THE INFORMATION CONTENT OF THE TEMPERATURE FIELD OF THE NORTHERN HEMISPHERE AND THE MEAN MONTHLY TEMPERATURES AT SZEGED

by

L. Pelle

Az északi félteke információtartalma és a szegedi havi középhőmérséklet. A tanulmány (Szeged és az északi félteke 83 állomása havi középhőmérsékletei 80 éves sorának felhasználásával) megállapitja, hogy bizonyos állomások előző adatainak ismeretében a klimatológiai előrejelzéshez képest csökkenteni lehet a Szegeden várható havi középhőmérséklet bizonytalanságát. Ezt a szegedi havi középhőmérséklet és a megelőző 12 hónapban az északi félteke 83 állomásán megállapított havi hőmérsékleti átlagok kölcsönös információtartalmának kiszámításával éri el.

The study finds out (using the 80-years line of the mean monthly temperatures of Szeged and 83 meteorological stations of the northern hemisphere) that the uncertainly of the mean monthly temperature to be expected at Szeged can be diminished as compared to the meteorological forecast — with knowledge of dates of certain stations relating to the previous period. This is achieved by computing of the reciprocal information contents of the mean monthly temperatures at Szeged and the mean monthly temperatures fixed at the 83 stations of the northern hemisphere over a period of 12 months preceding the months mentioned.

One of the most important targents of meteorology is to forecast the weather. This difficult and ungrateful work is not resolved yet. According to estimations the theoretic limit of expanding numerical forecasts is two-three weeks at the most. The short and middle-term prognostic based upon mathematical and physical calculations are relatively developed which is not so in the case of long-term forecasts where mostly statistical processes are applied. A method introduced below aims at a contribution of grounding long-term forecasts as well. Since the method involves clearly mathematical statistical features its limitations may not be disregarded when evaluating its results. Our method offers a possibility to correct climatological prognostics. A climatologic prognostic can be established in the following form: If the median of mean temperature in Szeged in January is e. g. m: $-1{,}0$ °C then with the help of climatologic data temperatures above and below $-1{,}0$ °C can be calculated with the same probability.

Test method.

The possible values of discrete random variable ξ should be $x_1, x_2, x_3, \ldots, x_n$, its distribution $p_1, p_2, p_3, \ldots, p_n$ where p_i constitutes the probability of supervention of the event x_i. The definition of entropy of ξ signed $H(\xi)$ is the following:

$$H(\xi) = -\sum_{i=1}^{n} p_i \log p_i . \qquad (1)$$

This quantity could be understood as a measure of instability of the random variable. The entropy of continuous random variable:

$$H(\xi) = -\int_{-\infty}^{+\infty} f(x) \log f(x)\, dx \qquad (2)$$

where $f(x)$ is a function of density ξ. These definitions can be given for multi-dimensional random variables. The next concept appears to be important from a practical point of view because it measures the closeness of the relation of two random variables, the reciprocal information content of two random variables signed as $I(\xi,\eta)$ is as follows:

$$I(\xi, \eta) = H(\xi)+H(\eta)-H(\xi, \eta). \tag{3}$$

In case of independence $I(\xi, \eta)=0$. The maximum of $I(\xi, \eta)$ is the smaller from the two entropies, $\max(I(\xi, \eta))=\min(H(\xi)), (H(\eta))$ and this is true if ξ and η are functions of each other, i. e. with the help of one of them the other can be expressed. We are of course more interested in stohastic connections occurring in reality than in extreme cases, the quantity (3) will be used to measure the previous one. In our test the temperature time sequences used as parting data were described with random variables of normal distribution. Consequently, with the help of (2) and (3) the reciprocal information content of two normal random variables can be calculated:

$$I(\xi, \eta) = -\frac{1}{2}\log(1-r^2) \tag{4}$$

where r is the correlation coefficient between normal random variables ξ and η, the demonstration of (4) can be found e. g. in [2]. If ξ and η are uncorrelated then $I(\xi, \eta)=0$ and if $|r|\to 1$ then $I(\xi, \eta)\to\infty$, so if can be stated that $I(\xi, \eta)$ measures the identity of the linear connection of the two probability variables.

As a data basis of this process the monthly mean temperatures of 83 stations over the northern hemisphere were applied with measurements taken between the period starting with 1880 and up to 1960.

The next step was to examine the information offered by the individual temperature fields (northern hemisphere) with regard to the monthly mean temperatures of specified station — Szeged. The reciprocal information content with the previous 12 months of the whole field was calculated for each monthly sequence in Szeged. For example we calculated the reciprocal information content of random variable describing the October temperature sequence in Szeged and that of the random variable describing the September, August, July, ..., October field over the hemisphere. In such way altogether $12\times 12=144$ information quantity fields were obtained. Each of them contains 83 information quantity this means altogether 11 952 calculations. A value acceptable on a 5% significance level was regarded as acceptable. (E. g. if $n=80$, $I(\xi, \eta)=0{,}0348$ bit.) As expected the decisive majority of information quantities appeared to be smaller. In each case, however, a number of values presented themselves higher than this. Maximal information quantities are around 0,14—0,15 bit. From the possible 144 information quantity maps the 14 more interesting ones were sketchd (Fig. 1—13). On these maps the fields with significant values were traced with thick lines. The signs "+" and "−" refer to the direction of the connection. On the basis of the maps it can be stated that certain territories have an important role, especially the south east part of North America, central and western part of Europe. Other territories, such as East India, the Persian Gulf, Mandjuria, Korea, Caribbean Islands represent realistic connections only in a few cases. The information quantities may be used for forecasts with two categories because proportionately with the obtained information the vagueness of forecast diminishes as well. Since 1 bit information is needed to a prognostic with two categories which undoubtedly comes true the maximal 0,15 bit calculated information quantity disallows this possibility, on the other hand, a better solution was found than the

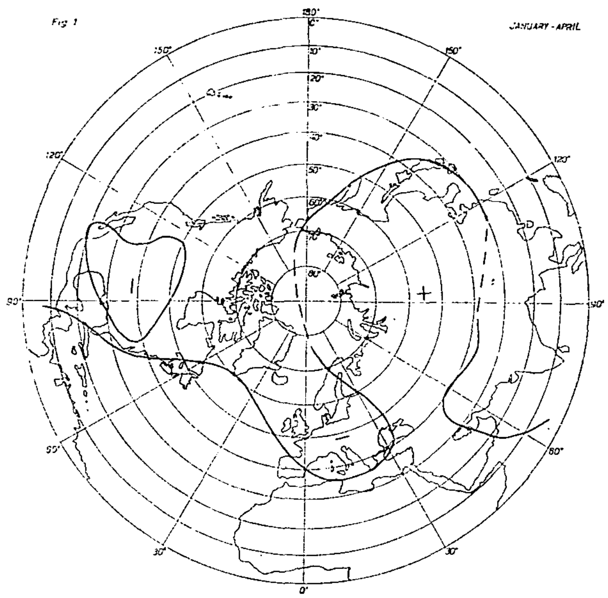

Fig. 1. The regions of the northern hemisphere at which the January mean monthly temperature has an information content relating to the April mean temperature at Szeged

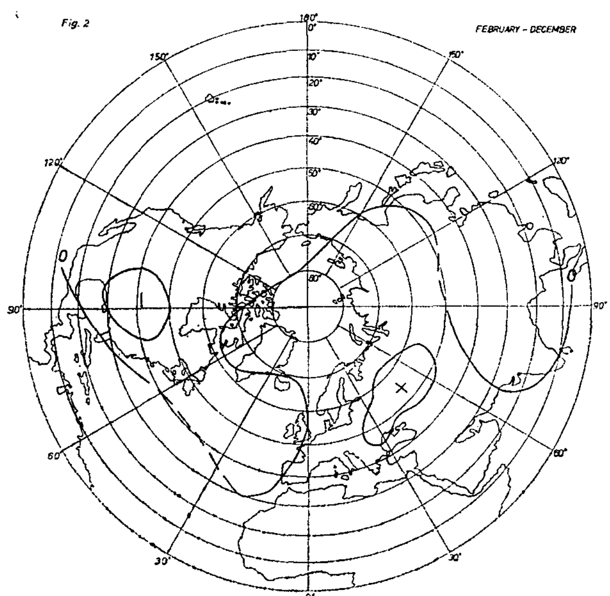

Fig. 2. The regions of the northern hemisphere at which February mean monthly temperature has an information content relating to the December mean temperature at Szeged

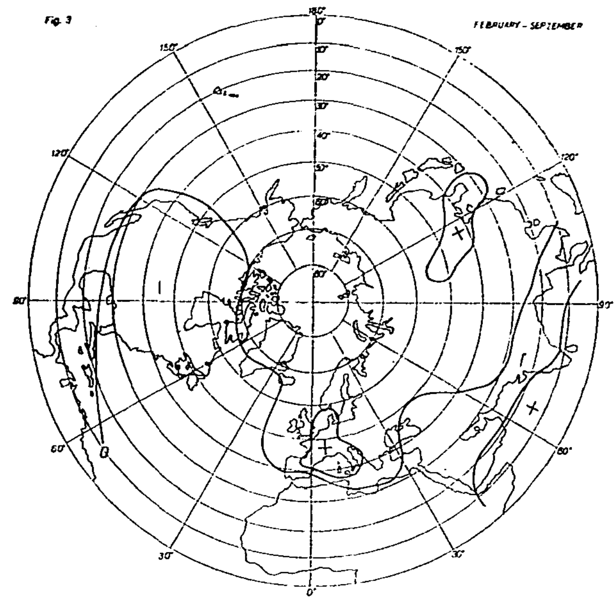

Fig. 3. The regions of the northern hemisphere at which the February mean monthly temperature has an information content relating to the September mean temperature at Szeged

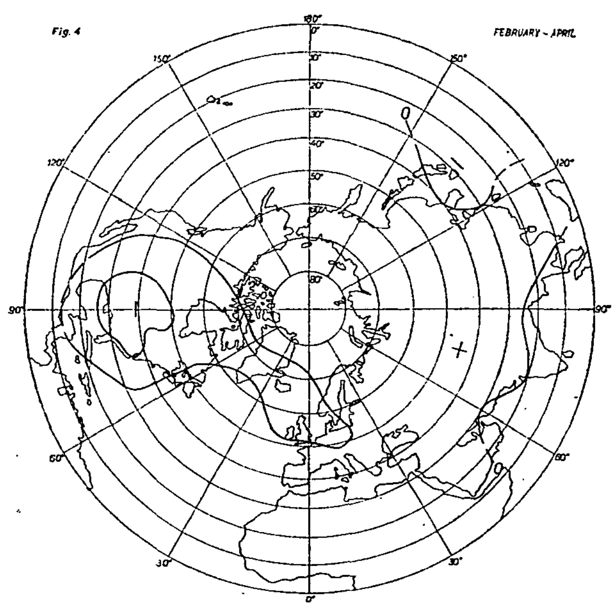

Fig. 4. The regions of the northern hemisphere at which the February mean monthly temperature has an information content relating to the April mean temperature at Szeged

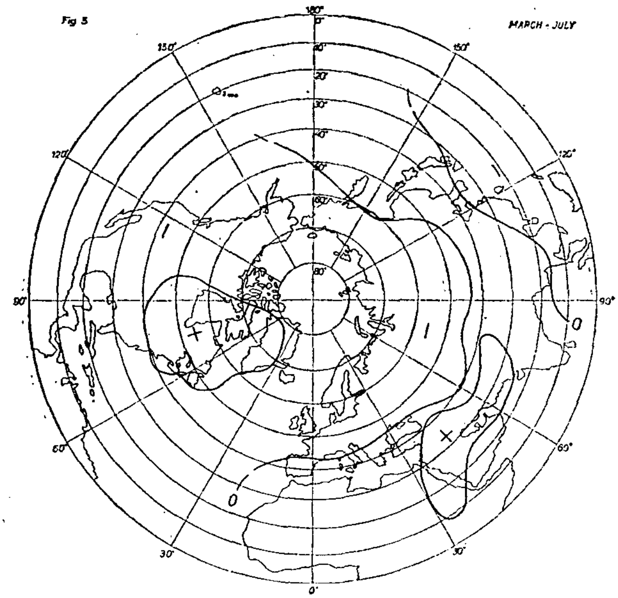

Fig. 5. The regions of the northern hemisphere at which the March mean monthly temperature has an information content relating to the July mean temperature at Szeged

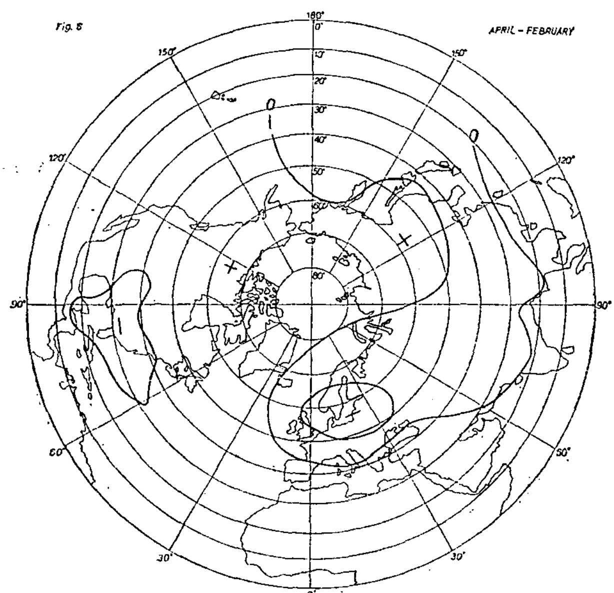

Fig. 6. The regions of the northern hemisphere at which the April mean monthly temperature has an information content relating to the February mean temperature at Szeged

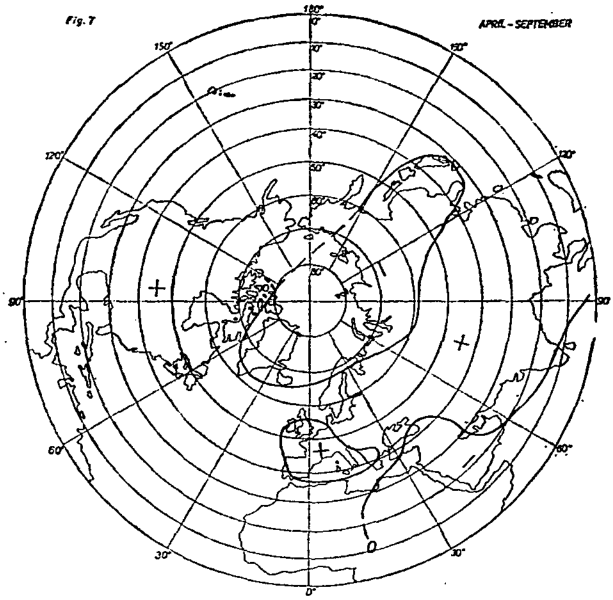

Fig. 7. The regions of he nor hern hemisphere at which he April mean monthly temperature has an information content relating to the September mean temperature at Szeged

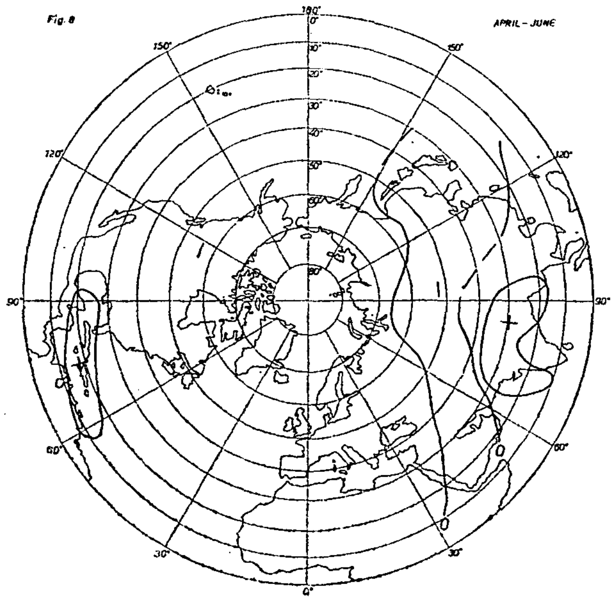

Fig. 8. The regions of the northern hemisphere at which the April mean monthly temperature has an information content relating to the June mean temperature at Szeged

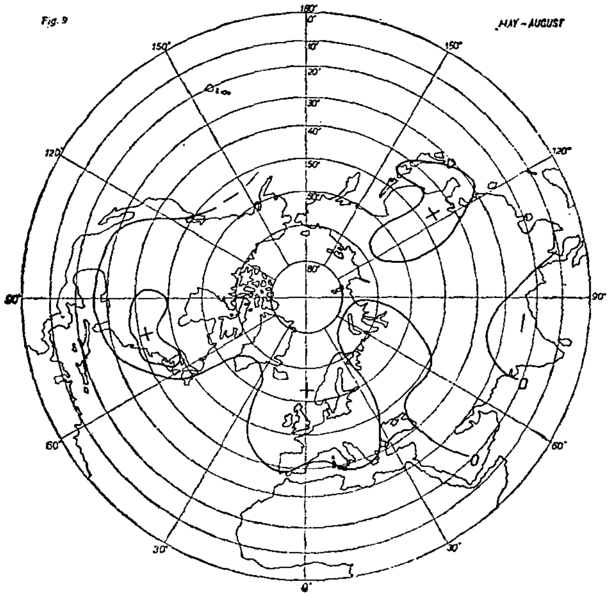

Fig. 9. The regions of the northern hemisphere at which the May mean monthly temperature has an information content relating to the August mean temperature at Szeged

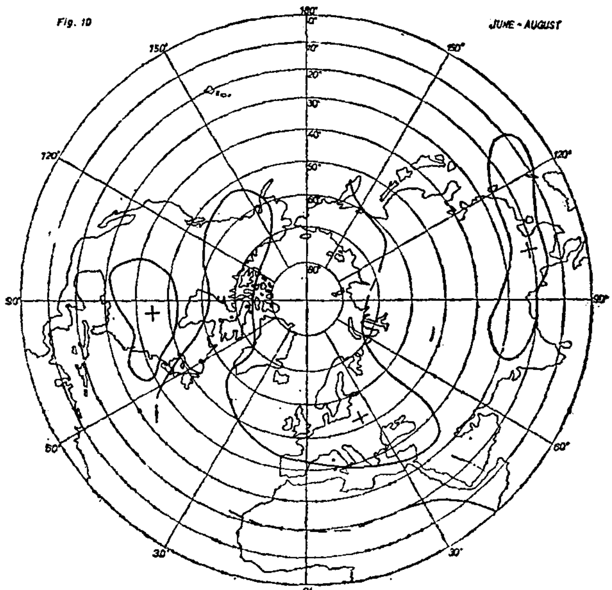

Fig. 10. The regions of the northern hemisphere at which the June mean monthly temperature has an information content relating to the August mean temperature at Szeged

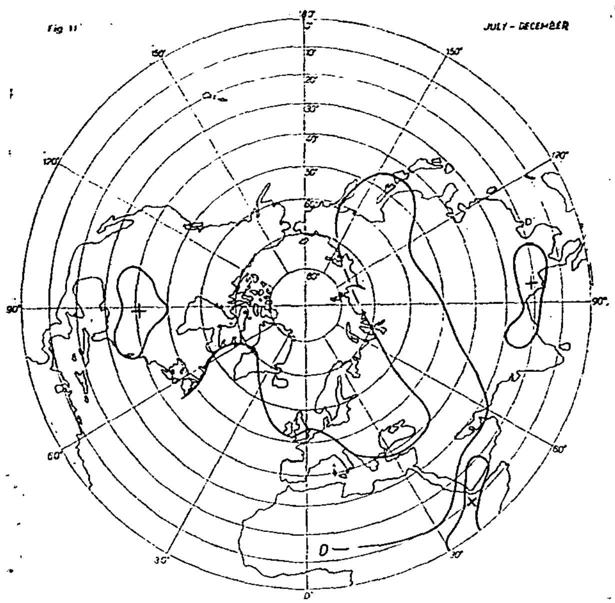

Fig. 11. The regions of the northern hemisphere at which the July mean monthly temperature has an information content relating to the December mean temperature at Szeged

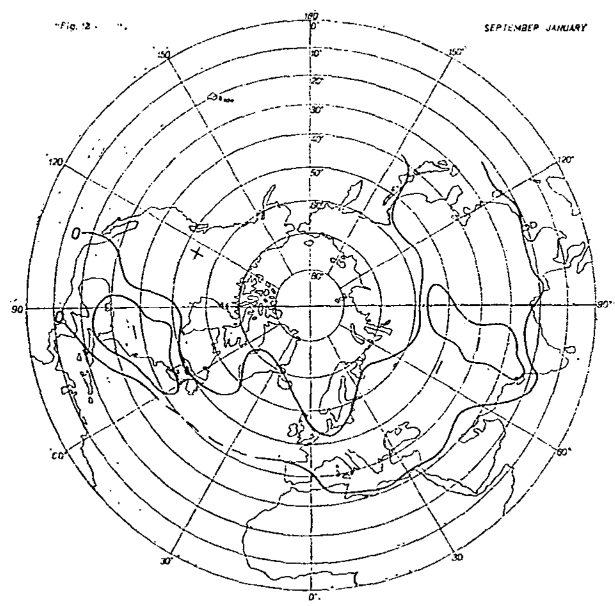

Fig. 12. The regions of the northern hemisphere at which the September mean monthly temperature has an information content relating to the January mean temperature at Szeged

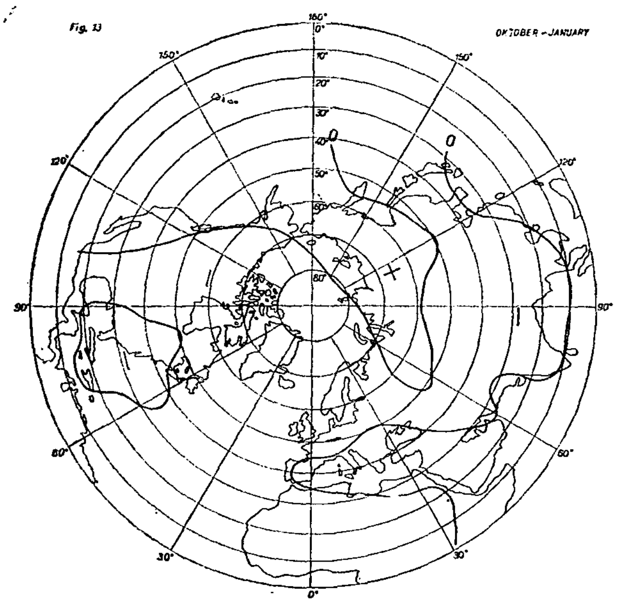

Fig. 13. The regions of the northern hemisphere at which the October mean monthly temperature has an information content relating to the January mean temperature at Szeged

climatologic prognostic. Because according to (1) the entropy of a forecast with two categories:

$$H = p \log_2 \frac{1}{2} + (1-p) \log_2 \frac{1}{1-p}$$

this has a maximum if $p=0{,}5$, $1-p=0{,}5$. $H_{\max}=1$ bit. With each information quantity gain the entropy diminishes

$$H = H_{\max} - I = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

this way with I the distribution $(p, (1-p))$ can be defined. E. g. concretely taking the first month before February, in this case the maximal information quantity is 0,14 bit, so

$$1 - 0{,}14 = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

from where $p=0{,}28$, $1-p=0{,}72$, consequently, since the actual January value of the temperature field is known, a $p=0{,}28$, $1-p=0{,}72$ prognostic can be given for mean temperature in Szeged, in February and because this relation is positive in the case of a January mean temperature below the average the probability of a below average February mean temperature is $p=0{,}72$, in the case of a January above average it is $p=0{,}28$. This method was applied for the period 1951—1960. This "archive" prognostic proved to be true in 65% of the cases as opposed to the 50% of a climatologic prognostic. The low efficiency of our forecast can be understood since a very plain predictor was applied though the presently available complicated proceedings which involve more calculations do not give better results either.