New Improved Brazilian Daily Weather Gridded Data (1961-2020)

Alexandre C. Xavier[1] Bridget R. Scanlon[2] Carey W. King[3] Aline I. Alves[1]

[1]Department of Rural Engineering, Federal University of Espírito Santo, Alegre, Espírito Santo, Brazil
alexandre.xavier@ufes.br
[2]Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas
bridget.scanlon@beg.utexas.edu
[3]Energy Institute, University of Texas at Austin, Austin, Texas, USA
careyking@energy.utexas.edu

May 2, 2022

The supplement material presents the data used in the paper; results of observed and the cross-validation for variables pr, Tmax, Tmin, Rs, RH and u2; information concern the NetCDF files.

Contents

1 Information of dataset update and examples

- Information of datasets and download at:

https://sites.google.com/site/alexandrecandidoxavierufes/brazilian-daily-weather-gridded-data?authuser=0

- Examples of codes using Python language to open, visualizing and manipulate the gridded at:

https://github.com/AlexandreCandidoXavier/BR-DWGD.

2 Biomes

Figure S1 presents the biomes in Brazil [IBGE, 2021a] and the basins boundary [Nacional de Águas, 2021].

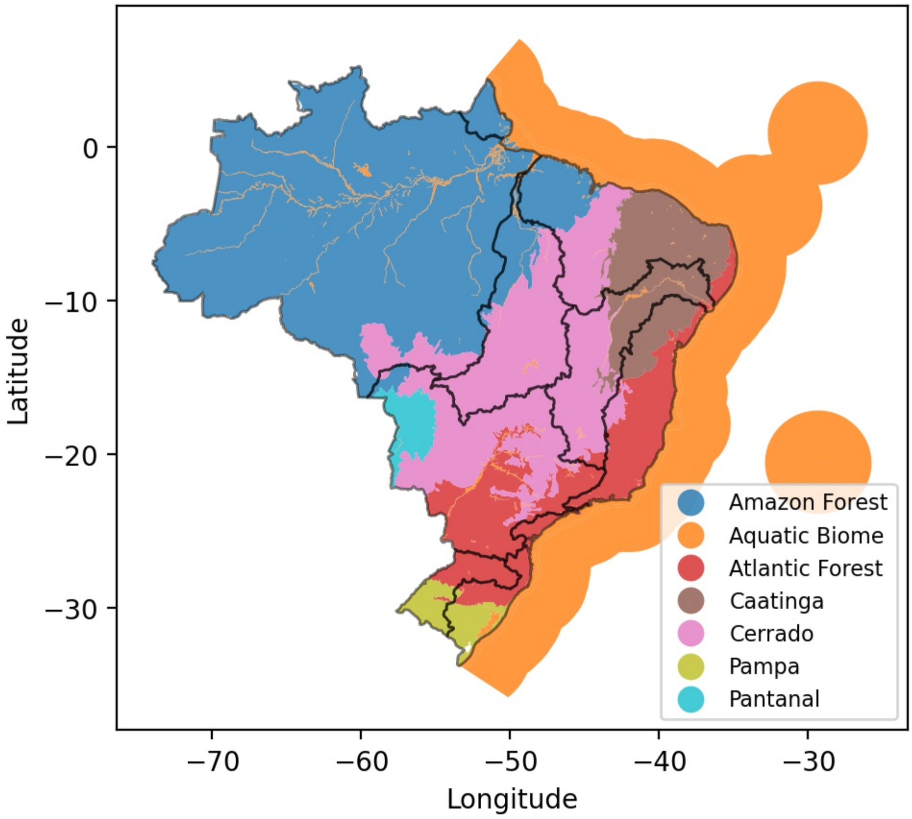

Figure S1: Biomes in Brazil and the limits of the major basins

3 Basins and theirs population, GDP and biomes

Table S1 presents the summary of population and GDP for year of 2018 [IBGE, 2021b].

Table S1: Features of the basins in relation of Gross domestic product (GDP) obtained from IBGE [2021b], where values in dollars were calculating with the ratio R$5.48/1$

Basin	Area*10^3km^2	Pop.*10^6	GDP (millions of $)						GPD per capita ($/year)	Biomes in (%) of basin area
			Total	Tax	Agriculture	Indus	Serv Pri	Serv Publ		
Amazon river	3 801.8 (44.7%)	11.5 (5.5%)	51.6 (4.0%)	5.9 (3.2%)	6.3 (11.1%)	10.2 (4.3%)	17.6 (2.9%)	11.6 (6.0%)	4 317.3	Amazon Forest: 93.2%; Aquatic Biome: 2.4%; Cerrado: 4.4%
Tocantins river	795.7 (9.4%)	5.5 (2.6%)	27.0 (2.1%)	2.1 (1.2%)	3.4 (6.0%)	7.6 (3.2%)	8.7 (1.4%)	5.1 (2.7%)	3 971.3	Amazon Forest: 20.4%; Aquatic Biome: 1.5%; Cerrado: 78.1%
North Atlantic region	1 059.6 (12.5%)	41.4 (19.9%)	130.6 (10.2%)	16.3 (9.0%)	7.3 (12.7%)	20.2 (8.5%)	56.4 (9.3%)	30.4 (15.9%)	2 162.4	Amazon Forest: 28.7%; Aquatic Biome: 2.0%; Atlantic Forest: 3.4%; Caatinga: 39.5%; Cerrado: 26.3%
Sao Francisco river	624,6 (7.3%)	15.5 (7.4%)	70.5 (5.5%)	8.2 (4.5%)	4.3 (7.6%)	14.8 (6.2%)	31.0 (5.1%)	12.1 (6.3%)	3 143.7	Aquatic Biome: 1.7%; Atlantic Forest: 3.4%; Caatinga: 37.6%; Cerrado: 57.3%
Central Atlantic region	586.3 (6.9%)	43.5 (20.9%)	255.8 (20.0%)	39.2 (21.6%)	4.2 (7.3%)	55.5 (23.3%)	115.2 (18.9%)	41.7 (21.8%)	3 191.3	Aquatic Biome: 1.5%; Atlantic Forest: 66.8%; Caatinga: 25.9%; Cerrado: 5.8%
Parana river	1 260.2 (14.8%)	71.3 (34.2%)	596.9 (46.7%)	89.2 (49.1%)	22.5 (39.4%)	99.5 (41.7%)	313.3 (51.4%)	72.4 (37.8%)	5 879.0	Amazon Forest: 2.3%; Aquatic Biome: 1.9%; Atlantic Forest: 36.0%; Cerrado: 48.0%; Pantanal: 11.7%
Uruguay river	175.9 (2.1%)	4.2 (2.0%)	28.9 (2.3%)	2.8 (1.5%)	4.9 (8.6%)	5.3 (2.2%)	11.9 (2.0%)	4.0 (2.1%)	6 804.1	Aquatic Biome: 0.4%; Atlantic Forest: 54.7%; Pampa: 45.0%
South Atlantic region	205.9 (2.4%)	15.5 (7.5%)	116.9 (9.1%)	17.9 (9.9%)	4.2 (7.4%)	25.1 (10.5%)	55.5 (9.1%)	14.1 (7.4%)	6 363.2	Aquatic Biome: 6.5%; Atlantic Forest: 53.5%; Cerrado: 0.0%; Pampa: 40.0%
Total	8 510.3	208.5	1278.1	181.7	57.2	238.2	609.8	191.2		

4 Statistics

With the observed and estimate daily data, we use some statistics to test the performance of the interpolation methods. We used the statistics and procedures described in Hofstra et al. [2008] and Xavier et al. [2016] (Equation 1-7).

$$\mathrm{R} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}\sqrt{(X_i - \bar{X})^2}\sqrt{(Y_i - \bar{Y})^2}} \tag{1}$$

$$\mathrm{bias} = \bar{Y} - \bar{X} \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{n}} \tag{3}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i| \tag{4}$$

$$\mathrm{CRE} = \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{5}$$

$$\mathrm{CSI} = \frac{a}{a+b+c} \tag{6}$$

$$\mathrm{PC} = \frac{a+d}{a+b+c+d}, \tag{7}$$

where: $\bar{X}$ and $\bar{Y}$ are the mean of X and Y, respectively, observed and estimate data; n is the number of observed data available; R is the coefficient of correlation; RMSE is the root mean square error, MAE is the mean absolute error; CRE is the compound relative error; CSI is the critical success index; and PC is percent correct. PC was calculated to evaluated the state of precipitation, wet or dry, were wet station is defined precipitation greater than 0.5 mm. CSI is tested to evaluated if the interpolated is able to forecast precipitation greater than 0.5 mm and the extreme values, those that fall above the 95th percentile (CSI high, CSIH) in the observed and estimated data (see Hofstra et al. (2008)).

5 Temporal behavior

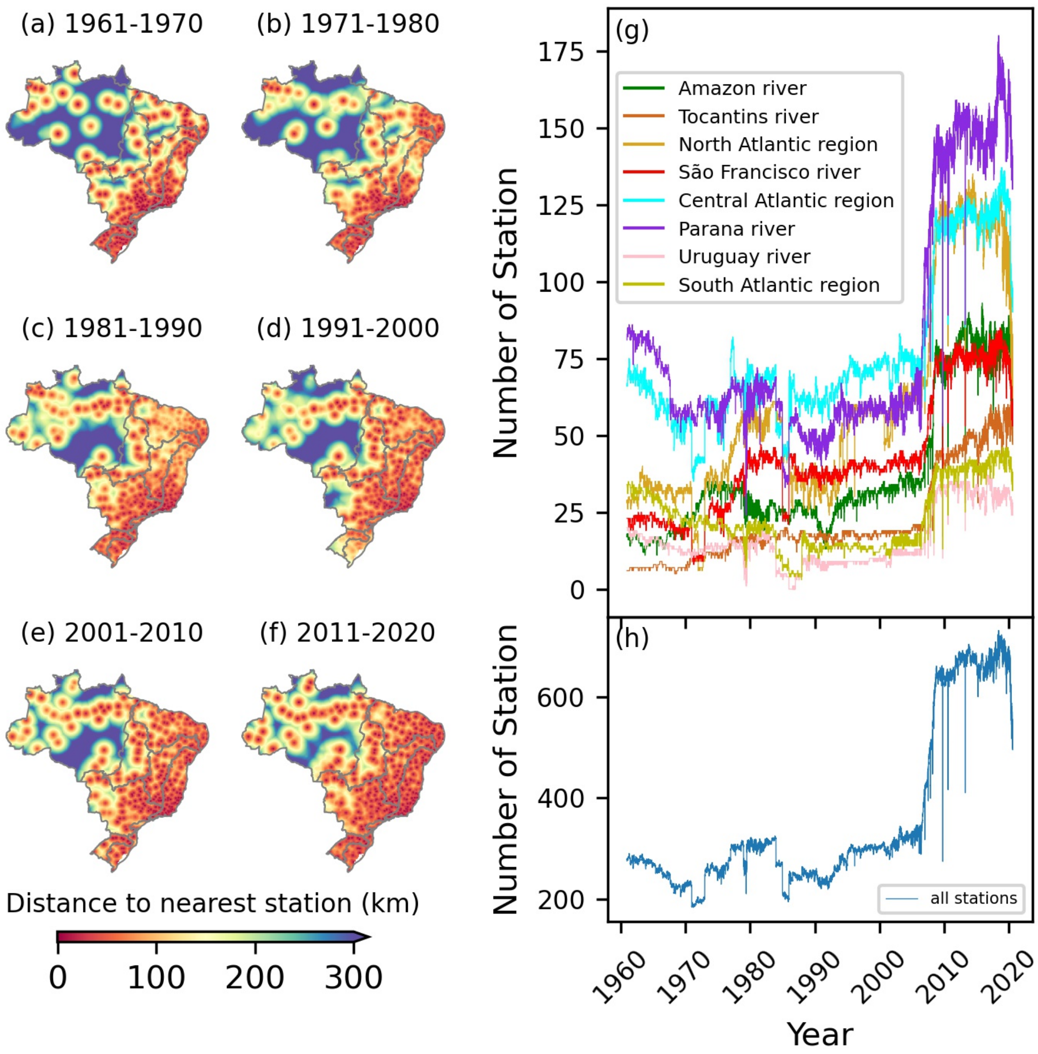

Figure S2: Average distance to the nearest weather station with available observed Tmin in the decades (a-f). Time series of total number of weather stations per basin over time (g) and the time series of total number of weather stations overall in Brazil (h).

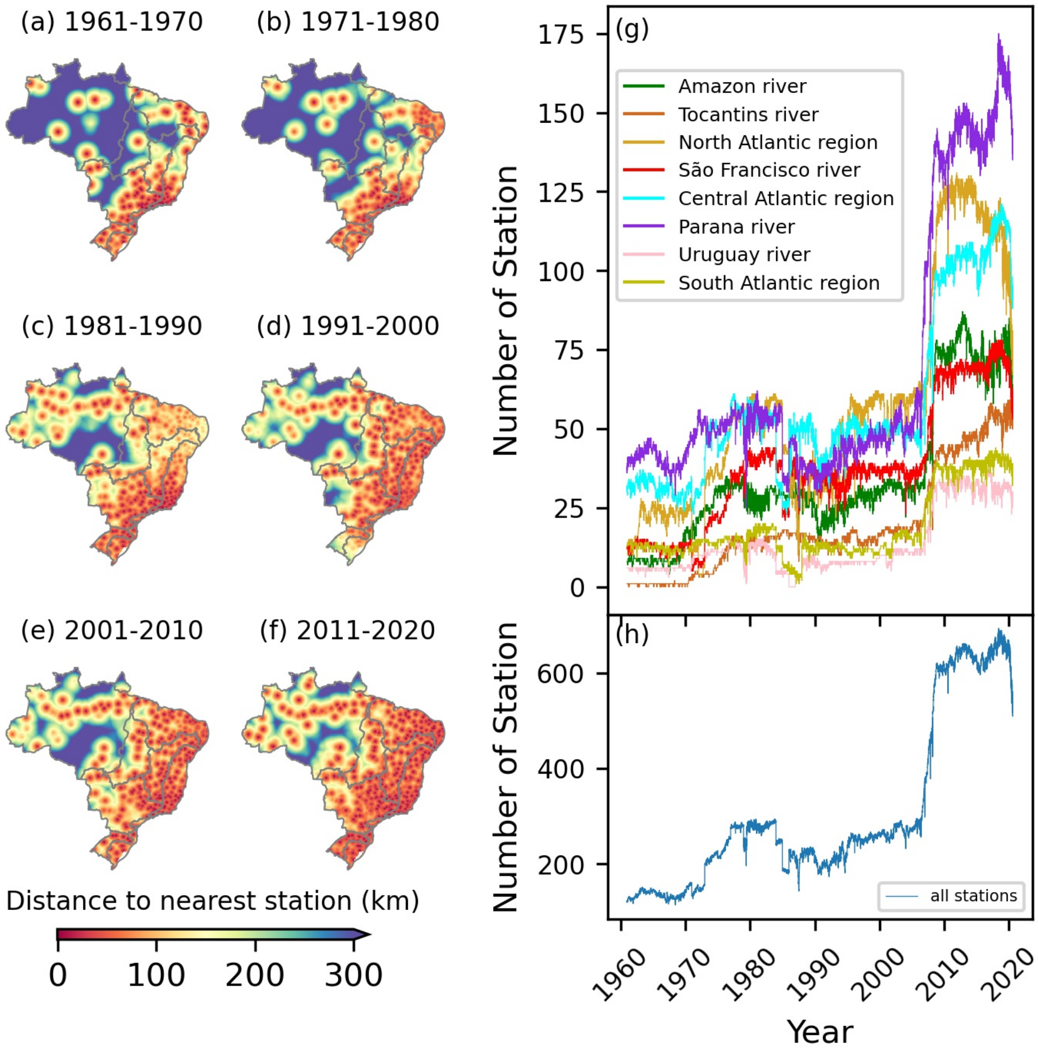

Figure S3: Average distance to the nearest weather station with available observed Rs in the decades (a-f). Time series of total number of weather stations per basin over time (g) and the time series of total number of weather stations overall in Brazil (h).

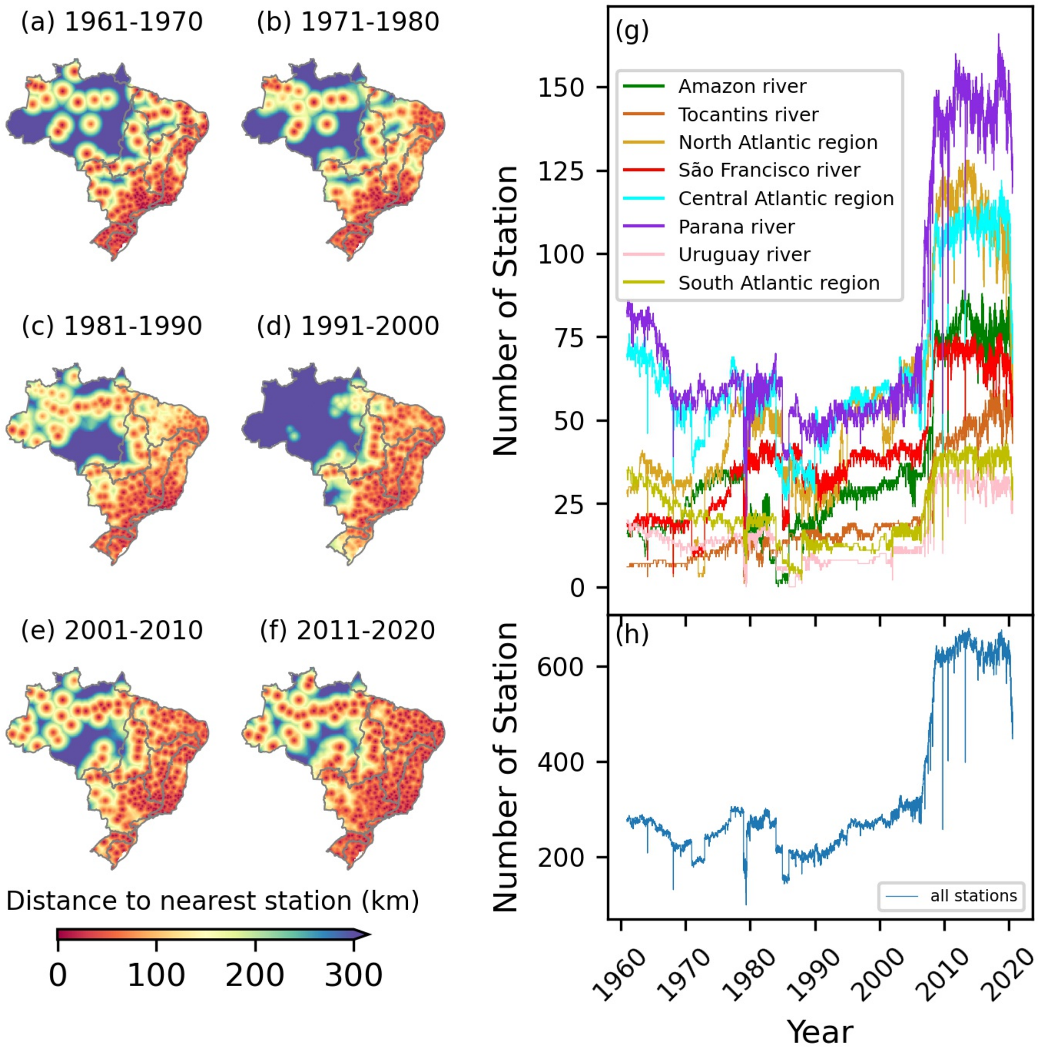

Figure S4: Average distance to the nearest weather station with available observed RH in the decades (a-f). Time series of total number of weather stations per basin over time (g) and the time series of total number of weather stations overall in Brazil (h).

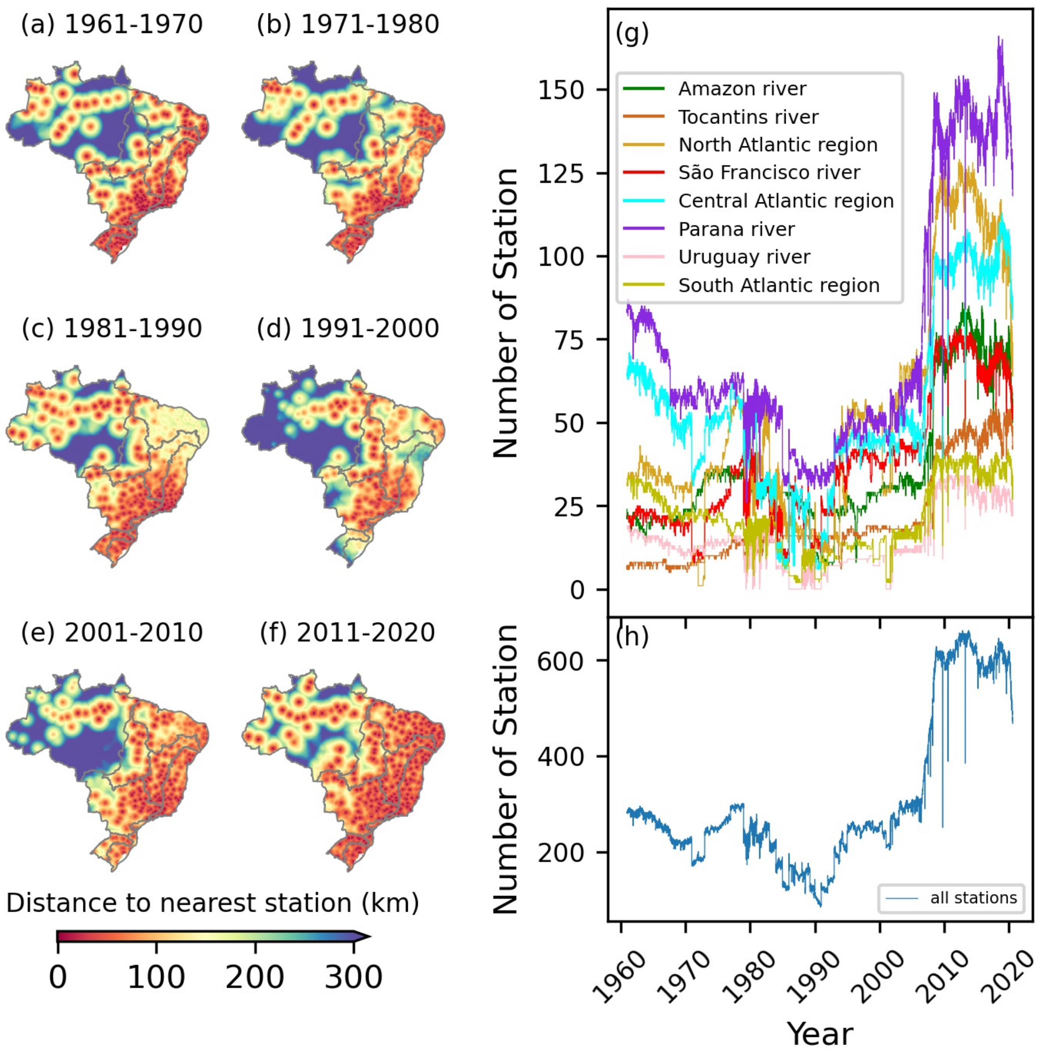

Figure S5: Average distance to the nearest weather station with available observed u2 in the decades (a-f). Time series of total number of weather stations per basin over time (g) and the time series of total number of weather stations overall in Brazil (h).

6 Cross-validation all data

Table S2: Rs cross-validation statistics and their respective skill score for each model.

Method	AV rank	R	#	Bias	#	RMSE	#	CRE	#	MAE	#	CSIL	#	CSIH	#
ADW	1.143	0.873	1	0.015	2	2.821	1	0.238	1	2.028	1	0.516	1	0.440	1
IDW	1.857	0.866	2	0.002	1	2.903	2	0.252	2	2.066	2	0.501	2	0.430	2

Table S3: RH cross-validation statistics and their respective skill score for each model.

Method	AV rank	R	#	Bias	#	RMSE	#	CRE	#	MAE	#	CSIL	#	CSIH	#
ADW	1.000	0.858	1	-0.075	1	6.971	1	0.264	1	5.292	1	0.486	1	0.274	1
IDW	2.000	0.852	2	-0.111	2	7.152	2	0.278	2	5.394	2	0.471	2	0.267	2

Table S4: u2 cross-validation statistics and their respective skill score for each model.

Method	AV rank	R	#	Bias	#	RMSE	#	CRE	#	MAE	#	CSIL	#	CSIH	#
ADW	1.000	0.534	1	0.003	1	0.897	1	0.751	1	0.661	1	0.081	1	0.192	1
IDW	2.000	0.518	2	-0.019	2	0.921	2	0.792	2	0.677	2	0.068	2	0.181	2

7 Cross-validation basin scale

Table S5: Tmin statistics results in each river basin, using IDW_{ElAdj} interpolation.

Basin	R	Bias	RMSE	CRE	MAE	CSIL	CSIH
Amazon river	0.779	0.005	1.417	0.409	1.029	0.526	0.223
Tocantins river	0.826	-0.115	1.635	0.339	1.135	0.390	0.312
North Atlantic region	0.817	0.010	1.338	0.345	0.968	0.444	0.265
São Francisco river	0.884	0.046	1.717	0.222	1.193	0.378	0.476
Central Atlantic region	0.895	0.041	1.639	0.202	1.177	0.483	0.354
Parana river	0.921	0.021	1.777	0.154	1.237	0.561	0.478
Uruguay river	0.949	-0.058	1.669	0.102	1.173	0.610	0.521
South Atlantic region	0.948	0.246	1.654	0.107	1.153	0.610	0.480

Table S6: Rs statistics results in each river basin, using ADW interpolation.

Basin	R	Bias	RMSE	CRE	MAE	CSIL	CSIH
Amazon river	0.689	0.081	3.544	0.545	2.676	0.229	0.244
Tocantins river	0.798	0.031	2.752	0.366	2.002	0.346	0.414
North Atlantic region	0.824	0.057	2.754	0.322	1.998	0.391	0.307
São Francisco river	0.879	-0.197	2.516	0.229	1.790	0.406	0.383
Central Atlantic region	0.880	0.079	2.795	0.226	2.044	0.443	0.373
Parana river	0.889	-0.029	2.651	0.210	1.872	0.534	0.450
Uruguay river	0.940	-0.094	2.602	0.117	1.835	0.476	0.445
South Atlantic region	0.920	0.281	2.881	0.154	2.021	0.455	0.496

Table S7: RH statistics results in each river basin, using ADW interpolation.

Basin	R	Bias	RMSE	CRE	MAE	CSIL	CSIH
Amazon river	0.762	-0.096	6.283	0.424	4.690	0.482	0.128
Tocantins river	0.901	-0.113	6.590	0.188	5.027	0.468	0.262
North Atlantic region	0.857	-0.323	7.468	0.267	5.664	0.460	0.245
São Francisco river	0.854	0.911	7.653	0.275	5.949	0.397	0.355
Central Atlantic region	0.756	-0.477	6.915	0.438	5.333	0.361	0.260
Parana river	0.869	0.124	6.735	0.246	5.067	0.491	0.342
Uruguay river	0.843	0.771	6.866	0.293	5.078	0.340	0.350
South Atlantic region	0.774	-1.115	6.931	0.420	5.241	0.342	0.277

Table S8: u2 statistics results in each river basin, using ADW interpolation.

Amazon river	0.287	0.028	0.763	1.140	0.561	0.000	0.101
Tocantins river	0.259	-0.008	0.762	1.065	0.575	0.000	0.105
North Atlantic region	0.623	0.013	0.894	0.625	0.657	0.113	0.146
São Francisco river	0.550	0.019	0.836	0.731	0.625	0.042	0.171
Central Atlantic region	0.442	-0.012	0.932	0.856	0.673	0.000	0.140
Parana river	0.406	0.010	0.842	0.904	0.643	0.059	0.156
Uruguay river	0.486	0.043	0.952	0.823	0.729	0.089	0.159
South Atlantic region	0.444	0.017	1.177	0.845	0.864	0.064	0.155

8 Plots of daily cross-validation statistics

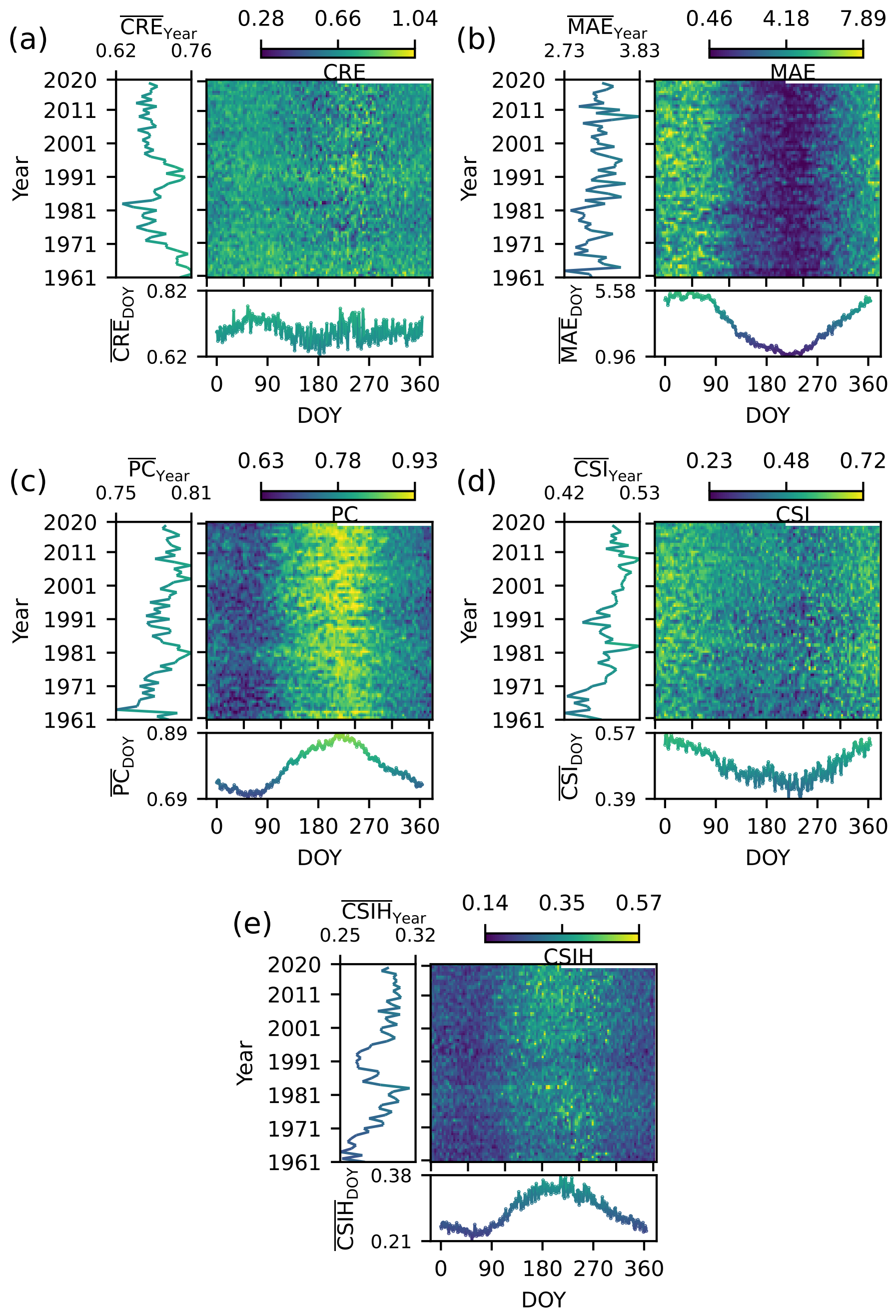

Figure S6: (a–e) precipitation cross-validation for entire daily data set for CRE, MAE, PC, CSI and CSIH with average statistics across all days of a given year (left hand side plot) and average statistics across all years for a given day of the year, DOY (bottom side plot).

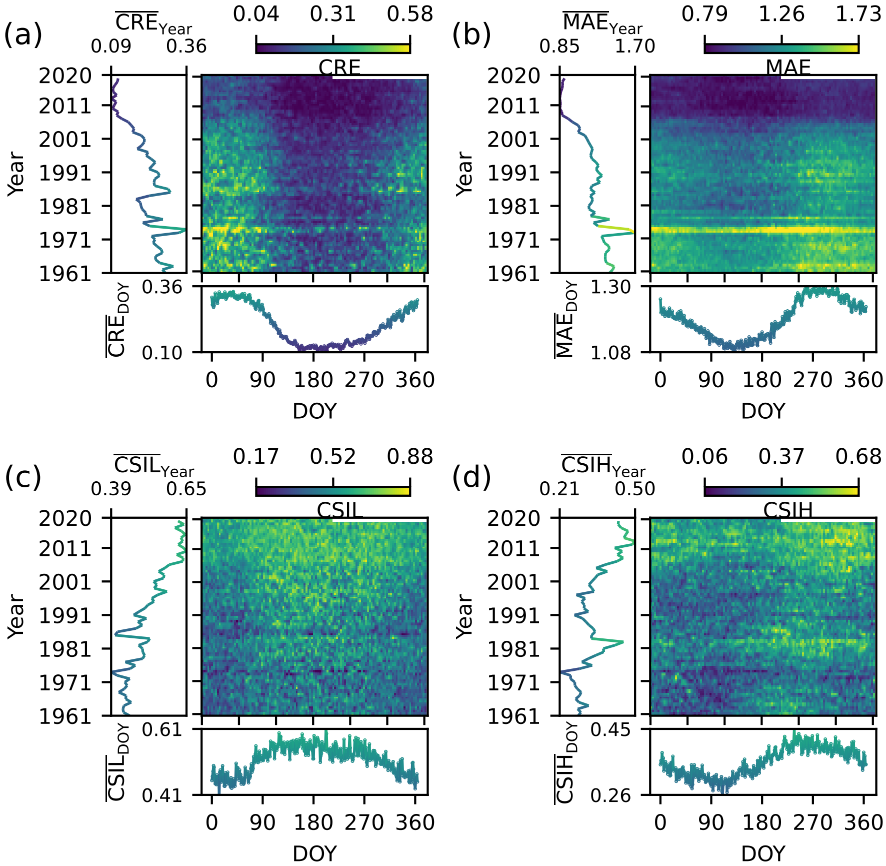

Figure S7: (a–d) Tmax cross-validation for entire daily data set for CRE, MAE, CSIL and CSIH with average statistics across all days of a given year (left hand side plot) and average statistics across all years for a given day of the year, DOY (bottom side plot).

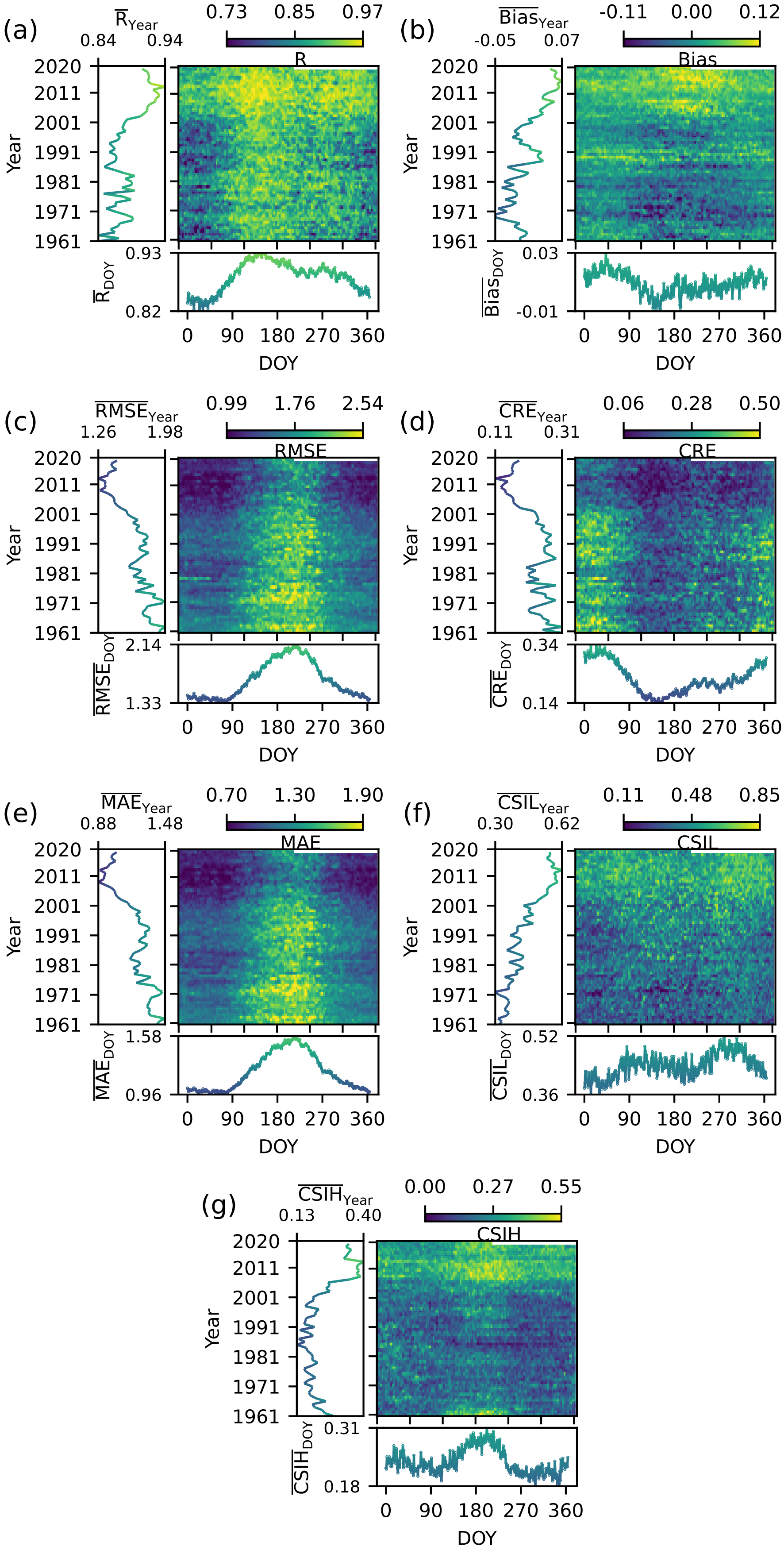

Figure S8: (a–g) Tmin cross-validation for entire daily data set for R, Bias, RMSE, CRE, MAE, CSIL and CSIH with average statistics across all days of a given year (left hand side plot) and average statistics across all years for a given day of the year, DOY (bottom side plot).

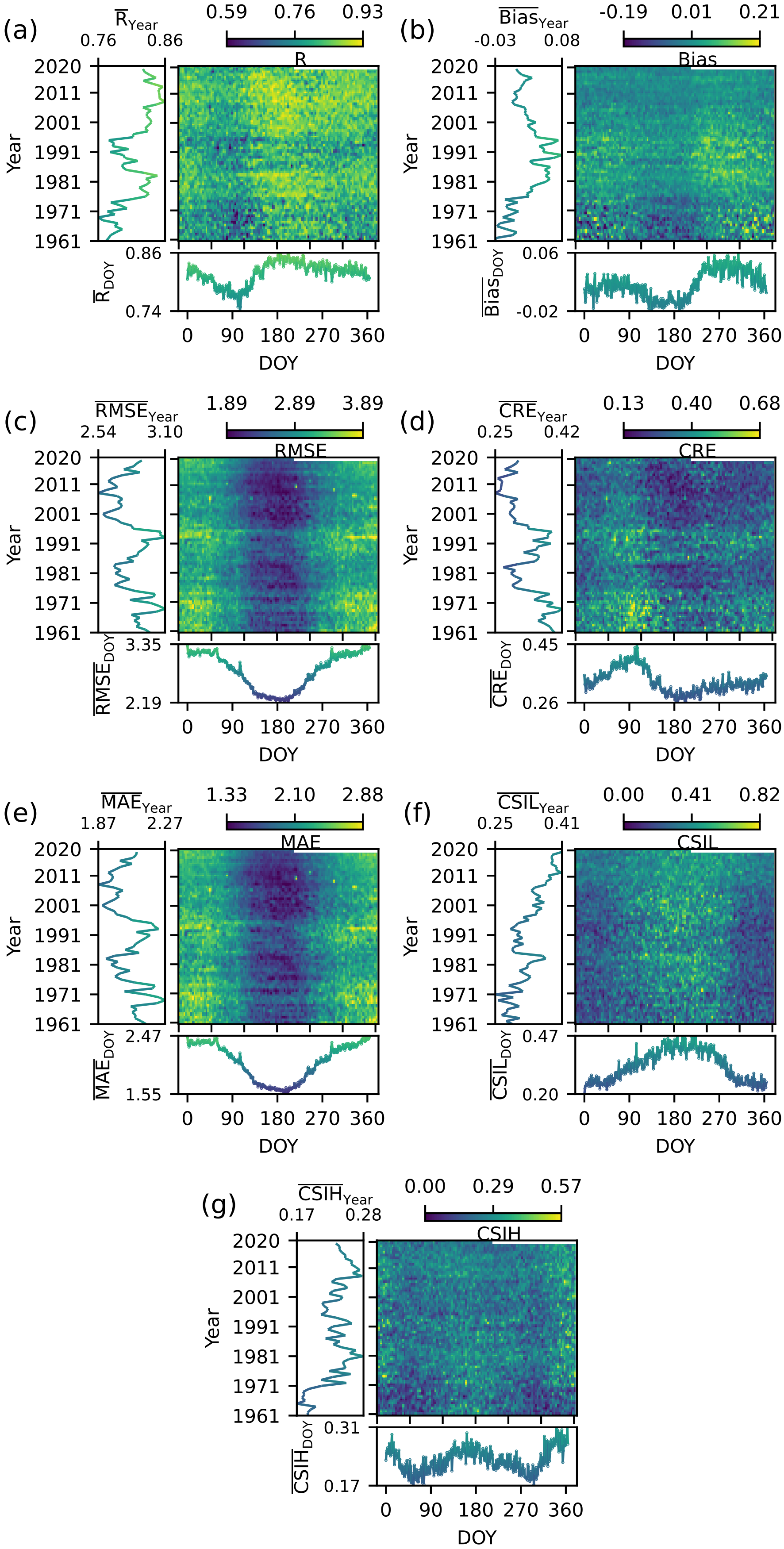

Figure S9: (a–g) Rs cross-validation for entire daily data set for R, Bias, RMSE, CRE, MAE, CSIL and CSIH with average statistics across all days of a given year (left hand side plot) and average statistics across all years for a given day of the year, DOY (bottom side plot).

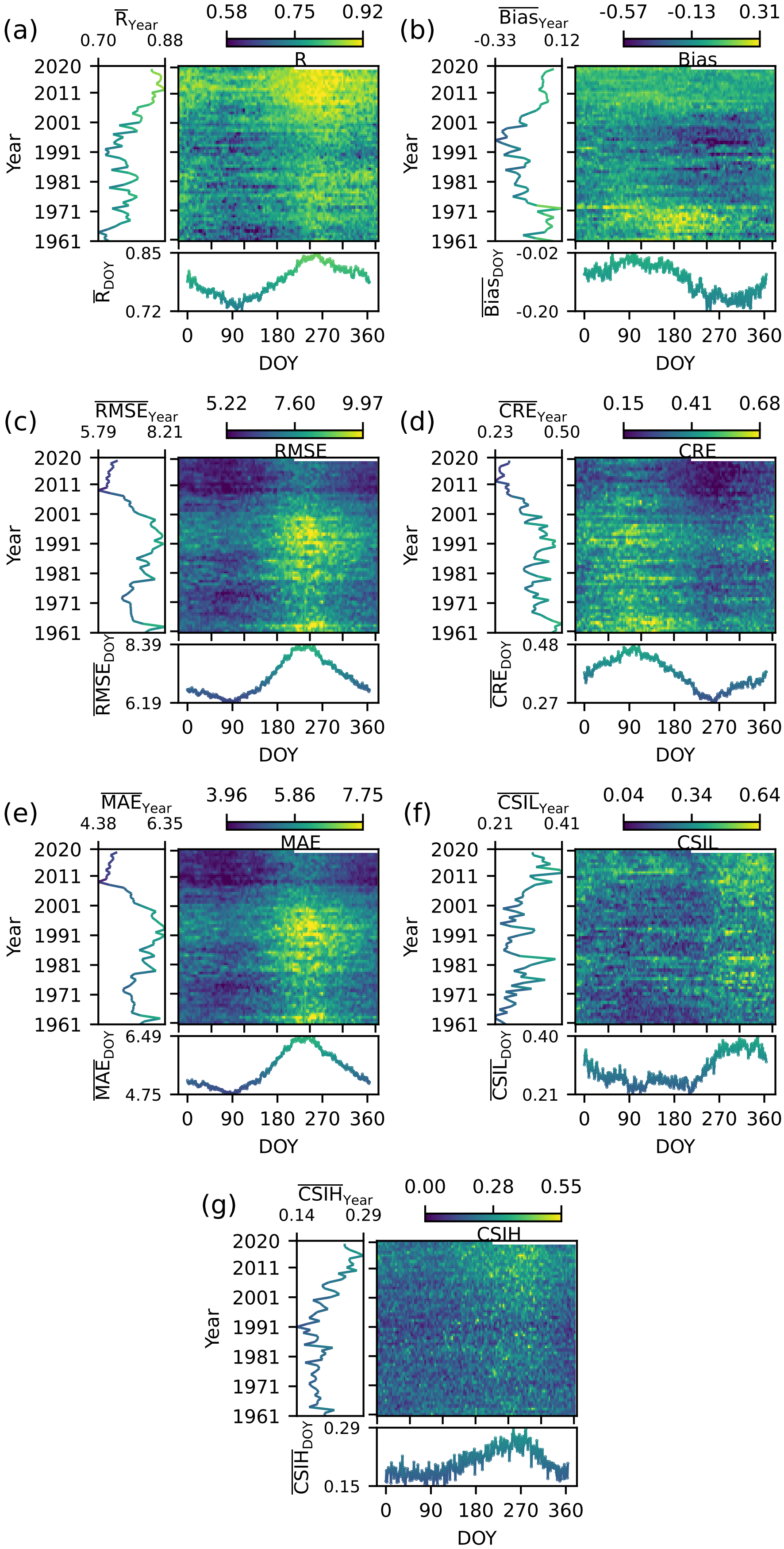

Figure S10: (a–g) RH cross-validation for entire daily data set for R, Bias, RMSE, CRE, MAE, CSIL and CSIH with average statistics across all days of a given year (left hand side plot) and average statistics across all years for a given day of the year, DOY (bottom side plot).

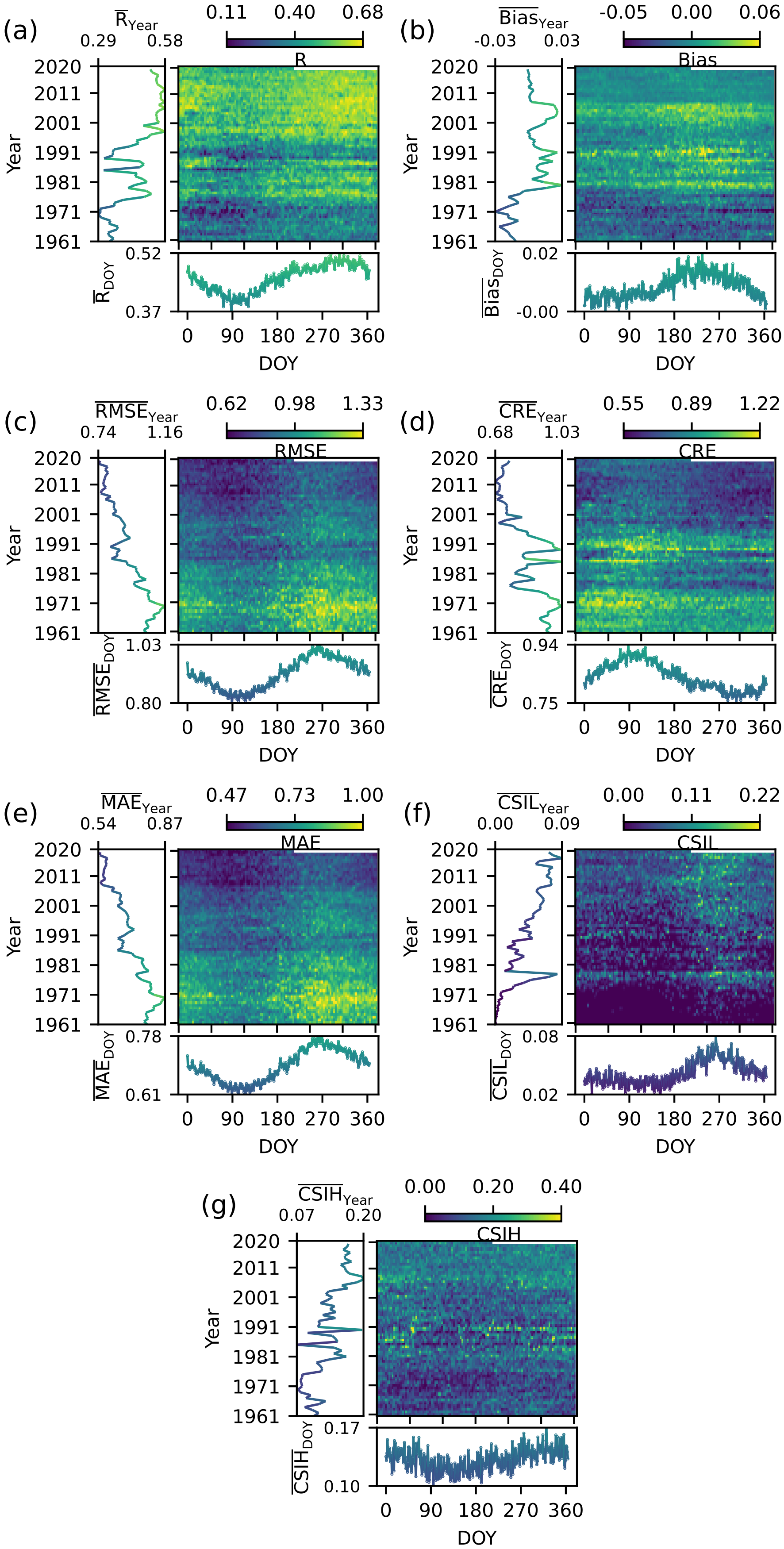

Figure S11: (a–g) u2 cross-validation for entire daily data set for R, Bias, RMSE, CRE, MAE, CSIL and CSIH with average statistics across all days of a given year (left hand side plot) and average statistics across all years for a given day of the year, DOY (bottom side plot).

9 Workflow stages

Observed data

Variables:

- precipitation (pr)
- minimum temperature (Tmin)
- maximum temperature (Tmax)
- solar radiation (Rs)
- wind speed at 10 m height (u2=0.75*u10, [Allen et al., 1998])
- relative humidity (RH)

Sources:

- Agência Nacional de Águas (ANA, here) (pr)
- Instituto Nacional de Meteorologia (INMET, here and here)

Organizing data

For each variable, it was create unique file with data of all station, an array with shape ($n \times m$), where n is the number of days and m is the number of stations. Latitude, longitude and altitude where save too. Automatic stations, with hourly observed data, had them resample to 24 hour, using the rules:

- Tmax is the maximum values of the hourly maximum temperature in the period of 0UTC⊣24UTC;
- Tmin is the minimum values of hourly minimum temperature in the period of 0UTC⊣24UTC;
- RH is the average of hourly RH in the period of 0UTC⊣24UTC;
- u2 is the average of hourly u2 in the period of 0UTC⊣24UTC;
- Rs is the sum of hourly Rs in the period of 0UTC⊣24UTC;
- pr is the sum of hourly pr in the period of 10UTC (previous day)⊣10UTC.

Data quality and homogeneity

1. Accepted data in accordance with:
   - pr (mm), $0 \leq$ pr $<$ 450 [Liebmann and Allured, 2005];
   - Tmax, Tmin ($^{\circ}$C), $-30 \leq$ Tmax, Tmin $<$ 50 [Shafer et al., 2000];
   - RH (%), $0 \leq$ RH$<$ 100;
   - Rs (MJ m^{-2}), $0.03Ra \leq$ Rs$<$Ra [Moradi, 2009];
   - u2 (m s^{-1}), $0 \leq$ u$<$ 100 [Shafer et al., 2000].
2. Visual homogeneity according to Xavier et al. [2016].

Cross-validation

- Procedure:
  1. select the interpolation method (IDW or ADW);
  2. choose one station/day that had observed data;
  3. with the data of five surrounding (nearest) stations on that station/day, we applied the interpolation at the position of the station selected in (1);
  4. save the interpolated value;
  5. repeat steps 1-4 for all station and days with observed data
- For Tmax and Tmin it was also used the observed values correct to the same altitude of the query point with the aid of the temperature lapse rate (0.0060°C m^{-1}).

Best interpolation

With observed and estimated daily data, we calculated the statistics:

- correlation coefficient (R);
- root mean square error (RMSE);
- mean absolute error (MAE);
- Bias;
- compound relative error (CRE);
- critical success index (CSI), CSI low (CSIL) and CSI high (CSIH);
- percent correct (PC).

For each interpolation method, their Statistics were ranking (1 = best, 2 = second best, etc.). We selected the interpolation method with lowest average rank to generated the gridded data.

Gridded data

The daily gridded data have spatial resolution of $0.1^{\circ} \times 0.1^{\circ}$, from 01/01/1961-07/31/2020. The interpolation was done according to:

1. for each variable to be interpolated (6 variables);
2. selected the best interpolation method for the variable;
3. for each cell in all cells inside Brazil;
4. for each day in all days to be interpolated (01/01/1961-07/31/2020);
5. selected the five nearest observed values to the cell;
   - if var is Tmax or Tmin, adjusted the five observed values to the altitude of the cell using the value of the temperature lapse rate;
   - applied the interpolation;
   - save the values of interpolation, the distance to the nearest station (dist_nearest) and the number of stations inside the cell (count).
6. repeat (1-5) for all cells and days;

Gridded for community

We generated NetCDF (Network Common Data Form) files for the variables (Tmax, Tmin, pr, RH, Rs, u2, dist_nearest and count):

- Data variables: variable (time, latitude, longitude)
- Coordinates: time, latitude, longitude;
- Dimensions: (time: 21762, latitude: 393, longitude: 391);
- Attributes: institution, institution_id, etc.

Main libraries:

- Pandas [The Pandas development team, 2020]
- Numpy [Harris et al., 2020]
- Joblib [Joblib Development Team, 2020]
- Request [Team, 2020]
- xarray [Hoyer and Hamman, 2017]
- numba [Lam et al., 2015]
- Matplotlib [Hunter, 2007]
- Cartopy [Met Office, 2010 - 2015]

Figure S12: General workflow stages to generated the gridded datasets.

10 Reference evapotranspiration gridded

The daily grid of the FAO Penman-Montheith evapotranspiration (ETo, Allen et al. [1998]) was calculated by means of Equation 8. The variables used to calculate ETo were those from BR-DWGD: $Tmax$, $Tmin$, RH, u_2 and, Rs.

$$ETo = \frac{0.408\Delta(Rn - G) + \gamma \dfrac{900}{T + 273} u_2(\mathrm{e}_s - \mathrm{e}_a)}{\Delta + \gamma(1 + 0.34u_2)}, \tag{8}$$

where ETo is reference evapotranspiration (mm day^{-1}), Rn is net radiation (MJ m^{-2} day^{-1}), G is the soil heat flux density (MJ m^{-2} day^{-1}), T is the air temperature at 2 m height (°C), u_2 is the wind speed at 2 m height (m s^{-1}), e_s is the saturation vapour pressure (kPa), e_a is the actual vapour pressure (kPa), $\mathrm{e}_s - \mathrm{e}_a$ is the saturation vapour pressure deficit (kPa), Δ is the slope vapour pressure curve (kPa°C^{-1}), and γ is the psychrometric constant (kPa °C^{-1}).

In our case, G was considered zero; T was the average of $Tmax$ and $Tmin$; e_a was estimated with the aid of RH, $Tmax$ and $Tmin$:

$$e_a = \frac{RH}{100}\left(\frac{e^\circ(Tmax) + e^\circ(Tmin)}{2}\right), \tag{9}$$

where $e^\circ(Tmax)$ and $e^\circ(Tmin)$ are respectively saturation vapour pressure at the air maximum and minimum temperature. Rn is calculated as:

$$Rn = Rns - Rnl \tag{10}$$

where Rns is the incoming net shortwave radiation and Rnl is the outgoing net longwave radiation. Rnl is $f(\mathrm{tasmax}, \mathrm{tasmin}, \mathrm{e}_a$, Rs, $Rso)$, where Rso is the clear-sky radiation (MJ m^{-2} day^{-1}) that is calculated.

11 NetCDF File

This is a metafile description of gridded data sets generated for the paper.

Download location, check here or URL:

`https://sites.google.com/site/alexandrecandidoxavierufes/brazilian-daily-weather-gridded-data?authuser=0`

The name of the files are formatted as:

VARNAME_STARTEDDATE_ENDDATE_GRIDABBREVIATION_UNIVERSITY1_UNIVERSITY2_VERSION.nc

Where:

VARNAME, the name of the variable (ETo, Tmax, Tmin, Rs, u2, RH and pr);
STARTEDDATE: data file started on yyyymmdd;
ENDDATE: data file finished on yyyymmdd;
GRIDABBREVIATION: BR-DWGD (Brazilian Daily Weather Gridded Data);
ABBREVIATION1: UFES (Federal University of Espírito Santo);
ABBREVIATION2: UT (University of Texas at Austin);

Example filename: RH_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc

Table S9: Files available and size.

File_name	Size GB
ETo_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
ETo_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
ETo_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.0
RH_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
RH_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
RH_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.0
RH_Control_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
RH_Control_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
RH_Control_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.0
Rs_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
Rs_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
Rs_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.0
Rs_Control_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
Rs_Control_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
Rs_Control_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.0
Tmax_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
Tmax_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
Tmax_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.0
Tmax_Control_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
Tmax_Control_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
Tmax_Control_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.0
Tmin_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
Tmin_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
Tmin_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.0
Tmin_Control_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
Tmin_Control_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
Tmin_Control_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.0
pr_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
pr_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
pr_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.0
pr_Control_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
pr_Control_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
pr_Control_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.0
u2_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
u2_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.1
u2_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	4.0
u2_Control_19610101_19801231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
u2_Control_19810101_20001231_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.1
u2_Control_20010101_20200731_BR-DWGD_UFES_UTEXAS_v_3.0.nc	3.0

References

IBGE. Instituto Brasileiro de Geografia e Estatistica - catálagos de Metadados, 2021a. URL http://www.metadados.geo.ibge.gov.br/geonetwork_ibge/srv/por/metadata.show?id=19581&currTab=simple. Accessed = 2021-05-26.

Agência Nacional de Águas. Catálago de metadados da ana, 2021. URL https://metadados.snirh.gov.br/geonetwork/srv/api/records/43539328-3a83-4bf2-9cea-2b47513f4b07. Accessed = 2021-05-26.

IBGE. Instituto Brasileiro de Geografia e Estatistica - sistema ibge de recuperação automática, 2021b. URL https://sidra.ibge.gov.br/tabela/5938. Accessed = 2021-05-26.

Nynke Hofstra, Malcolm Haylock, Mark New, Phil Jones, and Christoph Frei. Comparison of six methods for the interpolation of daily, european climate data. *Journal of Geophysical Research: Atmospheres*, 113(D21):n/a–n/a, 2008. ISSN 2156-2202. doi: 10.1029/2008JD010100. URL http://dx.doi.org/10.1029/2008JD010100.

Alexandre C. Xavier, Carey W. King, and Bridget R. Scanlon. Daily gridded meteorological variables in brazil (1980–2013). *International Journal of Climatology*, 36(6):2644–2659, 2016. doi: https://doi.org/10.1002/joc.4518. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.4518.

Brant Liebmann and Dave Allured. Daily precipitation grids for south america. *Bulletin of the American Meteorological Society*, 86:1567–1570, 2005. doi: http://dx.doi.org/10.1175/BAMS-86-11-1567.

M.A. Shafer, C.A . Fiebrich, D.S. Arndt, S.E . Fredrickson, and T.W. Hughes. Quality assurance procedures in the oklahoma mesonet. *Journal of Atmospheric and Oceanic Technology*, 17:474–494, 2000.

Isaac Moradi. Quality control of global solar radiation using sunshine duration hours. *Energy*, 34(1):1 – 6, 2009. ISSN 0360-5442. doi: http://dx.doi.org/10.1016/j.energy.2008.09.006. URL http://www.sciencedirect.com/science/article/pii/S0360544208002466.

The Pandas development team. pandas-dev/pandas: Pandas, February 2020. URL https://doi.org/10.5281/zenodo.3509134.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

Joblib Development Team. Joblib: running Python functions as pipeline jobs, 2020. URL https://joblib.readthedocs.io/.

Request Development Team. Requests: HTTP for humans^{tm}, 2020. URL https://docs.python-requests.org/en/latest/.

S. Hoyer and J. Hamman. xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5 (1), 2017. doi: 10.5334/jors.148. URL https://doi.org/10.5334/jors.148.

Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. Numba: A LLVM-based python JIT compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, pages 1–6, 2015.

J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science Engineering*, 9(3):90–95, May 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.55.

Met Office. *Cartopy: a cartographic python library with a matplotlib interface.* Exeter, Devon, 2010 - 2015. URL http://scitools.org.uk/cartopy.

R. G. Allen, D. R. Pereira, D. Raes, and M. Smith. *Crop evapotranspiration: guidelines for computing crop water requirements.* United Nations, FAO, Rome, Italy, 1 edition, 1998. URL http://www.kimberly.uidaho.edu/ref-et/fao56.pdf.