5. Foreman, A. J. E. & Makin, M. J. *J. nuclear Mater.* **79**, 43–57 (1979).
6. English, C. A., Eyre, B. L., Summers, J. & Wadley, H. *Proc. Int. Conf. on Fundamental Aspects of Radiation Damage in Metals* 751006-P2, 910–917 (ERDA Conf., 1975).
7. Evans, J. H. *Int. Conf. on Irradiation Behaviour of Metallic Materials for Fast Reactor Core Components*, Corsica 1979 (in the press).
8. Marwick, A. D. *J. Phys.* **F8**, 1849–1861 (1978).
9. Fikus, J. H. & Johnson, R. A. *Radiat. Effects* **40**, 63–70 (1979).

## Possible Sun–weather correlation

As I have recently published[1] a re-examination of some of Xanthakis' work[2], in which he claimed remarkably high correlations between solar activity and an index of zonal precipitation (for example, $r = +0.77$ in the zone 70°–80° N and $r = -0.71$ for 60°–70° N), I feel I should comment on the recent exchange between Xanthakis and Gerety[3,4].

My re-examination was prompted by the unusually high correlations claimed, the questionable discussion by Xanthakis of 'X-wise distributions'[5] and, more particularly, by the arbitrariness of the derivation of zonal mean $R - R_0$ values. Xanthakis defined $R - R_0$ for each year and station as the actual total precipitation, $R$, at that station in that year minus the lowest annual precipitation, $R_0$, reported for that same station over all the years considered. In the calculation of the zonal means Xanthakis averages together stations having very different mean $R - R_0$ values, with no attempt at normalisation (apart from assigning almost arbitrary $R_0$ values). Furthermore the stations come in or drop out in random fashion according to whether records are available. Thus, even if a perfect correlation between $R - R_0$ values and solar activity existed at each station (which is certainly not the case), the averaging process would be expected to inject discontinuities whenever a station record began or ended. In this respect Xanthakis' method should lead to poor correlations. This makes the claimed high correlations even more remarkable, and indeed puzzling.

The data for 70°–80° N were re-evaluated because that zone contains the least data, Xanthakis published a full tabulation of his data for that zone[2], and his result for that zone was an apparently highly significant correlation ($r = +0.77$) which did not change sign. Like Gerety *et al.*[6] I found that Xanthakis had used only part of the available data, and my analysis using additional data does not support the claimed highly significant correlation. I obtain a correlation of +0.38 which, considering that the data have been smoothed, is barely significant at the 95% confidence level. Xanthakis rightly points out that a different data sample will lead to a different result. However, the great reduction in the observed correlation using the larger sample certainly throws into question the significance of Xanthakis' result.

More importantly, unlike Gerety *et al.*, I chose only to use data as described by Xanthakis, who stated (or certainly implied) that he used all the available data published in *World Weather Records*[7] except that "only stations with more or less extensive series of precipitation data have been considered"[2]. An examination of his Table 2*a* shows that he used stations with as little as 6 years (Ingoy) or 8 years (Myggbukta) of data.

Examination of *World Weather Records* reveals a complete set of data for Isfjord Radio, Spitsbergen for 1912–24, which Xanthakis omitted even though this meant that for 1912–21 his 'zonal mean' was based on only one station. For the interval 1951–60 there are data for five additional stations—Isachsen (10 yr), Sachs Harbour (5 yr), Clyde (9 yr), Barter Island (10 yr), Chokurdakli (2 yr), and Vize Island (8 yr). All but Chokurdakli lie between the meridians 156° W through 0° to 80° E as stipulated by Xanthakis.

The issue is thus not whether different samples of stations were used, but rather why Xanthakis used the particular stations (which give such remarkable results) and did not use others which detract from the significance of the results. Furthermore, why was no mention made and no reason given for omission from the zonal mean of apparently good data (available from the same data source) from long-record stations such as Isfjord Radio, Isachsen and Barter Island, when shorter records were used from Ingoy and Myggbukta? Why does the choice of stations for inclusion in the zonal means consistently lead to higher correlations than those found using larger data samples?

In an attempt to answer these questions I have tested for the possibility that the correlation coefficients found by Xanthakis using his samples of stations and those found by using the larger available data samples are drawn from the same population by a random or chance selection of stations. According to Snedecor[8] this is done by applying a $Z$ transform to the correlation coefficients and a Student $t$ test to the difference between the $Z$ values.

Table 1 summarises the results for three latitude belts which have been independently assessed by Pittock[1] and Gerety *et al.*[6], following the procedures outlined by Xanthakis[2]. Assuming $N$ independent data pairs as indicated for each latitude belt, the probability $P$ that each pair of samples of stations is drawn at random from the same population is shown. $N$ has been reduced to allow for the computed autocorrelations in the data (including smoothing) according to the standard formula given by Quenouille[9]. Taken separately, the result for each of the three latitude zones suggests that Xanthakis' selection of stations (a selection which gives apparently significant correlations) was probably not by chance. Combining the results, it seems that the joint probability that Xanthakis' favourable selection was by chance is less than 1 in 30,000.

I am grateful to Dr E. J. Gerety for supplying the results of his computations.

A. Barrie Pittock
*Laboratory of Tree-Ring Research,
University of Arizona,
Tucson, Arizona 85721
and
CSIRO Division of Atmospheric Physics,
Mordialloc 3195, Australia*

1. Pittock, A. B. *Rev. Geophys. Space Phys.* **16**, 400–420 (1978).
2. Xanthakis, J. *Solar Activity and Related Interplanetary and Terrestrial Phenomena* (ed. Xanthakis, J.) 20–47 (Springer, New York, 1973).
3. Xanthakis, J. *Nature* **275**, 775 (1978).
4. Gerety, E. J. *Nature* **275**, 775 (1978).
5. Roberts, W. O. *WMO Bull.* **23**, 205–206 (1974).
6. Gerety, E. J., Olson, R. H. & Roberts, W. O. *Nature* **272**, 231–232 (1978).
7. *World Weather Records*, Smithsonian Miscellaneous Collections, Vols **79**, **90**, **105**; U.S. Department of Commerce, Vols **1–6**.
8. Snedecor, G. W. *Statistical Methods* 5th edn, 178 (Iowa State College Press, 1956).
9. Quenouille, M. H. *Associated Measurements*, 168 (Butterworths, London, 1952).

Xanthakis replies—Pittock claims that the X-wise distribution of the reported correlations between $R - R_0$ and $I_a$ (ref. 1) is questionable. The same dispute was put forward previously by Roberts[2]. However, in the light of recent analysis by Gerety *et al.*[3] it seems that in the zone 50°–60° N their correlation between $R - R_0$ and $I_a$ has changed sign from negative (1885–1913) to positive (1914–60)[5]. This change of sign of the correlation, which is

**Table 1** Correlation coefficients, $r$, between smoothed $\overline{R - R_0}$ and Xanthakis' solar activity index Ia, for three latitude zones over the years indicated, as obtained by Xanthakis[2] with his selection of stations, and those obtained by Pittock[1] (for 70°–80° N) and Gerety *et al.*[6] (for 60°–70° N and 50°–60° N) using larger available data samples

| | | $r$ | | | |
|---|---|---|---|---|---|
| Zone | Years | Xanthakis | Larger samples | $N$ | $P$ |
| 70°–80° N | 1912–60 | +0.77 | +0.38 | 19 | ~0.07 |
| 60°–70° N | 1882–1960 | −0.71 | −0.35 | 31 | ~0.05 |
| 50°–60° N | 1914–65 | +0.79 | +0.20 | 21 | ~0.009 |

$P$ is the probability that the data samples in each latitude zone are drawn at random from the same population, where $N$ independent data pairs are assumed to exist. $N$ has been reduced to allow for the computed autocorrelations in the data (including smoothing) according to the standard formula given by Quenouille[9].

0028–0836/79/290254—02$01.00