SHOULD WE WORRY ABOUT THE EARTH CALCULATED WARMING AT 0.7°C OVER THE LAST 100 YEARS WHEN THE OBSERVED DAILY VARIATIONS OVER THE LAST 161 YEARS CAN BE AS HIGH AS 24°C?

Darko Butina*
Chemomine Consultancy, Letchworth Garden City, Hertfordshire, UK

Abstract

Detailed analysis of thermometer based daily readings for Armagh, UK, has shown that there is nothing unusual in annual temperature patterns for the 1990 to 2004 period, in either warming or cooling trends that would separate them from the oldest years. Moreover, the most similar annual temperature patterns to 1990-2004 period are found in the pre-1880s. Annual patterns, or annual fingerprints, have been created by using the original daily thermometer-based readings with the first 365 cells of the fingerprint filled with Tmax, daytime readings, followed by 365 cells filled by Tmin, night-time readings. The similarity between the two annual patterns was calculated using Euclidean Distance, while two similarity-based algorithms used were proprietary clustering algorithm, dbclus, and k-NearestNeighbours, kNN, an algorithm that is standard in pattern recognition and machine learning sciences. It has been established that it is impossible to declare one year either warmer or colder than any other using annual temperature patterns, since on average, each year is 50% of the time warmer and 50% of the time colder than any other year. Furthermore, no two years could be found with the identical annual patterns, with the average distance between all possible pairwise combinations being 100 in Euclidean Distance space, equivalent of 4°C variations in daily temperature differences. Based on those findings, the Hockey Stick hypothesis had to be rejected since there is no obvious baseline in the annual patterns, and therefore no abnormal patterns could be detected. Observed individual daily temperature variations over a 161 year period can range between 10°C and 24°C without any obvious trends or patterns amongst them. Very similar daily patterns have been observed at two different weather stations based on two different continents, North America and Australia, therefore indicating that patterns observed in Armagh reflects overall temperature patterns around the globe.

Keywords: Global Warming, Thermometer Based Temperature Readings, Annual Temperature Patterns, kNN

* E-mail address: darko.butina@chemomine.co.uk (Corresponding Author)

Darko Butina has 40 years of experience in drug discovery and modelling complex bio-chemical processes. His CV can be found at his consultancy’s home web page, www.chemomine.co.uk.
1. Introduction and Background

If we do not understand the past or the present, what hope have we to project into the future?

This paper is about trying to understand the past and recent patterns of daily temperature readings and trying to find any significant changes in those patterns that would separate temperature trends observed in the 1800s, 1900s and 2000s.

Let us start this paper with a very simple question: “How is global warming being measured”?

When a group of scientists state that “recent abnormal warming is happening, it is real and will cause climate meltdown by 2100” any experimental scientist would assume that that thermometer or thermometer based device has been used as an experimental evidence of that warming.

However, the current approach used by the climate community in general is to look for warming or cooling trends in annual averages space. Daily thermometer readings over 365 day period are averaged into a single number which is then declared as the ‘annual temperature’. In that way, a purely theoretical number, the mean of set of numbers, is given physical property of temperature and becomes a measure of warming or cooling from the previous year. Since there is no experimental way to actually measure ‘the annual temperature’, any argument put forward for either warming or cooling is equally valid or wrong, since it cannot be either proved or disapproved by any experimental measurements.

To make the matters worse, to get ‘the global annual temperature’ one has then to combine annual averages from the -50C Antarctica weather stations, to the +50C hottest deserts on the Earth, with the rest of world’s weather stations in between.

What is very difficult to understand for any experimental scientist is the need to invent something like ‘global temperature’ which has absolutely nothing to do with physical reality. The Earth is host to huge biodiversity, and each of those living forms is defined by diversity of local temperature patterns. So, what would make eminent sense is to perceive ‘the global temperature’ not as a single number but as a network of local temperature patterns where all individual temperature patterns count or none does. In other words, if there is a global trend, either warming or cooling, every single weather station should move in the same direction. If they do not, then the pattern cannot be called global.

The climate scientific community is ignoring a very important paper by Essex, McKitrick and Andresen in 2007 [1], EMA07, entitled “Does Global Temperature Exist?” and concludes that it does not. The key argument that EMA07 put forward is that the concept where one calculates an average value across a set of temperature readings and then assigns the physical property of temperature to that number is both theoretically and physically nonsensical and cannot and should not be done. In addition, the recently published book “Slaying the Sky Dragon” by Ball et al [2], clearly demonstrates that the idea of the Earth’s temperature being represented by a single number is absolute nonsense and that the use of Green Gas Theory by global warming proponents to explain that imaginary warming is totally wrong.

The last but not the least part of this introduction is to discuss briefly the main failures of defining an abnormally high global warming by the Anthropogenic Global Warming community. Without going into whom and what might be causing it, the basic AGW hypothesis is relatively very simple:
1. The Earth has been abnormally warming for the last 20(?) years in a way not seen in last 1000 years
2. The evidence is here, it is undisputed and if it goes on uncorrected, by the 2100 the Earth’s climate will collapse with unimaginable consequences for humankind

As discussed at the start, the only ‘evidence’ of this global warming is trend analysis in the theoretical world of annual averaging which has nothing to do with the physical reality around us. The next problem for the global warming community is to reconcile the following facts when modelling experimental data: “No theoretical model of some experimental data can be more accurate than the accuracy of those experiments”. Or, “The output of a mathematical operation can’t have more significant digits than the smallest number of significant digits in any of the inputs”.

The official body attached to UN, the Intergovernmental Panel for Climate Change (IPCC) published its 2007 report for the governments around the world stating that, “Global Temperature has risen by 0.7C in the last 100 years” [3]. That implies that the warming trend was 0.007C per annum. The most optimistic estimate for the accuracy of the thermometer back in the 1800s is about +/- 0.5C which makes trend of 0.007C per annum 117 times more accurate than the accuracy of the thermometer would allow. In other words, there was no warming in the last 100 years and anyhow, since that statement by IPCC uses non-existing temperature space, it simply cannot be made.

If one uses an abstract term like ‘abnormal’, as in abnormal warming, one has first to establish what the ‘normal’ is. The second step then is then to establish the level where the normal trends stop and the abnormal levels begin. The main objective of this paper will be the attempt to define what is ‘normal’ in terms of experimentally observed daily temperatures at the Armagh weather stations over a period covering 161 years from 1844 to 2004.

If one wanders why this paper is dealing with something so elementary which should have been done 20 years ago, the matter of fact is that the author could not find a single paper analysing daily temperature patterns for a one or more weather stations without the use of averaging.

2. Formatting Armagh Daily Data, Pattern Recognition and Similarity

This part describes preparation of the daily data and introduces the approaches that will be used for the analysis of the experimental data, thermometer based annual temperature patterns.

2.1. Armagh Daily Data 1844-2004

The original archived data is in the public domain and downloaded from the Armagh Observatory web site [4]. There are two files to download, one labelled as Tmax set and another as Tmin set, corresponding to daily and nightly temperature readings respectively. Each year is presented as an array or table consisting of 32 rows and 13 columns. Row 1 has all the column labels, with row2 to row32 containing temperature readings. Column 1 has day
labels, from day 1 to day 31, representing days of the month, while the next 12 columns hold the temperature readings for each month.

After re-formatting, each year is represented by 730-bits fingerprint or cells, with the first 365 cells having Tmax temperatures for January 1 through December 31, followed by 365 cells containing corresponding Tmin temperatures. ALL cells are perfectly aligned for each of the 161 years so that the column labelled as Tmax1 has all the daily readings for January 1 for all of 161 years. Leap years were ignored and the extra day in February omitted so that all the years had same number of cells. All the missing values have been dropped together with the rest of that column so that only experimental data was present in the final annual fingerprint and not contaminated with anything that has been calculated. The final annual fingerprint has in total 649 bits of data, with the first cell or column as the year’s label, followed by 649 cells containing Tmax readings followed by Tmin readings.

Table 1. Formatted annual and day fingerprints within X,Y coordinate system.

Annual Fingerprints are all in X-space, 1x730 bits long, from X=1 to X=730 (Tmax + Tmin data), while Day Fingerprints are in Y space, 1x161 bits long, from Y=1 to Y=161 (number of years)

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>...</th>
<th>X365</th>
<th>X366</th>
<th>X367</th>
<th>...</th>
<th>X730</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tmax1</td>
<td>Tmax2</td>
<td>.....</td>
<td>Tmax365</td>
<td>Tmin1</td>
<td>Tmin2</td>
<td>...</td>
<td>Tmin365</td>
</tr>
</tbody>
</table>

As we can see from Table [1], Annual Fingerprints, AnnFing, move in imaginary X-axis space and have 649 bits of information, original temperature readings, while Day Fingerprints, DayFing, move in Y-axis space and have 161 bits of information. Through this paper, DayFing will only use Tmax data for simplicity.

2.2. Calculating Distance between Two Annual Temperature Patterns

There are many different ways of calculating the distance between two patterns and throughout this paper the golden standard in pattern recognition sciences, the Euclidean Distance, is used:
\[
\sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}
\]

The first step in calculating Euclidean Distance, EucDist, is to take the difference, column-wise, between two datapoints and take the square of the difference. This is then repeated across the rest of 649 datapoints. The next step is to sum all those squared differences and finally to take the square root of that sum and obtain the distance between the two patterns in EucDist space. The minimum distance that can be obtained is “0” and that happens when two patterns have identical values in each cell. The larger the EucDist is, the further apart the two patterns are. It is very important to emphasise at this stage, that due to taking the square of the difference, we do not know whether the differences have the same or opposite signs.

2.3. Use of Similarity Based Algorithms, dbclus and knn

Both algorithms use EucDist as measure of similarity, but they answer slightly different questions. The basic question that one wants to ask is how similar or different two patterns are? For example, dbclus, my own clustering algorithm that became in the last twenty years one of the standards in the pharmaceutical industry worldwide [5], is very good in partitioning groups of similar patterns in their unique clusters and also to identify individual patterns that do not fit with any other pattern and label them as singletons.

The analysis of the results from the clustering will give user an indication whether there is some pattern in the years that are clustered together. For example, if the years clustered together are all in chronological order, it would indicate no change in their annual temperature patterns. The years clustered together could also identify some cyclic repeating patterns. However, if the most recent years co-cluster with the oldest years, and if those oldest years are 160 years apart, then those results would indicate the most recent annual patterns are no different than the oldest annual patterns.

The kNearestNighbours algorithm [6], kNN, is one of the standards in pattern recognition and machine learning fields and the question that can be asked is: “find the 3 most similar patterns, or 3NN, for the year 2004 in all of the 161 years in the Armagh database”. If there was an abnormal warming going on, say since 1990, one would expect that the nearest 3 temperature patterns to 2004 would be in the 3 preceding years, i.e. 2003, 2002 and 2001. However, if 1 or more of those 3NN go back to 1800’s, the hypothesis of abnormal warming in the last 20 years cannot be true.

3. Experimental Errors and Temperature Patterns Baseline

Before one starts to do modelling of experimental data, the very first stage is to evaluate the level of experimental errors that are due to the accuracy of the instrument used to obtain that data. This information will determine the accuracy that any model can claim, since the model cannot be more accurate then the underlying experimental data will allow it. It is estimated that the accuracy of the thermometers used in 1800’s and early 1900’s could be at the best +/- 0.5 C [7], so any annual trends of either warming or cooling should be adjusted accordingly.
The second most important part of the evaluation of the experimental data to be analysed is to estimate the total ranges of the data, i.e. their spread. That is an absolutely crucially important step to do, since it defines the shape of the distribution curve of the observed data and allows one to identify extreme, say 5%, of the data in a statistically significant way.

One of the main problems with analysis of daily temperature data is that the daily temperature patterns are part of the two concurrently occurring time cycles:

- The day-night 24 hour cycle and
- The 365 days annual cycle, the seasonal cycle.

In addition, there is a third time cycle, galactic year, lasting 220 million years for the Sun and its solar system to make one rotation around the Milky Way Galaxy. While that time scale makes everything connected with the thermometer based recordings in practical terms totally insignificant, it plays probably the most significant contribution to the temperature patterns observed as the most recent experimental evidence is suggesting, CERN [8] and Svensmark [9]. The ‘Cloud’ experiment was done at CERN following the original work by Svensmark that deals with the effect that the cosmic rays have on the formation of the clouds, described in must-read book by Svensmark and Calder [10].

### 3.1. Observed Variations in Dayfing, Tmax Only, Over 161 Years Period

One way to overcome day-night and seasonal variations is to analyse each DayFing and treat them as 323 individual units. Please note that we have started with 365 Tmax datapoints, but after removing the entire DayFing if it contained even a single missing reading, the resulting matrix has 161 columns with 323 rows.

The observed variations in daily temperature readings over 161 years within 2 standard deviations on either side of the mean, also known as a 2 sigma range, covers 95% of the data and can fluctuate anywhere between 8°C and 14°C, with two extremes coloured in green:

![Figure 1. Normal variations in °C, i.e. baseline for day-time readings in 2-sigma range (95% of the data) for Tmax data over 161 years period. Tmax361 has highest daily range at 14°C and Tmax261 lowest range at 8°C, two datapoints highlighted in green.](image-url)
Plotting daily ranges for ALL the data, including 5% of extreme values, over 161 years, one can see daily variations as high as 23.8°C for Tmax124 and 9.9°C for Tmax302, both highlighted in green:

Figure 2. Difference in °C between the measured day-time minimum and maximum Tmax for each day over 161 years period, covering 100% of data. Daytime variations for individual days in 161 years period vary between 9.9 °C for Tmax302, October 29, and 23.8 °C for Tmax124, May 4.

Tmax207, which is July 26, was analysed in more details. It was chosen as an example since it is one of the more ‘ordered’ or ‘stable’ day with the minimum variance, i.e. the smallest ratio between the mean and standard deviation, than any other day in DayFing space.

As can be seen from the table below, the largest swing observed in Tmax207 readings within 2 consecutive years happened in 1867 and 1868, where July 26 day-time temperature went from 14.2°C to 25.5°C, a swing of +10.3°C. Another two most distinctive swings were between 1953 and 1957 with the warming swing of 10.5°C followed by the cooling swing of 8.9°C, all within 4 year period. A similar pattern was then observed between 1993 and 1997:

Figure 3. Variations for Tmax207 over 161 years with highlighted largest swings occurring within 1 or two years apart.
Table 2 quantifies those swings.

**Table 2. Largest swings observed for Tmax207 that occurred within up to 4 consecutive years**

<table>
<thead>
<tr>
<th>Year</th>
<th>July 26 Tmax</th>
</tr>
</thead>
<tbody>
<tr>
<td>1867</td>
<td>14.2</td>
</tr>
<tr>
<td>1868</td>
<td>24.5</td>
</tr>
<tr>
<td>1953</td>
<td>15.6</td>
</tr>
<tr>
<td>1955</td>
<td>26.1</td>
</tr>
<tr>
<td>1957</td>
<td>17.2</td>
</tr>
<tr>
<td>1993</td>
<td>16.6</td>
</tr>
<tr>
<td>1995</td>
<td>25.4</td>
</tr>
<tr>
<td>1997</td>
<td>17.7</td>
</tr>
</tbody>
</table>

The overall behaviour of Tmax207 is mirrored by all other days of the year. The only difference is the range of the swings between different years, but the overall pattern is the same, i.e. there seems to be a lack of any patterns and the level of the temperature swings observed is as high as 24C.

This totally chaotic behaviour of the daily temperature patterns is demonstrated in the figure below comparing daily patterns of July 26 and December 21. Since the temperature ranges between the two days are on different scales, one being in the summer and one in the winter, all the December 21 data were ‘lifted’ by 12.3C to match the centre point of the July 26 temperature ranges.

![Figure 4. Tmax361 (Dec 21) (added 12.3C to each reading to shift the mean into July mean) vs Tmax207 (July 26) 10 years moving averaging.](image)

Note how often cooling/warming trends between July 26 and December 21 are going into the opposite direction over 161 year period. As mentioned earlier, July 26 was chosen as one
of the more ‘stable’ day while December 21 was chosen as the beginning of winter. The largest temperature swings for those two days are in the Table below:

**Table 3. The swings recorded by thermometer for the single day, Tmax361, over 161 years are up to 18 C, while for the more ‘stable’ day, like Tmax207, it is close to 12 C.**

<table>
<thead>
<tr>
<th>Recorded in C</th>
<th>Tmax361 range (Dec 21)</th>
<th>Tmax207 range (July 26)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>-4.9</td>
<td>14.2</td>
</tr>
<tr>
<td>Max</td>
<td>13.1</td>
<td>26.1</td>
</tr>
<tr>
<td>Tot Diff</td>
<td>18</td>
<td>11.9</td>
</tr>
</tbody>
</table>

The following summarises daily fluctuations over 161 years of thermometer based readings:

1. Daytime variations for 95% of the data over 161 years are between 8 and 14 C around the corresponding daily mean, i.e. two standard deviations or 2 sigma, Figure [2]
2. There seems to be no obvious pattern to those variations, Figure [4], in terms of which part of the year might be heating or cooling
3. Since one cannot understand those patterns, they can only be described as a chaotic system
4. The observed ranges are way above the +/- 0.5 C sensitivities of the thermometer and therefore they are real

### 3.2. Datamining Annual Temperature Patterns, AnnFing

In this section, the comparison is done row-wise, i.e. comparing corresponding daily data between the two annual temperature patterns in AnnFing space.

The figure below is the result of taking the difference between all of the 649 datapoints between years 2004, the youngest year, and 1844, 160 years older:

![Figure 5. Difference between daily temperature readings of year 2004, the youngest year, and 1844, the oldest year. Note that the difference between two year’s daily readings can switch the sign from “+” to “-” on almost daily bases.](image-url)
What immediately becomes obvious is that there is no clear cut difference between the annual temperature patterns that are 160 years apart. 1844 is on 250 occasions hotter than 2004, while on 391 occasions colder. As an example, if we zoom into January variations between the two years, one can see that 9 switch-overs occurred between the two years, making one year warmer for few a days and then colder for few days within +/- 8C ranges, 16C in total:

![Figure 6. Nine switch-overs in January between 2004 and 1884 ranging between -8C and +8C.]

To determine whether this bi-polar behaviour between 2004 and 1844 is a rare event or part of the ‘normal’ behaviour, a new program has been written (in C-programming language) to calculate on how many occasions one year is warmer or colder than any other year. There are 161 x 160 possible combinations, in total 25760 calculations to perform.

It turns out that on average, each year is half the time warmer and half the time colder than any other year. The Figure below shows results for 1844 against all other 160 years with the Y-axis indicating the number of days that 1844 is warmer or colder than its counterpart:

![Figure 7. Number of times that 1844 is hotter/colder against any other year over 161 years period.]

It is also very important to emphasise, that there are as many years in 1800’s that have over 50% of the warmer days than the years in late 1990s as there are the other way around.
3.3. Quantifying Differences in Annfing Patterns

In this part of the paper two similarity-based algorithms, dbclus and kNN, will be used to further quantify similarity or differences in annual temperature patterns.

Both algorithms use EucDist to calculate similarity levels between two or more annual temperature patterns with the main difference between the two algorithms being:

- Clustering algorithm, dbclus, partitions the whole of the given dataset into clusters and singletons. Similarity within the clusters, defined by the user, is the only force driving the algorithm. For example “dbclus Armagh.csv 100” would cluster the whole of the Armagh database at EucDist=100 and report back number of clusters, membership and cluster centroid for each cluster plus the years that do not belong to any clusters, i.e. unique years, labelled as singletons

- kNN algorithm finds one, k=1, or n, k=n, most similar annual temperature patterns to one or more user specified target annual patterns. For example, if we want to know what are 3 of the most similar annual temperature patterns to year 1998 in the Armagh dataset at the given EucDist, one would type: “kNN 1998.csv Armagh.csv 100 3”. The algorithm would then compare the annual pattern of the year 1998 to each of the annual patterns in the Armagh database at EucDist=100 and report back 3 of the most similar

If, for example, there is something unique in the 1990-2004 period, both methods should find those annual patterns.

As discussed in the experimental section, the annual patterns consist of 649 datapoints per year, i.e. 649-bits fingerprint, starting with Tmax1 to Tmax365 (daytime readings) and followed by Tmin1 to Tmin365 (night-time readings). Euclidean Distance, EucDist, is used to calculate the distance between the two annual patterns, see formulae in the Experimental section.

The use of EucDist has one particular feature that is very useful in its application – the ability to ‘translate’ EucDist number to the average daily fluctuations into degrees of C.

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

It is important to emphasise that each and every datapoint is treated equally and to remind the reader that all the datapoints used are part of the original thermometer readings, i.e. the software used in this analysis does not adjust or modify any datapoint.
3.4. Clustering Armagh Annving Using dbclus

Use of dbclus could be seen as an unbiased assessment of the given dataset in terms of its overall similarity profile. To run the software all the user needs to specify is the distance between the cluster centroid and its members. For example “dbclus file.csv 0” will run software on file.csv at EucDist = 0. EucDist can be minimum 0, when two patterns are identical at every datapoint, to a very large positive number. The larger the number, the more distant the two patterns are. Therefore, running dbclus at EucDist=0 one would identify all the years that have identical annual patterns, i.e. with all the daily readings being the same.

The software output reports how many clusters and singletons are found and also gives the full list of cluster members and singletons. Running dbclus on the Armagh dataset from EucDist=0 to EucDist=120 in 5 EucDist increments the following table was generated:

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.

<table>
<thead>
<tr>
<th>EucDist</th>
<th>Number of clusters</th>
<th>Number of singletons</th>
</tr>
</thead>
<tbody>
<tr>
<td>d-0 to d-80</td>
<td>0</td>
<td>161</td>
</tr>
<tr>
<td>d-81</td>
<td>1</td>
<td>159</td>
</tr>
<tr>
<td>d-85</td>
<td>5</td>
<td>142</td>
</tr>
<tr>
<td>d-90</td>
<td>9</td>
<td>94</td>
</tr>
<tr>
<td>d-95</td>
<td>7</td>
<td>48</td>
</tr>
<tr>
<td>d-100</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>d-105</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>d-109</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>d-110</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

The original algorithm was designed to deal with up-to several millions of 2048-bits fingerprint within the few hours of computing time. Running a single run for the database consisting of 161 of 649-bits fingerprints was for all practical purposes an instantaneous process.

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, consisting of only two years (number of total singletons reduced by two, from 161 to 159) 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 3.14°C 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 4.32°C is obtained.

So when the average difference between daily data is 3.14°C all the years are seen as unique and diverse, while 1.18°C ‘later’, at 4.32°C, all the years are similar and part of the single cluster. This type of behaviour points to a very low level of identical datapoints in the annual fingerprint, the point that will be discussed in more detail later.

What other information could one obtain from the clustering run?

One could test whether the hockey stick scenario hypothesis is real or just a failed hypothesis. The main argument put forward by Mann et al in 1989 [11] requires the existence of some baseline followed by abnormally high warming in the most recent years. Table [4] indicates that running dbclus at EucDist 100 would produce a baseline with 145 years distributed between 6 clusters with 16 singletons still remaining. For the hockey stick scenario to be real, those 16 singletons should correspond to the youngest 16 years, i.e. years 1989-2004 which are very different from the rest of the years.

The table below lists 16 years labelled as singletons, i.e. unique years, with the youngest year been 1987 and oldest 1881. So no obvious trends, in terms of the last 16 years being all singletons, and the hockey stick scenario must be rejected using singletons argument discussed above:

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

<table>
<thead>
<tr>
<th>1844-1900</th>
<th>1900-1949</th>
<th>1950-1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1881</td>
<td>y1917</td>
<td>y1953</td>
</tr>
<tr>
<td>y1892</td>
<td>y1921</td>
<td>y1955</td>
</tr>
<tr>
<td>y1895</td>
<td>y1941</td>
<td>y1956</td>
</tr>
<tr>
<td></td>
<td>y1947</td>
<td>y1969</td>
</tr>
<tr>
<td></td>
<td>y1949</td>
<td>y1970</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y1971</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y1973</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y1987</td>
</tr>
</tbody>
</table>

Next, each year from the 1990-2004 period was traced through 6 different clusters and summarised in the Table 6.

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cl-3</td>
<td>cl-1</td>
<td>cl-0</td>
<td>cl-0</td>
<td>cl-0</td>
<td>cl-0</td>
<td>cl-0</td>
<td>cl-0</td>
<td>cl-0</td>
<td>cl-0</td>
</tr>
<tr>
<td>cl-0</td>
<td>cl-0</td>
<td>cl-0</td>
<td>cl-5</td>
<td>cl-0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Not only that the last 15 years were distributed across four different clusters, but each of those clusters has had the oldest years on record mixed-in within the same clusters.
Since all the years from the 1990-2004 period are part of the clusters that contain the oldest years as well as the youngest one, the hockey stick scenario therefore must be rejected using the cluster membership argument.

3.5. Analysis of the Armagh Daily Data Using k Nearest Neighbours Algorithm

The overall patterns detected by dbclus are confirmed and further quantified by the use of the kNN approach.

For each year in the 1990-2004 period, the 3 Nearest Neighbours were calculated, and all years but 2002, had at least 1NN in the pre-1880 period. For example, supposedly the hottest year globally, 1998 had all 3NN in 1850’s. Year 2002 has one of the most similar years 82 years back in the past, 1920, Figure 11:

![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.](https://example.com/figure8.png)

Based on the analysis of the Armagh data using kNN one would have to reject the Hockey Stick scenario. In addition, all experimental data points to a status quo in 161 years of annual temperature patterns in Armagh.

3.6. How Many Identical Readings Are There in Tmax and Annual Fingerprints?

Three consecutive Tmax readings have been extracted from the DayFing, Tmax1, 2 and 3 and when the difference between the two readings is equal to $0 \pm 0.5 \degree C$ (to account for thermometer accuracy) those readings are labelled as identical, Table [7]. Tmax1 and 2 were identical on 44 occasions, 27%, Tmax 1 and 3 on 43 occasions, 27%, while Tmax2 and 3 on 24 occasions, 15%. However, on only 7 out of 161 years, all three Tmax readings had the
same readings, 4% of time. The years that all three had the same readings were 1856, 1868, 1881, 1890, 1917, 1930 and 1964.

| Table 7. Number of identical temperature readings between 3 daily readings and 3 annual patterns |
|---------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| **DayFing**                     | **Tmax1-2-3**                 | **1vs2**                        | **1vs3**                        | **2vs3**                        | **Common to all 3**               |
| _Tmax_                           |                               | 44/161 (27%)                    | 43/161 (27%)                    | 24/161 (15%)                    | 7/161 (4%)                        |
| **AnnFing**                     | **y2004-03-02**               | 33/323 (10%)                    | 50/323 (15%)                    | 48/323 (15%)                    | 7/323 (2%)                        |

Identical experiment was performed, but this time in AnnAver fingerprints, with years 2004, 2003 and 2002. Years 2004 and 2003 had identical readings (in Tmax space only) on 33 occasions, 10%, 2004 and 2002 on 50 occasions, 15%, while 2003 and 2002 were identical on 48 occasions, 15% of time. All three years had identical Tmax reading on only 7 occasions, 2%. The identical daily readings were observed on Tmax4, 192, 241, 294, 299, 317 and 336.

What this exercise is telling us, that whether one looks for any patterns in either DayFing or AnnFing, they are not there. So, each day of the year has unique historical pattern over 161 years of measurements and should be treated as an independent unit. It is the same for the annual fingerprints, where for all practical purposes each year have its own individual pattern and each year should be treated as a single entity. As it can be seen from the table above, two different pairs of years might have the same similarity in EucDist space, but for very different reasons.

Looking for trends in weekly averages over period of years, in DayFing space, or ‘smoothing’ annual averages in 10, 20 or 30 years batches, which is quite a common practice in the climate community, should not be attempted. As any experienced pattern recognition specialist will tell you, merging together patterns that are not understood can only lead to erroneous conclusions and should not be done. Bearing in mind that the oldest dataset worldwide, the Central England Temperature set, is based on monthly averages going back to 1600’s, highlights the problem that the climate community has in accessing good quality thermometer based data!

### 3.7. Summary of What Is ‘Normal’ for the Armagh Weather Station:

**The Facts so Far**

- The lowest temperature on record is -15.1°C recorded on February 7 1895
- The highest temperature on record is +30.3°C recorded on July 10 1934
- Total range of 45.4°C
- Total range for 1844 recorded is 29.4°C, between -4.6°C and 24.8°C
- Total range for 2004 recorded is 29.1°C, between -5.1°C and 24.0°C
- Largest daytime variations for an individual day over 161 years observed is 23.8°C for May 4
- Smallest daytime variations for an individual day over 161 years observed is 9.9°C for Oct 29
- On average, each year is 50% of the time warmer and 50% colder than any other year

Those variations are real; they are way outside the thermometer errors and any abnormality in temperature patterns must be outside that ‘normal’ noise in statistically significant level.

So the question now is, could we detect any abnormality in daily temperature readings bearing in mind the chaotic behaviour of the system?

The two examples that follow, one real and one hypothetical, would tend to indicate that it is possible to detect abnormal patterns in the experimental data if they are there.

3.8. What Is so Unique about 1947 in Armagh?

The very last singleton to merge into a single cluster at EucDist = 110 was the year 1947. It follows therefore, that there are some extreme features in that year that make it the most unique/different from any other year. What are those features? To visually inspect the difference between 1947 and the rest of the years, the mean and the standard deviation has been calculated for each day across 161 years and 1947 placed between the hot and the cold boundaries of the mean +/- 2 standard deviations:

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.

It is worthwhile reminding the reader that the mean +/- 2 standard deviations covers 95% of the data, also known as 2-sigma range, and that any datapoint that crosses into +/- 3 standard deviations range will belong to 2.5% of the most extreme datapoints of the distribution range on either side of the mean. As we can see, 1947 has most of February around or just below the cold 2 sigma boundary, and also has one of the hottest couple of weeks in August, outside the hot 2 sigma boundary.
The fact that 1947 had one of the coldest Februaries and also one of the hottest Augusts demonstrates problems with the use of abstract terms like “the hottest year”. What is the best description for 1947: hottest, coldest or overall neutral year?

3.9. Is the Armagh Weather Station Very Unique or Does It Represent the Global Pattern?

Two weather stations on two different continents were sampled, Waterloo in Canada [12] and Melbourne in Australia [13] for years 2009 and 1998, and the following data obtained:

Table 8. Summary for daily data, Tmax, for years 2009 and 1998 at Waterloo, Canada and Melbourne Australia

<table>
<thead>
<tr>
<th>2009 vs 1998 (Tmax)</th>
<th>Warmer</th>
<th>Colder</th>
<th>EucDist</th>
<th>Daily Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterloo, Canada</td>
<td>144(37%)</td>
<td>193(63%)</td>
<td>118</td>
<td>6.7C</td>
</tr>
<tr>
<td>Melbourne, Australia</td>
<td>219(60%)</td>
<td>146(40%)</td>
<td>123</td>
<td>6.5C</td>
</tr>
</tbody>
</table>

The overall profiles for years 2009 and 1998 at two different Continents, Canada and Australia, follow the general patterns observed at Armagh in UK:

- EucDist in Tmax space are around 120
- Average daily variations calculated from the EucDist are 6.7C and 6.5C respectively
- Bi-polar behaviour is there, with 2009 been on 37% of time warmer and 63% of time colder than 1998 at Waterloo, while 60% warmer and 40% colder at the Melbourne weather station

3.10. Simulation of Warming at 0.1C Increments Across Daily Temperatures, Tmax Only

Let us design a computational experiment where we can make one year ‘truly’ warmer than the previous year by a fixed amount in °C. We start with the last year in the Armagh dataset, 2004, and add to each Tmax temperature reading 0.1C (Tmax data only). We then continue to add 0.1 C to each following year for the next 10 years. In this simulation we are trying to see the changes that would occur to the daily temperature profiles and how would it reflect on the overall similarity based profiles between the years.

If we take Tmax207 as an example and graph its temperature profile over period of 161 years plus 10 artificially created ones, we are beginning to see the major changes in daily patterns:

It becomes obvious that the experimentally observed chaotic patterns change into an ordered system when one applies systematic warming trends of only 0.1C per year.
Therefore, if the systematic warming of as little as 0.1C has started, the main worry would be not about the amount of warming but about the complete change in temperature patterns. The whole system would be changing from a chaotic system to a very ordered one and that would really be alarming.

Every single approach that this paper has used so far to compare annual and daily patterns would be detecting drastic changes in the annual temperature patterns behaviour.

For example, the Euclidean distance between 2004 and 2003 in Tmax space is 64.6, but the distance between 2004 and “2004-plus-0.1C” only 1.8:

Table 9. Changes in temperature patterns
for artificially created ‘continuous’ warming at 0.1C per annum

<table>
<thead>
<tr>
<th>Tmax only</th>
<th>EucDist</th>
<th>Daily Variations in C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-vs-2003</td>
<td>64.6</td>
<td>3.6</td>
</tr>
<tr>
<td>2004-vs-2004+0.1</td>
<td>1.8</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Figure 11. Number of years in clusters at 5 EucDist increments. All the years cluster between EucDist 80 and 110, as before, except 2004 which co-clusters with 10 artificially made years in a single cluster at EucDist 1.5.
Running dbclus on the modified Armagh dataset, shows identical behaviour for 1844-2003 annual patterns seen in the experimental data, but detects a new single cluster that contains 11 years at 1.5 EucDist and contains 2004 plus 10 artificially created years:

kNN algorithm shows that the last artificially made year has 10 NN in descending chronological order and a simple graph for January between 2004 and “2004-plus-1C” would show the warming patterns that if real, would identified unambiguous warming trends happening.

Figure 12. January Tmax data for year 2004 and “2004-plus-1C”. Note no cross-over lines.

4. Discussion and Conclusions

The only way to unequivocally declare any year as either the hottest or the coldest would be if that year has every single day either hotter or colder than any other year on record, as measured by thermometer.

However, it has been demonstrated that it is impossible to declare one year warmer or colder than another using thermometer based daily data due to the chaotic behaviour of the air-temperatures. The overall pattern is that on average, every year is on 50% occasions warmer and on 50% occasions colder than any other year.

The current practice in the climate community is to look for the trends of warming or cooling in a purely theoretical space of annual averages and global temperatures, where a single number obtained by averaging all daily temperatures from the coldest to the hottest places on the Earth is assigned the physical property of temperature and used to represent the Global Temperature for a given year. Since the global temperature cannot be measured by any thermometer-based device, this cannot be proven to be either correct or wrong, and therefore cannot be used as evidence of anything. Global temperature as a single number has nothing to do with the physical reality space where the Earth can be seen as a very complex
network of millions of local temperature patterns and where each local temperature pattern should be treated separately.

Two similarity based algorithms used in this paper, dbclus and kNN, both clearly show that there is nothing unique in the annual patterns of 1990-2004 period that would separate the youngest 15 years from any other 15 year period in last 161 years of historical data. It was also shown that the most similar annual patterns for the youngest years have been found among the oldest years, going back to pre-1880s.

Analysis of the individual day’s patterns over a 161 years’ period clearly indicates chaotic behaviour of daily temperatures, with swings up to 24°C observed, and therefore making claims that the warming of 0.7°C over the last 100 years is alarming, as claimed by IPCC, is nonsensical. A trend of that magnitude would imply annual warming at 0.007°C which would be 117 times more accurate than the accuracy of the thermometer used.

The results obtained by dbclus and kNN clearly show that the hockey stick hypothesis cannot be detected in the experimental data and therefore the hypothesis must be rejected.

The artificially generated 10 years of warming, where each day of a year is warmer by 0.1°C than the previous year, could be detected very easily by the computational methods used in this paper. The main reason for the easy detection has nothing to do with the amount of warming that has been added to each daily reading, but everything to do with complete change in daily and annual temperature patterns from the chaotic system to a very ordered system. While the current forecast of the local temperatures works to various success for 2-3 days in the future and cannot be used for more than 10 days ahead, the continuous warming trends, if real, would allow very easy and accurate future forecasting.

The general patterns observed at the Armagh (UK) weather station have been also observed at the weather stations on two different continents, Waterloo (Canada, North America) and Melbourne (Australia). Those results would strongly indicate that the chaotic behaviour of annual and daily patterns observed at Armagh is the norm across the globe and not an isolated behaviour of the Armagh weather station.

The author would expect that the only difference between Armagh and any other weather stations across the globe would be in the level of the daily temperature variation, where the levels of those variations would reflect the location of the weather station, while the overall patterns would remain chaotic.

Acknowledgments

I acknowledge the weather stations at Armagh (UK), Waterloo (Canada) and Melbourne (Australia) for free download of their daily temperature data.

I would also like to thank my wife Judy for her help with this manuscript and her willingness to learn about the huge problems associated with the global warming issues, the world so different from her familiar territory of arts and design.

References


