

Finescaled Weather Data Set for the US 1950-2005

Wolfram Schlenker

Michael J. Roberts

Columbia University and NBER North Carolina State University

1 Data Description

1.1 Data Construction

Our data is based on the same rectangular grid system underlying PRISM that covers the contiguous United States.¹ It consists of 1405 grids in the longitude direction and 621 grids in the latitude dimension, spaced in equidistance 1/24 degree steps (rough 2.5miles or 4km). We matched the centroid of each grid point to the fips codes of all counties in the United States, using ESRI's "detailed county" shape files.

Table 1 summarizes the number grid points in our data as well as PRISM and how they relate. Figure 1 displays them spatially. There are 471159 grid points with non-missing values in the PRISM data where the centroid could be matched to lie within a county. They are displayed in various shades of grey in Figure 1. Another 390,864 grid points had missing values in PRISM and the centroid could not be matched to a fips code. These grids are displayed in blue. There are 10,480 grid points with non-missing values in PRISM whose centroid could not be matched to a county. These are primarily around the border of the United States as well as over water bodies like the Chesapeake or San Francisco Bay. They are displayed in red in Figure 1. Finally, there are 2 grids that had missing values in PRISM but whose centroids could be matched to a county. They are displayed in green and can be found at the shoreline of the Great Lake region.

Grids are identified by the variable `gridNumber`, which ranges from 1 to 872505. The grids are ordered from the north-western most grid to the south-eastern most grid, moving first along the longitude dimension and second along the latitude direction. In other words, the first 1405 grids range from the north-west to the north-east.

1.2 Data Distribution

The files `stateXX_yearYYYY.csv` include the minimum and maximum temperature (Celsius) as well the total precipitation (cm) for each day of a year `YYYY` for all of the 471159 grids that fall within the state with state fips code `XX`. A list of states fips codes

¹For the original PRISM data see <http://www.prism.oregonstate.edu/>

is given in Table 2. A meta-file gridInfo.csv provides the link between gridNumber and fips codes as well as the cropland area within the grid cell. The longitude and latitude of each grid centroid is given by the STATA code:

```
gen longitude = -125 + mod(gridNumber-1,1405)/24
gen latitude = 49.9375+1/48 - ceil(gridNumber/1405)/24
```

There are 8 DVDs with zipped stateXX_yearYYYY.csv files by state. Figure 3 displays which states are on which disc, and a list of the states that are contained on each DVD (including the state fips code in brackets) is given in Table 3. Disc 8 also includes two STATA files: one the gives degree days (averaged over the cropland area within a county) by county and month using the most common degree days boundaries. For other degree days bounds or degree days that are averaged over different areas, see the next section. The second file gives monthly PRISM averages of minimum and maximum temperatures, averaged over the cropland area in a county. These files use the monthly data from PRISM and do not go through the interpolation procedure to first get daily data, which is required to calculate degree days. Hence, *for an analysis that uses monthly minimum, average, or maximum temperature, or total precipitation, it is preferable to use the PRISM data directly, as our procedure would first relate daily weather station data to these monthly PRISM averages to obtain interpolate daily values and then aggregate them back up into monthly totals. This procedure induces unnecessary noise in the monthly numbers.*

2 Customized Derivation of Degree Days

As outlines in the previous section, disc 8 includes two files with degree days variables as well as monthly averages of minimum and maximum temperature by county and month, where all grids within a county are weighted by the cropland area they contain.

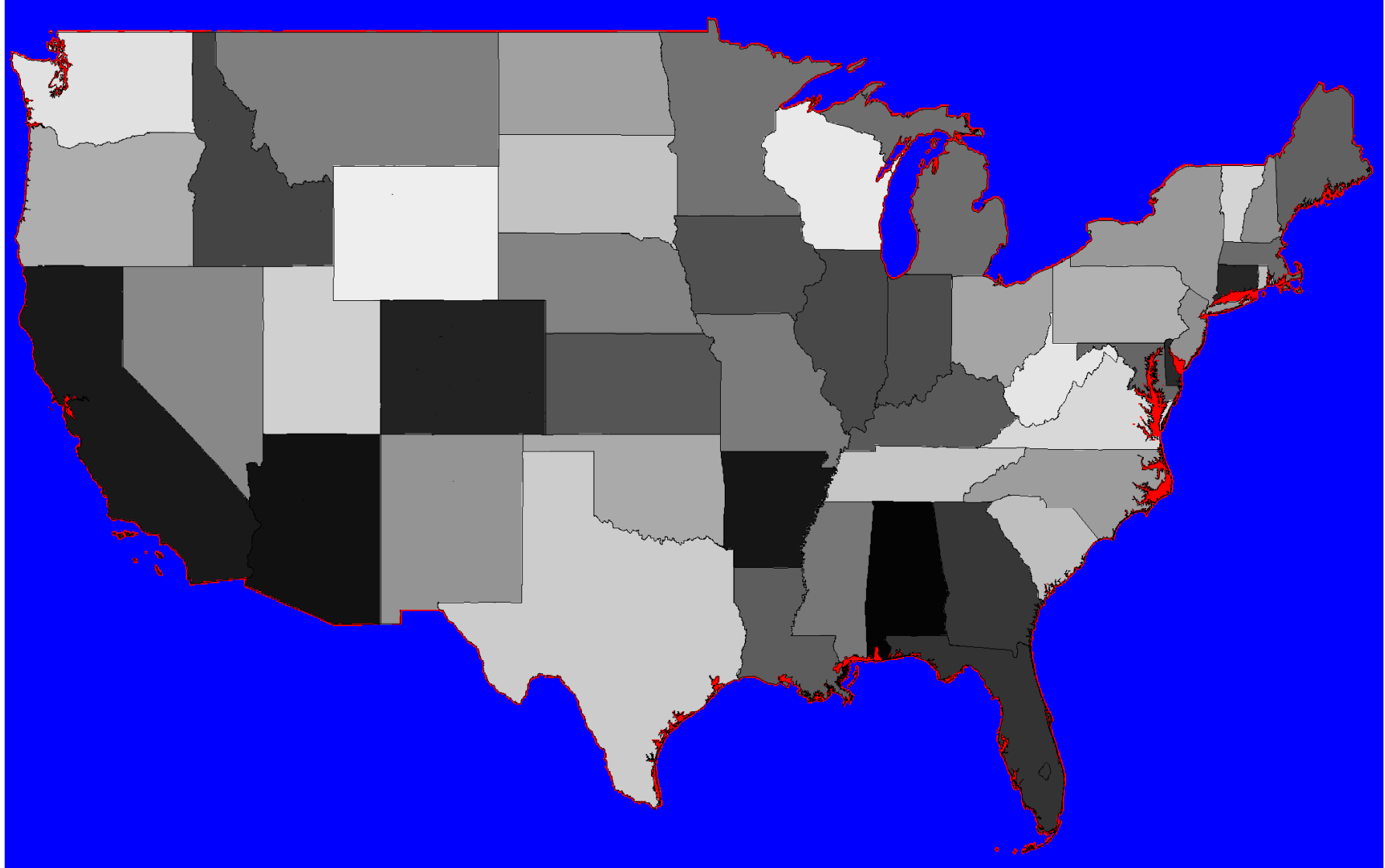
If you would like to (i) use different weights, (ii) aggregate over different spatial units (e.g., zip codes), the STATA-code ddayDays.do allows for the construction of such customized variables.

The file requires several inputs:

- Part 1: The first file requires you to specify the location of three directories: (i) the data directory contains the unzipped csv-files with the daily data that are contained on the DVDs as well as the meta-file gridInfo.csv. The latter has information about the the cropland area of each grid cell and the county in which it's centroid is located. (ii) The temporary directory specifies where temporary regression files are stored (they are automatically deleted once the script is completed). (iii) The output directory specifies where the newly created STATA file will be saved.

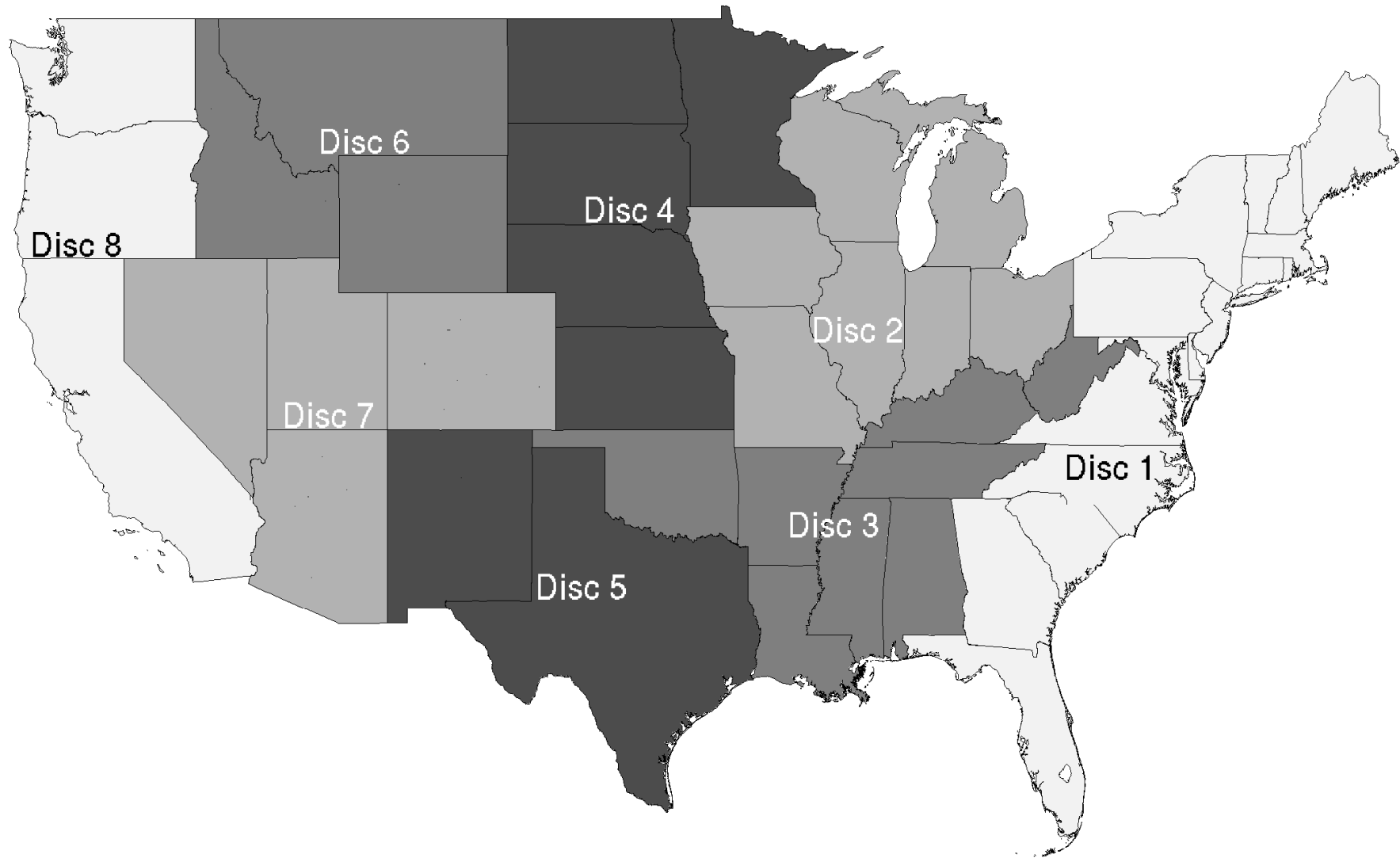
- Part 2: Specify the list of degree day bounds in degree Celsius
- Part 3: Specify the range of years for which the data should be derived (admissible range is 1950-2005).
- Part 4: Define the day-range within each year that should be considered. Days outside that range will be omitted in the final variable. For example, if you set the start date to April 15th and then choose monthly aggregates (see part 7 below), the output file will only include April 16-April 30 in the total for April).
- Part 5: Specify the list of states include in the analysis. To include a state, simply remove the star at the beginning of a line. To drop it again, add a star at the beginning of a line.
- Part 6: Define the weighting variable and aggregation variable. The aggregation variable is the variable within the file gridInfo.csv that specifies which grids should be aggregated together. For example, if you choose “fips,” all grids within a county will be aggregated. This setup allows for an arbitrary change in the spatial units. For example, if you want to aggregate the data by zip code, simply add a new variable called “zipCode” to gridInfo.csv that links each grid centroid to a zip code. The weighting variable gives relative weights assigned to each grid cell within the same aggregation variable (e.g., county). Only relative weights matter, i.e., multiply all weights by two will give identical results. If you want the average over the netire county area (not weighted by cropland), simply add a variable “constant” to the file gridInfo.csv that is equal to one for all observations and choose it as the weighting variable.
- Part 7: Choose whether you want the output by month or by year. If you choose monthly output, you will get one observation for all days within each month of the growing season defined in part 4 above. Yearly output gives one observation per year.

Figure 1: Grids in Data Set



Notes: Link between grid cells in this data and original PRISM data. Grids in our data that have non-missing PRISM values are displayed in grey. Grids which are *not* in our data but have non-missing PRISM values are displayed in red. Grids which are *not* in our data and have missing values in PRISM are displayed in blue.

Figure 2: States on Discs



Notes: Figure displays the set of states (all years) that are contained on each disc. Disc 8 also includes data that is aggregated over the cropland area within a county by month as well monthly PRISM totals average over the cropland area within a county.

Table 1: Link Between PRISM Grid Cells and Our Data Set

	Number of Grids
Non-missing value in PRISM and spatial match	471,159
Missing value in PRISM and no spatial match	390,864
Non-missing value in PRISM but no spatial match	10,480
Missing data in PRISM but spatial match	2
Total	872,505

Table 2: FIPS Code Lookup Table for Contiguous United States

Fips	State Name	Abbrev.	Fips	State Name	Abbrev.
1	Alabama	AL	31	Nebraska	NE
4	Arizona	AZ	32	Nevada	NV
5	Arkansas	AR	33	New Hampshire	NH
6	California	CA	34	New Jersey	NJ
8	Colorado	CO	35	New Mexico	NM
9	Connecticut	CT	36	New York	NY
10	Delaware	DE	37	North Carolina	NC
11	District of Columbia	DC	38	North Dakota	ND
12	Florida	FL	39	Ohio	OH
13	Georgia	GA	40	Oklahoma	OK
16	Idaho	ID	41	Oregon	OR
17	Illinois	IL	42	Pennsylvania	PA
18	Indiana	IN	44	Rhode Island	RI
19	Iowa	IA	45	South Carolina	SC
20	Kansas	KS	46	South Dakota	SD
21	Kentucky	KY	47	Tennessee	TN
22	Louisiana	LA	48	Texas	TX
23	Maine	ME	49	Utah	UT
24	Maryland	MD	50	Vermont	VT
25	Massachusetts	MA	51	Virginia	VA
26	Michigan	MI	53	Washington	WA
27	Minnesota	MN	54	West Virginia	WV
28	Mississippi	MS	55	Wisconsin	WI
29	Missouri	MO	56	Wyoming	WY
30	Montana	MT			

Table 3: List of States Contained on Each DVD

Disc	States
1: Atlantic	Maine (23); New Hampshire (33); Vermont (50); Massachusetts (25); Rhode Island (44); Connecticut (9); New York (36); New Jersey (34); Pennsylvania (42); Delaware (10); District of Columbia (11); Maryland (24); Virginia (51); North Carolina (37); South Carolina (45); Georgia (13); Florida (12)
2: Central Northeast	Ohio (39); Michigan (26); Indiana (18); Illinois (17); Wisconsin (55); Missouri (29); Iowa (19)
3: Central Southeast	Kentucky (21); West Virginia (54); Tennessee (47); Alabama (1); Mississippi (28); Arkansas (5); Louisiana (22); Oklahoma (40)
4: Central Northwest	Minnesota (27); North Dakota (38); South Dakota (46); Nebraska (31); Kansas (20)
5: Central Southwest	Texas (48); New Mexico (35)
6: Mountain North	Montana (30); Idaho (16); Wyoming (56)
7: Mountain South	Nevada (32); Utah (49); Colorado (8); Arizona (4)
8: Pacific	Washington (53); Oregon (41); California (6)
	Aggregates over cropland area by month (STATA files)