Arbeitsberichte der Schweizerischen Meteorologischen Zentralanstalt
Rapports de travail de l'Institut Suisse de Météorologie
Rapporti di lavoro dell'Istituto Svizzero di Meteorologia
Working Reports of the Swiss Meteorological Institute

Zürich

No. 29

TGS : Geo-Synoptical Quality Control of Transformed Climatological Data

(A proposed research program for the Swiss National Foundation for Scientific Research)

by

D. Elbashan
(Israel Meteorological Service Bet Dagan)

Climatology 551.501.45 : 551.583.15

September 1971

SUMMARY

Quality control of climatological data is done nowadays, at least partially, with the use of computers. The methods used for this control vary from simple-logical computer checks for finding suspicious values (such as the occurence of rain without the presence of any clouds), to the manual screening of data arranged in a geo-synoptical array (such as geographic listing of the maximum temperatures recorded on the same day at nearby stations).

The present proposal aims at using the computer itself also for a more thorough screening of this geo-synoptical array. This could be materialized if instead of listing the absolute data (temperatures in degrees, etc.) the computer will compute and screen transformed (z) data, such as standardized values. These z values are more easily comparable since they are free from most of the meteorological and geographical factors influencing the areal distributions of climatological elements on a specified time. Thus any kind of wrong data could be easily pointed out.

1. Introduction

As the present proposal deals with certain aspects of the variability of climatological elements, it is worthwhile to discuss the various causes of this variability.

1.1. The variability of climatological elements

To understand a certain measurement, taken in a specified place and time, the measured value (E) can be written as:

$$E = E_o + e^* + e_1 ,$$

where

- E_o - the long-term average annual value at that point in space;
- e^* - influences of technical factors on the reading;
- e_1 - astronomical and meteorological influences:

$$e_1 = e_{LT} + e_A + e_D + e_X , \quad \text{where}$$

- e_{LT} - the long-term climatic trend or fluctuation;
- e_A - the annual climatic cycle;
- e_D - the diurnal climatic cycle;
- e_X - all the other influencing factors, such as the meteorological process.

The e_A and e_D factors are well known, so their effect on the variation of E can be predicted. e_{LT} deserves more elaborate statistical treatment for predicting its effect on the variability of E; but e_X is little known quantitatively: It is effected mainly by the varying synoptic situations and perhaps by other unknown factors also. For illustration, a climatological series is presented in part A of the following table. This series is, by definition, free from the e_{LT}, e_A , e_D and e* influences:

A Temperatures at station A on 14 hr., Nov. 1st (^oC)	Year / reading no.	B Laboratory readings of a certain length (mm)
23	1	23
27	2	27
24	3	24
26	4	26
28	5	28

Part B of this table is given for comparison. Although the values are identical - the meaning of the variabilities are completely different. The cause for the variance of series B can be technical only and can be eliminated completely by technical improvements. In series A, on the other hand, such variations are characteristic, and reflect the variability that is inherent in the climatological process itself (Thom, 1954) - as the weather in the 1st year was rainy, in the 2nd - hot and dry, etc. Thus this inherent variance (e_X) must be accounted for before any further use is made of the climatological series.

The above background and definitions will be referred to in the following text to clarify the techniques suggested.

1.2. The purpose of the research program

As will be explained later (para 2.1.), quality control of climatological data is a very complicated task, even in this era of electronic computers and sophisticated schemes of work, because of the large amount of factors involved directly or indirectly - astronomical, geographical, topographical, technical and meteorological. The purpose of this research program is to carry this quality control problem one step forward by controlling and checking not the actual (absolute) value but some pure (transformed) index of this value, when these values are arranged in a geo-synoptical array.

1.3. The climatological concept of transformed data

The idea of using transformed values for climatological purposes is not new. The following examples will clarify the need and the practical use of such transformations.

1.3.1. The Isomer

For studying the annual cycle of rainfall over large geographical areas transformed values of the monthly rain amounts have been used instead of the absolute amounts, because the latter are influenced by many geographical and technical factors which are irrelevant in regard to studying the cycle:

The transformed value used is the Isomer (Elbashan, 1966) defined as

$$I(\%) = \frac{\text{monthly amount}}{\text{annual amount}} \, 100$$

in which geographical variations related to altitude, local features etc., are almost completely eliminated. The geographical factors beeing thus accounted for, this transformed value has a very low geographical variance - contrary to the variance of the absolute value. This characteristic can be made good use for further climatological work - including geo-synoptical quality control, prediction and interpolation.

1.3.2. The Relative Temperatures

For studying the annual cycle of temperature(Conrad L. Pollack, 1962) another transformed value - which is also free of the geographical influences - was introduced by W.Köppen. It is called "Relative Temperature" and it is computed as:

$$\text{Rel. T } (\%) = \frac{T_i - T_{min}}{T_{max} - T_{min}} \, 100$$

where T_i is the temperature of the month i (i = 1,.. 12), T_{min} = temperature of the coldestmonth and T_{max} = that of the hottest month).

Again,this transformed value is very conservative from the areal point of view.

1.3.3. The Isanomal

For studying the effects of continents and oceans on temperature, another transformed value - the Isanomal (Conrad L. Pollak, 1962) - is used. This index is free from the influence of the latitude, and it shows just the required information about the influences of land vs.sea. There are other kinds of Isanomals, such as those that show the differences (anomalies) from a standard "Element - Altitude" Curve. In this example the transformed values are free from the influence of absolute height - thus the effect of the exposure can be studied (e.g., comparing regions north and south of the Alps), and interpolations can be easily made in mountainous areas.

1.3.4. The standardized value

Other transformed values - achieved by statistical standardization - are advocated by Gringorten (1966), Sharon (1967) and Ginsburg (1970). As an example, Ginsburg defines transformed temperature as

$$Z_{i(j)} = \frac{T_{i(j)} - \bar{T}_{(j)}}{S_{(j)}}$$

where Z is the standardized value, T-temp. at a fixed hour, $\bar{T}$ - its mean, S-its standard deviation, j-the date and i-the year.
In this special case S is minimum since the series is free of e_{LT}, e_A and e_D, thus S is influenced mainly by e_X and sometimes also by e* (see para 1.1. above).
Standardization is the most general way of transformation because it can be applied to any element (even if the parent distribution is not Gaussian - see: Gringorten, 1966) and for any purpose - such as those for which the Isomers, the Relative Temperatures and the Isanomals were used.
In each application the purpose for the transformation should be strictly defined and according to this definition the mean and the standard deviation should be computed; for example, instead of isanomals for the "Element-Altitude" evaluations, a standardized value can be used in the following way:

$$Z_i = \frac{E_i - E'}{s'}$$

where E' is the estimated value according to a regression of E on altitude and s' is the standard deviation of E around the regression curve (standard error of estimate). Another advantage of the standardization is that the transformed values are pure numbers, so that comparisons of different elements or different regions can be easily made (contrary to the Isanomals). Moreover, these numbers are meaningful from the statistical point of view (contrary to any previous especially-invented transformation methods in climatology).

2. The Concept of Geo-Synoptical Quality Control

2.1. The present situation of climatological quality control

It is a well known fact that climatological data should be thoroughly checked and controlled, immediately after being received in the Met.Central Office, because of the huge amount of causes for mistakes. There are several methods for carrying out this quality control, most of them not efficient enough and demanding a heavy investment in labor. The following remarks are according to unpublished schemes of work, developed by Mr. S.Walther (Israel Met. Service).

In some trivial-logical problems of quality control (Filippov, 1968) the computer can be a great help in locating suspected values. There are three kinds of these problems:

a. Inner inconsistency in the observation (rain without clouds, wet bulb temp. higher than dry bulb temp. etc.)

b. Inconsistency in the behaviour of the element with time (such as temp. at 05hr. higher than both observations at 02 and 08hr (this check is applicable only for synoptic stations, observing every one to three hours all day long), or the obvious mistake of temp. at a certain hour higher than the maximum temp.)

c. Extreme values beyond certain limits (temp. above 45°C, wind speed above 34 knots etc.)

Both those kinds of obviously suspected values are only the minority in the whole spectrum of the suspected values. Most of the others can be found (Mané & Walther, 1967) when the data are tabulated according to geo-synoptical coordinates, and up-to the present the computer is used only as a printer of such tabulations.

2.2. The geo-synoptical quality control

A geo-synoptical array means that observations of the same element at the same time of measurement are tabulated or mapped geographically; e.g.

Stations		Max.temp., °C, on 1 July 1971
station	1	18.2
"	2	18.4
"	3	23.5
"	4	20.8
"	5	17.5
"	6	17.9
	.	.
	.	.

Manual screening of the array is the next step. As stations 1 to 6 are in the same geographical region it seems that the values 23.5° (station 3) and 20.8° (station 4) are suspect and should be checked. The first was really found wrong by 5° (the correct value of station 3 is 18.5°) but the other is correct (station 4 is situated in a topographical depression).

All the values in the geo-synoptical array are similar from the point of view of the e_{LT}, e_A, e_D and even of the e_X influences. The differences between them are caused by:

1) the geographical conditions, and
2) by technical and observational errors (e*)

As only the last cause is the subject of the quality control - there should be found an objective way to eliminate the influence of the geographical, topographical and exposure factors.

3. The proposed control of Transformed Data (TGS)

3.1. The Control of standardized data

If standardized values are enlisted in a geo-synoptical array then the influence of all those geographical (fixed) factors will be mostly eliminated because they are effecting also the mean and the standard deviation of the variate in each station. If this is done there will then remain only one factor influencing the differences between the standardized values, namely e*. After standardization the array (para 2.2 above) would look thus:

stations	absolute, °C	max.temp. on 1 July 1971 standardized
1	18.2	-1.4
2	18.4	-1.6
3	23.5	+1.2
4	20.8	-1.4
5	17.5	-1.6
6	17.9	-1.5
.	.	.
.	.	.

Then, the suspected value (station 3) is clear and could be detected either manually or by simple operational statistical test; thus it is also hoped that <u>the computer could be used more efficiently than being just a printer</u>.

3.2. <u>The influence of weather on TGS</u>

However it may not be correct to assume that all the geographical influences vanish completely when the data are transformed to standardized numbers. Although the geographical situation practically never changes - the influences of these fixed factors can vary a lot. e.g. in a clear-calm inversion-weather the minimum (night) temperature of a valley station may be lower than usual (negative standardized value) while exact opposite phenomenon occurs on stations on the top of a hill (positive standardized number). This point, and may be many others, should be thoroughly studied experimentally before the TGS could be operationally used.

3.3. <u>Synoptic climatology and TGS</u>

TGS could be made more sensitive and meaningful by reducing the standard deviation (e_X). This might be done by breaking down the historical climatological series into partial series according to the general synoptical situations ("Grosswetterlagen"; see Hess & Brezowsky, 1952) and by computing separate means and standard deviations for each partial series. This is a huge task, but it may be a worthwhile project since it can solve the problem stated in the preceeding paragraph (3.2). On the other hand the amount of data for each partial series may be so small as to prevent calculation of stable estimates of the required statistical parameters.

4. Pre-requisite information

As at present only for one element (temp.) from only one station (Basel, 1901-1963) standardized values for each day of the year and for each time of observation were computed (Ginsburg, 1970), there remains a burdensome amount of calculations to be done before TGS can be operational.

4.1. The amount of calculations needed

First, there is the obvious task of checking the possibility of standardizing data for all the Swiss stations, for all the climatological elements and for all the times of observations. Only if this task is accomplished, TGS can really become practical routine.

4.2. The station's net-work problem

As the station's net-work is in a constant change (old stations stop reporting, active stations are slightly moved, new stations are established etc.) it is not a realistic assumption that for each active station all the estimates of the statistical parameters required could be calculated by its own series. The only solution is that for each "element/data/time of observation" the geographical distribution of all the parameters all over the country would be known mathematically and thus could be fed in the computer. This is also a huge task, and may be it is a definite subject for a research project of its own.

4.3. The scientific justification of TGS

As the demands of all these pre-requisites are so high, may be it should be tested if the whole TGS project is worthwhile. However, from the scientific point of view, such a project is justified.

5. A proposal for a pilot-project

As a result of all the difficulties expected (chap. 4 above) it is best to suggest that a pilot project should be the first actual step. It seems appropriate to check TGS on one known and familiar element only, namely - on the maximum temperature. As this element is relatively conservative from the geographical point of view, the first pilot project will not be worked out on partial series (see para 3.3 above). Only if this pilot project is a success, then all the improvements discussed above could be tried.

6. References

Conrad, V. and L.W. Pollack, 1962 :
Methods in Climatology
Harvard Univ. Press, 459 pp.

Elbashan, D., 1966 :
Monthly Rainfall Isomers in Israel
Isr.Jl. Earth Sc., V.15, pp.1-7.

Filippov, v.v., 1968 :
Quality Control Procedures for Meteorological Data
World Met.Org. (WWW-PR 26), 38 pp. + Appendices

Ginsburg, Th., 1970 :
Die Statistische Auswertung von langjährigen Temperaturreihen
Schweiz.Met.Zentralanst. (MZA-19), 41 pp.

Gringorten, I.I, 1966:
A stochastic Model of the Frequency and Duration of Weather Events
Jl.App.Met., V.5, pp. 606-624.

Hess, P. and H. Brezowsky, 1952 :
Katalog der Grosswetterlagen Europas.
Ber.Deut.Wetterdienst in der U.S.-Zone, No. 33.

Mané, U., and S. Walther, 1967 :
The Collection of Climatological Data and their Control
Israel Met.Service, 17 pp.

Sharon, D., 1967 :
On the further Development of Gringorten's Stochastic Model for Climatological Prediction.
Jl. App. Met., v.6, pp. 625-630.

Thom, H.C.S., 1954 :
The Analytical Foundations of Climatology
U.S.W.B., Washington, 26 pp. (memog.)