Table 1 Simple correlation coefficients between the independent variables

	$\ln\{O_3\}_{t-1}$	$\ln T_t$	$\ln I_t$	$\ln H_t$	$\ln W_t$*
$\ln\{O_3\}_{t-1}$	1.00	0.38	0.28	−0.23	−0.09
$\ln T_t$		1.00	0.47	−0.26	−0.31
$\ln I_t$			1.00	−0.68	−0.04
$\ln H_t$				1.00	0.12
$\ln W_t$					1.00

Only 427 cases were actually used in the comparison. Any cases with missing data were eliminated completely from the analysis.

* W_t is the average wind speed (ms^{-1}) between 1000 h and 1600 h at the London Weather Centre.

when converted into arithmetic units this is the factor by which the mean value must be multiplied or divided by to find the range for one standard error. In fact the standard error lines shown in Fig. 2 are apparently drawn at 2 standard errors from the mean. If this is the case then one standard error $\simeq 0.375$ (units $\log_e$) or a factor of 1.45 in arithmetic units. Thus the actual limits on the mean value are +45% and −31%. Stewart *et al.* state a confidence limit of ±26% for predicting elevated levels above 8.0 p.p.h.m. This is incorrect not just because of the above reasoning but also because the standard error lines should not be drawn parallel to the regression lines as shown in Fig. 2. The standard error is a minimum in the 'middle' of the observed range and increases to either side. In other words for elevated ozone levels the standard error will be greater than for the mean value.

(2) The meteorological predictors used in the model should strictly speaking be independent. However, Table 1 shows that they are far from independent. Multicollinearity can mean that computed regression and correlation coefficients vary markedly from sample to sample. One test for multicollinearity is to randomly split a data set into two values and compare the computed coefficients. For the ozone data it was thought that it might be more instructive to split the data into days when the peak hourly concentration was above and below the background peak level of ≈4.0 p.p.h.m. The results of this analysis are discussed in detail elsewhere[3]. Suffice it to say that the regression coefficients for temperature, insolation and wind speed (neglected by Stewart *et al.*) were different for the two data sets as shown in Table 2 below. The strong positive correlation between low ozone levels (less than 4.0 p.p.h.m.) and wind speed, may be because higher wind speeds result in lower concentrations of locally emitted nitric oxide, and hence a smaller loss of background ozone. On the other hand, in light winds a city such as London will act as an effective ozone sink, and very little will be brought in from outside the city. In stronger winds, natural background levels are more likely to persist because of the import of natural levels from outside the city, even though the city itself is still acting as a sink. Whatever the mechanism involved it is obvious that wind speed should not be ignored.

(3) One must be extremely careful in any statistical analysis to fulfil all the assumptions of the statistical model used. I do not present my equations given in Table 2 as substitutes for the model of Stewart *et al.*, but merely to illustrate the point that this type of 'black box' analysis is not applicable to the prediction of ozone levels in London.

JOHN E. THORNES

Department of Geography,
University College,
London, UK

1. Stewart, E. J. *et al. Nature* **263**, 582–584 (1976).
2. Ball, D. J. *G.L.C. Scientific Branch. Rep.* No. E59/EA/R49 (1977).
3. Thornes, J. E. *Prog. phys. Geog.* **1**, 506–517 (1977).

WILLIAMS AND SULLIVAN REPLY—The first of Thornes' main points concerns the underprediction of ozone levels on several days during the summer of 1976. In response to this we feel that some general comments on statistical models are appropriate. Any statistical model is historical, in the sense that its range of applicability to future conditions is limited by the range of conditions used in its formulation. Given this fact and bearing in mind the abnormal meteorological conditions prevalent in the summer of 1976, it is not surprising that ozone levels were underpredicted on some days in that period. The low prediction is most probably due to considerations such as these, rather than as Thornes suggests to any distinction between locally generated and imported ozone which is not appropriate to a gross statistical model such as the one in question. For example, the summer of 1977 has proved to more closely resemble the period for which the model was formulated than did summer 1976. For instance, on one of the days of highest ozone levels, 28 May 1977, the model predicts a maximum value of 8.9 p.p.h.m. compared with an observed 9.5 p.p.h.m.

Table 2 Correlation coefficients and regression equations for $\{O_3\}_t < 4.1$ and $\geqslant 4.1$ and for the total data set

For ozone	$\ln T_t$	$\ln I_t$	$\ln H_t$	$\ln\{O_3\}_{t-1}$	$\ln W_t$	No. of cases
⩾4.1	0.47	0.23	−0.30	0.38	−0.34	220
<4.1	0.09	0.36	−0.24	0.27	0.24	207
All	0.47	0.55	−0.47	0.55	0.13	427

		Multiple r	Standard error of mean (%)
⩾4.1	$\{O_3\}_t = 5.7\ \{O_3\}_{t-1}^{0.19}\ T_t^{0.44}\ I_t^{-0.007}\ H_t^{-0.3}\ W_t^{-0.21}$	0.62	+29%; −22%
<4.1	$\{O_3\}_t = 0.75\ \{O_3\}_{t-1}^{0.19}\ T_t^{0.001}\ I_t^{0.2}\ H_t^{-0.12}\ W_t^{0.24}$	0.50	+38%; −28%
All	$\{O_3\}_t = 0.69\ \{O_3\}_{t-1}^{0.38}\ T_t^{0.35}\ I_t^{0.28}\ H_t^{-0.34}\ W_t^{-0.03}$	0.71	+48%; −32%

On the point of the error of prediction and confidence limits we agree with Thornes comments, but would mention that the definition of standard error used in the note was 1.96 ($=t_{5\%}$) times that used by Thornes. The axis labels in Fig. 2 should be interchanged.

The condition of the independence of the predictor variables has been fully recognised by workers in this field as being not strictly fulfilled so that all such analyses are performed with this in mind and all results are naturally subject to the extent to which, in practice, this is found to be a problem. Indeed in the majority of such regression analyses of pollutant/meteorology relationships it would be difficult to satisfy this condition fully.

The comment that the wind speed was neglected is not strictly correct as we stated that this variable was considered and was found not to contribute significantly, when all the ozone data were taken into account. The findings of Thornes are in agreement with this conclusion. As our purpose was to investigate the gross correlations of all the ozone data with meteorological parameters, it was not felt to be appropriate to analyse in greater detail in the original note.

Notwithstanding these considerations we feel that regression analyses can prove useful in highlighting the important factors in this type of problem as was the intention in our note, but would agree that considerable caution needs to be exercised in attempting to draw detailed conclusions from such studies particularly regarding in this case, mechanisms leading to elevated ozone levels. In the time that has elapsed since the note was published it has become more apparent that more detailed physical models of ozone formation which treat atmospheric chemistry as well as the diffusion and advection of pollutants, are more likely to provide a better understanding of such mechanisms.

M. L. WILLIAMS
E. J. SULLIVAN

Department of Industry,
Warren Spring Laboratory,
PO Box 20, Gunnels Wood Road,
Stevenage, Herts, UK