Choice and Acquisition: ‘climate’ variables for time-series analyses

Sadie J. Ryan

Choice

- Before you dive into products, or just do what someone else did:
  - Think about the biological mechanisms you might be exploring
  - Think about what meteorological or climate data corresponds to the mechanism.

- Example: mosquitoes (pick your species of choice)
  - When does temperature limit your mosquito of interest, and how?
  - Is minimum temperature likely to be most important? Maximum?
  - Create reasonable hypotheses for mechanistic processes
  - Temperate vs. tropical vs. boreal

Choice, cont.

- Scale
  - You've decided minimum temperature will be important – but when and for how long?
    - Coldest month? Minimum temperature in an hour in a day?
    - First month in which a daily minimum temperature is exceeded? (Thresholds)
    - Average v. cumulative measures (think temperature and precipitation)
  - A bit of Tobler's law – how far does the effect carry (derived from the 'law' that things that are closer are more similar)
    - Is the nearest weather station useful for your organism of interest, is interpolation of data reflecting the likely response at the location?
  - More "geographic" consideration for met/climate variables
    - Where are measurements made? Is ground surface temperature equivalent to air temperature, and when does it matter? Does rain absorb into the surface or sit on it?

Acquisition

- Making the best of things
  - Unless you specifically have a weather station logging your variables of interest next to your trapping/collecting design, you will use proxies in some way
  - Time series, so we need regularly spaced, consistent observation or modeled products
  - (Happy to talk about climate and season descriptors, other environmental variables and proxies later)
- More choices
  - Most ‘weather’ products are modeled (interpolated) in some way – read the documentation carefully, and know what it is – imperfect is often still useful, but may have important limitations
  - EOS products are also modeled, and spatiotemporal aggregations will also determine their utility – don’t be fooled by apparent consistency

Acquisition, cont.

- Two examples:
  - Point extraction of Daymet data – useful for USA-based studies, high frequency data availability, consistency already worked out for you.
  - GEE extraction of MODIS products (EOS) – proxies for rainfall, temp, and NDVI, which can proxy both, in some circumstances

- Before you even start
  - What is the location of your vector time-series?
    - We assume absolutely perfect data, and chances are you have a coordinate pair
    - What projection is it in? Whose GPS unit was it reported from? How accurate or precise is it?
    - Throw it on Google maps and make sure it seems to be where you think it is.

Daymet point extraction example

- Put your point in google maps, or choose an address
- 29.6557572689658, -82.32141674790877
- What is this? This is my house – 405 NE 5th Avenue, Gainesville, FL.

Right click to get coordinates

Earthdata profile with NASA – make an account - https://urs.earthdata.nasa.gov/

Single Pixel Extraction Tool

Coordinates for a latitude, longitude pair (in decimal degrees) can be manually entered below or automatically filled by dragging the balloon in the map. Click on a tile (within the Google Map) to see the latitude and longitude bounds for that tile.

Coordinates:

Latitude:

29.655757268

Longitude:

-82.32141674

Variables:

ALL
DAYL
PRCP
SRAD
SWE
TMAX
TMIN
VP

Date Range **Years**

Start Date:

2023-01-01

End Date:

2023-12-31

Browser URL requests are populated as variables and time subsets are selected in the GUI. See the Web Services page for additional information.

curl -J 'https://daymet.ornl.gov/single-pixel/api/data?lat=29.6557572689658&lon=-82.32141674790877&vars=dayl,prcp,srad,swe,tmax,tmin,vp&start=2023-01-01&end=2023-12-31' -O

wget --content-disposition 'https://daymet.ornl.gov/single-pixel/api/data?lat=29.6557572689658&lon=-82.32141674790877&vars=dayl,prcp,srad,swe,tmax,tmin,vp&start=2023-01-01&end=2023-12-31'

Download Data | Visualize Data | Preview Data

DON’T OPEN YOUR DATA IN EXCEL – demo only
Annoying/informational header lines

	A	B	C	D	E	F	G	H	I
1	Latitude: 29.6557572689658 Longitude: -82.32141674790877								
2	X & Y on Lambert Conformal Conic: 1663970.19 -1195191.03								
3	Tile: 10669								
4	Elevation: 41 meters								
5	All years; all variables; Daymet Software Version 4.0								
6	How to cite: Thornton; M.M.; R. Shrestha; Y. Wei; P.E. Thornton; S-C. Kao; and B.E. Wilson. 2022. Daymet: Daily Surface Weather Data on a 1-km Grid for North America; Version 4 R1. ORNL DAAC; Oak Ridge; Tenne								
7	year	yday	dayl (s)	prcp (mm/	srad (W/m	swe (kg/m^2)	tmax (deg c)	tmin (deg c)	vp (Pa)
8	2023	1	36485.81	0	209.45	0	24.78	14.49	1649.28
9	2023	2	36512.74	0	269.55	0	26.46	12.32	1431.96
10	2023	3	36541.88	0	258.33	0	27.37	13.89	1586.09
11	2023	4	36573.21	13.63	142.41	0	25.96	16.64	1892.82
12	2023	5	36606.72	0	213.62	0	23.96	13.66	1563.08
13	2023	6	36642.38	0	298.69	0	19.44	4.41	836.76
14	2023	7	36680.17	0	340.31	0	20.92	1.9	700.48
15	2023	8	36720.07	0	350.5	0	23.6	2.88	750.9
16	2023	9	36762.04	0	306.36	0	23.47	7.53	1038.08
17	2023	10	36806.07	0	302.26	0	20.83	5.67	913.64
18	2023	11	36852.14	0	354.46	0	23.99	3.16	766.32
19	2023	12	36900.2	13.44	258.02	0	25.63	5.92	929.53
20	2023	13	36950.23	0	237.29	0	18.02	7.39	1028.28
21	2023	14	37002.21	0	239.45	0	11.3	1.04	658.37
22	2023	15	37056.1	0	342.01	0	15.13	-2.13	522.29
23	2023	16	37111.87	0	373.9	0	19.75	-1.86	532.83

Last notes on a Daymet pull

- For a single point, the question about what the projection is in the Daymet model should be ok, because you are not overlaying rasters, and you can see it all on google maps, both before and during your extraction
  - If you are doing this for multiple points, DEFINITELY THINK ABOUT THE PROJECTION

- User choice about how to wrangle it into R
  - Be super careful about csv formats if you open in Excel, but you probably know this at this stage in your career.
  - Generally, don't. Just pull it into R and chop off rows or subset what you need, check the classes of each variable field you'll use, make sure the units of measurement (temperature, VP, etc) are as you expect, and enjoy!

Pause – were you able to navigate this so far?

- Why is Daymet a bit deluxe? – high resolution (pixels are small), high frequency (daily availability).
- Easy UIs for data extraction

- Limitations:
  - Only continental North America (plus HI from 1980 and PR from 1950)
  - Seems to get released by calendar year (up to end of last year)

Quick example of pulling a local weather station’s data using NOAA access

George and Wesley

How to Download Historical Weather Data from NOAA

14 Easy Steps

1. Go to https://www.ncei.noaa.gov/maps/lcd/

2

- **Find the area on the map that you want weather data from. Click and drag to move around and scroll to zoom in or out.**
  In this example, we will retrieve weather data from South Bend.

3.

- **In the “Layers” tab on the left panel, select the tools icon.**

4.

Choose an option from the tools menu to select an area to get weather data from.

- In this case, we will use the rectangle tool.

5.

Use the selected tool to select a point or area.

With the rectangle tool, simply click and drag to form a rectangle over the area of interest.

6.

- **In the "Results" tab of the left panel, click the checkbox for "Select All"** or select the appropriate stations that you would like to receive data from.

7. Click the “Add to Cart” button.

8.

A new tab will open.

- *In this specific step, you may likely experience technical difficulties. If you see an error page, simply reload the page until this error ceases to occur.*
- *Furthermore, if the page loads but you do not see the "LCD CSV" option listed, reload until you do. The page should look like this:*

9.

Select the “LCD CSV” option.

10.

Select the start and end date from which you would like to receive

- **weather data from.**
- Be sure to select not only a year and a month but also a day, or the blank will not update.
- Click the “APPLY” button when finished.

11.

Click the “Continue” button.

If nothing happens, scroll up to see the error message. The most likely error is that you selected more than ten years of data.

12.

Enter your email address so that the download link can be sent to you.

- **Then click "SUBMIT ORDER".**

Enter email address

Please enter your email address. This is the address to which your data links and information regarding this order will be sent. Please read NOAA's Privacy Policy if you have any concerns.

Email Address

Verify Email Address

☐ **Remember my email address**

NOAA will not share your email address with anyone. The email address will not be used for any purpose other than communicating the order status.

EDIT ORDER SUBMIT ORDER

13.

You will immediately receive an email that you don’t need to do anything with. This email can be ignored.

14.

About one minute later, you should receive another email with a download link for the CSV file. Click the link.

Success! You have now downloaded historical weather data from NOAA!

Next up: Google Earth Engine

GEE MODIS product pull

- Walking you through an example – thanks to Dr. Cat Lippi (former VByte Postdoc), and Dr. Nique Etienne for this!

What is this example about? NEON sites

- We will look at 13 NEON terrestrial sites, for which there are observation towers for data, and also tick plot sites
- Terrestrial sites
- Spanning a chunk of the Northeastern US

We realized the tick plots and the NEON climate data may not overlap a lot

OSBS

NEON tower

Tick plots

UNDE
BART
HARV
BLAN
SERC
MLBS
SCBI
ORNL
GRSM
TALL
JERC
OSBS
DSNY
N
0 125 250 500 Km
NEON tower
Tick drag plot
MODIS LST (°C)
Mar
May
Jul
Sep
Nov
Date
2.27 Km
0 0.4 0.8 1.5 Km
Temperature (°C)
Ticks
MODIS LST (°C)
2018
2019
2020
2021
2022
2023

UNDE
BART
HARV
BLAN
SERC
MLBS
SCBI
ORNL
GRSM
TALL
JERC
OSBS
DSNY
N
0 125 250 500 Km

BART
BLAN
DSNY
GRSM
HARV
JERC
MLBS
ORNL
OSBS
SCBI
SERC
TALL
UNDE
Land Surface Temperature (°C)
log2 Daily Tick Count
2018
2020
2022
2024
Date

BART
BLAN
DSNY
GRSM
HARV
JERC
MLBS
ORNL
OSBS
SCBI
SERC
TALL
UNDE
Normalized Difference Vegetation Index
log2 Daily Tick Count
2018
2020
2022
2024
Date

To start

- Need to create GEE account
- Can add this to existing Google/Gmail account
- Sign up for noncommercial/ academic use

Earthengine.google.com

More things to keep in mind

This is all in python, which is pretty good about highlighting when you have a syntax issue.

- Still good to familiarize yourself with conventions (e.g., creating objects with “var”, ending functions with “;”, etc)

The worst part of this is that the script runs all at once (i.e., you can’t run individual lines like in R).

- This can make it tricky to test code/troubleshoot, but you can still use “//” to “turn off” blocks of code (i.e., use like # in R)

Also be mindful of saving script often.

- If you click to open a different script, for example, to copy and paste some code, any unsaved changes will be lost. This is dumb.

Hit the ‘Save’ button often in code editor

Navigating to the Code Editor

Code Editor

- Similar enough to RStudio

Code Editor

Start a script under the ‘Scripts’ tab by clicking NEW

Can also create new project groups and folders

Docs

Docs tab has a directory of functions

Provides definitions, arguments, and snippets of code that you can copy and paste into your script

Assets – for this, 13 sites are our asset

- If there are spatial datasets that you'll use a lot, it can be more efficient to upload them as Assets
- These are stored in GEE and can be called on in your code directly, as opposed to reading in with script

***If uploading csv, MAKE SURE YOU SAVE A COPY IN utf8 FORMAT FIRST OR IT WON'T WORK

Loading EOS data

- Use the search bar to shop EOS products
- Click on product to open product

You want to give it a try?

- In Materials, you can find the script “NEON_SJR_MODIS_LST.txt”, which will pull time series of Land Surface Temperature (LST) for the 13 NEON sites – both for tower and plot locations
- You will find a csv for the 13 sites (your asset) – “NEON_tickplots_13sites_utf8”
- See if you can get it to run
- See if you can export the data and find it again!
- If you recognize the site names from other data in this workshop, you may find interesting things
- NB: beware of data missingness; even perfect data is imperfect sometimes

Using R to pull GEE

- RGEE is a package – it provides a wrapper for the python code, you'll need to deal with the API piece anyway, but could be fun.

- I have not used it, so I'm not your expert! However, it looks like a good and supported option, which is nice.

EOS vs. weather station data caveats

- EOS – comes from satellites, ‘looking down’ onto things
  - LST – land surface temperature – reflectance converted to temperature
  - What if it’s a forest? Is that what your vector is experiencing?
  - What if there are lots of clouds? Worse, what if cloud cover occurs more in a specific season?
  - Lots of products available, but need to catch up to real time – not a common issue, but processing raw data is a whole different set of skills

- Weather station data
  - Point based data that gets interpolated to represent irregular region shapes – great if you have lots of them, less great when they are sparse contributors
  - Require people, so data gaps can occur during holidays and natural disasters; know the missing data protocol for whatever you are using for your specific area
  - Very much influenced by geography – hard to have globally consistent coverage, which is obscured by global products
  - Tracking down the nearest station to your observations is possible, data availability (to you) can be very mixed.

There are many products and data sources out there!

- **NOAA** site with lots of gridded data - https://psl.noaa.gov/data/gridded/
- **Daymet**
- **Merraclim** – climate products, rather than weather variables
- **Worldclim** – similarly to Merraclim, climate products and projections
- **ERA**
- **Copernicus**

- *Let's generate some sources and sites together*