Bulgarian Geophysical Journal, 2010, Vol. 36
Geophysical Institute, Bulgarian Academy of Sciences

OUTLIERS IN THE SEASONAL AND ANNUAL TEMPERATURE SERIES ON THE TERRITORY OF BULGARIA IN THE PERIOD 1900-2000

M. Syrakova

Sofia University, Faculty of Physics, Dept. of Meteorology and Geophysics, 5, J.Bourchier blvd., Sofia 1164, Bulgaria, e-mail: marsyr@phys.uni-sofia.bg

Abstract. The mean seasonal and annual temperature series at 31 stations, representative for the lowlands of Bulgaria, are checked for outliers, i.e. for observations very distant from the mean value. The data cover the period 1900-2000. Winter outliers are only negative, spring outliers are only positive. Summer, autumn and annual outliers are both positive and negative with dominance of positive outliers in summer and negative outliers in autumn. The magnitudes of the negative outliers reach grater absolute values than those of the positive outliers. The number and frequency of the outliers, except for some cases in the series, are greater in the first half of 20th century.

Key words: seasonal temperature series, outliers, Bulgaria, 20th century

Introduction

The long-term climate time series contain some values, which are rather distant from the mean value – the so called “outliers”. On the one hand, such values that are not characteristic of the bulk of the data may be erroneous, the errors being caused by corrupted measurements, misprints in the log books or other reasons. That is why the identification of outliers is the basis of the quality control work (e.g. Trenberth and Paolino, 1980; Lanzante, 1996; Peterson and Vose, 1997; Peterson et al., 1998). On the other hand, however, the outliers may be due to extreme meteorological events – extremely heavy precipitation or exceptionally cold/warm temperatures in certain year. In these cases they are completely valid values.

Here, the outliers as indicators of extreme meteorological events are considered. Temperature series at stations, representative of the lowlands of Bulgaria, are inspected. Data and method are described in the next section. Section 3 contains the results – the size, temporal and annual distribution of the outliers. In the end, the conclusions are summed up.

Data and method

The original data used in this work are the monthly mean temperature series at 31 stations distributed over the lowlands of Bulgaria, as it is shown in Fig.1. The data are provided by the National Institute of Meteorology and Hydrology (NIMH). Stations with the longest series of temperature observations, containing the least number of missing data and representative for different parts of the territory are chosen. The list of the stations with their geographical position (latitude, longitude, altitude) and the abbreviations of the names, used in Fig.1, is given in Table 1 (the arrangement is according to the latitude, from north to south).

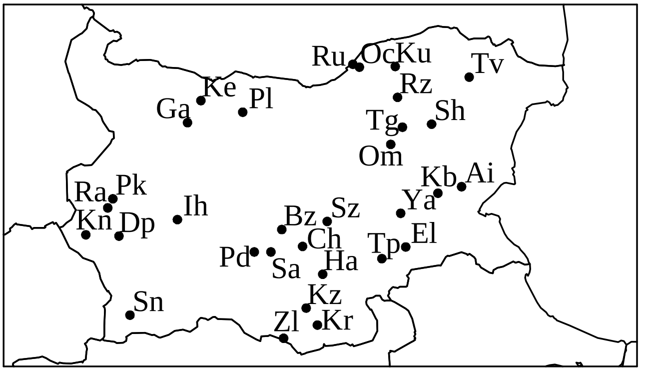

Fig. 1. Spatial distribution of the stations. For the names, see Table 1.

Temperature data, covering the entire period of 20th century, are available in very few stations (seven in the list). In fact, observations in some stations start in the end of 19th century, but in different years, that is why the beginning of these series is put in 1900. A big part of the records begin in 1920-1930, but there are stations where temperature observations are started even later – in 1940s and 1950s. Unfortunately, at some stations, observations are interrupted in 1990s. The time evolution of the number of stations is shown in Fig.2.

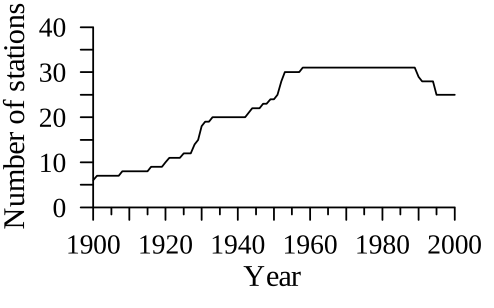

Fig. 2. Time evolution of the number of stations

A preliminary treatment of the data, before calculating the seasonal and annual values, consists in filling in the gaps. A missing monthly value would cause a gap both in the respective seasonal and in the annual series and these gaps can be found to be larger than the gaps in the monthly series. The gaps are filled in following the procedure described in Boehm (1992). It is based on the stability of differences between the series at neighbouring highly correlated stations.

Table 1. List of stations. The abbreviations of the names are used in Fig. 1.

	Station		Latitude (N)	Longitude (E)	Altitude (m, a.s.l.)	Period (year)
1.	Russe	(Ru)	43° 51'	25° 57'	37	1928-2000
2.	Obr.chiflik	(Oc)	43° 48'	26° 02'	157	1900-2000
3.	Kubrat	(Ku)	43° 48'	26° 30'	230	1949-1991
4.	Tervel	(Tv)	43° 45'	27° 25'	215	1952-1995
5.	Razgrad	(Rz)	43° 31'	26° 31'	206	1916-2000
6.	Knezha	(Ke)	43° 30'	24° 05'	117	1920-2000
7.	Pleven	(Pl)	43° 24'	24° 36'	156	1900-2000
8.	Gabare	(Ga)	43° 18'	23° 55'	120	1944-1990
9.	Shumen	(Sh)	43° 16'	26° 56'	218	1900-2000
10.	Targovishte	(Tg)	43° 15'	26° 34'	226	1952-2000
11.	Omurtag	(Om)	43° 06'	26° 25'	530	1953-2000
12.	Aitos	(Ai)	42° 42'	27° 16'	90	1925-1995
13.	Karnobat	(Kb)	42° 39'	26° 59'	196	1930-2000
14.	Pernik	(Pk)	42° 36'	23° 02'	694	1947-2000
15.	Radomir	(Ra)	42° 33'	22° 58'	691	1929-1990
16.	Yambol	(Ya)	42° 29'	26° 31'	143	1930-2000
17.	Ihtiman	(Ih)	42° 26'	23° 49'	636	1921-2000
18.	St. Zagora	(Sz)	42° 25'	25° 38'	229	1943-2000
19.	Brezovo	(Bz)	42° 21'	25° 05'	257	1952-1995
20.	Kjustendil	(Kn)	42° 16'	22° 43'	527	1900-2000
21.	Dupnitsa	(Dp)	42° 16'	23° 07'	555	1908-2000
22.	Chirpan	(Ch)	42° 12'	25° 20'	173	1928-2000
23.	Elhovo	(El)	42° 11'	26° 34'	138	1951-2000
24.	Plovdiv	(Pd)	42° 09'	24° 45'	160	1901-2000
25.	Sadovo	(Sa)	42° 09'	24° 57'	155	1900-2000
26.	Topolovgrad	(Tp)	42° 05'	26° 17'	305	1958-2000
27.	Haskovo	(Ha)	41° 57'	25° 34'	230	1900-2000
28.	Kardzhali	(Kz)	41° 39'	25° 22'	331	1930-2000
29.	Sandanski	(Sn)	41° 34'	23° 17'	203	1931-2000
30.	Krumovgrad	(Kr)	41° 30'	25° 30'	231	1953-2000
31.	Zlatograd	(Zl)	41° 22'	25° 06'	430	1933-2000

After the monthly series are completed, seasonal and annual temperature series are calculated. The four standard climatological seasons are used: winter – December, January, February (DJF); spring – March, April, May (MAM); summer – June, July, August (JJA); autumn – September, October, November (SON). Mean annual temperature corresponds to the period January-December (J-D).

The seasonal and annual temperature series are tested for abrupt changes and are homogenized using the Standard Normal Homogeneity Test (SNHT) developed by

Alexandersson (1986) and Alexandersson and Moberg (1997). The procedure of homogeneity and the results are described in Syrakova and Stefanova (2008). The homogenized series are checked for outliers.

As it is said in the Introduction, the outliers are values rather distant from the bulk of the data. In our case, these are very low/high temperatures. The identification of the outliers is based on certain threshold values, calculated for each time series – outliers are the values, which trespass the threshold. The threshold values depend on the variability of the series. Examples of using different measures of the variability of a series for calculating the threshold values are given in Peterson *et al.* (1998). The most widely used measure of the variability is the standard deviation and the threshold values are assumes to be $(\bar{x} \pm k\sigma)$, where $\bar{x}$ is the mean value of the series, σ is the standard deviation and different values of k are used in different works. According to Peterson *et al*. (1998), a common practice is using 3σ limit for temperature series. Foe example, this limit is used in Trenberth and Paolino (1980) for sea-level pressure series.

Here, the threshold values are assumed to be $(\bar{t} \pm 2.5\sigma)$, where $\bar{t}$ is the mean temperature. With the narrower 2.5σ limit instead of the 3σ limit, the number of outliers will be increased. Some of them will not be as distant from the bulk of the data as the 3σ outliers, but being at the extreme ends of the distribution they will also indicate exceptionally cold or warm months.

It is obvious that the width of the limits depends only on the standard deviation, which has well manifested seasonal variation and can be different at different stations. Because of the greater intensity of the atmospheric processes in the cold seasons, the values of σ are greater in winter than in summer. The most stable (with the least values of σ) are the annual series, because of the larger time span over which the temperatures are averaged to obtain the annual values. The limits themselves, however, depend not only on σ, but on the mean temperatures $\bar{t}$, as well, which have much higher seasonal and spatial variability. Besides, the mean temperatures depend to a greater extent than σ on the concrete time period over which they are calculated. Having in mind that not all of the temperature series inspected here have equal lengths, the values of $\bar{t}$ and σ are calculated for the period 1951-1995, when observations are available at most of the stations.

The assessment that an outlier is not an erroneous value, but is a valid extreme value is based on the comparison between neighbouring highly correlated stations. This comparison is done for the monthly series. The outliers for a given month, if they are not errors, appear in one and the same year at these stations or are in agreement with rather high/low values which lie within the limits. Some values suspected as erroneous are checked in the log books and corrected when necessary.

Results

In Fig.3, the temporal distribution of the number of outliers in the seasonal and annual temperature series is shown. Black bars represent the negative outliers, empty bars represent the positive outliers. Because of the different length of the series at different stations, the relative number of outliers instead of the real number is given as percentage of the number of stations in different years. The same approach is applied in Syrakova and Dimitrova (2009) for the monthly series as a way to put the short and long series into accordance.

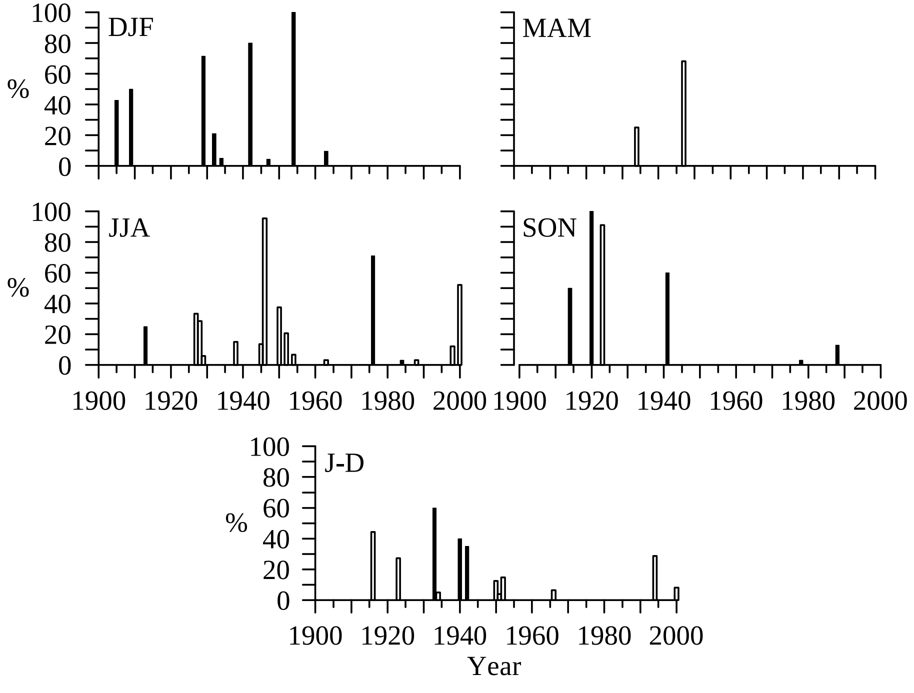

Fig. 3. Temporal distribution of the number of the seasonal and annual outliers (percentage of the number of stations). Positive – empty bars, negative – black bars.

In winter, only negative outliers are indicated. This is a consequence of the dominance of such outliers in the three winter months (see Syrakova and Dimitrova, 2009). The last year with exceptionally cold mean winter temperatures is 1963, but the stations at which these outliers are indicated are very few. In spite of the cold December and January temperatures in the end of the century (1998 and 2000, respectively) they do not emerge as outliers in the seasonal series. The year with outliers manifested at all available stations is 1954, which results from the extremely low temperatures in all three winter months, especially in February. The other two years with widespread outliers are 1942 and 1929 with outlying values at about 80% and 70% of the stations, respectively. In 1942, the seasonal outliers are consequences mainly of the low January temperatures, in 1929 – of the low February temperatures (in both cases the cold monthly temperatures emerge as outliers at all available stations).

In the beginning of the century, outliers are indicated in separate months (January, 1905 and February, 1909) at about half of the stations, but the relative number of the outliers remains the same over the whole season. This implies that the temperatures in the other winter months, although remaining within the limits, are cold enough to produce exceptionally low seasonal temperatures in many places. On the contrary, in spite of the exceptionally cold December 1948 (outliers at all stations), the seasonal temperatures remain within the limits, being compensated by the January and February temperatures (1949), which are far from the extremes. The situation is almost the same in 1947, when only one seasonal outlier is registered in spite of the well pronounced peak of outliers in January.

In spring, outliers are indicated only in two years (1934 and 1947) and there is no one in the second half of the century. In both cases the outliers are positive. At that, in 1947, the exceptionally warm spring comes after the exceptionally cold winter. The outliers in 1947 result mainly from the warm temperatures in March (widespread positive outliers) and to some extent in April. In 1934, however, the situation is different – the monthly temperatures, although within the limits in all three months (except for two outliers in April), are high enough to produce seasonal outliers at one quarter of the stations. On the contrary, other cases with monthly outliers, both positive and negative, some of which are widespread (e.g. warm March 1922, cold April 1997, cold May 1919, or warm May 1968) do not emerge as seasonal outliers, because of the mutual compensation of the monthly temperature values, which are averaged to form the seasonal ones.

In summer, outliers are dispersed over the whole century (which resembles the situation in the separate summer months, especially in June and July), but an increased number and frequency is characteristic for the second quarter of the century. Positive outliers prevail. The period of exceptionally warm summers in the second half of 20th century is typical for July and partly for August. The best pronounced peak in 1946 results from the extremely warm temperatures in all three summer months, especially in July. However, the July well pronounced peak in 1950 and the August peak in 1952, although existing in some seasonal series, are not so well pronounced as the respective monthly outliers (emerge in much fewer stations). The cases of exceptionally warm summers in the end of 1920s can also be revealed as monthly outliers (June and July 1927, July and August 1928, August 1929). The cases with positive outliers in the end of the century reflect to some extent the situation in separate months (especially in June and July) – this is the increased number of extremely warm temperatures after 1980. The seasonal outliers, however, are well pronounced only in 2000, when the July outliers are combined with relatively warm, but within the limits, temperatures in June and August. On the contrary, the widespread warm outliers in July 1988 are not so well manifested in the seasonal temperatures, because are not "supported" by the June and August temperatures.

The well pronounced peak of cold outliers in 1976 results not only from extremely low temperatures in August (cold outliers at all available stations), but is "supported" by relatively low temperatures (yet within the limits) in the other summer months, especially in June. The same effect – the influence of relatively low temperatures in all three months – produces the negative summer outliers in 1913 and the few ones 1984.

The autumn outliers are indicated in fewer years compared to the winter and summer ones and are concentrated mainly in the first half of the century, except for the few cases in 1978 and 1988. Negative outliers dominate. The best pronounced cases are the exceptionally cold autumn of 1920 (outliers at all available stations) and the exceptionally warm autumn of 1923 (outliers at 90% of the stations), the latter being the only one case (year) with positive outliers. In all cases (years) until 1950 the seasonal outliers, both the positive and the negative ones, are better pronounced (indicated at more stations) than the monthly outliers in the respective years, which implies that they result from the temperature conditions in two or all three autumn months. Something more, the outliers in 1914, which are manifested at 50% of the stations, are not indicated in any one of the autumn months. The same is valid for the few cold outliers in 1978.

On the contrary, there are cases (years) when the monthly outliers, even though widespread in certain month, do not produce seasonal outliers. For example, these are the warm spells in September 1918, 1946, 1952, and 1944 (in the last case outliers emerge at about 95% of the stations), as well as the warm spells in October 1932 (covering 100% of the stations), 1935 and 1966.

Discussing the seasonal outliers, it is interesting to notice that the extremely cold winters of 1928-1929, 1953-1954, or 1962-1963 (outliers at numerous stations, especially in 1928-1929 and 1953-1954) are followed by exceptionally warm summers, though the summer outliers are not as widespread as the winter ones.

The temporal distribution of the annual outliers shows some peculiarities characteristic for different seasons. In the first place, this is the greater number and frequency of outliers in the first half of the century. Positive outliers dominate. The tendency of high temperatures at the end of the century results from the influence of the summer season. The three exceptionally cold years are in the second quarter of the century: 1933, 1940 and 1942. There is some coincidence of years with annual and seasonal outliers. These are, for instance, the extremely warm 1923 which coincides with the warm autumn, the extremely cold 1942 – cold winter, the warm 2000 and early 1950s – compare with the summer outliers. In all these cases, however, the annual outliers are manifested at a less number of stations compared to the seasonal outliers.

In other cases annual outliers coincide with outliers in certain months, but not with the seasonal ones – for example, 1916 when the warm annual temperatures coincide with the warm December temperatures; 1933, 1940 – coincidence with the cold December temperatures; 1994 – coincidence with the warm September temperatures. The fact that these monthly outliers have not affected the seasonal ones, but annual outliers emerge in the same years, shows that the latter result from the influence of the temperatures in some other months (in other seasons).

Compared to the number of the monthly outliers the seasonal ones are much less, because the series of the mean seasonal temperatures are more stable than the series of monthly temperatures. This is a consequence of the larger time span over which the temperatures are averaged to obtain the seasonal values. Due to the same reason the annual series are more stable than the seasonal ones.

In Fig. 4, the number of outliers with opposite signs in different seasons and for the year as a whole can be seen more clearly. Here, in contrast to Fig. 3, the real (not relative) number of outliers is given. As it is said above, the winter outliers are only negative, the spring ones are only positive. In summer, warm outliers prevail, while in autumn the situation is the opposite. For the year as a whole, the number of positive outliers is greater compared to the number of negative ones. The number of the seasonal outliers is greatest in summer and one possible explanation of this peculiarity can be the narrower limits (the least values of σ) in this season compared to the other seasons.

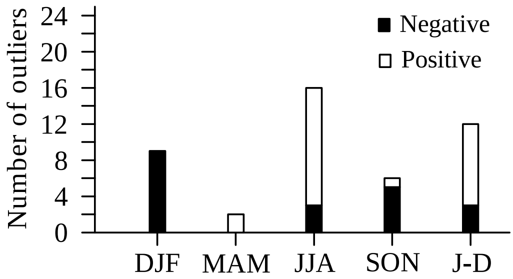

Fig. 4. Total number of the outliers in different seasonal and annual series

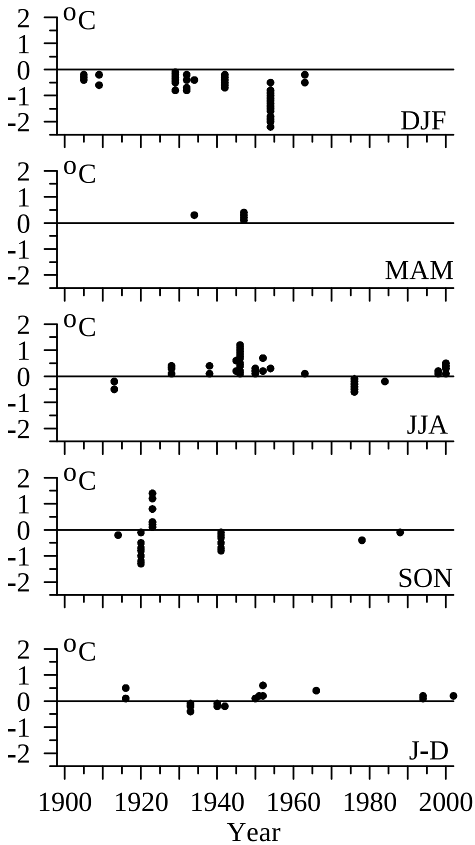

Fig. 5. Temporal distribution of the magnitude of the seasonal and annual outliers

In Fig. 5, the temporal distribution of the magnitude of the outliers, expressed as deviations Δt from the lower and upper limits $(\bar{t} \pm 2.5\sigma)$ is given. Some points represent more than one case with the respective value of Δt because of coinciding values at different stations. The magnitude of the outliers in different seasons and for the year as a whole can be easily compared in Fig. 6.

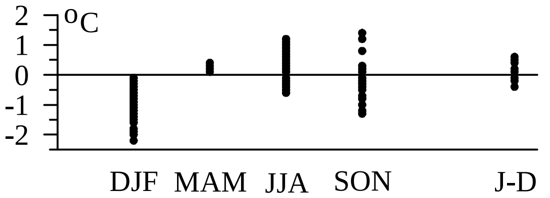

Fig. 6. Magnitude of the outliers in different seasonal and in the annual series

As the mean seasonal temperature series are more stable than the monthly ones, the magnitudes of the seasonal outliers reach smaller absolute values than the magnitudes of the monthly outliers. The negative outliers reach greater absolute values than the positive ones, which is valid for the monthly outliers, as well (Syrakova and Dimitrova, 2009). The greatest negative deviation (-2.2°C) is reached in winter 1954. In this coldest winter, outliers are registered at all available stations and the smallest negative deviation (by absolute value) is –0.5°C. No peculiarities in the spatial distribution of these outliers can be revealed. The next coldest winters are 1929 (with Δt reaching -0.8°C), 1932 (Δt also up to -0.8°C, the coldest temperatures – in the western part of the country), and 1942 (Δt up to -0.7°C, the deviations from the limits being greater at southern stations).

The next season according to the magnitude of the negative outliers is autumn. The coldest autumns temperatures are reached in 1920 (Δt up to -1.3°C, outliers at all available stations, greater deviations at northern stations) and 1941 (Δt up to -0.8°C, greater deviations in northern regions). The warm autumn outliers, although emerging in only one year (1923), are manifested at most stations and reach the greatest values (Δt up to 1.4°C) compared to the other seasons; the deviations are greatest in North Bulgaria.

The summer outliers, although emerging in more years, have smaller magnitudes. The most outstanding positive ones are in 1946 when the summer temperatures are beyond the limits at almost all stations and the values of Δt reach 1.2°C, being greater in North Bulgaria. High magnitude of Δt (0.7°C) is registered in 1952, but the outliers in this year are very few (at very few stations). The next warmest summer is in 2000 with deviations from the limits up to 0.5°C and outliers at about half of the stations (not well manifested spatial distribution). The coldest summer is in 1976 with Δt up to -0.6°C.

The season with the smallest magnitude of the outliers is spring. The year with the warmest spring is 1947 with Δt up to 0.4°C.

The annual outliers, though not very few in number, do not reach very high magnitudes compared with the seasonal ones (except for spring). The greatest deviation from the upper limit is 0.6°C, in 1952 but, as it can be seen in Fig. 3, the number of annual outliers in this year is rather small (they are registered in the western region). The next warmest year is 1916 with Δt up to 0.5°C and outlying values mainly in western region. The coldest annual temperatures are registered in 1933 with deviations from the lower limits up to -0.4°C and outliers at about 60% of the stations, better pronounced in North Bulgaria. In 1940 and 1942, in spite of the relatively great number of negative annual outliers, their magnitudes are not very high.

Summary and conclusion

The mean seasonal and annual temperature series at 31 stations, representative for the lowlands of Bulgaria, are checked for outliers. The data cover the period 1900-2000. These observations that are very distant from the mean value are due to exceptionally cold/warm temperatures in some years. The threshold values for identification of the outliers are assumed to be $(\bar{t} \pm 2.5\sigma)$, the mean temperature $\bar{t}$ and the standard deviation σ being calculated over the period 1951-1995, which is common for almost all stations.

The cases in which the outliers in a given season are manifested at all available stations (or at most of them) are very rare: the cold winters in 1954 (outliers at all stations), 1942 (outliers at 80% of the stations), and 1929 (outliers at 70% of the stations); the warm spring of 1947 (outliers at about 70% of the stations); the warm summer of 1946 (outliers at 95% of the stations); the cold summer of 1976 (outliers at 70% of the stations); the cold autumn of 1920 (outliers at all stations), followed by the warm autumn of 1923 (outliers at 90% of the stations).

Winter outliers are only negative, spring outliers are only positive. Summer, autumn and annual outliers are both positive and negative with dominance of positive outliers for summer temperatures and negative outliers for autumn temperatures. It is interesting to notice that the exceptionally cold winters in 1929, 1954 and 1963 are followed by exceptionally warm summers, but the summer outliers are much fewer than the winter ones. The number and frequency of the outliers are greater in the first half of 20th century (up to 1954). Outliers in the second half of the century are characteristic mainly for the summer and annual series with prevalence of positive ones.

The magnitudes of the negative outliers expressed by the deviations from the limits reach greater absolute values than those of the positive outliers. The greatest negative deviation is reached in the winter series of 1953-1954, followed by the autumn series of 1920. The greatest positive deviation is reached in the autumn series of 1923, followed by the summer series of 1946. The deviations in the annual series reach smaller values compared to the seasonal series. The warmest year as a whole turns out to be 1926 (widespread outliers reaching almost the highest magnitude of the deviations from the limits), the coldest one is 1933 with the exceptionally cold temperatures better pronounced in North Bulgaria.

Acknowledgments. The author wishes to thank the National Institute of Meteorology and Hydrology (NIMH), Bulgaria, for providing the temperature data and Prof. D. Syrakov for the computer programs used in this work and Mrs. P. Dimitrova for her kind assistance in preparing the data for the analysis.

References

Alexandersson, H., 1986. A homogeneity test applied to precipitation data, *J. Climatol*., **6**, 661-675.

Alexandersson, H., and A. Moberg, 1997. Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends, *Int. J. Climatol*., **17**, 25-34.

Boehm, R., 1992. *Description of the Procedure of Homogenizing Temperature Time Series in Austria*, CERI. Proj. Gr. Met. WP2, Vienna, 14 pp.

Lanzante, J.R., 1996. Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde stations data, *Int. J. Climatol.*, **16**, 1197-1226.

Peterson, T.C., and R.S. Vose, 1997. An overview of the Global Historical Climatology Network temperature database, *Bull. Am. Meteorol. Soc*. **78**, 2837-2849

Peterson, T.C., R. Vose, R. Schmoyer, and V. Razuvaev, 1998. Global Historical Climatology Network (GHCN) quality control of monthly temperature data, *Int. J. Climatol.*, **18**, 1169-1179.

Syrakova, M., and M. Stefanova, 2009. Homogenization of Bulgarian Temperature Series, *Int. J. Climatol.*, **29**, 1835-1849.

Syrakova, M., and P. Dimitrova, 2009. Outliers in Bulgarian temperature series in the period 1900-2000, *Bulgarian Geophysical Journal*, **35**, 63-76.

Trenberth K.E., and D.A. Paolino, Jr, 1980. The Northern Hemosphere sea-level pressure data: Trends, errors and Discontinuities, *Mon. Wea. Rev.* **108,** 855-872.

Екстремни стойности в сезонните и годишни температурни редици на територията на България за периода 1900-2000 г

М. Сиракова

Резюме. Редиците от средни сезонни и годишни температури в 31 станции, представителни за извънпланинската част на България, са проверени за екстремно големи стойности, т.е. за стойности, твърде отдалечени от средната. Данните обхващат периода 1900-2000 г. В зимните редици отклоненията от граничните стойности са само отрицателни, а в пролетните – само положителни. В летните, есенните и годишните редици има както положителни, така и отрицателни отклонения, като в летните преобладават положителните, а в есенните – отрицателните. Отрицателните отклонения достигат по-големи абсолютни стойности от положителните. Броят и честотата на екстремните стойности, с малки изключения за летните редици, са по-големи през първата половина на XX век.