Pan Arctic Outlook

Christopher Randles

Public outlook, statistical method

1. Extent Projection 4.4m km^2

2. **Method** – multiple linear, single non-linear regression

   A gompertz fit of the NSIDC September extent figures is used as a starting point. Multiple linear regression is then used to predict the residual from the Gompertz fit. Two predictors have been used which are:

   a) The residual of the end of June Cryosphere Today area numbers at the end of June from a gompertz fit of those end of June area numbers.
   b) The residual of the end of June PIOMAS volume numbers at the end of June from a gompertz fit of those end of June volume numbers.

3. **Rationale**

   Several contributors have used multiple linear regression. This felt inadequate when there appears to be a curved shape that other contributors have used quadratic, exponential, logistic or gompertz fits to approximate.

   The predictors used in the multiple linear regression are 1. Area because of the direct implications for albedo feedback. 2. Volume because ice is more likely to disappear faster if there is less ice to melt. Testing showed that using the residual from a gompertz fit worked better for predicting the residual than using the raw area and volume numbers which is not surprising. These predictor variables are likely to work better using July data when that is available. Other predictor variables may well exist to further reduce the error.

   Hamilton's Contribution used a gompertz fit and yielded an estimate of 4.4m km^2. This prediction updates that prediction with two effects:
   Area at end of June of 7.28m km^2 is higher than the gompertz fit value of 7.14, suggesting that area has not declined by as much as expected by the gompertz fit. The multiple regression factor of 0.5236 applied to the residual of 0.14m suggests a prediction of 0.07m more than Hamilton's gompertz fit.

   The other factor is the PIOMAS volume. This year's end of June data of 12.261k km^3 is well below the Gompertz fit of the end of June volumes of 13.694k km^3. This suggests we will likely see more rapid decline in extent. The multiple regression factor of 0.1078 is much smaller than the factor for area suggesting that area is more important than volume. Multiplying by the residual of 1.43k km^3 gives a larger

0.15m reduction from Hamilton's gompertz fit. The multiple regression intercept figure is negligible as I am working with residuals meaning I am working with de-trended data. Hence the overall effect is for a 0.08m km^2 lower prediction than Hamilton's Gompertz fit prediction. The Gompertz fit of 4.44m km^2 is reduced to 4.36m km^2 but both of these get rounded to 4.4m km^2 to avoid suggesting too much precision. The method is appears to be predicting only just above the 2007 record low but considerable uncertainty remains.

I have not seen anyone attempting any sort of multiple non-linear regression and this approach which de-trends the extent, area and volume data in a non-linear manner prior to a multiple linear regression to predict the residual in the extent that we are trying to estimate appears to be novel in the context of SEARCH predictions.

4. Executive Summary

The data appears to have a curved shape which it appears advantageous to recognise and adapt multiple linear regression to predicting the residuals from the curved shape which has been approximated using a Gompertz fit. – See Hamilton's Contribution. This model yields an average September extent prediction of 4.4m km^2 with a 95% confidence interval in the region of +/- 1m (though RMSE is as low as 0.36m).

5. Estimate of Forecast Skill

A 95% confidence interval of +/- 1m is calculated though there are some indicators that this understates the uncertainty. This estimate is substantially higher than the inappropriately tuned RMSE figures of as low as 0.36m.

The RMSE of estimates reduces as follows:
Linear regression of September average extent =0.508m
Gompertz fit of September average extent = 0.438m

Gompertz fit then linear regression prediction of residual with CT area residual from gompertz fit = 0.372m

Gompertz fit then linear regression prediction of residual with PIOMAS volume residual from gompertz fit = 0.396m

Gompertz fit then multiple linear regression prediction of residual with both CT area and PIOMAS volume residuals from gompertz fits = 0.36m

Note however that these RMSE numbers are likely to underestimate the likely error as they have the advantage of the method being tuned with data that cannot be available at the time of making a true prediction.

Removing that advantage

Year	**Prediction**	**Actual**	**Error**
1991	6.940	6.55	-0.395
1992	7.000	7.55	0.550
1993	6.293	6.5	0.207
1994	6.885	7.18	0.295
1995	6.408	6.13	-0.278
1996	7.528	7.88	0.352
1997	6.793	6.74	-0.053
1998	6.775	6.56	-0.215
1999	6.666	6.24	-0.426
2000	6.440	6.32	-0.120
2001	6.638	6.75	0.112
2002	6.864	5.96	-0.904
2003	6.269	6.15	-0.119
2004	6.310	6.05	-0.260
2005	5.703	5.57	-0.133
2006	5.324	5.92	0.596
2007	5.148	4.3	-0.848
2008	5.026	4.68	-0.346
2009	4.872	5.36	0.488
2010	3.697	4.9	1.203

Average absolute error	0.395
RMSE without tuning to unavailable data	0.492

A 95% confidence interval is calculated at +/- 1m and only one year of 20 above has a larger error supporting that size for the confidence interval.

However, the average of the absolute errors for the first 10 year is only 0.29 whereas the average in the last 10 years is higher at 0.50. So there may be some growth in the expected size of errors and therefore a 95% credible interval may need to be higher than +/-1m.

In the format

	A	B	C	D	E	F
1	m_n	m_{n-1}	...	m_2	m_1	b
2	se_n	se_{n-1}	...	se_2	se_1	se_b
3	r_2	se_y				
4	F	d_f				
5	ss_{reg}	ss_{resid}				

The multiple regression factors and data are

Multiple Regression Factors - Area and Volume

0.107838	0.523576	-2.09026E-05
0.073398	0.206417	0.06655474
0.330561	0.37649	
7.159937	29	
2.02977	4.110604	

6. Review of formula arising from model for possible bias

As explained earlier, the full formula for average September 2011 extent can be expressed as:

=gompertz fit of 4.438 +0.5236*(area-7.14) + 0.1078*(volume-13.694)

Which can be simplified to = -0.776 + 0.5236 * area + 0.1078 *volume

If the end of June area and volume was as absurdly low as half the expected figures (say 3.5m km^2 and 7k km^3), we should be certain that the vast majority of the ice would melt by the beginning of September. The above formula would calculate to a September average extent of 1.8m km^2. The formula clearly calculates too much ice extent when it is taken outside of the ranges where it is hoped that the linear regression might work. This suggests that we might expect a non-linear response to the predictors used in the linear regression. This year it looks like the volume is well below the expected gompertz fit value. Thus trying to account for the expected non linear response to the volume predictor would seem to suggest that this method will predict too high a level for the September 2011 average extent. However the same could be said to apply even more to 2010 and that effect does not seem to have been observed. It could of course be there but hidden by random error but this would mean that the random error would have to be an even larger unprecedented size.