

WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas

Stephen E. Fick^{a*}  and Robert J. Hijmans^b

^a Department of Plant Sciences, University of California, Davis, CA, USA

^b Department of Environmental Science and Policy, University of California, Davis, CA, USA

ABSTRACT: We created a new dataset of spatially interpolated monthly climate data for global land areas at a very high spatial resolution (approximately 1 km²). We included monthly temperature (minimum, maximum and average), precipitation, solar radiation, vapour pressure and wind speed, aggregated across a target temporal range of 1970–2000, using data from between 9000 and 60 000 weather stations. Weather station data were interpolated using thin-plate splines with covariates including elevation, distance to the coast and three satellite-derived covariates: maximum and minimum land surface temperature as well as cloud cover, obtained with the MODIS satellite platform. Interpolation was done for 23 regions of varying size depending on station density. Satellite data improved prediction accuracy for temperature variables 5–15% (0.07–0.17 °C), particularly for areas with a low station density, although prediction error remained high in such regions for all climate variables. Contributions of satellite covariates were mostly negligible for the other variables, although their importance varied by region. In contrast to the common approach to use a single model formulation for the entire world, we constructed the final product by selecting the best performing model for each region and variable. Global cross-validation correlations were ≥ 0.99 for temperature and humidity, 0.86 for precipitation and 0.76 for wind speed. The fact that most of our climate surface estimates were only marginally improved by use of satellite covariates highlights the importance having a dense, high-quality network of climate station data.

KEY WORDS interpolation; climate surfaces; WorldClim; MODIS; land surface temperature; cloud cover; solar radiation; wind speed; vapour pressure

Received 5 July 2016; Revised 12 March 2017; Accepted 14 March 2017

1. Introduction

Spatially interpolated gridded climate data, here referred to as ‘climate surfaces’, are used in many areas of work, particularly in the environmental, agricultural and biological sciences. For many applications, data at a high (≤ 1 km²) spatial resolution are preferred to capture environmental variation that can be lost at lower spatial resolutions, particularly in mountainous and other areas with steep climate gradients. Hijmans *et al.* (2005) provided climate surfaces, referred to as the ‘WorldClim version 1 database’, for global land areas (excluding Antarctica), consisting of long-term average monthly temperature and precipitation. Here we present a refined and expanded version of this database. Over the past decade, there has been an increase in the number of stations for which data is available, including a number of stations located at high latitudes and elevations. Furthermore, while the original dataset was limited to monthly precipitation and temperature, we also included solar radiation, windspeed and vapour pressure, which may be necessary to model important processes such as plant growth. These variables were

previously only available at a lower spatial resolution (10 arc-minutes) and for a different base period (New *et al.*, 2002).

We also wanted to improve estimates for areas with low station density and in areas with unusual gradients (inversions) and sharp gradients such as rain-shadows or abrupt temperature changes due to ocean-land transitions (Hijmans *et al.*, 2005; Daly *et al.*, 2008). Estimates for regions where weather station density (or quality) is insufficient to resolve local gradients may be improved using remotely sensed meteorological data from satellites (Jin and Dickinson, 2010; Mildrexler *et al.*, 2011). Gridded time-series of meteorological variables such as land surface temperature (LST; the earth’s ‘skin temperature’) and cloud cover are now available from a number of satellite-borne instruments, and these have the potential to inform estimates of the variables of interest. Previous efforts to incorporate satellite data in climate interpolations have been focused on interpolating daily temperature data for limited spatial and temporal extents (e.g. Kilibarda *et al.*, 2014). We examined the utility of including satellite-derived and other covariables (distance to coast, extraterrestrial solar radiation) for interpolating several climate variables across global land areas.

We compiled monthly average climate data for weather stations from a large number of global, regional, national,

* Correspondence to: S. E. Fick, Stockholm Environment Institute, Linnégatan 87 D, 115 23 Stockholm, Sweden. E-mail: sfick@ucdavis.edu

and local sources, mostly for the 1970–2000 period (Table 1). Utilizing satellite-derived data and other covariables, we interpolated these data using the thin-plate smoothing spline algorithm implemented in ANUSPLIN (Hutchinson and Xu, 2013) and created global climate surfaces for monthly precipitation and minimum, mean, and maximum temperature, solar radiation, wind speed and vapour pressure (that is, 12 gridded surfaces per variable). Our surfaces have a 30 arc-seconds spatial resolution; this is equivalent to about 0.86 km² at the equator (and less elsewhere) and commonly referred to as ‘1-km’ spatial resolution. The data are referred to as the ‘WorldClim version 2’ database and are available for download from <http://worldclim.org/>.

2. Methods

2.1. Climate data compilation and processing

Weather station data were obtained from multiple sources, and are summarized in Table 1 and Table S1, Supporting information. Station data were checked for correspondence between their reported elevation and the elevation obtained from a global elevation raster data (Hijmans *et al.*, 2005). Stations with large deviations (>several 100 m) between reported and actual elevation were mapped and evaluated relative to available geographic information and neighbouring station data. We checked the coordinates of these stations by searching for the station name in Google Earth (Google Inc., 2013) and elsewhere. Unless the data of either the elevation or location could be corrected with a high degree of certainty, the stations with large elevation discrepancies were left out of the analysis. Further quality control was done by inspecting outliers in the interpolation (see below).

Average temperature was calculated as the mean of maximum and minimum of tabulated station-wise monthly temperatures. If vapour pressure was not reported, but dew-point temperature was available, vapour pressure was calculated by the following formula (Tetens, 1930; vapour pressure equals saturated vapour pressure at dew-point temperature):

$$\text{svp} = 0.611 \times 10^{\frac{7.5 \times T}{237.7 + T}}$$

where svp is saturated vapour pressure (in hPa) and T is the dew-point temperature (°C).

If neither vapour pressure nor dew-point temperature were recorded, but relative humidity was available, vapour pressure was calculated as:

$$\text{vp} = \text{rh} * \frac{\text{svp}(T)}{100}$$

Where rh is mean relative humidity, T is mean temperature (°C) and svp is from the formula above. If multiple derived values for vapour pressure were available for a station, preference was given to dewpoint-temperature derived estimates. In all cases, if directly observed data was available for stations identified as duplicates, derived data was not used. Approximately 12% (1239 records) of data

used came from directly reported vapour pressure, the rest being derived from either dew-point temperature (79%) or relative humidity (8%).

2.2. Data aggregation and weighting

The climate data sources used included databases with long-term average values (WMO, 1996; FAO, 2001), time-series of monthly averages by year (Lawrimore *et al.*, 2011; Rohde *et al.*, 2013; Harris *et al.*, 2014), as well as daily weather data (NCEI, 2015) (Table 1). All data were aggregated to monthly climate averages. The target temporal range for station data was between 1970 and 2000. Stations with observations for at least 25 years within this period were selected for surface fitting (see details below). From the remaining stations, those with at least 10 years of data between 1960 and 2010 were considered for inclusion if they were located at a minimum distance from all stations selected in the first pass. This threshold was set, somewhat arbitrarily, at 25 km for precipitation, and 50 km for all other climate variables to reduce redundancy and noise in regions with high station density. These stations were then processed to remove duplicates and included for model fitting. In a final pass, stations were included with at least 10 years of data (or 5 years for solar radiation) for which only climatic averages, without exact temporal range information available, and exceeding the distance threshold from stations in the previous two passes. We added these stations under the assumption that in remote locations, low spatial data density would contribute more to error than lack of temporal data density and/or climate change during this period.

To facilitate interpolation we removed duplicate stations as much as possible. We identified stations duplicated among datasets by checking for identical locations (coordinates rounded to three decimal places). To detect remaining duplicates with larger discrepancies in location, the Levenshtein distance (number of character additions or deletions needed to match a word) between each station's name and the names of its five nearest neighbours was calculated. Distances less than 35% of the name length were considered matches. Finally, all stations less than 3000 m horizontal distance and 500 m vertical distance apart were also marked for merging. The climate data for duplicate stations records did not always match exactly. We averaged climate values for stations, but assigned location information based on the record from the dataset with the most stringent error-checking procedures (see Table S1 for rankings). Weights were assigned based on the period of record. When period of record was not available, an intermediate value was assigned. The number of stations for each variable is reported in Table 2, and by region in Figure S1.

2.3. Covariates from satellite sensors

Global extent surfaces of mean monthly cloud cover, maximum (13:30 h local time) and minimum (1:30 h local time) land surface temperature (LST) were compiled from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data archives using Google Earth Engine (Google

Table 1. Major data sets used for covariates and climate elements. An expanded list for climate elements in Table S1.

Dataset	Variables	Dataset type/range	Distribution/resolution	Citation	URL
Covariates					
Oxford Malaria Atlas Project	LST day, LST night	Raster time-series	Global, 5 km	Weiss <i>et al.</i> , 2014	http://www.map.ox.ac.uk/map-earth-engine-meta-data/
Gap-Filled LST	Cloud cover	Raster time-series	Global, 1 km	NASA LP DAAC, 2014	https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod09ga
MOD09GA					http://srtm.csi.cgiar.org/
SRTM	Elevation	Raster	60°N–60°S, 30 arc-second	Farr <i>et al.</i> , 2007	
GTOPO 30	Elevation	Raster	Global, 30 arc-second	USGS, 1996	https://lta.cr.usgs.gov/GTOPO30
Climate elements					
Climate Research Unit Time-series v3.22 observations	t_{\min} , t_{\max} , t_{avg} , prec	Time-series 1901–present	Global	Harris <i>et al.</i> , 2014	http://www.cru.uea.ac.uk/data
The Berkeley Earth project	t_{\min} , t_{\max} , t_{avg}	Time-series 1750–present	Global	Rohde <i>et al.</i> , 2013	http://www.berkeleyearth.org
Global Historic Climate Network Dataset (GHCN) version 3	t_{\min} , t_{\max} , t_{avg}	Time-series 1850–present	Global	Lawrimore <i>et al.</i> , 2011	http://www.ncdc.noaa.gov/ghcnm/v3.php
Global Historic Climate Network Dataset (GHCN) version 2	prec	Time-series 1850–present	Global	Peterson and Vose, 1997	http://www.ncdc.noaa.gov/ghcnm/v2.php
WMO climatological normals (CLINO) for 1961–1990	t_{\min} , t_{\max} , t_{avg} , prec, vp, sol, wind	Summary 1961–1990	Global	WMO, 1996	http://gcmd.nasa.gov/records/GCMD_gov.noaa.ncdc.C00058.html
FAOCLIM 2.0 global climate database	t_{\min} , t_{\max} , t_{avg} , prec, vp, sol, wind	Summary 1960–1990	Global	FAO, 2001	http://geonetwork3.fao.org/climpag/agroclimdb_en.php
Center for Tropical Agriculture (CIAT) in Colombia	t_{\min} , t_{\max} , t_{avg} , prec	Summary	Tropics	See Hijmans <i>et al.</i> , 2005	
European Climate Assessment and Dataset monthly (ECAm)	t_{\min} , t_{\max} , t_{avg} , prec, vp	Time-series 1850–present	Europe	Tank <i>et al.</i> , 2002	http://www.ecad.eu
WorldClim v1	t_{\min} , t_{\max} , t_{avg} , prec	Summary 1960–1990	Global	Hijmans <i>et al.</i> , 2005	https://databasin.org/datasets/15a31dec689b4c958ee491ff30fce75
Global Summary of Day (GSOD)	t_{\min} , t_{\max} , t_{avg} , prec, vp, sol, wind	Time-series 1920–present	Global	NCEI, 2015	https://data.noaa.gov/dataset/global-surface-summary-of-the-day-gsod

Table 2. Climate element summary and covariates used in model building. Raw N stations indicates the number of stations accumulated before duplicate removal. N used indicates the number of stations used in interpolations after removal of duplicates and selection criteria. N Tier 1 indicates the number of stations meeting the target temporal range (> 25 years between 1970 and 2000). Covariates include mean MODIS cloud cover (Cld), distance to oceanic coast (cdist), elevation (elev), MODIS daytime land surface temperature (LST max), nighttime land surface temperature (LST min), average of night and daytime land surface temperatures (LST mean) and top-of-atmosphere incident solar radiation calculated from latitude (ETR).

Element	Raw N stations	N used	N tier 1	Unit	Covariates examined
Precipitation	60 419	34 542	13 763	mm	Cld, cdist, elev
Minimum temperature	36 772	16 883	9967	°C	LST min, cdist, elev
Maximum temperature	37 092	16 988	10 142	°C	LST max, cdist, elev
Mean temperature	47 453	20 268	10 273	°C	LST mean, cdist, elev
Vapour pressure	13 873	9541	3148	kPa	Cld, cdist, elev
Wind speed	12 831	10 149	77	m s ⁻¹	Cld, cdist, elev
Solar radiation	9118	5489	2820	MJ m ⁻²	Cld, cdist, elev, ETR

Inc., 2015). Monthly cloud cover was calculated as the percentage of time a pixel contained a positive flag in the cloud state band (bits 1 and 2) across all MOD09GA images (NASA LP DAAC, 2014) from a given month from 2001 to 2013, not counting any missing values. Minimum and maximum surface temperature were derived from the Oxford Malaria Atlas Project Gap-Filled LST dataset based on MODIS data (Weiss *et al.*, 2014) which has a spatial resolution of ~ 5 km resolution, averaged across years 2001–2013 by month and resampled via bilinear interpolation to our ~ 1 km spatial resolution. This product was used to due to the exaggerated levels of local variation (noise) and persistent gaps found in the raw MODIS LST data (e.g. Southeast Asia in monsoon season; Frey and Kuenzer, 2014). For complex shorelines and islands not captured at the Oxford dataset's resolution, mean night and day LST values were extracted from ~ 1 km resolution MODIS MOD11A1 images, averaged by month from 2001 to 2013. An average of the (aggregated) maximum and minimum LST values was used as a covariate for mean temperature.

We used elevation data from the Shuttle Radar Topography Mission (SRTM) aggregated to 30 arc-second spatial resolution, using the median value (data available at <http://srtm.csi.cgiar.org/>). We used the GTOPO30 (USGS, 1996) database for the areas north of 60°N and south of 60°S where no SRTM data was available. For each grid cell we computed the distance to the nearest cell that was an ocean or sea. Bays and inlets have a subdued marine effect compared to open ocean, but we did not consider that.

2.4. Creation of climate surfaces

Climate variables were interpolated with thin-plate splines using the program SPLINA from ANUSPLIN 4.4 (Hutchinson and Xu, 2013). Processing time for spline estimates using this method is proportional to the square of the number of data points, thus a series of models were fit to 23 geographically overlapping subsets of station data (hereafter 'regions', Figure 2) to reduce computation time and to allow for regional model selection. Regional surfaces were merged by weighting estimates in overlapping regions inversely proportional to distance from each region's border (New *et al.*, 2002, Hijmans *et al.*,

2005). Regions were chosen to ensure at least 1 degree of overlap, span bioclimatic zones, and to a target number of interior stations between 1000 and 4000. The set of regions is different for each variable (Figure 2). In the case of precipitation, the set included some relatively small regions (e.g. two to cover most of Mexico and three for eastern Australia) with about 1000–2000 rainfall stations, as well as much larger regions (Canada and parts of the USA; Russia and areas to its west and south) with almost 4000 stations (Figure 2).

In ANUSPLIN, spline models of the N observed data values z_i are fit following the form:

$$z_i = f(x_i) + b^T y_i + e_i \quad (i = 1, \dots, N)$$

where f is a smooth function of the spline independent variables x_i , b is a vector of linear coefficients for the independent covariates y_i and e_i is an independent, zero mean error term (Hutchinson and Xu, 2013). For each climate variable, multiple model formulations were tried using different combinations of covariates either as independent spline variables or linear covariates (Table 1). Elevation (in m) was divided by 1000 following scaling recommendations by Hutchinson (1995) and the distance to coastline was hyperbolic tangent transformed prior to model fitting to emphasize variation near the coast relative to the interior. For solar radiation, computed extraterrestrial radiation (i.e. the solar radiation in the absence of atmospheric effects) was also included as a predictor variable. Precipitation, radiation, vapour pressure and wind speed values were square root transformed prior to fitting following recommendations in Hutchinson and Xu (2013). The order of each spline surface was set at the lower limit as determined by the program, based on the number of variables in each model.

After initial model fitting, we looked for outliers by examining stations with high residuals (difference between observed and predicted values). In many cases, these stations had different units from what was reported, or wrong location information. Where possible, these errors were corrected. Outliers (that is, stations for which the model residuals were very large) for which the source of possible error was not clear were only removed if climate values were highly unlikely, such as if a station's time-series

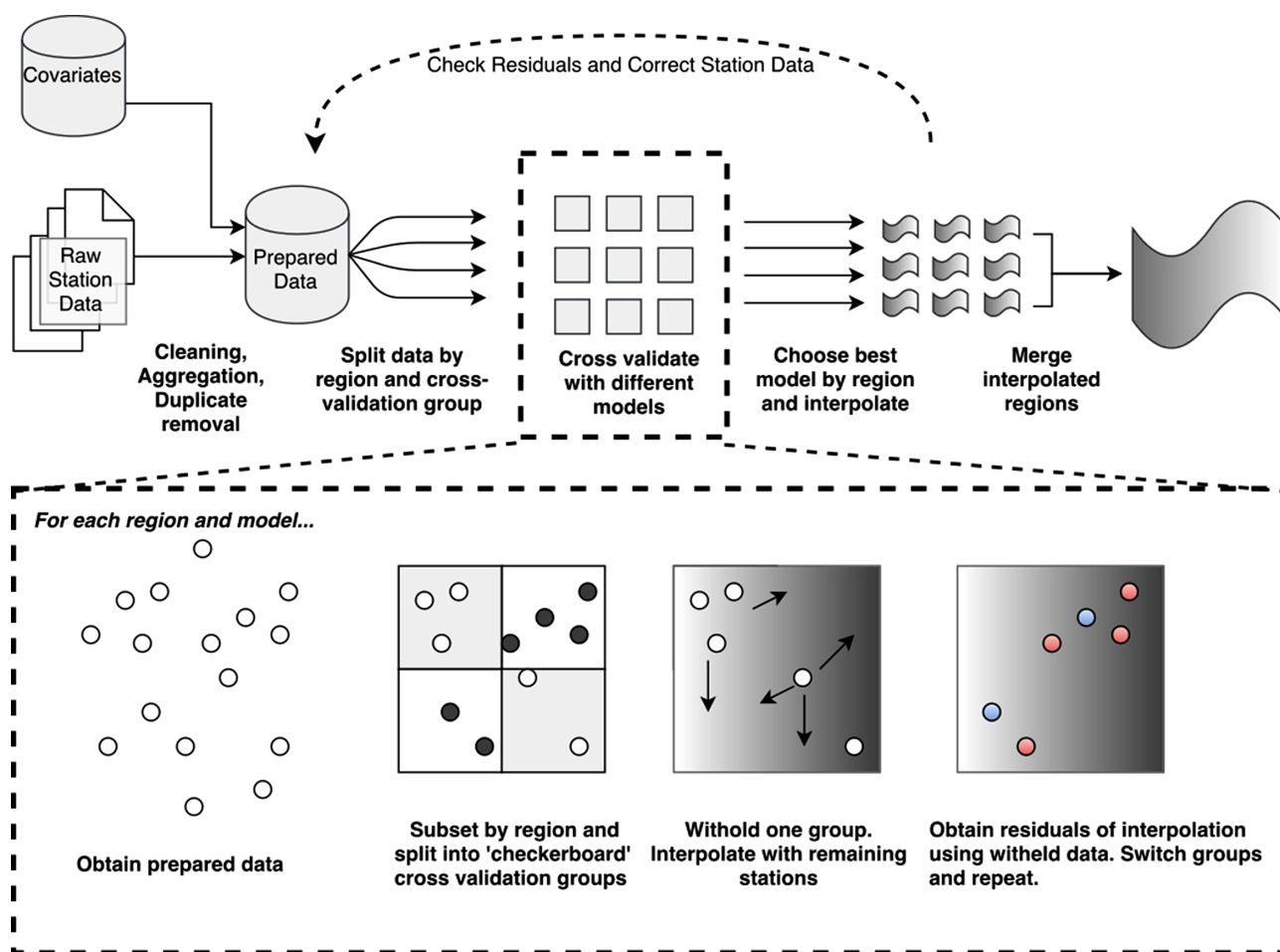


Figure 1. Workflow for climate surface creation. The twofold spatially stratified cross-validation approach detailed within dashed box. [Colour figure can be viewed at wileyonlinelibrary.com].

was strongly incongruous with nearby stations, or clearly wrong for a given location. Although scanning for large residuals is helpful for identifying individual outlying stations, this method may not be effective for discovering systematic errors in datasets (Hijmans *et al.*, 2005). We looked for such systematic discrepancies by visually comparing our first-pass climate surfaces with those previously published (New *et al.*, 1999; Hijmans *et al.*, 2005).

2.5. Evaluation

Cross-validation of models built with spatial point datasets using randomly selected subsets tends to be problematic because the high spatial autocorrelation at close ranges will lead to test-score inflation if the points are geographically clustered (Hijmans, 2012). To characterize model performance in the absence of nearby ‘control’ stations, models were evaluated using spatially stratified twofold cross-validation (Figure 1). A $3^\circ \times 3^\circ$ ‘checkerboard’ grid was used to determine membership in cross-validation groups, based on station location (either in a ‘black’ tile or ‘white’ tile). This assured that testing data was generally far away from training data. For each region, separate spline surfaces were fit using only stations from one group or the other (approximately half for each fold), then

predicted values from these surfaces were compared with observed values in the withheld group for both sets. We computed root mean square error (RMSE) and Pearson’s correlation coefficient (ρ) between predicted and observed estimates and report the average of these for the two test data sets.

Final surfaces were created by selecting only the model with the lowest cross-validation RMSE value for each region, fit to the full dataset (Table S2). This resulted in combinations of surfaces from different model formulations across regions of the earth’s surface. We did not consider model averaging or selecting different models for each month.

3. Results and discussion

3.1. Station data varied in quality and distribution through space and time

The spatial distribution of stations for all climate variables was uneven, with higher station densities generally corresponding to regions of higher population density (Figure 2). There was an apparent effect of country-wise investment in climate data collection and distribution, with countries such as Mexico, Australia and Germany, for

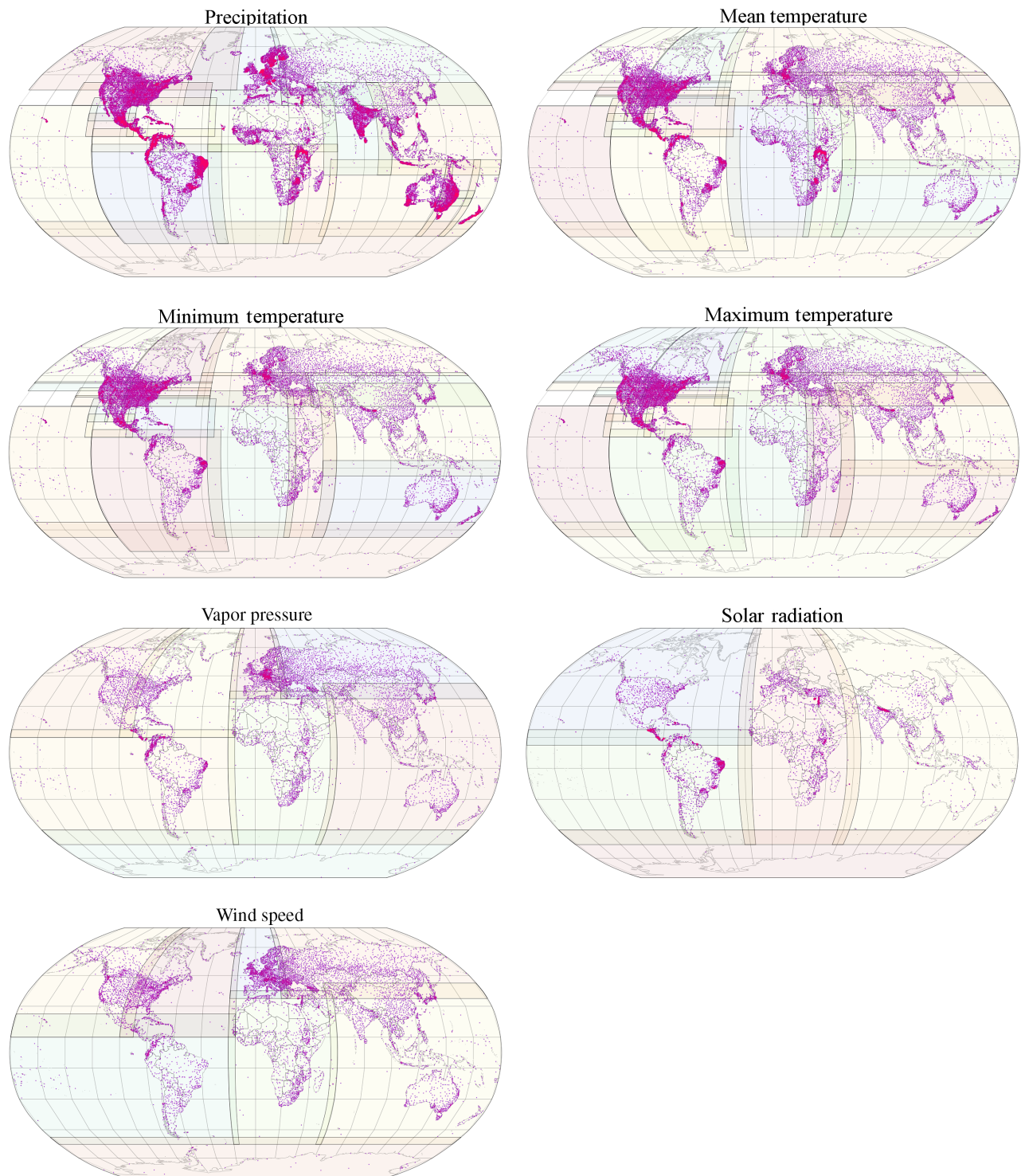


Figure 2. Spatial distribution of weather stations used for the different climate variables: Coloured boxes indicate boundaries of regions used for creating spline surfaces. See Figure S1 for details. [Colour figure can be viewed at wileyonlinelibrary.com].

example, having much higher densities of precipitation stations than neighbouring countries. Regions with sparse station data included islands in the Pacific, areas with high elevations, Greenland and Antarctica (Figure 2).

The temporal density of available climate observations for all variables peaked around 1980, with a sharp decline in number of observations approaching 2010 (Figure S2), as observed in other studies (Harris *et al.*, 2014). For all

variables except solar radiation, the number of stations meeting the target temporal criteria (between 1970 and 2000, ≥ 25 years of data) ranged from approximately 30–50% of stations used for interpolation (Table 2, Figure S2). Solar radiation data had a particularly low number of stations meeting our highest criteria (only 77 stations) reflecting the paucity of high-quality solar radiation data available.

Table 3. Global twofold cross-validation statistics for selected models. RMSE indicates root-mean-squared error, MAE is mean absolute error, and relative error was calculated as the average absolute log proportion of observed to predicted value, corresponding approximately to percent change.

Climate variable	Correlation coefficient (ρ)	RMSE	MAE	Relative error	Unit
Minimum temperature	0.993	1.39	0.949	-	C
Maximum temperature	0.995	1.29	0.850	-	C
Average temperature	0.996	1.12	0.715	-	C
Precipitation	0.861	49.46	21.290	30.4	mm
Vapour pressure	0.990	0.12	0.073	7.4	kPa
Wind speed	0.759	1.14	0.796	31.8	m s^{-1}
Solar radiation	0.965	1.45	0.981	6.7	MJ m^{-2}

3.2. Overall model accuracy

Global model accuracy was very high for the temperature variables, when selecting the best model from each region. All temperature variables had a global correlation coefficient (between estimated and observed values) of 0.99 or higher and an average RMSE between 1.1 and 1.4 °C (Table 3). Model accuracy was also very high for solar radiation and vapour pressure both with correlation coefficients greater than 0.95. Accuracy was lowest for precipitation ($\rho = 0.86$) and windspeed ($\rho = 0.76$). This ordering is similar to what was reported for previously published climate surfaces (New *et al.*, 2002; Hijmans *et al.*, 2005; Table S3). Precipitation can be highly variable in time and space and some regions have abrupt changes (rain-shadows) whereas temperature generally follows relatively simple gradients of latitude and elevation; and even where modified by, for example, coastal effects, transitions may be gradual (but there are exceptions to this such as the abrupt change observed in summer temperatures along the coast of central California). Such very strong gradients tend to come out too smooth in the climate surfaces.

Global cross-validation statistics may obscure important variation in prediction accuracy at high altitudes and remote locations where there are relatively few stations to evaluate (Figure 3). In general, prediction error increased with station elevation and distance to nearest neighbouring station (in the training set) for all variables (Figure 3). Poor interpolation accuracy at higher elevations might be associated with lower station densities (Hijmans *et al.*, 2005; Kilibarda *et al.*, 2014), although more complex mountain terrain is also a potential explanatory factor (though not all high elevations have topographic complexity). Generalized additive models (GAMs) of cross-validation errors showed that higher elevations tended to be associated with lower interpolation accuracy, even after accounting for the effects of isolation and spatial variation in errors, although this effect differed among variables. Error tended to decline as a function of distance from the coast for precipitation, vapour pressure, wind speed and solar radiation, but not for temperature. This may be indicative of relatively high spatial variation in these variables near the coast, and of the lack of oceanic (buoy) weather stations.

3.3. Spatial variation in prediction error

The spatial distribution of prediction errors, as measured in RMSE (root-mean-squared error; Figure 4) and

proportional change (Figure 5) generally reflect the sensitivity of interpolations to both station density and topographic complexity, with much higher uncertainty in isolated, mountainous regions. For example in the continental United States, there is a notable east–west split in terms of RMSE for all variables, with higher RMSEs found in the relatively remote and mountainous western part of the country. Isolated locations on oceanic islands, Greenland and Antarctica consistently had high RMSE values, as well as the band of mountain ranges across southern Siberia and the central Andes.

For variables other than temperature (precipitation, vapour pressure, wind speed and solar radiation) absolute RMSE was high in regions where the magnitude of observed values are also high. RMSE for precipitation was high in the tropics, and errors in wind speed were highest on the coast, in locations with a large range and variance of recorded wind speeds. Solar radiation and vapour pressure also exhibited largest absolute errors near the equator where incident radiation and humidity are highest.

Relative error in these variables reveals a different pattern, with highest values in areas with complex topography and low station density. Relative error in precipitation is notably high in dry regions (Figure 5(a)), which is a consequence of small absolute differences in predicted precipitation being large relative to total rainfall. However, regions such as the Western Gats of India and the coastal north-west United States also have high relative error rates, indicating higher uncertainty in these typically wet and topographically complex areas. Relative error in wind speed was very small in the great plains of the United States, corresponding to the dense station network in this area and flat terrain. Solar radiation error (in both absolute and relative terms) is highest in tropical montane regions, particularly in Peru and Ecuador. Sharp gradients in solar radiation in tropical montane regions related to cloud cover may drive much of this uncertainty, in spite of the inclusion of mean cloud cover as a covariate for interpolation.

3.4. LST improved estimates of temperature

For all temperature variables, interpolations including satellite observed LST as either a covariate or independent spline were consistently superior to basic trivariate splines of latitude, longitude and elevation (Table S2). The benefit of including LST varied between regions and was

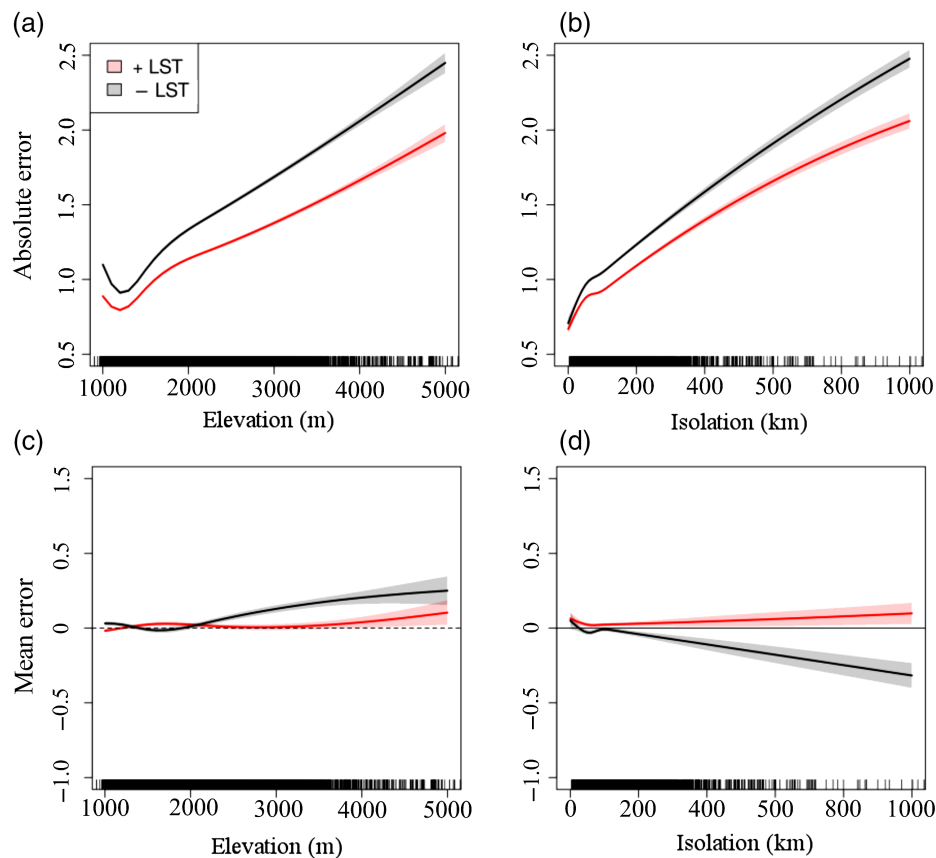


Figure 3. Cross-validation mean and absolute errors ($^{\circ}\text{C}$) as a function of elevation (a, c) and isolation (b, d) for minimum temperature across all regions. Errors from models including MODIS land surface temperature (with the lowest RMSE for each region) indicated by dashed red lines. Shaded regions indicate 95% confidence interval for tensor product smooths. Isolation was measured as the distance between a station and its nearest control point in the corresponding cross-validation set. Observations indicated by vertical ticks below graph. [Colour figure can be viewed at wileyonlinelibrary.com].

particularly low in for average and maximum temperature in the eastern United States. In these cases inherent bias in the relationship between LST and air temperature (Figure 6) may have introduced error, perhaps because variation or bias in LST may have been high relative to station density and/or climatic gradients.

The relationship between LST and recorded temperature is noticeably biased for maximum (and, by derivation average) temperature, with systematically higher and lower LSTs recorded at high and low maximum temperatures, respectively (see Figure 6). This phenomenon is attributable to increased radiative energy at the earth's surface under hot, dry conditions (in the absence of forest canopy cover) and to exaggerated estimates due to radiative cooling if LST estimates are from clear-sky days in the winter (Hulley and Hook, 2009; Mildrexler *et al.*, 2011; Van De Kerchove *et al.*, 2013).

Addition of LST as a covariable was effective for improving estimates at high elevations and remote locations, particularly for minimum temperature (Figure 3). Although error remained high for such locations, the magnitude of error was reduced when satellite data was included (Figures 3(a) and (b)). Bias at high elevations and remote locations was also reduced by including satellite covariates (Figures 3(c) and (d)).

Inclusion of LST data for spatial interpolation of temperatures in other studies has provided mixed results. Parmentier *et al.* (2014) found models using only latitude, longitude and elevation to be consistently superior, although LST provided fine grained structure to interpolations. Kilibarda *et al.* (2014) found that LST improved estimates for maximum but not minimum or average temperature. Similarly Parmentier *et al.*, (2015) found that LST slightly improved daily maximum temperature estimates in summer months in Oregon, likely due to a stronger correlation between LST and maximum air temperature than for elevation and air temperature during these months. Hengl *et al.* (2011) and Yao and Zhang (2013) suggested that LST data improved interpolation of temperature data but they did not directly test the effects of LST on model accuracy.

Despite strong correlations between LST and air temperature, the amount of information LST provides for interpolation may be marginal relative to that of the information provided by geographically proximal stations in the spline model, even when those stations are located far away. A comparison of LST estimates to naïve predictions based on the values from the nearest station (i.e. k -nearest neighbour estimates for $k=1$) suggests that even at distances greater than 250 km away, nearest

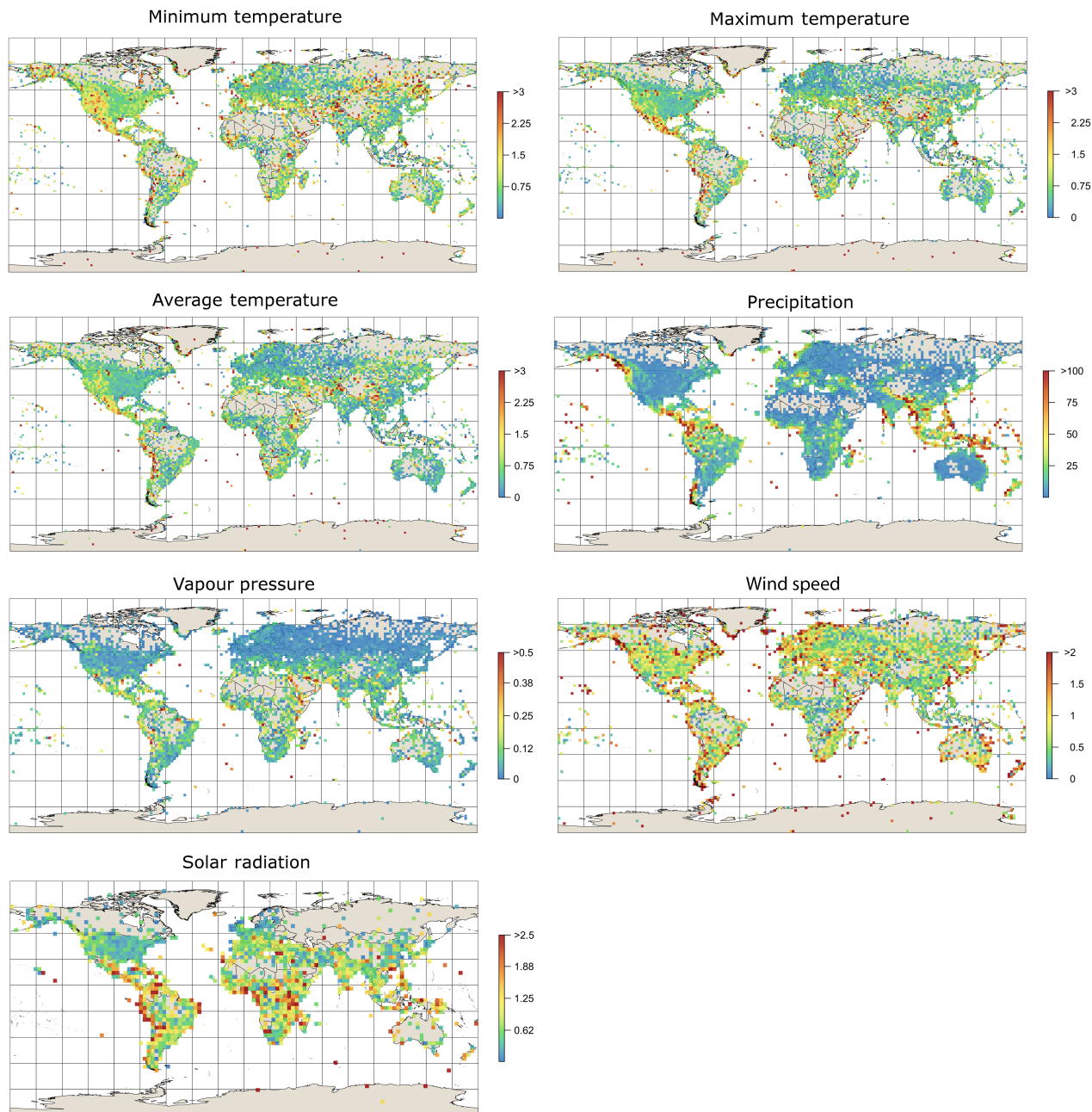


Figure 4. Spatially aggregated RMSE values obtained with twofold cross-validation. Values aggregated to 1.5° tiles (temperature and precipitation), 2° tiles (vapour pressure and wind speed), and 3° (solar radiation) to aid visualization of sparse data. [Colour figure can be viewed at wileyonlinelibrary.com].

neighbour values are generally more accurate than LST. It should also be noted that there is a temporal mismatch between the data used at stations (1970–2010) and average LST (2000–2013). LST may become more useful as the time period over which there is data increases, particularly in places with very few weather stations and in areas with strong gradients or topographically complex terrain.

3.5. Cloud cover had marginal effect on precipitation estimates

Models with cloud cover as a covariate or independent spline variable were virtually indistinguishable from

baseline trivariate spline models of latitude, longitude and elevation only (Table S2). Although satellite covariate models were selected in a number of regions as superior in terms of RMSE, the difference was marginal, usually only around ~ 1 mm improvement in precipitation RMSE values (Table S2). There was no apparent improvement in cross-validation error or bias for stations as either a function of elevation or isolation.

One notable improvement in accuracy occurred in the Pacific region where an interpolation based on latitude, longitude and cloud cover had a cross-validation correlation coefficient of 0.6 compared to 0.36 for a model based on coordinates and elevation (Table S2). Climate

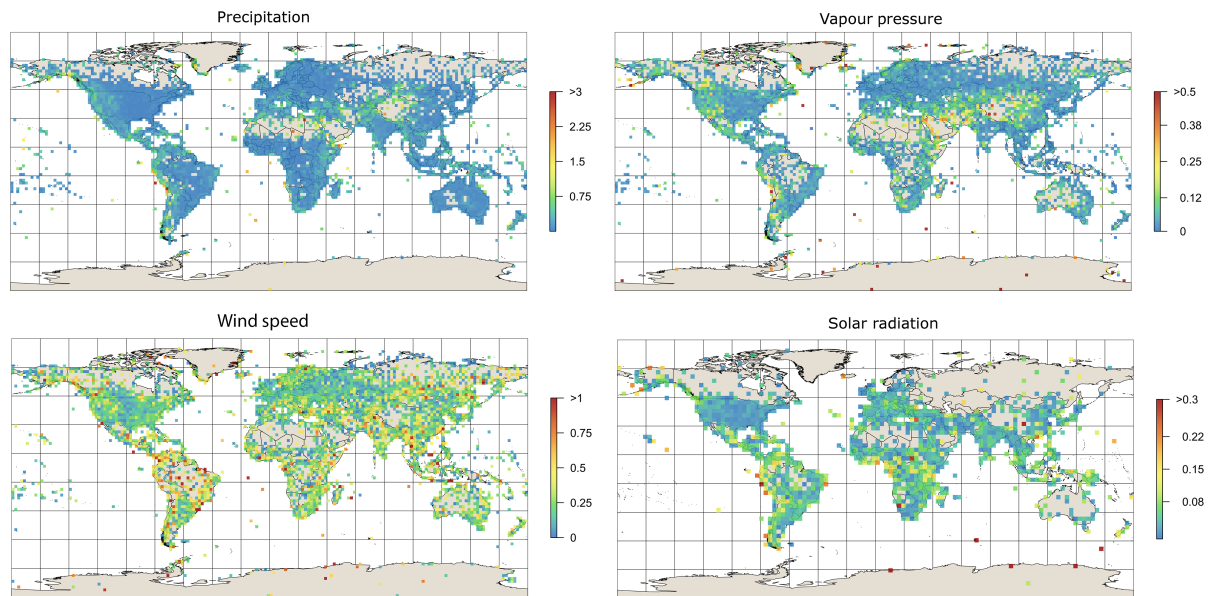


Figure 5. Spatially aggregated proportional change between observed and predicted values in twofold cross-validation. Calculated as the average absolute value of the logarithm of predicted/observed values for each grid cell in cross-validation. Values aggregated to 1.5° tiles (temperature and precipitation), 2° tiles (vapour pressure and wind speed), and 3° (solar radiation) to aid visualization of sparse data. [Colour figure can be viewed at wileyonlinelibrary.com].

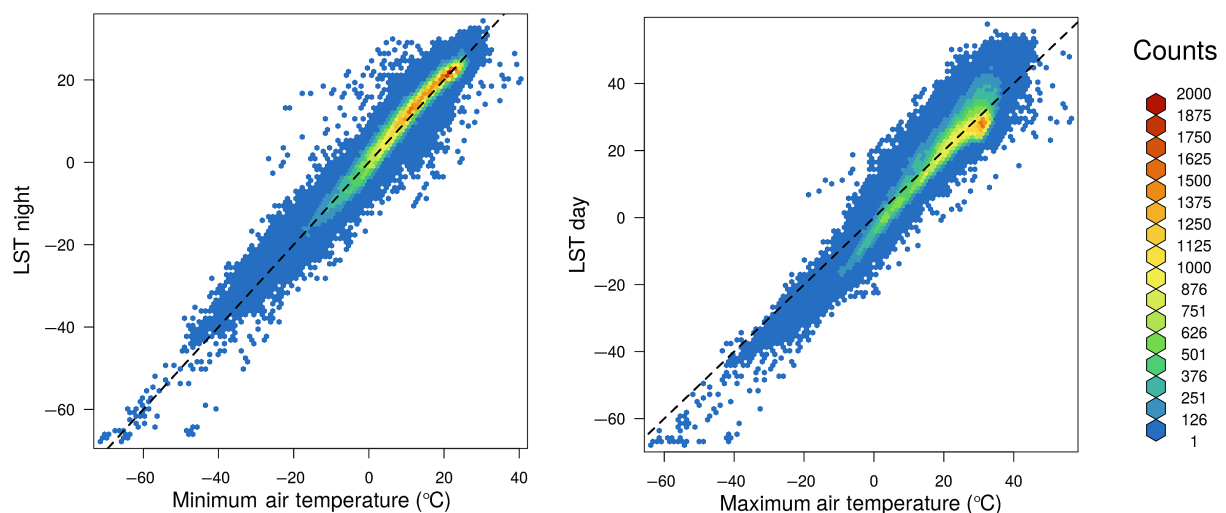


Figure 6. Relationship between air temperature at weather stations and average Land surface temperature (LST) for the same locations. The dashed line indicates a 1 : 1 relationship. The larger bias and scatter for LST day is related to higher surface temperatures under hot conditions (except under forest canopies), and clear-sky bias for low temperatures. Correlations coefficients were 0.947 and 0.974 for maximum temperature and minimum temperature respectively (using 200 000 observations for each). [Colour figure can be viewed at wileyonlinelibrary.com].

interpolation in this region is difficult due to very low station density and sharply contrasting microclimates (e.g. rainshadows on the Hawaiian islands, Figure 7). Given these constraints the addition of cloud cover emphasized these topographic effects and avoided over-smoothing relationships between elevation and precipitation with stations at equivalent elevations across rain-shadows (Daly *et al.*, 2008).

Lack of model improvement due to inclusion of cloud cover may be due to the noisy relation between cloud cover and precipitation. Although precipitation generally increased with cloud cover (Figure 8), the relationship is extremely noisy: precipitation varies in almost five orders

of magnitude for a given level of cloud cover. This is not entirely surprising, as conditions where average cloudiness is high and precipitation is relatively low (for example, winter conditions along the Pacific coast in Peru) and vice versa (thunderstorms at the end of otherwise clear days) are common. In addition, there are inconsistencies in satellite cloud cover quality (Leinenkugel *et al.*, 2013; Wilson *et al.*, 2014).

3.6. Average MODIS-derived cloud cover marginally improved solar radiation estimates

Solar radiation estimates benefitted from satellite-derived information, with the greatest reduction in

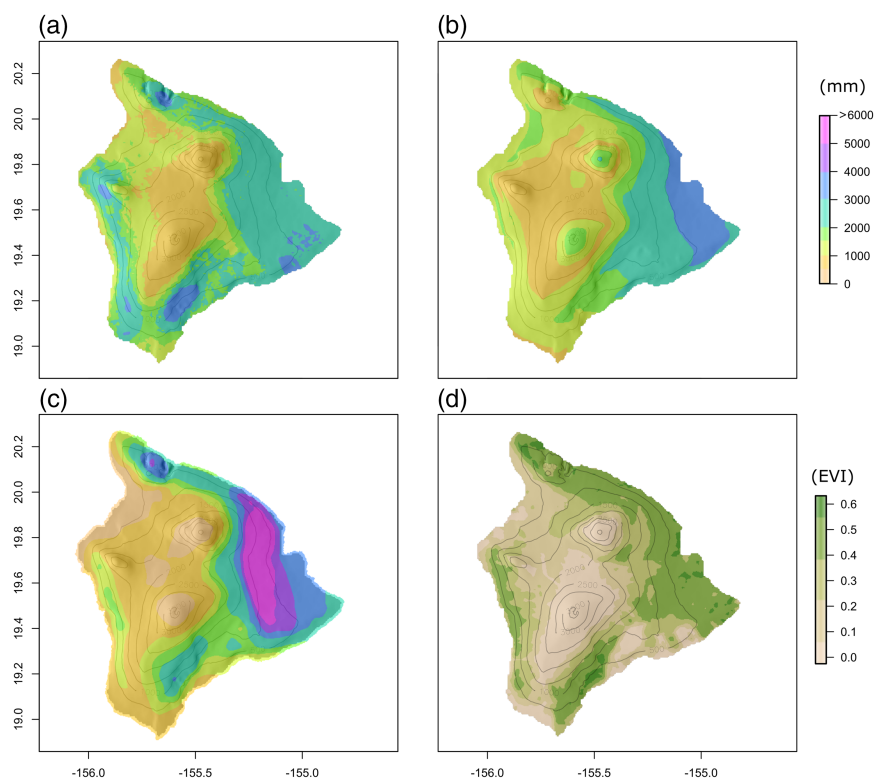


Figure 7. (a) Cumulative annual rainfall interpolated with latitude, longitude and cloud cover as independent spline variables, (b) cumulative annual rainfall interpolated with latitude, longitude and elevation as independent spline variables, (c) Hawai'i rainfall atlas (Giambelluca *et al.*, 2013), and (d) average MODIS Enhanced Vegetation Index (EVI) values on the island of Hawai'i highlighting the sharp rainshadows on the island. Note how interpolations with cloud cover maintain sharp rainshadows (e.g. on the north of the island) but introduce artefacts related to fog or cloud pooling (dark spots), while the spline with elevation smooths over these features and overestimates rainfall at the highest elevations. Compared to the Hawai'i rainfall atlas, both surfaces miss the precipitation maximum of over 6000 mm on the lower northeastern (windward) slope, just below the top of the well-known trade wind inversion. This maximum does not correlate well with the cloudiness layer used for interpolating (a). [Colour figure can be viewed at wileyonlinelibrary.com].

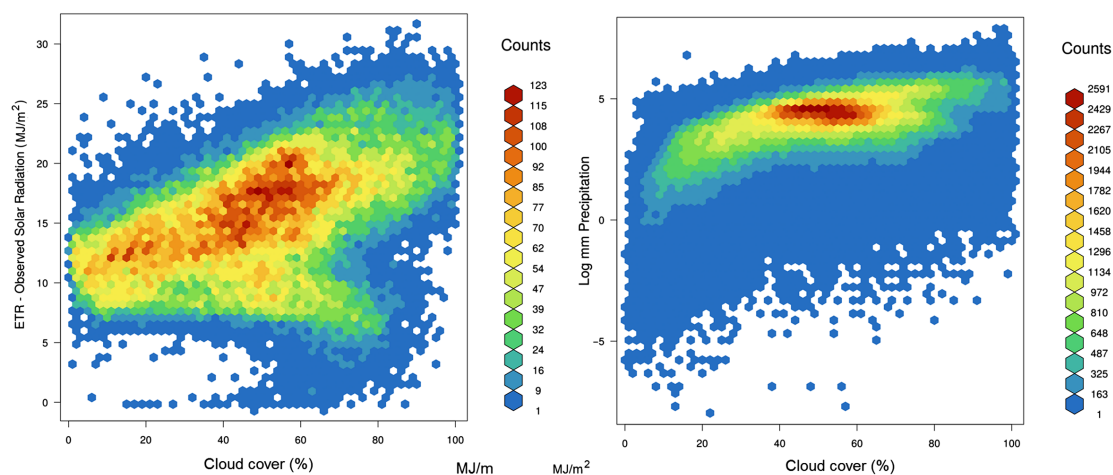


Figure 8. Relationships between climate variables and cloud cover obtained from MODIS satellite images. (a) Extraterrestrial solar radiation (ETR) minus observed solar radiation and (b) Log transformed precipitation (mm). [Colour figure can be viewed at wileyonlinelibrary.com].

root-mean-square error (RMSE) and mean absolute error (MAE) obtained when average cloud cover was used as a linear covariate (Table S2). However the changes in RMSE compared to baseline models were very small (about 2%). Extraterrestrial radiation (ETR) did not tend to improve interpolation estimates, perhaps due to the fact that variation in ETR is a direct function of latitude,

which was already included in the models. Splines, which included ETR generally, tended to perform worse than a simple model based on latitude, longitude and elevation (Table S2).

One consequence of including satellite data in interpolations was the propagation of errors related to artefacts originating from the satellite data itself. For solar

radiation interpolations which included cloud cover as a covariate, predictions tended to be erroneously low in areas with expansive white-coloured ground surfaces, such as the Salar de Uyuni salt flat in the Bolivian Altiplano or the White Sands area in New Mexico, USA. The MODIS cloud cover algorithm has a tendency to overestimate coverage over high-albedo surfaces (Wilson *et al.*, 2014), which apparently occurred for salt flats in the mean cloud coverage surfaces.

3.7. Distance from oceanic coast improved estimates for vapour pressure and wind speed

Including satellite predictors (cloud cover) did not improve estimates for either vapour pressure or wind speed, and often made them worse (Table S2). Given the weak relationship between cloud cover and these variables (Figure 8), this is not surprising. Adding distance from the coast, however, tended to improve estimates for these variables (Table S2). For both variables, models consisting of latitude, longitude, elevation and coast distance were the most frequently selected across regions, however differences in aggregated RMSE statistics were marginal, rarely exceeding approximately 0.05 m s^{-1} and 10 Pa over baseline models for wind speed and vapour pressure respectively. We used simple distance to the coast as proxy for coastal processes, and interpolation accuracy could likely be improved with covariates that are more directly related to coastal climatic gradients due to topography and marine layer penetration.

4. Conclusions

The climate surfaces generated in this study represent an expansion upon previous efforts in terms of spatial and temporal scope, climate stations used, covariates tested and climate variables interpolated. Final surfaces generally had high correspondence with observations, although interpolation of some variables (e.g. precipitation and wind speed) was not as accurate as others (notably temperature), as was also reported in previous studies.

Adding satellite-derived data either as covariates or independent spline variables had the greatest positive impact on accuracy for temperature variables, particularly in isolated and/or high elevation locations. For other climate variables (precipitation, vapour pressure, wind speed and solar radiation) addition of satellite surfaces had marginal or even negative effects, which was somewhat surprising. Lack of strong correlations between climate variables and satellite predictors, lack of station density to resolve these relationships across space, noise and bias within satellite predictors, as well as inability of spline algorithms to properly fit these relationships may be contributing to this result. Several studies have noted diminishing returns when increasing the number of predictor variables (such as distance to coast, aspect) in interpolating climate surfaces, using splines and other methods (Hutchinson, 1995; Jarvis and Stuart, 2001a; Daly *et al.*, 2008; Neteler, 2010; Van De Kerchove *et al.*, 2013; Kilibarda *et al.*, 2014).

In our study, optimal spline model formulations and useful variables varied from region to region. Local context has a strong influence on climate processes (e.g. in the effect of elevation on precipitation; Brunsdon *et al.*, 2001), and the importance of predictor variables (e.g. cloud cover or distance to the coast) may vary regionally. In this study we used a novel approach by using different spline model formulations across different regions. Although differences in global statistics were similar between this method and single model formulations, this adaptive method allowed for better model fits in remote regions, such as with precipitation in the Pacific. Similarly, Hofstra and New (2009) found that optimal model formulation for angular-distance weighted interpolation varied between regions and seasons. Identifying the optimal delineation of regions for different interpolations is a potential area for future research. Future work could also look more at alternative algorithms and specifications of the spline model and its effect on accuracy for locations that are nearby and further away from weather stations. Instead of selecting the best model, model averaging could be used.

Satellite-derived predictors improved in interpolation accuracy at the extremes of elevation and isolation (as well as distance to coast), particularly for temperature variables. This was not always apparent in global statistics of interpolation accuracy. Global cross-validation statistics may obscure important differences between model formulations in prediction accuracy at high altitudes and remote locations as there are very few stations there to evaluate (Figure 3). Although we reduced this problem by using spatially stratified cross-validation as that increases the average distance between model training and testing data, the majority of the testing stations were still located near populated (and developed) regions of the world, which are associated with high station densities and, consequently, high interpolation accuracies.

While there is significant emphasis on uncertainty for future climate model projections (e.g. Foley, 2010), much less attention has been paid to uncertainty in spatially extensive estimates of current climate (the climate of the recent past) at scales relevant for environmental research. Poor climate station data quality affects resulting climate surfaces and consumes resources for quality-control efforts. Moreover, the need to aggregate vast amounts of climate data from disparate sources is an added hurdle for progress and comparisons across modelling studies, since climate surfaces will inherently vary based on underlying data. Previous efforts to standardize and compile global climatological data are noteworthy, however these efforts have tended to be limited to select institutions or groups of researchers. For these reasons we appeal for the creation of a cooperatively managed, standardized and open-source database of climatological data, analogous to highly successful efforts for other types of global data such as the Global Biodiversity Information Facility (GBIF; <http://www.gbif.org/>) database.

The fact that most of our climate surface estimates were only marginally improved by use of satellite covariates

highlights the importance having a dense, high-quality network of climate station data. It has been noted that for standard spatial interpolation methods, there appears to be a limit to which additional spatial covariates improve model estimates (Jarvis and Stuart, 2001a, 2001b). For the data and methods employed in the creation of these surfaces, inherent noise in satellite spectral data and lack of correspondence with on-the-ground measurements likely contributed to this unexpected result. However, the temporal record of MODIS is still relatively short, and we did not explore all available satellite data, and more satellite products will become available at high spatial or temporal resolutions. These products, in combination with modelling methods appropriately sensitive to local context should be evaluated for improved global interpolations. Improving the quality of precipitation estimates is particularly important, and future work could evaluate the use of data from the Tropical Rainfall Measuring Mission (TRMM) and Global Precipitation Measurement (GPM) satellites as covariates.

Acknowledgements

Funding for this project was provided by the Feed the Future Sustainable Intensification Innovation Lab (SIIL) through the United States Agency for International Development (grant # AID-OOA-L-14-00006). We would like to thank Alex Mandel for technical and logistical support as well as George Hsu and Charmaine Bonifacio for assistance with data preparation. We thank Andrew Latimer and Truman Young for helpful suggestions. We thank Peter G. Jones for his visionary work on the CIAT climate database which provided the original starting point for the WorldClim project. This manuscript was improved by the comments of five anonymous reviewers.

Supporting information

The following supporting information is available as part of the online article:

Figure S1. Region boundaries, number of stations, and chosen models used for interpolation.

Figure S2. Distribution of climate stations by temporal frequency.

Table S1. Data sources.

Table S2. Cross-validation root-mean-square errors (RMSEs) by climate element, model and region.

Table S3. Comparison of RMSE and correlation values for WorldClim 1 and WorldClim 2.

References

Brunsdon C, McClatchey J, Unwin DJ. 2001. Spatial variations in the average rainfall–altitude relationship in Great Britain: an approach using geographically weighted regression. *Int. J. Climatol.* **21**: 455–466.

Daly C, Halbleib M, Smith JJ, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.* **28**: 2031–2064.

FAO. 2001. *FAOCLIM 2.0 A World-Wide Agroclimatic Database*. Food and Agriculture Organization of the United Nations: Rome, Italy.

Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Alsdorf D. 2007. The shuttle radar topography mission. *Rev. Geophys.* **45**(2): RG2004.

Foley AM. 2010. Uncertainty in regional climate modelling: a review. *Prog. Phys. Geogr.* **34**: 647–670.

Frey C, Kuenzer C. 2014. Land-surface temperature dynamics in the Upper Mekong Basin derived from MODIS time series. *Int. J. Remote Sens.* **35**: 2780–2798.

Giambelluca TW, Chen Q, Frazier AG, Price JP, Chen Y-L, Chu P-S, Eischeid JK, Delporte DM. 2013. Online rainfall atlas of Hawai'i. *Bull. Am. Meteorol. Soc.* **94**: 313–316.

Google Inc.. 2013. *Google Earth Pro*. Google Inc.: Mountain View, CA, USA.

Google Inc. 2015. Google Earth Engine. earthengine.google.com.

Harris I, Jones PD, Osborn TJ, Lister DH. 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol.* **34**: 623–642.

Hengl T, Heuvelink GBM, Tadić MP, Pebesma EJ. 2011. Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theor. Appl. Climatol.* **107**: 265–277.

Hijmans RJ. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* **93**: 679–688.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**: 1965–1978.

Hofstra N, New M. 2009. Spatial variability in correlation decay distance and influence on angular-distance weighting interpolation of daily precipitation over Europe. *Int. J. Climatol.* **29**: 1872–1880.

Hulley GC, Hook SJ. 2009. Intercomparison of versions 4, 4.1 and 5 of the MODIS land surface temperature and emissivity products and validation with laboratory measurements of sand samples from the Namib desert, Namibia. *Remote Sens. Environ.* **113**: 1313–1318.

Hutchinson MF. 1995. Interpolating mean rainfall using thin plate smoothing splines. *Int. J. Geogr. Inform. Syst.* **9**: 385–403.

Hutchinson MF, Xu T. 2013. *ANUSPLIN Version 4.4 User Guide*. Australian National University: Canberra.

Jarvis CH, Stuart N. 2001a. A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part I: the selection of “guiding” topographic and land cover variables. *J. Appl. Meteorol.* **40**: 1060–1074.

Jarvis CH, Stuart N. 2001b. A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part II: The interaction between number of guiding variables and the type of interpolation method. *J. Appl. Meteorol.* **40**: 1075–1084.

Jin M, Dickinson RE. 2010. Land surface skin temperature climatology: benefitting from the strengths of satellite observations. *Environ. Res. Lett.* **5**: 44004.

Kilibarda M, Hengl T, Heuvelink GBM, Gräler B, Pebesma E, Perčec Tadić M, Bajat B. 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *J. Geophys. Res. Atmos.* **119**: 2294–2313.

Lawrimore JH, Menne MJ, Gleason BE, Williams CN, Wuertz DB, Vose RS, Rennie J. 2011. An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J. Geophys. Res.: Atmospheres* **116**: D19121.

Leinenkugel P, Kuenzer C, Dech S. 2013. Comparison and enhancement of MODIS cloud mask products for Southeast Asia. *Int. J. Remote Sens.* **34**: 2730–2748.

Mildrexler DJ, Zhao M, Running SW. 2011. A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests. *J. Geophys. Res. Biogeo.* **116**: G03025.

NASA LP DAAC. 2014. MOD09GA Surface Reflectance Daily L2G Global 1km and 500m. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center.

NCEI. 2015. Global Surface Summary of the Day. National Oceanic and Atmospheric Administration.

Neteler M. 2010. Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data. *Remote Sens. (Basel)* **2**: 333–351.

New M, Hulme M, Jones P. 1999. Representing twentieth-century space-time climate variability. Part I: development of a 1961–90 mean monthly terrestrial climatology. *J. Climate* **12**: 829–856.

New M, Lister D, Hulme M, Makin I. 2002. A high-resolution data set of surface climate over global land areas. *Climate Res.* **21**: 1–25.

- Parmentier B, McGill B, Wilson AM, Regetz J, Jetz W, Guralnick RP, Tuanmu M-N, Robinson N, Schildhauer M. 2014. An assessment of methods and remote-sensing derived covariates for regional predictions of 1 km daily maximum air temperature. *Remote Sens. (Basel)* **6**: 8639–8670.
- Parmentier B, McGill BJ, Wilson AM, Regetz J, Jetz W, Guralnick R, Tuanmu M-N, Schildhauer M. 2015. Using multi-timescale methods and satellite-derived land surface temperature for the interpolation of daily maximum air temperature in Oregon. *Int. J. Climatol.* **35**: 3862–3878.
- Peterson TC, Vose RS. 1997. An overview of the Global Historical Climatology Network temperature database. *Bull. Am. Meteorol. Soc.* **78**: 2837–2849.
- Rohde R, Muller RA, Jacobsen R, Muller E, Perlmutter S, Rosenfeld A, Wurtele J, Groom D, Wickham C. 2013. A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinform. Geostat.: An Overview* **1**: 1–7.
- Tank AK, Wijngaard J, Engelen A van. 2002. *Climate of Europe; Assessment of observed daily temperature and precipitation extremes*. KNMI: De Bilt, The Netherlands, 1–36.
- Tetens O. 1930. Über einige meteorologische Begriffe. *Zeitschrift für Geophysik* **6**: 297–309.
- USGS. 1996. Global 30 Arc-Second Elevation (GTOPO30). United States Geological Survey.
- Van De Kerchove R, Lhermitte S, Veraverbeke S, Goossens R. 2013. Spatio-temporal variability in remotely sensed land surface temperature, and its relationship with physiographic variables in the Russian Altay Mountains. *Int. J. Appl. Earth Observation Geoinform.* **20**: 4–19.
- Weiss DJ, Atkinson PM, Bhatt S, Mappin B, Hay SI, Gething PW. 2014. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS J. Photogrammetry Remote Sens.* **98**: 106–118.
- Wilson AM, Parmentier B, Jetz W. 2014. Systematic land cover bias in Collection 5 MODIS cloud mask and derived products – A global overview. *Remote Sens. Environ.* **141**: 149–154.
- WMO. 1996. Climatological Normals (CLINO) for the period 1961–1990. World Meteorological Organization: Geneva, Switzerland.
- Yao Y, Zhang B. 2013. MODIS-based estimation of air temperature of the Tibetan Plateau. *J. Geogr. Sci.* **23**: 627–640.