

Report on my paper for non-scientists: Part 2

Should We Worry About the Earth's Calculated Warming at 0.7°C Over Last the Last 100 Years
When the Observed Daily Variations Over the Last 161 Years Can Be as High as 24°C?

Darko Butina

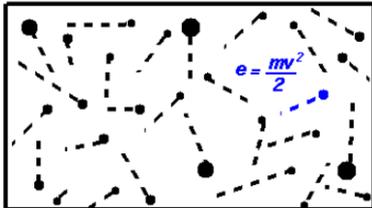
In Part 1 of my contribution I have discussed part of the paper which describes first step of data analysis known as 'get-to-know-your-data' step. The key features of that step are to established accuracy of the instrument used to generate data and the range of the data itself. The importance of knowing those two parameters cannot be emphasised enough since they pre-determine what information and knowledge can one gain from the data. In case of calibrated thermometer that has accuracy +/- 0.5C it means that anything within 1C difference in data has to be treated as 'no information' since it is within the instrumental error, while every variation in data that is larger than 1C can be treated as real. As I have shown in that report, daily fluctuation in the Armagh dataset varies between 10C and 24C and therefore those variations are real. In total contrast, all fluctuations in theoretical space of annual averages are within errors of thermometer and therefore it is impossible to extract any knowledge out of those numbers.

So let me start this second part in which I will quantify differences between the annual temperature patterns with a scheme that explains how thermometer works. Please note that this comes from NASA's engineers, specialists who actually know what they are doing in contrast to their colleagues in modelling sections. What thermometer is detecting is kinetic energy of the molecules that are surrounding it, and therefore thermometer reflects physical reality around it. In other words, data generated by thermometer reflect physical property called temperature of the molecules (99% made of N2 and O2 plus water) that are surrounding it:

 **Gas Temperature** 

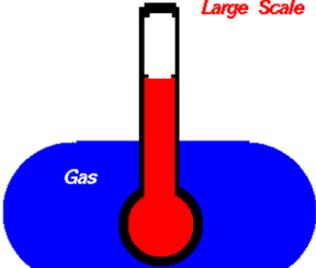
Small Scale $m = \text{mass}$ $v = \text{velocity}$
 $e = \text{kinetic energy}$

$e = \frac{mv^2}{2}$



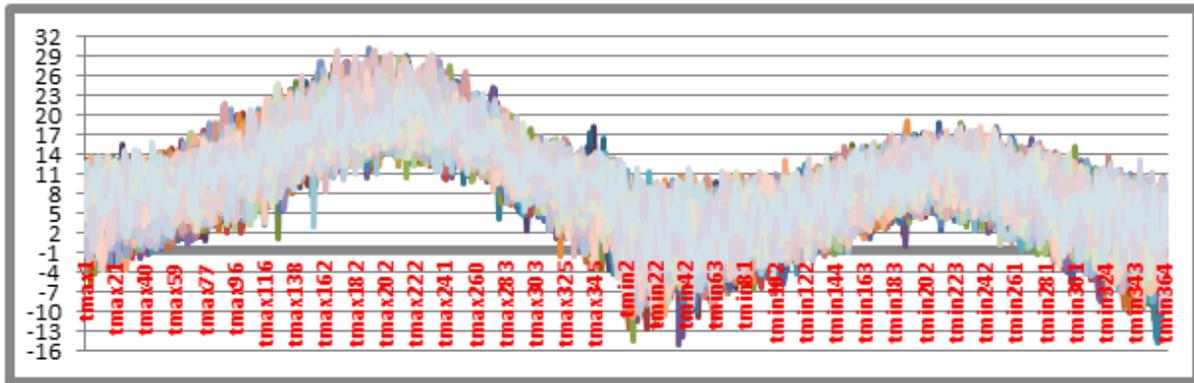
Temperature is a measure of the average kinetic energy of translation of the gas molecules.

Large Scale

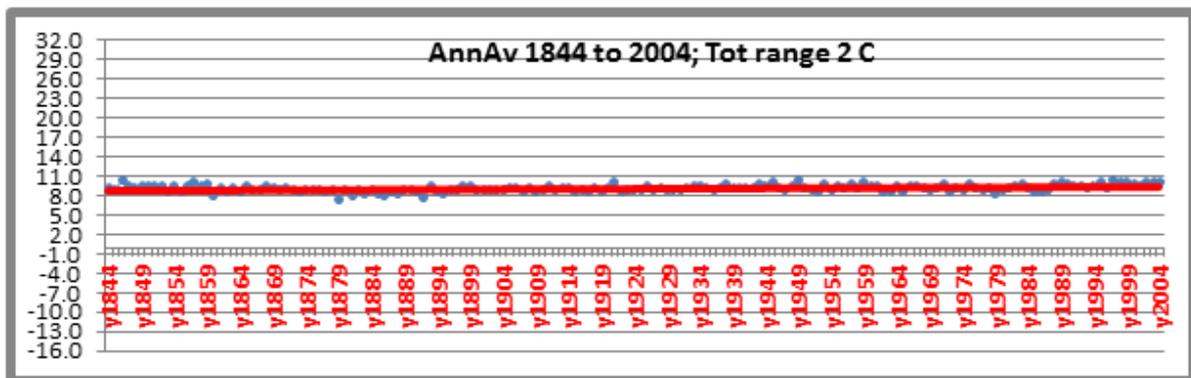


Objects in thermal equilibrium have the same temperature.

Let us now plot all of Armagh data in their annual fingerprints and compare them with annual averages that are obtain from them:



Graph 1. All years (1844-2004) in Armagh dataset, as original daily recordings, displayed on a single graph with total range between -16C and +32C



Graph 2. All years (1844-2004) in Armagh dataset, in annual averages (calculated) space with trend line in red

Please note that I am not using any of ‘Mike’s tricks’ in Graph 2 where Y-axis range is identical to the Y-axis range in Graph 1. Since Graph 2 is created by averaging data in Graph 1 it has to be displayed using the same temperature ranges to demonstrate what happens when 730-dimensional space is reduced to a single number by ‘averaging-to-death’ approach. BTW, I am not sure whether anyone has realised that not only a paper that analyse thermometer data has not been written by AGW community, but also not a single paper has been written that validates conversion of Graph 1 to Graph 2 – NOT A SINGLE PAPER! I have quite good idea, actually I am certain why that is the case but will let reader make his/her mind about that most unusual approach to inventing new proxy-thermometer without bothering to explain to wider scientific community validity of the whole process.

The main reason for displaying the two graphs above is to help me explain the main objective of my paper, which is to test whether the Hockey stick scenario of global warming, which was detected in theoretical space of annual averages, can be found in the physical reality of the Earth atmosphere, i.e. thermometer data. The whole concept of AGW hypothesis is based on idea that the calculated

numbers are real and thermometer data are not, while the opposite is true. Graph 1 is reality and Graph 2 is a failed attempt to use averages in order to represent reality.

The hockey stick scenario can be represented as two lines graph consisting of baseline and up-line:



The main problem we now have is to ‘translate’ 730-dimensional problem, as in Graph 1, into two-line problem without losing resolution of our 730-bit fingerprints. The solution can be found in scientific field of pattern recognition that deals with finding patterns in complex data, but without simplifying the original data. One of the standard ways is to calculate distance between two patterns and one of the golden standards is Euclidean distance, let’s call it EucDist:

$$\sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

There are 3 steps involved to calculate it: square difference between two datapoints, sum them up and take square root of that sum. The range of EucDist can be anywhere between ‘0’ when two patterns are identical and very large positive number – larger the number, more distant two patterns are. One feature of using EucDist in our case is that it is possible to translate that distance back to the temperature ranges by doing ‘back-calculating’. For example, when EucDist = 80.0 it means that an **average** difference between any two daily temperatures is 3.14C:

1. 80 comes from the square root of 6400
2. 6400 is the sum of differences squared across 649 datapoints: $6400/649=9.86$
3. 9.86 is an average **squared** difference between any two datapoints with the square root of 9.86 being 3.14
4. Therefore, when two annual temperature patterns are distant 80 in EucDist space, their baseline or normal daily ‘noise’ is 3.14C

Let me now introduce very briefly two algorithms that will be used, clustering algorithm dbclus, my own algorithm that I published in 1999 and since then has become one of the standards in field of similarity and diversity in space of chemical structures, and k Nearest Neighbours, or kNN, which is standard in fields of datamining and machine learning.

Basic principle of dbclus is to partition given dataset between clusters and singletons using ‘exclusion circles’ approach in which user gives a single instruction to the algorithm – the radius of that circle. Smaller the radius, tighter the clusters are. Let me give you a simple example to help you in visualising how dbclus works. Let us build matrix of distances between every planet in our solar system, where each planet’s fingerprints contain distance to all other planets. If we start with clustering run at EucDist=0, all planets will be labelled as singletons since they all have different grid

points in space. If we keep increasing the radius of the (similarity) circle, at one stage we will detect formation of the first clusters and would find cluster that has the Earth as centroid and only one member – the Moon. And if we keep increasing that radius to some very big number, all planets of our solar system would eventually merge into a single cluster with the Sun being cluster centroid and all planets cluster members. BTW, due to copyrights agreement with the publisher, I can only link my papers on my own website which will go live by mid-May where free PDF files will be available. My clustering algorithm has been published as ‘pseudo-code’ so any of you with programming skills can code in that algorithm in any language of your choice. Also, all the work involving dbclus and kNN was done on Linux-based laptop and both algorithms are written in C.

Let us now go back to hockey stick and work out how to test that hypothesis using similarity based clustering approach. For the hockey stick scenario to work you need two different sets of annual temperature patterns – one set of almost identical patterns which form the horizontal line and one set that is very different and form up-line. So if we run clustering run at EucDist=0 or very close to it, all the years between 1844 up to, say 1990, should be part of a single cluster, while 15 years between 1990 and 2004 should either form their own cluster(s) or most likely be detected as singletons. If the hockey stick scenario is real, youngest years MUST NOT be mixed with the oldest years:

Table 4. Running automated script at 5 EucDist increments indicates that all years are unique, i.e. singletons, up to EucDist of 80. The first cluster is formed at EucDist=81 containing only two (most similar) years, with 159 years remaining as singletons. ALL the years merge into a single cluster at EucDist=110 and no singleton remains.

EucDist	Number of clusters	Number of singletons
d-0 to d-80	0	161
d-81	1	159
d-85	5	142
d-90	9	94
d-95	7	48
d-100	6	16
d-105	3	3
d-109	2	1
d-110	1	0

The very first thing that becomes clear from Table [4] is that there are **no** two identical annual patterns in the Armagh dataset. The next things to notice is that up to EucDist of 80 all the annual patterns still remain as singletons, i.e. all the years are perceived to be unique with the minimum distance between any two pairs being at least 80. The first cluster is formed at EucDist=81 (d-81), consisting of only two years, 1844 and 1875. At EucDist 110, **all** the years have merged into a single cluster. Therefore, the overall profile of the dataset can be summarised as follows:

- All the years are unique up to EucDist of 80
- All the years are part of a single cluster, and therefore ‘similar’ at EucDist 110

Now we are in a position to quantify differences and similarities within the Armagh historical data.

The fact that any two years are distant by at least **80 in EucDist** space while remaining singletons, translates into minimum average variations in daily readings of **3.14C** between any two years in the database.

At the other extreme, all the years merge into a single cluster at **EucDist of 110**, and using the same back-calculation as has been done earlier for EucDist of 80, the average variation between daily readings of **4.32C** is obtained.

The first place to look for the hockey stick’s signal is at the run with EucDist=100 which partitioned Armagh data into 6 clusters and 16 singletons and to check whether those 16 singletons come from the youngest 16 years:

Table 5. Distribution of 16 singletons clustered at EucDist=100 distributed at 50 years intervals

1844-1900	y1881	y1892	y1895					
1900-1949	y1917	y1921	y1941	y1947	y1949			
1950-1989	y1953	y1955	y1956	y1969	y1970	y1971	y1973	y1987

As we can see, those 16 singletons come from three different 50-years periods, 3 in 1844-1900 period, 5 in 1900-1949 period and 8 in 1950-1989 period. So, hockey stick scenario cannot be detected in singletons.

What about clusters – are any ‘clean’ clusters there, containing only youngest years in the dataset?

Table 6. Distribution of the 1990-2004 period across four different clusters with each cluster containing years from 1800s mixed together with late 1900s

y1990	y1991	y1992	y1993	y1994	y1995	y1996	y1997	y1998	y1999
cl-3	cl-1	cl-0	cl-0	cl-0	cl-0	cl-0	cl-5	cl-0	cl-1
y2000	y2001	y2002	y2003	y2004					
cl-0	cl-0	cl-0	cl-5	cl-0					

No hockey stick could be found in clusters either! Years from 1990 to 2004 period have partitioned between 4 different clusters and each of those clusters was mixed with the oldest years in the set. Therefore the hockey stick hypothesis has to be rejected on bases of the clustering results.

Let me now introduce kNN algorithm which will give us even more information about the youngest years in dataset. Basic principle of kNN is very similar to my clustering algorithm but with one difference: dbclus can be seeing a un-biased view of your dataset where only similarity within a cluster drives the algorithm. kNN approach allows user to specify which datapoints are to be compared with which dataset. For example, to run the algorithm the following command is issued:

“kNN target.csv dataset.csv 100.00 3” which translates – run kNN on every datapoint in target.csv file against the dataset.csv file at EucDist=100.00 and find 3 nearest neighbours for each datapoint in the

target.csv file”. So in our case, we will find 3 most similar annual patterns in entire Armagh dataset for 15 youngest years in the dataset:

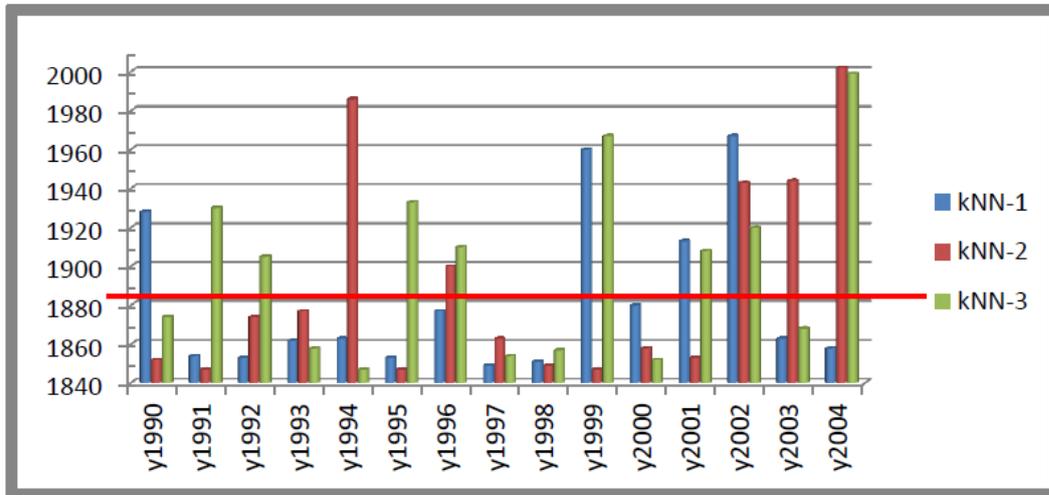


Figure 8. kNN was run at EucDist=100 and found the 3 most similar annual patterns for each of the 15 years in the 1990-2004 period. All the years have the most similar patterns way back in pre 1880s except 2002 which has the oldest analogue in 1920.

Let me pick few examples from Figure 8: year 1990 has the most similar annual patterns in years 1930, 1850 and 1880; supposedly the hottest year, 1998 is most similar to years 1850, 1848 and 1855, while 2004 is most similar to 1855, 2000 and 1998. So kNN approach not only confirms the clustering results, which it should since it uses the same distance calculation as dbclus, but it also identifies 3 most similar years to each of the 15 youngest years in Armagh. So, anyway you look at Armagh data, the same picture emerges: every single annual fingerprint is unique and different from any other; similarity between the years is very low; it is impossible to separate the oldest years from the youngest years and the magnitude of those differences in terms of temperatures are way outside the error levels of thermometer and therefore real. To put into context of hockey stick hypothesis – since we cannot separate oldest years from the youngest one in thermometer data it follows that whatever was causing daily variations in 1844 it is causing the same variations today. And that is not due to CO2 molecule.

Let us now ask a very valid question – is the methodology that I am using sensitive enough to detect some extreme events? First thing to bear in mind is that all that dbclus and kNN are doing is simply calculating distance between two patterns that are made of original readings – there is nothing inside those two bits of software that modify or adjust thermometer readings. Anyone can simply use two years from the Armagh data and calculate EucDist in excel and will come up with the same number that is reported in the paper, i.e. I am neither creating nor destroying hockey sticks inside the program, unlike some scientists whose names cannot be mentioned. While the primary objective of the cluster analysis and the main objective of the paper were to see whether hockey stick signal can be found in instrumental data, I have also look into the results to see whether any other unusual pattern can be found. One year that has ‘stubbornly’ refused to merge into the final cluster was year 1947, the same year that has been identified as ‘very unique’ in 6 different weather stations in UK, all at lower resolution than Armagh, either as monthly averages or Tmax/Tmin monthly averages. So what is so unusual about 1947? To do analysis I created two boundaries that define ‘normal’ ranges in statistical terms know as 2-sigma region and covers approximately 95% of the dataset and placed 1047 inside those two boundaries. Top of 2-sigma region is defined by adding 2 standard deviations to the mean

and bottom by taking away 2 standard deviation from the mean. So any datapoints that venture outside 2-sigma boundaries is considered as ‘extreme’:

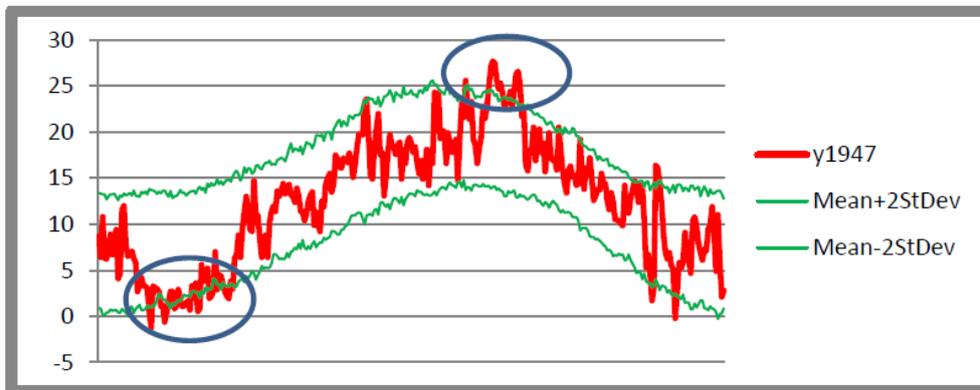


Figure 9. Year 1947 has most of February at or outside the 2 standard deviations ‘cold’ region, and two weeks of a very hot August, outside the 2 standard deviations ‘hot’ region. It also has excursion into 3sigma territory in November and December.

As we can see, 1947 has most of February in 3 sigma cold region and most of August in 3 sigma hot region illustrating the problem with using abstract terms like abnormally hot or cold year. So is 1947 extremely hot or extremely cold or overall average year?

Let me finish this report with a simple computational experiment to further demonstrate what is so horribly wrong with man-made global warming hypothesis. Let us take a single day-fingerprint, in this case Tmax207 and use the last year, 2004 as an artificial point where the global (local) warming starts by adding 0.1C to 2004, then another 0.1C to the previous value and continue that for ten years. So the last year is 1C hotter than its starting point, 2004. When you now display daily patterns for 2004+10 artificial years that have been continuously warming at 0.1C rate you can immediately see a drastic change in the overall profile of day-fingerprints:

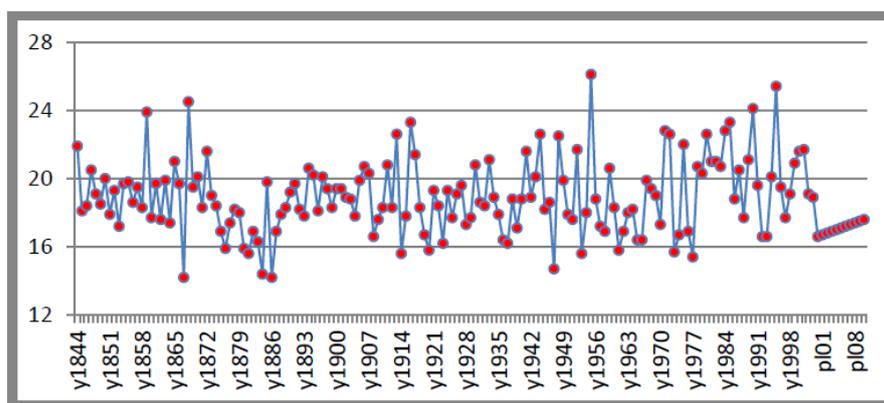


Figure 10. Tmax207 over 161 years plus 10 years of ‘warming’ after 2004 at 0.1C per year.

What would be worrying, if Figure 10 is based on real data is that a very small but continuous warming trend of only 0.1C per annum would completely change the whole system from being chaotic and with large fluctuation into a very ordered linear system with no fluctuations at all.

So let me now summarise the whole paper: not a single evidence of either alarming warming or cooling could be found at the Armagh (UK), Waterloo (Canada) or Melbourne (Australia); we do not understand temperature patterns of the past or the present and therefore we cannot predict temperature patterns of the future; all temperature patterns across the globe are unique and local. Every single aspect of man-made global warming is wrong and is based on large number of assumptions that cannot be made and arguments that cannot be validated: alarming trends in annual global temperatures are all within thermometer's error levels and therefore have no statistical meaning; not a single paper has been published that have found alarming trends in thermometer data; and not a single paper has been published validating reduction of 730-dimensional and time dependent space into a single number.

Let me finish this report on a lighter note and suggest very cheap way of detecting arrival of global warming, if it ever does come to visit the Erath: let us stop funding any future work on global warming and instead simply monitor and record accuracy of the next day temperature predictions instead! If you look at Figure 10, it becomes obvious that once the next day temperature predictions become 100% accurate it will be clear and unequivocal sign that the global warming has finally arrived to our Planet. The last paragraph can be represented by the following equations:

- chaotic system = no warming = 0% accuracy when predicting next day temperatures
- ordered-linear system = global warming = 100% accuracy when predicting next day temperatures

And let me leave you with two take-home messages:

- All knowledge is in instrumental data that can be validated and none in calculated data that can be validated only by yet another calculation
- We must listen to data and not force data to listen to us. As they say, if you torture data enough it will admit anything