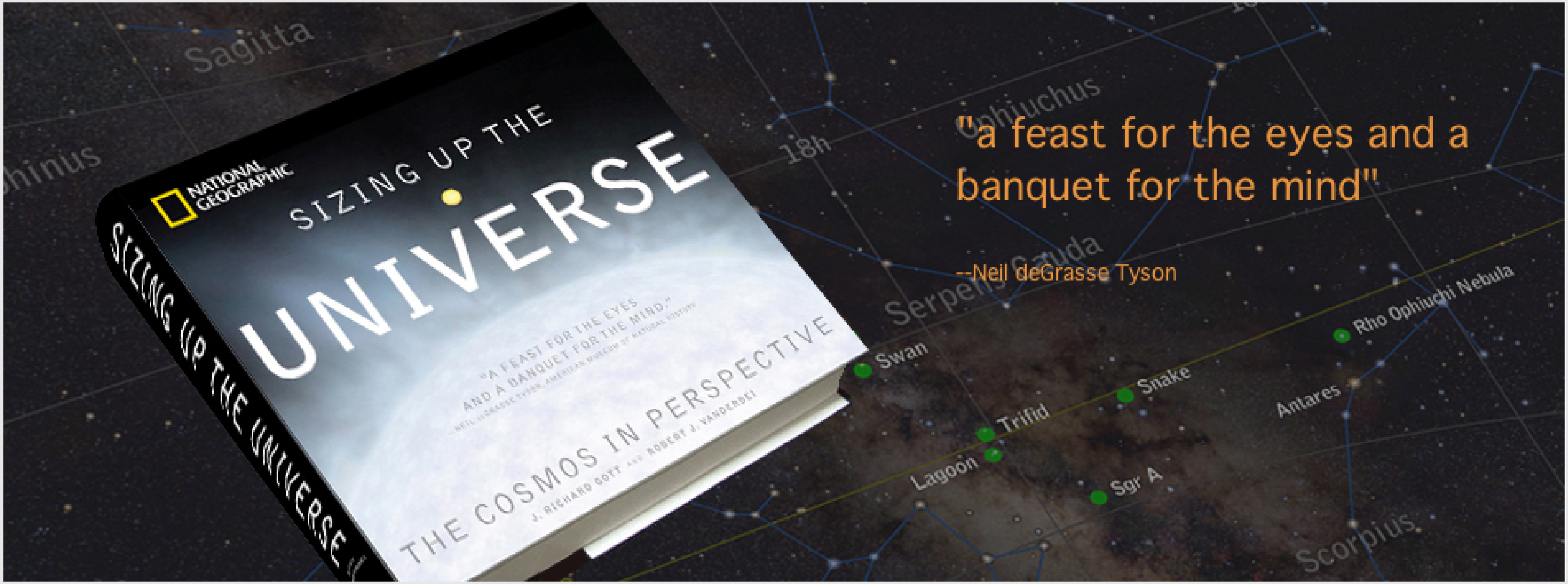
Local Warming

Robert J. Vanderbei

2011 January 11

ICS, Monterey CA

http://www.princeton.edu/∼rvdb



www.sizinguptheuniverse.com

Introduction

There has been so much talk about global warming.

Is it real?

Is it human-made?

Global warming starts at home.

So, let's address the question of *local warming*.

Has it been getting warmer in NJ?

The Data

The *National Oceanic and Atmospheric Administration* (NOAA) collects and archives weather data from thousands of collection sites around the globe. The data format and instructions on how to download the data can be found on this NOAA website:

ftp://ftp.ncdc.noaa.gov/pub/data/gsod/readme.txt

The list of the roughly 9000 weather stations is posted here:

ftp://ftp.ncdc.noaa.gov/pub/data/gsod/ish-history.txt

Perusing this list, I discovered that McGuire Air Force Base, located not far from Princeton NJ, is one of the archived weather stations. Since, the data is archived in one year batches, I wrote a UNIX shell script to grab the 55 annual data files for McGuire and then assemble the relevant pieces of data into a single file. Here is the shell script...

http://www.princeton.edu/~rvdb/ampl/nlmodels/LocalWarming/McGuireAFB/data/getData.sh

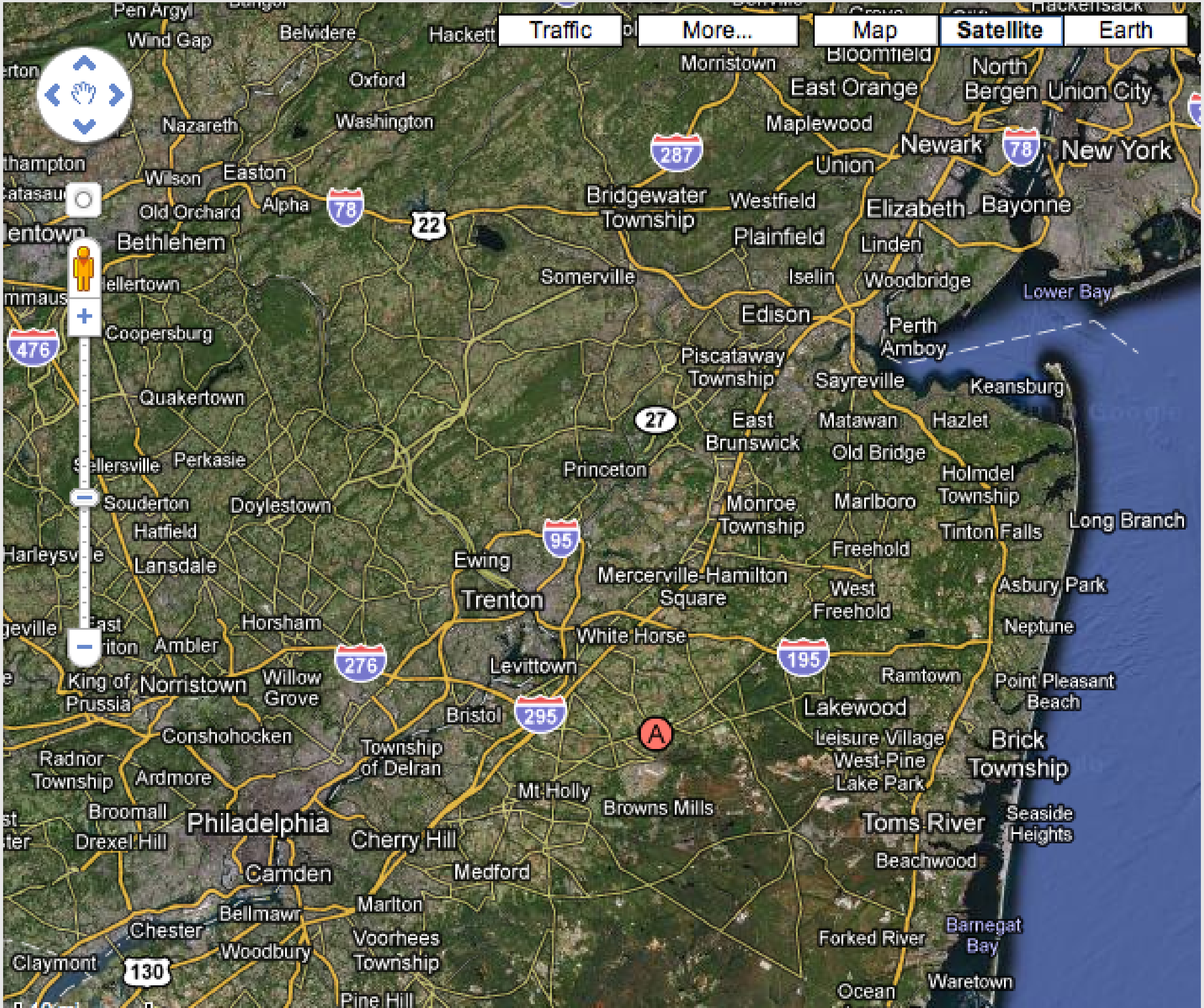
The resulting pair of data files that I used as input to my local climate model are posted at...

http://www.princeton.edu/~rvdb/ampl/nlmodels/LocalWarming/McGuireAFB/data/McGuireAFB.dat

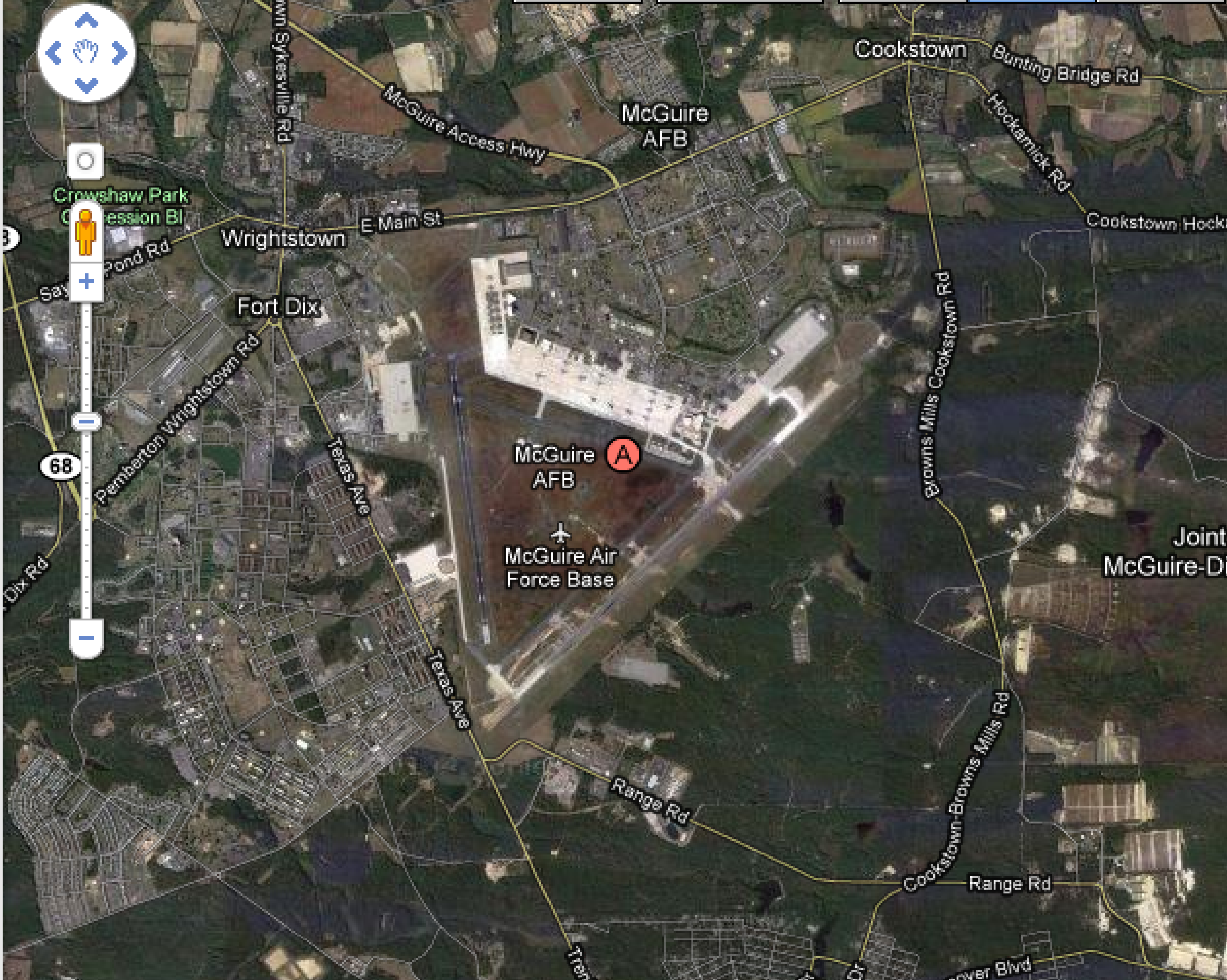
and

http://www.princeton.edu/~rvdb/ampl/nlmodels/LocalWarming/McGuireAFB/data/Dates.dat

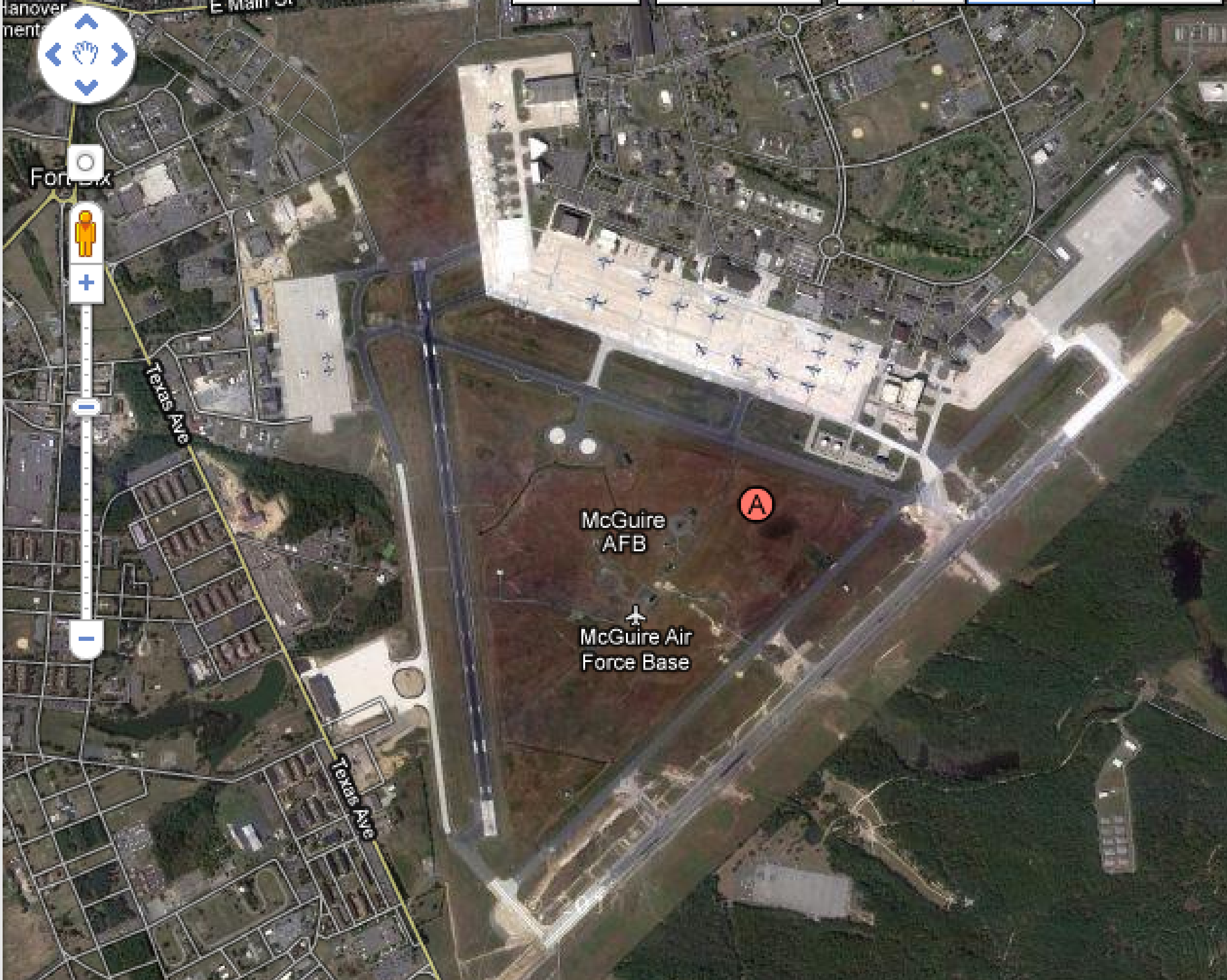
McGuire AFB



Traffic
More...
Map
Satellite
Earth
Pen Argyl
Wind Gap
Belvidere
Hackett
Morristown
Bloomfield
North Bergen
Union City
East Orange
Oxford
Nazareth
Washington
Maplewood
Newark
New York
Union
Wilson
Easton
Old Orchard
Alpha
Bridgewater Township
Westfield
Elizabeth
Bayonne
Bethlehem
Plainfield
Linden
Hellertown
Somerville
Iselin
Woodbridge
Lower Bay
Coopersburg
Edison
Perth Amboy
Piscataway Township
Sayreville
Keansburg
Quakertown
East Brunswick
Matawan
Hazlet
Old Bridge
Holmdel Township
Perkasie
Princeton
Souderton
Doylestown
Monroe Township
Marlboro
Tinton Falls
Long Branch
Hatfield
Ewing
Freehold
Lansdale
Mercerville-Hamilton Square
West Freehold
Asbury Park
Trenton
Horsham
White Horse
Neptune
Ambler
Levittown
Ramtown
Point Pleasant Beach
King of Prussia
Norristown
Willow Grove
Lakewood
Bristol
Conshohocken
Leisure Village
Brick Township
West-Pine Lake Park
Radnor Township
Ardmore
Township of Delran
Mt Holly
Browns Mills
Toms River
Seaside Heights
Broomall
Philadelphia
Drexel Hill
Cherry Hill
Medford
Beachwood
Camden
Marlton
Bellmawr
Chester
Voorhees Township
Barnegat Bay
Forked River
Woodbury
Claymont
Waretown
Ocean Acres
Pine Hill
287
78
22
476
27
95
276
295
195
130
A



Cookstown
Bunting Bridge Rd
McGuire AFB
McGuire Access Hwy
Sykesville Rd
Hockamick Rd
Crowshaw Park
E Main St
Wrightstown
Pond Rd
Cookstown Hockan
Fort Dix
Pemberton Wrightstown Rd
Browns Mills Cookstown Rd
68
Texas Ave
McGuire AFB
A
McGuire Air Force Base
Joint
McGuire-Dix
Dix Rd
Range Rd
Cookstown-Browns Mills Rd
Range Rd


E Main St
Fort Dix
Texas Ave
McGuire
AFB
McGuire Air
Force Base
Texas Ave
A

The Model

Let T_d denote the average temperature in degrees Fahrenheit on day $d \in D$ where D is the set of days from January 1, 1955, to August 13, 2010 (that's 20,309 days).

We wish to model the average temperature as a *constant* a_0 plus a *linear trend* $a_1 d$ plus a sinusoidal function with a one-year period representing *seasonal changes*,

$$a_2 \cos(2\pi d/365.25) + a_3 \sin(2\pi d/365.25),$$

plus a sinusoidal function with a period approximately equal to 10.7 years to represent the *solar cycle*,

$$a_4 \cos(2a_6\pi d/(10.7 \times 365.25)) + a_5 \sin(2a_6\pi d/(10.7 \times 365.25)).$$

The parameters $a_0, a_1, \ldots, a_6$ are unknown regression coefficients. We wish to find the values of these parameters that minimize the sum of the absolute deviations:

$$\begin{aligned}\min_{a_0,\ldots,a_6} \sum_{d \in D} \Big| \quad & a_0 + a_1 d \\ & + a_2 \cos(2\pi d/365.25) + a_3 \sin(2\pi d/365.25) \\ & + a_4 \cos(a_6 2\pi d/(10.7 \times 365.25)) + a_5 \sin(a_6 2\pi d/(10.7 \times 365.25)) \\ & - T_d \quad \Big|.\end{aligned}$$

Linearizing the Solar Cycle

If the unknown parameter a_6 is *fixed at* 1, forcing the solar-cycle to have a period of exactly 10.7 years, then the problem can be reduced to a *linear programming problem*.

If, on the other hand, we allow a_6 to vary, then the problem is *nonlinear* and even *nonconvex* and therefore much harder in principle. Nonetheless, if we initialize a_6 to one, then the problem might, and in fact does, prove to be tractable.

Note: This *least-absolute-deviations* (LAD) model automatically ignores "outliers" such as the record heat wave of 2010.

AMPL Model

```
set DATES ordered;
param avg {DATES};
param day {DATES};
param pi := 4*atan(1);

var a {j in 0..6};
var dev {DATES} >= 0, := 1;

minimize sumdev: sum {d in DATES} dev[d];
subject to def_pos_dev {d in DATES}:
    a[0] + a[1]*day[d] + a[2]*cos( 2*pi*day[d]/365.25)
                       + a[3]*sin( 2*pi*day[d]/365.25)
                       + a[4]*cos( a[6]*2*pi*day[d]/(10.7*365.25))
                       + a[5]*sin( a[6]*2*pi*day[d]/(10.7*365.25))
        - avg[d]
    <= dev[d];
subject to def_neg_dev {d in DATES}:
    -dev[d] <=
    a[0] + a[1]*day[d] + a[2]*cos( 2*pi*day[d]/365.25)
                       + a[3]*sin( 2*pi*day[d]/365.25)
                       + a[4]*cos( a[6]*2*pi*day[d]/(10.7*365.25))
                       + a[5]*sin( a[6]*2*pi*day[d]/(10.7*365.25))
        - avg[d];
```

AMPL Data and Variable Initialization

```
data;

set DATES := include "data/Dates.dat";
param: avg := include "data/McGuireAFB.dat";
let {d in DATES} day[d] := ord(d,DATES);

let a[0] := 60;
let a[1] := 0;
let a[2] := 20;
let a[3] := 20;
let a[4] := 0.01;
let a[5] := 0.01;
let a[6] := 1;
```

The nice thing about AMPL and LOQO is that anyone can use these programs via the NEOS server at Argonne National Labs...

http://www-neos.mcs.anl.gov/

The Results

The linear version of the problem solves in a small number of iterations and only takes a minute or so on my MacBook Pro laptop computer. The nonlinear version takes more iterations and more time but eventually converges to a solution that is almost identical to the solution of the linear version. The optimal values of the parameters are

$$
\begin{aligned}
a_0 &= 52.6\ ^\circ\text{F} \\
a_1 &= 9.95 \times 10^{-5}\ ^\circ\text{F/day} \\
a_2 &= -20.4\ ^\circ\text{F} \\
a_3 &= -8.31\ ^\circ\text{F} \\
a_4 &= -0.197\ ^\circ\text{F} \\
a_5 &= 0.211\ ^\circ\text{F} \\
a_6 &= 0.992
\end{aligned}
$$

From a_0, we see that the nominal temperature at McGuire AFB was 52.56 °F (on January 1, 1955).

We also see, from a_1, that there is a positive trend of 0.000099 °F/day. That translates to 3.63 °F per century—in amazing agreement with results from global climate change models.

A 95% *confidence interval* for a_1 is $[3.6132$ °F, 3.7133 °F$]/100$ yrs.

Magnitude of the Sinusoidal Fluctuations

From a_2 and a_3, we can compute the amplitude of annual seasonal changes in temperatures...

$$\sqrt{a_2^2 + a_3^2} = 22.02 \text{ °F}.$$

In other words, on the hottest summer day we should expect the temperature to be 22.02 degrees warmer than the nominal value of 52.56 degrees; that is, 77.58 degrees. Of course, this is a daily average—daytime highs will be higher and nighttime lows should be about the same amount lower.

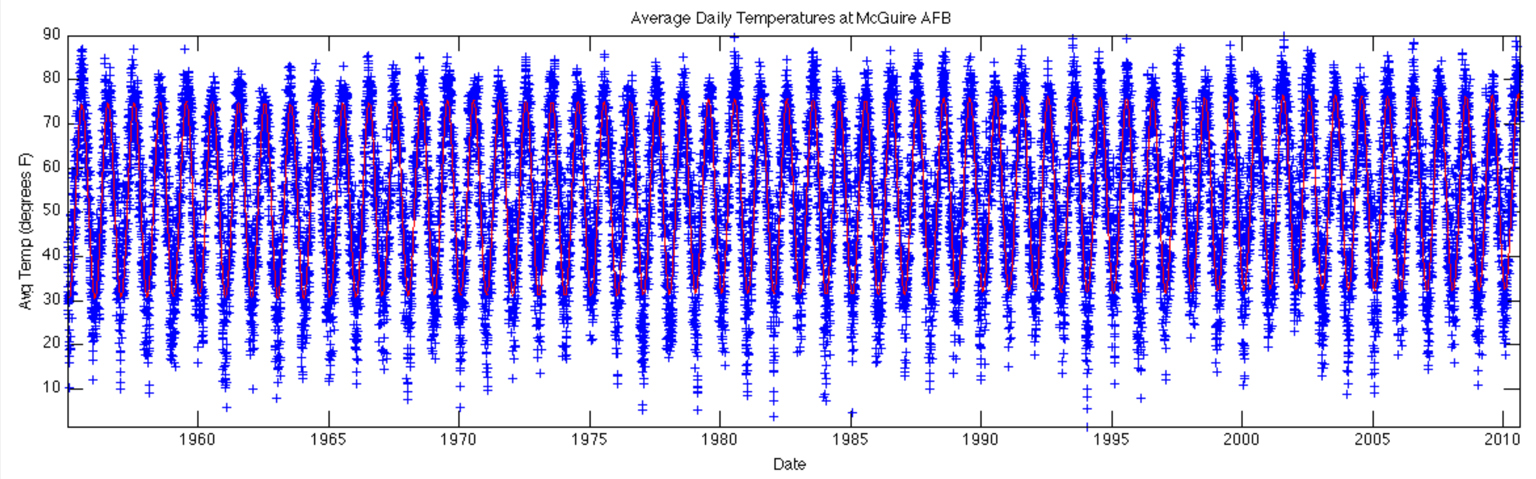
Similarly, from a_4 and a_5, we can compute the amplitude of the temperature changes brought about by the solar-cycle...

$$\sqrt{a_4^2 + a_5^2} = 0.2887 \text{ °F}.$$

The effect of the *solar cycle* is real but relatively small.

The fact that a_6 came out slightly less than one indicates that the solar cycle is slightly longer than the nominal 10.7 years. It's closer to $10.7/a_6 = 10.78$ years.

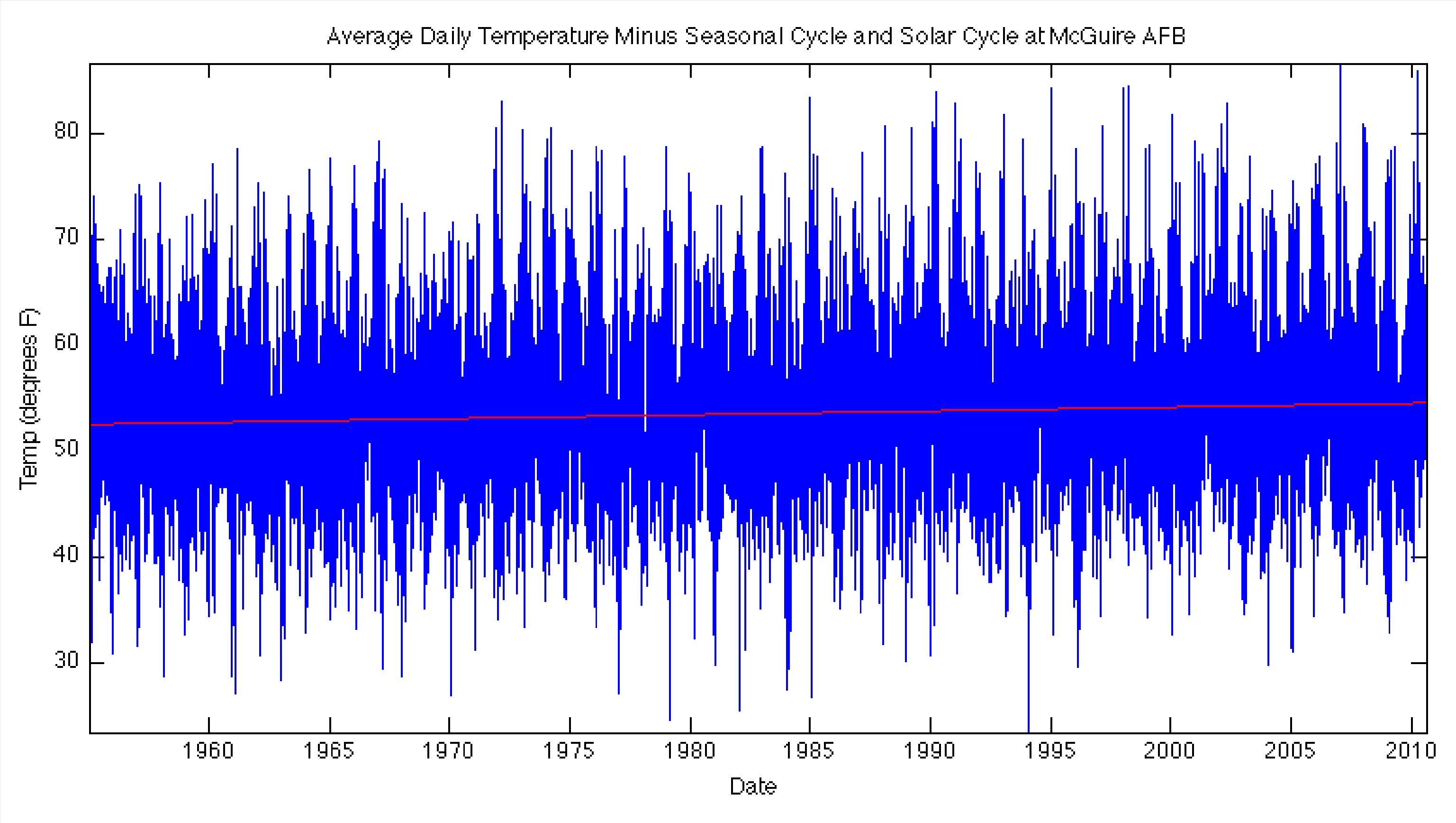
Plot Showing Actual Data and Regression Curve



Blue: Average daily temperatures at McGuire AFB from 1955 to 2010.
Red: Output from least absolute deviation regression model.

Seasonal fluctuations completely dominate other effects.

Subtracting Out Seasonal Effects

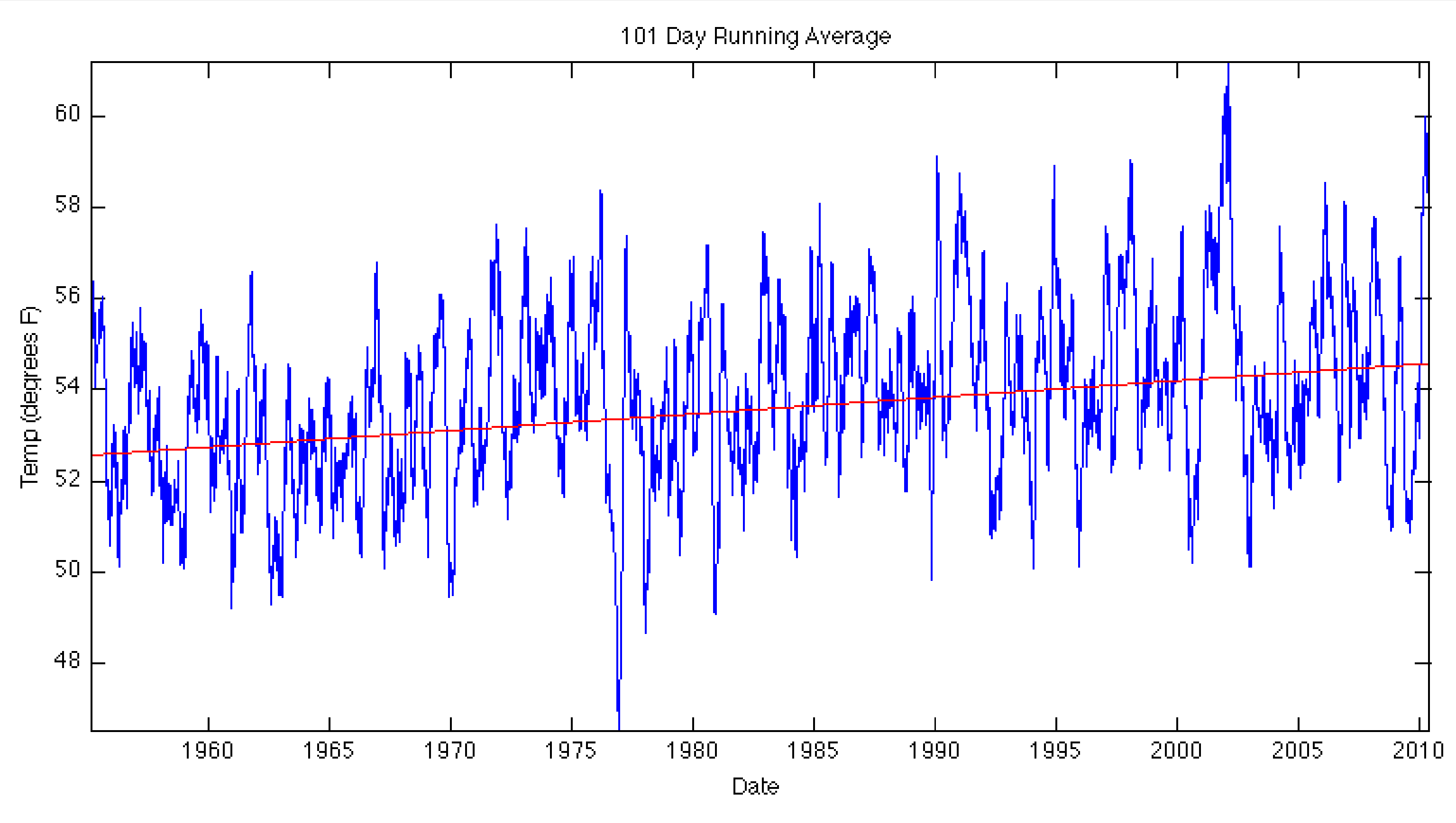


As before but with sinusoidal seasonal variation removed and sinusoidal solar-cycle variation removed as well.

Even this plot is noisy simply because there are many days in a year and some days are unseasonably warm while others are unseasonably cool.

Smoothed Seasonally Subtracted Plot

To smooth out high frequency fluctuations, we use 101 day rolling averages of the data.



In this plot, the long term trend in temperature is clearly seen. In NJ we have *local warming*.

Means, Medians, and Optimization

Let $b_1, b_2, \ldots, b_n$ denote a set of measurements.

Solving

$$\text{argmin}_x \sum_i (x - b_i)^2$$

computes the *mean*.

Solving

$$\text{argmin}_x \sum_i |x - b_i|$$

computes the *median*.

Medians correspond to *nonparametric statistics*. Nonparametric confidence intervals are given by percentiles. The p-th percentile is computed by solving the following optimization problem:

$$\text{argmin}_x \sum_i \left(|x - b_i| + (1 - 2p)(x - b_i) \right).$$

Confidence Intervals For Medians

Assume that $B_1, B_2, B_3, \ldots, B_n$ are independent identically distributed with median m.

Let

$$B_{(1)} < B_{(2)} < B_{(3)} < \cdots < B_{(n)}$$

denote the *order statistics*, i.e., the original variables rearranged into increasing order.

Note: $B_{(k)}$ is the (k/n)-th *sample percentile*.

Then,

$$\begin{aligned} \mathbb{P}(B_{(k)} \leq m \leq B_{(k+1)}) &= \mathbb{P}(B_j \leq m \text{ for } k \text{ indices and} \\ &\qquad B_j \geq m \text{ for the remaining } n-k \text{ indices}) \\ &= \binom{n}{k} \left(\frac{1}{2}\right)^n . \end{aligned}$$

Hence,

$$\mathbb{P}(B_{(k)} \leq m \leq B_{(n-k+1)}) = \sum_{j=k}^{n-k} \binom{n}{j} \left(\frac{1}{2}\right)^n .$$

For any given n, it is easy to choose k so that $\sum_{j=k}^{n-k} \binom{n}{j} \left(\frac{1}{2}\right)^n \approx 0.95.$

Relaxing the Length of a Year

The number 365.25 days per year is not exactly correct.

One extra variable, $a[7]$, can be added to the model. It is analogous to $a[6]$.

The new variable makes the problem even more nonconvex but miraculously it becomes easier to solve.

Here's the output:

$$
\begin{aligned}
a_0 &= 52.5 \text{ °F} \\
a_1 &= 1.02 \times 10^{-4} \text{ °F/day} \\
a_2 &= -20.1 \text{ °F} \\
a_3 &= -9.02 \text{ °F} \\
a_4 &= -0.194 \text{ °F} \\
a_5 &= 0.181 \text{ °F} \\
a_6 &= 0.998 \\
a_7 &= 1.00021
\end{aligned}
$$

The correct value of the correction factor a_7 is 1.0000068 (one day every 400 years). The value of a_7 from the regression model makes a correction in the right direction but it is too big. Clearly this effect was too small to pull out of the data.

Least Squares Solution (Mean instead of Median)

Suppose we change the objective to a sum of squares of deviations:

```
minimize sumdev: sum {d in DATES} dev[d]^2;
```

The resulting model is a *least squares model.*
The objective function is now convex and quadratic and the problem is still easy to solve.
The solution, however, is *sensitive* to outliers.
Here's the output:

$$
\begin{aligned}
a_0 &= 52.6 \text{ °F} \\
a_1 &= 1.2 \times 10^{-4} \text{ °F/day} \\
a_2 &= -20.3 \text{ °F} \\
a_3 &= -7.97 \text{ °F} \\
a_4 &= 0.275 \text{ °F} \\
a_5 &= 0.454 \text{ °F} \\
a_6 &= 0.730
\end{aligned}
$$

In this case, the rate of local warming is 4.37 °F per century.
However, the model produces the *wrong answer* for the period of the solar cycle.

Final Remarks

Close inspection of the output shows that the January 22 is the coldest day in the winter, July 24 is nominally the hottest day of summer, and February 12, 2007, was the day of the last minimum in the 10.78 year solar cycle.

The ampl model and the shell scripts are available on my webpage.

Everyone is encouraged to grab data for any location they like.

Send me the results and I'll compute a *global average*.