CLIMATIC RESEARCH AND THE WEATHER-FORECASTS

by

G. Koppány

Éghajlatkutatás és az időjárás-előrejelzés. Az időjárást illető bizonytalanság meghatározható a Shannon-entrópia segítségével, ha ismeretesek az egyes időjárási jelenségek éghajlati valószínűségei. A dolgozat példákat mutat be a prognózis előtti és a prognózis utáni bizonytalanság kiszámítására.

The uncertainty about the weather is defined by means of Shannon-entropy when the climatological probabilities of weather phenomena are available. The paper presents examples for the calculation of the uncertainties before and after the weather-forecasts.

The purpose of classical climatology was the statistical analysis of the several decade long measured data sets of a given geographical point. Statistical characteristics have been set up in order to characterize the climate of the given place, such as: 1) averages of many years or normal values, 2) the relative frequencies of the deviations from normal values, 3) the average value or the standard deviation of the deviations from the normal values (anomalies), 4) the size of absolute amplitudes (records), 5) the annual variation of the normal values of climatological parameters.

The above characteristics have proved to be very useful in the following fields of practical usage: the acclimatization of agri-cultures, the planning of water-economy, the planting of wind or water power stations and so on. The growing lenght of data sets and the spreading of the non-classical means of climatological research (dendrochronology, pollen-analysis and so on) have made the examination of the fluctuation of weather important.

The interest of meteorologists has turned from climate-research to weather-forecast since the 1930-ies. It is not our aim to deal here with the historical reasons of this fact here. However, we are sorry to say that because of the separation of climatology and synoptic meteorology, a really deep connection could not have been created between the two major territories of meteorology. In our opinion synoptic meteorology, which is dealing with weather-forecast could profit much more from the climatological characteristics than it practically does.

In the 1950-ies *G. Péczely* was the first among the Hungarian researchers who was eager to build a bridge between climatology and synoptic meteorology (1957). He found the base of *synoptic climatological* research by elaborating the 13 Hungary oriented large-scale weather situations. Synoptic climatology is a special adaptation of the more generally so-called *conditional climatology.* The essence of conditional climatology is that we examine a clima-factor (e. g. temperature) in the condition of another (wind direction), because we suppose that there is a statistical connnection between the two factors (*Fig. 1*).

We get a new way of approach of the prognostical importance of the results of weather-forecast if it is defined on the basis of information theory:

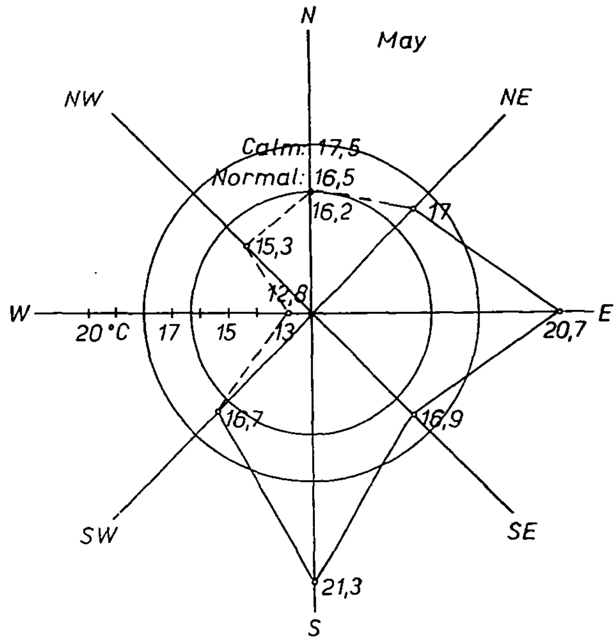

Fig. 1. The daily mean temperatures in Budapest in May as functions of wind direction and wind speed (after A. Réthly). Legend: the temperature is measured by distance from the origo; the monthly average and mean value for calm days are denoted with concentric circles, the expected temperatures for the 8 wind directions are also denoted.

Only that forecast is considered real which aims to reduce some kind of uncertainty. In *contrary* to this definition here are some examples of wrong forecast:
1. The July mean temperature next year in Hungary will be higher than the January mean temperature, 2. the number of sunny hours will be higher in the summer months than in the period from November to January, 3. there won't be significant precipitation in Israel next summer.

There is no need to prove that the above statements are highly certain on the basis of our knowledge (The coldest July in Budapest was in 1913 with the average of 18.4 °C and the mildest January was in 1983, with an average of 5.1 °C. The number of sunny hours in our country is much higher in summer than between November and January, which is due partly to astronomical reasons and partly to the yearly variation of clouds. In Israel there are precipitation data from Jerusalem back to 1861, and during the 120 years they never got a precipitation more than 1 mm in the summer time.) That's why in our examples the predictions to the future can't be considered as forecasts, namely they lack the obligatory uncertainty.

But we may put such a question concerning the future where there is a smaller or bigger uncertainty connected with the correct answer. Let our question be: what will be the mean monthly temperature in Budapest next December?

The answers to the question are determined by the collected clima data lines of Budapest back to 1780. According to this there were as low as -10 °C (1879) and as high as $+5.6$ °C (1826) mean temperatures in December. During the past 204 years the amplitude of monthly temperature comes up to almost 16 degrees. The uncertainty connected with our problem depends on the probability of answers on the bases of climatology.

In our table there is a survey of the relative frequency of monthly mean temperatures in Budapest in 1780—1983 in 16 one-degree intervalls.

Table 1

The relative frequency of monthly mean temperatures in December in Budapest between 1780—1983

−10 −9	−9 −8	−8 −7	−7 −6	−6 −5	−5 −4	−4 −3	−3 −2	centigrades
0.5	0.5	—	1.0	0.5	2.9	3.9	6.4	%
−2 −1	−1 0	0 1	1 2	2 3	3 4	4 5	5 6	centigrades
7.4	9.3	19.6	16.2	12.2	12.2	5.0	2.4	%

The *Shannon-entropy* should be used to measure uncertainty. Let Q be the well-defined question, to which we expect an answer from the forecast. A question is called well-defined if the place the time and the needed information are given. All the information we have concerning the Q question (the climatological knowledge) should be signed X. Let's suppose that „m" is the number of answers that can be given to the question according to our climatological knowledge, and the climatological probabilities of an answer (relative frequencies) $p_1, p_2, \dots p_m$ are known. We are uncertain about which of the m possibilities will be realized.

The value of uncertainty can be estimated in the following way using the Shannon-entropy:

$$S(Q/X) = \sum_{i=1}^{m} p_i \cdot \mathrm{ld}\, 1/p_i, \quad \text{(bit)} \tag{1}$$

where S stands for the Shannon-entropy concerning the well-defined Q question, of which we have X information, p_i stands for the possibility of the i-th answer that can be given to the question Q, while ld $1/p_i$ stand for the logarithm to the base 2 of this probability with a minus sign. Using logarithm to the base 10, equation (1) can be written in the form:

$$S(Q/X) = 1/\log 2 \sum_{i=1}^{m} p_i \cdot \log 1/p_i \quad \text{(bit).} \tag{2}$$

We can interpolate from our table that 50% of the December mean temperatures happened to be lower than +0.4 °C and 50% higher than that. If we expect from our forecast the information whether the monthly mean temperature is below or above the median (+0.4 °C) then according to formulae (1) or (2) the uncertainty is one bit. If we expect from our forecast however to tell the monthly mean temperature in one-degree interval, then putting the relative frequencies in our table into formula (1) the uncertainty can be exactly calculated: *3,3278 bit.*

We can avoid the long-lasting calculation if we divide the weather-data sets into 4 or 8 categories with equal probabilities using the quartiles or octiles. In this case the uncertainly will be 2 or 3 bits. When trying to forecast the amount of precipitation, three categories with equal probabilities, the so-called tertiles, are used very often. In this case the uncertainty before the forecast according to equasion (1) is *1,5849 bit*.

Thus uncertainty to be decreased by the forecast depends on one hand on our expectations about the forecast, in our example on the broadness of the intervalls in which we want to foretell the temperature. The narrower the intervalls are in which we want to forecast the bigger the uncertainty is. On the other hand the uncertainty depends on the climatological probabilities. It is evident that the uncertainty increases if the climatological probabilities of the answers given to the question Q are almost equal.

The uncertainty is almost maximal when in the case of m possible answers:

$$p_1 = p_2 = \ldots = p_m = 1/m.$$

We know from experience that the accuracy of weather-forecasts turns out to be limited. In other words: the forecasts are approximations of reality with some error. The relative frequencies of errors can be determined by the verification of a comperatively high number of forecast.

The information gain obtained from the forecast reports is defined as the difference between two entropy values: one of them reflecting our knowledge before the forecast (X), the other one (X') is that after the forecast, and is obtained by the verification of a number of forecasts so that we determine the relative frequency of the errors of prognosis. The information gain is:

$$L = S - S'. \tag{3}$$

Where S and S' are the uncertainties before and after the forecast, respectively.

The more we can reduce the uncertainty after the forecast (S'), the bigger the information gain is. In order to determine S' we use the probabilities (p_i') of the errors of the forecasts.

It is evident that the smaller is the standard deviation of errors the bigger is the information gain and the smaller the value of S'. This is also true if there is a systematical error in the forecast but we know this error from experience. In this case we can correct the forecasts according to our knowledge of systematical errors (*Koppány, G*, 1975.).

On the contrary if the standard deviation of the errors of the forecasts is bigger than that of the climatological data, the uncertainty increases after the forecast and the information gain is negative.

Let us have a simple example for the sake of better understanding.

Let our task be the forecast of temperature in five equally probable categories. The limits of categories are determined in this case by the quinteles. The uncertainty before the forecast is:

$$S = 5 \cdot 0{,}2 \cdot \mathrm{ld}\, 5 = 2{,}322 \quad \text{bit.}$$

Let us have an enough number of verified forecasts and let us suppose that the probability of the correct forecasts: $p_k'=0.5$, and that of the errors referring to two categories: $p_k' \pm 2=0$, and that of the errors referring to one category is $p_k' \pm 1=0.25$. In this case:

$$S' = 0.5\, \mathrm{ld}\, 2 + 2 \cdot 0.25 \cdot \mathrm{ld}\, 4 = 1.5 \quad \text{bit}$$

and the information gain is:

$$I = S - S' = 2.322 - 1.5 = 0.822 \ \text{bit.}$$

Informative results are obtained by the following modifications. Let $p'_{k+1}=0.6$, $p'_k=0.4$ and $p'_{k+2}=p'_{k-1}=0$. So the probability of the correct forecast is 40%, that of the false estimation of plus one category is 60% and the probability of all the other errors is 0. In this case uncertainty after the forecast is:

$$S' = 0.6 \cdot \text{ld } 1/0.6 + 0.4 \cdot \text{ld } 1/0.4 = 0.9709 \text{ bit.}$$

The information gain is:

$$I = S - S' = 2.322 - 0.9709 = 1.3511 \text{ bit.}$$

In the latter case there is a systematical error in our forecasts: in 60% of the cases next higher category was realized instead of the expected one. The uncertainty of the realibility of our forecast can be increased by correcting with this systematical error. The information gain has increased as compared to the previous example, because the standard deviation of errors has decreased.

Finally, let us consider in brief the importance of the synoptic climatology as reflected in the information theoretical adaptation of forecast. According to the long data sets from Budapest the absolute amplitude of the January daily mean temperature is 32 °C and 19 °C, respectively. The standard deviation of meteorological data is comparatively high. The standard deviation of the daily mean temperatures categorized by the macrosynoptical categories decreases. (*Péczely, Gy.*, 1961.) So if the forecast of the atmospherical pressure, based on the service of the Meteorological Regional Centers are at our disposal for the next few days then the forecast field can be identified as a macrosynoptical type. Since the standard deviation of the temperatures belonging to a macrosynoptical type is smaller than that of the general climatological data the uncertainty connected with the expected daily mean temperature decreases. This is naturally true not only in the case of temperature but also for the other meteorological parameters.

Attention should be paid to a quantitative study in which from the data sets of certain meteorological parameters (wind direction, wind speed, clouds, rain, etc.) the relative frequency of these would be provided in the given scale and also for the categorized data sets according to Péczely's types. Having all these data it would be possible to determine the uncertainty connected with the expected figure of the individual meteorological parameters and also the decrease of this uncertainty supposing the knowledge of Péczely's macrosynoptical types for the next few days.

References

Koppány, Gy. (1975): A meteorológiai előrejelzések verifikációjának módszerei. (Verification methods of meteorological forecasts) Meteorológiai Tanulmányok, No. 13.

Péczely, Gy. (1961): Magyarország makroszinoptikus helyzeteinek éghajlati jellemzése. (Climatological characteristics of the large-scale synoptic patterns of Hungary) OMI Kisebb Kiadványai, 32. K.

Réthly, A. (1947): Budapest éghajlata. (Climate of Budapest) Budapest.