The Summer Index

FOR OFFICIAL USE ONLY

I.M. 47/66

DEPARTMENT OF TRANSPORT AND POWER

METEOROLOGICAL SERVICE

INTERNAL MEMORANDUM

THE SUMMER INDEX.

- by -

C J GILLMAN.

Received 16/6/66

Issued 20/12/66

THE SUMMER INDEX.

Introduction.

Although the instantaneous relations which exist between physical measures in the atmosphere follow known physical laws, it is the mathematical difficulties which arise through dealing with a large number of measures that necessitate so many simplifying assumptions that our answers are only approximate. In dealing with meteorological averages, whether over space or over time, the exact laws are weakened by the mathematical process of averaging, and no fixed relationship exists.

In the 1920s Sir Gilbert Walker and Prof Franz Baur were led to using averages to explain the evolution of large scale weather types on a continental basis. Both workers leaned heavily on statistical and probabilistic methods to determine the empirical relationships involved. It was found that these relationships varied with locality, and also with season.

Baur was able to deduce a number of empirical theorems as a result of extended work on large-scale meteorological measures. Nevertheless, he warned of the need for sound statistical analysis, and the use of rigorous mathematical criteria. It was Baur who formulated the idea of Grosswetter, the existence of certain broad scale features which are distinct from the day to day picture provided by the usual weather chart. His First Empirical Theorem is :-

> "A Gross-wetter exists, in which there are governing complexes comprising variable conditions; because of these complexes, the probabilities of the occurrence of certain indices that characterise the weather for longer periods vary from year to year." [Baur 1951.]

Baur further stated :- "the first theorem ... forms the logical foundation for the exploration of the problem of extended-range weather forecasting." [Baur 1951.]

Nevertheless, C.S. Durst, in the same volume commented that "the methods by which the mechanism of climate is investigated have been sadly lacking in a reasoned approach."

Baur does not define the term "indices" in his First Theorem, which is stated in very broad terms. This is probably due largely to the fact that in other sciences the applications of index theory left much to be desired. (The type known as consumer price indices, in particular, is still the butt of much uninformed criticism.) Indices in another sense resulted from a development of measures of certain common factors in phychology. However, in the early stages of the theory, there was a great deal of unseemly controversy among the psychologists. This lessened the interest of mathematicians in the basic problems involved.

In attempting to develop a measure of the quality of summers it is appropriate to use multivariate analysis, which provides the sound mathematical basis which is needed. In this way we not only determine the best index, but can also estimate its effectiveness as a measure, on a single scale, of the complex it purports to represent.

The Problem.

A simple and attractive method of obtaining an index of summer weather has been given by Poulter (1962), using data for Kew which had been published by Rogers (1960).

Poulter did not give any theoretical background for the index. He left unanswered the question whether his proposed index gave a meaningful figure, and also whether the method of obtaining it was efficient.

A preliminary examination suggests that the method is inefficient, as he uses the range, i.e. the difference between the highest and lowest values for each of the three variables considered - temperature, sunshine and rainfall. It is known that the range is not an efficient measure, and the larger the sample the less efficient it becomes. (Quenouille, 1958, p 61).

The purpose of the present note is to investigate these questions.

Analysis.

We use Yule's notation, and denote

temperature by x_1

sunshine by x_2 and

rainfall by x_3

Kew Data.

The correlation matrix for the Kew mean summer temperature, total summer sunshine and total summer rainfall for the 80 years 1880 to 1959 is :-

$$\begin{bmatrix} 1. & 0.678\ 02 & -0.610\ 58 \\ 0.678\ 02 & 1. & -0.444\ 28 \\ -0.610\ 59 & -0.444\ 28 & 1. \end{bmatrix} = R_k$$

and $|R_k| = 0.337\ 94.$

We first test whether the eigenvalues of this matrix are significantly different, by using a criterion due to Bartlett (1948) and given by Kendall (1961, p 88).
It is that

$$-\left[(n - 1) - \frac{1}{6} (2p + 5)\right] \ln |R|$$

is distributed approximately as χ^2 with $\frac{1}{2}p\,(p - 1)$ degrees of freedom, where

n = number of sets of observations, and

p = total number of eigenvalues.

On applying the test we obtain $\chi^2 = 140$, which for three degrees of freedom is significant beyond $P = 0.001$.

The eigenvalues of the correlation matrix are :-

$$\lambda_1 = 2.160\ 56,$$
$$\lambda_2 = 0.560\ 15$$
$$\lambda_3 = 0.279\ 29.$$

If we use these eigenvalues to determine component values Z_1, Z_2, Z_3, then each of these components may be regarded as determining certain indexes, which, when taken together represent the original data completely. However, the three indexes represent different proportions of the total information in the original figures.

That derived from λ_1 represents 72.0% of the total variance, while those derived from λ_2 and λ_3 represent 18.7% and 9.3% respectively.

<u>λ_1 provides a satisfactory index,</u>

and is essentially equivalent to Poulter's index.

In working with the original data, the variables were transformed as follows:-

$$x_1 = (X_1 - 58.0), \qquad X_1 \text{ in } {}^\circ F.$$
$$x_2 = (X_2 - 300)/50 \qquad X_2 \text{ in hours.}$$
$$x_3 = X_3 / 50 \qquad X_3 \text{ in mm.}$$

The coefficients of Z_1 when applied to the normalised data are:-

$$(0.673\ 87, \quad 0.534\ 50, \quad - 0.510\ 12),$$

or, when applied to x_1, x_2, x_3,

$$Z_1 = (0.422\ 61\ x_1 + 0.277\ 47\ x_2 - 0.416\ 66\ x_3 - 1.734),$$

which has zero mean, and unit variance.

If we wish to avoid negative numbers we may use

$$Z_1' = (0.422\ 61\ x_1 + 0.277\ 47\ x_2 - 0.416\ 66\ x_3 + 3.266)$$

which has mean 5.0 and unit variance.
Its range is from 1.8 for 1888 to 8.3 for 1911.

In terms of the original data, X_1, X_2 and X_3, this index corresponds to the factors:-

$$1. : 0.013\ 13 : - 0.019\ 92$$

or

$$1 : 76.16^{-1} \quad -50.20^{-1},$$

whereas Poulter's index corresponds to

$$1 : 60^{-1} \quad -50^{-1}$$

The loss of efficiency by applying 1/60 to X_2 instead of the correct theoretical value of 1/76 is small.

We conclude that the fitting is reasonably efficient, and that the index represents 72% of the variance of the original data.

<u>Dublin Data.</u>

The same calculations have been done for Dublin for the period 1880 - 1964. The original values X_1, X_2, X_3 were transformed as follows:-

$$x_1 = (X_1 - 10), \qquad X_1 \text{ in } {}^\circ C$$

$$x_2 = (X_2 - 300)/100 \qquad X_2 \text{ in hours}$$

$$x_3 = X_3/100 \qquad X_3 \text{ in mm.}$$

The correlation matrix is:

$$\begin{bmatrix} 1. & 0.351\,80 & -0.357\,44 \\ 0.351\,80 & 1. & -0.331\,03 \\ -0.357\,44 & -0.331\,03 & 1. \end{bmatrix} = R_d$$

and $|R_d| = 0.722\,66.$

On applying Bartlett's criterion we obtain $\chi^2 = 27.03$, which is significant beyond $P = 0.001$. We conclude that the eigenvalues are significantly different.

These eigenvalues are

$$\lambda_1 = 1.693\,52, \text{ and}$$

$$\lambda_2 = \lambda_3 = 0.653\,24.$$

Thus a component based on λ_1 will account for only 56.45% of the variance. As the other eigenvalues are equal, no further improvement is possible by considering a second component, and we must have 43.55% of the variance unexplained.

The component based on λ_1, i.e. Z_1 is given by

$$Z_1 = 0.839\,x_1 + 0.890\,x_2 - 0.933\,x_3 - 3.462$$

with zero mean and unit variance.

If however we take :-

$$Z_1' = X_1 + x_2 - x_3 + 0.8$$

we obtain the expression given by Morgan (1965), transformed to have a mean of 5.0. Its range is from 0.7 to 9.6. However, as the index offers a poor explanation of the quality of summers as measured by the three quantities temperature, sunshine and rainfall, it is probably realistic to calculate it only to $\frac{1}{4}$ units.

In terms of X_1, X_2 and X_3, the factors obtained by multivariate analysis are in the ratio:-

$$1 : 0.010\,61 : 0.011\,12,$$

i.e.

$$1 : 94.3^{-1} : 89.9^{-1}$$

to which the ratios for Z_1' approximate reasonably well.

Comparison of the Kew and Dublin Indexes.

The clue to the difference in representativeness of the two indexes - 72.0% and 56.5% - lies in the correlation matrices.

For Kew, the correlations r_{12} and r_{13} are moderate with r_{23} low.

For Dublin, the three correlations are low.

The difference between r_{12} for Kew and r_{12} for Dublin is very highly significant - $P = 0.0023$.

The difference between r_{13} for both places is significant, $P = 0.018$, while there is no significant difference between r_{23} for both places.

It is likely that the observed higher correlations in the Kew matrix is due to the greater continentality of the London area. This point is worth further investigation.

Relation between the Dublin and Kew Indices for the same year.

It is of especial interest to compare the Dublin and Kew Indices for the same year, to see how far one of them may be used as a predictor for the other.

The correlation coefficient over the period 1880 to 1961 is

$$r = 0.82.$$

Thus, 67% of the variance of one index would be explained by the other, with 33% unexplained.

The relationship is also shown on the attached scatter diagram. It is clear that the index for one place would not be a useful predictor for that at the other.

On comparing the components of the Indices, we find that the correlation between the mean summer temperature in Dublin and that in Kew (over the period 1880 to 1959) is:-

$$r = 0.93.$$

For sunshine the correlation coefficient is:-

$$r = 0.46,$$

while for rainfall the correlation is:-

$$r = 0.43.$$

These values suggest that from the viewpoint of attempting to give a regional measure of summer weather in this part of the globe, temperature alone will give a better measure than any combination of temperature, sunshine and rainfall.

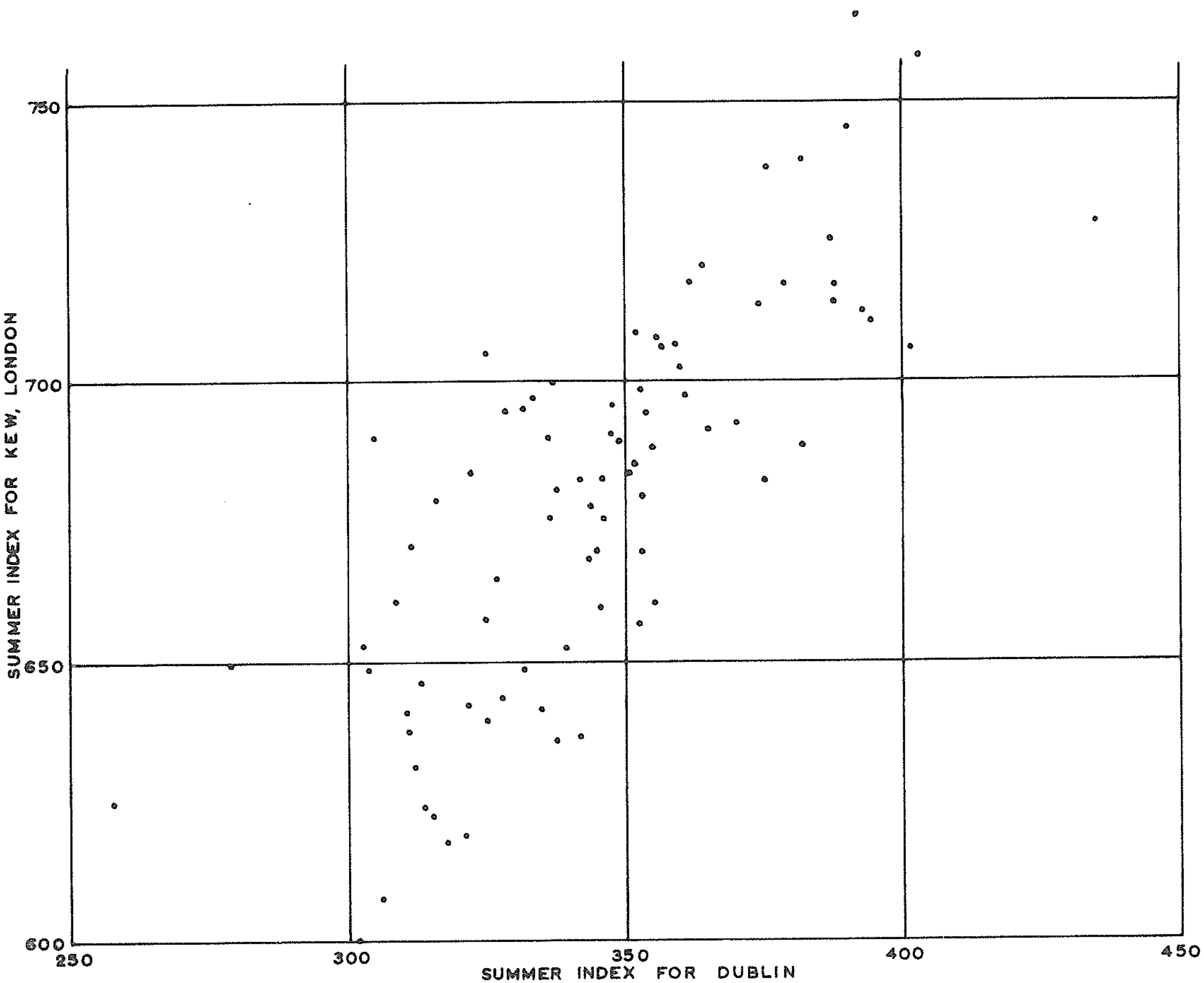
Conclusions.-

The results of this analysis show that the Poulter Index for Kew explains 72% of the variance of the original three measures of temperature, sunshine and rainfall.

For Dublin, the corresponding figure, is $56\frac{1}{2}$%, and its index cannot be rated as a useful measure.

References.

Bartlett, M.S.	1948	Brit. J. Psych. Stat. Section, 1, 73.
Baur, F.	1951	Compendium of Meteorology. Boston. Article on Extended-Range Weather Forecasting.
Durst, C.S.	1951	Compendium of Meteorology. Boston. Article on Climate - The synthesis of Weather.
Kendall, M.G.	1961	A Course in Multivariate Analysis. Griffin, London.
Morgan, W.A.	1965	Summers in Dublin. Irish Meteorological Service, Dublin.
Poulter, R.M.	1962	The next few summers in London. Weather, London.
Quenouille, M.H.	1958	Fundamentals of Statistical Reasoning. Griffin, London.
Rogers, P.	1960	London Summers of the 1950s. Weather, London.



Scatter Diagram showing the relationship between Poulter's Summer Index for Kew, London, and the Summer Index for Dublin, for the years 1880 - 1961