

# SCIENTIFIC DATA

OPEN

## Data Descriptor: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015

John T. Abatzoglou<sup>1</sup>, Solomon Z. Dobrowski<sup>2</sup>, Sean A. Parks<sup>3</sup> & Katherine C. Hegewisch<sup>1</sup>

Received: 7 August 2017

Accepted: 30 October 2017

Published: 9 January 2018

We present TerraClimate, a dataset of high-spatial resolution ( $1/24^\circ$ , ~4-km) monthly climate and climatic water balance for global terrestrial surfaces from 1958–2015. TerraClimate uses climatically aided interpolation, combining high-spatial resolution climatological normals from the WorldClim dataset, with coarser resolution time varying (i.e., monthly) data from other sources to produce a monthly dataset of precipitation, maximum and minimum temperature, wind speed, vapor pressure, and solar radiation. TerraClimate additionally produces monthly surface water balance datasets using a water balance model that incorporates reference evapotranspiration, precipitation, temperature, and interpolated plant extractable soil water capacity. These data provide important inputs for ecological and hydrological studies at global scales that require high spatial resolution and time varying climate and climatic water balance data. We validated spatiotemporal aspects of TerraClimate using annual temperature, precipitation, and calculated reference evapotranspiration from station data, as well as annual runoff from streamflow gauges. TerraClimate datasets showed noted improvement in overall mean absolute error and increased spatial realism relative to coarser resolution gridded datasets.

<b>Design Type(s)</b>	data integration objective • observation design • source-based data analysis objective
<b>Measurement Type(s)</b>	climate
<b>Technology Type(s)</b>	digital curation
<b>Factor Type(s)</b>	temporal_interval
<b>Sample Characteristic(s)</b>	Earth • temperature of air • pressure of air • hydrological precipitation process • atmospheric wind • porosity of soil

<sup>1</sup>University of Idaho, Department of Geography, 875 Perimeter Dr, Moscow, ID 83844, USA. <sup>2</sup>University of Montana, Department of Forest Management, Missoula, MN 59812, USA. <sup>3</sup>USDA Forest Service, Rocky Mountain Research Station, Aldo Leopold Wilderness Research Institute, 790 E. Beckwith Ave, Missoula, MT 59801, USA. Correspondence and requests for materials should be addressed to J.T.A. (email: jabatzoglou@uidaho.edu).

Background & Summary

Global environmental questions that invoke climate as a driver of a specific phenomenon require spatially and temporally consistent datasets. An extensive and growing collection of historical observed climate datasets exist including those that interpolate data from surface observations<sup>1</sup>, reanalysis<sup>2</sup>, and some combination thereof<sup>3–5</sup> with each dataset offering advantages in the variables provided, temporal and spatial resolution, as well as the geographic extent and time period covered. Increased interest in applying climate data to multidisciplinary problems has prompted the development of climate datasets at more meaningful spatial scales. While time varying (monthly data spanning multiple decades) and high spatial resolution (< 5-km) climate datasets have been developed at national and continental scales<sup>6–8</sup>, and high-spatial resolution climate normals have been developed for the globe<sup>9–11</sup>, we are unaware of time varying (i.e., by month and year), high-resolution climate data that covers all global land surfaces and encompasses many of the essential surface climate variables.

There is increasing recognition of the mismatch between the scale of existing climate data and the scale at which organisms experience their environment<sup>12</sup>. This has motivated a large body of work aimed at downscaling gridded climate fields for use in ecological and global change studies, the results of which may be sensitive to the spatial scale of data used<sup>13</sup>. High spatial resolution climate data is often necessary for ecological and hydrological analyses and modeling<sup>14</sup>. The newest version of WorldClim<sup>15</sup> used station data, remotely sensed land surface temperature, and topographic covariates to produce high spatial resolution climate surfaces of temperature, precipitation, wind speed, vapor pressure, and downward shortwave flux. While WorldClim provides a complete set of variables to assess long-term monthly normals for several climate variables, it does not lend itself to temporal analysis that may be important for linking climate variability and climate impacts in ecological, agricultural, and hydrological systems.

In addition to temperature and precipitation, there has been interest in measures of climate that are more directly linked to ecosystem productivity and water resources. Water balance models integrate climate and biophysical factors to produce a set of variables that account for the concurrence of water and energy. Water balance models have been used to estimate runoff<sup>16,17</sup>, ecosystem productivity<sup>18</sup>, ecosystem distribution<sup>19</sup>, and ecological disturbance<sup>20,21</sup>. Whereas some global water balance datasets exist from simple water balance models<sup>22–24</sup> or more sophisticated land-surface hydrologic models<sup>25</sup>, these models typically provide output at relatively coarse spatial scales (>50-km) or lack extensive time series components. Higher resolution (< 5-km) country and continental scale water balance datasets have been developed<sup>21</sup>, yet global, time varying, high-resolution datasets do not currently exist.

This paper outlines the development of a global monthly high-resolution climate dataset from 1958–2015, herein referred to as TerraClimate, that includes the requisite variables for calculating energy-based reference potential evapotranspiration and a water balance model. We discuss the procedures for downscaling coarser resolution temporal anomalies from Climate Research Unit time series data version 4.0 (CRU Ts4.0)<sup>1</sup> and the Japanese 55-year Reanalysis (JRA-55)<sup>2</sup> from 1958–2015 with high-resolution climatological fields from WorldClim<sup>11,15</sup> using climatic aided interpolation. Furthermore, we discuss the procedures used for developing the monthly water balance dataset for the period 1958–2015.

Methods

Three global gridded climate datasets were used to develop TerraClimate, each offering distinct spatial or temporal qualities (Table 1).

The high-spatial resolution climatology from WorldClim version 2 (<http://worldclim.org/version2>) used thin-plate splines to interpolate station observations using covariates including MODIS derived land surface temperature and cloud cover, along with topographic features to develop monthly climate normal surfaces for global land surface<sup>15</sup>. Monthly climate normals for the 1970–2000 period were developed for maximum and minimum temperature, precipitation, solar radiation, vapor pressure and wind speed at four different spatial grains. We used the 2.5 arc-minute (1/24°) resolution to accommodate a high-spatial resolution (< 5-km) product that is computationally tractable for developing longer time series.

We used the monthly diurnal temperature range (DTR) from WorldClim version 1.4 (<http://worldclim.org/version1>) rather than the WorldClim version 2 dataset, as the latter showed widespread

Dataset	Spatial Resolution Used	Temporal Resolution Used	Variables
WorldClim v2 <sup>15</sup>	1/24°	Monthly, 1970–2000 normals	Tas, vap, pr, srad, ws
WorldClim v1.4 <sup>11</sup>	1/24°	Monthly, 1960–1990 normals	Tmax, Tmin, Tas
CRU Ts4.0 <sup>1</sup>	0.5°	Monthly, 1958–2015	Tmax, Tmin, vap, pr, Tas
JRA-55 <sup>2</sup>	1.25°	Monthly, 1958–2015	Tas, vap, pr, srad, ws
Root zone storage capacity <sup>32</sup>	0.5°	Time invariant	Water storage capacity

**Table 1. Datasets used in producing the TerraClimate dataset.** Variable abbreviations listed include average monthly maximum (Tmax), minimum (Tmin) and mean temperature (Tas), vapor pressure (vap), wind speed (ws), downward shortwave flux at the surface (srad), and accumulated monthly precipitation (pr). Note that the WorldClim datasets provide only monthly climate normals averaged over a period of record.

biases in mid latitudes despite monthly mean temperature showing nominal bias (Supplementary Figs 1 and 2). The monthly DTR from WorldClim version 1.4 for the 1960–1990 period was further modified to account for the change in DTR between the 1960–1990 base period and 1970–2000 base period of WorldClim v2.0. This was facilitated by interpolating differences in monthly DTR for these two base periods in the CRU Ts4.0 data. Finally, monthly average maximum and minimum temperature climatologies for the 1970–2000 period were reconstructed using WorldClim v2.0 monthly mean temperatures and the DTR from WorldClim v1.4 as modified by CRU Ts4.0.

Time varying climate data were derived from two sources: (i) Climate Research Unit (CRU) time series data version 4.0 (ref. 1) (Data Citation 1), and (ii) the Japanese 55-year Reanalysis (JRA-55 (ref. 2); [http://jra.kishou.go.jp/JRA-55/index\\_en.html](http://jra.kishou.go.jp/JRA-55/index_en.html)). The CRU Ts4.0 data provide monthly average maximum and minimum temperature, vapor pressure, and cumulative precipitation at a 0.5° grid from 1901–2015 for global land surface and were developed based on station data. The CRU Ts4.0 data however does not cover all islands and does not include Antarctica. Furthermore, CRU Ts4.0 data assumes climatological values for pixels where there are no long-term stations within a decorrelation distance, resulting in voids of climate signals in poorly sampled land areas of the globe. The JRA-55 is the longest running full-observing-system modern reanalysis product, providing spatially and temporally complete data covering the period 1958–present. We calculated monthly average 10-m wind speed, downward surface shortwave flux, temperature, and precipitation data from 3-hourly products at a 1.25° grid. For both CRU Ts4.0 and JRA-55 datasets, we calculated monthly anomalies for all variables relative to the 31-year baseline period (1970–2000) defined by WorldClim v2.0.

Climatologically aided interpolation was used to superimpose monthly climate anomalies from CRU Ts4.0 and JRA-55 with monthly climate normals from WorldClim to estimate monthly time series from 1958–2015. Climatologically aided interpolation is a simple spatial downscaling approach that employs bilinear interpolation of temporal anomalies from a higher-temporal, lower-spatial resolution dataset to a lower-temporal, higher-spatial resolution dataset<sup>26,27</sup>. All anomalies were calculated relative to the 1970–2000 base period to adhere with the WorldClim climatology. Monthly average maximum temperature, minimum temperature, vapor pressure, and precipitation anomalies were prioritized from CRU Ts4.0 where these data cover land surface pixels. The CRU Ts4.0 data assigns pixels to their climatological normal when there are no stations within decorrelation distances<sup>28</sup> in a given month. These represent a very small fraction of the temperature and precipitation record from 1958–2015. Rather than use a potentially artificial null climate signal such instances, we bilinearly interpolate JRA-55 anomalies to infill data voids. CRU Ts4.0 anomalies were further extrapolated to a 0.5° buffer from land masses by interpolating anomalies from the 8 nearest neighbors. Monthly anomalies from JRA-55 were used for pixels on islands located outside this buffer, as well as in Antarctica which is not covered by CRU Ts4.0. Anomalies of 10-m wind speed and downward surface shortwave flux were taken exclusively from JRA-55. Anomalies were computed as additive (i.e., departure from normal) for temperature, wind speed, and solar radiation, and multiplicative (i.e., percent of normal) for precipitation and vapor pressure. We capped monthly precipitation anomalies at 2,000% of normal to avoid unrealistic precipitation in exceptionally dry regions and months where climatological precipitation from CRU Ts4.0 or JRA-55 approaches zero. There were rare occurrences (~0.1% of pixels) where monthly mean maximum temperature was less than 0.5 °C higher than minimum temperature due to our method independently accounting for anomalies in maximum and minimum temperature. In such cases, we set the monthly minimum temperature equal to 0.5 °C below the monthly maximum temperature for internal consistency.

Monthly reference evapotranspiration ( $ET_0$ ) was calculated using the Penman Monteith approach<sup>29</sup>.  $ET_0$  has been argued to be a more appropriate measure of potential water loss over purely temperature based metrics for estimating potential evapotranspiration as it uses an energy balance approach.  $ET_0$  assumes a reference grass surface across space, but can yield biases where potential vegetation water use departs substantially from this assumption. Following prior studies, we modified  $ET_0$  to account for reduced surface water flux when snow cover exists or prior to the onset of the growing season and active transpiration using an empirical relationship with temperature that accounts for precipitation phase changes<sup>16,30</sup>.

A one-dimensional modified Thornthwaite-Mather climatic water-balance model (WBM)<sup>22,31</sup> was used to calculate monthly water balance from 1958–2015. The WBM is a single bucket model applied consistently across global land surfaces that operates on a monthly time step and considers the interplay between precipitation,  $ET_0$ , as well as soil and snowpack water storage. The WBM accounting scheme considers runoff as the excess of liquid water supply (precipitation and snowmelt) used by monthly  $ET_0$  and soil moisture recharge. Soil water is extracted during months where  $ET_0$  exceeds liquid water supply, with the extraction efficiency of soil water declining exponentially with the ratio of soil water to extractable soil water capacity. Under such conditions, actual evapotranspiration is counted as the liquid water supply plus the soil water utilized and climatic water deficit is the difference between  $ET_0$  and actual evapotranspiration. The WBM requires data on the plant extractable soil water capacity. We used extractable soil water storage capacity data at a 0.5° grid from Wang-Erlandsson *et al.*<sup>32</sup> that were derived from estimates of satellite based evaporation,  $ET_0$ , and precipitation. Wang-Erlandsson *et al.*<sup>32</sup> provided estimates for varying return intervals. We used the 2-year period given our use of  $ET_0$ , as it best matched with evapotranspiration variability in grassland biomes, corresponding with our use of a reference grass

surface in the WBM. To adhere to the spatial resolution of TerraClimate, we first extrapolated data using a 0.5° buffer from existing data using the 8 nearest neighbors, and then bilinearly interpolated plant extractable soil water from its native 0.5° grid. We set a lower bound on plant extractable soil water of 10 mm and set a default value of 50 mm for places void of data.

We use the WBM outlined by Dobrowski *et al.*<sup>31</sup>, with two modifications to better account for runoff and snowpack dynamics on a global scale. First, we use an empirical temperature-based transformation for precipitation phase to determine the proportion of precipitation falling as liquid, as well as the fraction of snowpack melted each month<sup>30</sup>. Although snowmelt is a more complicated process that could be improved by incorporating shortwave fluxes and snowpack ripeness, temperature based indices have been shown to be sufficient for macroscale monthly water balance calculations<sup>33</sup>. Secondly, we allow for 5% of the cumulative rainfall and snowmelt each month to occur as direct runoff irrespective of soil conditions<sup>17</sup>. The resultant datasets from the WBM are monthly total runoff (Q), climatic water deficit, actual evapotranspiration, as well as snow water equivalent and soil moisture at the end of each month.

## Data Records

TerraClimate is available to the public through an unrestricted data repository hosted by the University of Idaho's Northwest Knowledge Network (Data Citation 2). The full list of TerraClimate datasets are provided in Table 2. Files were created separately for each variable and year in NetCDF format following the Climate and Forecast metadata standards. A compressed archive of NetCDF files for the period of record covering all variables can be accessed at <http://doi.org/10.7923/G43J3B0R>. However, due to the sheer size of the data once uncompressed (~800 GB), we provide access to individual annual files for each variable at <https://climate.northwestknowledge.net/TERRACLIMATE/>. Temporal extensions to TerraClimate will be updated periodically as additional years of data become available.

## Technical Validation

We conducted a validation of the TerraClimate product for monthly temperature and precipitation using station-based data obtained from the Global Historical Climatology Network (GHCN) database<sup>34</sup>. We examined time series of temperature data that have been corrected for climate inhomogeneities that arise due to non-climatic artifacts, such as station relocation or observational practice, from the GHCNv3 dataset; whereas we used monthly precipitation data from the GHCNv2 dataset. We constrained our validation to GHCN stations that had complete monthly data for at least 30 calendar years from 1958–2015 resulting in a total of 3,230 stations for temperature and 6,102 stations for precipitation. As a caveat, we note that many of these stations were used in both the WorldClim and CRU datasets making the comparison not completely independent.

Data from TerraClimate were extracted at grid cells co-located with individual stations. We calculated linear validation statistics of Pearson's correlation coefficients, mean absolute error (MAE), and bias for each station. We constrained our validation to calendar year summaries rather than individual months to avoid artificially inflating the correlation by incorporating the seasonal cycle. To demonstrate potential added value from TerraClimate, we compare our validation statistics to CRU Ts4.0, acknowledging that by using CRU Ts4.0 in climatically aided interpolation, we should expect similar results for temporal variability.

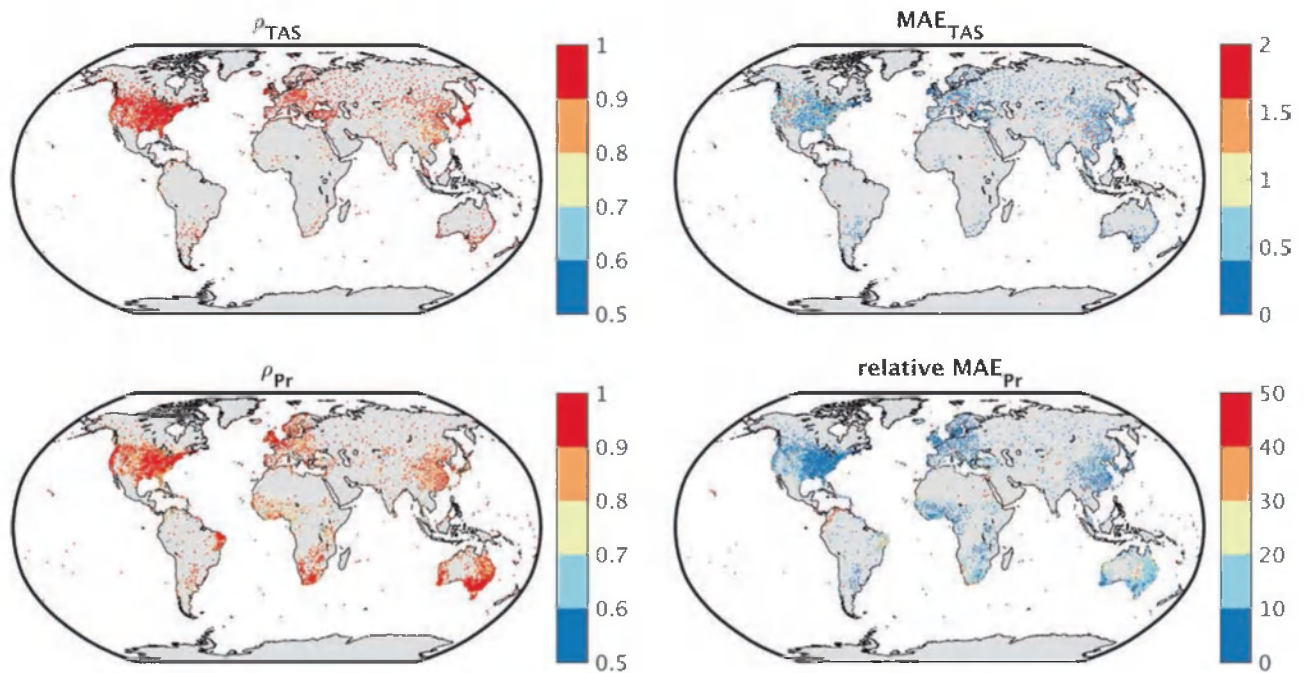
The TerraClimate dataset largely captured interannual variability in mean annual temperature across GHCN stations with a median Pearson's correlation coefficient ( $\rho$ ) of 0.95, with  $\rho > 0.9$  for approximately 80% of the stations (Fig. 1). Most of the subpar correlations ( $\rho < 0.8$ ) were found equatorward of 30° latitude, where interannual variability in mean temperature is lesser in magnitude. Similarly, TerraClimate precipitation fields captured interannual variability observed at GHCN stations globally with a median correlation of  $\rho = 0.90$ . Subpar correlations for annual precipitation were often found adjacent to stations with high correlation suggesting that interannual precipitation variability can vary across relatively short spatial scales, particularly where convective or orographic features are involved that influence the spatial variability of precipitation anomalies at scales finer than those resolved by CRU Ts4.0. Alternatively, subpar correlations may also derive from potential inhomogeneities or errors in station-based precipitation records.

Correlation statistics were nearly identical to those found using the CRU Ts4.0 dataset, as expected, given its use in the development of the TerraClimate dataset (Table 3). However, the MAE was

Dataset	Variables
Essential Climate Variables	2-m Maximum Temperature, 2-m Minimum Temperature, 2-m Vapor Pressure, 10-m Wind Speed, Downward Solar Radiation Flux at the Surface, Accumulated Precipitation
ET <sub>0</sub>	Reference Evapotranspiration
Water Balance Variables	Runoff, Actual Evapotranspiration, Climate Water Deficit, Soil Moisture, Snow Water Equivalent

**Table 2. Datasets provided by TerraClimate.** Datasets are monthly and time varying from 1958–2015 and produced at a 1/24° spatial resolution.





**Figure 1.** Validation of TerraClimate temperature and precipitation time series using GHCN stations.

(a) Pearson's correlation coefficient ( $\rho$ ) and (b) mean absolute error (MAE, units of  $^{\circ}\text{C}$ ) between GHCN stations and co-located pixels from TerraClimate for time series of annual mean temperature (TAS) from 1958–2015. (c, d) Show the correlation and relative mean absolute error (units of % of mean annual Pr) for calendar year accumulated precipitation (Pr). Statistics are reported for GHCN stations with at least 30 years of complete data.

substantially improved in the TerraClimate dataset, a consequence of bias correcting the data using the WorldClimv2 data. For mean temperature, the median MAE fell from  $0.53^{\circ}\text{C}$  for CRU to  $0.32^{\circ}\text{C}$  for TerraClimate, whereas for precipitation the median MAE fell from 11% to 9%. We also note that unaccounted-for error and associated bias inherent in comparing gridded climate data to station data is a function of mismatches in the scale of the data being compared. This is particularly apparent in regions of complex terrain where the elevation of a station may differ substantially from the average elevation of a co-located grid cell. For example, approximately 40% of the variance in the mean annual temperature bias was linearly explained by biases in elevation (i.e., grid cell elevation minus station elevation).

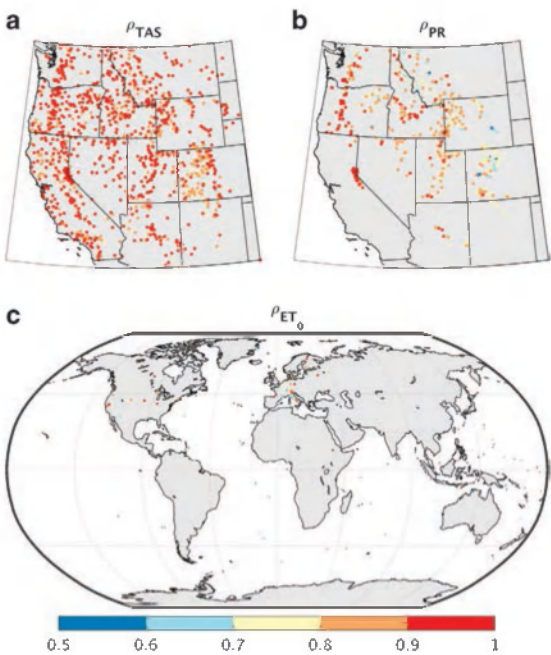
Given the potential fidelity problems when validating TerraClimate with GHCN stations, we perform a complementary validation of temperature and precipitation data using a network of 2,100 automated climate stations located across mountains of the western United States from the Snowpack Telemetry (SNOTEL) and Remote Automated Weather Stations (RAWS) networks. Due to suspected inhomogeneities with minimum temperature records from SNOTEL stations arising from instrumentation changes<sup>35</sup>, we used a set of homogenized daily temperature data for both RAWS and SNOTEL stations<sup>36</sup>. Additionally, we used daily records of accumulated precipitation from a more limited set of 513 SNOTEL stations that were quality controlled<sup>37</sup>. Station-years where more than 5% of daily observations were missing were omitted from the validation. Due to the short period of record for most stations, we used a more liberal criterion of at least 10 years of complete data to compute interannual correlations, resulting in 1,292 stations for temperature and 503 stations for precipitation.

TerraClimate largely captured interannual variability in mean annual temperature across the Snotel and Raws network with a median Pearson's correlation coefficient ( $\rho$ ) of 0.90 (Fig. 2a). Interannual correlations between calendar year precipitation from TerraClimate and recorded precipitation at SNOTEL sites was also lower than seen for neighboring GHCN stations across the western US with a median  $\rho$  of 0.79 (Fig. 2b), with a dry bias ( $\sim 30\%$ ) across interior mountains where correlations were generally lower. The lower correlations compared with GHCN could be a function of station quality control measures, but also factors that influence temperature and precipitation variability in mountainous areas such as inversions, snow-albedo feedbacks, and orographic precipitation ratios. In addition, the incongruence of the scales of gridded data and in-situ measurements are heightened in regions of complex terrain thus leading to an additional challenge in validation efforts. Validation statistics using CRU data showed generally similar values for correlation as those for TerraClimate, but with CRU showing slightly larger errors as noted with MAE (Table 3).

Comparisons of interannual variability in annual  $\text{ET}_0$  from TerraClimate were made using a global network of FLUXNET stations<sup>38</sup>. Hourly and bi-hourly observations of temperature, humidity, wind

Variable	# Stations	Network	TerraClimate		CRU Ts4.0	
			$\rho$	MAE	$\rho$	MAE
Tas	3,230	GHCN	0.95	0.32 °C	0.95	0.53 °C
Pr	6,102	GHCN	0.90	9.1% (62.9 mm)	0.89	11% (76 mm)
Tas	1,292	SNOTEL+RAWS	0.90	0.84 °C	0.90	1.21 °C
Pr	503	SNOTEL	0.78	30% (251 mm)	0.78	41% (316 mm)
ET <sub>0</sub>	50	FLUXNET	0.77	8.3% (57 mm)	0.76	8.5% (59 mm)
Q	587	GRDC	0.80	36.0% (101 mm)	0.76	42.0% (118 mm)

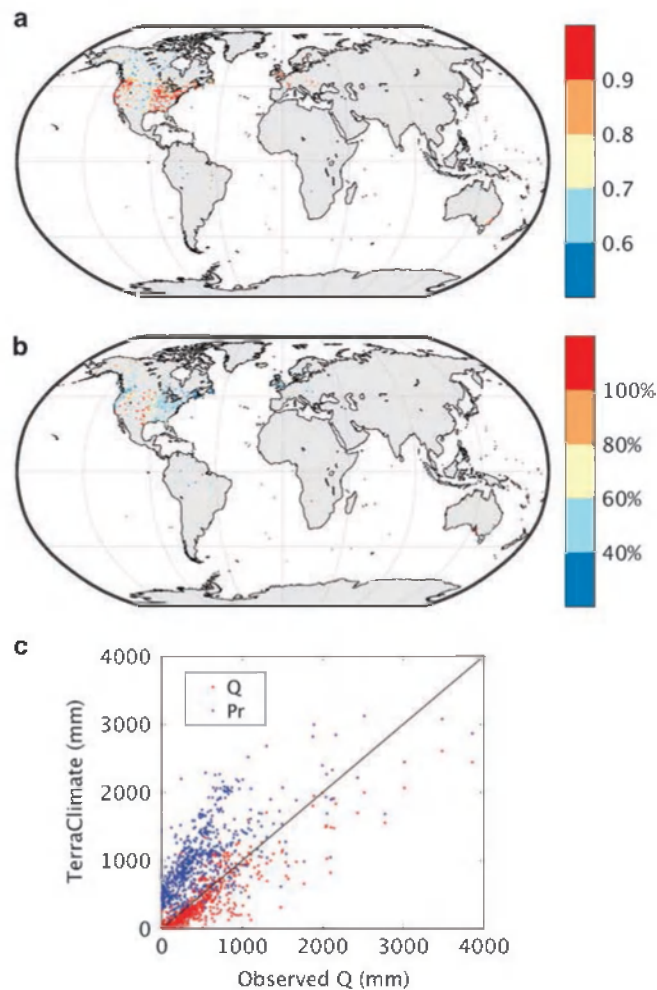
**Table 3.** Comparison of validation statistics for TerraClimate and CRU Ts4.0 for annual mean temperature (Tas), cumulative precipitation (Pr), reference evapotranspiration (ET<sub>0</sub>), and cumulative streamflow (Q). The median value of validation statistics for Pearson’s correlation coefficient ( $\rho$ ) and mean absolute error (MAE) are reported. Relative MAE is reported for Pr, ET<sub>0</sub>, and Q with absolute MAE shown in parentheses.



**Figure 2.** Validation of TerraClimate temperature, precipitation, and reference evapotranspiration time series. Pearson’s correlation coefficient ( $\rho$ ) with data from co-located pixels from TerraClimate and (a) annual mean temperature (TAS) from 1980–2015 from homogenized RAWS and SNOTEL stations, (b) annual mean precipitation (PR) from quality controlled SNOTEL stations, and (c) annual mean reference evapotranspiration (ET<sub>0</sub>) from FLUXNET stations from 1994–2014. Statistics are only reported for stations that had at least 10 years of complete data.

speed, and solar radiation from these stations were used to calculate ET<sub>0</sub>. Observational data were aggregated to monthly temporal scales to ensure that we used the same procedures for calculating ET<sub>0</sub> as TerraClimate. Interannual correlations were computed for stations that had at least 10 years of complete data, missing no more than 10% of data in any calendar year.

Pearson’s correlations of annual ET<sub>0</sub> from TerraClimate showed a median  $\rho$  of 0.77 over the 50 FLUXNET stations (Fig. 2c), with a MAE of approximately 57 mm (8.5%). The spatial correlation between observed annual ET<sub>0</sub> and ET<sub>0</sub> as modeled by TerraClimate was  $\rho = 0.98$  highlighting that the spatial pattern was well captured. A slight overall bias (median +8%) in TerraClimate ET<sub>0</sub> was seen relative to ET<sub>0</sub> calculated at FLUXNET sites, which may arise due to the non-standard heights relative to canopies measured at FLUXNET sites. Overall validation statistics were slightly better with TerraClimate than those seen for CRU (Table 3). In addition, less of the spatial variability in climatological ET<sub>0</sub> was captured with CRU ( $\rho = 0.94$ ).

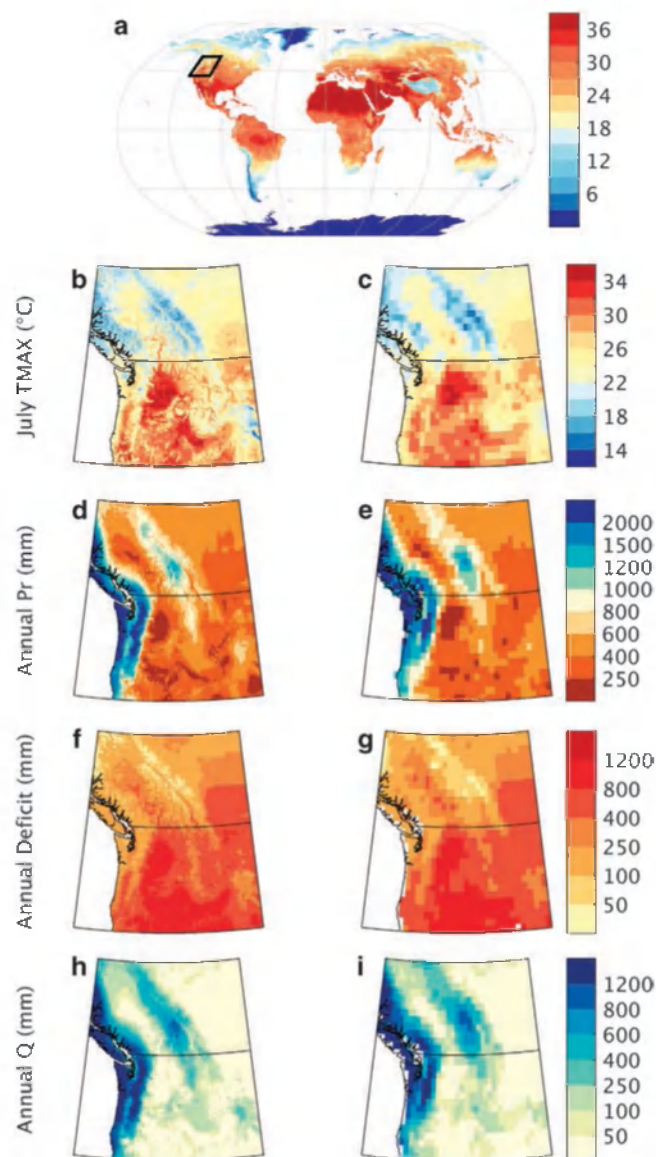


**Figure 3. Validation of TerraClimate runoff data.** Validation statistics of TerraClimate estimated runoff from the WBM showing (a) Pearson's correlation coefficient, and (b) mean absolute error (mm) for annual water-year (Oct-Sep) runoff (Q) for 587 streamflow from BGDC's Climate Sensitive Stations Dataset covering at least 30 complete years from 1958–2015. (c) Shows a scatterplot of observed annual mean Q from streamflow stations versus TerraClimate estimated Q and water-year accumulated precipitation (P). The 1:1 line is shown for reference.

To demonstrate potential utility of the water balance component of the dataset, we compared interannual variability in water-year runoff from our water balance model (WBM) to streamflow from 587 pristine river basins that are part of the Climate Sensitive Stations Dataset from the Global Runoff Data Centre (The Global Runoff Data Centre, 56068 Koblenz, Germany, 2013). We considered water-year (Oct-Sep) rather than calendar year comparisons to account for the contributions of snowmelt runoff from watersheds in the northern hemisphere. These stations were selected based on having nominal water extraction or diversions, consistency of data records, at least 30 complete water-years of data from 1958–2015, and existing watershed boundary data. We extracted raster based runoff from the WBM for each gauge by delineating pixels that occurred within the upstream area of the respective watershed. The identical WBM was also run using CRU Ts4.0 datasets for comparative purposes.

Pearsons' correlation coefficients were calculated between water-year cumulative runoff (Q) simulated by the WBM and observations. Streamflow data were scaled to units of mm per year by accounting for the size of the upstream contributing watershed. We considered temporal correlations to assess interannual variability, as well as spatial correlations to assess the performance of the dataset to capture the spatial variability in annual Q. Interannual variability in estimated runoff was reasonably captured by the WBM, with a median  $\rho = 0.80$  (Fig. 3a). The WBM captured over 70% of the interannual variability in Q of the mid-latitude basins in the United States, Europe and southern Australia. Lower explained variance was seen for parts of the tropics, boreal systems, and across the Great Plains of North America. Overall, the WBM exhibited a tendency for less runoff than was observed with a mean bias of  $-24\%$ . The median relative MAE of water-year runoff was 36% (101 mm; Fig. 3b). However, MAE exceeded 100% for



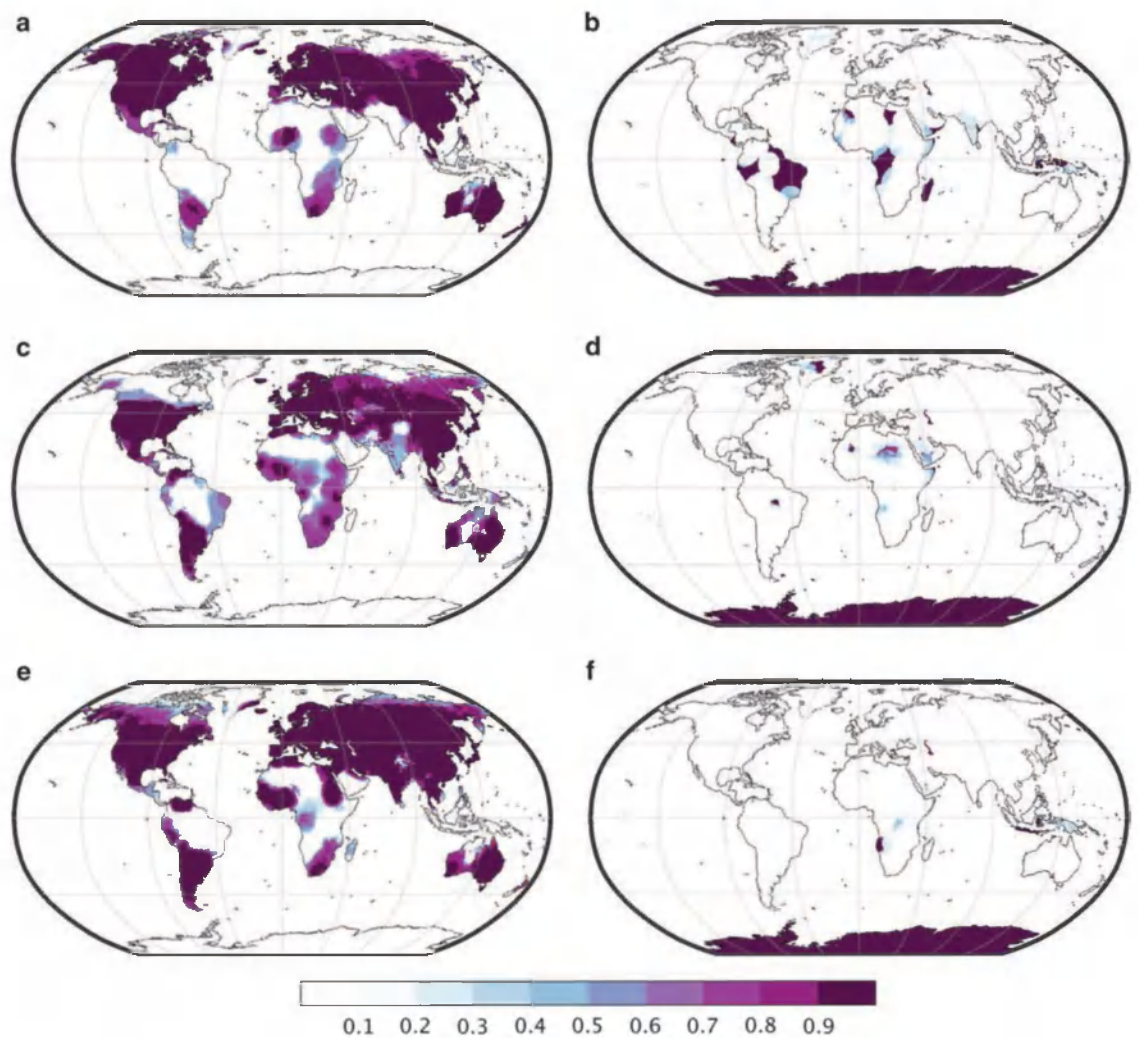


**Figure 4.** Illustration of added value from the TerraClimate dataset. Top panel shows July 2015 monthly average maximum temperature (Tmax) from TerraClimate with the inset highlighting the region for (a–h). Comparison of (a,b) July 2015 Tmax, (c,d) 2015 calendar year accumulated precipitation, (e,f) 2015 calendar year climatic water deficit, and (g,h) 2015 runoff between the TerraClimate dataset (left) and CRU Ts4.0 dataset (right).

6% of the watersheds, predominantly arid watersheds across the Great Plains and Australia where annual mean  $Q < 25$  mm. The spatial correlation between observed and simulated annual mean  $Q$  was  $\rho = 0.87$ , suggesting that the patterns were well represented by the WBM (Fig. 3c). By comparison, these validation results were improvements over the CRU Ts4.0 dataset using the same WBM framework in terms of the median interannual correlation ( $\rho = 0.765$ ), relative MAE (42%, 118 mm), and the spatial correlation ( $\rho = 0.84$ ).

Water balance models can be biased due to several factors. Errors can arise through inaccurate estimates of climate forcings, by not accounting for vegetation dynamics within a watershed, and by the overall simplicity of the single bucket model approach which does not account for subsurface routing of water. Notably, by using  $ET_0$ , we assume a constant reference vegetation type that may substantially deviate from observations. Additionally, plant water soil storage capacity likely deviates at scales finer than those mapped here, which could contribute to uncertainty. We conducted an additional validation analysis by comparing cumulative water-year precipitation data to observed  $Q$  for each watershed. The WBM runoff data explained substantially more variance of the interannual variability in  $Q$  than using precipitation data alone (median  $\rho = 0.68$ ). Observed mean runoff exceeded estimated mean annual precipitation for 30 of the 587 watersheds, further illustrating the challenges of interpolating estimates of





**Figure 5.** Coverage of CRU Ts4.0 station influence used in TerraClimate anomalies. Proportion of monthly data with (left) at least 8 and (right) 0 stations from CRU Ts4.0 contributing to anomalies from 1958–2015. (a,b) Show coverage for mean temperature, used in TerraClimate maximum and minimum temperature, (c,d) show coverage for precipitation, and (e,f) show coverage for vapor pressure. TerraClimate uses anomalies from JRA-55 rather than assume climatological averages from CRU for voxels where 0 stations contribute.

precipitation data, particularly in mountainous watersheds as highlighted by previous studies<sup>15,39</sup>. Underestimates of runoff and precipitation in some mountain watersheds highlights the need for continued improvement of mapping climate at fine spatial scales globally<sup>10,15</sup> and investment in observational systems in historically data sparse regions such as mountains.

We illustrate some of the value-added information from the TerraClimate dataset compared to CRU Ts4.0 for a portion of northwestern North America in 2015 (Fig. 4). Specifically, we highlight the spatial detail added by the TerraClimate dataset that is absent in CRU across a region with substantial geographic gradients in energy and moisture that are apparent in July 2015 average maximum temperature (Fig. 4b,c), and 2015 annual precipitation totals (Fig. 4d,e). Integrating climate metrics through the WBM improves the biophysical relevance of the gridded products for use in global-change and ecological studies, as can be seen by the additional detail in TerraClimate for annual climatic water deficit (Fig. 4f,g), and annual Q in 2015 (Fig. 4h,i).

The use of climate data has increased in the past decade due to increasing availability of data, broadening of the disciplines that use such data, and increased interest in assessing local impacts from climate change. The potential applications of TerraClimate span a variety of spatial scales from users wanting local climate and water balance time series in regions void of easily accessible or sufficiently long station-based data, to users conducting broader spatial analyses to better understand and refine regional and global climate impacts to water resources and agriculture<sup>40</sup>. In addition, TerraClimate has many potential uses in ecology, macroecology, and global change science. For example, water balance data are

increasingly used in explaining and predicting patterns of wildfire activity<sup>20,41,42</sup>, vegetation distribution<sup>19,43</sup> and structure<sup>44</sup>, climate velocity<sup>31</sup>, and drought induced tree mortality<sup>45</sup>.

### Usage Notes

Climatological data for several voxels from the WorldClim dataset near perennial sea ice or glaciers primarily in Antarctica and Greenland had exceptionally large, and likely unrealistic, vapor pressure deficits ( $>0.5$  kPa). These voxels were primarily located near the coastline and had nominal differences in diurnal temperature range ( $<2$  °C). Moreover, we caution users that output from the water balance model, as well as estimated  $ET_0$ , are likely suspect in perennially glaciated landscapes including Greenland and Antarctica. Secondly, there are a few voxels near the South Pole and in the Sahara Desert where climatological 1970–2000 precipitation from WorldClim v2.0 is zero. The use of multiplicative anomalies in climatologically aided interpolation sets all values to zero, which may be suspect.

Uncertainty in observational gridded climate datasets that use station based data can arise from a variety of sources ranging from procedural choices for interpolating and extrapolating data spatially, to changes in data availability through time, to climate inhomogeneities that lead to structural uncertainty in climate datasets<sup>46,47</sup>. TerraClimate uses three different datasets, each of which contain their own sources of uncertainty. We provide a measure of uncertainty in TerraClimate data using estimates of the number of stations that contribute to anomaly fields from CRU Ts4.0 based on decorrelation distances<sup>28</sup>. CRU Ts4.0 provides estimates of the number of stations that contribute to estimates of monthly anomalies at the native scale of the data and provides these data for temperature, precipitation, and vapor pressure fields. The number of stations (between 0 and 8) contributing to monthly CRU Ts4.0 precipitation, mean temperature, and vapor pressure fields were interpolated to the TerraClimate grid and provided as ancillary data for precipitation, monthly maximum and minimum temperature, and vapor pressure. Anomalies from JRA-55 were used for voxels where zero stations contributed to the anomaly fields in CRU Ts4.0 data. The fraction of months from 1958–2015 with 8 contributing stations (maximum number of stations influencing) and 0 contributing stations (anomalies strictly derived from JRA-55) from CRU Ts4.0 for temperature, precipitation, and vapor pressure is shown in Fig. 5. These ancillary data are provided to help users of TerraClimate identify the robustness of interannual variability in data for specific geographic locations, as well as the data source of the interannual variability used in TerraClimate for these four variables. For example, the low fraction of months with high station density from CRU may limit the robustness of TerraClimate across regions such as Africa and South America, where temporal anomalies instead rely more heavily on JRA-55. Ancillary data are not provided for downward shortwave radiation or 10-m winds which exclusively use JRA-55 anomalies, or for  $ET_0$  and the water balance variables which use combinations of the essential climate variables.

Long-term trends in primary climatological fields of TerraClimate are preserved from their parent datasets in JRA-55 and CRU Ts4.0 and hence TerraClimate should not be used to provide an independent estimate of trends. Interannual variability in primary climate variables from TerraClimate will not incorporate spatial variability in anomalies at scales finer than their parent datasets, including those associated with orographic precipitation ratios, inversions, and near coastal maritime influences. Likewise, while WorldClim data incorporated into TerraClimate allows for a high-resolution dataset, such approaches may not adequately resolve microclimate features, particularly in complex terrain or heterogeneous land-cover. Some approaches have been used to better resolve microclimate features to finer spatial scales<sup>48</sup>, yet such relationships have been typically applied to regional domains.

### References

- Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
- Kobayashi, S. *et al.* The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteorol. Soc. Jpn. Ser. II* **93**, 5–48 (2015).
- Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**, 3088–3111 (2006).
- Abatzoglou, J. T. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.* **33**, 121–131 (2013).
- Beck, H. E. *et al.* MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.* **21**, 589–615 (2017).
- Daly, C. *et al.* Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.* **28**, 2031–2064 (2008).
- Wang, T., Hamann, A., Spittlehouse, D. & Carroll, C. Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. *PLoS ONE* **11**, e0156720 (2016).
- Thornton, P. E., Running, S. W. & White, M. A. Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.* **190**, 214–251 (1997).
- Ficklin, D. L., Abatzoglou, J. T., Robeson, S. M. & Dufficy, A. The influence of climate model biases on projections of aridity and drought. *J. Clim.* **29**, 1269–1285 (2015).
- Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 170122 (2017).
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
- Potter, K. A., Arthur Woods, H. & Pincebourde, S. Microclimatic challenges in global change biology. *Glob. Chang. Biol.* **19**, 2932–2939 (2013).
- Root, T. L. & Schneider, S. H. Strategic cyclical scaling: bridging five orders of magnitude scale gaps in climatic and ecological studies. *Integr. Assess* **3**, 188–200 (2002).

14. Franklin, J. *et al.* Modeling plant species distributions under future climates: how fine scale do climate projections need to be? *Glob. Chang. Biol.* **19**, 473–483 (2013).
15. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
16. Abatzoglou, J. T., Barbero, R., Wolf, J. W. & Holden, Z. A. Tracking interannual streamflow variability with drought indices in the US Pacific Northwest. *J. Hydrometeorol.* **15**, 1900–1912 (2014).
17. Wolock, D. M. & McCabe, G. J. Explaining spatial variability in mean annual runoff in the conterminous United States. *Clim. Res.* **11**, 149–159 (1999).
18. Schimel, D. S., Braswell, B. H. & Parton, W. J. Equilibration of the terrestrial water, nitrogen, and carbon cycles. *Proc. Natl. Acad. Sci.* **94**, 8280–8283 (1997).
19. Stephenson, N. L. Climatic control of vegetation distribution: the role of the water balance. *Am. Nat.* **135**, 649–670 (1990).
20. Abatzoglou, J. T. & Williams, A. P. Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl. Acad. Sci.* **113**, 11770–11775 (2016).
21. Williams, A. P. *et al.* Temperature as a potent driver of regional forest drought stress and tree mortality. *Nat. Clim. Chang.* **3**, 292–297 (2013).
22. Willmott, C. J., Rowe, C. M. & Mintz, Y. Climatology of the terrestrial seasonal water cycle. *J. Climatol.* **5**, 589–606 (1985).
23. McCabe, G. J. & Wolock, D. M. Temporal and spatial variability of the global water balance. *Clim. Change* **120**, 375–387 (2013).
24. Haddeland, I. *et al.* Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results. *J. Hydrometeorol.* **12**, 869–884 (2011).
25. Adam, J. C., Hamlet, A. F. & Lettenmaier, D. P. Implications of global climate change for snowmelt hydrology in the twenty-first century. *Hydrol. Process.* **23**, 962–972 (2009).
26. Willmott, C. J. & Robeson, S. M. Climatologically aided interpolation (CAI) of terrestrial air temperature. *Int. J. Climatol.* **15**, 221–230 (1995).
27. Mosier, T. M., Hill, D. F. & Sharp, K. V. 30-Arcsecond monthly climate surfaces with global land coverage. *Int. J. Climatol.* **34**, 2175–2188 (2014).
28. New, M., Hulme, M. & Jones, P. Representing Twentieth-Century Space-Time Climate Variability. Part II: Development of 1901–96 Monthly Grids of Terrestrial Surface Climate. *J. Clim.* **13**, 2217–2238 (2000).
29. Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. Crop evapotranspiration—Guidelines for computing crop water requirements—FAO Irrigation and drainage paper 56. *FAO, Rome* **300**, D05109 (1998).
30. Dai, A. Temperature and pressure dependence of the rain-snow phase transition over land and ocean. *Geophys. Res. Lett.* **35**, L12802 (2008).
31. Dobrowski, S. Z. *et al.* The climate velocity of the contiguous United States during the 20th century. *Glob. Chang. Biol.* **19**, 241–251 (2013).
32. Wang-Erlandsson, L. *et al.* Global root zone storage capacity from satellite-based evaporation. *Hydrol. Earth Syst. Sci.* **20**, 1459 (2016).
33. Ohmura, A. Physical Basis for the Temperature-Based Melt-Index Method. *J. Appl. Meteorol.* **40**, 753–761 (2001).
34. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. & Houston, T. G. An overview of the global historical climatology network-daily database. *J. Atmos. Ocean. Technol.* **29**, 897–910 (2012).
35. Oyler, J. W., Dobrowski, S. Z., Ballantyne, A. P., Klene, A. E. & Running, S. W. Artificial amplification of warming trends across the mountains of the western United States. *Geophys. Res. Lett.* **42**, 153–161 (2015).
36. Oyler, J. W., Ballantyne, A., Jencso, K., Sweet, M. & Running, S. W. 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 (2015).
37. Lute, A. C. & Abatzoglou, J. T. Role of extreme snowfall events in interannual variability of snowfall accumulation in the western United States. *Water Resour. Res.* **50**, 2874–2888 (2014).
38. Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC). *FLUXNET Web Page*. Available online [http://fluxnet.ornl.gov] from ORNL DAAC, Oak Ridge, Tennessee, USA.
39. Henn, B., Clark, M. P., Kavetski, D. & Lundquist, J. D. Estimating mountain basin-mean precipitation from streamflow using Bayesian inference. *Water Resour. Res.* **51**, 8012–8033 (2015).
40. Lobell, D. B., Schlenker, W. & Costa-Roberts, J. Climate trends and global crop production since 1980. *Science* **333**, 616–620 (2011).
41. Parks, S. A. *et al.* How will climate change affect wildland fire severity in the western US? *Environ. Res. Lett.* **11**, 35002 (2016).
42. McKenzie, D. & Littell, J. S. Climate change and the eco-hydrology of fire: will area burned increase in a warming western U.S.? *Ecol. Appl.* **27**, 26–36 (2017).
43. Piedallu, C., Gégout, J.-C., Perez, V. & Lebourgeois, F. Soil water balance performs better than climatic water variables in tree species distribution modelling. *Glob. Ecol. Biogeogr.* **22**, 470–482 (2013).
44. Klein, T., Randin, C. & Körner, C. Water availability predicts forest canopy height at the global scale. *Ecol. Lett.* **18**, 1311–1320 (2015).
45. Anderegg, W. R. L. *et al.* Tree mortality predicted from drought-induced vascular damage. *Nat. Geosci.* **8**, 367–371 (2015).
46. Morice, C. P., Kennedy, J. J., Rayner, N. A. & Jones, P. D. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res. Atmos.* **117**, D08101 (2012).
47. Dai, A., Fung, I. Y. & Del Genio, A. D. Surface Observed Global Land Precipitation Variations during 1900–88. *J. Clim.* **10**, 2943–2962 (1997).
48. Holden, Z. A. *et al.* Development of high-resolution (250 m) historical daily gridded air temperature data using reanalysis and distributed sensor networks for the US Northern Rocky Mountains. *Int. J. Climatol.* **36**, 3620–3632 (2016).

## Data Citations

1. Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. *Climatic Research Unit* <http://dx.doi.org/10.5285/edf8febfdad48abb2cbaf7d7e846a86> (2017).
2. Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. & Hegewisch, K. C. *Northwest Knowledge Network* <http://doi.org/10.7923/G43J3B0R> (2017).

## Acknowledgements

This work was partially supported by Google Earth Engine faculty research awards to J.T.A. and S.Z.D.

## Author Contributions

J.T.A. conceived of the development of the dataset and the development of the code, validation efforts, and manuscript writing. S.Z.D. helped conceive the development of the dataset and writing of the manuscript. S.A.P. helped conceive the development of the dataset and writing of the manuscript. K.C.H. aided in the coding and data processing issues.

## Additional Information

**Supplementary Information** accompanies this paper at <http://www.nature.com/sdata>

**Competing interests:** The authors declare no competing financial interests.

**How to cite this article:** Abatzoglou, J. T. *et al.* TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* 5:170191 doi: 10.1038/sdata.2017.191 (2018).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.

© The Author(s) 2018