International Journal of Chemical Modeling ISSN: 1941-3955
Volume 9, Number 1

LOOKING FOR PATTERNS IN EXTREME COLD AND HOT PERIODS ACROSS THE USA IN DAILY TMAX TEMPERATURES FOR THE LAST 125 YEARS USING 2 STANDARD DEVIATIONS ON EITHER SIDE OF THE MEAN AS REFERENCE POINTS

Darko Butina*

ABSTRACT

This paper has applied one of the gold standards in statistics, the normal distribution bell curve, to identify extreme cold and hot daily temperatures across the USA since the late 1800s. The normal distribution curve was used to design a classification protocol in which datapoints were partitioned between three classes, an extreme cold class, labelled as e. cold which has the z-scores below -2.0; an extreme hot class, labelled as e.hot, which has z-scores above 2.0; and the major class, labelled as normal with z-scores within +/- 2 standard deviations from the mean. Statistical analysis has been done on the datasets from 14 weather stations at different geographical locations representing East, West and South coast and also continental USA. The datasets were downloaded, free of charge, from the NCDC/NOAA global daily station data that covers 23,717 different geographical locations across the globe using KNMI Climate Explorer to access the data. Systematic statistical analysis was performed on over 570,000 daily tmax original thermometer readings, ranging from temperatures as low as -36.1C and as high as 53.9C, with the total range of 90.0C.

The main conclusion of the paper is that air temperatures across the USA over the last 125 years are dominated by extreme cold temperatures, while extreme hot temperatures cannot be differentiated from normal variations which are within +/- 2 standard deviations from the mean. This conclusion is based on the fact that every single weather station tells the same story - each weather station's distribution curve is heavily skewed to the left of the mean, i.e., to the cold tail of the curve. Since all data used in the paper is in the public domain, the results and the conclusions can be independently validated by the specialists in this field.

---

* Dr Darko Butina is a retired scientist with 40 years of working in experimental and computational fields of drug discovery. For any enquiry please contact the author at darko.butina@chemomine.co.uk.

INTRODUCTION

One of the notoriously difficult problems to solve in physical and biological sciences is to design a statistically valid protocol that will partition a set of numbers into some predefined number of classes. For example, in drug discovery where the author of this paper has spent his professional career, we very often want to assign 3-class labels, for example, ***very inactive***, ***active*** and ***very active*** to numbers generated by some biological screen.

Using a trial and error approach, a consensus has been reached that the concept of the normal distribution curve can be used to design classification protocol by which all datapoints that lie between 2 standard deviations on either side of the mean are labelled as normal, while those that lie outside 2 standard deviations on either side of the mean are labelled as extreme [1, 2, 3]:

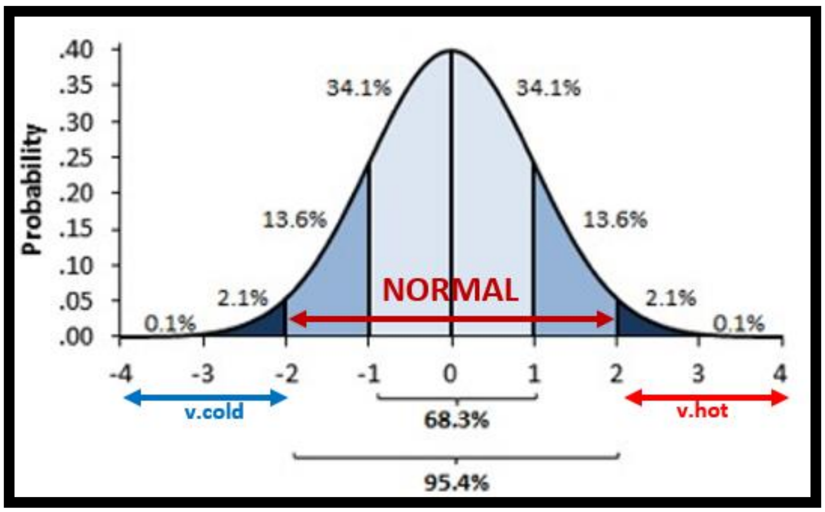

Figure 1. Normal distribution bell-shaped curve where 95% of data reside within +/- 2 standard deviations from the mean and are labelled as normal, while 5% of the datapoints which are outside that range will be labelled as extreme.

The most important fact about the concept of normal distribution is that the datapoints outside the normality boundaries, i.e. those outside 2 standard deviations on either side of the mean, are considered to be statistically significant numbers and therefore require further analysis.

Before we continue any further, let us briefly summarise the basic concept of normal distribution. The key step in generating the distribution curve is to transform any given dataset into the universal distance space in which each datapoint is expressed in a number of standard deviations away from the mean. That *distance from the mean in standard deviations is known as* the ***z-score***. Since the distance from the mean can be either left or right from the mean, a negative sign is assigned to all datapoints that are left of the mean, while those to the right have a positive sign, or no sign at all. The mean, by definition, has the z-score = 0. Please note that very often the terms below or above the mean are used instead of left or right. Also, the term SD will be used as short for standard deviation.

So, if the z-score is, say -2.3, that means that the datapoint is 2.3 standard deviations to the left of the mean, while the z-score = 1.8 means that it is 1.8 standard deviations to the right of the mean.

Since the concept of a normal distribution curve is a key part of the classification protocol that will be used in this paper, let us simplify Figure 1 and define and declare the key reference points for the classification scheme:

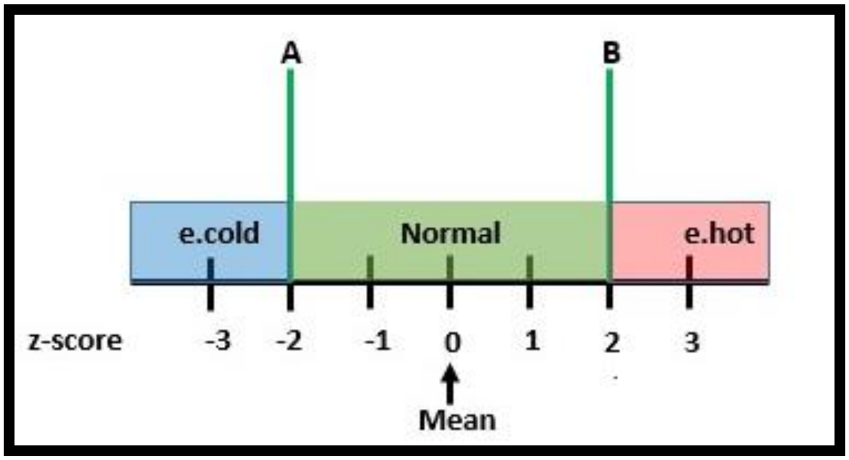

Figure 2. Classification rules to assign each datapoint into one of the three classes, **e.cold**, **normal** or **e.hot.**

The points A and B (Fig.2) represent ***normality boundaries*** and all the datapoints **within**, or inside, those two boundaries will be assigned the label ***normal***. Datapoints that are outside the lower normality boundary A (A has z-score = -2.0) will be labelled as e.cold, while those above the upper boundary B (B has z-score = 2.0) will be labelled as e.hot:

Normal (between points A and B): -2.0 > = z-score < = 2.0 (1)

e.cold: z-score < -2.0 (2)

e.hot: z-score > 2.0 (3)

So, the datapoint with a z-score = -3.4 will be labelled as e.cold, the datapoint with z-score = 1.2 will be labelled as normal, while the datapoint with z-score = 2.1 will be labelled as e.hot.

One of the main features of the use of z-scores that cannot be emphasised enough, is the ability to back-transform z-score to the datapoint's original value. For example, the transform formula to calculate the z-score is

z-score = (X-Mean)/SD (F1)

where **X** is the original datapoint, while **Mean** and **SD** are the mean and the standard deviation of the dataset. In excel, the mean and the standard deviation can be calculated by using functions ***average*** and ***stdevp*** respectively.

If one wants to back-transform the z-score to its original value, the following formula is used:

X = Mean+(z-score*SD) (F2)

For example, if a dataset has the mean = 15.0, SD = 10.0 and datapoint X = 20.0C, the z-score would be:

z-score = (20.0–15.0)/10.0 = 0.5 using formula (F1).

However, if we want to find the original datapoint for z-score = 2.1 using the same mean and SD as above, **X = 15.0+(2.1*10.0) = 36.0C** using formula (F2).

Now that all the definitions and the classification protocol have been established let us move to the issue of the datasets that will be analysed.

DATASETS

The paper will be analysing datasets from 14 weather stations which are based at 14 different geographical locations across the USA. Each weather station's dataset contains maximum daytime air temperatures, tmax, for approximately 125 years. All the data is available free of charge from the NCDC/NOAA global daily station data that covers 23717 different geographical locations. The data can be accessed for download via KNMI Climate Explorer (http://climexp.knmi.nl/start.cgi).

The datasets were selected based on two simple criteria:

- The weather stations should be representative of East, West and South coastal regions and the continental parts of the USA
- Each weather station should have at least 100 years of data

A few words are needed here about the history of collecting daily temperature data and the importance of the calibrated thermometer. The first systematic recording of the maximum and minimum daily temperatures across the globe started in the mid-1800s, where the highest daily temperature was labelled as tmax while the lowest as tmin. The observations protocol designed by the meteorologists of that time makes scientific sense for the following reasons:

– the highest temperature during daytime, tmax, corresponds to the kinetic energy of the air molecules that surround the thermometer. That in turn tells us about the amount of heat energy generated by the sun that has reached the fixed-to-the-ground thermometer. *In the thermodynamic terms, the tmax reflects the process of warming*

– on the other hand, tmin observed during night time, when the sun does not heat that part of our planet, *reflects the process of cooling*, thermodynamically a very different process from the warming process that is captured by tmax

In order to make sense of any temperature patterns generated during the day (tmax) and those generated during the night (tmin), we need to separately treat tmax data from tmin data.

In this paper, only maximum daytime temperatures will be analysed.

Format and the Size of Datasets

Table 1. Typical table that comes from download of tmax data using KNMI Climate Explorer

order	yy	mm	dd	tmax
1	1893	1	1	-2.8
2	1893	1	2	-15
3	1893	1	3	-8.9
4	1893	1	4	1.7
5	1893	1	5	-13.3
	……	……	……	……
43762	2016	9	29	20
43763	2016	9	30	21.1
43764	2016	10	1	18.3
43765	2016	10	2	24.4

All the downloads from KNMI Climate Explorer consist of 4 columns that are labelled here as yy (year), mm (months), dd (day) and tmax (daytime maximum temperatures in degrees C).

The first column labelled '*order'* was created by the author to help with sorting and allowing re-setting the dataset to its original order.

It is very important that any sorting on any individual column must be done on all columns since the internal relationship between each column must be maintained throughout the analysis.

A Detailed Statistical Analysis for Willow City (ND)

The main part of the paper begins with a detailed statistical analysis of a single weather station, Willow City in North Dakota, a small town near the Canadian border which has 125 years of data from 1892 to 2016:

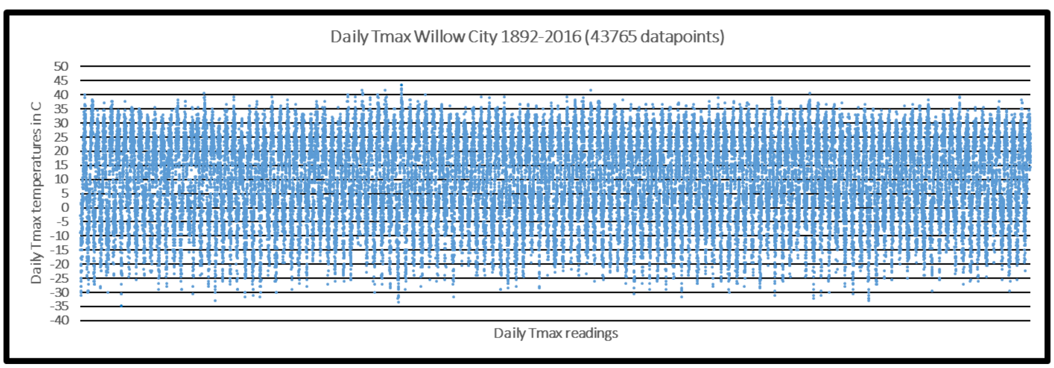

Figure 3. Daily tmax temperatures for Willow City between 1892 and 2016 with 43765 daytime readings.

There are 43,765 observations of tmax observations at Willow City between 1892 and 2016 with temperatures ranging between -35.0C in the winter and 45.0C in the summer, with a total range of 80C.

After calculating the mean and the standard deviation, extracting minimum and maximum tmax the following table is created:

Table 2. Key reference points for Willow City daily tmax data between 1892 and 2016

WS	Mean	SD	Min Tmax	z-score min	A (z-score=-2)	B (z-score=+2)	Max Tmax	z-score max
WS-11	10.1	15.0	-35.0	-3.0	-19.9	40.1	43.3	2.2

The statistical analysis in this paper will be concentrated on two parameters that best describe the extreme tails of the distribution curve: **the range** and **the frequency** of the extreme cold, e.cold, and extreme hot, e.hot regions.

The first step in the analysis, after Table 2 is created is to transform each temperature reading into its corresponding z-score using formula z-score = (X – Mean)/SD followed by sorting the whole dataset on z-scores:

Table 3. Tmax data for Willow City transformed to z-scores and sorted on z-scores in descending order

order	yy	mm	dd	tmax	z-score
14794	1936	7	11	43.3	2.21
14789	1936	7	6	42.2	2.14
14793	1936	7	10	42.2	2.14
12973	1931	6	17	41.7	2.11
14029	1934	5	30	41.7	2.11
......	......	......	......	......	......
36326	1996	2	2	-32.2	-2.82
6216	1912	1	11	-32.8	-2.86
36313	1996	1	20	-32.8	-2.86
14645	1936	2	12	-33.3	-2.89
1900	1899	1	31	-35.0	-3.01

The table above shows the five highest and five lowest z-scores for Willow City, or in terms of temperatures, the top five highest and lowest temperatures. The maximum tmax was recorded on 11/7/1936 with the reading of 43.3C and z-score of 2.2 (top row), while minimum tmax was recorded on 31/1/1899 with the reading of -35.0C and z-score of -3.0.

The difference between the z-score for the maximum tmax, 2.21 and the upper normality boundary which is set at z-score = 2.0 (point B in Figure 2), identifies the total range for the e.hot datapoints, which is 0.21 SD. In other words, the extreme hot region extends only 0.21 standard deviations outside the normal region. In terms of actual tmax temperatures, the difference between the maximum tmax, 43.3C, and the temperature back-transformed from z-score = 2.0 which is 40.1C (column 7 in Table 2), gives the total range outside the upper normality boundary at 3.2C.

If we repeat the same process for the e.cold region, where z-score = -3.1 for the minimum tmax and lower normality boundary with z-score = -2.0, the range of e.cold datapoints in z-scores is 1.1 SD, while in temperature terms, it is 15.6C (temperature at z-score = -3.1 is -35.0C, while at lower normality boundary it is -19.9 and therefore the difference is 15.1C).

What can be concluded for Willow City is that the total range of z-scores for the e.cold datapoints is 1.1 SD, while those for e.hot is only 0.21 SD outside the normal region.

The second parameter that we need in order to make some definitive conclusions about the extreme temperature patterns for that location, is the frequency or number of datapoints that are labelled either e.cold or e.hot. There are many ways that one can count the number of datapoints below or above a given temperature, and excel offers a very simple macro called 'countif' that does the job.

Since we are using the classification protocol where all datapoints below z-score = -2.0 will be labelled as e.cold and those above z-score = 2.0 will be labelled as e.hot, the ***countif*** macro will return number 1102 as number of datapoints with z-sores < -2.0, and only 12 with z-score > 2.0. We are now able to construct the distribution curve for Willow City:

Table 4. Distribution curve for Willow City between 1892 and 2016

	Counts	Percent
normal	**42651**	**97.45**
e.cold	**1102**	**2.52**
e.hot	**12**	**0.03**

Whichever way we looked at the datapoints outside the normality boundaries, we come to the same conclusion: the history of temperatures recorded in Willow City is dominated by the cold extremes, both in terms of range and frequency, i.e. the distribution curve is skewed to the left of the mean.

Daily Tmax Temperature Patterns for 14 Weathers Stations Across the USA

Table 5. Summary table for 14 weather stations across the USA with the historical data since the late 1890s

Order	WS	Weather Station, USA	Years	n total	Mean	SD	Max Tmax	Min Tmax	TotRange (max - min)
1	WS-1	Sparta, IL	1893-2015	42443	19.8	11.1	45.6	-19.4	65.0
2	WS-2	Lincoln, VA	1900-2016	40751	19.4	10.4	42.8	-13.9	56.7
3	WS-3	Urbana, OH	1895-2016	40398	16.3	11.4	43.3	-23.9	67.2
4	WS-4	Death Valley, CA	1961-2016	20209	32.9	10.6	53.9	3.3	50.6
5	WS-5	Norton, KS	1893-2016	41996	19.1	12.2	46.7	-23.3	70.0
6	WS-6	New Braunfels, TX	1893-2016	42544	26.8	7.9	44.4	-6.7	51.1
7	WS-7	Miami, TX	1905-2006	35673	21.9	10.8	45.6	-5.0	50.6
8	WS-8	Dillion, MT	1895-2016	41647	14.4	10.9	38.9	-27.8	66.7
9	WS-9	Hanford, CA	1899-2016	39842	25.0	9.3	46.7	-1.1	47.8
10	WS-10	Chappel Hill, NC	1891-2016	43476	21.8	9.1	41.7	-13.3	55.0
11	WS-11	Willow Hill, ND	1892-2016	43765	10.1	15.0	43.3	-35.0	78.3
12	WS-12	Bozeman, MT	1892-2016	44236	13.0	11.4	40.6	-31.3	71.9
13	WS-13	Bottineau, ND	1893-2016	42770	9.6	15.0	43.9	-36.1	80.0
14	WS-14	New Your, NY	1876-2016	51254	16.5	10.4	41.1	-16.7	57.8

Brief description of the table above:

- columns 1, 2 and 3 are the codes designed for lookup, sorting purposes and geographical locations of **W**eather **S**tations (WS)
- column 4 tells us which years of daily temperatures are covered by the weather station
- column 5 contains number of datapoints that are recorded at individual weather stations
- columns 6 – 9 hold the key reference points from which all other relevant numbers for analysis can be generated
- the last column holds total ranges of temperatures recorded at each weather station

The numbers highlighted in yellow are maximum and minimum values for each column for quick visual assessment.

The last column reflects the size of the natural variations in daytime temperatures observed, with 47.8C being lowest (WS-9) and 80.0C the largest (WS-13). One way to look at those numbers is to divide them by 2 and then we can state that, for example, the natural variations at the WS-13 can be 40.0C on either side of the mean, while those at the WS-9 can be 23.9C on either side of the mean.

The Ranges of Datapoints Classified as e.cold and e.hot

The table below summarises the maximum and minimum observed temperatures at each weather station (labels WS-1 to WS-14) with their corresponding z-scores:

Table 6. Maximum and minimum recorded tmax daily temperatures at 14 weather stations expressed in z-scores

WS	Min Tmax	z-score min	Max Tmax	z-score max
WS-1	-19.4	-3.5	45.6	2.3
WS-2	-13.9	-3.2	42.8	2.3
WS-3	-23.9	-3.5	43.3	2.4
WS-4	3.3	-2.8	54.2	2.0
WS-5	-23.3	-3.5	46.7	2.3
WS-6	-6.7	-4.2	44.4	2.2
WS-7	-5.0	-2.5	45.6	2.2
WS-8	-27.8	-3.9	38.9	2.2
WS-9	-1.1	-2.8	46.7	2.3
WS-10	-13.3	-3.9	41.7	2.2
WS-11	-35.0	-3.0	43.3	2.2
WS-12	-31.3	-3.9	40.6	2.4
WS-13	-36.1	-3.0	43.9	2.3
WS-14	-16.7	-3.2	41.1	2.4

The above Table lists the maximum and minimum recorded temperature at each weather station and also corresponding z-scores. Since the extreme labels are assigned to datapoints outside +/- 2SD distances from the mean, it follows that subtracting 2.0 from the z-scores listed in columns 3 and 5 will reflect the maximum ranges of z-scores in the extreme hot and cold region. For example, the lowest z-score was observed at WS-6 at -4.2 (highlighted in yellow) and subtracting 2.0 from it gives us the total range for that weather station at -2.2 standard deviations in the cold extreme range. At the other extreme, weather station 3 (WS-3) has the largest z-sore at 2.4 and subtracting 2.0 gives us the total range at 0.4 standard deviations in the hot extreme range.

On average, the total range for e.cold class is 1.35 standard deviations in the cold direction, while for the e.hot class only 0.26 standard deviations in the hot direction.

If we back-transform the z-score ranges into the actual temperature ranges, using formula

X = Mean + (z-score*SD) we get the following ***temperature*** ranges for e.cold and e.hot classes:

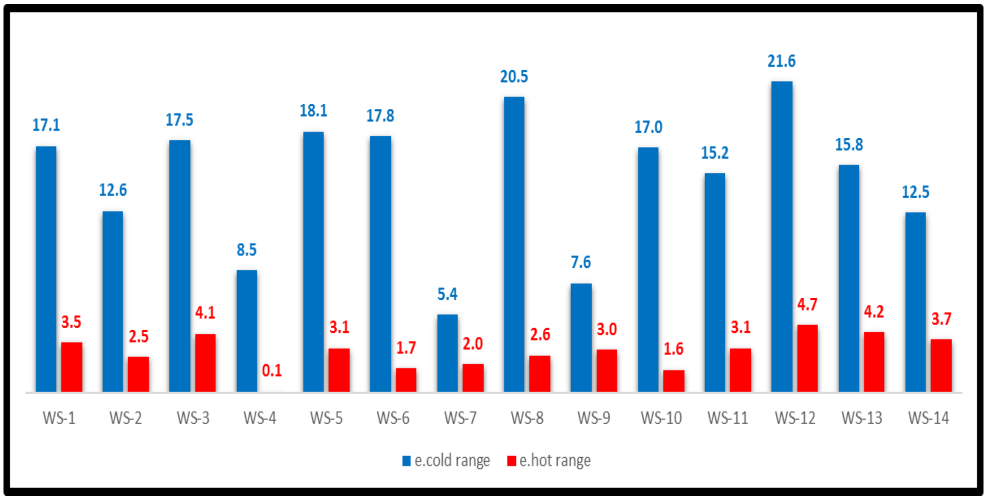

Figure 4. Temperature ranges (in degrees C) observed in e.cold (blue) and e.hot (red) classes.

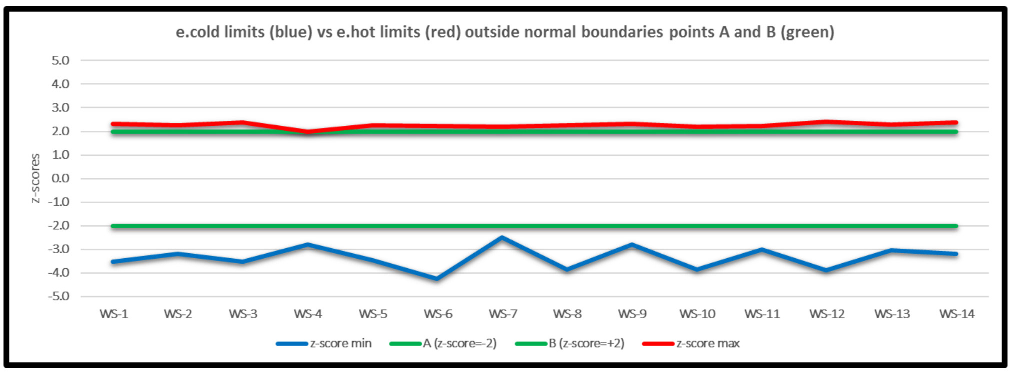

Figure 5. Maximum observed temperatures for each weather station expressed in **z-scores** in red, and minimum in blue. Green lines represent normality boundaries, lower one with z-sore = -2.0, upper one with z-score = 2.0.

The cold extreme temperature ranges are the dominant features for each weather stations across the USA.

The figure summarises the e.cold/e.hot patterns (***in z-scores***) for the USA since the late 1890s:

The most important point of Figure 5 is that the ***maximum*** z-scores (red) cannot be distinguished from the upper normality boundary (B), while the ***minimum*** z-scores (blue) are clearly separated from the lower normality boundary.

If we plot the same data but instead of the z-scores, the original tmax daily temperatures are used, we should get the same patterns as in Figure 5 but the differences between 3 classes would be expressed in degrees C:

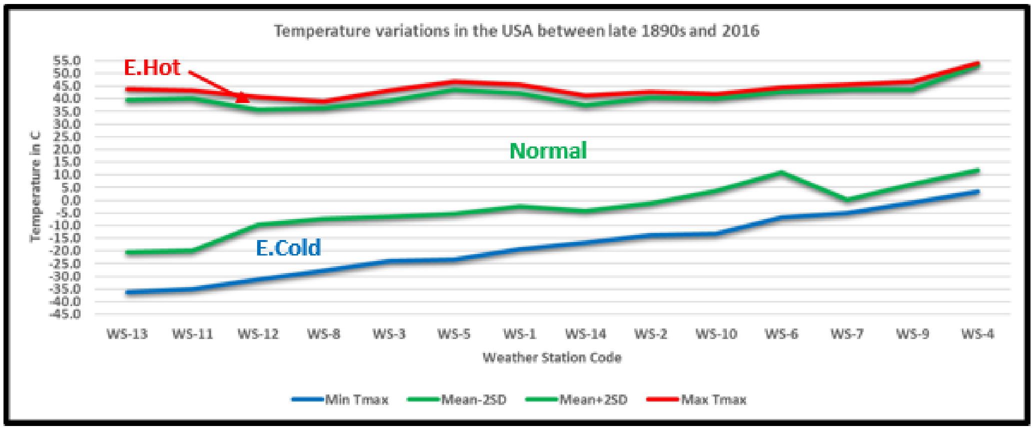

Figure 6. Maximum temperatures in red, upper and lower normality boundaries in green and minimum observed temperatures in blue.

The main difference between Figure 5 and Figure 6 is that the latter is displayed in the original units, degrees C, which was sorted on the minimum observed tmax temperature for each weather station. In this way, it can be clearly seen that the coldest temperature detected in the USA was observed at WS-13 (bottom left corner) -36.1C (Bottineau, North Dakota), while the hottest observed at WS-4 (top right corner) at 54.2C (Death Valley, California).

The average temperature difference between minimum observed temperatures and the lower normality boundary, class **e.cold**, is 15.6C, while the other extreme is 3.2C, class **e.hot**. The average range for class **normal** is 44.4C.

The main reason for generating Figures 5 and 6 is to re-emphasise the importance of z-scores and the concept of the normal distribution curve where the normality boundaries always have the same value with z-score = -2.0 for the lower boundary, and 2.0 for the upper. That fact alone enables a direct comparison of datasets that might have huge variations in their original units, like degrees C.

On average, each weather station spends **934** days (***out of total days recorded***) in the cold extreme (blue) while only **23** days in the hot extreme (red). If we take into account the total number of records at each weather station, we can generate the distribution curve for the extreme tails at the opposite side of the mean in percent of datapoints (out of the total number of datapoints) in two extreme classes.

The frequencies of datapoints classified as e.cold and e.hot

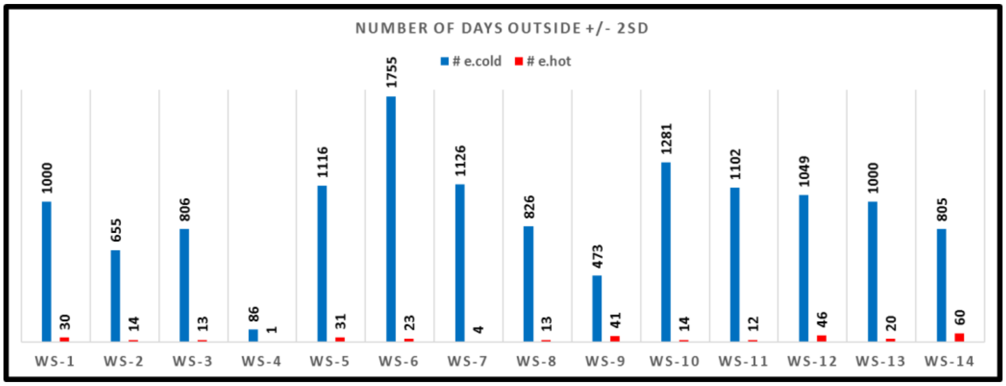

Figure 7. Number of days that each weather station spent outside +/- 2 SD: class e.cold (blue) and e.hot (red).

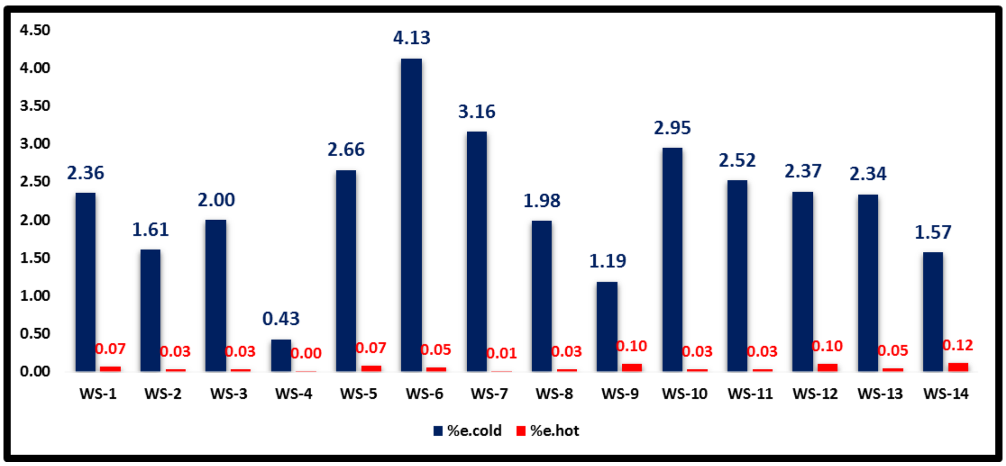

Figure 8. Tails distribution as % of total number of datapoints: e.cold class (blue) and e.hot class (red)

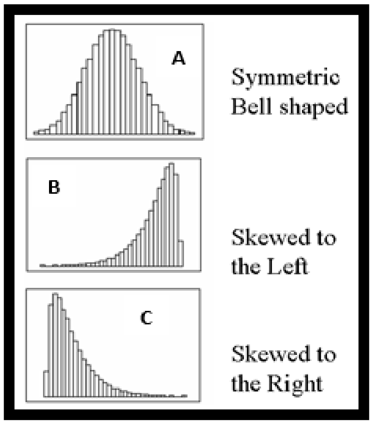

Figure 9. Three basic types of the distribution curves: Symmetric (A), Skewed to the left (B) and Skewed to the right (C).

So, again, a very clear pattern is emerging where the frequency of the class **e.cold** is on average **2.2%**, the class **e.hot** is **0.05%** while the class **normal** covers **97.75%** of the total population.

While this paper's starting point has been the concept of the symmetrical distribution curve, the science of statistics tells us that a given dataset can be best described by one of the three distribution curve shapes, Figure 9.

As we have seen from our analysis, every single weather station can be best described by the shape B (Figure 8), i.e., skewed to the left, or in our case skewed to the extreme cold part of the curve.

SUMMARY AND CONCLUSION

This paper has applied one of the gold standards in statistics, the normal distribution bell-shaped curve, to identify the extreme cold and hot daily temperatures across the USA since the late 1890s. The concept of the normal distribution curve was used to design a classification protocol in which datapoints were partitioned between 3 classes:

- e.cold class which was defined by z-score < -2.0
- e.hot class defined by z-score > 2.0 and
- normal class defined by z-score > = -2.0 and < = 2.0

We have analysed 14 weather stations across the USA with the oldest one, New York going back to 1876, analysed in a total of 571,004 daily tmax records with a total range of temperatures of 90.0C. ***The paper did not start by trying to prove or disprove any model or hypothesis published in the different fields of climate sciences, but simply perform a systematic statistical analysis of the original records that are in the public domain.*** Since the paper uses a clearly defined classification protocol and the original data without any modifications, all the results and conclusions can be easily validated by any scientist *who is familiar with the statistical tools used in this paper and most importantly understands the physical meaning of the numbers generated by a calibrated thermometer.*

In conclusion:

- every weather station's distribution is significantly skewed to the extreme cold tail of the distribution curve (Figure 8)
- on average, the e.cold class is a 1.35 standard deviation outside the lower normality boundary, while the e.hot class is only a 0.26 standard deviation outside the upper normality boundary. The ratio between the ranges of e.cold/e.hot is 5.2 (Figure 5)
- on average, the temperature ranges for e.cold class are 15.6C while temperature ranges for e.hot class are 3.2C (Figure 6)
- on average, each weather station has recorded 934 days (out of the total number of records) that are classified as e.cold, while only 23 days as e.hot (Figure 7)
- in percentage terms, e.cold class covers on average 2.2%, the e.hot class 0.05% while the class normal covers 97.75% of the total
- the overall trend for the temperatures in the USA over the last 125 years is of cooling, while the extreme hot temperatures cannot be separated from the normal daily tmax variations

Note About the Author

Dr Darko Butina is a retired scientist with 20 years of experience in experimental organic and medicinal chemistry plus a further 20 years of working in the field of pattern recognition and datamining of experimental data. He was a member of the team that designed and synthesised the first effective drug, Sumatriptan, for treatment of migraine – an achievement for which the team at Glaxo (UK) received The Queens Award. For more than twenty years Sumatriptan has improved the quality of life for millions of migraine sufferers worldwide. While working in the computational side of drug discovery, the author developed a novel clustering algorithm, dbclus, that became a de facto standard for quantifying diversity in world of molecular structures.

Since his retirement, he has applied various numerical tools developed in the fields of pattern recognition, datamining, machine learning to mention but a few, to the analysis of the thermometer-generated data and has published 4 papers since 2012 [4,5,6,7] with two awaiting publication in early 2018. He also runs his own website at www.l4patterns.com.

Acknowledgment

I acknowledge KNMI Climate Explorer for free access to the global temperatures database which was used in this paper. I also acknowledge my wife, Judy Glasman for proof reading this paper and the moral support

References

[1] Altman D., Why We Need Confidence Intervals, *World J. Surg.*, 2005, 29, 554–556.

[2] Copeland B., A practical means for measurement and control of the precision of clinical laboratory determinations, *Precision in Clinical Laboratories*., 1957, 553, 551-558.

[3] Finney D., *Experimental Design and its Statistical Basis*, University of Chicago Press, 1955, p. 169.

[4] Butina D., Should we worry about the Earth calculated warming at 0.7C over last 100 years?, *Int. J. of Chemical Modeling*., 2012, 4, Number 2-3, 233-253.

[5] Butina D., Quantifying the effect that N2, O2 and H2O have on night-to-day warming trends at ground level, *Int. J. of Chemical Modeling*., 2013, 5, Number 4, 457-478.

[6] Butina D., Is Arctic Melting? Theory vs. Observations, *Int. J. of Chemical Modeling*., 2015, 7, Number 1, 91-113.

[7] Butina D., New algorithm to identify coldest and hottest time periods. Case study: Coldest winters recorded at Armagh Observatory over 161 years between 1844 and 2004, *Int. J. of Chemical Modeling*., 2015, 7, Number 3, 212-227.