L. Hamilton August 6, 2012

PAN-ARCTIC OUTLOOK — Hamilton

1. Extent Projection

A simple regression model for NSIDC mean September *extent* as a function of mean daily sea ice *area* from August 1 to 5, 2012 (and a quadratic function of time) predicts a mean September 2012 extent of **4.02 million km²**, with a confidence interval of plus or minus .32. This supersedes an earlier year-in-advance prediction based on a Gompertz (asymmetrical S curve) model that used data only through September 2011.

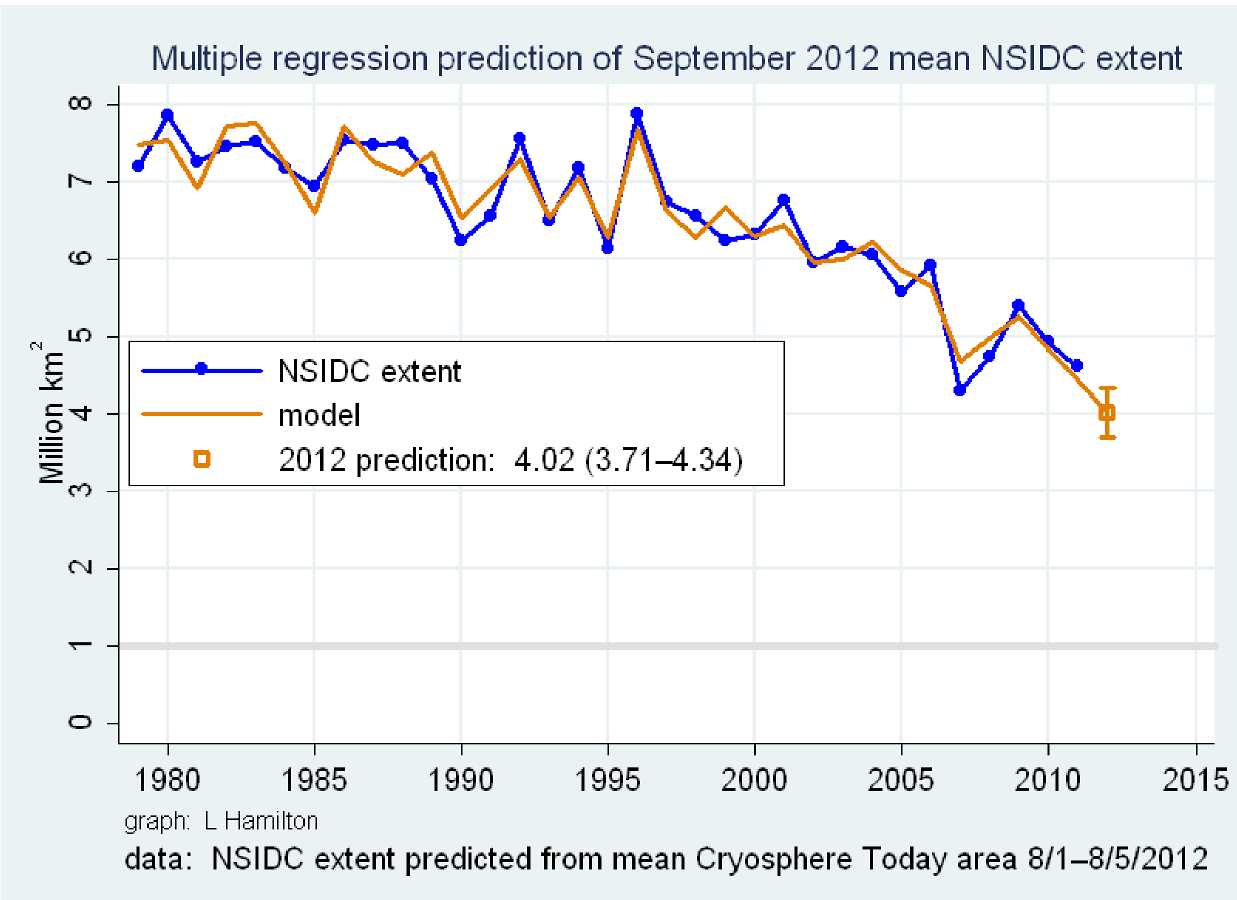

Figure 1

2. Methods / Techniques

The prediction uses only the most recent daily estimates of sea ice area, together with a quadratic time trend. The multiple regression analysis is shown below. In this model, *extent* refers to mean September NSIDC extent, in millions of km^2. L1.*areact* refers to the mean sea ice area for August 1 through 5, based on Cryosphere Today reports (Polar Research Group at the University of Illinois). *year0* is calender year minus 1995, and *year2* is the square of this value. Thus, we have the regression of September extent (1979–2011) on a quadratic time trend, and on the August 1–5 mean sea ice area.

. regress *extent year0 year2* L1.*areact* if *month*==9

```
      Source |       SS       df       MS              Number of obs =      33
-------------+------------------------------           F(  3,    29) =  130.40
       Model |  27.6611663     3  9.22038876           Prob > F      =  0.0000
    Residual |  2.05052842    29  .070707877           R-squared     =  0.9310
-------------+------------------------------           Adj R-squared =  0.9238
       Total |  29.7116947    32  .928490459           Root MSE      =  .26591

------------------------------------------------------------------------------
      extent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       year0 |  -.0069473   .0110912    -0.63   0.536    -.0296314    .0157367
       year2 |  -.0010332   .0006373    -1.62   0.116    -.0023365    .0002702
             |
      areact |
         L1. |   1.109613   .1432935     7.74   0.000     .8165448    1.402681
             |
       _cons |    .417073   .8287226     0.50   0.619    -1.277855    2.112001
------------------------------------------------------------------------------
```

This multiple regression model yields a prediction of **4.02 million km^2, with a confidence interval of plus or minus .32**.

3. Rationale

The rationale for this model is purely statistical. Least-squares regression seeks the best linear equation for predicting observed values of the endogenous variable, based on observations of exogenous variables. Sea ice area estimates have, in the recent past, been more effective than sea ice extent in predicting September ice extent. The decline in September ice extent has been distinctly faster than linear; the Gompertz curves described later provide attractive models, especially for extrapolation. Over the range of values from 1979 to 2011, however, a quadratic time trend is quite close to the Gompertz curve, and has the advantage of simple incorporation within a linear multiple regression model.

The model shown above combines a quadratic time trend with information on sea ice area in the first few days of August to predict the September 2012 extent.

4. Executive Summary

A multiple regression model that combines a quadratic time trend with information on sea ice area during the first few days of August leads to a **September 2012 extent prediction of 4.02 million km^2, with a confidence interval of plus or minus .32**. This model explains about 92% of the variance in September mean extent from 1979 through 2011.

5. Estimate of Forecast Skill

The adjusted R^2 indicates that the model explains about 92% of the variance in observed September ice extent. Its residuals (tested by Ljung–Box *Q* statistics) show no significant autocorrelation: they are not different from white noise.

. corrgram *extres*

```
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        -0.1115  -0.1123   .44881  0.5029          |                 |
2        -0.2771  -0.2959    3.309  0.1912        --|               --|
3        -0.0947  -0.2039   3.6541  0.3013          |                -|
4         0.0293  -0.1008   3.6883  0.4498          |                 |
5         0.0179  -0.0735   3.7015  0.5931          |                 |
6        -0.1282  -0.2295    4.404  0.6222         -|                -|
7        -0.0298  -0.1989   4.4433  0.7275          |                -|
8         0.1610   0.0678   5.6415  0.6873          |-                |
9         0.0174  -0.0685   5.6562  0.7738          |                 |
10       -0.0732  -0.1405   5.9255  0.8215          |                -|
11       -0.1170  -0.0924   6.6443  0.8271          |                 |
12       -0.0648  -0.2862   6.8755  0.8657          |               --|
13        0.1222  -0.3162   7.7376  0.8603          |               --|
14       -0.0055  -0.1947   7.7394  0.9025          |                -|
```

Figure 1 above visualizes how closely the model tracks past data. Its narrow 95% confidence interval results from this historically close fit

6. Discussion

Figure 2 shows the naive year-in-advance model data that was the basis for my earlier predictions. This naive model predicts September mean extent from a Gompertz curve representing the trend over previous years. Estimation data are the **NSIDC monthly mean extent** reports from September 1979 through September 2011.

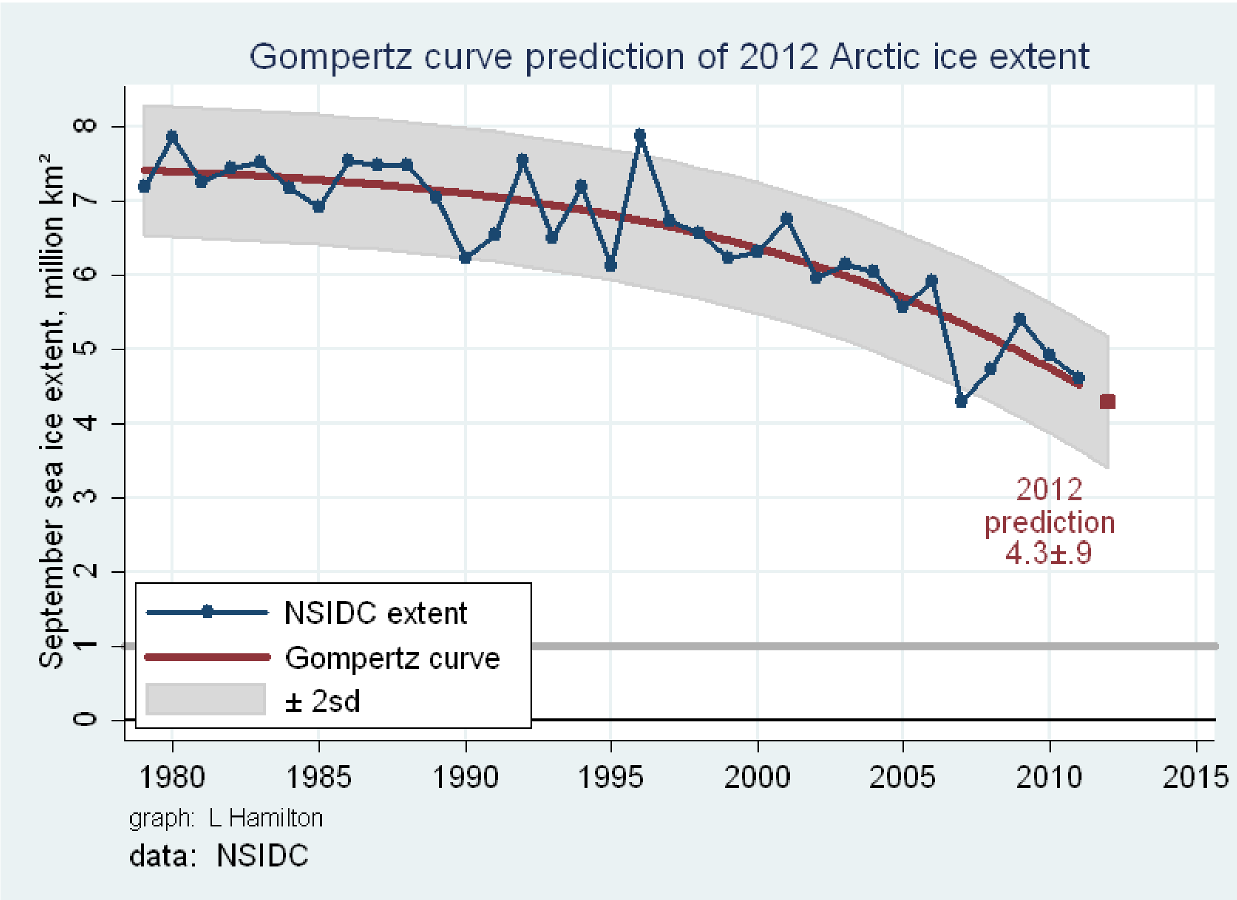

Figure 2

Parameters for the model are estimated via iterative least squares, using the **nl** procedure of Stata 12.1. Figure 2 also shows confidence bands calculated as the prediction plus or minus twice the standard deviation of the residuals.

In the command below, **gom3** specifies a 3-parameter Gompertz curve. ***extent*** refers to September mean NSIDC sea ice extent, in millions of km^2. ***year*** refers to the calendar year.

```
. nl gom3: extent year, nolog
(obs = 33)

      Source |       SS       df       MS
-------------+------------------------------         Number of obs =        33
       Model |  1425.43798     3  475.145994         R-squared     =    0.9957
    Residual |  6.15941312    30  .205313771         Adj R-squared =    0.9953
-------------+------------------------------         Root MSE      =  .4531156
       Total |   1431.5974    33  43.3817393         Res. dev.     =  38.25858

3-parameter Gompertz function, extent = b1*exp(-exp(-b2*(year - b3)))
------------------------------------------------------------------------------
      extent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b1 |   7.580278    .291652    25.99   0.000     6.984645    8.175911
         /b2 |  -.0995915   .0271646    -3.67   0.001     -.155069    -.044114
         /b3 |   2017.531   2.173212   928.36   0.000     2013.093    2021.969
------------------------------------------------------------------------------
```

The squared correlation between observed and predicted values (not shown) is $r^2 = .79$. There is no significant autocorrelation among the residuals, as tested by Ljung–Box *Q* statistics.

```
. predict resid, resid
. corrgram resid, lag(6)
                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1      -0.2135  -0.2138   1.6449  0.1996          -|                 -|
2      -0.0599  -0.1094   1.7787  0.4109           |                  |
3      -0.2277  -0.3000   3.7748  0.2868          -|                --|
4       0.0233  -0.1280   3.7963  0.4343           |                 -|
5       0.1641   0.1840   4.9073  0.4273           |-                 |-
6      -0.1815  -0.2761   6.3169  0.3886          -|                --|
```

This naive model was estimated at the end of the 2011 melt season. Most trend-line analyses of Arctic sea ice have used linear, quadratic, exponential or logistic models. The Gompertz curve appears preferable to these alternatives in several respects.

- It follows the observed pattern of gradually accelerating decline in the 1970s and 80s.
- The decline later steepens at an accelerating rate, as observed since the mid-2000s.
- Model predictions do not cross or exactly reach zero extent. Rather they approach this limit asymptotically.
- The asymmetrical-S shape bears a qualitative resemblance to results from much more elaborate physical models, such as those reported by the IPCC (2007); other sigmoid projections in Holland et al. (2006), Wang & Overland (2009), or Boe, Hall & Qu (2009).

Although out-of-sample extrapolation of this non-physical model is purely speculative, it is interesting to note the suggestion of extent falling below 1 million km^2 by 2025. More realistically, if we add Gaussian noise with the same standard deviation as past residuals to the projected future curve, we see behavior like the four examples in **Figure 3**.

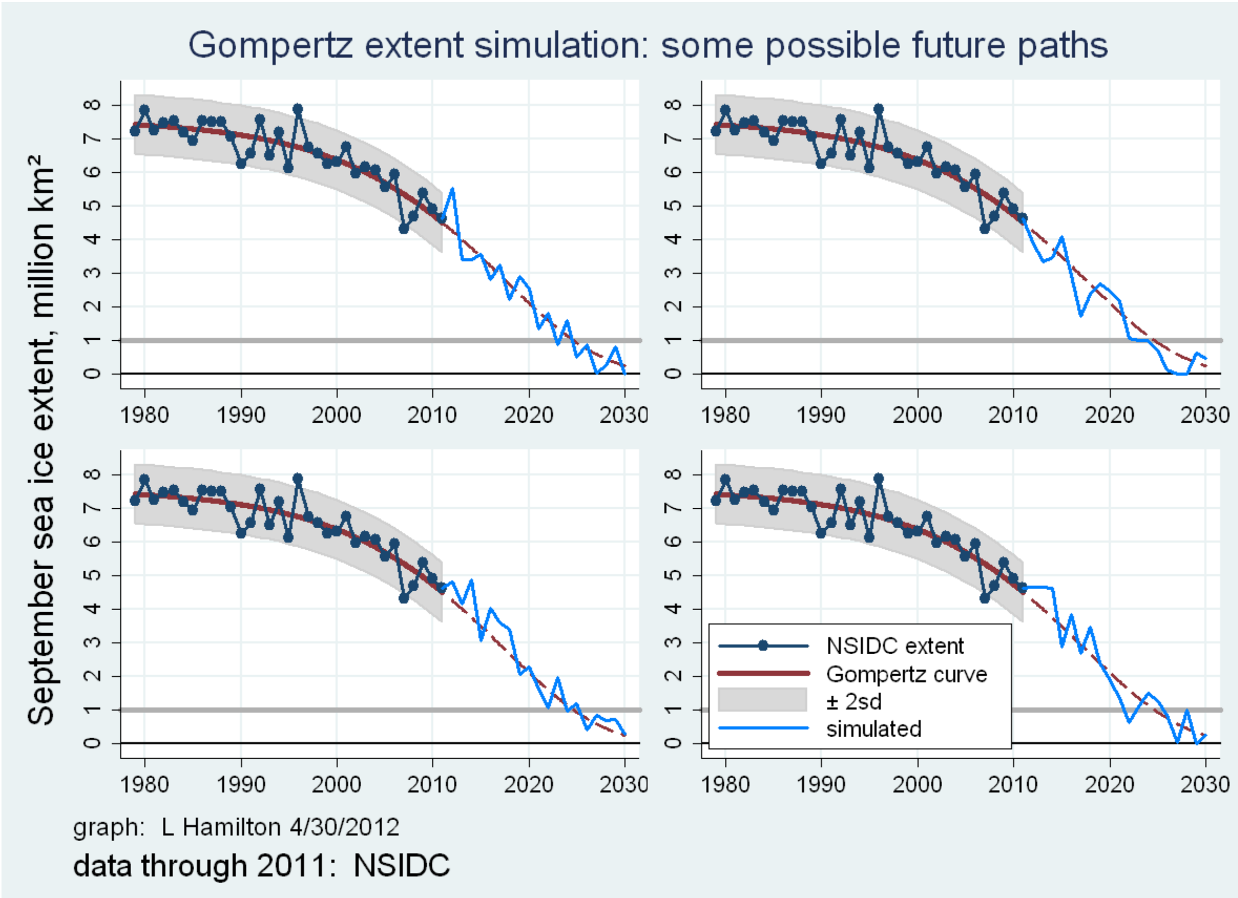

Figure 3

The same curve-fitting approach yields a year-in-advance September 2012 mean **NSIDC area** prediction of 3 million km^2, with confidence interval from 2.2 to 3.7 (**Figure 4**).

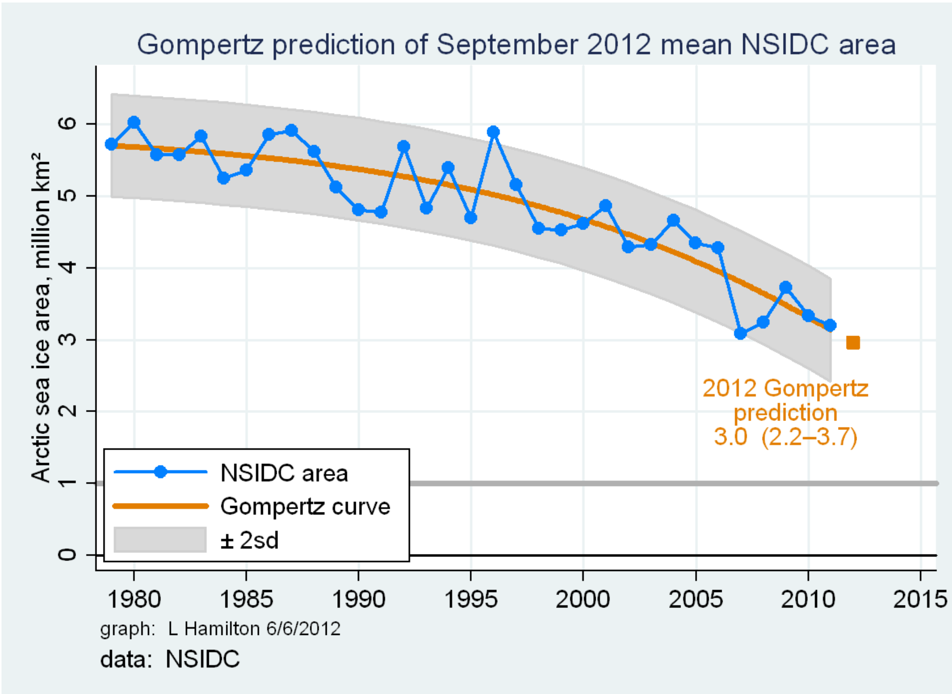

Figure 4

The year-in-advance **PIOMAS volume** prediction is 4 thousand km^3, with confidence interval from 2 to 5.9 (**Figure 5**).

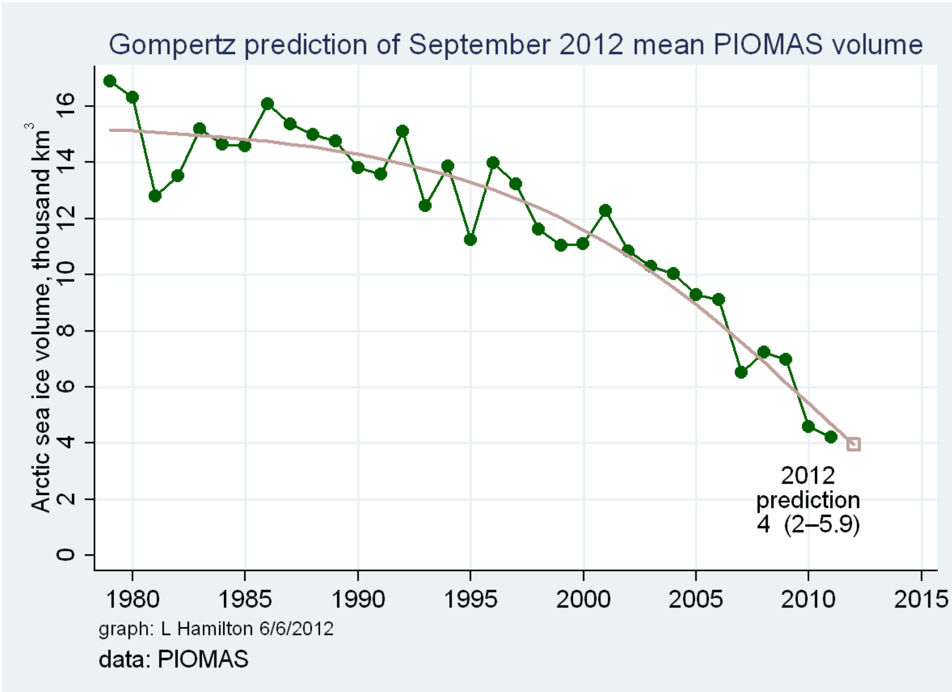

Figure 5

But all these are monthly means. What about daily values? They are often disdained too random to reasonably predict. I tried this out anyway, coming up with a minimum 1-day **Cryosphere Today extent** prediction of 2.7 million km^2, with confidence interval from 2.1 to 3.3 (**Figure 6**).

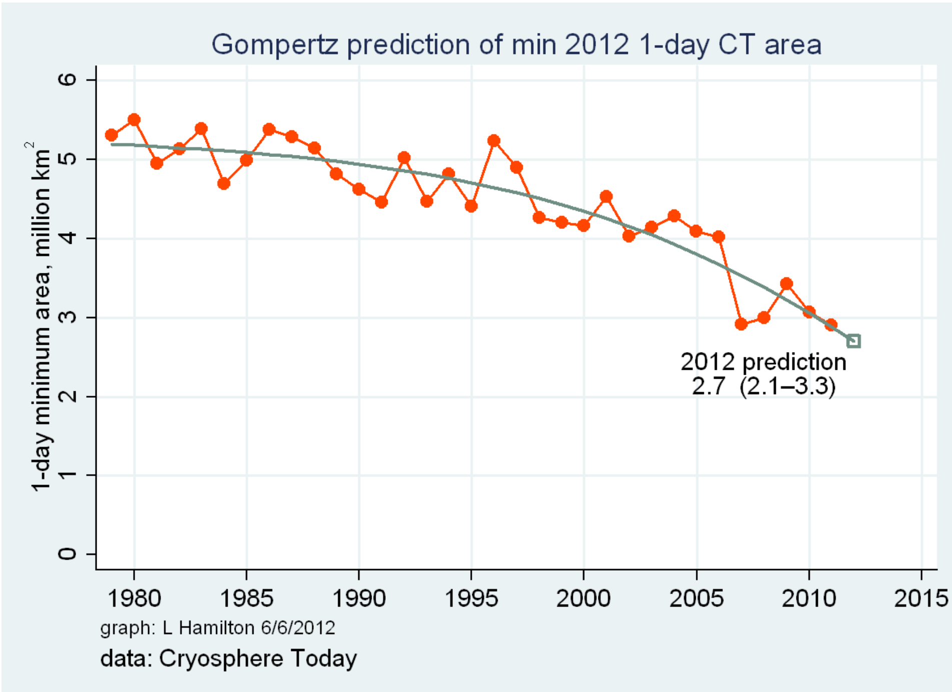

Figure 6

Are minimum daily values really more difficult to predict? The surprising answer is no, at least with respect to Cryosphere Today area. The Gompertz model for minimum daily area graphed above fits with a squared correlation of .984, and a residual standard deviation of .292 million km^2. A very similar model for CT September mean has a slightly lower r^2 (.95) and higher residual sd (.343).

The models in Figures 2–6 all base their predictions exclusively on information available in early October 2011. That is, they are year-in-advance predictions, appropriately compared with other year-in-advance prediction. From an elementary statistical perspective, however, we might guess that significantly better prediction of September 2012 ice extent could be achieved using data from the summer of 2012, instead of just through fall of 2011, as done in Figure 1. This model too is purely statistical, but could be appropriately compared with the performance of equally up-to-date physical models.