Correctly applying lapse rates in ecological studies: comparing temperature observations and gridded data in Yellowstone

MICHAEL T. TERCEK,1,† ANN RODMAN,2 SHANNON WOOLFOLK,2 ZACHARY WILSON,2 DAVID THOMA,3 AND JOHN GROSS4

1Walking Shadow Ecology, PO Box 1085, Gardiner, Montana 59030 USA
2Yellowstone Center for Resources, US National Park Service, PO Box 168, Yellowstone National Park, Wyoming 82190 USA
3US National Park Service, Inventory and Monitoring Program, 2327 University Way Suite 2, Bozeman, Montana 59715 USA
4US National Park Service, Climate Change Response Program, 1201 Oakridge Drive Suite 200, Fort Collins, Colorado 80525 USA

Citation: Tercek, M. T., A. Rodman, S. Woolfolk, Z. Wilson, D. Thoma, and J. Gross. 2021. Correctly applying lapse rates in ecological studies: comparing temperature observations and gridded data in Yellowstone. Ecosphere 12(3):e03451. 10.1002/ecs2.3451

Abstract. Researchers often use gridded datasets to develop statistical relationships between climate and natural resources at locations that are distant from weather stations. These gridded datasets may provide inaccurate estimates of temperature at sites with elevations that differ significantly from the elevation assigned to the corresponding grid cell. We assess the accuracy of three gridded climate datasets commonly used in ecological studies (800-m resolution PRISM, 4-km GRIDMET, and 1-km Daymet) compared with a network of 153 temperature dataloggers arranged along elevation transects in Yellowstone National Park (Wyoming, Montana, Idaho). Measured lapse rates for monthly average daytime high temperatures were generally steeper (more cooling per unit of elevation increase) than lapse rates in gridded datasets and steeper than those in similar mountainous regions. Measured lapse rates for monthly average nighttime lows were similar to gridded data lapse rates. Our measured lapse rates are most useful for the adjustment of daytime highs from weather stations or gridded data during warmer months. Temperatures during cooler months sometimes are strongly affected by other factors. Ecologically relevant metrics calculated from temperatures adjusted with constant lapse rates can increase multiplicatively and by varying amounts from year to year. For example, high-elevation estimates of climatic water deficit (CWD) calculated from gridded data that were not adjusted with our measured lapse rates were 2.5–4 times greater (depending on the year) than calculations from temperatures that were adjusted. Our results emphasize the importance of correcting grid-based climate estimates for elevation in complex terrain when accurate, site-specific data are required. When the elevation assigned to grid cells differs significantly from the elevation of points within the cell, the lapse rate obtained for data extracted from grids can be shallower than the true rate. We illustrate the implications of these findings with case studies in Yellowstone.

Key words: accuracy assessment; Daymet; elevation; gridded climate data; GRIDMET; lapse rate; PRISM; temperature; Yellowstone National Park.

Received 28 November 2020; accepted 9 December 2020; final version received 24 January 2021. Corresponding Editor: Debra P. C. Peters.
Copyright: © 2021 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
†E-mail: Tercek@YellowstoneEcology.com
INTRODUCTION

As primary drivers of ecological processes, weather and climate have influences that vary greatly on fine spatial scales, particularly in areas with complex, mountainous terrain (Ricklefs 1990). Since these same mountainous areas often have weather stations that disproportionately represent lower elevations (Davey et al. 2006, Barry 2008), ecological studies are often forced to use somewhat imperfect estimates of temperature for locations that are distant from weather stations, particularly at higher elevations (Oyler et al. 2014, 2015).

Many datasets provide temperature estimates as continuous, gridded surfaces, in which higher elevation temperatures are extrapolated or derived from primarily lower elevation weather station data. The methods for estimating higher elevation temperatures range from assuming fixed lapse rates (e.g., 6.5°C/km) to spatial interpolation, to regression-based calculations of lapse rates that incorporate elevation, aspect, and other parameters, to dynamic corrections of higher elevation data based on satellite data. The available datasets, their methods, and the differences in their estimates are reviewed and evaluated in Behnke et al. (2016), Walton and Hall (2018), and Alder and Hostetler (2019). These reviews clearly show that datasets vary in accuracy across geographic regions and that the selection of best dataset depends on which variables are most important and how they are to be used (e.g., capturing temperature inversions in coastal areas vs. accuracy of high-elevation temperatures in mountains vs. accurately capturing trends over time). Researchers need to carefully evaluate the strengths and weaknesses of the gridded climate data products that they use, rather than treating them as a black box (Behnke et al. 2016). Importantly, none of the reviews just cited compare the gridded datasets against measurements that are independent of the data used to create the gridded products, leaving open the possibility that the accuracy of the estimates declines in locations that are more distant from weather stations.

The US National Park Service (NPS) in Yellowstone National Park (YNP; Wyoming, Montana, Idaho, USA) routinely uses three gridded climate data products in their ecological research: 800-m resolution Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 2008), 4-km resolution GRIDMET (Abatzoglou 2013), and 1-km resolution Daymet (Thornton et al. 1997, Hasenauer et al. 2003). These three gridded datasets were chosen because they (1) are routinely and relatively rapidly updated, making them suitable for retrospective analysis of field data and near-term forecasts of resource conditions, (2) have fine spatial scale, and (3) have reasonable accuracy when gridded data are compared to co-located weather stations (Behnke et al. 2016).

Since ecological and management questions at higher elevations are of particular interest to management in YNP (discussed below), and lapse rates are known to vary widely among regions (Barry 2008), we designed this study to independently measure lapse rates at a series of transects in YNP. This allowed us to evaluate the suitability of the gridded data products currently in use for ecological research purposes, particularly at higher elevation locations distant from the weather stations that are used as inputs in gridded data creation. Holden et al. (2011, 2015) used dataloggers to create high-resolution (250 m) gridded datasets for the northern Rocky Mountains, including YNP, but their gridded dataset is not regularly updated, making it unsuitable for ongoing research use, and their study design did not explicitly include elevation in a way that allows us to routinely estimate temperatures for locations of research and management interest in YNP.

We deployed a datalogger network modeled on the methods of Holden et al. (2011, 2013) along elevation gradients in YNP. Our goals were to:

1. Assess the point accuracy (gridded vs. in situ measurement) of the three gridded products currently used in YNP against independent temperature measurements, seasonally and across elevation gradients.
2. Empirically calculate temperature lapse rates, particularly for areas in northern YNP that have a large number of long-term research sites.
3. Evaluate and illustrate some of the implications of gridded vs. measured temperature differences for ecological studies in YNP.
and other mountainous areas that use temperature time series extracted from single grid cells.

**METHODS**

**Temperature data collection**

One hundred and fifty-three onset pendant dataloggers (Onset Corporation Part # UA-001-64, accuracy ± 0.53°C from 0°C to 50°C, resolution 0.14°C at 25°C) were deployed along roads throughout YNP, and off-road elevation transects were established in six mountainous regions (regions labeled in Fig. 1). In order to measure localized spatial variability in lapse rates, one of the mountainous regions (SR: Specimen Ridge) had three replicate transects, that is, three lines of dataloggers ascending nearby slopes. Each elevation transect was designated as north- or south-facing, and care was taken to choose sites along the slopes that corresponded to the target aspect within ± 10 degrees bearing. Regions BP and W (Fig. 1) had north- and south-facing transects, while the other regions had transects with only one aspect. In total, there were 10 elevation transects. All of the dataloggers in our elevation transects were on ridgelines rather than ravines subject to cold air drainage or other topographic effects. The purpose of the road-based dataloggers was to evaluate gridded data accuracy as a function of horizontal distance from weather stations within an elevation band. The lowest elevation transect (SM, north aspect; Fig. 1) spanned the range 1659–2886 m. The highest transect (W, north aspect) spanned 1910–3125 m. About 60% of the data loggers were distributed in the middle-elevation band (2148–2636 m; Fig. 2). In each transect, dataloggers were placed at approximately 100-m vertical distance increments. Elevations were determined with a research-grade GPS unit. Each datalogger was mounted in a solar radiation shield (Holden et al. 2013), which minimized the effects of variable sun exposure as a result of patchy vegetation cover, and attached to the north side of a tree at 2.5 m height above ground. This height was chosen because it was consistently above the maximum height of the snow pack in our study area. Measurements were collected every hour from June 2014 to November 2019.

**Calculation of lapse rates and gridded data accuracy**

Temperature observations were quality-screened to eliminate problems that included days in which the sensors were covered with snow, direct sun hitting the sensor (temperature spikes indicating radiation shield failure), physically impossible readings, and progressive thermistor failure (detected as a gradual shift into physically impossible readings). When days with these problems were detected, the entire 24-h period (midnight to midnight) was replaced with missing value markers. The Python code for conducting this screening is available from the authors. For each 24-h period containing valid measurements, maximum and minimum temperatures were extracted and averaged to produce monthly average daytime highs and monthly average nighttime lows at each location during each month, which matches the procedure followed by manual weather stations in YNP, one of the primary data sources for the gridded data products evaluated here. Averages were not calculated for months with more than five missing days of measurements, matching the protocol routinely used in YNP to calculate monthly averages from weather stations. Comparisons between datalogger and gridded data were performed by extracting data from the grid cells containing each datalogger and calculating the difference as datalogger monthly average minus gridded monthly average. The three gridded datasets compared were 800-m resolution Parameter-elevation Regressions on Independent Slopes Model (PRISM, version LT81; Daly et al. 2008), 4-km resolution GRIDMET (Abatzoglou 2013), and 1-km resolution Daymet (Thornton et al. 1997, Hasenauer et al. 2003). Monthly values were obtained directly from PRISM. For Daymet and GRIDMET, monthly values were calculated from daily data.

Observed lapse rates (OLRs) were estimated from datalogger observations. OLRs were stratified by aspect (north-facing vs. south-facing) and calculated as the slope of non-parametric linear regressions performed on the monthly values. Non-parametric regressions are less biased by extreme values (Sen 1968), but comparisons with ordinary least squares regressions found only fractions of a degree difference in slopes. OLRs were only estimated from locations within a
transect, for example, all the W locations (Fig. 1). We did not calculate lapse rates by regressing data from low- vs. high-elevation locations from different transects (Fig. 1). Following a review of the initial results, estimation of OLRs was restricted to months at a particular transect in which monthly averages could be calculated from at least three dataloggers that spanned a minimum of 400-m elevation and at least two grid cells.

There are two ways to calculate lapse rates from gridded data. In both methods, the same gridded temperatures along an elevation transect are used, but the elevations used to calculate lapse rate differ. The first method estimates a gridded-modeled lapse rate (GMLR) from elevations used by the gridded climate dataset. The second method estimates a user-obtained lapse rate (OLR) from elevations used by the gridded climate dataset. The second method estimates a user-obtained lapse rate (OLR) from elevations used by the gridded climate dataset.

Fig. 1. Map showing the locations of dataloggers in Yellowstone National Park, located in northwest Wyoming and parts of Montana and Idaho, USA. Letters indicate elevation transects (SM = Sepulcher Mountain, BP = Bunsen Peak, W = Mt. Washburn, DP = Druid Peak, SR = Specimen Ridge, AP = Avalanche Peak).

Fig. 2. Elevational distribution of dataloggers in this study (orange) and weather stations (gray) among evenly spaced elevation bands in Yellowstone National Park.
rate (UOLR), using elevations accurately measured at the datalogger locations. The GMLR is the lapse rate used by the algorithm that created the gridded data. The UOLR is the effective lapse rate, or the practical result, obtained by an ecological researcher that uses gridded climate data as-is, without considering the assigned elevation of the grid cells, merely extracting temperatures for a point location of interest with a known or independently measured elevation, and uncritically using those temperatures as covariates for field data or to answer research questions (examples cited and described below). An incorrect UOLR is important whether or not a gridded data user is interested in lapse rates. Since high-elevation temperatures in gridded data are usually extrapolated from lower elevation weather stations, an incorrect UOLR will usually result in an incorrect temperature estimate for high-elevation locations (illustrated below).

The difference between GMLR and UOLR calculations is significant because gridded data rarely capture the full range of elevations in mountainous terrain such as found in YNP, particularly the tops of mountain peaks. This can be seen in Fig. 3 (top) for GRIDMET data, which captures the least elevation range among the three datasets considered because of its coarser 4-km spatial resolution. The blue dots show that the north-facing Washburn transect, which is used for the example water balance calculations below, spans approximately 1910–3125 m (1.2 km) vertical distance when measured by a GPS unit but only 2000–2600 m (0.6 km) in the GRIDMET dataset (Fig. 3, top). Furthermore, the lowest elevation datalogger at Tower Junction (left-most blue dot, Fig. 3, top) falls within a GRIDMET cell that has an assigned elevation that is greater than the next higher datalogger in the transect (blue dot second from left, Fig. 3, top). The difference between the assigned grid cell elevation and the elevation of the measured locations within the high-elevation grid cells consistently lowers the UOLR calculated from GPS elevations, relative to the lapse rate assumed by the gridded data algorithm, that is, the GMLR (Fig. 3, bottom, average ratio of values on x-axis vs. y-axis = 0.53). Another way to think about this difference is that the gridded data algorithm may have a correct lapse rate (°C/km), but when ecological researchers use the data as-is to obtain temperature estimates at high-elevation point locations (such as mountain tops) within the cells, they are unwittingly applying this lapse rate over too short of a vertical distance, making the UOLR too shallow. Most ecological users of gridded data do not think in terms of lapse rates but the difference between the GMLR and the UOLR is an important contributor to the differences between empirical temperature measurements at point locations and gridded temperature estimates. In the case of Mount Washburn (blue dots, Fig. 3, top), the GRIDMET GMLR of approximately -10°C/km during a particular month is being applied over 600 m of elevation gain (2000–2600 m) from Tower Junction to Mount Washburn rather than the known 1.2 km (1910–3125 m) distance from the nearest weather station to the mountain top study site, making the UOLR (effective lapse rate) for estimating Mount Washburn temperatures at 3125 m approximately one half the GMLR (~5°C/km, Fig. 3, bottom, lower left of figure).

Ecological implications of user-obtained lapse rates (UOLRs) differing from gridded-modeled lapse rate (GMLR)

To illustrate potential consequences of the UOLR being too shallow, we used a water balance model to calculate temperature-dependent metrics that are more proximally related to ecological processes than temperature or precipitation by themselves (Thorntwhaite and Mather 1955, Lutz et al. 2010, Kiem 2010). Water balance variables in the model included climatic water deficit (CWD in mm; evaporative demand from the atmosphere not met by available soil water; Stephenson 1998), and accumulated growing degree-days (AGDD, °C with a base temperature of 4.4°C). Potential evapotranspiration (PET) in the model, which is a precursor calculation for CWD, was calculated with the temperature-based Hamon method as described in Lutz et al. (2010).

To drive the water balance model, we used OLRs (datalogger-derived lapse) and gridded data UOLRs to estimate temperatures for Mount Washburn (3125 m, location W in Fig. 1), which is a site of long-term research in YNP. Temperatures from the Tower Junction COOP weather station (elevation 1910 m, station ID 489025), located at the base of the mountain, were
adjusted with lapse rates calculated separately for each month, producing estimated temperatures for the top of the mountain (3125 m). We applied the lapse rates over a vertical distance of 1.2 km, the known elevation difference between the weather station and the mountain top. The results of these adjustments were a collection of four synthetic datasets with weather station precipitation data common to all and four different lapse-adjusted temperature time series. These four synthetic datasets were then used to drive the water balance model at the

Fig. 3. Top: Scatterplot of datalogger (GPS-measured) elevations vs. elevations assigned to the corresponding grid cell by GRIDMET. Blue squares are dataloggers located on the north-facing W transect (Fig. 1), which were the basis for lapse rates used in our illustrative water balance calculations (Figs. 9, 10). Black dots are all the other dataloggers in locations where temperature vs. elevation regressions were calculated from the GRIDMET cells. Red dashed line = 1:1 ratio of axes. Bottom: Lapse rates calculated from GRIDMET data with the same ΔT among cells in a transect but different elevations in the divisor. x-axis = measured GPS elevations as divisors for lapse rate. y-axis = elevations assigned to the GRIDMET cells as divisors. y-axis lapse rates are those assumed by the GRIDMET algorithm.
summit of Mount Washburn. Rather than merely extracting temperature data from the grid cells corresponding to Mount Washburn summit and using them as inputs for the water balance model, we used lapse rates in the way described to estimate Mount Washburn temperatures because we wanted commensurate estimates for the datalogger sites over a longer time period than the datalogger observational period of 2014–2019, so that the ecological implications of the differences could be more completely evaluated (described below). We focused on this exemplar location to illustrate the potential effects of the gridded data biases on water balance metrics that are linked to key resource conditions and park management decisions. A spatially exhaustive analysis of these effects is beyond the scope of this study.

The four sets of CWD and AGDD estimates for Mount Washburn were evaluated with respect to the uses of these metrics in YNP to support management decisions. CWD is strongly linked to historical wildfire patterns and is currently used by NPS as a metric for estimating real-time, daily fire risk (Tercek 2019). Locations where the AGDD (with a base temperature of 4.4°C) exceed 833 in a year are regarded as having sufficiently sustained warm periods to allow mountain pine beetles (Dendroctonus ponderosae), which parasitize white pines in YNP and cause widespread deforestation, to become univoltine, completing a reproductive cycle in one year rather than several and thus allowing populations to rapidly increase (Carroll et al. 2006, Shanahan et al. 2016).

To quantify the differences between datalogger-collected and weather station-collected temperatures, we placed dataloggers within 5 m of 12 weather stations during the entire study period and quantified the temperature differences as mean absolute error and mean error. Details are in Appendix S1.

RESULTS

Comparison of gridded data to observations

Estimates of monthly average daytime high temperatures from gridded data (PRISM, GRIDMET, and Daymet) were usually warmer than the in situ datalogger measurements (Fig. 4). This difference was greatest in the cooler months and less in the warm months, and GRIDMET had consistently greater differences than the other two datasets (Fig. 4, top). Median difference for daytime highs from November to March ranged from $-3.7^\circ$ to $-1.1^\circ$C (datalogger value minus grid value), while median difference was near $-1^\circ$C during the summer months (Fig. 4). Monthly average nighttime lows estimated from GRIDMET were often less similar to observations than the other two gridded products, particularly during the summer months (Fig. 5, bottom). Monthly differences for daily maximum temperatures averaged across all months for all datasets were $-1.6^\circ$C. For nighttime lows, the same average was $+0.74^\circ$C. Tabular versions of the figures appear in Appendix S2.

A closer examination of these differences revealed that gridded estimates for monthly average daytime highs differed more from dataloggers at higher elevations (Fig. 4). As described in Methods, higher elevation point locations such as mountain tops are consistently underrepresented in gridded data because average cell elevations are virtually always less than the higher peaks within a grid cell. The differences between datalogger elevations and grid cell elevations are shown in Fig. 3 for GRIDMET. In the 2637- to 3125-m elevation band, which was not included at all in the GRIDMET dataset for these locations (i.e., no GRIDMET cells were assigned these elevations; Fig. 3), median temperature difference for daytime highs ranged from $-8.4^\circ$C (GRIDMET, March) to $-1^\circ$C (Daymet, July). In the 1657- to 2147-m elevation band, temperature differences ranged from $-2.1^\circ$C (GRIDMET, December) to $+0.3^\circ$C (GRIDMET, July; Fig. 5). There was a consistent pattern of increasing difference in monthly average daytime highs with elevation in all datasets (Fig. 5). Particular locations had much greater temperature differences (see dots of the boxplots; Fig. 5). The pattern for monthly average nighttime lows was more complex than the monthly average daytime high pattern (Fig. 6). Cooler months (November–March) often had more accurate nighttime estimates at higher elevations instead of at lower elevations, while warmer months had nighttime low estimates that were less accurate at higher elevations, though with a positive bias—that is, the in situ measurements were warmer than the gridded data (Fig. 6).
Lapse rates

Monthly average daytime high temperature OLRs (observed lapse rates, datalogger lapse rates) were consistently steeper (i.e., more rapid temperature decline with elevation) on south-facing slopes than on north-facing slopes, but north- and south-facing slopes had similar lapse rates for nighttime low temperatures (Fig. 7). OLRs

Fig. 4. Differences between gridded monthly averages and datalogger monthly averages, calculated as datalogger value minus gridded value. Top: monthly average daytime high temperatures. Bottom: monthly average nighttime lows. Boxplots show the distribution of differences at all locations in a given month, with data from 2014 to 2019 merged into corresponding months. Boxes = 25th percentile (bottom), median (middle line), and 75th percentile (top of box). Whiskers = 1.5 × the distance between the 25th and 75th percentiles (interquartile range, IQR), or, if no points exceed the IQR, whiskers = extreme values. Points = differences that exceeded the IQR.

Lapse rates

Monthly average daytime high temperature OLRs (observed lapse rates, datalogger lapse rates) were consistently steeper (i.e., more rapid temperature decline with elevation) on south-facing slopes than on north-facing slopes, but north- and south-facing slopes had similar lapse rates for nighttime low temperatures (Fig. 7). OLRs
Fig. 5. Differences between gridded monthly average daytime high temperatures and corresponding datalogger monthly averages, stratified by elevation. Top: GRIDMET. Middle: PRISM. Bottom: Daymet. Negative values indicate that the gridded monthly averages were warmer than the datalogger averages.
for both aspects were steepest in the spring and shallower during the warmer months, and lapse rates for daytime highs were generally steeper than lapse rates for nighttime lows (Fig. 7). Tabular versions of all the lapse rate graphs appear in Appendix S2.

OLRs calculated from monthly average daytime high temperatures were consistently steeper than UOLRs (user-obtained lapse rates) from any of the gridded datasets on both north- and south-facing slopes (Fig. 8). The differences between OLRs and UOLRs were less pronounced for monthly average nighttime low temperatures (Fig. 8). For daytime high temperatures, GRIDMET-based UOLRs were the shallowest and consistently differed the most from the datalogger lapse rates (Fig. 8).

**Ecological implications of incorrectly estimated lapse rates**

Even though the lapse rate adjustments of the Tower weather station data resulted in linear changes in temperature (i.e., a fixed, consistent temperature change for each unit of elevation gain during each month in order to estimate Mount Washburn summit temperatures), estimates of climatic water deficit (CWD) derived from the different lapse-adjusted datasets exhibited non-linear differences. In other words, even though the lapse rate adjustments produced constant differences in temperature estimates for Mount Washburn (3125 m) from year to year among the datalogger vs. gridded synthetic datasets, the magnitude of differences among the CWD estimates derived from these temperatures varied among years. The ratios of gridded:datalogger CWD varied from year to year, that is, the lines in Fig. 9, bottom are not flat across the time series. This reflected the complex and ecologically relevant interactions between incident energy, evapotranspiration, and precipitation. When temperatures were adjusted with the PRISM or Daymet UOLRs, they produced estimates of climatic water deficit (CWD) in the water balance model that ranged from 1.5 to 2.2 times greater (varying from year to year) than estimates produced by temperatures that were adjusted with the datalogger-derived OLRs (Fig. 9). GRIDMET-derived estimates of CWD had greater differences, ranging from 2.2 to 4.5 times greater than the datalogger-derived estimates (Fig. 9).

Unlike CWD, estimates of AGDD produced by the different lapse-adjusted datasets had constant differences among each other from year to year (Fig. 10). The number of years exceeding the univoltine beetle cutoff of 833 AGDD varied greatly depending on which lapse rate adjustments (OLR or OULR) were used to estimate temperatures for Mount Washburn from Tower weather station data (Fig. 10). During the period from 1987 to 2019, datalogger-based estimates (OLR-adjusted) never exceeded the cutoff value, whereas estimates using UOLRs based on Daymet, PRISM, and GRIDMET exceeded the threshold 5, 12, and 29 times, respectively (Fig. 10).

**Gridded data differences from observations as a function of distance to the nearest weather station**

When only the middle-elevation band was considered (2148–2636 m, the band containing the most roads), there was no significant linear relationship between the gridded data bias (difference between datalogger temperatures and corresponding grid cell temperatures) and distance to the nearest weather station (tests not shown), either during particular months or when all months in the data were aggregated. Furthermore, scatter plots showed no clear non-linear patterns between weather station distance and bias (not shown).

**Difference between weather station temperature measurements and co-located datalogger temperature measurements**

Mean error (mean of all the differences between datalogger and station measurements) calculated at 12 co-located weather stations for monthly data, which is the time interval at which the lapse rates were calculated, was $-0.7^\circ$C for monthly average daytime highs and $-0.1^\circ$C for monthly average nighttime lows. The negative values indicate that the datalogger values were cooler on average. More detail on weather station vs. datalogger differences is in Appendix S1.

**DISCUSSION**

**Differences between gridded data and datalogger observations**

Our results illustrate two kinds of systematic differences in gridded climate data (Daymet,
Fig. 6. Differences between gridded monthly average nighttime low temperatures and corresponding datalogger monthly averages, stratified by elevation. Top: GRIDMET. Middle: PRISM. Bottom: Daymet. Negative values indicate that the gridded monthly averages were warmer than the datalogger averages.
PRISM, GRIDMET) relative to independent temperature observations for the YNP region. First, at lower elevations, the gridded data provided estimates of monthly average daytime highs that were often warmer compared with our in situ datalogger measurements and estimates of monthly average nighttime lows that were often cooler (Fig. 5). Second, the user-obtained lapse rates (UOLRs; i.e., the effective lapse rate that would be obtained by an ecological researcher that extracts data for a point location from gridded data), were less steep than the more accurate, observed lapse rates (OLRs) calculated from the datalogger data (Fig. 8). Combined, these two factors result in estimates for daily high temperature that generally become increasingly different from observations with elevation (Fig. 5) and gridded monthly average nighttime low temperature estimates that often converged with observations at higher elevations (Fig. 6). In other words, the too-cool gridded estimates of nighttime lows at lower elevations were extrapolated to higher elevations using UOLRs that did not cool quickly enough. The two factors affecting nighttime low temperature estimates compensate for each other as elevation increases. Conversely, the too-warm daytime high estimates at lower elevations were not cooled quickly enough by the too shallow UOLRs, exacerbating the warm bias at higher elevations.

The incorrect UOLRs were in large part due to the elevational variation contained within a cell of the gridded data, and the fact that in the YNP region, the elevation assigned to the grid cells is usually lower than any mountain peaks that might be included within the cell, meaning that the lapse rate assumed by the gridded data algorithm is applied over a smaller vertical distance than needed to reach the mountain tops, effectively lowering the UOLR for higher elevation locations within the grid cell (Fig. 3, bottom). In the case of GRIDMET, the upper elevation band of logger-measured temperatures (2637–3125 m) was completely absent from grid cells for the YNP region (i.e., no grid cells were assigned elevations in this band; Fig. 3), which contributed to GRIDMET’s larger differences from observations for these high-elevation locations (Fig. 5). Another contributing factor is the fact that the weather stations that provide the inputs for the gridded datasets considered here are disproportionately located at lower elevations (Fig. 2). If more weather stations were located at higher elevations, the gridded datasets would be better able to interpolate among observations (Behnke et al. 2016, Walton and Hall 2018).

YNP seems to have unusually steep lapse rates compared with what would be expected from physical principals and observations from other regions. For example, when empirical measurements are unavailable, temperatures are sometimes estimated by assuming a mean lapse rate of approximately \(-6.5\,^{\circ}{\text{C}}/{\text{km}}\), which is a rough average of the dry adiabatic lapse rate \((-9.8\,^{\circ}{\text{C}}/{\text{km}}\); i.e., cooling experienced by a parcel of dry air rising with no external energy inputs) and moisture-saturated adiabatic lapse rates, which are
Even though mountainous regions often have steeper rates than this assumed environmental lapse, the datalogger-measured values reported here (Fig. 7, maximum value = $-16.6^\circ$C/km; Appendix S2 contains a table of values) are often much greater than those recorded in similar studies. For example, lapse rates reported for the European Alps had a steepest value of $-7^\circ$C/km (Rolland 2003), while the Cascade Mountains of Washington had a steepest value of $-7.5^\circ$C/km (Minder et al. 2010), and central Idaho had a steepest value of $-7^\circ$C/km (Blandford et al. 2008). Pepin and Losleben (2002) and Barry (2008) reported lapse rates of approximately $-12^\circ$C/km for the Colorado Rockies and British mountains, respectively.

Several factors likely contribute to the steep lapse rates observed in this study and explain why they often exceed the dry adiabatic lapse rate. First, the physical environment of YNP is conducive to steep lapse rates. In particular, dry air, colder temperatures, and high levels of solar radiation all cause steeper lapse rates (reviewed in Barry 2008). Second, mountains in YNP often retain snow for months after the valleys have melted out. As a result, the valleys absorb more heat, while higher elevations have greater albedo and reflect more solar radiation back into space. This phenomenon is particularly relevant to the very steep rates observed in April (Fig. 7). Gardner et al. (2009) referred to such lapse rates, calculated across a snow vs. snow-free boundary as artificial or inflated because they do not reflect the free air temperatures above the near-ground boundary layer. However, for our purposes, the actual temperatures experienced near the ground, regardless of mechanism causing them, are of interest because they affect modeling such as shown in Figs. 9, 10. Third, changes in land cover as measurements ascend a slope,
particularly shifts from forest to rocky scree, have been shown to create sharp, non-linear drops in temperature as a function of elevation (Barry 2008). Similarly, large plateaus sometimes have thermal belts, higher winds, or distinct microclimates that can cause non-linear temperature drops as measurements cross the boundary between the steeper part of a slope and the flatter plateau. Raw data from our dataloggers (not shown) did indeed show evidence of non-linear temperature drops and thermal belts, particularly on Specimen Ridge (Transect SR, Fig. 1), which borders a plateau. However, since our goal was to calculate lapse rates for estimating temperatures for higher elevation sites such as Mount Washburn, the end-point calculations (i.e., the total temperature drop from highest to lowest) were more germane to our purposes. More detailed modeling efforts, similar to those of Holden et al. (2015), could use our datalogger

Fig. 9. Top: July–August climatic water deficit (CWD) calculated from a water balance model that used temperatures estimated with lapse rates from dataloggers and three user-obtained lapse rates (UOLRs, see Methods) from gridded temperature datasets. Bottom: The ratio of gridded estimates of CWD to datalogger-derived estimates of CWD on Mount Washburn (3125 m), 1987–2019. Even though constant, linear lapse rate adjustments were used consistently for every year, the ratios show interannual variation, indicating non-linear effects of incorrect lapse rates on CWD.
data to correct gridded data on finer scales for middle elevations.

Some of the variability in measured lapse rates was due to the range of elevations sampled, which changed from month to month as dataloggers failed or had their data screened out during the QC process. For example, in the three replicated transects on Specimen Ridge (SR in Fig. 1, transects separated from each other by ~3 km horizontal distance), the greatest differences among transects in lapse rates occurred during months in which the dataloggers at the elevational extremes (top or bottom) failed. Spatial variability (among transect variability) during those months was roughly double (~4°C) temporal (year to year) variability (~2°C) at single transects. We minimized this source of error during analysis by excluding data collected during months when datalogger failures created conditions in which there were less than 400-m elevation span or fewer than three dataloggers in a transect. Even with this quality control, the non-linear nature of the elevation-temperature relationship in locations such as transect SR means that lapse rates calculated here should not be applied without consideration of the underlying topography and its possible non-linear effects on elevation vs. temperature relationships.

**Ecological implications**

Our calculations of climatic water deficit (CWD) and accumulated growing degree-days (AGDD) were designed from an ecological perspective to illustrate and evaluate the potential effects of extracting temperatures from gridded data for a high-elevation point location and using them without consideration of whether that location is beyond the elevational range of the grid cell. We were interested in the implications for existing research questions and the proper use of data to inform management decisions.

A key insight from our CWD example was that even though UOLR (gridded) and OLR (datalogger) lapse rates were applied consistently to the low-elevation weather station data from year to year, the high-elevation estimates of CWD differ from each other by varying amounts from year to year. This conclusion was unchanged when we repeated the CWD estimates using the more complex Penman-Montieth method (Allen et al. 1998) to calculate the potential evapotranspiration (PET) term of CWD (Appendix S3). Using the Hamon
PET method, CWD estimates from temperatures adjusted by gridded UOLRs were 1.5–4.5 times greater than CWD estimates from datalogger OLR-adjusted temperatures (Fig. 9). Using the Penman-Montieth method, CWD estimates from gridded UOLRs were 1.0–1.9 times greater (Appendix S3). This non-linear change in CWD results from a linear lapse rate correction because temperature interacts with precipitation and other site factors in the water balance model, and temperature inputs to the water balance model are used in non-linear equations describing evapotranspiration and soil drying (Lutz et al. 2010). The non-linear differences in CWD estimates were less pronounced when the Penman-Montieth PET method was used, suggesting that there may be advantages to using the method in some circumstances (Allen et al. 1998; Appendix S3). Nevertheless, the Hamon method for calculating PET is commonly used in ecological studies (e.g., Lutz et al. 2010, Ray et al. 2019, Tercek 2019) because the non-temperature data needed as inputs are often not available, forcing the method to use approximations that reduce its advantages (Appendix S3).

Complex relationships between temperature and ecological processes, such as water balance, indicate that applied, site-specific research that uses data extracted for high-elevation point locations will likely contain multiplicative (Fig. 9) biases, the magnitude of which would be difficult to evaluate fully without a complete adjustment of the input temperature data and a recalculation of the relevant metrics for the time periods of interest. The methods of Tercek (2019), for example, might be over-estimating CWD, which would result in higher estimates of fire risk at higher elevations. Similarly, Ray et al. (2019) might be underestimating spring runoff and the drought-related (CWD-driven) risk to amphibians in high-elevation wetlands because they based their runoff calculations on Daymet.

There are research implications also for temperature-derived metrics that show interannually constant differences produced by measured lapse vs. gridded UOLR-adjusted temperatures (Fig. 10). For example, Shanahan et al. (2016), following Carroll et al. (2006), attributed the peak of the most recent mountain pine beetle outbreak in the Greater Yellowstone Ecosystem to growing seasons exceeding 833 AGDD at high-elevations, but our calculations (Fig. 10) show that exceeding this threshold depends on the choice of data used. If datalogger data are used for the calculation, then Mount Washburn has not exceeded this cutoff during any year since 1987, while at the other extreme, GRIDMET UOLR-based estimates suggest that the threshold has been exceeded during most years in the time period (Fig. 10). This highlights the fact that research questions or management decisions relying on specific magnitudes or thresholds of climate variables, particularly if they are taken from single point locations, are sensitive to the accuracy of the temperatures used to assess the threshold exceedance. In contrast, calculations that rely on trends over time, such as the rate or relative change in CWD between time periods, based on estimates of slopes from regressions on time series, might be less affected.

**CONCLUSIONS**

Ecological researchers and other users of gridded climate data should be aware that the elevation assigned to grid cells may differ significantly from the elevation of locations (e.g., mountain tops) of research interest within the grid cells (Fig. 3, top). Because of this discrepancy, the effective lapse rate, that is, the user-obtained lapse rate (UOLR) that results when data are extracted from grid cells for point locations and used in modeling (Figs. 9, 10), can be significantly lower than the OLR calculated from in situ measurements (Fig. 8) or the lapse rate used to create the gridded data (Fig. 3, bottom). In the case of YNP, more accurate temperature estimates can be obtained for these high-elevation point locations by applying the observed lapse rates (OLRs) presented here (Fig. 7, Appendix S2) to lower elevation weather station data. In contrast, ignoring the flattening of the lapse rates caused by the elevation averaging in the gridded data results in a pattern of elevation-dependent bias, with daily high temperatures becoming increasingly inaccurate and nighttime lows becoming more accurate with elevation (Figs. 5, 6). In areas outside YNP where lapse rates have not been measured, researchers using gridded climate data should at least make note of the difference between the grid cell elevation and the elevation of their study site. One possible
solution in the absence of empirically measured lapse rates would be to correct the site’s temperatures with the lapse rates taken from the gridded set (Fig. 3, bottom) but applied over the measured elevation range to the site rather than the elevation range specified by the grid cells.

In general, the lapse rates presented here are most useful for the adjustment of daytime highs during warmer months. Lapse rates have less influence on local microclimates during periods when air mixing is poor, and at these times, lapse rates are often near zero, making factors such as cold air pools and solar radiation potentially more important (Figs. 7, 8; Dobrowski et al. 2009). This is seen prominently in our data for nighttime lows during the winter (Fig. 7, nighttime lows during winter months).

There appears to be no absolutely right or wrong choice of gridded climate dataset. Instead, the data should be carefully evaluated with regard to the intended purpose of the research and the geographic region in which it is being applied. In our evaluation of monthly temperatures in YNP, PRISM and Daymet were generally more accurate (approximately 0.5°C to −1.7°C median warm bias across all months) compared with GRIDMET (1.5°C to 3.7°C warm bias), but in an evaluation of eight datasets across the entire conterminous United States, Behnke et al. (2016) found PRISM and Daymet to be within the middle of the pack in terms of accuracy. The coarser spatial resolution of GRIDMET (4 km) makes the elevation estimates less accurate for point locations in mountainous terrain, which affects the user-obtained lapse rates (UOLRs).

Insights generated by evaluating spatial and temporal differences in lapse rates have important implications for ongoing research and management decisions in YNP and other mountainous regions, which often rely on water balance modeling or other calculations driven by temperature data. Critical ecological processes such as wildland fire, forest pest dynamics, and species distribution changes are ubiquitous, and understanding their relationship to past, present, and projected climates is essential to proactively managing for the future.

ACKNOWLEDGMENTS

Funding for this work was provided by Yellowstone National Park and the National Park Service Climate Change Response Program. We thank John Abatzoglou for helpful comments on an earlier draft of this paper. The authors declare that they have no conflict of interest in the publication of this work.

LITERATURE CITED

Abatzoglou, J. T. 2013. Development of gridded surface meteorological data for ecological applications and modeling. International Journal of Climatology 33:121–131.

Alder, J. R., and S. W. Hostetler. 2019. The dependence of hydroclimate projections in snow-dominated regions of the western United States on the choice of statistically downscaled climate data. Water Resources Research 55:2279–2300.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration – guidelines for computing crop water requirements. FAO irrigation and drainage paper 56. https://www.fao.org/3/X0490E/x0490e00.htm#Contents

Barry, R. 2008. Mountain climate and weather. Third edition. Cambridge University Press, Cambridge, UK.

Behnke, R., S. Vavrus, A. Allstadt, T. Allbright, W. E. Thogmartin, and V. C. Radeloff. 2016. Evaluation of downscaled, gridded climate data for the conterminous United States. Ecological Applications 26:1338–1351.

Blandford, T., K. Humes, B. Harshburger, B. Moore, and V. Walden. 2008. Seasonal and synoptic variations in near-surface air temperature lapse rates in a mountain basin. Journal of Applied Meteorology and Climatology 47:249–261.

Carroll, A. L., J. Regniere, J. A. Logan, S. W. Taylor, and J. A. Powell. 2006. Impacts of climate change on range expansion by the mountain pine beetle. Mountain Pine Beetle Initiative working paper 2006-14. Pacific Forestry Center, Natural Resources Canada, Canadian Forest Service, Victoria, British Colombia, Canada.

Daly, C., M. Halbleib, J. L. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. Pasteris. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology 28:2031–2064.

Davey, C. A., K. T. Redmond, and D. B. Simeral. 2006. Weather and Climate Inventory, National Park Service, Greater Yellowstone Network. Natural Resource Technical Report NPS/GRYN/NRTR—2006/001. National Park Service, Fort Collins, Colorado, USA.

Dobrowski, S., J. Abatzoglou, J. Greenberg, and S. Schadlow. 2009. How much influence does landscape-scale physiography have on air temperature
in a mountain environment? Agricultural and Forest Meteorology 149:1751–1758.

Gardner, A., M. Sharp, R. Koerner, C. Labine, S. Boon, S. Marshall, D. O. Burgess, and D. Lewis. 2009. Near-surface temperature lapse rates over Arctic glaciers and their implications for temperature down-scaling. Journal of Climate 22:4281–4298.

Hasenauer, H., K. Merganicova, R. Petritsch, S. Pietsch, and P. Thornton. 2003. Validating daily climate interpolations over complex terrain in Austria. Agricultural and Forest Meteorology 119:87–107.

Holden, Z., J. C. Abatzoglou, C. H. Luce, and L. Baggett. 2011. Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain. Agricultural and Forest Meteorology 151:1066–1073.

Holden, Z., A. E. Klene, R. F. Keefe, and C. G. Moisen. 2013. Design and Evaluation of an inexpensive radiation shield for monitoring surface air temperatures. Agriculture and Forest Meteorology 180:281–286.

Holden, Z., A. Swanson, A. Klene, J. Abatzoglou, S. Dobrowski, S. Cushman, S. Cushman, J. Squires, G. Moisen, and J. Oyler. 2015. Development of high-resolution historical daily gridded air temperature data using reanalysis and distributed sensor networks for the US Northern Rocky Mountains. International Journal of Climatology. https://doi.org/10.1002/joc.4580

Keim, B. 2010. The lasting scientific impact of the Thornthwaite water-balance model. Geographical Review 100:295–300.

Lutz, J. A., J. W. van Wagendonk, and J. F. Franklin. 2010. Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. Journal of Biogeography 37:936–950.

Minder, J., P. Mote, and J. Lundquist. 2010. Surface temperature lapse rates over complex terrain: lessons from the Cascade Mountains. Journal of Geophysical Research 115:D14122.

Oyler, J., A. Ballantine, K. Jenso, M. Sweet, and S. Running. 2014. Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. International Journal of Climatology. https://doi.org/10.1002/joc.4127

Oyler, J., S. Z. Dobrowski, A. P. Ballantyne, A. E. Klene, and S. W. Running. 2015. Artificial amplification of warming trends across the mountains of the western United States. Geophysical Research Letters 42:153–161.

Pepin, N., and M. Losleben. 2002. Climate change in the Colorado rocky mountains: free air vs. surface air temperatures. International Journal of Climatology 22:311–329.

Ray, A. M., A. J. Sepulveda, K. M. Irvine, S. K. Wilmoth, D. Thoma, and D. Patla. 2019. Wetland drying linked to variations in snowmelt runoff across Grand Teton and Yellowstone national parks. Science of the Total Environment 666:1188–1197.

Ricklefs, R. 1990. Ecology. Verlag. New York, New York, USA.

Rolland, C. 2003. Spatial and seasonal variations of air temperature lapse rates in alpine regions. Journal of Climate 16:1032–1046.

Sen, P. 1968. Estimates of the regression coefficient based on Kendall’s Tau. Journal of the American Statistical Association 63:1379–1389.

Shanahan, E., K. Irvine, D. Thoma, S. Wilmoth, A. Ray, K. Legg, and H. Shovic. 2016. Whitebark pine mortality related to white pine blister rust, mountain pine beetle outbreak, and water availability. Ecosphere 7:Article e01610.

Stephenson, N. 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. Journal of Biogeography 25:855–870.

Terek, M. 2019. Nowcasting and forecasting fire severity in Yellowstone. Yellowstone Science 27:27–32.

Thornthwaite, C. W., and J. R. Mather. 1955. The water balance. Publications in Climatology, Centerton, Arkansas, USA.

Thornton, P. E., S. W. Running, and M. A. White. 1997. Generating surfaces of daily meteorology variables over large regions of complex terrain. Journal of Hydrology 190:214–251.

Walton, D., and A. Hall. 2018. An assessment of high-resolution gridded temperature datasets over California. Journal of Climate 31:3789–3810.

**Supporting Information**

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3451/full