding author: William Rudisill, williamrudisill@lbl.gov
Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/BAMS-D-23-0082.s1.

Manuscript received 10 April 2023, in final form 28 March 2024, accepted 2 April 2024

AFFILIATION: [a] Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, California

1. Introduction

In this study, we compile evidence suggesting that many types of high-resolution atmospheric models (Gutowski et al. 2020) report cold biases in 2-m air temperature ("T2m") measurements across major mountain regions. The inaccuracies in modeled T2m beg the question of how such biases should be treated and whether key processes are being adequately represented at the scales required for the skillful projection of climate change in Earth's water towers (Viviroli and Weingartner 2004; Immerzeel et al. 2020; Siirila-Woodburn et al. 2021). The bias is defined here as the difference between the model and the observation, so a cold bias of 1°C indicates that the model is colder than the observations by that amount. In the following sections, we describe 44 studies from the last decade, including both limited-area models (LAMs) and variable resolution general circulation models (Table A1). Most notably, at the mountain-range scale, we were unable to find examples of warm biases in the published literature (though, as we will show, warm biases can occur in valley subregions of mountain ranges). For simplicity, we refer to all of these previous studies as Models of Applicable Resolution for Mountain Meteorology across Time Scales (MARMOTS), based on their common goal of producing meteorological information at scales applicable to questions related to mountain hydroclimates.

a. Mountains in a warming world. Though they occupy a small percentage of Earth's landmass (between 13% and 30%; Kapos et al. 2000; Körner et al. 2011, 2017; Snethlage et al. 2022), mountains have an outsized global impact as the world's water towers (Viviroli and Weingartner 2004; Immerzeel et al. 2020; Siirila-Woodburn et al. 2021). Theoretically, mountains warm from anthropogenic climate change at rates different from low-lying regions through a variety of mechanisms related to both cryospheric changes (e.g., snow albedo feedbacks) and atmospheric thermodynamic considerations (Mountain Research Initiative EDW Working Group 2015; Palazzi et al. 2019; Hock et al. 2019). However, observational determination of elevation-dependent warming has proven more elusive (Pepin et al. 2022), and such assessments are limited by data quality, continuity, and coverage in high-elevation areas (Oyler et al. 2015b; McAfee et al. 2019; Ma et al. 2019). At the same time, T2m is the first-order control of whether precipitation falls as rain or snow (Harpold et al. 2017; Jennings et al. 2018) including rain-on-snow events (Heggli et al. 2022), snowmelt timing (Musselman et al. 2021), and streamflow drought (Udall and Overpeck 2017; Gangopadhyay et al. 2022) in many mountain regions.

A number of factors complicate the spatiotemporal patterns of T2m in complex mountain terrain compared to flat, low-elevation areas. Figure 1 illustrates idealized depictions of some of these processes. Mountains are generally high elevation, so the total mass of the atmosphere above them (pressure) is less than in low-lying areas. As a consequence, there is less diffuse radiation from scattering and more intense direct beam radiation (Smith 2019). At the surface, the incident radiation also depends on slope, aspect, and terrain shadowing and terrain reflection (Fig. 1a), so sun-facing aspects may have warmer temperatures (Strachan and Daly 2017). Additional complications of T2m in mountain regions may arise

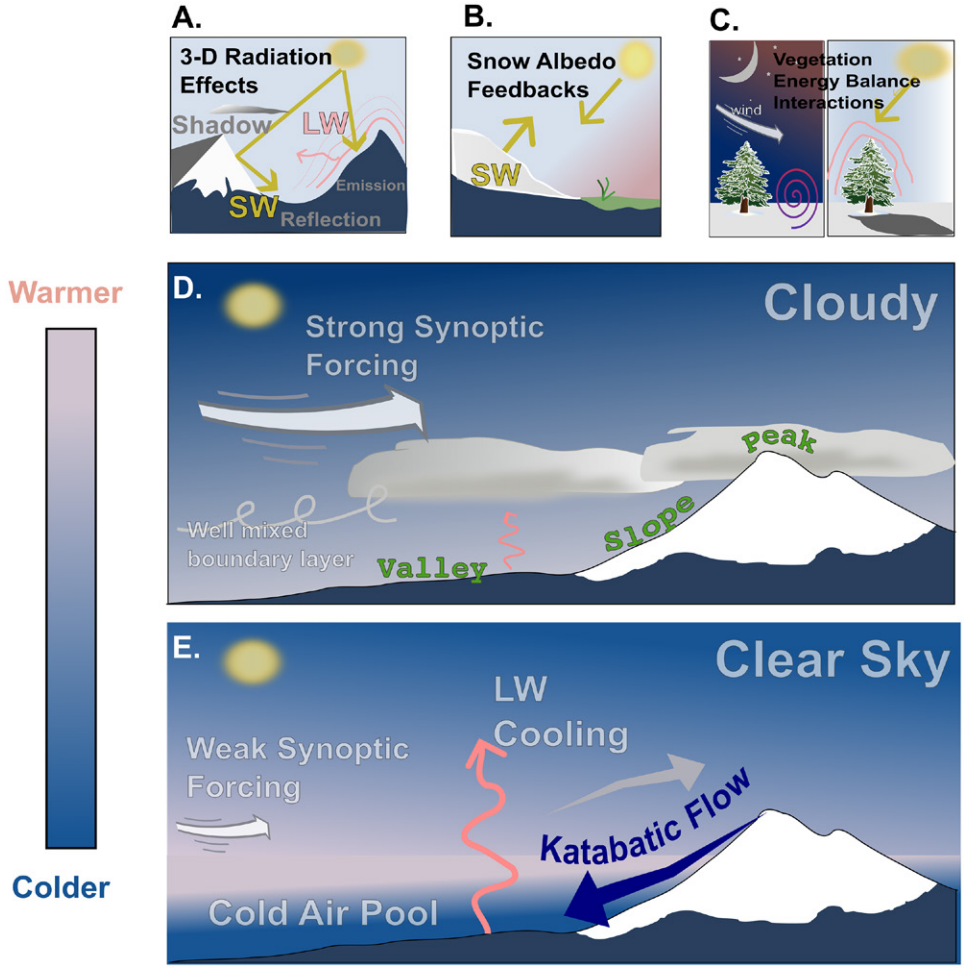

Fig. 1. Depiction of processes influencing mountain T2m. (a) Incident solar radiation at the ground surface depends on terrain aspect and shadowing. Mountain slopes can also reflect SW and emit LW radiation; (b) snow reflects solar radiation, cooling surface temperatures through the snow-albedo feedback; (c) vegetation's high surface roughness and lower albedos warm surrounding air relative to nonvegetated snow surfaces; (d) moderate synoptic forcing leads to upper-air and boundary layer temperature mixing, with an uninverted vs elevation profile; (e) katabatic flows develop valley temperature inversions during clear-sky, cloud-free conditions when surface LW cooling is strong.

from the patchwork of land-cover types, namely, snow and vegetation, which impact the surface energy balance and therefore T2m. Snow has a high albedo and a high emissivity (Fig. 1b), so it both very effectively reflects incoming radiation and emits heat in the longwave (LW) (Armstrong and Brun 2008). Vegetation has high surface roughness and a lower albedo than snow (Fig. 1c) and likewise influences the surface energy balance through both radiative and turbulent exchange mechanisms (Lee et al. 2011; Schultz et al. 2017; Burakowski et al. 2018).

Cold-air pools are common features in mountain climates that result from the topography and occur especially during periods of light winds, clear skies, during winter, and at night (Figs. 1d,e; Daly et al. 2010; Whiteman 2000; Lundquist et al. 2008). In such cases, cold, dense air drains from aloft and settles in valley bottoms, leading to stable stratification with relatively warm air overlying the cold air near the surface. Cold-air pools may also form even with relatively minimal cold-air drainage, in cases where topography limits mixing with ambient air (Clements et al. 2003) and radiative cooling dominates. In these cases, T2m observed within the valley cold pool may be colder than T2m observed at higher altitudes (an inverted temperature profile; Fig. 1e). Such features may mix out during the day or persist for days or weeks (Fig. 1d). The intricacies of the mountain planetary boundary layer are complex and

have only begun to be explored (Lehner and Rotach 2018; Serafin et al. 2018). Serafin et al. (2018) provide a thorough review of many unique features of the mountain planetary boundary layer and challenges for models therein.

b. MARMOTS: The keys to understanding mountain climates. High-resolution atmospheric models are nonetheless the best methods for assessing climate impacts in the world's mountain regions, as uniform-resolution general circulation models are currently too coarse to resolve the mountain topography that is so fundamental for shaping mountainous climates (Rhoades et al. 2018a; Gutowski et al. 2020; Demory et al. 2014). Prein et al. (2015) discuss some of the relatively recent advancements in MARMOTS development. Numerical weather prediction models often share similar dynamical cores and parameterizations with their climate model counterparts but are intended to operate on shorter time scales with continually updated initial and boundary conditions and even state information (through nudging) that make extensive use of existing data assimilation datasets. Thus, output from numerical weather prediction models is increasingly used as inputs into models for mountain hydrological research (e.g., Currier et al. 2017; Reynolds et al. 2021; Meyer et al. 2023), so numerical weather prediction studies are also considered in this review.

c. Goals and outline. The paper is structured as follows. We start by reviewing papers evaluating MARMOTS temperature biases. We augment the literature review by analyzing T2m data from NCAR's "high-resolution CONUS (HRCONUS)" model dataset, presented in Liu et al. (2017), which covers the entire United States at a 4-km grid spacing. We then pose the question: Is T2m bias truly a bias, or a by-product of model-to-observation resolution mismatches? To answer these questions, we review observational capabilities across the globe and examine some of the gridded reference datasets that are frequently used to compute T2m biases.

Finally, we examine model T2m biases in the 300-km^2 upper East River watershed (ERW), located in the Colorado Rockies, using data collected during the Surface Atmosphere Integrated Field Laboratory (SAIL) field campaign (Feldman et al. 2023). We examine T2m from nine stations located throughout the ERW valley (spanning ~600 m of elevation) in addition to single-site measurements of T2m covariates, namely, near-surface wind speed, 2-m specific humidity, snow skin temperature, precipitable water vapor (PWV), and cloud cover fraction. We compare observed T2m to output from the High-Resolution Rapid Refresh (HRRR) model (Benjamin et al. 2016) and the reanalysis-forced Weather Research and Forecasting (WRF; Powers et al. 2017) Model configuration described in Xu et al. (2023). We do not propose a solution to the problem of winter season T2m cold biases, as the solutions will undoubtedly require community-wide efforts, but instead seek to demonstrate the nature of the problem and illuminate paths forward for solutions.

2. MARMOTS are cold biased over mountains, particularly in winter

A review of the recent literature shows that 44 studies report winter season cold biases across the world's major mountain regions (Fig. 2 and Table A2). Mentions of model cold bias were found by first searching major studies, such as multimodel ensemble evaluation studies and those that have been widely shared and cited in the mountain hydroclimate field. We uncovered the majority of the studies by looking at the chain of references from those studies. Additional studies were found using search terms on Google Scholar such as "regional climate model evaluation," or "WRF Model evaluation mountains" and related terms. While it is possible that some published model configurations have shown warm biases over mountain regions, none were found from this analysis.

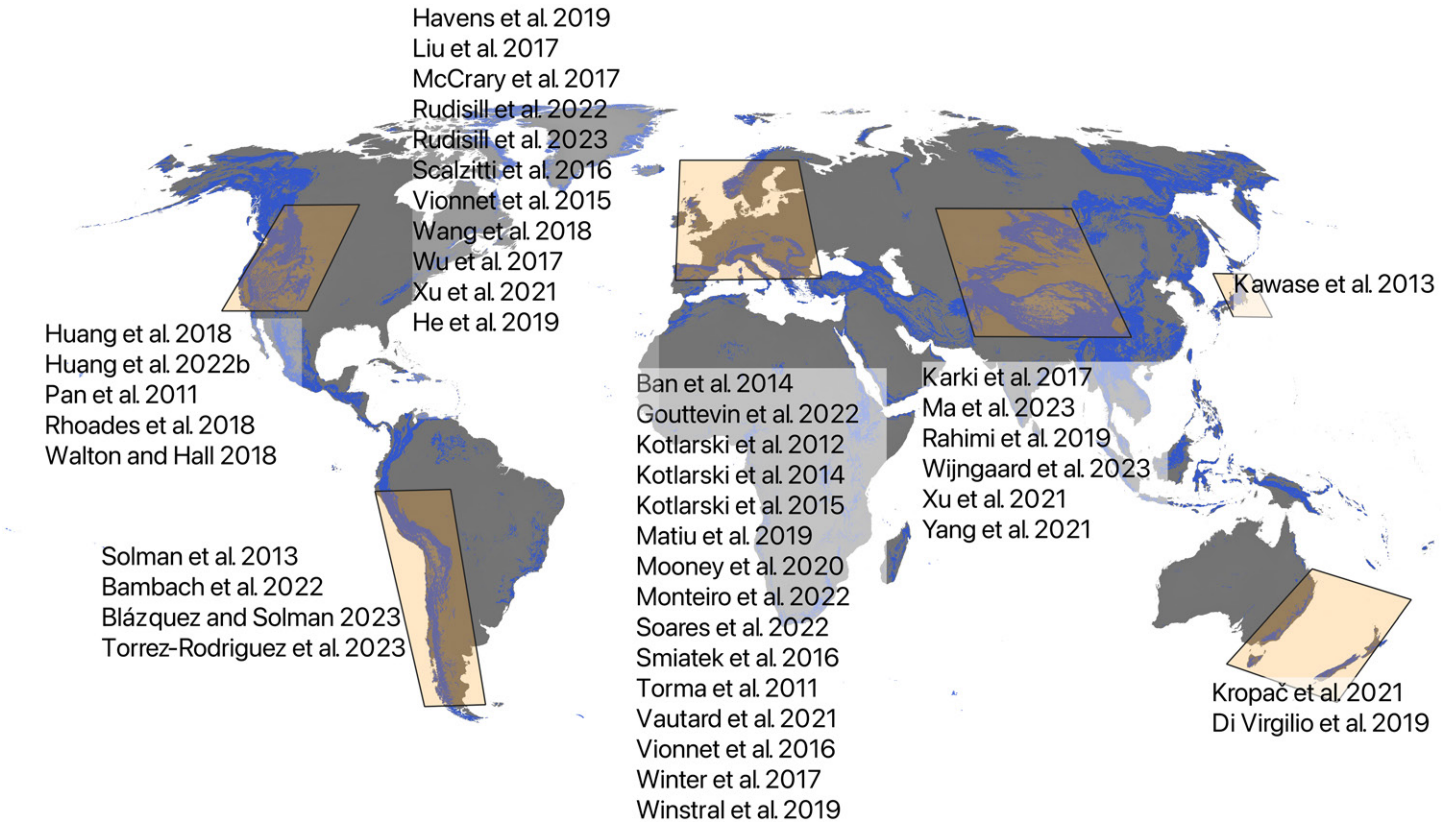

Fig. 2. The geographic locations (yellow boxes) of studies reporting mountain cold biases. The global mountain classifications from Snethlage et al. (2022) are shown in blue. The 44 studies are all reported in the peer-reviewed literature with publication dates spanning 2011–23 and include both single-model experiments and multimodel ensemble evaluations.

The model biases reported in these studies are typically determined by evaluating model output against gridded meteorological observations covering the same model extent or by comparing individual weather stations to the closest model grid cells. T2m tends to reach a minimum at night (TMIN) and a maximum during the day (TMAX) and is often reported in terms of the daily average (TAVG). Unless specified, T2m refers to TAVG. The magnitude of cold biases generally ranges from 1° to 5°C (Table A2), though not all studies report a quantitative value of the bias.

a. Examples of model cold bias. Vautard et al. (2021) present a historical climate bias analysis of temperature from the 0.11° horizontal resolution European Coordinated Regional Downscaling Experiment (EURO-CORDEX) ensemble consisting of 8 global climate model drivers and 11 independent regional models. The median model is cold biased over the Alps, Pyrenees, and Scandinavian ranges. Even the "hottest" models (95th percentile), with positive temperature biases in lowlands, are too cold in mountain regions. They further show that the TMIN bias is dominated by model structural variability rather than boundary conditions. Earlier EURO-CORDEX analyses showed similar cold biases in the Alps and Scandinavian ranges (Kotlarski et al. 2014). South American CORDEX (SA-CORDEX) experiments are similarly cold biased in the Andes (Blázquez and Solman 2023; Solman and Blázquez 2019; Torrez-Rodriguez et al. 2023). Similar biases are also found in SA-CORDEX and North American CORDEX (NA-CORDEX) evaluations (Torrez-Rodriguez et al. 2023; Xu et al. 2019). Results from Australasia-CORDEX show extensive cold biases that may be related to elevation (Di Virgilio et al. 2019). NA-CORDEX is cold biased in the Sierra Nevada (Xu et al. 2019) and possibly the Southern Rockies (McCrary et al. 2017).

Variable-resolution global models likewise demonstrate cold biases. Rhoades et al. (2018b) found extensive cold biases in variable-resolution CESM (VR-CESM) (Table A1) over the Sierra Nevada. Similar results are found again in the Sierra Nevada (Fig. 13 of Xu et al. 2018; Xu et al. 2021) as well as the Rockies (Fig. 3 of Wu et al. 2017), the Tibetan Plateau (Xu et al. 2021), and the Andes (Bambach et al. 2022). The System for Integrated Modeling of the Atmosphere-MPAS (SIMA-MPAS) model (Table A1) likewise shows a cold bias for high peaks in the western United States (WUS; X. Huang et al. 2022). The recently evaluated regionally refined mesh Energy Exascale Earth System Model (RRM-E3SM) (Table A1) demonstrates pervasive cold biases, though elevation-specific analyses are not presented (Tang et al. 2023).

The WRF Model has been applied across the globe as both a regional climate model and a numerical weather prediction model. WRF's utility for supporting snow-water resource applications has been strongly argued given its skill in simulating orographic precipitation (Ikeda et al. 2010; Gutmann et al. 2012; He et al. 2019; Liu et al. 2011; Lundquist et al. 2019), but fewer studies scrutinize WRF's T2m performance in mountains. Nonetheless, cold biases in that model have been articulated or shown in the Sierra Nevada (Fig. 5 of Pan et al. 2011; Huang et al. 2018; Walton and Hall 2018), the Wasatch (Scalzitti et al. 2016), and the Idaho–Bitterroot (Rudisill et al. 2022; Havens et al. 2019) ranges and other interior WUS mountain ranges (Fig. 13 of Wang et al. 2018). Similar biases in that model are found in Japan, the Himalayas, and the Southern Alps (Kawase et al. 2013; Karki et al. 2017; Kropač et al. 2021). NCAR's convection-permitting WRF simulations described in Liu et al. (2017), hereafter L2017, cover all of CONUS at a 4-km horizontal grid spacing between 2000 and 2013 using ERA-Interim lateral boundary conditions. L2017 is cold biased over major mountain ranges, particularly on snow-topped peaks (He et al. 2019). As this dataset is publicly available and one of very few covering the entire WUS at a 4-km grid spacing, we analyze L2017 in greater detail in the next section.

b. Scrutinizing T2m biases from the L2017 4-km WRF dataset. To better illustrate the nature of T2m biases, we evaluate data from L2017 for January–March 2008 (Fig. 3a). This specific time period is chosen because it was also examined in L2017 (Fig. 11 of their paper), and they note that it was also analyzed in several other studies. We group temperature biases by selected mountain regions using the regional definitions from Snethlage et al. (2022). The T2m biases are computed against the 4-km Parameter-Elevation Regressions on Independent Slopes Model (PRISM) AN81d daily temperature (https://prism.oregonstate.edu/). In addition, we perform a landform classification that groups each site as either a slope, a valley, or ridge-top grid cell using standard terrain position metrics (Lindsay 2016). The supplemental material provides additional information about the processing steps. We also group T2m biases by the entire dataset (every grid cell evaluated in the model against PRISM) and only those grid cells that encompass a weather station observation. We do this because the PRISM data should be very close to the underlying observations for those grid cells with weather stations, so we can test whether biases persist for areas with observations or whether they are primarily in locations where PRISM does not have an observation.

We find T2m biases are approximately normally distributed across the entire CONUS (Figs. 3b,i), with a mean near zero, but each mountain range of the WUS shows a cold bias (Figs. 3c–i) of between 0.8° and 1.4°C averaged across all mountain-range grid cells. The dry continental interior ranges (e.g., Figs. 3h,c) have more extreme cold biases than the coastal ranges (e.g., Figs. 3e,f), reaching cold bias values of over 3°C in certain regions. Cold biases persist even when only the grid cells containing weather station observations are compared between WRF and PRISM (and in some cases become worse), suggesting that biases are not merely a product of out-of-sample predictions made by PRISM. The geographic pattern of the T2m biases suggests that they are not purely a function of

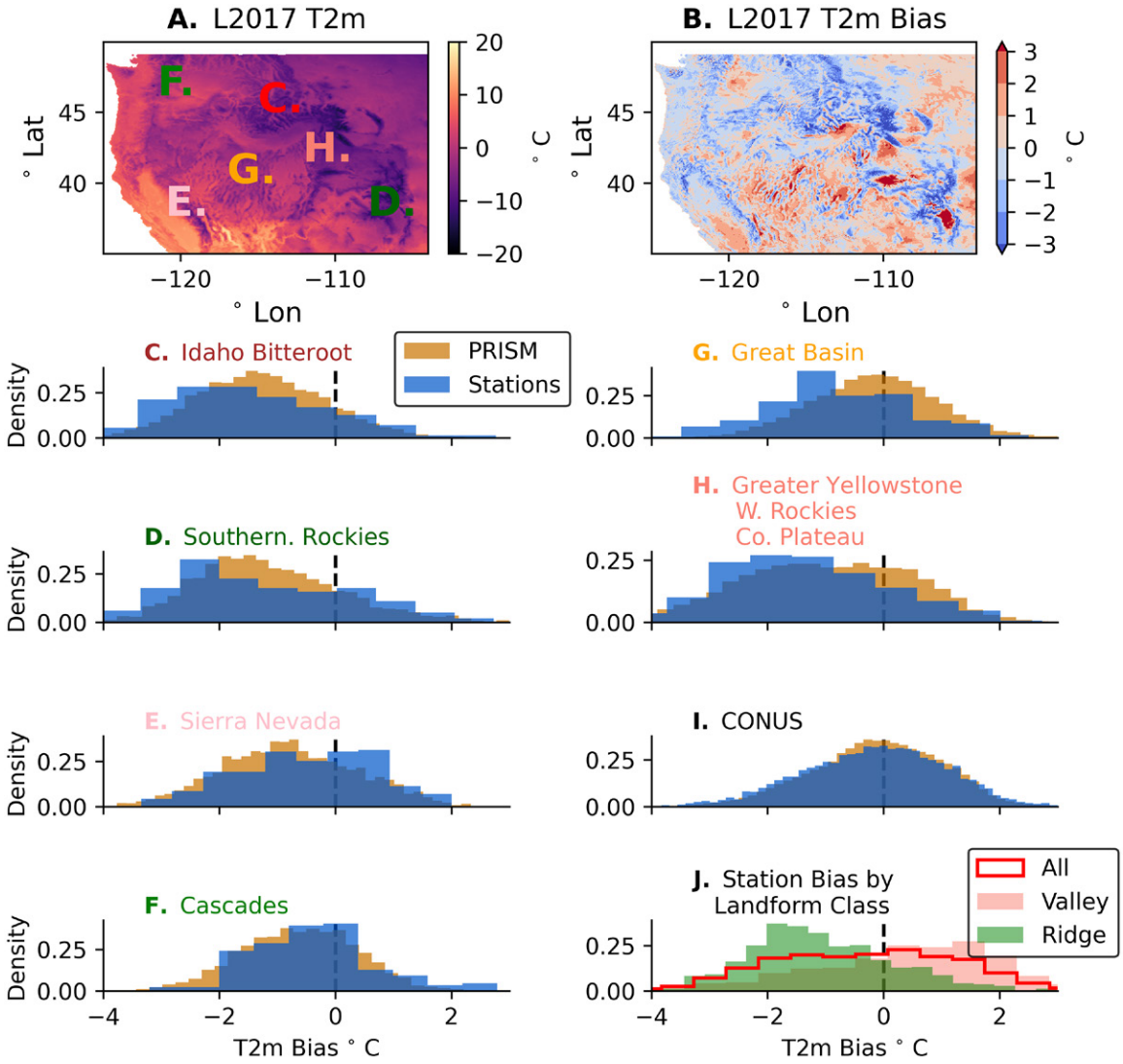

Fig. 3. (a) Average T2m from the Liu et al. (2017) 4-km WRF dataset for January through March 2008 across select mountain ranges of the WUS in (b) T2m bias evaluated using the PRISM AN81d product as reference, (c)–(h) T2m biases grouped by mountain range for all PRISM cells (orange) and for only cells containing station observations (blue), (i) T2m bias for all of CONUS, and (j) T2m biases for western U.S. stations greater than 1000 m of elevation (all) grouped by valley and ridge landform types. All units are in degrees Celsius.

elevation or snow cover. When we group the T2m biases at weather station grid cells by land classification (Fig. 3j), we find not only that the model cold biases are primarily at ridge locations but also that the model is typically warm biased in valleys greater than 1000 m above mean sea level.

From the studies examined, a common thread emerges: Aggregating across entire mountain ranges, model T2m fields are cold biased on average, though mountain valleys are sometimes too warm. T2m biases are generally more significant in the winter, and they do not necessarily decrease in magnitude as the model resolution increases. The following sections evaluate the biases and their potential causes in greater detail.

3. How are MARMOTS biases determined? Notes on observations and gridded data products in mountains

Finding persistent model biases begs the question of how well we actually understand the reference data, particularly in mountains where measurements are challenging. This section reviews current challenges and limitations of T2m observations in mountains.

a. Observation and measuring challenges of T2m networks. While measuring T2m in mountains is easier than measuring precipitation (Lundquist et al. 2019; Goodison et al. 1998), observations can themselves suffer from calibration challenges (Oyler et al. 2015b), temporal inhomogeneities (McAfee et al. 2019), and representation biases. For instance, less than 1% of the stations of the well-vetted and globally spanning Global Historical Climatology Network version 4 (GHCNv4) dataset (Menne et al. 2018) are above 3000 m (Mountain Research Initiative EDW Working Group 2015) and are systematically biased toward more accessible, less-rugged regions of mountain ranges (Thornton et al. 2022).

Figure 4a shows the temperature-observing stations across the WUS used by the PRISM dataset, which includes nearly all available data streams. These include Natural Resources Conservation Service (NRCS) Snowpack Telemetry (SNOTEL; Serreze et al. 1999; Strachan and Daly 2017), National Weather Service Cooperative Observer Program (COOP), Remote Automatic Weather Stations (RAWS; Zachariassen et al. 2003), and a variety of other local networks with data distributed by the MesoWest network (Horel et al. 2002). The primary high-elevation T2m observations in the mountains are from SNOTEL (Fig. 4d) which is preferentially located in tree-sheltered, midvalley regions that accumulate deep snowpacks (Strachan and Daly 2017). The majority of sites are located on slopes, with the exception of COOP stations, which are almost exclusively in valley locations (Fig. 4d). The station densities

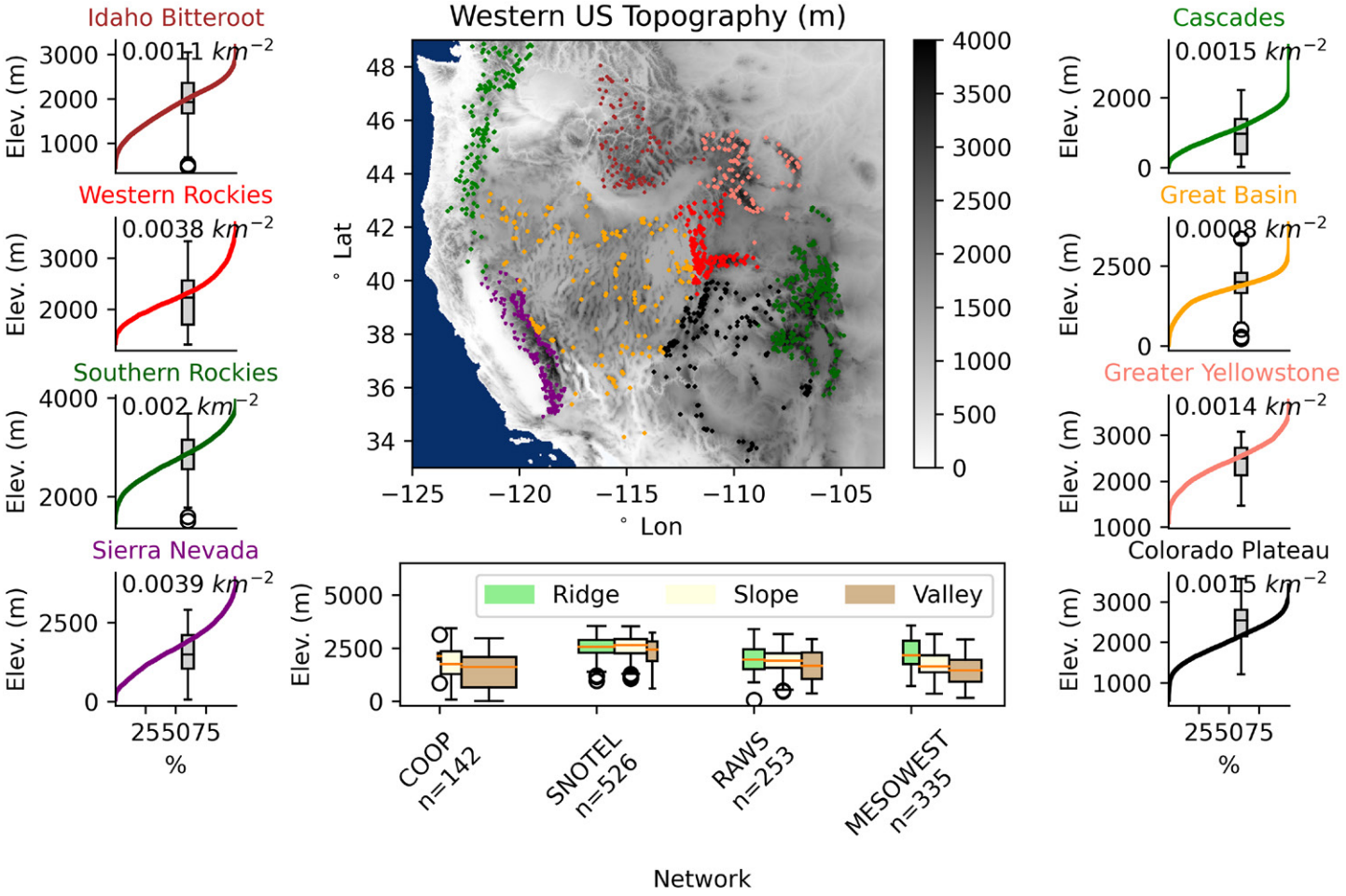

Fig. 4. (a) T2m observing stations used by the PRISM dataset at the time of publication in the WUS, including the National Weather Service COOP, RAWS (Zachariassen et al. 2003), NRCS SNOTEL stations (Serreze et al. 1999), and "other" sites included in the MesoWest network (Horel et al. 2002). All station points are denoted by blue circles, and stations within one of the eight mountain-range zones examined are denoted by marker color. (b),(c) The elevation distribution of stations (box and whiskers) compared against the parent mountain-range cumulative area–elevation curves. The number of observing stations per 1000 square kilometers is also shown in bold type in the upper left. (d) Stations classified by elevation and landform type positions (ridge, slope, and valley).

are between 1 and 4 stations per 1000 km^2 in the major ranges of the WUS, with the lowest density in the Great Basin and Idaho–Bitterroot range. At the same time, the median station elevation for western mountain ranges is fairly well aligned with the distribution of topography, covering between 40% and 85% of the cumulative area of each mountain range. The Greater Yellowstone and the Sierra Nevada have the worst observational coverage over their highest elevations, with 9%–10% of the total (high-elevation) land area completely unsampled by weather stations (Figs. 4b,c).

b. The challenges of "gridding" sparse T2m observations in complex terrain. The development of gridded observational products from station observations is a widely utilized practice and facilitates direct comparisons with model outputs. Because there are many steps associated with gridding, understanding the physical and statistical assumptions made in gridded observational products is central to diagnosing and quantifying sources of model T2m biases. Gridding approaches generally create two-dimensional fields by attempting to capture the salient processes that can be predicted with each grid box in the domain given a sparse series of T2m values from stations. This prediction can be based on statistical models such as fitting smooth surfaces or interpolating between observation points using various approaches such as regression kriging or locally weighted interpolations (Thornton et al. 1997; Maurer et al. 2002; Livneh et al. 2013; Dodson and Marks 1997; Fick and Hijmans 2017). Terrain elevation is typically used as a covariate, while other covariates such as distance from the coast and terrain aspect are used less commonly, and some algorithms explicitly model cold-air pool formation in mountain valleys based on topographic factors (Daly et al. 2008). Other gridded products incorporate remote sensing observations (Oyler et al. 2015a; Alvarez-Garreton et al. 2018) or use reanalysis products as starting points (Cosgrove et al. 2003; Weedon et al. 2014). Table A3 provides a list of gridded T2m datasets that are used as reference in the studies we examined in this paper. The PRISM dataset is frequently used in the United States, whereas the E-OBS version 17 (E-OBS17) dataset is commonly used in Europe and was explicitly designed for the purpose of MARMOTS evaluation (Haylock et al. 2008). Additionally, many datasets may still use PRISM data as part of their gridding procedure (Behnke et al. 2016), so there may be common sources of skill and/or error in different gridding procedures.

If there were dense observations in mountain terrain, then elevational gradients of T2m would simply be calculated from those observations—but this is not the case. In some gridded datasets, it is assumed that T2m decreases with terrain elevation at a rate equivalent to the lapse rate of the international standard atmosphere (6.5°C km^{-1}; note that the lapse rate is typically defined as the rate of the decrease in temperature with altitude, so positive lapse rate values indicate decreasing temperatures above the surface). This assumption has been shown to produce too cold temperatures at mountain peaks (Walton and Hall 2018; Mizukami et al. 2014; Newman et al. 2015), and an abundance of evidence suggests that the true T2m elevational gradient is generally less steep or even of the opposite sign (i.e., a temperature inversion, where T2m increases with elevation) particularly during calm conditions (Dobrowski et al. 2009; Lute and Abatzoglou 2021; Blandford et al. 2008; Minder et al. 2010; Lundquist and Cayan 2007), while also exhibiting seasonal variation (Liston and Elder 2006; Immerzeel et al. 2014; Kattel et al. 2013; Rozante et al. 2022). Consequently, the elevational gradient assumptions of any gridded dataset should be carefully considered before diagnosing a model simulation as biased.

There have been relatively few independent, mountain-focused validations of global or regional T2m datasets either among products or between products and independent observations. PRISM and the Daymet (Thornton et al. 2016) datasets differ in TMIN January temperatures by as much as 1°–3°C in WUS mountains (Daly et al. 2008), and some comparisons show

absolute differences as high as 6°C (Walton and Hall 2018) for reasons that are sometimes attributable to T2m elevational gradient assumptions (Newman et al. 2015). Even fewer studies have compared gridded datasets against out-of-sample observational sources. Those that do provide such comparisons show that dataset skill varies by season and depends on physiographic position and other factors (Strachan and Daly 2017).

Last, the relationships between TMIN, TMAX, and TAVG in mountain environments and over snow cover may warrant additional scrutiny. For instance, PRISM creates grids of TMIN and TMAX and computes TAVG as the arithmetic mean of the two fields, implicitly assuming that diurnal air temperatures are symmetrical about the mean. Few studies have investigated the assumption of symmetry of diurnal temperature in complex terrain, though Wang et al. (2018) intentionally compared the average of WRF TMIN and TMAX (rather than the average of all sub-daily outputs) against TAVG from PRISM, and still found a cold bias in the interior WUS mountains. Furthermore, the time of day that observations are recorded poses a well-known challenge to analyzing temperature records, and model evaluation procedures need to be cognizant of this issue (Wang and Zeng 2014).

4. Possible causes of T2m cold bias

Many of the studies we have surveyed also include hypotheses on the cause(s) of modeled T2m cold biases, which we report without judgment in Table 1. The hypotheses include errors in cloud cover, snow properties, surface albedo, low wind speeds, radiation [shortwave (SW) and longwave], and surface-layer parameterizations. Some of the proposed causes cannot necessarily be separated (i.e., too much snow accumulation may lead to a high-albedo bias). Each of the causes is, at the outset, plausible, but the relative and absolute importance of each of these as a function of location and time is central to addressing this bias. Therefore, the following section investigates possible causes in greater detail.

a. Observation density, quality, and model scale mismatch. A number of studies have noted that data sparsity, quality, or related issues with the gridded reference data itself explain, in part, the observed model T2m biases (Table 1). However, we find that the reported MARMOTS cold biases are likely true biases and not just artifacts of under-sampling or data representation issues. Model T2m biases do not differ between grid cells with weather station observations and ones without those observations, as they would if the biases were

Table 1. Potential causes of model T2m cold biases.

Proposed cause	Citation
Too little cloud cover	Vionnet et al. (2016), Monteiro et al. (2022)
High snow bias	Kotlarski et al. (2014), Vionnet et al. (2016), Ma et al. (2023), Liu et al. (2017), McCrary et al. (2017), Careto et al. (2022)
Surface albedo	Yang et al. (2021), Xu et al. (2019), He et al. (2019), Careto et al. (2022), Torrez-Rodriguez et al. (2023)
Snow-covered area parameterization	Liu et al. (2017)
Wind speeds too slow	Xu et al. (2021), Gouttevin et al. (2022)
Downwelling SW low biased	Xu et al. (2021), Vionnet et al. (2016)
Downwelling LW low biased	Gouttevin et al. (2022), Vionnet et al. (2016)
Surface-layer parameterization	Rontu et al. (2016), Monteiro et al. (2022)
Observation sparsity	Torma et al. (2011), Huang et al. (2018), Fernández et al. (2021), Torrez-Rodriguez et al. (2023)
Observation quality	Rasmussen et al. (2023)
Model-to-observation elevation differences	Kotlarski et al. (2014)

the result of the gridding procedure. However, L2017 biases persist even when only the grid cells containing a weather station observation are compared to corresponding WRF grid cells (Fig. 3) suggesting this is not the case. Moreover, high-resolution (1–4 km) model configurations are still cold biased when the station-to-gridcell elevation differences are compared, and in such cases, the elevation mismatch between the coarse model grid topography and the true elevation grid is likely minimal. Other studies have found that T2m cold biases have persisted after applying elevation corrections (Kotlarski et al. 2014). Rasmussen et al. (2023) recently attributed WRF cold biases to well-known sensor calibration issues at SNOTEL stations (Oyler et al. 2015a). However, cold biases are also found in areas outside of the United States, and the RAWS network in the United States still provides information from high-elevation regions (Fig. 4d) that do not suffer the same systematic calibration issue. Therefore, it seems difficult to ascribe all model biases to T2m observing errors alone, though the case is certainly far from closed.

b. Snow representation in land surface models. Snow on the ground actively influences T2m across continental (Dutra et al. 2011) to local scales (Mott et al. 2018; Rudisill et al. 2021). As such, errors in modeled snowpack thermodynamics are suspected to be the cause of T2m model bias in many studies (Table 1). The atmospheric layer between the surface and the lowest model level can become very stable, and turbulent transfer coefficients become quite low (Lapo et al. 2019). In these conditions, the snow receives little heating from sensible heat flux and further decouples the snowpack from the atmosphere, causing excessively cold surface temperatures (Slater et al. 2001). Modeling highly stable boundary layers is a well-known challenge even for flat terrain (Sterk et al. 2013), and assumptions of canonical surface-layer exchange formulations are violated in complex terrain (Rotach et al. 2022, 2008; Serafin et al. 2018; Lehner and Rotach 2018). As such, errors in snow energy budgets, the stable atmospheric layers above them, and the formulae used to interpolate between skin temperature and the lowest model levels [Eq. (A1)] are very much connected and are areas that need further attention.

Gouttevin et al. (2022) investigated many of the hypotheses concerning the causes of T2m biases in Table 1 with the AROME model (Table A1). They used single-column and offline land surface model frameworks for a modeling domain with a 1.3-km grid spacing across the French Alps and concluded that the contribution from downwelling shortwave radiation is minimal and that most of the bias is caused by AROME exhibiting both too low wind speed errors and too low downwelling longwave radiation relative to observations. The remainder of the T2m bias is attributable to the surface-layer parameterization in clear-sky, low wind speed conditions. Whether or not these findings are applicable to other regions and models remains to be seen, but these results show how multiple observations that constrain the processes that control T2m enable straightforward diagnostics that identify the cause(s) of T2m model biases.

c. Model grid spacing. While finer model grid spacings should, in principle, better resolve the processes that control T2m and therefore reduce model T2m biases, the evidence collected by this study presents a more complicated picture. Orographic precipitation biases tend to decrease at finer horizontal grid spacings (Ikeda et al. 2010, 2021; Ban et al. 2021), as do, in principle, other climate variables (Lucas-Picher et al. 2021). However, Soares et al. (2022) and Careto et al. (2022) quantified the added value of finer grid spacings for T2m and found little improvement in bias by downscaling at 2–3 km compared to coarser 12–25-km simulations (Coppola et al. 2020).

VR-CESM simulations from Rhoades et al. (2018b) demonstrate that "all VR-CESM simulations exhibited a systemic mountain cold bias that worsened with elevation…" even when resolution is increased from 55 to 7 km, a finding recently corroborated by Wijngaard et al. (2023). WRF simulations with grid spacings on the order of 500 m–2 km (Xu et al. 2023;

Rudisill et al. 2022; Havens et al. 2019) do not appear to report less significant T2m biases than the 4-km L2017, though precise model configuration differences preclude direct comparison. Similarly, the cold biases in the COSMO Model (Table A1) are not mitigated going from grid spacings ranging from 22 to 2 km (Kotlarski et al. 2012; Ban et al. 2014; Winstral et al. 2019). Whether or not subkilometer simulations solve T2m biases remains to be seen. However, Vionnet et al. (2015) report that 0.25-km grid spacing improved high-elevation T2m biases but still performed poorly in valleys.

d. Other factors contributing to T2m bias. In addition to the previously mentioned factors that contribute to T2m biases, there are process representations within models that are known to be deficient. One known deficiency in models is associated with their representations of radiative fluxes in complex terrain (Fig. 1a). For shortwave radiation, terrain shadowing and surface reflection can lead to large and systematic differences in surface and intra-atmospheric shortwave radiative fluxes in complex terrain relative to a plane-parallel atmosphere and surface. The emission of longwave radiation by heterogeneous surfaces with variable skin temperatures can likewise lead to systematic differences in the downwelling longwave relative to a plane-parallel atmosphere and surface assumptions. A number of models do account for 3D terrain-related shortwave effects through straightforward parameterization (e.g., Lee et al. 2019; Steger et al. 2022), but fewer account for 3D longwave effects that likely contribute significantly to surface energy budgets (Plüss and Ohmura 1997; Zhu et al. 2017; Feldman et al. 2022). Missing 3D longwave effects could contribute to low-longwave biases that have been attributed to T2m biases in some studies (Table 1).

5. Case study: Diagnosing T2m biases with observations from the SAIL campaign

Resolving T2m model biases with a limited number of atmosphere and surface observations is ill posed because T2m is the manifestation of many unconstrained and interacting processes that impact the surface energy balance (Figs. 1a–c). However, the recently deployed U.S. Department of Energy Atmospheric Radiation Measurement program's SAIL campaign (Feldman et al. 2023) from 2021 to 2023 near Crested Butte, Colorado, and partner campaigns including NOAA's Study of Precipitation, the Lower Atmosphere and Surface for Hydrometeorology (SPLASH; de Boer et al. 2023), in Colorado's 300-km^2 ERW, provide the means to diagnose T2m biases with many more simultaneous observational constraints than are typically found in mountain watersheds. These include measurements of fractional cloud cover made using a hemispheric sky imager, precipitable water vapor retrievals using ground station microwave radiometry, snow skin temperature measurements using an infrared thermometer, and downwelling longwave radiation measured by a pyrgeometer. Other quantities were measured using standard meteorological instruments (additional descriptions are provided in Table A4). All of these observations were collected at the main SAIL facility located in an open clearing in the upper reach of the ERW valley (Fig. 5a; 38°57′22.35″N, 106°59′16.66″W at 2885 MSL). The main SAIL observing site is classified as a "valley" using the criteria employed previously (Figs. 3 and 4). Additional automated weather station data located throughout the ERW (Fig. 5) were used in this analysis, as were topographic elevation information at 50-m resolution that was acquired from an airborne lidar survey of the region (https://data.airbornesnowobservatories.com/).

Using SAIL and partner organization data, we examine T2m biases and their potential causes in the 3-km HRRR model version 4 (Benjamin et al. 2016) and the CFSR reanalysis-forced WRF configuration described in Xu et al. (2023), henceforth Xu-WRF. Xu-WRF has a 500-m innermost grid spacing which is used for analysis. This comparison is performed over the period 1 January–1 April 2022 during which the entire ERW is snow covered. Here, the zeroth hour of the HRRR forecast (also called the "analysis" in

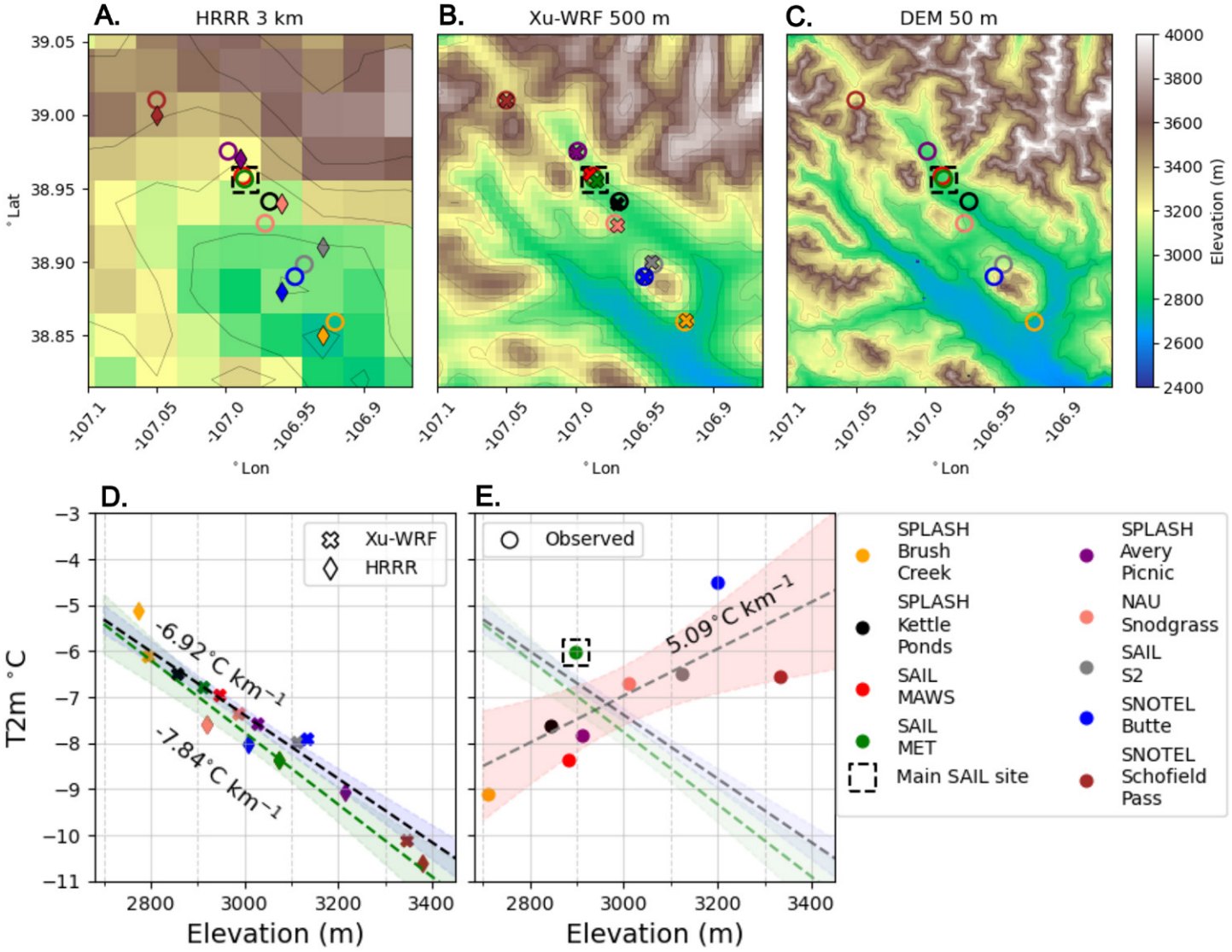

FIG. 5. Terrain elevation from the (a) 3-km HRRR model, (b) 500-m Xu-WRF Model, and (c) 50-m Airborne Snow Observatory Inc. DEM. The locations of nine temperature-observing stations are depicted by open circles. Filled markers (diamond and "x" for HRRR and Xu-WRF, respectively) show the location of model grid cells closest to the station. The dashed box indicates the location of the primary SAIL observing site. (d) The average T2m from 1 Jan to 1 Apr 2022 plotted against the corresponding terrain elevation for Xu-WRF and HRRR grid cells. The color of each marker denotes the station. The dashed lines depict linear regressions with 90% mean confidence intervals highlighted. (e) As in (d), but showing observed T2m. The HRRR and Xu-WRF linear regressions from (d) are replotted in the second panel for comparison purposes between the models and observations.

the operational community) is concatenated into a continuous time series for comparison against observations.

a. Elevational gradients of T2m in observations and models. Figure 5 illustrates the relationship between T2m and terrain elevation in both observations and models, recognizing that the elevation grid in the model contains coarsened topography. Observational data were visually inspected for anomalies prior to analysis, and both observational and model data span an elevation range of ~600 m. The grid cells from the Xu-WRF and HRRR models that are closest to each observation station were selected for analysis (Figs. 5a–c). Since the stations are close together, there are only six unique HRRR grid cells (but nine for Xu-WRF) that represent the nine station locations. The coarser HRRR model (Fig. 5a) shows more substantial offsets between the closest gridcell center and station location than in the Xu-WRF Model (Fig. 5b).

During the analysis period (1 January–1 April 2022), we find that the observed T2m actually increases with increasing terrain elevation on average (a temperature inversion).

Both Xu-WRF and HRRR fail to produce the observed T2m inversion; rather, the models exhibit elevational gradients in T2m that closely follow the mean environmental lapse rate of the standard atmosphere—lapse rates of 6.92° and 7.84°C km^{-1} for Xu-WRF and HRRR, respectively (Fig. 5d). As found by other studies (e.g., Lute and Abatzoglou 2021; Blandford et al. 2008), the mean environmental lapse rate proves to be a poor representation of the actual T2m–elevation relationship, and consequently, both models are cold biased at high-elevation stations (Butte, Schofield, and SAIL S2; Fig. 5e), but warm biased at most valley sites. This pattern matches Fig. 4 of Xu et al. (2023) that evaluates the same model against PRISM, as well as the patterns found across the entire WUS in L2017 (Fig. 3). It is important to keep in mind that the positive T2m elevational gradient from the observations spans a relatively short range of elevations, is confined to one particular valley, and would likely not be appropriate for larger spatial scales. Gridcell elevation mismatches play a small role in the spatial patterns of this bias, as the model gridcell elevations are within 100 m or less of the true elevation in the Xu-WRF Model, though the coarser HRRR model has some larger elevational deviations.

b. HRRR and reanalysis-driven WRF evaluation of T2m biases. To scrutinize the drivers of T2m biases, we examine covariate observations of cloud cover, precipitable water vapor, wind speed, and snow skin temperature at the main SAIL site (referred to as "MET" Fig. 5) located in the upper reaches of the ERW valley. Downwelling longwave radiation is also examined, as this has been suspected as a cause of T2m bias (Table 1).

HRRR and Xu-WRF T2m biases exhibit many common symptoms (Fig. 6). The biases are not always evenly distributed over the diurnal cycle and depend on background meteorological factors. For instance, Xu-WRF is cold biased by 2.1°C during the night hours and slightly warm biased (0.1°C) during the daytime (Fig. 6a). Meanwhile, HRRR is cold biased over 2.3°C for both periods (Fig. 6b). Both models show warm biases during the daytime when the skies are clear and wind speeds are low, but cold biases when there are even moderate cloud fractions

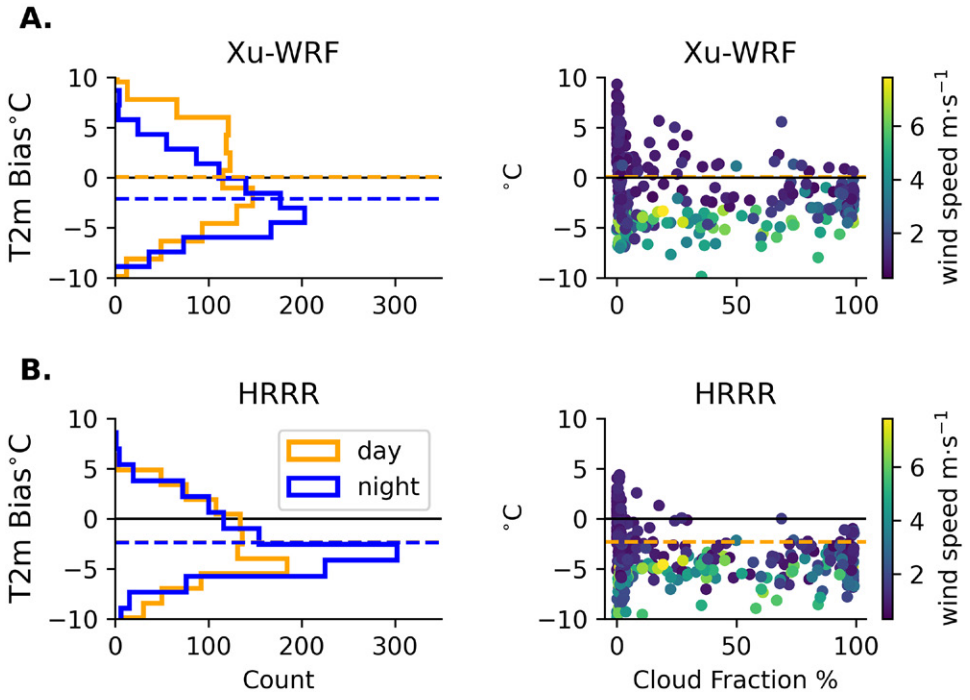

Fig. 6. Corresponding grid cells from the (a) Xu-WRF and (b) HRRR T2m biases computed against SAIL meteorological station T2m observations for day and night hours between 1 Jan and 1 Apr 2022. Horizontal lines show the mean values of each distribution. The day and night means of the HRRR bias distribution are visually indistinguishable. The right column depicts the relationship between the T2m bias, the daytime cloud fraction as measured by the SAIL total sky imager, and the observed wind speed. The solid black lines show the line of zero bias.

and moderate winds. These results show the models contain compensating errors and that a decomposition of T2m bias analysis between clear-sky and non-clear-sky days and day and night is likely warranted. Furthermore, the cold bias is uncorrelated with cloud fraction when cloud fraction exceeds 10%, which points to nonradiative processes causing T2m biases under most non-clear-sky conditions.

c. Diagnosing source(s) of T2m biases with T2m and skin temperature correlations. To investigate these relationships further, and the potential contribution of the surface-layer formulation to the observed T2m bias, we examine 2D histograms of skin temperature, T2m, and wind speed (Fig. 7). SAIL observations show that the land surface skin temperature and T2m are in reality more closely correlated (with linear model R^2 values of 0.81) than they are in either model (R^2 values of 0.46 and 0.63, respectively) especially at temperatures below −24°C. SAIL data also show that T2m almost never exceeds the skin temperature (Fig. 7a). Both models have a positive wind speed bias, which is a longstanding challenge in complex terrain (Jiménez and Dudhia 2012). The 500-m Xu-WRF Model appears to generate light winds better than the coarser HRRR model, possibly due to Xu-WRF having more realistic topography.

In reality, the weakest winds lead to the coldest observed temperatures (Fig. 7b). The 2D histograms and the locally estimated scatterplot smoothing curves (Cleveland 1979) of observations (shown as a dashed black line) indicate that the coldest temperatures occur with light winds (<2 m s^{-1}), while both models indicate that the coldest temperatures occur with moderate winds (~5 m s^{-1}). Whether or not the weak winds represent drainage flows that contribute cold

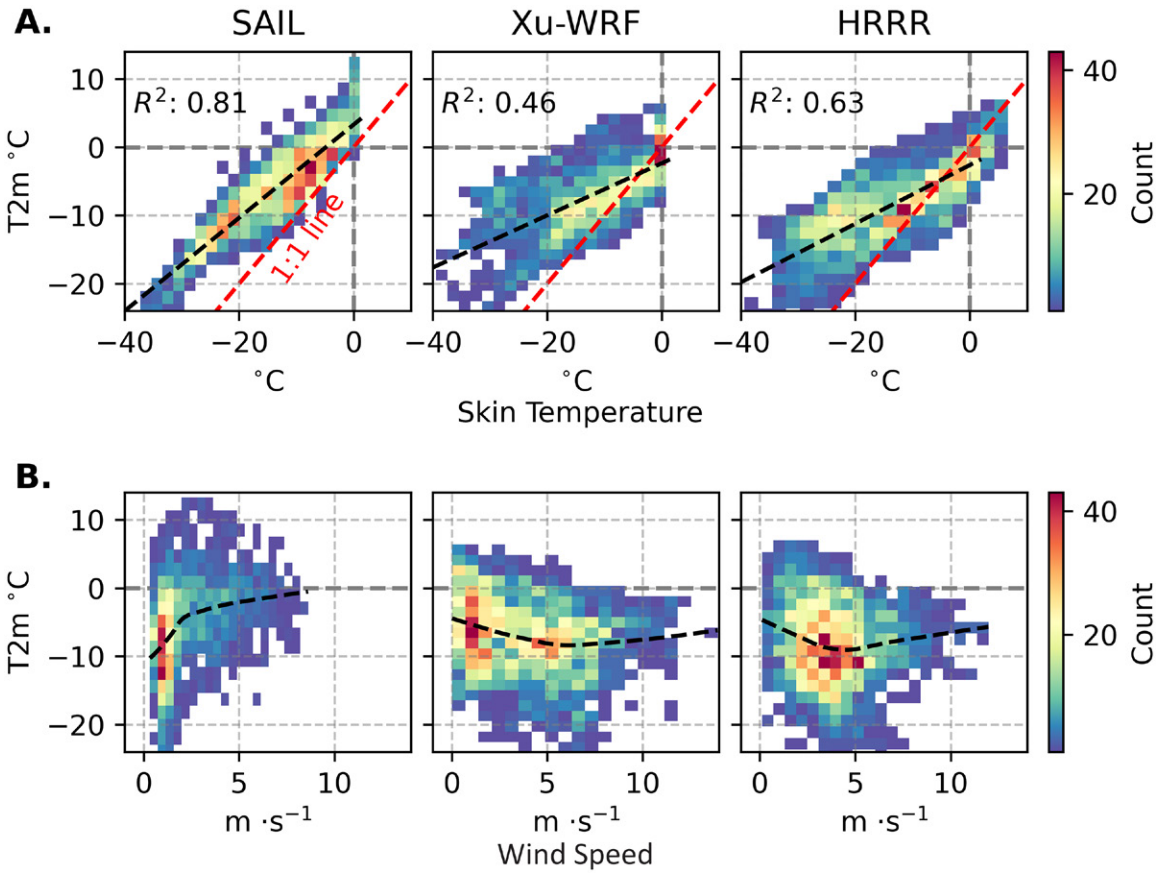

Fig. 7. The 2D histograms showing relationships between (a) T2m vs land skin temperature and (b) T2m vs near-surface wind speed for the HRRR model, Xu-WRF Model, and SAIL observations. Color count denotes the number of hourly data points from between 1 Jan and 1 Apr 2022 that fall within each bin. Gray dashed lines denote 0°C. In (a), the red lines denote the 1:1 line, the dashed black line shows a fitted linear regression model, and the R^2 is reported for that regression model. In (b), the curved black lines are locally estimated scatterplot smoothing regression lines.

air into the basin from upslope, or possibly that weak background winds simply inhibit the mixing of warm ambient air, allowing the cold pool to grow through radiative cooling (e.g., Clements et al. 2003) remains to be seen. Regardless of the cause, models do not capture the correlation structure between T2m and wind speed (Fig. 7b).

d. Diagnosing contribution(s) to T2m biases from humidity and longwave radiation biases. Downwelling LW (DLW) radiation is suspected as a cause of T2m bias (Table 1), so we examine the relationship between DLW, specific humidity at two meters, and column precipitable water vapor (Fig. 8). The net longwave radiation budget is a major component of the snow surface energy balance (Slater et al. 2001; Lapo et al. 2015) and influences valley T2m structure (Clements et al. 2003; Rauchöcker et al. 2023), and model errors in the atmospheric component of net longwave may cause, or be caused by, errors in other parameterized processes or state variables (e.g., the surface layer, planetary boundary layer, or biases in boundary conditions). We leverage the well-established functional relationships between DLW and near-surface humidity in cold, high-elevation environments (Rangwala 2013; Rangwala and Miller 2012) to evaluate model longwave biases and their causes during the SAIL period. Only 6-hourly profiles of precipitable water vapor were saved from the Xu-WRF data, but all other quantities are examined hourly. Empirically derived power-law curves between specific humidity and DLW for cloud-covered (top) and cloud-free conditions (bottom) developed previously by Naud et al. (2013), which represent a radiative transfer parameterization in complex terrain, are found to fit the SAIL observations well (black dashed curves; Fig. 8a). The SAIL site is very dry, but even still, the 2-m specific humidity is low biased in both models by 0.36 and 0.40 g kg^{-1} for Xu-WRF and HRRR, respectively (Fig. 8b), though the modeled total column precipitable water generally matched observations.

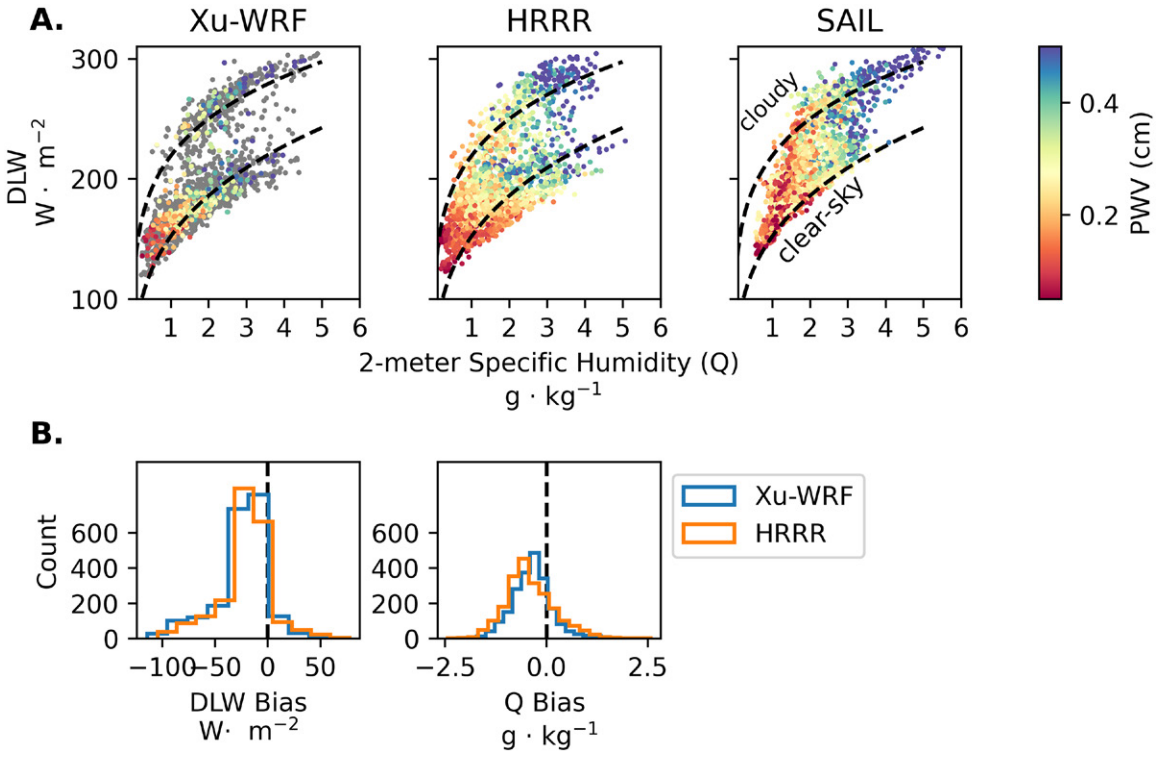

Fig. 8. (a) Relationship between 2-m specific humidity *Q* and DLW radiation. Empirical curves from Naud et al. (2013) are shown as dashed lines for (a) cloudy and (b) cloud-free conditions. Grid cells are colored by the modeled (Xu-WRF and HRRR) and observed (SAIL) PWV content (cm). All data points are hourly, but only 6-hourly precipitable water from Xu-WRF was available. Gray dots are shown at time steps where PWV observations are not available. (b) Downwelling LW and 2-m specific humidity biases for Xu-WRF and HRRR evaluated against observations from the main SAIL site.

The models, meanwhile, exhibit radiative transfer errors that produce specific humidity/DLW relationships that are too linear than what is observed, particularly during clear-sky conditions. Ultimately, the DLW is biased in both models by −20 and −24 W m^{-2}, with long tails of the distribution up to −100 W m^{-2}. This relationship serves as a useful metric to guide model development, since it indicates that the radiation and surface-layer schemes need to capture the observed, scale-independent relationship between DLW and humidity.

e. Transferability of SAIL observations. As with any field campaign, it is essential to establish the transferability of findings to other regions. Some indicators of transferability are already clear. Ultimately, a warm-valley and cold-ridge bias pattern found in L2017 (Fig. 3) across the WUS is confirmed in the ERW as well (Fig. 5). Both the Xu-WRF and HRRR models are cold biased significantly at the two nearby SNOTEL sites (also found in Rudisill et al. 2023). Given that SNOTELs are not generally located in valley sites (Fig. 4), it is possible that some model evaluations that only report biases at these locations have likely failed to report nearby warm biases in adjacent valleys (Rudisill et al. 2022; Havens et al. 2019), as the valley data are simply not present.

The SAIL data show that the model biases are related to a variety of compensating errors related to the surface layer, clouds, radiation, and winds. The magnitude and even the sign of the bias (either cold or warm) depend on elevation, time of day, and background meteorology. Therefore, it is crucial for model evaluators to keep in mind that T2m biases may not be monocausal nor purely a function of elevation alone. Model evaluators should first investigate whether the interactions between the PBL and surface schemes are producing wrong-signed T2m elevational gradients in high-elevation mountain valleys (Fig. 5), while the interrelationships between T2m, skin temperature, and winds suggest model developers should critically evaluate, using benchmarks and other simple cases, the surface-layer scheme performance under clear/calm conditions (Figs. 6 and 7). Negative downwelling longwave biases point to the need for radiative transfer parameterization updates (Fig. 8).

6. Discussion and conclusions

We have shown that MARMOTS of many types are cold biased in midlatitude mountain ranges, particularly during winter. The exact causes of cold biases currently remain unknown but are likely related to a combination of multiple, interacting parameterization deficiencies, and surface-layer exchange processes over snowpacks in particular. Model biases can depend on the physiographic classification (i.e., slope, ridge, or valley) suggesting a connection to cold-air pools and local wind circulations. In many cases, valleys are warm biased with ridges and slopes too cold (Figs. 3 and 5). As such, there are likely compensating error mechanisms that produce the ultimate T2m bias (Figs. 6 and 7).

a. The need for direct and ancillary observations of T2m in mountains. While the density of observations is typically quite low in mountains (1–4 stations per 1000 km^2 in the WUS and lesser in other mountain ranges) and biased toward lower-elevation and less-rugged terrain (Fig. 4), we have argued that cold biases are still true biases and not products of resolution mismatches or the assumptions introduced by reference datasets. Nevertheless, targeted observing campaigns can constrain processes and allow for systematic T2m bias diagnostics. Measuring T2m is easier to measure than other climate variables like snowfall (Rasmussen et al. 2012; Goodison et al. 1998) and can be done accurately with low-cost sensors (Hubbart et al. 2005; Minder et al. 2010). Deploying targeted, dense sensor networks and adequately sampling ridges, valleys, and slopes at the scale of a MARMOTS grid cell is one clear, though logistically challenging, target for model evaluation improvement. Engaging with winter recreationists and related industries, such as ski areas, avalanche forecasting

centers, and backcountry huts, could expand existing networks of T2m measurements in mountains. Gridded T2m reference datasets are the standard for model evaluation, but the associated uncertainties introduced by different gridding methodologies are not always clear. As such, intercomparisons of gridded datasets in mountains are warranted.

b. Suggestions for modelers: How to improve diagnoses of T2m bias. Modelers should recognize that finer horizontal model resolutions alone may not be sufficient to improve T2m in mountains (e.g., Soares et al. 2022). If biases are found in model results, it is important to first understand the limitations and assumptions made in reference datasets, particularly assumptions about T2m elevational gradients which typically decrease at a rate less than the standard free atmosphere (6.5°C km^{-1}), as well as issues related to data quality (e.g., Rasmussen et al. 2023). That understanding is best achieved through communication with those who collected/constructed those datasets and/or are very familiar with the observational information contained within those datasets and their strengths and limitations. Being aware of the details of how the data were collected, including when using the datasets is warranted or alternatively when caution must be exercised in using the datasets, is an essential prerequisite for model diagnostics and ultimately the model development that will be required to mitigate T2m biases. Differentiating biases by physiographic category (slope, valley, ridge), synoptic classification (strong/weak forcing, clear/cloudy), hour of day, and covariates with other meteorological variables (skin temperature, wind speed) will undoubtedly uncover information about the circumstances leading to T2m bias and further point to pathways for model and process parameterization improvement.

c. Looking forward: Cross-disciplinary collaborations are needed to improve mountain climate understanding. Improving T2m in models ultimately will require both specific attention to its root cause(s) in the mountains and a collaboration between atmospheric model developers, model users, and mountain climate and regional experts. Intensive field campaigns have been essential for moving the science forward (Smith 2019), and observing campaigns such as SAIL (Feldman et al. 2023) and the upcoming Multi-Scale Transport and Exchange Processes in the Atmosphere over Mountains–Programme and Experiment (TEAMx) (Rotach et al. 2022) campaigns show that observational density enables multivariate model diagnostics and preliminary determination of potential model development targets. The systematic evaluation of model parameterizations (such as radiation and the surface layer) through the use of single-column models (e.g., Massey et al. 2016; Bogenschutz et al. 2020; M. Huang et al. 2022; Gouttevin et al. 2022) is also warranted and likely essential for improving understanding and enhancing model's predictive powers in mountains.

Acknowledgments. This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, and the Atmospheric System Research Program under U.S. Department of Energy Contract DE-AC02-05CH11231. Coauthor Rhoades was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research through the Regional and Global Model Analysis (RGMA) Program under "the Calibrated and Systematic Characterization, Attribution and Detection of Extremes (CASCADE)" Science Focus Area (Award DE-AC02-05CH11231) and "An Integrated Evaluation of the Simulated Hydroclimate System of the Continental US" project (Award DE-SC0016605).

Data availability statement. The SAIL data used in this study are publicly available from https://adc.arm.gov/discovery/#/. WRF data from the Liu et al. (2017) study are publicly available from https://rda.ucar.edu/datasets/ds612.0/. The 4-km version of the PRISM dataset is publicly available from https://prism.oregonstate.edu/. HRRR forecast data were downloaded using the Herbie library (https://doi.org/10.5281/zenodo.4567540).

APPENDIX

Additional Information for Interpreting MARMOTS T2m Biases

a. How is T2m diagnosed in models? T2m is typically a diagnostic quantity, rather than prognostic quantity, as the lowest model level of an atmospheric model is typically higher than 2 m above the land surface. In the WRF Model, T2m is diagnosed using the following equation:

$$\theta_{2\mathrm{m}} = \theta_{z_\theta} + (\theta_{z_\theta} - \theta_z)\frac{\ln\left(\frac{2}{z_\theta}\right) - \psi\left(\frac{2}{L}\right)}{\ln\left(\frac{z}{z_\theta}\right) - \psi\left(\frac{z}{L}\right)}, \tag{A1}$$

where θ_{z0} is the potential temperature at the ground (surface) level, θ_z is the potential temperature at the lowest atmospheric level z, $z0$ is the roughness length, L is the Obukhov length (Jiménez et al. 2012), and ψ is the similarity function from Monin–Obukhov theory, the shape of which depends on the stability regime and the specific formulation employed (Brutsaert 2013). Thus, errors in T2m depend on errors of temperature at the lowest model level, errors of the skin temperature of the ground/snow surface, and errors in the interpolation formula that depends on surface-layer parameterizations [Eq. (A1)], all of which are related and depend on the near-surface energy budget, surface roughness, and meteorological conditions.

b. Tables. Tables A1–A3 describe the models examined in this review, specific model studies that have reported cold biases, and the gridded temperature datasets used to evaluate those models. Direct quotes about biases are provided in Table A2 if they are available. Table A4 describes the SAIL instrument data used to produce in Figs. 5–8 in the main text.

Table A1. MARMOTS evaluated in this study. Some models are part of larger CORDEX ensembles or may be part of studies that evaluate single-model configurations. LAM = limited-area model; RRM = regionally refined mesh.

Model	Grid type	Citation
AROME	LAM	Seity et al. (2011)
ARPEGE	LAM	Roehrig et al. (2020)
CanRCM4	LAM	Scinocca et al. (2016)
COSMO	LAM	Schättler et al. (2021)
COSMO-CLM (CCLM)	LAM	https://www.clm-community.eu/
CRCM5	LAM	Zadra et al. (2008)
GEM-LAM	LAM	Zadra et al. (2008)
HiRAM5	LAM	Christensen et al. (2007)
HRRR	LAM	Benjamin et al. (2016)
RACMO	LAM	van Meijgaard et al. (2008)
RCA4	LAM	Jones et al. (2011)
RegCM4	LAM	Giorgi et al. (2012)
REMO	LAM	Jacob et al. (2012)
RRM-E3SM	RRM	Tang et al. (2023)
SIMA-MPAS	RRM	Skamarock et al. (2012), X. Huang et al. (2022)
VR-CESM	RRM	Zarzycki et al. (2014)
WRF	LAM	Powers et al. (2017), Skamarock et al. (2019)

Table A2. MARMOTS studies examined in this review and used to produce Fig. 2. The model platform, geographic region, grid spacing, and reference dataset used to compute the T2m bias are listed. Direct quotes from the paper describing cold biases are given in the notes columns when applicable. Otherwise, a description/interpretation of the cold bias is provided. Table A3 in the appendix lists the study for the reference dataset used to identify the model T2m bias. Otherwise, "stations" refers to comparing single grid cells against weather station observations.

Study	Model	Region	Horiz. grid spacing	Reference dataset	Notes
Ma et al. (2023)	WRF	Tibetan Plateau	9 km	Stations	Cold bias of 4.5°C in southern Tibetan Plateau
Huang et al. (2018)	WRF	Sierra Nevada	9 km	Unstated (cites previous evaluation study)	"A slight cold bias exists in WRF at higher elevations, partly due to the sparse observations over mountain peaks"
Yang et al. (2021)	WRF	Tien Shan	9 km	Stations	T2m bias has a range between +2° and −6°C, with a mean below zero (a cold bias) for each region examined
Walton and Hall (2018)	WRF	Sierra Nevada	9 km	Multiple gridded temperature datasets	Model T2m is colder across the entire domain, but particularly in the Sierra Nevada
Kawase et al. (2013)	WRF	Japan	4.5 km	Stations	"Low temperature bias around 1.0 to 1.5 K. In particular, the minimum temperature below 0°C has the lower temperature bias"
Liu et al. (2017)	WRF	CONUS	4 km	PRISM	Cold bias over mountains in the WUS
He et al. (2019)	WRF	CONUS	4 km	Multiple	Uses the same data as Liu et al. (2017); finds cold biases over mountain ranges particularly in the interior WUS for one year of evaluation
Scalzitti et al. (2016)	WRF	Western Rocky Mountains	4 km	PRISM	Cold bias of 5°C
Pan et al. (2011)	WRF	Sierra Nevada	4 km	PRISM	Cold bias in maximum temperatures (Fig. 5 of their paper)
Mooney et al. (2020)	WRF	Scandinavian Mountains	3 km	E-OBS17; seNorge	Cold bias of approximately 2°C and stronger in winter compared against E-OBS17; possible warm bias when compared against seNorge product
Kropač et al. (2021)	WRF	Southern Alps	2 km	Stations	Cold bias and wind speed overstimulation at a glacier observing site
Havens et al. (2019)	WRF	Idaho–Bitterroot Mountains	2 km	Stations	Cold bias at lower temperatures compared against NRCS SNOTEL during winter
Karki et al. (2017)	WRF	Himalaya	1 km	Stations	"Winter and pre-monsoon seasons feature a high cold bias for high elevations while lower elevations show a simultaneous warm bias"
Rudisill et al. (2022)	WRF	Idaho–Bitterroot Mountains	1 km	Stations	Cold bias of 1°C compared against NRCS SNOTEL during winter
Rudisill et al. (2023)	WRF	Rockies	1 km	Stations	Cold bias of 1°–3°C compared against NRCS SNOTEL during winter
Wang et al. (2018)	WRF	Interior WUS Mountains	4 km	PRISM	"A cold bias of 4°C is observed over high mountain ridges in winter"
Vionnet et al. (2016)	AROME	European Alps	2.5 km	Stations	Cold bias of 1.5°C increasing with elevation
Gouttevin et al. (2022)	AROME	European Alps	1.3 km	Stations	Cold biases attributed to LW radiation, wind, and turbulence parameterizations
Monteiro et al. (2022)	AROME	European Alps	2.5 km	S2M	Cold bias increases with elevation and during winter. Convection-permitting resolution does improve bias, but it still remains.
Kotlarski et al. (2012)	COSMO	European Alps	22 km	E-OBS17	Cold bias "Strongly negative above 1000 m"
Winstral et al. (2019)	COSMO	European Alps	2.2 km	Stations	Cold bias of 1°–2°C

(*Continued*)

Table A2. (*Continued*).

Study	Model	Region	Horiz. grid spacing	Reference dataset	Notes
Ban et al. (2014)	COSMO	European Alps	12, 2 km	E-OBS17	"In the winter season, a weak cold bias is found over most of the domain in both models"
Vionnet et al. (2015)	GEM-LAM	Canadian Rockies	1–0.25 km	Stations	0.5°C cold bias at high elevations. Finer model resolution improves temperatures on peaks, but cold pools still do not form well in valleys.
Kotlarski et al. (2015)	ENSEMBLES	Alps, Pyrenees	25 km	E-OBS17	Models show "an overestimated temperature decrease with elevation"
Vautard et al. 2021	EURO-CORDEX	Europe	0.11°	E-OBS17	Winter cold bias in the Alps and Scandinavian Mountains
Kotlarski et al. (2014)	EURO-CORDEX	Europe	0.11°	E-OBS17	Winter cold bias in the Alps and Scandinavian Mountains
Matiu et al. (2019)	EURO-CORDEX	Italian Alps	0.11°	E-OBS17	Winter cold bias in the Italian Alps
Smiatek et al. (2016)	EURO-CORDEX	Italian Alps	0.11°	E-OBS17	Winter cold bias in the Italian Alps (Fig. 2 of their paper)
Soares et al. (2022)	CORDEX-FPS	Alps	3–4 km	E-OBS17	Higher model resolution does not improve biases in the Alps
Blázquez and Solman (2023)	SA-CORDEX	Andes	0.44°–0.22°	CPC	Cold bias in DJF and JJA (Figs. 1 and 2)
Solman et al. (2013)	SA-CORDEX	Andes	0.44°	UDEL, CRU	2°C cold bias
Torrez-Rodriguez et al. (2023)	CORDEX-CORE	Chilean Andes	10–14 km	CR2METv1	"RCMs and ERA5 have a common prominent cold bias at high elevations"
Xu et al. (2019)	NA-CORDEX	Sierra Nevada	44–11 km	PRISM	Cold bias across models and resolutions
McCrary et al. (2017)	NARCCAP/ NA-CORDEX	Rockies	0.5°	UDEL	Areal averaged cold bias in three of six MARMOTS experiments in the Southern Rockies region
Di Virgilio et al. (2019)	CORDEX-Australasia	Australian Alps	0.44°	AGCD	Large cold bias across most of Australia; possibly more significant in the Eastern Australian Highlands
Rhoades et al. (2018b)	VR-CESM	Sierra Nevada	55–9 km	PRISM	Increasing cold bias with elevation across multiple model grid spacings and configurations
Wu et al. (2017)	VR-CESM	Northern Rocky Mountains	0.125°	PRISM, stations	Winter cold bias in Tmin of 1°–6°C
Bambach et al. (2022)	VR-CESM	Andes	14 km	UDEL	Cold bias in the Andes during winter
Xu et al. (2021)	VR-CESM	Tibetan Plateau; WUS	0.125°	PRISM	Cold bias in the Tibetan Plateau and WUS
Rahimi et al. (2019)	VR-CESM and WRF	Tibetan Plateau	0.1° and 12 km	AIRS, APHRODITE	Both WRF and VR-CESM are cold biased in the Tibetan Plateau during DJF
Wijngaard et al. (2023)	VR-CESM	High Mountain Asia	7 km	WFDEI	"Western HMA subregions have cold temperature biases during winter; these biases grow with increasing resolution"
X. Huang et al. (2022)	SIMA-MPAS	WUS	3–4 km	PRISM	SIMA-MPAS is cold biased "over very high mountain top ranges" (Fig. 8)
Torma et al. (2011)	RegCM3	Carpathians	10 km	E-OBS17; Additional station data	"A systematic cold bias is found over the mountainous areas of the alpine region and the coastal Balkans"
Winter et al. (2017)	Multiple RCMs	Alps	25 km	Stations	"Cold temperature biases at high elevations seem to contribute to a "snow albedo feedback" overestimation in several RCMs"

Table A3. Gridded T2m reference datasets used in the studies examined in Table A2.

Abbreviation	Horizontal resolution	Extent	Reference
AGCD	0.05°	Australia	Jones et al. (2009)
APHRODITE	0.5°–0.25°	Asia	Yatagai et al. (2012)
CR2METv1	0.05°	South America	Alvarez-Garreton et al. (2018)
CRU	0.5°	Global	Mitchell and Jones (2005)
Daymet	1 km	United States	Thornton et al. (2016)
E-OBS17	0.25°	Europe	Haylock et al. (2008)
ERA5-Land	9 km	Global	Muñoz-Sabater et al. (2021)
Hiebl et al.	1 km	Europe	Hiebl et al. (2009)
PRISM	800 m–4 km	United States	Daly et al. (2008)
UDEL	0.5°	Global	Matsuura and Willmott (2009)
S2m	Multiple	Europe	Vernay et al. (2022)
seNorge	1 km	Scandinavia	Lussana et al. (2019)
WFDEI	0.5°	Global	Weedon et al. (2014)
WorldClim	1 km	Global	Fick and Hijmans (2017)

Table A4. SAIL and partner organization instrumentation used to evaluate Xu-WRF and HRRR models.

Instrument	Quantity	Instrument citation	Campaign or organization
Microwave radiometer	Atmospheric PWV (cm)	Morris (2019)	SAIL
SAIL Automated weather stations	T2m, wind speed, relative humidity, barometric pressure	Ritsche (2011)	SAIL
Ground facing IR thermometer	Snow surface temperature	Morris (2018)	SAIL
Surface energy balance system	Downwelling broadband SW irradiance, downwelling broadband SW irradiance	Cook and Sullivan (2019)	SAIL
Total sky imager	Percent of the upward-facing sky view that is covered by opaque clouds	Morris (2005)	SAIL
SNOTEL temp	T2m	Serreze et al. (1999)	NRCS
Northern Arizona University (NAU) automated weather station	T2m	Simonpietri et al. (2022)	NAU

c. List of Abbreviations.

CONUS	Contiguous United States
CORDEX	Coordinated Regional Climate Downscaling Experiment
DLW	Downwelling longwave
EURO-CORDEX	European Coordinated Regional Downscaling Experiment
ERW	East River watershed
HRRR	High-Resolution Rapid Refresh model
MARMOTS	Models of Applicable Resolution for Mountain Meteorology across Time Scales
NA-CORDEX	North American Coordinated Regional Downscaling Experiment
PRISM	Parameter-Elevation Regressions on Independent Slopes Model
PWV	Precipitable water vapor
SA-CORDEX	South American Coordinated Regional Downscaling Experiment
SAIL	Surface Atmosphere Integrated Field Laboratory
T2m	Two-meter air temperature
VR-CESM	Variable-resolution Community Earth System Model
WRF	Weather Research and Forecasting Model
WUS	Western United States of America

References

Alvarez-Garreton, C., and Coauthors, 2018: The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies – Chile dataset. *Hydrol. Earth Syst. Sci.*, **22**, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018.

Armstrong, R. L., and E. Brun, 2008: *Snow and Climate: Physical Processes, Surface Energy Exchange and Modeling.* Cambridge University Press, 222 pp.

Bambach, N. E., A. M. Rhoades, B. J. Hatchett, A. D. Jones, P. A. Ullrich, and C. M. Zarzycki, 2022: Projecting climate change in South America using variable-resolution Community Earth System Model: An application to Chile. *Int. J. Climatol.*, **42**, 2514–2542, https://doi.org/10.1002/joc.7379.

Ban, N., J. Schmidli, and C. Schär, 2014: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *J. Geophys. Res. Atmos.*, **119**, 7889–7907, https://doi.org/10.1002/2014JD021478.

——, and Coauthors, 2021: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, Part I: Evaluation of precipitation. *Climate Dyn.*, **57**, 275–302, https://doi.org/10.1007/s00382-021-05708-w.

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.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Blandford, T. R., K. S. Humes, B. J. Harshburger, B. C. Moore, V. P. Walden, and H. Ye, 2008: Seasonal and synoptic variations in near-surface air temperature lapse rates in a mountainous basin. *J. Appl. Meteor. Climatol.*, **47**, 249–261, https://doi.org/10.1175/2007JAMC1565.1.

Blázquez, J., and S. A. Solman, 2023: Temperature and precipitation biases in CORDEX RCM simulations over South America: Possible origin and impacts on the regional climate change signal. *Climate Dyn.*, **61**, 2907–2920, https://doi.org/10.1007/s00382-023-06727-5.

Bogenschutz, P. A., S. Tang, P. M. Caldwell, S. Xie, W. Lin, and Y.-S. Chen, 2020: The E3SM version 1 single-column model. *Geosci. Model Dev.*, **13**, 4443–4458, https://doi.org/10.5194/gmd-13-4443-2020.

Brutsaert, W., 2013: *Evaporation into the Atmosphere: Theory, History and Applications.* Springer Science & Business Media, 302 pp.

Burakowski, E., A. Tawfik, A. Ouimette, L. Lepine, K. Novick, S. Ollinger, C. Zarzycki, and G. Bonan, 2018: The role of surface roughness, albedo, and Bowen ratio on ecosystem energy balance in the eastern United States. *Agric. For. Meteor.*, **249**, 367–376, https://doi.org/10.1016/j.agrformet.2017.11.030.

Careto, J. A. M., P. M. M. Soares, R. M. Cardoso, S. Herrera, and J. M. Gutiérrez, 2022: Added value of EURO-CORDEX high-resolution downscaling over the Iberian Peninsula revisited—Part 2: Max and min temperature. *Geosci. Model Dev.*, **15**, 2653–2671, https://doi.org/10.5194/gmd-15-2653-2022.

Christensen, O. B., M. Drews, J. H. Christensen, K. Dethloff, K. Ketelsen, I. Hebestadt, and A. Rinke, 2007: The HIRHAM Regional Climate Model. Version 5 (beta). Danish Meteorological Institute Tech. Rep. 06-17, 23 pp., https://www.dmi.dk/dmi/tr06-17.

Clements, C. B., C. D. Whiteman, and J. D. Horel, 2003: Cold-air-pool structure and evolution in a mountain basin: Peter Sinks, Utah. *J. Appl. Meteor.*, **42**, 752–768, https://doi.org/10.1175/1520-0450(2003)042<0752:CSAEIA>2.0.CO;2.

Cleveland, W. S., 1979: Robust locally weighted regression and smoothing scatterplots. *J. Amer. Stat. Assoc.*, **74**, 829–836, https://doi.org/10.1080/01621459.1979.10481038.

Cook, D. R., and R. C. Sullivan, 2019: Surface Energy Balance System (SEBS) Instrument Handbook. R. Stafford, Ed., ARM Climate Research Facility, DOE/SC-ARM/TR-092, 16 pp., https://doi.org/10.2172/1004944.

Coppola, E., and Coauthors, 2020: A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean. *Climate Dyn.*, **55**, 3–34, https://doi.org/10.1007/s00382-018-4521-8.

Cosgrove, B. A., and Coauthors, 2003: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res.*, **108**, 8842, https://doi.org/10.1029/2002JD003118.

Currier, W. R., T. Thorson, and J. D. Lundquist, 2017: Independent evaluation of frozen precipitation from WRF and PRISM in the Olympic Mountains. *J. Hydrometeor.*, **18**, 2681–2703, https://doi.org/10.1175/JHM-D-17-0026.1.

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.

——, D. R. Conklin, and M. H. Unsworth, 2010: Local atmospheric decoupling in complex topography alters climate change impacts. *Int. J. Climatol.*, **30**, 1857–1864, https://doi.org/10.1002/joc.2007.

de Boer, G., and Coauthors, 2023: Supporting advancement in weather and water prediction in the Upper Colorado River basin: The SPLASH Campaign. *Bull. Amer. Meteor. Soc.*, **104**, 1853–1874, https://doi.org/10.1175/BAMS-D-22-0147.1.

Demory, M.-E., P. L. Vidale, M. J. Roberts, P. Berrisford, J. Strachan, R. Schiemann, and M. S. Mizielinski, 2014: The role of horizontal resolution in simulating drivers of the global hydrological cycle. *Climate Dyn.*, **42**, 2201–2225, https://doi.org/10.1007/s00382-013-1924-4.

Di Virgilio, G., and Coauthors, 2019: Evaluating reanalysis-driven CORDEX regional climate models over Australia: Model performance and errors. *Climate Dyn.*, **53**, 2985–3005, https://doi.org/10.1007/s00382-019-04672-w.

Dobrowski, S. Z., J. T. Abatzoglou, J. A. Greenberg, and S. G. Schladow, 2009: How much influence does landscape-scale physiography have on air temperature in a mountain environment? *Agric. For. Meteor.*, **149**, 1751–1758, https://doi.org/10.1016/j.agrformet.2009.06.006.

Dodson, R., and D. Marks, 1997: Daily air temperature interpolated at high spatial resolution over a large mountainous region. *Climate Res.*, **8**, 1–20, https://doi.org/10.3354/cr008001.

Dutra, E., C. Schär, P. Viterbo, and P. M. A. Miranda, 2011: Land-atmosphere coupling associated with snow cover. *Geophys. Res. Lett.*, **38**, L15707, https://doi.org/10.1029/2011GL048435.

Feldman, D. R., M. Worden, N. Falco, P. J. Dennedy-Frank, J. Chen, B. Dafflon, and H. Wainwright, 2022: Three-dimensional surface downwelling longwave radiation clear-sky effects in the Upper Colorado River basin. *Geophys. Res. Lett.*, **49**, e2021GL094605, https://doi.org/10.1029/2021GL094605.

——, and Coauthors, 2023: The Surface Atmosphere Integrated Field Laboratory (SAIL) campaign. *Bull. Amer. Meteor. Soc.*, **12**, E2192–E2222, https://doi.org/10.1175/BAMS-D-22-0049.1.

Fernández, A., V. Schumacher, I. Ciocca, A. Rifo, A. A. Muñoz, and F. Justino, 2021: Validation of a 9-km WRF dynamical downscaling of temperature and precipitation for the period 1980–2005 over Central South Chile. *Theor. Appl. Climatol.*, **143**, 361–378, https://doi.org/10.1007/s00704-020-03416-9.

Fick, S. E., and R. J. Hijmans, 2017: WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*, **37**, 4302–4315, https://doi.org/10.1002/joc.5086.

Gangopadhyay, S., C. A. Woodhouse, G. J. McCabe, C. C. Routson, and D. M. Meko, 2022: Tree rings reveal unmatched 2nd century drought in the Colorado River basin. *Geophys. Res. Lett.*, **49**, e2022GL098781, https://doi.org/10.1029/2022GL098781.

Giorgi, F., and Coauthors, 2012: RegCM4: Model description and preliminary tests over multiple CORDEX domains. *Climate Res.*, **52**, 7–29, https://doi.org/10.3354/cr01018.

Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation measurement intercomparison. WMO/TD 872, 318 pp., https://library.wmo.int/idurl/4/28336.

Gouttevin, I., V. Vionnet, Y. Seity, A. Boone, M. Lafaysse, Y. Deliot, and H. Merzisen, 2022: To the origin of a wintertime screen-level temperature bias at high

altitude in a kilometric NWP model. *J. Hydrometeor.*, **24**, 53–71, https://doi.org/10.1175/JHM-D-21-0200.1.

Gutmann, E. D., R. M. Rasmussen, C. Liu, K. Ikeda, D. J. Gochis, M. P. Clark, J. Dudhia, and G. Thompson, 2012: A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. *J. Climate*, **25**, 262–281, https://doi.org/10.1175/2011JCLI4109.1.

Gutowski, W. J., and Coauthors, 2020: The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information. *Bull. Amer. Meteor. Soc.*, **101**, E664–E683, https://doi.org/10.1175/BAMS-D-19-0113.1.

Harpold, A. A., M. L. Kaplan, P. Z. Klos, T. Link, J. P. McNamara, S. Rajagopal, R. Schumer, and C. M. Steele, 2017: Rain or snow: Hydrologic processes, observations, prediction, and research needs. *Hydrol. Earth Syst. Sci.*, **21** (1), 1–22, https://doi.org/10.5194/hess-21-1-2017.

Havens, S., D. Marks, K. FitzGerald, M. Masarik, A. N. Flores, P. Kormos, and A. Hedrick, 2019: Approximating input data to a snowmelt model using Weather Research and Forecasting Model outputs in lieu of meteorological measurements. *J. Hydrometeor.*, **20**, 847–862, https://doi.org/10.1175/JHM-D-18-0146.1.

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, https://doi.org/10.1029/2008JD010201.

He, C., F. Chen, M. Barlage, C. Liu, A. Newman, W. Tang, K. Ikeda, and R. Rasmussen, 2019: Can convection-permitting modeling provide decent precipitation for offline high-resolution snowpack simulations over mountains? *J. Geophys. Res. Atmos.*, **124**, 12 631–12 654, https://doi.org/10.1029/2019JD030823.

Heggli, A., B. Hatchett, A. Schwartz, T. Bardsley, and E. Hand, 2022: Toward snowpack runoff decision support. *iScience*, **25**, 104240, https://doi.org/10.1016/j.isci.2022.104240.

Hiebl, J., and Coauthors, 2009: A high-resolution 19611990 monthly temperature climatology for the greater Alpine region. *Meteor. Z.*, **18**, 507–530, https://doi.org/10.1127/0941-2948/2009/0403.

Hock, R., and Coauthors, 2019: High mountain areas. *Ocean and Cryosphere in a Changing Climate*, H.-O. Pörtner et al., Eds., Cambridge University Press, 131–202.

Horel, J., and Coauthors, 2002: MesoWest: Cooperative mesonets in the western United States. *Bull. Amer. Meteor. Soc.*, **83**, 211–226, https://doi.org/10.1175/1520-0477(2002)083<0211:mcmitw>2.3.co;2.

Huang, M., P.-L. Ma, N. W. Chaney, D. Hao, G. Bisht, M. D. Fowler, V. E. Larson, and L. R. Leung, 2022: Representing surface heterogeneity in land–atmosphere coupling in E3SMv1 single-column model over ARM SGP during summertime. *Geosci. Model Dev.*, **15**, 6371–6384, https://doi.org/10.5194/gmd-15-6371-2022.

Huang, X., A. D. Hall, and N. Berg, 2018: Anthropogenic warming impacts on today's Sierra Nevada snowpack and flood risk. *Geophys. Res. Lett.*, **45**, 6215–6222, https://doi.org/10.1029/2018GL077432.

——, A. Gettelman, W. C. Skamarock, P. H. Lauritzen, M. Curry, A. Herrington, J. T. Truesdale, and M. Duda, 2022: Advancing precipitation prediction using a new-generation storm-resolving model framework–SIMA-MPAS (V1.0): A case study over the western United States. *Geosci. Model Dev.*, **15**, 8135–8151, https://doi.org/10.5194/gmd-15-8135-2022.

Hubbart, J., T. Link, C. Campbell, and D. Cobos, 2005: Evaluation of a low-cost temperature measurement system for environmental applications. *Hydrol. Processes*, **19**, 1517–1523, https://doi.org/10.1002/hyp.5861.

Ikeda, K., and Coauthors, 2010: Simulation of seasonal snowfall over Colorado. *Atmos. Res.*, **97**, 462–477, https://doi.org/10.1016/j.atmosres.2010.04.010.

——, and Coauthors, 2021: Snowfall and snowpack in the western U.S. as captured by convection permitting climate simulations: Current climate and pseudo global warming future climate. *Climate Dyn.*, **57**, 2191–2215, https://doi.org/10.1007/s00382-021-05805-w.

Immerzeel, W. W., L. Petersen, S. Ragettli, and F. Pellicciotti, 2014: The importance of observed gradients of air temperature and precipitation for modeling runoff from a glacierized watershed in the Nepalese Himalayas. *Water Resour. Res.*, **50**, 2212–2226, https://doi.org/10.1002/2013WR014506.

——, and Coauthors, 2020: Importance and vulnerability of the world's water towers. *Nature*, **577**, 364–369, https://doi.org/10.1038/s41586-019-1822-y.

Jacob, D., and Coauthors, 2012: Assessing the transferability of the regional climate model REMO to different COordinated Regional Climate Downscaling EXperiment (CORDEX) regions. *Atmosphere*, **3**, 181–199, https://doi.org/10.3390/atmos3010181.

Jennings, K. S., T. S. Winchell, B. Livneh, and N. P. Molotch, 2018: Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere. *Nat. Commun.*, **9**, 1148, https://doi.org/10.1038/s41467-018-03629-7.

Jiménez, P. A., and J. Dudhia, 2012: Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. *J. Appl. Meteor. Climatol.*, **51**, 300–316, https://doi.org/10.1175/JAMC-D-11-084.1.

——, ——, J. Fidel González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, 2012: A revised scheme for the WRF surface layer formulation. *Mon. Wea. Rev.*, **140**, 898–918, https://doi.org/10.1175/MWR-D-11-00056.1.

Jones, C. G., P. Samuelsson, and E. Kjellström, 2011: Regional climate modelling at the Rossby Centre. *Tellus*, **63A** (1), 1–3, https://doi.org/10.1111/j.1600-0870.2010.00491.x.

Jones, D., W. Wang, and R. Fawcett, 2009: High-quality spatial climate datasets for Australia. *Aust. Meteor. Oceanogr. J.*, **58**, 233–248, https://doi.org/10.22499/2.5804.003.

Kapos, V., J. Rhind, M. Edwards, M. F. Price, and C. Ravilious, 2000: Developing a map of the world's mountain forests. *Forests in Sustainable Mountain Development: A State of Knowledge Report for 2000*, Task Force on Forests in Sustainable Mountain Development, CABI Publishing, 4–19.

Karki, R., S. ul Hasson, L. Gerlitz, U. Schickhoff, T. Scholten, and J. Böhner, 2017: Quantifying the added value of convection-permitting climate simulations in complex terrain: A systematic evaluation of WRF over the Himalayas. *Earth Syst. Dyn.*, **8**, 507–528, https://doi.org/10.5194/esd-8-507-2017.

Kattel, D. B., T. Yao, K. Yang, L. Tian, G. Yang, and D. Joswiak, 2013: Temperature lapse rate in complex mountain terrain on the southern slope of the central Himalayas. *Theor. Appl. Climatol.*, **113**, 671–682, https://doi.org/10.1007/s00704-012-0816-6.

Kawase, H., M. Hara, T. Yoshikane, N. N. Ishizaki, F. Uno, H. Hatsushika, and F. Kimura, 2013: Altitude dependency of future snow cover changes over Central Japan evaluated by a regional climate model. *J. Geophys. Res. Atmos.*, **118**, 12 444–12 457, https://doi.org/10.1002/2013JD020429.

Körner, C., J. Paulsen, and E. M. Spehn, 2011: A definition of mountains and their bioclimatic belts for global comparisons of biodiversity data. *Alp. Bot.*, **121**, 73–78, https://doi.org/10.1007/s00035-011-0094-4.

——, W. Jetz, J. Paulsen, D. Payne, K. Rudmann-Maurer, and E. M. Spehn, 2017: A global inventory of mountains for bio-geographical applications. *Alp. Bot.*, **127** (1), 1–15, https://doi.org/10.1007/s00035-016-0182-6.

Kotlarski, S., T. Bosshard, D. Lüthi, P. Pall, and C. Schär, 2012: Elevation gradients of European climate change in the regional climate model COSMO-CLM. *Climatic Change*, **112**, 189–215, https://doi.org/10.1007/s10584-011-0195-5.

——, and Coauthors, 2014: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci. Model Dev.*, **7**, 1297–1333, https://doi.org/10.5194/gmd-7-1297-2014.

——, D. Lüthi, and C. Schär, 2015: The elevation dependency of 21st century European climate change: An RCM ensemble perspective. *Int. J. Climatol.*, **35**, 3902–3920, https://doi.org/10.1002/joc.4254.

Kropač, E., T. Mölg, N. J. Cullen, E. Collier, C. Pickler, and J. V. Turton, 2021: A detailed, multi-scale assessment of an atmospheric river event and its impact on extreme glacier melt in the southern alps of New Zealand. *J. Geophys. Res. Atmos.*, **126**, e2020JD034217, https://doi.org/10.1029/2020JD034217.

Lapo, K., B. Nijssen, and J. D. Lundquist, 2019: Evaluation of turbulence stability schemes of land models for stable conditions. *J. Geophys. Res. Atmos.*, **124**, 3072–3089, https://doi.org/10.1029/2018JD028970.

Lapo, K. E., L. M. Hinkelman, M. S. Raleigh, and J. D. Lundquist, 2015: Impact of errors in the downwelling irradiances on simulations of snow water equivalent, snow surface temperature, and the snow energy balance. *Water Resour. Res.*, **51**, 1649–1670, https://doi.org/10.1002/2014WR016259.

Lee, W.-L., K.-N. Liou, C.-C. Wang, Y. Gu, H.-H. Hsu, and J.-L. F. Li, 2019: Impact of 3-D radiation-topography interactions on surface temperature and energy budget over the Tibetan Plateau in winter. *J. Geophys. Res. Atmos.*, **124**, 1537–1549, https://doi.org/10.1029/2018JD029592.

Lee, X., and Coauthors, 2011: Observed increase in local cooling effect of deforestation at higher latitudes. *Nature*, **479**, 384–387, https://doi.org/10.1038/nature10588.

Lehner, M., and M. W. Rotach, 2018: Current challenges in understanding and predicting transport and exchange in the atmosphere over mountainous terrain. *Atmosphere*, **9**, 276, https://doi.org/10.3390/atmos9070276.

Lindsay, J. B., 2016: Whitebox GAT: A case study in geomorphometric analysis. *Comput. Geosci.*, **95**, 75–84, https://doi.org/10.1016/j.cageo.2016.07.003.

Liston, G. E., and K. Elder, 2006: A meteorological distribution system for hHigh-resolution terrestrial modeling (MicroMet). *J. Hydrometeor.*, **7**, 217–234, https://doi.org/10.1175/JHM486.1.

Liu, C., K. Ikeda, G. Thompson, R. Rasmussen, and J. Dudhia, 2011: High-resolution simulations of wintertime precipitation in the Colorado headwaters region: Sensitivity to physics parameterizations. *Mon. Wea. Rev.*, **139**, 3533–3553, https://doi.org/10.1175/MWR-D-11-00009.1.

——, and Coauthors, 2017: Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dyn.*, **49**, 71–95, https://doi.org/10.1007/s00382-016-3327-9.

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Climate*, **26**, 9384–9392, https://doi.org/10.1175/JCLI-D-12-00508.1.

Lucas-Picher, P., D. Argüeso, E. Brisson, Y. Tramblay, P. Berg, A. Lemonsu, S. Kotlarski, and C. Caillaud, 2021: Convection-permitting modeling with regional climate models: Latest developments and next steps. *Wiley Interdiscip. Rev.: Climate Change*, **12**, e731, https://doi.org/10.1002/wcc.731.

Lundquist, J., 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, https://doi.org/10.1029/2006JD007561.

——, M. Hughes, E. Gutmann, and S. Kapnick, 2019: Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. *Bull. Amer. Meteor. Soc.*, **100**, 2473–2490, https://doi.org/10.1175/BAMS-D-19-0001.1.

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, https://doi.org/10.1029/2008JD009879.

Lussana, C., O. E. Tveito, A. Dobler, and K. Tunheim, 2019: seNorge_2018, daily precipitation, and temperature datasets over Norway. *Earth Syst. Sci. Data*, **11**, 1531–1551, https://doi.org/10.5194/essd-11-1531-2019.

Lute, A. C., and J. T. Abatzoglou, 2021: Best practices for estimating near-surface air temperature lapse rates. *Int. J. Climatol.*, **41**, E110–E125, https://doi.org/10.1002/joc.6668.

Ma, C., S. R. Fassnacht, and S. K. Kampf, 2019: How temperature sensor change affects warming trends and modeling: An evaluation across the state of Colorado. *Water Resour. Res.*, **55**, 9748–9764, https://doi.org/10.1029/2019WR025921.

Ma, M., J. Tang, T. Ou, and P. Zhou, 2023: High-resolution climate projection over the Tibetan Plateau using WRF forced by bias-corrected CESM. *Atmos. Res.*, **286**, 106670, https://doi.org/10.1016/j.atmosres.2023.106670.

Massey, J. D., W. James Steenburgh, J. C. Knievel, and W. Y. Y. Cheng, 2016: Regional soil moisture biases and their influence on WRF Model temperature forecasts over the intermountain west. *Wea. Forecasting*, **31**, 197–216, https://doi.org/10.1175/WAF-D-15-0073.1.

Matiu, M., M. Petitta, C. Notarnicola, and M. Zebisch, 2019: Evaluating snow in EURO-CORDEX regional climate models with observations for the European Alps: Biases and their relationship to orography, temperature, and precipitation mismatches. *Atmosphere*, **11**, 46, https://doi.org/10.3390/atmos11010046.

Matsuura, K., and C. J. Willmott, 2009: Terrestrial air temperature: 1900–2008 gridded monthly time series. Center for Climatic Research, Dept. of Geography, University of Delaware, https://rims.unh.edu/data/read_me.cgi?category=7&subject=9.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237–3251, https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

McAfee, S. A., G. J. McCabe, S. T. Gray, and G. T. Pederson, 2019: Changing station coverage impacts temperature trends in the Upper Colorado River basin. *Int. J. Climatol.*, **39**, 1517–1538, https://doi.org/10.1002/joc.5898.

McCrary, R. R., S. McGinnis, and L. O. Mearns, 2017: Evaluation of snow water equivalent in NARCCAP simulations, including measures of observational uncertainty. *J. Hydrometeor.*, **18**, 2425–2452, https://doi.org/10.1175/JHM-D-16-0264.1.

Menne, M. J., C. N. Williams, B. E. Gleason, J. Jared Rennie, and J. H. Lawrimore, 2018: The global historical climatology network monthly temperature dataset, version 4. *J. Climate*, **31**, 9835–9854, https://doi.org/10.1175/JCLI-D-18-0094.1.

Meyer, J., J. Horel, P. Kormos, A. Hedrick, E. Trujillo, and S. M. Skiles, 2023: Operational water forecast ability of the HRRR-iSnobal combination: An evaluation to adapt into production environments. *Geosci. Model Dev.*, **16**, 233–250, https://doi.org/10.5194/gmd-16-233-2023.

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, https://doi.org/10.1029/2009JD013493.

Mitchell, T. D., and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, **25**, 693–712, https://doi.org/10.1002/joc.1181.

Mizukami, N., M. P. Clark, A. G. Slater, L. D. Brekke, M. M. Elsner, J. R. Arnold, and S. Gangopadhyay, 2014: Hydrologic implications of different large-scale meteorological model forcing datasets in mountainous regions. *J. Hydrometeor.*, **15**, 474–488, https://doi.org/10.1175/JHM-D-13-036.1.

Monteiro, D., C. Caillaud, R. Samacoïts, M. Lafaysse, and S. Morin, 2022: Potential and limitations of convection-permitting CNRM-AROME climate modelling in the French Alps. *Int. J. Climatol.*, **42**, 7162–7185, https://doi.org/10.1002/joc.7637.

Morris, V. R., 2005: Total Sky Imager (TSI) Handbook. U.S. Department of Energy, DOE/SC-ARM/TR-017, 12 pp., https://www.arm.gov/publications/tech_reports/handbooks/tsi_handbook.pdf.

——, 2018: Infrared Thermometer (IRT) Instrument Handbook. R. Stafford, Ed., ARM Climate Research Facility, DOE/SC-ARM-TR-015, 30 pp., https://doi.org/10.2172/1417312.

——, 2019: Microwave Radiometer (MWR) Handbook. R. Stafford, Ed., U.S. Department of Energy, DOE/SC-ARM/TR-016, 23 pp., https://doi.org/10.2172/1020715.

Mooney, P. A., S. Sobolowski, and H. Lee, 2020: Designing and evaluating regional climate simulations for high latitude land use land cover change studies. *Tellus*, **72A** (1), 1–17, https://doi.org/10.1080/16000870.2020.1853437.

Mott, R., V. Vionnet, and T. Grünewald, 2018: The seasonal snow cover dynamics: Review on wind-driven coupling processes. *Front. Earth Sci. China*, **6**, 197, https://doi.org/10.3389/feart.2018.00197.

Mountain Research Initiative EDW Working Group, 2015: Elevation-dependent warming in mountain regions of the world. *Nat. Climate Change*, **5**, 424–430, https://doi.org/10.1038/nclimate2563.

Muñoz-Sabater, J., and Coauthors, 2021: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data*, **13**, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021.

Musselman, K. N., N. Addor, J. A. Vano, and N. P. Molotch, 2021: Winter melt trends portend widespread declines in snow water resources. *Nat. Climate Change*, **2021**, 418–424, https://doi.org/10.1038/s41558-021-01014-9.

Naud, C. M., Y. Chen, I. Rangwala, and J. R. Miller, 2013: Sensitivity of downward longwave surface radiation to moisture and cloud changes in a high-elevation region. *J. Geophys. Res. Atmos.*, **118**, 10 072–10 081, https://doi.org/10.1002/jgrd.50644.

Newman, A. J., and Coauthors, 2015: Gridded ensemble precipitation and temperature estimates for the contiguous United States. *J. Hydrometeor.*, **16**, 2481–2500, https://doi.org/10.1175/JHM-D-15-0026.1.

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

——, S. Z. Dobrowski, A. P. Ballantyne, A. E. Klene, and S. W. Running, 2015b: 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.

Palazzi, E., L. Mortarini, S. Terzago, and J. von Hardenberg, 2019: Elevation-dependent warming in global climate model simulations at high spatial resolution. *Climate Dyn.*, **52**, 2685–2702, https://doi.org/10.1007/s00382-018-4287-z.

Pan, L. L., S. H. Chen, D. Cayan, M. Y. Lin, Q. Hart, M. H. Zhang, Y. Liu, and J. Wang, 2011: Influences of climate change on California and Nevada regions revealed by a high-resolution dynamical downscaling study. *Climate Dyn.*, **37**, 2005–2020, https://doi.org/10.1007/s00382-010-0961-5.

Pepin, N. C., and Coauthors, 2022: Climate changes and their elevational patterns in the mountains of the world. *Rev. Geophys.*, **60**, e2020RG000730, https://doi.org/10.1029/2020RG000730.

Plüss, C., and A. Ohmura, 1997: Longwave radiation on snow-covered mountainous surfaces. *J. Appl. Meteor.*, **36**, 818–824, https://doi.org/10.1175/1520-0450-36.6.818.

Powers, J. G., and Coauthors, 2017: The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Prein, A. F., and Coauthors, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.*, **53**, 323–361, https://doi.org/10.1002/2014RG000475.

Rahimi, S. R., C. Wu, X. Liu, and H. Brown, 2019: Exploring a variable-resolution approach for simulating regional climate over the Tibetan plateau using VR-CESM. *J. Geophys. Res. Atmos.*, **124**, 4490–4513, https://doi.org/10.1029/2018JD028925.

Rangwala, I., 2013: Amplified water vapour feedback at high altitudes during winter. *Int. J. Climatol.*, **33**, 897–903, https://doi.org/10.1002/joc.3477.

——, and J. R. Miller, 2012: Climate change in mountains: A review of elevation-dependent warming and its possible causes. *Climatic Change*, **114**, 527–547, https://doi.org/10.1007/s10584-012-0419-3.

Rasmussen, R., and Coauthors, 2012: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc.*, **93**, 811–829, https://doi.org/10.1175/BAMS-D-11-00052.1.

Rasmussen, R. M., and Coauthors, 2023: CONUS404: The NCAR-USGS 4-km long-term regional hydroclimate reanalysis over the CONUS. *Bull. Amer. Meteor. Soc.*, **104**, E1382–E1408, https://doi.org/10.1175/BAMS-D-21-0326.1.

Rauchöcker, A., A. Rudolph, I. Stiperski, and M. Lehner, 2023: Cold-air pool development in a small Alpine Valley. *Quart. J. Roy. Meteor. Soc.*, **150**, 1243–1266, https://doi.org/10.1002/qj.4644.

Reynolds, D. S., J. M. Pflug, and J. D. Lundquist, 2021: Evaluating wind fields for use in basin-scale distributed snow models. *Water Resour. Res.*, **57**, e2020WR028536, https://doi.org/10.1029/2020WR028536.

Rhoades, A. M., P. A. Ullrich, and C. M. Zarzycki, 2018a: Projecting 21st century snowpack trends in western USA mountains using variable-resolution CESM. *Climate Dyn.*, **50**, 261–288, https://doi.org/10.1007/s00382-017-3606-0.

——, ——, ——, H. Johansen, S. A. Margulis, H. Morrison, Z. Xu, and W. D. Collins, 2018b: Sensitivity of mountain hydroclimate simulations in variable-resolution CESM to microphysics and horizontal resolution. *J. Adv. Model. Earth Syst.*, **10**, 1357–1380, https://doi.org/10.1029/2018MS001326.

Ritsche, M. T., 2011: ARM Surface Meteorology Systems Handbook. U.S. Department of Energy, DOE/SC-ARM/TR-086, 19 pp., https://doi.org/10.2172/1007926.

Roehrig, R., and Coauthors, 2020: The CNRM global atmosphere model ARPEGE-Climat 6.3: Description and evaluation. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002075, https://doi.org/10.1029/2020MS002075.

Rontu, L., C. Wastl, and S. Niemelä, 2016: Influence of the details of topography on weather forecast—Evaluation of HARMONIE experiments in the Sochi Olympics domain over the Caucasian Mountains. *Front. Earth Sci. China*, **4**, 13, https://doi.org/10.3389/feart.2016.00013.

Rotach, M. W., M. Andretta, P. Calanca, A. P. Weigel, and A. Weiss, 2008: Boundary layer characteristics and turbulent exchange mechanisms in highly complex terrain. *Acta Geophys.*, **56**, 194–219, https://doi.org/10.2478/s11600-007-0043-1.

——, and Coauthors, 2022: A collaborative effort to better understand, measure, and model atmospheric exchange processes over mountains. *Bull. Amer. Meteor. Soc.*, **103**, E1282–E1295, https://doi.org/10.1175/BAMS-D-21-0232.1.

Rozante, J. R., E. Ramirez, and A. A. Fernandes, 2022: A newly developed South American mapping of temperature with estimated lapse rate corrections. *Int. J. Climatol.*, **42**, 2135–2152, https://doi.org/10.1002/joc.7356.

Rudisill, W., A. N. Flores, and J. McNamara, 2021: The impact of initial snow conditions on the numerical weather simulation of a northern Rockies atmospheric river. *J. Hydrometeor.*, **22**, 155–167, https://doi.org/10.1175/JHM-D-20-0018.1.

——, K. E. Kaiser, and A. N. Flores, 2022: Evaluating long-term one-way atmosphere-hydrology simulations and the impacts of two-way coupling in four mountain watersheds. *Hydrol. Processes*, **36**, e14578, https://doi.org/10.1002/hyp.14578.

Rudisill, W. J., A. N. Flores, H. P. Marshall, E. Siirila-Woodburn, D. R. Feldman, A. M. Rhoades, Z. Xu, and A. Morales, 2023: Cold-season precipitation sensitivity to microphysical parameterizations: Hydrologic evaluations leveraging snow lidar datasets. *J. Hydrometeor.*, **25**, 143–160, https://doi.org/10.1175/JHM-D-22-0217.1.

Scalzitti, J., C. Strong, and A. K. Kochanski, 2016: A 26 year high-resolution dynamical downscaling over the Wasatch Mountains: Synoptic effects on winter precipitation performance. *J. Geophys. Res. Atmos.*, **121**, 3224–3240, https://doi.org/10.1002/2015JD024497.

Schättler, U., G. Doms, and C. Schraff, 2021: A description of the nonhydrostatic regional COSMO-Model Part VII: User's guide. Deutscher Wetterdienst Rep. COSMO-Model, 144 pp., https://doi.org/10.5676/DWD_pub/nwv/cosmo-doc_6.00_VII.

Schultz, N. M., P. J. Lawrence, and X. Lee, 2017: Global satellite data highlights the diurnal asymmetry of the surface temperature response to deforestation. *J. Geophys. Res. Biogeosci.*, **122**, 903–917, https://doi.org/10.1002/2016JG003653.

Scinocca, J. F., and Coauthors, 2016: Coordinated global and regional climate modeling. *J. Climate*, **29**, 17–35, https://doi.org/10.1175/JCLI-D-15-0161.1.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, https://doi.org/10.1175/2010MWR3425.1.

Serafin, S., and Coauthors, 2018: Exchange processes in the atmospheric boundary layer over mountainous terrain. *Atmosphere*, **9**, 102, https://doi.org/10.3390/atmos9030102.

Serreze, M. C., M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. S. Pulwarty, 1999: Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resour. Res.*, **35**, 2145–2160, https://doi.org/10.1029/1999WR900090.

Siirila-Woodburn, E. R., and Coauthors, 2021: A low-to-no snow future and its impacts on water resources in the western United States. *Nat. Rev. Earth Environ.*, **2**, 800–819, https://doi.org/10.1038/s43017-021-00219-y.

Simonpietri, A., M. Carbone, and A. Richardson, 2022: Raw soil carbon dioxide, moisture, temperature and micrometeorological data in the East River Watershed, Colorado June 2021–October 2022. DE-SC0021139, accessed 1 October 2023, https://doi.org/10.15485/1909712.

Skamarock, W. C., J. B. Klemp, M. G. Duda, L. D. Fowler, S.-H. Park, and T. D. Ringler, 2012: A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering. *Mon. Wea. Rev.*, **140**, 3090–3105, https://doi.org/10.1175/MWR-D-11-00215.1.

——, and Coauthors, 2019: A description of the Advanced Research WRF Model version 4. NCAR Tech. Note NCAR/TN-556+STR, 145 pp., https://doi.org/10.5065/1dfh-6p97.

Slater, A. G., and Coauthors, 2001: The representation of snow in land surface schemes: Results from PILPS 2(d). *J. Hydrometeor.*, **2**, 7–25, https://doi.org/10.1175/1525-7541(2001)002<0007:TROSIL>2.0.CO;2.

Smiatek, G., H. Kunstmann, and A. Senatore, 2016: EURO-CORDEX regional climate model analysis for the Greater Alpine Region: Performance and expected future change. *J. Geophys. Res. Atmos.*, **121**, 7710–7728, https://doi.org/10.1002/2015JD024727.

Smith, R. B., 2019: 100 years of progress on mountain meteorology research. *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, *Meteor. Monogr.*, No. 59, https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0022.1.

Snethlage, M. A., and Coauthors, 2022: A hierarchical inventory of the world's mountains for global comparative mountain science. *Sci. Data*, **9**, 149, https://doi.org/10.1038/s41597-022-01256-y.

Soares, P. M. M., and Coauthors, 2022: The added value of km-scale simulations to describe temperature over complex orography: The CORDEX FPS-Convection multi-model ensemble runs over the Alps. *Climate Dyn.*, 1–24, https://doi.org/10.1007/s00382-022-06593-7.

Solman, S. A., and J. Blázquez, 2019: Multiscale precipitation variability over South America: Analysis of the added value of CORDEX RCM simulations. *Climate Dyn.*, **53**, 1547–1565, https://doi.org/10.1007/s00382-019-04689-1.

——, and Coauthors, 2013: Evaluation of an ensemble of regional climate model simulations over South America driven by the ERA-Interim reanalysis: Model performance and uncertainties. *Climate Dyn.*, **41**, 1139–1157, https://doi.org/10.1007/s00382-013-1667-2.

Steger, C. R., B. Steger, and C. Schär, 2022: HORAYZON v1.2: An efficient and flexible ray-tracing algorithm to compute horizon and sky view factor. *Geosci. Model Dev.*, **15**, 6817–684, https://doi.org/10.5194/gmd-15-6817-2022.

Sterk, H. A. M., G. J. Steeneveld, and A. A. M. Holtslag, 2013: The role of snow-surface coupling, radiation, and turbulent mixing in modeling a stable boundary layer over Arctic sea ice. *J. Geophys. Res. Atmos.*, **118**, 1199–1217, https://doi.org/10.1002/jgrd.50158.

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

Tang, Q., and Coauthors, 2023: The fully coupled regionally refined model of E3SM version 2: Overview of the atmosphere, land, and river results. *Geosci. Model Dev.*, **16**, 3953–3995, https://doi.org/10.5194/gmd-16-3953-2023.

Thornton, J. M., N. Pepin, M. Shahgedanova, and C. Adler, 2022: Coverage of in situ climatological observations in the world's mountains. *Front. Climate*, **4**, 814181, https://doi.org/10.3389/fclim.2022.814181.

Thornton, P. E., S. W. Running, and M. A. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.*, **190**, 214–251, https://doi.org/10.1016/S0022-1694(96)03128-9.

——, M. M. Thornton, B. W. Mayer, Y. Wei, R. Devarakonda, R. S. Vose, and R. B. Cook, 2016: Daymet: Daily surface weather data on a 1-km grid for North America, version 3. ORNL DAAC, Oak Ridge, Tennessee, USA. USDA-NASS, 2019. 2017 Census of Agriculture, Summary and State Data, Geographic Area Series, Part 51, AC-17-A-51, accessed 3 March 2024, https://doi.org/10.3334/ORNLDAAC/1328.

Torma, C., E. Coppola, F. Giorgi, J. Bartholy, and R. Pongrácz, 2011: Validation of a high-resolution version of the regional climate model RegCM3 over the Carpathian Basin. *J. Hydrometeor.*, **12**, 84–100, https://doi.org/10.1175/2010JHM1234.1.

Torrez-Rodriguez, L., K. Goubanova, C. Muñoz, and A. Montecinos, 2023: Evaluation of temperature and precipitation from CORDEX-CORE South America and Eta-RCM regional climate simulations over the complex terrain of subtropical Chile. *Climate Dyn.*, **61**, 3195–3221, https://doi.org/10.1007/s00382-023-06730-w.

Udall, B., and J. Overpeck, 2017: The twenty-first century Colorado River hot drought and implications for the future. *Water Resour. Res.*, **53**, 2404–2418, https://doi.org/10.1002/2016WR019638.

van Meijgaard, E., L. H. Van Ulft, and W. J. Van de Berg, 2008: The KNMI regional atmospheric climate model RACMO, version 2.1. Royal Netherlands Meteorological Institute TR-302, 50 pp., https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubTR/TR302.pdf.

Vautard, R., and Coauthors, 2021: Evaluation of the large EURO-CORDEX regional climate model ensemble. *J. Geophys. Res. Atmos.*, **126**, e2019JD032344, https://doi.org/10.1029/2019JD032344.

Vernay, M., M. Lafaysse, D. Monteiro, P. Hagenmuller, R. Nheili, R. Samacoïts, D. Verfaillie, and S. Morin, 2022: The S2M meteorological and snow cover reanalysis over the French mountainous areas, description and evaluation (1958—2020). *Earth Syst. Sci. Data*, **14**, 1707–1733, https://doi.org/10.5194/essd-14-1707-2022.

Vionnet, V., S. Bélair, C. Girard, and A. Plante, 2015: Wintertime subkilometer numerical forecasts of near-surface variables in the Canadian Rocky Mountains. *Mon. Wea. Rev.*, **143**, 666–686, https://doi.org/10.1175/MWR-D-14-00128.1.

——, I. Dombrowski-Etchevers, M. Lafaysse, L. Quéno, Y. Seity, and E. Bazile, 2016: Numerical weather forecasts at kilometer scale in the French Alps: Evaluation and application for snowpack modeling. *J. Hydrometeor.*, **17**, 2591–2614, https://doi.org/10.1175/JHM-D-15-0241.1.

Viviroli, D., and R. Weingartner, 2004: The hydrological significance of mountains: From regional to global scale. *Hydrol. Earth Syst. Sci.*, **8**, 1017–1030, https://doi.org/10.5194/hess-8-1017-2004.

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, A., and X. Zeng, 2014: Range of monthly mean hourly land surface air temperature diurnal cycle over high northern latitudes. *J. Geophys. Res. Atmos.*, **119**, 5836–5844, https://doi.org/10.1002/2014JD021602.

Wang, Y., B. Geerts, and C. Liu, 2018: A 30-year convection-permitting regional climate simulation over the interior western United States. Part I: Validation. *Int. J. Climatol.*, **38**, 3684–3704, https://doi.org/10.1002/joc.5527.

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo, 2014: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.*, **50**, 7505–7514, https://doi.org/10.1002/2014WR015638.

Whiteman, D. C., 2000: *Mountain Meteorology: Fundamentals and Applications.* Oxford University Press, 376 pp.

Wijngaard, R. R., A. R. Herrington, W. L. Lipscomb, G. R. Leguy, and S.-I. An, 2023: Exploring the ability of the variable-resolution Community Earth System Model to simulate cryospheric-hydrological variables in High Mountain Asia. *Cryosphere*, **17**, 3803–3828, https://doi.org/10.5194/tc-17-3803-2023.

Winstral, A., J. Magnusson, M. Schirmer, and T. Jonas, 2019: The bias-detecting ensemble: A new and efficient technique for dynamically incorporating observations into physics-based, multilayer snow models. *Water Resour. Res.*, **55**, 613–631, https://doi.org/10.1029/2018WR024521.

Winter, K. J.-P. M., S. Kotlarski, S. C. Scherrer, and C. Schär, 2017: The Alpine snow-albedo feedback in regional climate models. *Climate Dyn.*, **48**, 1109–1124, https://doi.org/10.1007/s00382-016-3130-7.

Wu, C., X. Liu, Z. Lin, A. M. Rhoades, P. A. Ullrich, C. M. Zarzycki, Z. Lu, and S. R. Rahimi-Esfarjani, 2017: Exploring a variable-resolution approach for simulating regional climate in the Rocky Mountain region using the VR-CESM. *J. Geophys. Res. Atmos.*, **122**, 10 939–10 965, https://doi.org/10.1002/2017JD027008.

Xu, Y., A. Jones, and A. Rhoades, 2019: A quantitative method to decompose SWE differences between regional climate models and reanalysis datasets. *Sci. Rep.*, **9**, 16520, https://doi.org/10.1038/s41598-019-52880-5.

Xu, Z., A. M. Rhoades, H. Johansen, P. A. Ullrich, and W. D. Collins, 2018: An Intercomparison of GCM and RCM dynamical downscaling for characterizing the hydroclimatology of California and Nevada. *J. Hydrometeor.*, **19**, 1485–1506, https://doi.org/10.1175/JHM-D-17-0181.1.

——, A. Di Vittorio, J. Zhang, A. Rhoades, X. Xin, H. Xu, and C. Xiao, 2021: Evaluating variable-resolution CESM over China and western United States for use in water-energy nexus and impacts modeling. *J. Geophys. Res. Atmos.*, **126**, e2020JD034361, https://doi.org/10.1029/2020JD034361.

——, E. R. Siirila-Woodburn, A. M. Rhoades, and D. Feldman, 2023: Sensitivities of subgrid-scale physics schemes, meteorological forcing, and topographic radiation in atmosphere-through-bedrock integrated process models: A case study in the Upper Colorado River basin. *Hydrol. Earth Syst. Sci.*, **27**, 1771–1789, https://doi.org/10.5194/hess-27-1771-2023.

Yang, T., Q. Li, X. Chen, R. Hamdi, P. De Maeyer, and L. Li, 2021: Variation of snow mass in a regional climate model downscaling simulation covering the Tianshan Mountains, Central Asia. *J. Geophys. Res. Atmos.*, **126**, e2020JD034183, https://doi.org/10.1029/2020JD034183.

Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh, 2012: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, **93**, 1401–1415, https://doi.org/10.1175/BAMS-D-11-00122.1.

Zachariassen, J., K. F. Zeller, N. Nikolov, and T. McClelland, 2003: A review of the Forest Service Remote Automated Weather Station (RAWS) network. General Tech. Rep. RMRS-GTR119, Rocky Mountain Research Station, U.S. Forest Service, 153 pp., https://www.fs.usda.gov/rm/pubs/rmrs_gtr119.pdf.

Zadra, A., D. Caya, J. Côté, B. Dugas, C. Jones, R. Laprise, K. Winger, and L.-P. Caron, 2008: The next Canadian regional climate model. *Phys. Canada*, **64**, 75–83.

Zarzycki, C. M., C. Jablonowski, and M. A. Taylor, 2014: Using variable-resolution meshes to model tropical cyclones in the Community Atmosphere Model. *Mon. Wea. Rev.*, **142**, 1221–1239, https://doi.org/10.1175/MWR-D-13-00179.1.

Zhu, M., T. Yao, W. Yang, B. Xu, and X. Wang, 2017: Evaluation of parameterizations of incoming longwave radiation in the high-mountain region of the Tibetan Plateau. *J. Appl. Meteor. Climatol.*, **56**, 833–848, https://doi.org/10.1175/JAMC-D-16-0189.1.