MODIS Daily Land Surface Temperature Estimates in Google Earth Engine as an Aid in Geothermal Energy Siting

Franklin G. Horowitz

Dept. EAS, Snee Hall, Cornell University, Ithaca, NY 14853, U.S.A.

frank.horowitz@cornell.edu

ABSTRACT

The four-samples-per-day MODIS thermal infrared observations from the EOS Terra and Aqua satellite systems have been ingested into the Google Earth Engine platform. This means that a potent dataset -- with coverage from satellites of orbital inclination 98.1 degrees, deployed since the year 2000, and roughly 1km resolution -- is easily available. This is a massive dataset of many tens to hundreds of terabytes, available spinning online for the first time in a platform that allows easy computation using Google servers. For the geothermal community, simple Temperature statistics such as the Mean Annual Surface Temperature (e.g. Horowitz and Regenauer-Lieb, 2009) or daily T range statistics can then be calculated for any region covered by data. Having such statistics easily available should be of aid in making siting decisions for geothermal electrical generation or direct use applications. Examples of such calculations will be shown (hopefully live) at the meeting.

1. INTRODUCTION

Google Earth Engine (GEE; https://earthengine.google.org) is a remote sensing processing platform under development by Google. It has (at least) petascale distributed online storage as well as many tens of thousands (at least) of distributed Google CPUs available to its users. At present, GEE is limited to participants in its "trusted tester" program, but access to that program is widely available – currently at no cost.

Quoting from the MODIS website (http://modis.gsfc.nasa.gov/about/): "MODIS (or Moderate Resolution Imaging Spectro-radiometer) is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths."

Of particular interest in the geothermal arena, one of the calibrated "Level 3" remote sensing products from MODIS is the "Land Surface Temperature" (LST) estimate available twice daily with approximately 1km pixel resolution from each satellite (http://modis.gsfc.nasa.gov/data/dataprod/dataproducts.php?MOD_NUMBER=11). GEE has ingested these LST data from both the Terra and Aqua satellites, for the entire duration of their (ongoing) missions.

For geothermal work, because our resource temperatures are so low compared to temperatures available from fossil fuel combustion, Carnot (and real-world) efficiencies for electricity generation as well as the exergy available for direct use applications depend heavily on the rejection temperatures of the thermodynamic system. This means that knowing the historical LST distribution everywhere on the planet can be quite useful in assessing the economics of a proposed site. Additionally, availability of the time series of LST measurements holds the tantalizing promise of robust estimation of other spatial quantities of interest to geothermal explorationists – perhaps even the spatial distribution of geothermal heat flux?

Past work with these data (Horowitz and Regenauer-Lieb, 2009) estimated the Mean Annual Surface Temperature (MAST) over Australia and New Zealand from the Terra satellite with data up through 2006. The dominant component of the workload there was simply the logistics of acquiring the data from the tape libraries at the EROS data center (https://eros.usgs.gov/) and organizing the data locally. Actually writing programs for exploratory data analysis was a relatively minor part of the workload. Note that this was the traditional style of large computation; "moving the data to the compute engine".

Now that GEE has ingested the LST data, all of the data management logistics have in essence already been taken care of (for the whole planet, and for the entire duration of the ongoing data acquisition). Hence in GEE, the dominant part of the former workload has already been performed "once and only once" by Google as a service to its users. A price paid by users for this service is the need to learn the Google "map-reduce" style of large computation (i.e. "moving the compute power to the data"). Fortunately, for experienced users of array languages such as MATLAB or numpy, much of that should be easily digestible.

2. METHODS

While slightly nontraditional, this section is the heart of the paper.

There are multiple ways of interacting with the MODIS LST data in GEE, but arguably the style most appropriate for geothermal work is simply to write a script telling the servers what computation to perform. The native development environment for GEE is JavaScript (e.g. https://en.wikipedia.org/wiki/JavaScript), although Python scripting is available too.

Given that we have a time series of LST measurements for each ~1km pixel in a Region Of Interest (ROI), a standard way from basic statistics to assess the variability of the measurements in each pixel would be to calculate a histogram – i.e. a discrete version

of the probability density function (PDF). A slightly different way, representing exactly the same information, would be to estimate the Cumulative Distribution Function (CDF) – the integral of the PDF.

2.1 Google Earth Engine Computation

Listing 1 displays the entirety of the code required to perform the LST CDF calculation for a portion of Upstate New York surrounding Cornell University. Lines 9-18 are defining and setting up the boundaries for the ROI, needed in 2 different fashions in the code. Lines 20-23 instruct GEE to access the Terra (MODIS/MOD11A1) instrument data for a 4 year period from 2002 through 2005 inclusive. The "bounds" filter gives a hint to the GEE infrastructure as to which spatial portion of the data will be accessed, so that the computation can be moved to the appropriate location in Google's infrastructure nearby (i.e. with high data rate transmission) to where the data are stored. The user neither knows nor has to care about the mechanics of this behind the scenes operation – GEE's infrastructure simply takes care of it! The "select" filter on line 23 picks band 0 from the collection, corresponding to the daytime pass LST measurement of the Terra satellite. Line 26 is an example of a "map/reduce" style operation used to parallelize the computation across GEE's infrastructure. The result in the variable "total" contains an image whose pixel values count the total number of valid pixels from the ROI and duration previously specified. It is used to normalize the rest of the calculation so that the probability estimates scale to unity. Lines 29 through 35 set up some needed data structures for the actual histogram counts. Lines 37 through 41 set up a suite of constant-valued images – boundaries of the histogram "bins" – that are then collected into a single multi-band image in Line 43. Lines 46 through 49 are the heart of the calculation. For each LST estimate in the time series in each pixel, the LST is compared to the histogram boundaries in "bins" and then counted and normalized. Lines 50 through 54 simply output the resulting multiband CDF image – at 926 meters per pixel resolution – into the specified Google Drive folder as a GeoTiff. Finally, lines 57 through 59 display aspects of the calculation in the GEE "playground" – the website where such code as this is interactively composed and saved, debugged, executed from, and communicated to others via URLs.

```
1: // ExportHistogramImagesForUpstateNY
2:
3: // The first comment seems to get placed into the filename during SAVE...
4:
5: // Google Drive foldername for export (shows up in the account associated with
   this username)
6: var drive_folder = 'Google EE Exports';
7:
8: // ROI for clipping
9: var west = -79.0;
10: var east = -73.5;
11: var south = 42.0;
12: var north = 44.0;
13: var upstateNyROI = ee.Feature.Rectangle(west, south, east, north);
14: var upstateNyROIString = '[[' + west + ',' + north + '], ' +
15:                          '[' + east + ',' + north + '], ' +
16:                          '[' + east + ',' + south + '], ' +
17:                          '[' + west + ',' + south + ']]';
18:
19: // MODIS LST Image
20: var collection = ee.ImageCollection('MODIS/MOD11A1')
21:     .filterDate('2002-01-01', '2005-12-31')
22:     .filterBounds(upstateNyROI)
23:     .select(0);
24:
25: // Total number of values at each pixel.
26: var total = collection.reduce('count');
27:
28: // Make histogram bins
29: var bins = [];
30:
31: // Normalization for MODIS LST is to multiply the int by 0.02 to get Kelvin
32: // Hence this range is for -30 to 60 deg C in increments of 1 degree C
33: var i_beg = 11650;
34: var i_end = 16150;
35: var i_incr = 50;
36:
37: for (var i = i_beg; i <= i_end; i += i_incr) {
38:   // Create a constant image and give it a nice name.
39:   var b = ee.Image(i).select([0], ["bin_" + i*0.02]);
40:   bins.push(b);
41: }
42: // Combine all the bins into 1 image.
43: bins = ee.Image(bins);
44:
45: // Compute the normalized counts.
46: var counts = collection.map(function(img) {
47:   return img.lte(bins);
48: }).reduce('sum').divide(total);
49:
50:
51: exportImage(counts,'counts_UpstateNY',
52: {'region':upstateNyROIString,
53:  'scale':926,
54:  'driveFolder':drive_folder});
55:
56:
57: addToMap(counts,null,'Upstate_NY_CDF');
58: addToMap(ee.Feature(upstateNyROI),null,'Upstate NY ROI');
59: centerMap((west+east)/2.0,(north+south)/2.0, 8);
```

Listing 1: JavaScript code required in GEE to generate and export each pixel's Cumulative Distribution Function from Land Surface Temperature time series measured by the MODIS Terra instrument. See text for a discussion of the details.

Noel Gorelick, the founder of GEE, says his stated goal for the GEE environment was to make remote sensing calculations "ridiculously easy" for users. With the possible exception of Lines 46 through 48 – kindly provided by him as an example when I was first tackling this problem and asking for help on GEE's discussion group – I think that the clarity of Listing 1 attests to how close they are to attaining that goal. In particular, I contrast this 1 page of code with my previous efforts behind the computations of Horowitz and Regenauer-Lieb (2009) – there is simply no contest.

Note also that to repeat the calculation for a different region, time duration, or day/night simply requires substituting appropriate values in Lines 9 through 12 and 20 through 23. Using data from both satellites and both day and night observations is equally straightforward. Some other simple statistics – such as mean or median annual surface temperature – single calls to appropriate routines in the GEE environment.

2.2 Visualizing the Result

The LST data are fundamentally spatiotemporal (2D space by 1D time), and methods to visualize even such basic statistics as PDFs or CDFs of such data are not well described – or perhaps even well known – in the scientific visualization literature. There are hints about placing a violin plot glyph (e.g. https://en.wikipedia.org/wiki/Violin_plot) over each pixel, or plotting isosurfaces at the inter-quartile boundaries, but I found no code or examples available during a cursory search. I hope to have some kind of more appealing visualization by the time of the meeting, but for now, I simply plot the CDF value for a specified temperature over the ROI. Figure 1 shows such an image resulting from the code of Listing 1.
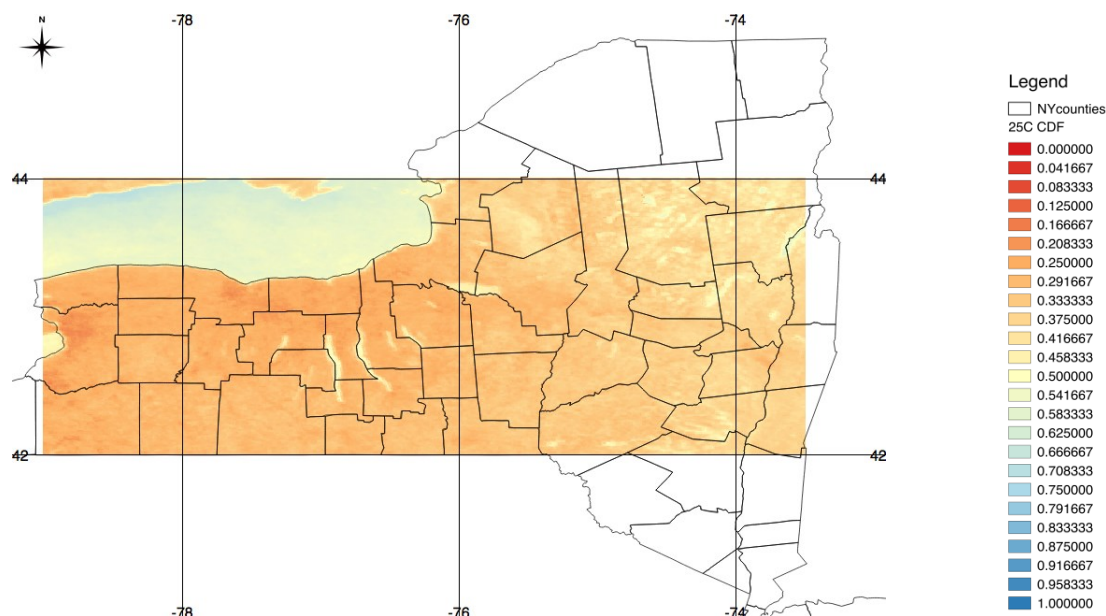
Figure 1: The CDF value for 25 ºC (77 ºF) over the ROI defined in Listing 1. Colors, keyed in the Legend, represent the fraction of the daytime MODIS Terra LST measurements found below 25 ºC occurring from 2002 through 2005 inclusive. The computations are at a resolution of slightly less than 1km per pixel. Urban "heat island" effects are clearly visible in built-up environments (e.g. Buffalo, Rochester, Albany, and even Ithaca). Even though the MODIS LST product is not calibrated for water temperatures (to the best of my knowledge – a separate product is available for sea surface temperatures: http://modis.gsfc.nasa.gov/data/dataprod/dataproducts.php?MOD_NUMBER=28) bodies of water still are easily discernable. The color scale was chosen to display bodies of water in cool colors and land in warmer colors. Also overlaid are the county boundaries of New York state and a grid for geographical reference.

Figure 2 shows the mean temperature in Kelvin for the same ROI and duration, while Figure 3 shows the corresponding median value.

3. CONCLUSION

For geothermal applications, such statistical information can be quite useful. Rather than simply assuming a MAST of 20 ºC for (e.g.) geothermal gradient estimates or for powerplant/direct-use exergy analyses, actual data are now readily available anywhere on the planet. Economic decisions – dependent on the surface temperature for e.g. water requirements in cooling tower designs, air-cooling designs, or district-heating designs – now can be based on long-term rather than short-term measurements.

I expect that other quantities of potential geothermal interest are still buried in these data. Preliminary work on the Afar region of Ethiopia demonstrates that year-to-year seasonal statistical variations can be used to locate ash eruptions from the dike intrusion events (e.g. Ayele et al., 2009), of the mid 2000's decade. Indeed, re-implementing the MODVOLC algorithm (e.g. http://modis.higp.hawaii.edu/, or Wright et al., 2004) is likely a matter of only a few lines of code in the GEE environment – if the raw band values have already been ingested. Finally, I still hope that inverse problems using the LST time series information along with other calibrated remotely sensed quantities may ultimately become useful in geothermal exploration to estimate quantities such as geothermal heat flux and very near surface thermal conductivities.
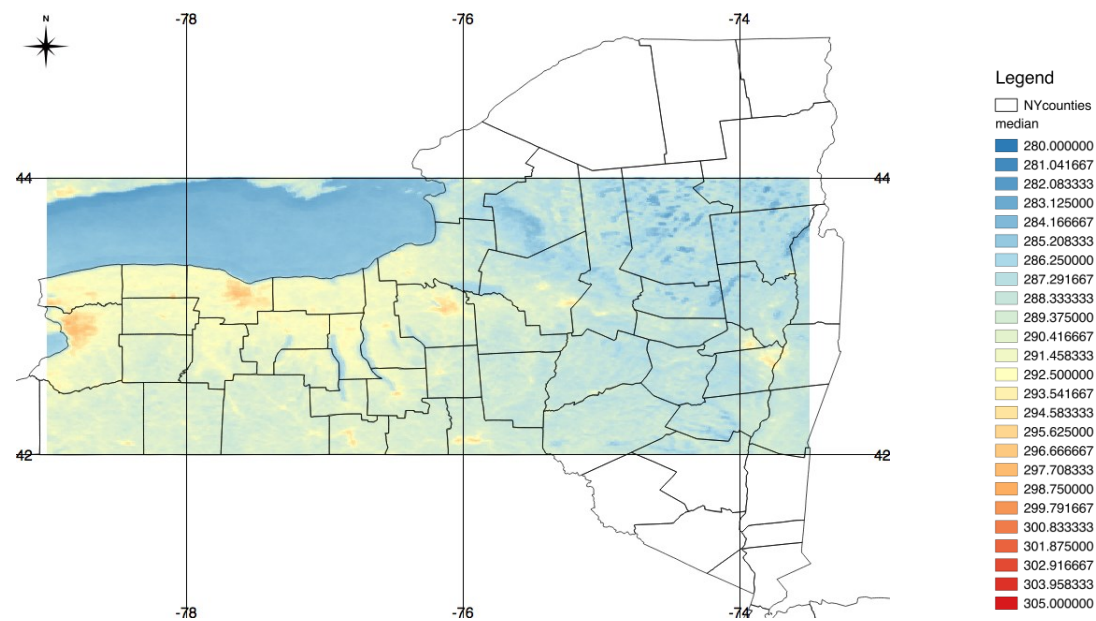
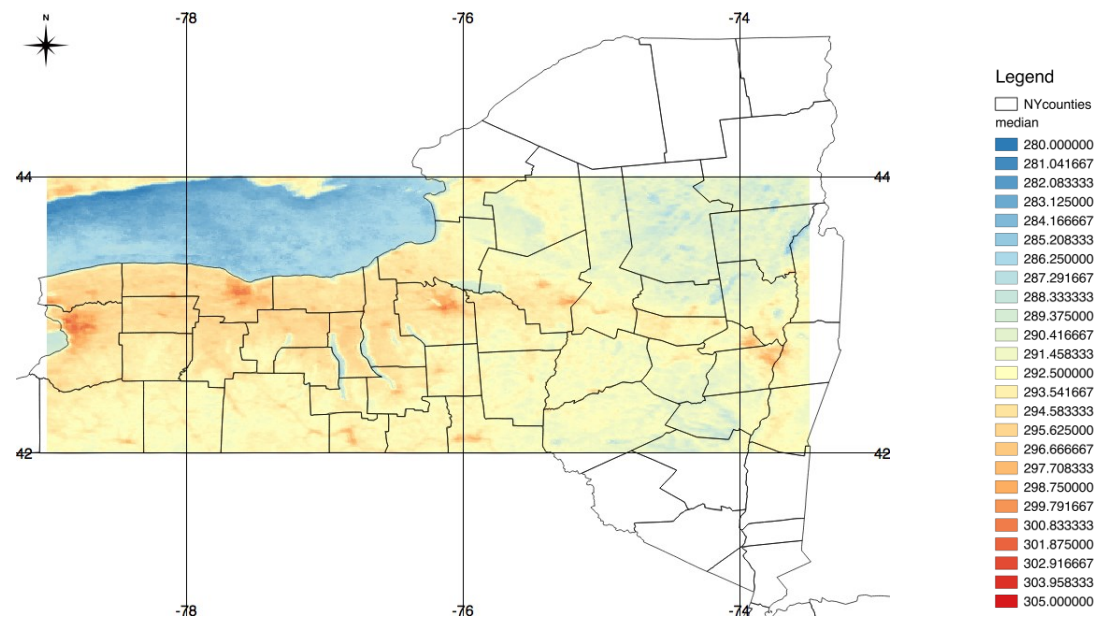**Figure 2: The mean T (in Kelvin) for the same region and duration as in Figure 1.**

Figure 3: The median T (also in Kelvin) for the same region and duration as in Figure 1 and Figure 2. The same color scale is used here as in Figure 2.

REFERENCES

Ayele, A., Keir, D., Ebinger, C., Wright, T. J., Stuart, G. W., Buck, W. R., Jacques, E., Ogubazghi, G., & Sholan, J. (2009). September 2005 mega-dike emplacement in the Manda-Harraro nascent oceanic rift (Afar depression). Geophysical Research Letters, 36 (20). http://dx.doi.org/10.1029/2009gl039605

Horowitz, F. G., & Regenauer-Lieb, K. (2009). Mean annual surface temperature (MAST) and other thermal estimates for Australia and New Zealand from 6 years of remote sensing observations. In Proceedings, NZ Geothermal Workshop, (pp. 65-68). The University of Auckland. http://www.geothermal.org.nz/nzgeothermal2009/1a2456x3.cfm

Wright, R., Flynn, L. P., Garbeil, H., Harris, A. J. L., & Pilger, E. (2004). MODVOLC: near-real-time thermal monitoring of global volcanism. *Journal of Volcanology and Geothermal Research*, *135* (1-2), 29-49. http://dx.doi.org/10.1016/ j.jvolgeores.2003.12.008