Bulletin of the *Transilvania* University of Braşov
Series II: Forestry • Wood Industry • Agricultural Food Engineering • Vol. 6 (55) No.2 - 2013

A CASE STUDY OF MONTHLY VALUES PROBABILITY USING GLOBAL HISTORICAL CLIMATOLOGY NETWORK DATA

Victor-Dan PĂCURAR[1]

Abstract: *The paper presents a study of a monthly temperatures time series, using data available from the Global Historical Climatology Network for Sibiu. Only by analysing the time series variability it is possible to correctly identify the range of "normal" climate conditions and the fluctuations or eventually climate changes that could affect forest dynamics. A simple empirical method, the quantile plot, was used for estimating the probabilities associated with different temperature monthly values, which are very useful and possible to be easily estimated using this procedure by researchers and managers from forestry, agriculture, land planning etc.*

Key words: *monthly temperatures, global databases, quantile plot.*

1. Introduction

Climate conditions play a vital role for life on Earth, the human society making no exception. The importance of climatology studies, aimed to describe, understand and eventually predict the climate of an area is presently widely acknowledged [9]. The main reason for this is certainly the increasing concern of the large public regarding the possible climate changes.

In the past, a very simple definition of climate, as representing the average weather, was accepted [6]. But climate is never constant and change is actually the main characteristic of both weather and climate. Thus, nowadays even descriptive climatology considers both average and extreme conditions as climate characteristics.

A clear distinction has to be made between climate change and climate variability. The main difference relates to the time scales, since variability is observed from season to season and year to year, fluctuations span for decades and climate change for centuries [1].

When climate variability is not properly considered the analysis could be confusing. This is quite common in the public opinion that often considers any difference from the so called "normal" values as anomalies and proofs of climate change. There are also studies of applied climatology in forestry, agriculture, tourism etc. using only the average data (not even completed by standard deviations), with low descriptive quality and very limited utility in planning and design.

The quality and utility of climate variability analysis could be increased by estimating the probabilities associated with the different values of climate parameters.

[1]Silviculture Department, *Transilvania* University of Braşov.

For this, researchers have to consider the time series of climate parameters for long intervals, preferably more than 30 years (the standard period). There is a wide range of study possibilities, spanning from rigorous and sophisticated methods of time series modelling to simple empirical procedures, such as the one presented in this paper.

2. Materials and methods

A clear aim of this study was to use easily available data and simple data processing techniques enabling the researchers from different domains, forestry, agriculture, land planning etc. to use it with minimal additional efforts as compared with the traditional approach.

Presently, there are online sources of climate data among which the global historical databases [3], [8] for weather station data and reanalysis (retrospective analysis) databases for gridded data [2].

The data used in this study were extracted from the Global Historical Climatology Network (GHCND)-Monthly Summaries, where one can find data from more than 40000 weather stations world wide. These are derived from the GHCN-Daily database which compiles climate data from numerous sources synthesized and carefully reviewed for quality assurance. The Monthly Summaries database includes 18 meteorological elements referring to temperature, precipitation, snowfall, maximum snow depth, and degree days.

For data downloading the Climate Data Online portal of the NOAA was used, where monthly values could be freely ordered [12], a fee being only charged for certifying data to be used in law or insurance matters.

In this case study, the weather station chosen for the monthly temperature values analysis was Sibiu. This is one of the oldest in Romania, where systematic observations began in 1851, with data representative for the southern Transylvanian region. The monthly values were extracted for 40 years, from 1970 to 2009. It is very important to mention that in this database similar data could be found for several other stations distributed across the country, allowing a similar study for other regions.

As mentioned before, an empirical simple method was used in order to facilitate its application and also to provide additional parameters easy to understand and use. This is the quantile plot or the empirical distribution chart method, based on calculating the cumulative probabilities (or plotting positions) as simple functions of the value rank in the sorted series and the total number of observations [5], [10]. The basic manner of calculating the empirical probabilities in this study was by dividing the rank index (in the descending sorted data set) by the number of distinct observations supplemented by one [7]. Afterwards these were multiplied by 100, obtaining percent values (expressing the chances of exceeding the threshold in a century) and plotted against the corresponding monthly temperature values.

3. Results and Discussion

The first step in analysing the data set of monthly temperature values (1970-2009) was that of calculating the multi-annual averages for each month and the whole year, presented in Table 1 (AVG-1) along with the corresponding standard deviations. In the table there are also included the monthly averages (AVG-2) for a previous time interval (1896-1955)[11]. By comparing the two sets of long term averages one could observe a slight increase of 0,2^{o}C for the whole year and more accentuated positive differences in January (1,3^{o}C), February (0,8^{o}C) and March (0,5^{o}C).

Monthly average temperatures and standard deviations [^{o}C] Table 1

Month	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Year
AVG-1	-2,5	-0,4	4,6	9,8	14,7	17,7	19,4	19,0	14,9	9,7	3,8	-1,1	9,1
STDEV	2,3	3,0	2,3	1,6	1,5	1,3	1,3	1,3	1,6	1,3	2,2	2,4	0,7
AVG-2	-3,8	-1,2	4,1	9,7	14,7	17,7	19,6	18,9	14,8	9,4	3,6	-0,9	8,9

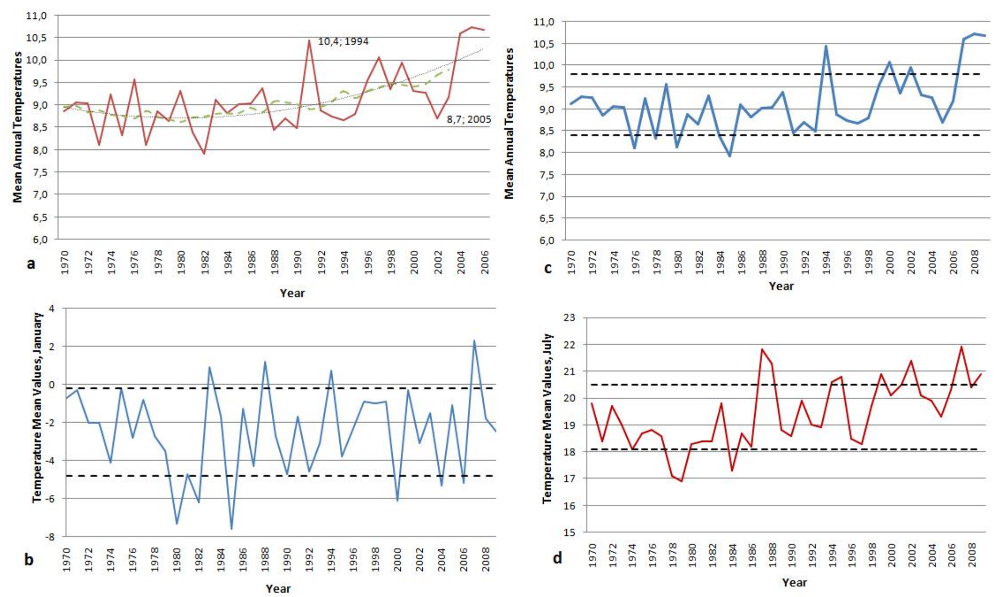

Fig. 1. *Air temperatures time series (1970-2009) at Sibiu. a-Mean annual values along with 7 years moving averages(dash line) and polynomial trend line (dotted line), b.Mean monthly values-January, c- Mean annual values, d-Mean monthly values-July (in b,c and d the dash lines are marking the thresholds calculated by subtracting and adding the standard deviation from the series average)*

The standard deviation values (STDEV) are much higher in the winter months, almost double, than in summer indicating an increased variability in this season. It is also important to mention the lowest standard deviation (0,7), corresponding to the series of annual averages.

As regards the warming trend, the analysis of the annual averages series, presented in Figure 1a and c, brings more information. The seven years moving averages and the polynomial trend adjusted lines (Figure 1a) show that the warming began after 1990. This could be explained by looking at the warm years, with annual averages exceeding the multi-annual mean with more than the standard deviation (over the threshold dash line marked at 9,8^{o}C in Figure 1c). These years are 2009 (10,7^{o}C), 2008(10,7^{o}C), 2007(10,6^{o}C), 2002(10,0^{o}C), 2000(10,1^{o}C) and 1994 (10,4^{o}C). It is noteworthy that this situation is quite different from the global one [4]. For example, 2005 was one of the warmest years at global level (0,84^{o}C over the mean) and for the northern hemisphere (1,08^{o}C over the mean) but it was colder than usual at Sibiu (8.7^{o}C). Also 1998, another very warm year for the globe was relatively cold in this location (8.8^{o}C).

Regarding the monthly average in January (Figure 1b), after 2000, there are values exceeding the average standard deviation of the period (2007) and years with values below the dashed line that marks the lower threshold (2000, 2004 and 2006). The cold January of 2000 (-6.1 ° C) is close to that in 1982 (-6.2 ° C) but still warmer than the extreme months of the entire period from 1980 (-7.3 ° C) and 1985 (-7.6 ° C), which is evident in the chart notes.

In July, the monthly averages (Figure 1d) peak in 2007(21,9^{o}C) but a very close value was recorded in 1987 (21,8^{o}C). Other warm July values (above the threshold) occur in 1988, 1994, 1995, 1999, 2002 and 2009. The cold Julies, under the threshold (18,1^{o}C), were less frequent, occurring in 1978, 1979 and 1984.

Certainly, by considering the standard deviation thresholds together with the multi-annual averages, the deviations from the “normal” climate could be more accurately identified, as shown in the short analysis presented above.

A better judgement of a particular situation, with its corresponding temperature value, could be made by calculating its associated probability, respectively the chances to have similar values in other years. A simple way of doing this was by using the quantile plot method, briefly described in the previous chapter, applied in this case study.

Examples of such quantile plots are presented in Figure 2. In the charts, the dash lines are marking the polynomial adjusted series of cumulative probabilities. As regards the mean annual values (Figure 2a), from the chart one can observe that, for instance, a temperature value equal or greater than 10^{o}C corresponds to a cumulative probability of 20%, meaning that this event could happen twenty times in a century or in other words it has a return period of 5 years. The 10,6^{o}C annual average temperature, which was exceeded only in the last 3 years of the studied 40 years period, has a cumulative probability of 9,1% and an expected return period of more than a decade.

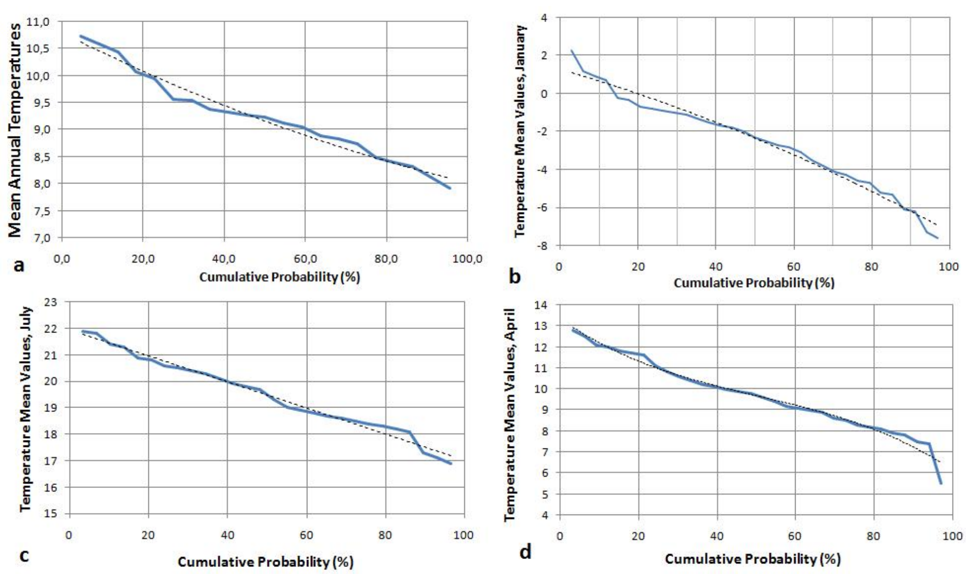

Fig.2. *Quantile plots for annual and monthly air temperature averages*
(a – Annual means, b-January means, c- July means, *d- April mean values)*

Looking at the other end of the chart, a value of 8,4^{o}C (the threshold calculated by subtracting the standard deviation from the average) has a probability of 81.8%, the chances to get a lower value being 18,2%, an event that is expected to happen once every 5,5 years.

One of the definitions of “normal” climate conditions [1], widely accepted, considers the values associated with 33% and 67% (delimiting upper and lower thirds). For the Sibiu station 40 years annual averages series, these limits approximately correspond to 8,8^{o}C and respectively 9,4^{o}C. This interval is narrower than that defined by taking into account the standard deviation (8,4^{o}C to 9,8^{o}C) and consequently some more years could be considered outside the “normal” range (for instance 1979 or 1991).

As concerns January, by analyzing the quantile plot (Figure 2b) it is possible to assess the chances to get a very warm month, with an average over 0,7^{o}C, corresponding to 11,8%. Most of the years will probably have a monthly average over -5^{o}C (80%, thus eighty in a century) and the chances to get values under -7^{o}C (as in 1980 and 1985) are under 8%. The limits of the “normal” interval, defined as the middle third of the observations, could be estimated at -3,7^{o}C respectively -1,2^{o}C, obtaining a range reduced with almost 2^{o}C as compared with the one delimited by the standard deviation around the multi-annual average (-4,8^{o}C to -0,2^{o}C). A month average of -1,5^{o}C, recorded in 2003, one degree higher than the 40 year mean (-2,5 ^{o}C), lies in the normal conditions range, similarly with the colder previous year 2002 (-3,1^{o}C), despite the subjective impression one could get by simply comparing the two consecutive years.

A similar analysis could be made for the July chart of cumulative probabilities (Figure 2c). One could observe the upper limit of the graph under 22^{o}C, corresponding to the Koppen classification threshold for the warmest month. The chances to get a July average over 20^{o}C are quite high (40%) and a cool month with a mean under 17^{o}C is not very probable (under 10%, less than one in 10 years). The “normal” third of the observations lies between 18,6^{o}C and 20,3^{o}C.

Another example of quantile plot analysis refers to April, a spring month (Figure 2d). The chances to have a very warm month, with an average over 12^{o}C are less than 10% and a cold one, with a mean below 7,5^{o}C is expected less than once in a decade. The middle third of the observed range is delimited between 8,9^{o}C and 10,4^{o}C. This example also illustrates one of the problems of this simple method, which could occur with extreme values (5,5^{o}C in 1997). The accurate estimation of probabilities for such outliers requires more rigorous and complex models.

4. Conclusions

For analysing the climate dynamics and particularly the temperatures departures from the “normal” situation it is not acceptable to simply compare a value with the multi-annual average. Only values that are over or under the average with more than one standard deviation could be considered as special situations.

There is a common saying that climate is what we expect and weather is what we get. By estimating the probabilities associated with certain values, one could be more aware of what to expect. It is important to correctly establish the range of normal climate conditions, in order to identify the abnormal situations and eventually the possible climate changes. In analysing the forest dynamics, the estimation of the chances to have a year or more outside the optimum range for a species is very useful for both researchers and managers.

The proper approach of temperature series variability is presently facilitated by the data available from global historical climatology databases. By applying this simple method it is possible to easily estimate useful parameters.

References

1. Bridgman H.A., Oliver, J.E. et al., 2006. The Global Climate System. Patterns, Processes, and Teleconnections. Cambridge University Press.
2. Compo G.P., J.S. Whitaker P.D. et al., 2011. The Twentieth Century Reanalysis Project. In: Quarterly J. Roy. Meteorol. Soc., 137, 1-28.
3. Lawrimore J. H., Menne M. J. et al., 2011. An Overview of the Global Historical Climatology Network Monthly Mean Temperature Data Set, version 3, In: J. Geophys. Res., 116.
4. Lugina K.M., Groisman P.Y. et al., 2006. Monthly surface air temperature time series area-averaged over the 30-degree latitudinal belts of the globe, 1881-2005. In Trends: A Compendium of Data on Global Change. Oak Ridge National Laboratory.
5. Makkonen L., 2008. Bringing Closure to the Plotting Position Controversy, Communications in Statistics. In: Theory and Methods 37 (3): 460-467.
6. O`Hare G., Sweeney J. et al., 2005. Weather, Climate and Climate Change. Human Perspectives. Pearson Education Limited.
7. Păcurar V.D., 2004. Sunshine Duration Probabilities Estimation for the Brasov Area. In: Bulletin of the Transilvania University of Brasov, vol.11 (46): 241-244.
8. Peterson T.C., Vose R.S., 1997. An Overview of the Global Historical Climatology Network Temperature Database. In: Bulletin of the American Meteorological Society, 78 (12), 2837-2849.
9. Robinson P.J., Henderson-Sellers A. 1999. Contemporary Climatology. Pearson Education Limited.
10. Salas J.D., 1992. Analysis and Modelling of Hydrologic Time Series. In: Handbook of Hydrology, McGraw-Hill.
11. *** Romanian Climate. Climate Data, vol. II, 1961 (in Romanian). Bucuresti,
11. *** Climate Data Online. Available at: http://www.ncdc.noaa.gov/cdo-web/datasets/GHCNDMS/stations/GHCND: ROE00108901.