20TH CENTURY VARIATIONS OF THE SOIL MOISTURE CONTENT IN EAST-HUNGARY

SZILVIA HORVÁTH

PHD DISSERTATION

THESES

SUPERVISORS:
LÁSZLÓ MAKRA AND JÁNOS MIKA

SZEGED

2003

Scope and objectives

Hungary and the eastern part of the Great Hungarian Plain have always been characterised by significant hydrological extremities. Drought and inundation events followed each other, sometimes even within a year. On the other hand, couple of decades proved to be either dry or wet during the 20th century.

The most serious inundation events occurred between 1939-1942 and extended over 570,000 hectares. The whole region threatened by inundation exceeds 1.8 million hectares, which reaches nearly 60 % of the cultivated lands of the Great Hungarian Plain (Pálfai, 2001). Considering the 20th century, droughts at the beginning and the middle of the 1900s, as well as those in the first part of the 1990s were the strongest ones. Furthermore, the drought of similar strength in 2000 can also be listed here. Frequency of drought and inundation events are similar in the Great Plain (total area: 45,000 km^2, cultivated lands: 32,000 km^2); however, typical spatial extension of the former phenomenon is much larger than that of the inundation (Pálfai, 2001).

The aim of *the Dissertation* is to give a modern climatographical analysis on the varied hydrometeorological relations of the region, based on reliable observations of meteorological stations. The analysis includes statistical characteristics of the inter-annual variability, spatial and temporal correlation of the available soil moisture content and long-range changes, as well as their possible relation with climatic trends for greater regions, respectively. The object and – hopefully – result of our work is to help further inter-disciplinary study of agro-ecological problems, influenced by the soil moisture content, by performing objective regionalization based on soil moisture anomalies in the region; by publishing characteristics and calendar of the objective year-types; and, also, by determining climatically representative, long dry and wet periods.

The above-mentioned aims are intended to realize on the basis of the *Palmer's Drought Severity Index* (PDSI) data series, known world-wide in the agroclimatological analysis (Palmer, 1965; Alley, 1984; Briffa et al., 1994; Scian and Donnari, 1997; Cook et al., 1999). Monthly PDSI data series of five stations were determined for the 20th century (99 years), in three versions: *a)* plant-covered surface (maize plant) and homogenised (i.e. against possible non-representative peculiarities of the stations or the observations); *b)* bare (not plant-specific) surface, homogenised; and *c)* plant-covered surface, but non-homogenised meteorological data. Among them, we consequently consider the first alternative, i.e. the homogenised data sets for plant-covered surface as the most important, basic version. Analysis of the other two versions aimed to examine whether our consequences could be influenced by the homogenisation and by some hidden features connected to the evapotranspiration of the selected plant.

The following questions were answered in the Dissertation:

- ❖ Is the seasonal and spatial independence of the PDSI, elaborated for another region, valid in the Great Hungarian Plain, too?
- ❖ Can the distributions of the monthly PDSI sets be considered as normal ones?
- ❖ Have the PDSI series got significant correlation to parallel independent estimations of the soil moisture content, according to which the index can be reliably interpreted as a soil moisture indicator?
- ❖ Do the PDSI anomalies show regional differences or definite structures in the examined region of Eastern Hungary (36,000 km^2)?
- ❖ Until what time-lag is the autocorrelation, coming from the recursive definition of the PDSI, significant, and what kind of year-types can be defined on the basis of the significant autocorrelation?
- ❖ What kind of slow changes occurred in the course of PDSI in the 20th century?
- ❖ Did this local changes show statistical connection to the synchronous temperature characteristics of the Northern Hemisphere?
- ❖ Did any sufficiently long periods, occur in the 20th century, PDSI indices of which differed significantly from those of the whole period in the given month?

Data basis

Most of our analysis was performed on the Palmer's Drought Severity Index data sets of five meteorological stations found in Eastern Hungary (Miskolc, Nyíregyháza, Debrecen, Kecskemét and Szeged) for the period between 1901 and 1999. Input data of the index consist of monthly mean temperatures and monthly precipitation totals. The data sets mentioned above are disturbed by inhomogeneity, coming from different time of observations, from transfer of the stations, and from other factors. Hence, the Hungarian Meteorological Service, who observed and collected the data, decided to homogenise the monthly temperature and precipitation data sets. The basis of our calculations was the homogenised monthly data sets (Szentimrey, 1999).

The given process of homogenisation compares the anomalies of ten Hungarian meteorological stations, having long data sets. It supposes that, according to the break point analysis, significant breaks of the examined series, relative to the course of data at the neighbouring stations, are not natural climatic developments. They are considered as inhomogeneities, found in the circumstances of the measurements, which have to be corrected. For comparison, not only the homogenised data, but the original ones were used, as well.

The basis of the analysis is the Palmer's Drought Severity Index, produced from the monthly temperature and precipitation data sets (Palmer, 1965). The PDSI was elaborated by its author for the climatic conditions of North-America, in order to develop a numerically defined index, which is partly independent from local and seasonal features, and which partly shows the extremities of the soil moisture content always and everywhere on the same scale. (Owing to the different climatic conditions, the examinations started with the questions of place- and season-independence and with interpretation of the PDSI as a parameter of the soil moisture content.)

The Palmer-index numerically defines the soil moisture anomalies of the given month in five steps, beginning from the PDSI value of the preceding month. This PDSI value is multiplied by 0.9 and added to a term which is proportional to the difference between the precipitation of the given month and the computed precipitation that could definitely keep the actual soil moisture conditions unchanged. Proportionality coefficient weighting this difference before adding to the PDSI is a temperature-related empirical factor calculated for ~~the~~ climatic conditions of the USA.

The above-mentioned steps of the calculation are performed in order to determine the precipitation amount, which is needed for preserving the soil moisture content. The stages of this process are the calculation of the elements of the potential (maximal, if there is enough water content) water balance and then establishing multiplicative factors for making normalisation, against the differences of local and seasonal climates.

A crucial point of calculating the PDSI is the choice of the formula and the reference-plant to estimate the potential evapotranspiration, which is performed in two versions. The basic solution was to choose the Blaney-Criddle's plant-specific procedure, by considering maize on the soil. This plant is also characteristic for our region, and its evapotranspiration is fairly similar to that of many other plants, due to its quickly extending leaf surface. As an alternative, the Thornthwaite's plant-independent PDSI series, based on climatic calculations, were also taken into account.

Considering the five stations used for representation of the region, the soil-types seemed to be not important to differ. Hence, the water capacity of 150 mm was taken into account in all versions of PDSI computation.

For studying the spatial correlation, the five stations are not enough, of course. Consequently, short PDSI sets of 17 stations were calculated for the period between 1951 and 1992, with the Thornthwaite's plant-independent method and, in proper sense, without homogenisation. Differences among the arable land's water capacities of the characteristic soil-types were taken into consideration at these 17 stations.

Besides the PDSI, two further data sets estimating the soil moisture content kindly provided by the authors, were used. Both sets use various combinations of the daily meteorological elements, influencing the water balance of the surface. Monthly correlation coefficients between the monthly values of the two series are between 0.7-0.9 in the most part of the year, while they are only between 0.4-0.7 from May to September. The Dunkel's data set (1994) was used between 1901-1990, while that of Bussay (1989) was applied between 1901-1988. These parallel series were used to certify the PDSI as a quantitative indicator of the soil moisture content, and to establish the conversion of the PDSI into the soil moisture, as a physical parameter.

Besides these parameters, a version of the Pálfai's Drought Severity Index was produced and used, operating with the monthly temperature and precipitation data, without numerically defined corrections of the daily extremities (Pálfai, 1990). The author, himself also suggests it as the PAI_0 Index. This index is a favourable tool for agro-hydrological examinations, the only disadvantage of which is that each year is characterised by one number, only, integrating the period from October to the following August. The PAI_0, as a further independent index, supports our conclusion that PDSI can really be related to the soil moisture content.

Finally, at one point of our analysis, when studying the connection of the soil moisture content with the regional climatic anomalies, the data sets of the hemispheric mean temperatures and the continent-ocean air temperature contrast were also used (Jones et al., 2000; Mika, 1988).

Methods

In the dissertation, the classical uni- and multivariate methods of the mathematical statistics are applied. Firstly, the first and second momentums of the monthly PDSI sets are determined, then the place- and season-independence of the latter ones are examined with *F*-test, pair by pair. (The averages differ from zero because the reference period is not the whole 99 years but only the first 80 ones.)

The monthly PDSI sets, amalgamating all months and stations in one sample, as it is known from the special literature, do not fit the normal distribution. When dividing the whole set into monthly subsets for the separate stations, their normality is improving, examined by the Kolmogorov-Smirnov- and the χ^2-tests.

The spatial and temporal correlation coefficients, calculated for short data sets of 17 stations in the region, are all significant when using the *Z*-test. On the basis of them, objective regions are determined using the rotated factor analysis. The results are compared to the classification of the hierarchical cluster analysis based on Ward-method, using the Euclidean distances. The cluster analysis is also applied for determining objective year-types utilising the significant temporal autocorrelation. However, in the latter case, the method of *k*-means is chosen to classify the years into three year-types – dry, medium and wet.

The smoothing for detection of slow changes in the long data series is performed by the traditional moving averages and by the Gaussian-filter, which latter does not deform the effect of any intermediate frequencies. The calculation window was 11 years in both cases.

After all, 1-1 methods of the supervisors are applied, which do not exceed the field of the methods represented in the first sentence of the current chapter. The parameters of the local changes and the hemispheric temperature; namely, the connections between the mean temperature in the Northern Hemisphere and the continent-ocean air temperature contrast are defined numerically by the method of “slices” (Mika, 1988).

The essence of the method is slicing the original data sets into sub-periods of same length (5, 9, 13, 17 and 21 elements) and then, using the averages of the sub-periods, a regression analysis is performed.

The method of "slices" is applied to the data sets between 1901 and 1988 and the conservation of the coefficients is tested for the independent period of 1989-1999. The correlation between the two hemispheric variables is negligible in the basic period, which makes it possible to avoid the so called "multi-collinearity" of the two independent variables. The regression coefficients are estimated by the method of the "least squares", while their statistical significance is checked with the Student's *t*-test.

Both the examination of the slow changes and the regression analysis performed with the hemispheric temperature characteristics referred to the substantial changes in the values of the PDSI, parallel with the global warming, considering the 20th century. Hence, the closing chapter of the dissertation aimed to select climatically representative long periods, which can be used for impact studies on any problems depending on the soil moisture content.

In order to identify dry or wet sub-periods, averages of which are significantly lower or higher than that of the whole 99-year long data set, a new interpretation of the classical two-sample test was performed (Makra, 2000). The basic question of this test is whether or not a significant difference can be found between the averages of an arbitrary sub-sample of a given time series and the whole sample.

Normality of the distribution of monthly PDSI, which is found to be valid only partially (see point 2 of the *Results*), is a satisfactory condition to the application of the method. At the same time this is not a necessary condition, since in case of very large samples (99-99 data) the distribution of the density function for the sum of the elements is nearly normal, apart from the distribution of the basic sample. Besides that, a further condition is that the random variables be independent, which is realized for the succeeding PDSI values.

The most part of our analysis was fulfilled on the monthly PDSI data sets of the whole year, while a smaller part of it was made on the four selected months of the growing season (April, June, August and October). The odd months in the latter case were omitted considering the strong and significant autocorrelations.

The calculations were performed using the Windows-based, STATISTICA 5.1 and SPSS 9.0 software products. When preparing the data and drawing the figures, the EXCEL 5.0 software was also used.

Results

In the first eight thesis points the results are summarised, based on the Palmer's Drought Severity Index data sets, calculated by the Blaney-Criddle method. Input data of this calculation are homogenised temperature and precipitation data series.

The differences of the PDSI data, generated by the Thornthwaite's method of the potential evapotranspiration and homogenised data, from the above-mentioned sets are represented in the thesis point 9. On the other hand, the comparison of the PDSI based on partly homogenised, partly non-homogenised data, both calculated by the Blaney-Criddle's method, are found in the thesis point 10.

The results for the whole year are contained in the thesis points 1-5, while those for the growing season between April and October are found in the thesis points 6-8. Discussions of both the thesis points 9 and 10 follow this structure.

1. Place-independence of the PDSI, as a desirable condition when elaborating the index for different climate conditions, is realised for the monthly values of the five stations in the Great Hungarian Plain. According to the *F*-test, the standard deviations of only one pair of stations differ significantly, which is far below the random proportion.

Season-independence of PDSI, as another desirable feature, is only partly realised in the Great Plain, since the proportion of significantly different standard deviations is three times higher than that comes from the pure chance. Standard deviation of the indices is considerably lower in May and June, than those of the other months. On the other hand, none of the standard deviations of the other ten months differ significantly from each other.

2. When putting the index values of *i)* the five stations, *ii)* the twelve months, *iii)* both aspects into one sample, distribution of the PDSI differ significantly from the normal one in all of the three cases. The main reason of this result is the rare occurrence of the near-zero values, which is a well-known feature of the index experienced in other regions, as well. On the other hand, when analysing the PDSI sets of each station and month separately, then, according to the Kolmogorov-Smirnov test, distributions of all the 5 x 12 = 60 samples are normal; while, according to the χ^2-test, distributions of only ten samples are not normal at the 95 % probability level. This value (ten from the all sixty) is three times higher than the chance; nevertheless, the distribution of the samples for the most stations and months can be considered normal.

3. The PDSI, as a standardised index without unit, has a close relation with the two (Dunkel's and Bussay's) calculations of the monthly soil moisture content, as well as with the Pálfai-index, characterising the water supply of the growing season from October to next August. According to this, the PDSI can correctly be interpreted as a characteristic of the soil moisture content. Furthermore, on the basis of the regression coefficients of the linear connection, the index values can be expressed in physical unit of the water content of the upper one meter soil layer. Hence, unit increase of the PDSI, corresponds to 8-19 mm surplus according to the Dunkel's, and 5-17 mm surplus concerning the Bussay's soil moisture content, respectively, depending also on season and place.

Further feature of the regression coefficients, referring to the deep sense of the PDSI, is that if they are divided by the standard deviation of the soil moisture content, the following coefficients having no units are received: 0,35 ± 0,05 in the data sets of the Dunkel's soil moisture content, while 0,25 ± 0,07 in those of the Bussay's one, respectively. Consequently, unit change of the PDSI in each month and station equals to almost the same change in the unit of the soil moisture content standardised by the standard deviation.

4. The spatial correlation of the PDSI in each of the 15 station-pairs formed of the five stations and the 12 months exceed the significance threshold value of 0.3 but do not reach the value of 0.9 in any case. From November to April all vales are over 0.6, while from May to October they are lower than the winter values. In this period even the hydrological extremities of the region develop partly independently from each other.

   The significant spatial correlation make it possible to determine objective sub-regions for the examined area. In the enlarged network with 17 stations, the rotated factor analysis of the Thornthwaite's non-homogenised index series with shorter data sets results in at least two sub-regions in each month. The sub-regions are located along the north-east – south-west axis; whereas, more stations and larger area belong to the southern sub-region. At the same time, during the seasonal alternation from drying out to filling up, between September and November, the procedure isolates an intermediate sub-region, too (Horváth, Sz., 2002: Physics and Chemistry of the Earth, 27, 1051-1062.; Horváth, Sz., Makra, L. and Mika, J., 2000: The 20th Conference of the Danubian countries on Hydrological forecasting and the hydrological basis of water management. Bratislava, Slovakia, 313-320; Horváth, Sz., Makra, L. és Mika, J., 2001: 1st Hungarian Conference of Geographers. Szeged, Hungary.). The independent method of the cluster analysis also confirmed the classification of the stations to the sub-regions received from the factor analysis.

5. The high autocorrelation, coming from the recursive definition of the of PDSI, remains significant even at the time difference of two (0,7-0,9) and six months (0,3-0,7). Within these ranges, the somewhat lower values of the autocorrelation in the summer half-year come from the higher variability of convective precipitation.

   The high autocorrelation make it possible to define so called year-types objectively. The twelve months from November to the next October were classified into three types by cluster analysis. Each type occurred in a considerable part (14-57 %) of the 98 examined years. At all stations 1-1 year-types can be isolated, each month of which is definitely dryer or wetter than the average. At three stations, each month of the third type is around the average, while this type is characterised by a wet summer after a dry winter at Miskolc and Debrecen. The knowledge of the fact that a given month of a given year fell into which year-type, reduces the monthly variance of the PDSI with somewhat more than 50 %, as an average (Horváth, Sz., 2002: Physics and Chemistry of the Earth, 27, 1051-1062.).

6. The PDSI data sets, calculated for the even months of the growing season, show significant linear trends only at three stations in April in the whole period of the 99 years. At the same time, the trends for each month and station are negative. Considering a hundred year time-span, its values are -0.7 and -3.1, respectively. Namely, the 20th century was characterised by slow drying out. This definite tendency is reflected also in the data sets, smoothed with the moving average and Gaussian filters, detecting the non-monotonous and non-linear details of the slow inter-annual fluctuation.

7. When searching for regression connection between the PDSI data sets in the growing season, on the one hand, and, the average temperature in the northern hemisphere and the continent-ocean temperature contrast, on the other hand, then 80 % of the months and stations show significant connection with at least the one hemispherical variable. The most coefficients, relating to

the hemispheric temperature, are negative. Namely, the drying out in the basic period between 1901-1988 is connected not only with the time, but the warming characterising the average of the northern hemisphere, too.

The partial regression coefficients relating to the continent-ocean contrast (the rate of the warming) are mostly negative, as well. The significant coefficients calculated for a hemispheric warming of 0.5°C involve a decrease between -0.5 and -2.9 in the values of the PDSI, if the connection can generally be related to other periods, too. At the same time, this assumption is not supported by the behaviour of PDSI in the following independent 11 years (1989-1999), when the agreement of the actual PDSI anomalies and those calculated with the regression is not better than the chance (Horváth, Sz., 2002: Physics and Chemistry of the Earth, 27, 1051-1062.).

8. The break-point analysis makes it possible to detect sub-periods, PDSI averages of which differ significantly from that of the examined 99 years. Considering the growing season, the length of these significant sub-periods is many decades in the most of the examined months and stations. The maximum difference of the sub-period's averages from that of the whole data set decreases according to an exponential function, as the length of the sub-period increases. There is a number of sub-periods characterised by significant deviation from the overall mean being higher than the unit of the PDSI. Significant wet sub-periods are detected in the first half of the 20th century, while dry breaks occurred in the second half of the century. The averages of significant sub-periods with opposite signs show even more definite difference, than those between one-sided deviations and the overall mean (Makra, L., Horváth, Sz., Pongrácz, R. and Mika, J., 2002: Physics and Chemistry of the Earth, 27, 1063-1071; Horváth, Sz., Makra, L. and Mika, J., 2000: The 20th Conference of the Danubian countries on hydrological forecasting and the hydrological basis of water management. Bratislava,

Slovakia, 313-320; Horváth, Sz., Makra, L. és Mika, J., 2001: 1st Hungarian Conference of Geographers. Szeged, Hungary).

9. The difference of the homogenised PDSI data set, based on the Thorntwaite's potential evapotranspiration from that calculated with the Blaney-Criddle method (see: the chapter of Results, points 1-8.), are generally not substantial, but they can be summarised as follows:
   - i) *The standard deviations in each month are somewhat higher than those determined with considering plants. There is no significant difference between the standard deviations of the month-pairs even including May and June.*
   - ii) *Distributions consisting of the largest data sets (5 stations* 12 months*99 years; 5 stations, separately, 12 months*99 years) differ significantly from the normal distribution. The most monthly distributions per stations can be considered normal; nevertheless, the number of those monthly data sets, distributions of which are not normal, is definitely higher than the chance.*
   - iii) *The strong linear connections between these PDSI data sets and the soil moisture content are to some extent weaker in the most part of the year; though in May and June they are stronger.*
   - iv) *The most spatial correlation coefficients are higher than those determined by considering plants.*
   - v) *The most of the temporal correlation coefficients are higher than those determined by considering plants.*
   - vi) *The linear trends for the whole period decrease more intensely and the drying out according to the smoothed curves is steeper in this version. In other words, existence of plants slightly reduced the drying out in the 20th century.*
   - vii) *The regressions relating to the hemispheric temperature are mostly negative in this version, too. The significant coefficients refer to a drying out higher than 20 %.*

viii) The limits (the starting and the ending years) of the significant sub-periods are not independent of the method of calculating the PDSI. The ratio of the concurrently significant sub-periods in both versions is very little. Hence, before the possible application, it is worth to choose with care which method of PDSI to use for identification of the analogous periods.

10. The difference of the indices, using the Blaney-Criddle method and non-homogenous data series, from the above-mentioned basic behaviour are generally not substantial, but they can be characterised as follows (see: the chapter of Results, points 1-8.):

i) There is no important difference comparing with the results for the standard deviations of the homogenised data set.

ii) The distributions consisting of the largest data sets differ significantly from the normal one in this case, too. The most of the monthly distributions per stations can be considered normal; nevertheless, the number of those monthly data sets, distribution of which is not normal, is definitely higher than the chance in this case, as well.

iii) The correlation coefficients between these PDSI data sets and the soil moisture content are stronger in this version, the probable reason of which is that this latter estimation is also performed on non-homogenised data.

iv) The most of the spatial correlation coefficients are lower than those calculated for the homogenised data sets.

v) The most of the temporal correlation coefficients are somewhat lower than those calculated for the homogenised data sets.

vi) The linear trends for the whole period decrease to a lesser extent and the drying out, according to the smoothed curves, is also slighter in this version. In other words, the given homogenisation increased the drying out characterising the 20th century.

vii) *The regressions relating to the hemispheric temperature are mostly negative in this version, as well. However, the significant coefficients refer to a drying out lower than 20 %. Hence, the homogenisation increased the drying out in this respect, too.*

viii) *The limits (the starting and the ending years) of the significant sub-periods are influenced by the homogenisation. The ratio of the concurrently significant sub-periods in both versions is very little. Hence, before the possible application, it is worth to choose with care which method of PDSI to use for identification of the analogous periods.*