Chapter 9
Climate Change Assessment Using Statistical Process Control Methods

Branko Vučijak, Tarik Kupusović, Sanda Midžić-Kurtagić and Admir Ćerić

Abstract Statistical process control (SPC) uses the application of statistical methods and procedures to monitor and control a process, in order to evaluate two possible causes of variation in the process: natural (common) and assignable (special) causes. The aim of this activity is to improve the process' capabilities. If the variability of a process is within the range of natural causes, the process is said to be under statistical control. When that variability exceeds the expected natural causes range, it is a signal to look for, and to correct, assignable causes. SPC may even be used to "control" climate change, through comparison of present day variations with the natural variation capacity for change in air temperature, precipitation and sea levels in the past. Are today's frequent floods, tornados, warm winter periods or cold summer days actually caused by "natural" causes (should they be statistically expected), or has the "capability" of the natural processes changed? This paper will demonstrate the potential use of SPC methods in evaluating variations in temperature and precipitation that should be expected, based on the assessment of the statistical behaviour of data for these natural indicators during different periods. "Warning" and "action" lines will be assessed and compared for the selected periods. Also, the number of records below or above warning and action lines will be compared. This approach could be useful for spatial planners, even if the causes of the changes are global or not human-induced.

B. Vučijak (✉) · T. Kupusović · S. Midžić-Kurtagić · A. Ćerić
Hydro Engineering Institute Sarajevo, Stjepana Tomica 1, 71000 Sarajevo, Bosnia and Herzegovina
e-mail: branko.vucijak@heis.com.ba

T. Kupusović
e-mail: tarik.kupusovic@heis.com.ba

S. Midžić-Kurtagić
e-mail: sanda.midzic@heis.com.ba

A. Ćerić
e-mail: admir.ceric@heis.com.ba

W. Leal Filho et al. (eds.), *Climate-Smart Technologies*,
Climate Change Management, DOI: 10.1007/978-3-642-37753-2_9,

Keywords Statistical process control · Control charts · Climate change · Monitoring data · Process capability

Short Introduction

Statistical process control (SPC) presents the application of statistical methods and procedures of monitoring and control of a process, in order to evaluate two possible causes of process variation: natural (common) and assignable (special) causes. This paper will demonstrate the potential uses of SPC methods in evaluating variations in temperature and precipitation that should be expected, based on assessment of the statistical behaviour of data for these natural indicators during different periods.

Introduction

There is no full scientific consensus regarding the issue of climate change. While temperature increase has been noticed and validated over the last 10–20 years, there are still different opinions on whether this increase is related to permanent effects on the climate, or whether it is yet another climate variation. Opinions also differ on the source of such change—is it human-induced, or just another whim of nature?

This paper presents a potential response to the question of whether the temperature increase confirms that there has been a change from the former average temperature values and that the natural processes "capability" has changed, or the "temperature process" is still in control and the increase is actually within the range of values that could statistically be expected.

This paper demonstrates the potential uses of selected Statistical Process Control (SPC) methods in evaluating variations in temperature. It is based on the assessment of the statistical behaviour of temperature data for the period 1961–1990 when the "process was considered to be incontrol", and then used for evaluating the behaviour of temperatures during the period 2000–2010. Data relates to the city of Sarajevo and the same measurement site.

The structure of the paper is the following: after the introduction, the next section is a reminder of the main aspects of climate change and its basis. Section Climate Change introduces the statistical process control concept, while the fourth one presents results of the application of the SPC mean chart to the series of temperatures in Sarajevo, Bosnia and Herzegovina, during the periods 1961–1990 and 2000–2010. Finally, paper ends with some conclusions both on the effects of climate change and on the applicability of SPC in this area.

Climate Change

Scientists define climate as the average weather for a particular region and time period, including temperature, precipitation, humidity, sunshine, cloudiness and wind, usually taken over 30-year period (NASA 2005). It is really an average pattern of weather for a particular region. The Earth's climate has changed many times during the past in response to natural causes, due to interaction between the sun, land, oceans and atmosphere, and the effects that they have on each other. But the term "climate change" is used to refer to the changes in the climate caused by human activities, mostly during the 20th and 21st centuries, where "greenhouse effect" is specifically emphasised.

For thousands of years, the atmosphere contained relatively stable levels of greenhouse gases, but during the 20th and 21st centuries, human influence has changed the previous balance, resulting in climate change. America's Climate Choices report of 2011 states that the average temperature of the Earth's surface has increased by about 1.4 °F (0.8 °C) over the past 100 years, with about 1.0 °F (0.6 °C) of this warming occurring over just the past three decades. Most of the world scientists agree that global temperatures will continue to rise (Committee on America's Climate Choices 2011).

Nearly all European regions are anticipated to be negatively affected by certain future impacts of climate change, and these will pose challenges to many economic sectors. Climate change is expected to magnify regional differences in Europe's natural resources and assets (IPCC, Summary for Policymakers 2007).

For the south-east European countries, including Albania, Bosnia and Herzegovina, Montenegro, Serbia and Turkey, the impacts of climate change may include: increased temperatures; a rise in the frequency of extreme weather events; increased coastal erosion; a rise in sea levels; an impact on marine biodiversity; rising water levels in tidal rivers; increased flooding; severe pressure on water resources; increased forest and scrubland fires; changing agricultural landscapes, including crop failure; changes in habitat composition and species distribution, richness and diversity; and increasing problems caused by invasive alien species. In addition, there will be more problems for local and regional communities that depend on the services provided by ecosystems (in the form of food, drinking water, fuel, building materials or as a harvestable resource) to sustain acceptable living conditions and welfare (Laušević et al. 2008).

Following the scenario of partially applied measures for the reduction of greenhouse gas emissions, the territory of Bosnia and Herzegovina could expect an increase in air temperature of 3–4 °C by the end of the 21st century. The next several decades could bring a significant reduction in the number of days with snow, as well as a reduction in rainfall during the warmer part of the year, which would result in a reduction in soil humidity and the availability of water resources (Spasova et al. 2007).

The concept of statistical process control Vilfredo Pareto (1848–1923), who was trained as an engineer but is best known for his economic and sociological works, has set one of the basic optimisation postulates of statistical process control (SPC). He noticed that many failures in a system result from a small number of causes, and that in the production process it is rarely some "general malaise" causing problems. Pareto found that even though some companies show both diligence and hard work, and even strong motivation in some cases, the quality of the product or service was still poor. Thus, in order to improve this system for production, management or providing services, it is necessary to find and correct these causes, also called "Pareto glitches" (Thompson and Koronacki 2002).

During the 1920s, Walter Shewhart developed the basic theory of SPC, which was later popularised worldwide by Edwards Deming. Both of them noticed that repeated measurements of a single process will exhibit some level of variation. Even though Shewhart originally started working with manufacturing processes, both he and Deming understood that such observations could be applied to any sort of process. If a process is stable, its variation will be predictable and it is possible to describe it with some of the several statistical distributions (whereby normal distribution is most often used). Today, methods are used in different types of issues besides manufacturing processes, like healthcare (Benneyan et al. 2003), software processes (Jalote and Saxena 2002), statistical inference at work (Bakker et al. 2008) and others. Still, it seems that SPC has so far been used only in a limited manner for environmental processes.

It is considered that the inherent nature of any process has some common variations in cause that it is not possible to alter without changing the process itself. But 'assignable' or 'special' causes of variation are unusual disruptions to the process, the causes of which can, and should, be removed, of course after being recognised as such. One key purpose of SPC is to distinguish between these two types of variation, aiming to avoid both overreaction and under reaction, leading to a lack of need to respond to the process. It assists in recognising situations where reaction relates to a cause that has sufficient impact, and which is practical and economic to remove, in order to improve quality (Woodall 2000).

Application of the concept of statistical process control aims to enable steady improvement in the quality of a product, even while dealing with the everyday crises which are an unavoidable part of any production or service process. It needs to be underlined that statistical process control is completely different from the end product inspection conveniently associated with "quality assurance". It generally consists of the three following phases:

1. Provision of a flowchart of the process, clearly separating process functions and steps;
2. Random sampling and measuring, usually at regular temporal intervals, during different phases/functions of the process;
3. Provision of "control chart(s)" aiming to recognise such "Pareto glitches", in order to discover and remove their causes.

All processes can be supervised and brought 'under control' by collecting and using their data. This includes process quality-performance measurements, which will provide the feedback required for corrective action, but only where found necessary. Statistical process control methods provide an objective means of controlling the quality in any transformation process. W.E. Deming wrote that quality and productivity increase as variability decreases, and, because variations are unavoidable, statistical methods of quality control must be used to measure and gain understanding of the causes of the variation.

The essence of statistical process control is to differentiate the various causes of process variation. Some variations belong to the category of chance or random variations, considered as inherent to the process, and they could be removed only by revising the whole process. However, other causes of variation, relatively large in magnitude and identifiable, are conveniently classified as 'assignable' or 'special' causes. When special causes of variation are present, variation is excessive and the process is classified as 'unstable' or 'out of statistical control'.

Thus, SPC tries to respond to the following two key questions: (1) "Is the process incontrol?", and (2) "What is the extent of process variability?" A response to these questions actually relates to the potential presence of any special causes of variation, and identifies whether or not variability is due only to natural process capability, in which case only the common causes of variation are present.

To control a process using data, it is necessary to monitor the current state of the accuracy and precision of the distribution of the data, which is done using control charts. A control chart has a function similar to that of the traffic signal, where the green light is given when the process is running properly and does not need any adjustment (process is under control), meaning that only the common causes of variation are present. The next level is the amber light, which signals that some discrepancy to the natural process might be possible. The red light clearly shows that assignable or special cause(s) of variation appeared before the occurrence of such data and the process is definitely out of control. However, such a control mechanism may only be used when the process itself is "in statistical control", meaning that it did not change its main behavioural characteristics, such as the mean value or variance.

Results of the Application of the SPC Mean Chart

The control chart used within this paper relates to the mean chart of data samples. Periodically, samples of a given size (e.g. four paint cans, five bottles of beverages, seven mobile phones etc.) are taken from the production process at decided intervals. They are considered to be representative for a given production process, when this process is believed to be incontrol and with no need for any adjustments. Specific value (e.g. volume, weight, etc.) is to be measured for each item of the sample. For each sample, the arithmetic mean (average) is calculated, and after taking the last sample, two more key variables are calculated. These key values are process mean value X, evaluated as an average of sample means, and process Standard Error SE, evaluated as standard deviation of sample means (relation between SE and standard deviation σ of the whole dataset or population is

SE = $\sigma/\sqrt{n}$, where n is sample size). If the process is stable, it may be expected that most of the individual sample means lie within the range X ± 3SE (Oakland 2003), based on the assumption of a normal distribution of data.

Which signals show that the process is not incontrol? The first such signal, having the probability of occurrence of less than 0.3 % in the case of a normal distribution of data (which should be expected for a sample of 28–31 in size), is that the individual sample mean lies outside the range X ± 3SE. The second signal is when the probability that the individual sample mean lies outside the range X ± 2SE is about 4.5 %, as this makes it very unlikely for two consecutive samples to have such characteristics. This is thus considered as an indication that the process may be out of control. The third signal is when there is the case that several consecutive sample means lie at the same side of the Process Mean Value, being consistently lower or higher. With 6 or 7 such values, probability of such an occurrence drops to around 1.4 or 0.7 %, which is again a signal that the process is getting out of control. That is the basis to establish "warning lines" LWL (lower warning line) and UWL (upper warning line) at X ± 2SE and "action lines" LAL (lower action line) and UAL (upper action line) at X ± 3SE (Oakland 2003).

For this research, every single month in one year is considered as one sample, and a sequence of 30 samples for the same month are taken for the period 1961–1990. The measurement station is in Sarajevo. Temperatures were recorded by the Hydro-Meteorological Institute, and can be considered as valid. Sample means for each of the months provide another data set. Their standard deviations were calculated and they represent the SE for that specific month. Finally, this process is repeated for the whole year, month by month. The control chart is then extended with the sample means for the period 2000–2010, still using the same warning and action lines to access its variations.

The results of these evaluations are presented in the following Figs. 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 9.10, 9.11 and 9.12:

As may be seen from the control chart for the month of January during the observed period of 2000–2010, only one year (2007) had an average temperature higher than the UWL, which is not an indication that the process is out of control.

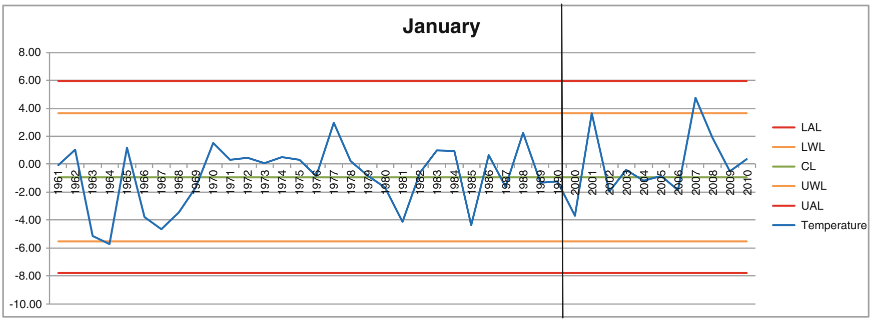

Fig. 9.1 Control chart for January temperatures

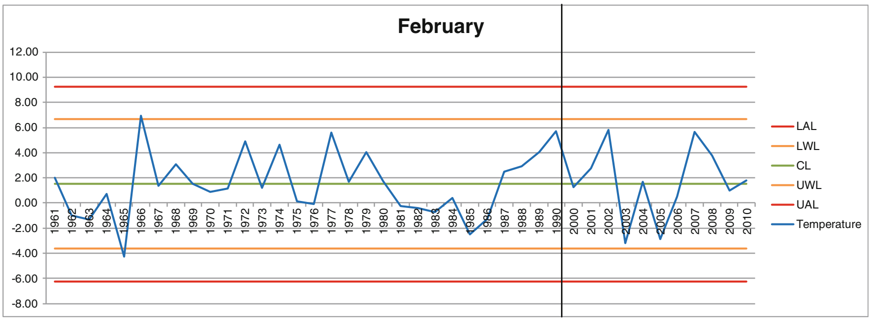

Fig. 9.2 Control chart for February temperatures

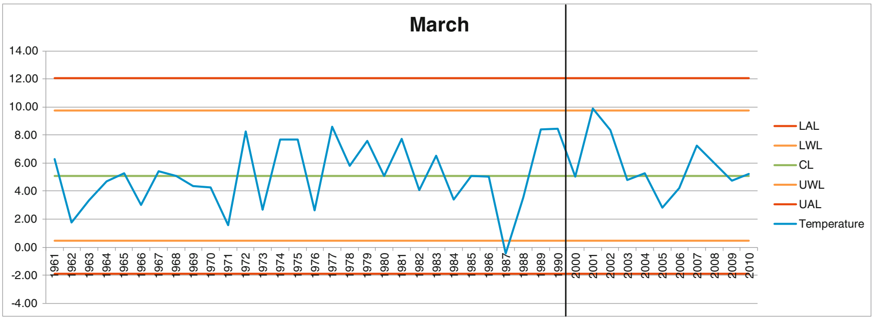

Fig. 9.3 Control chart for March temperatures

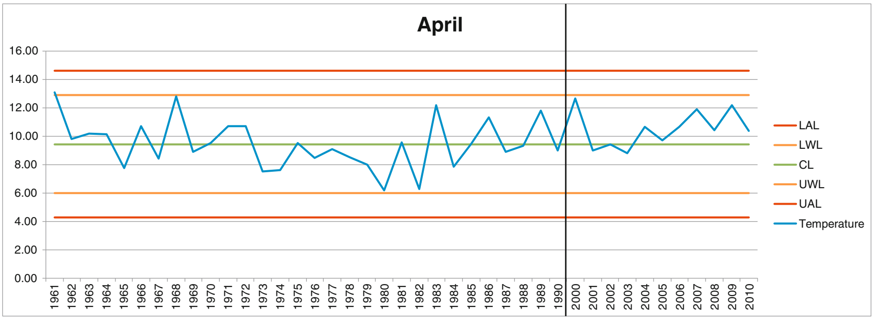

Fig. 9.4 Control chart for April temperatures

The control chart for the month of February shows that, in the period 2000–2010, the average temperature was fully within the range between the two warning lines, meaning that “the process” is fully in control.

The control chart for the month of March shows that, in the period 2000–2010, only one year (2001) had an average temperature higher than the UWL, which is not an indication that the process is out of control.

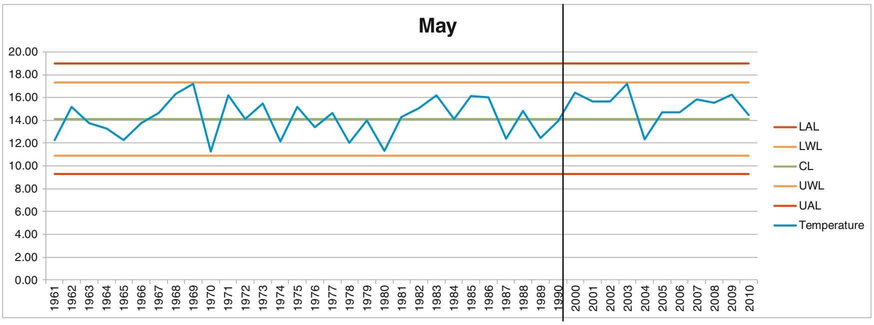

Fig. 9.5 Control chart for May temperatures

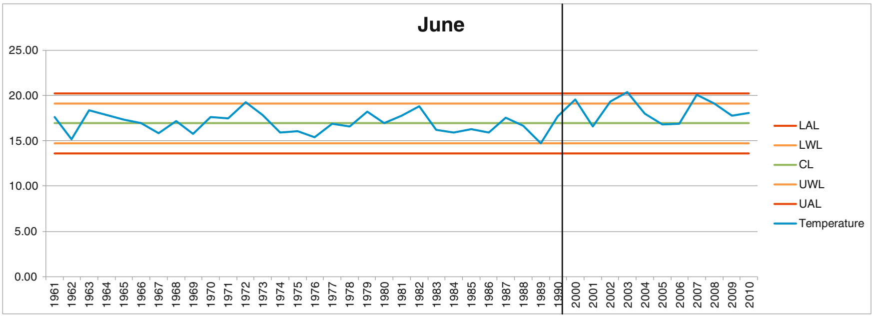

Fig. 9.6 Control chart for June temperatures

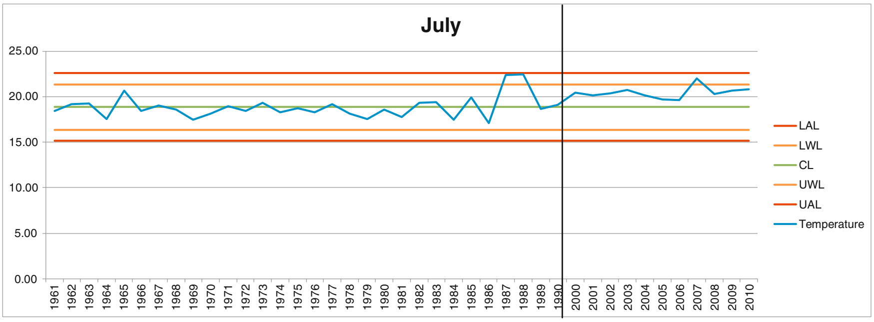

Fig. 9.7 Control chart for July temperatures

The control chart for the month of April shows that, during the period 2000–2010, the average temperature was fully within the range between the two warning lines, which again means that "the process" is fully in control.

Similar to the previous charts, the control chart for the month of May also shows that, in the period 2000–2010, average temperature was fully within the

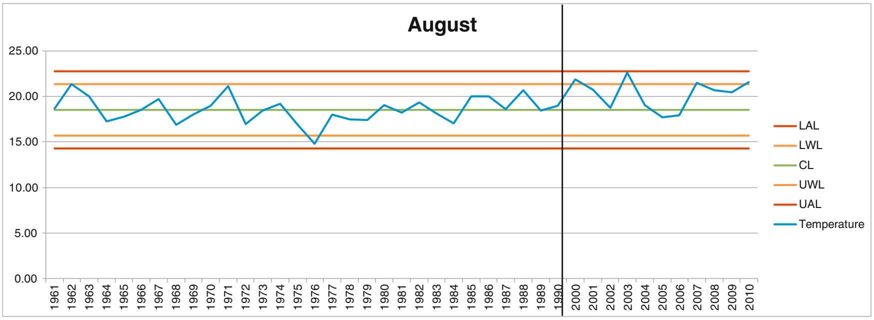

Fig. 9.8 Control chart for August temperatures

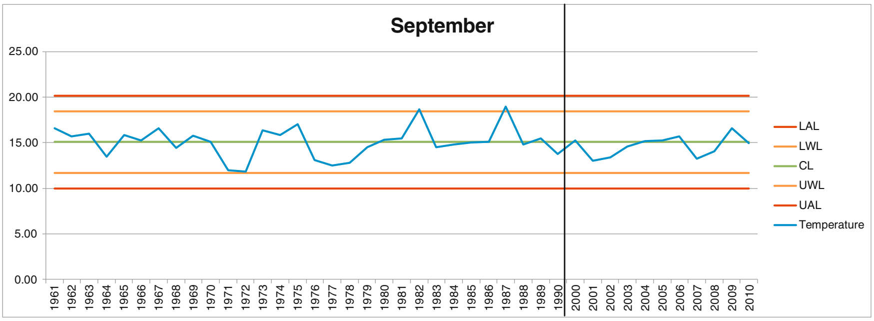

Fig. 9.9 Control chart for September temperatures

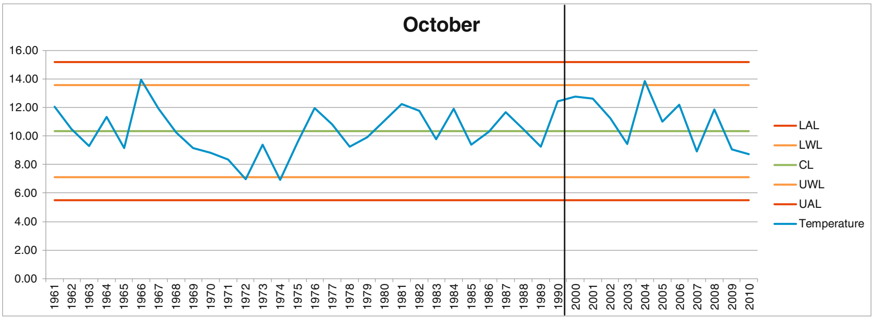

Fig. 9.10 Control chart for October temperatures

range between the two warning lines, which again means that “the process” is fully in control.

The control chart for the month of June shows different temperature behaviour during the observed period of 2000–2010. Namely, during four years (2002, 2003, 2007 and 2008) the average temperature was higher than the UWL, but this is still not an indication that the process is out of control. The average temperature from

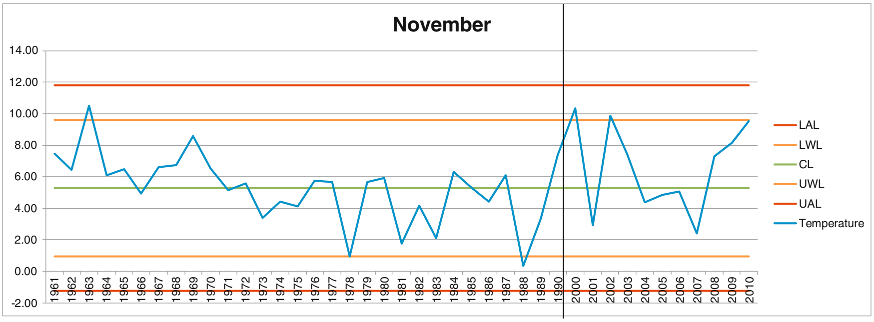

Fig. 9.11 Control chart for November temperatures

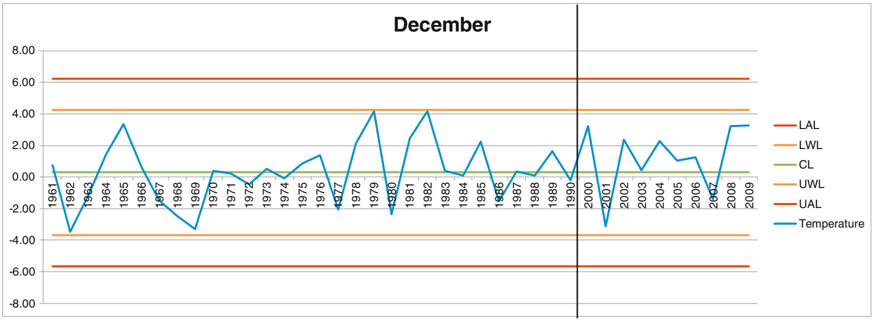

Fig. 9.12 Control chart for December temperatures

2003 was also higher than the upper action line (UAL), which is an indication that for this month and the year of 2003, the temperature has indicated possible special causes of variation, that were not existing during the period of 1961–1990.

The control chart for the month of July shows that, during the observed period of 2000–2010, only one year (2007) had an average temperature higher than the UWL, which is not an indication that the process is out of control. There was similar behaviour in the Julys of 1987 and 1988, which is not relevant for this study, even though this variation was included in the calculation of the control lines.

The control chart for the month of August again shows different temperature behaviour during the observed period of 2000–2010. During three years (2003, 2007 and 2010), the average temperature was higher than the UWL, however, this is still not an indication that the process is out of control.

Similarly to some of the previous charts (February, April, May), the control chart for the month of September also shows that, during the period 2000–2010, the average temperature was fully within the range between the two warning lines, which again means that the process is fully in control.

The control chart for the month of October shows that, in the period 2000–2010, only one year (2004) had an average temperature higher than the UWL, which is an indication that the process is in control.

The control chart for the month of November shows that during the observed period of 2000–2010, only two years (2000 and 2002) had average temperatures higher than the UWL, which is an indication that the process is still in control.

Similar to some of the previous charts (February, April, May, September), the control chart for the month of December also shows that, during the period 2000–2010, the average temperature was fully within the range between the two warning lines, which again means that the process is fully in control.

The comparison of the average monthly temperatures for the periods 1961–1990 and 2000–2010 in Fig. 9.13 clearly shows that, except for the month of September, all other months had higher average temperature values during the latter period. Variation in the average temperature increase is presented in Fig. 9.14. The average temperature increase for the whole year is 0.95 °C (red line), ranging from −0.43 °C in September to 1.58 °C in July (blue line).

Even though both Figs. 9.13 and 9.14 clearly show that the average temperature increased in the latter period observed, it may also be seen that the temperature values still mostly fall within the range of X ± 3SE, and there is only one example of a sample mean being higher than the UAL, and that is for the June of 2003.

When referring to consecutive individual sample means lying outside the range X ± 2SE in the period 2000–2010, which may be noticed in the month of June for the years 2002–2003 and 2007–2008, it may be seen that, both times, the values are actually at the UWLs (the same is the case for July 1987–1988).

When several consecutive sample means that lie at the same side of or on the centre line (being constantly higher) are concerned, the following periods have 6 or 7 such consecutive values: the month of April in the period 2004–2010 and the month of July in the whole period 2000–2010. The month of May in the period of

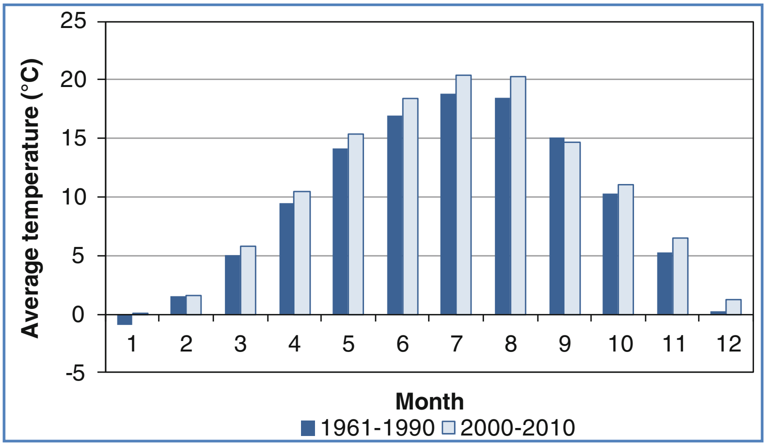

Fig. 9.13 Temperature comparison by month

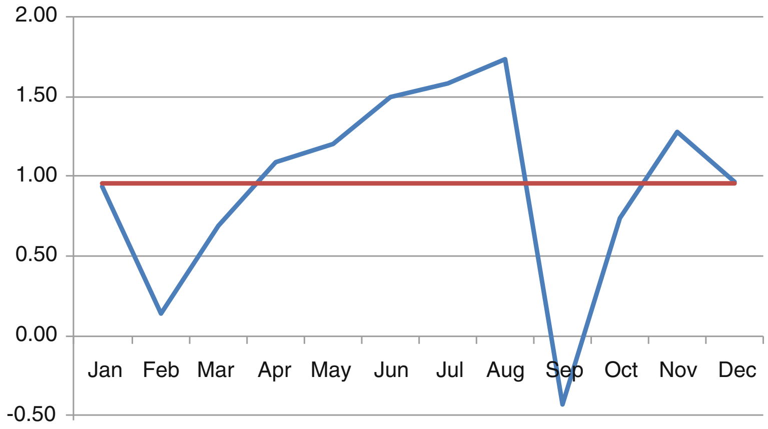

Fig. 9.14 Temperature differences between periods 1961–1990 and 2000–2010 by month

2000–2010 can also be stressed, even though the year 2004 does not meet this condition. The case of December in the period 2002–2009 is similar, where the year 2007 is an exception.

Thus, SPC clearly indicates temperature increase. However, this increase is not evenly distributed throughout the year, with the spring months of April and partly May the most affected, the summer months of June and July and the winter month of December.

Conclusion

The objective of this paper is to demonstrate the potential uses of SPC methods in evaluating variations in temperature, in order to check the most commonly accepted hypothesis that climate change is an ongoing process. The results achieved clearly show that SPC methods, in this case the sample means control chart, indicate specific behaviour of natural processes. However, the use of other types of control chart should also be implemented.

At the same time, results for Sarajevo prove that temperature increase, but it is not evenly distributed throughout the year. The behaviour indicating that the process is out of control may only be noticed during certain months of the year. This is primarily during the warmer months, April to July, and to some extent also during the winter month of December. Further research would be needed to justify why such distribution is occurring, and if it may be expected to continue as such.

This paper provided the first results using this approach, which has not been applied in this context before. It should be reapplied to data from other locations or types, e.g. precipitation. Also, other control charts, such as range charts showing differences in temperature range behaviour, should be applied.

References

Bakker A, Kent P, Derry J, Noss R, Hoyles C (2008) Statistical inference at work: statistical process control as an example. Stat Educ Res J 7(2):130–145

Benneyan JC, Lloyd RC, Plsek PE (2003) Statistical process control as a tool for research and healthcare improvement. Qual Saf Health Care 12:458–464

Committee on America's Climate Choices (2011) Board on atmospheric sciences and climate, America's climate choices. The National Academies Press, Washington, D.C

IPCC (2007) Summary for Policymakers. In: Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson CE (eds) Climate change 2007: impacts, adaptation and vulnerability. Contribution of working group II to the fourth assessment Report of the Intergovernmental panel on climate change. Cambridge University Press, Cambridge, UK

Jalote P, Saxena A (2002) Optimum control limits for employing statistical process control in software process. IEEE Trans Software Eng 28(12):1125–1133

Laušević R, Jones-Walters L, Nieto A (2008) Climate change and biodiversity in South-East Europe—impacts and action. REC Szentendre Hungary, ECNC, Tilburg, Netherlands

National Aeronautics and Space Administration (NASA) (2005) What's the difference between weather and climate?. Available at http://www.nasa.gov/mission_pages/noaa-n/climate/climate_weather.html

Oakland JS (2003) Statistical process control, 5th edn. Butterworth-Heinemann, Burlington MA (An imprint of Elsevier Science)

Spasova D, Trbić G, Trkulja V, Majstorovic Z (2007) Study for climate change impact assessment on agriculture and adaptations strategy development in Bosnia and Herzegovina. Government of Republika Srpska Ministry of Physical Planning, Civil Engineering and Ecology, Regional Environmental Center, Banja Luka

Thompson JR, Koronacki J (2002) Statistical process control: the Deming paradigm and beyond. Chapman & Hall/CRC, USA

Woodall WH (2000) Controversies and contradictions in statistical process control. J Qual Technol 32(4):341–350