edge in these compounds, showed that Cu 1 sites are populated by Cu^+ and Cu^{2+}, depending on oxygen deficiency.

What about the other two Cu ions in the 1,2,3 system? Being Cu^{2+} even in $YBa_2Cu_3O_6$, these ions (a pair) may be considered equivalent to the type-3 Cu present in laccases (antiferromagnetically coupled but without a bridging ligand). Though there is some evidence by extended absorption fine structure (EXAFS) and XANES at Y edge for a partial occupancy of the oxygen sites in the Y plane[12], such a result is excluded by diffraction studies.

EXAFS work on haemocyanin active Cu sites bears out some similarity between these structural units and those of $YBa_2Cu_3O_{7-x}$ (refs. 5,9). Comparison of the Cu-Cu distances in haemocyanins and these compounds[9,13] shows that, in oxyhaemocyanins, it is 3.67 Å; upon deoxygenation it decreases to 3.39 Å. On the other hand, Cu 1–Cu 1 distance in the O_7 compound is 3.885 Å, and decreases only slightly to 3.857 Å in the O_6 compound, indicating structural rigidity in the perovskite.

Turning to the spectral information in $YBa_2Cu_3O_{7-x}$, bands are noted[14] at 309 and 335 cm^{-1}. In oxyhaemocyanins drawn from different sources, shoulders are seen at these positions[10]. A special feature of oxyhaemocyanins is the appearance of a band at 1075 cm^{-1}, assigned to the singlet–triplet transition of the peroxide-bridged binuclear Cu^{2+} unit. Arguments for the possible presence of peroxide-like species in 1,2,3 when x is close to zero were already alluded to[6-8]. Raman studies by Czerwosz *et al.*[15] in $YBa_2Cu_3O_{7-x}$ and related compounds (where Ba is partially substituted by Ca or Sr) revealed a significant band at 1100 cm^{-1}. Is this one due to the above feature in oxyhaemocyanins? On the other hand this could also be due to a superoxide. In KO_2 it is seen at 1145 cm^{-1} and in BaO_2 at 1061 cm^{-1}. The 1100 cm^{-1} band could be due to the attachment of a superoxide ion to Cu^+ as an asymmetrical mononuclear complex[16].

In conclusion, it may be seen from the above discussion that the active copper sites in $YBa_2Cu_3O_{7-x}$ do resemble the binuclear, antiferromagnetically coupled and EPR-silent Cu sites in haemocyanins in several features. Up to an oxygen stoichiometry of $O_{6.4}$, only oxide ions are present. On uptake of additional oxygen, i.e. in the range $O_{6.4}$ to $O_{6.8}$, where T_c is composition-dependent, another chemical species in the lattice, like a superoxide ion, would make its presence felt. Further oxidation to O_7 would lead to a peroxide species, with T_c becoming composition-independent. In the overall process of oxidation, the binuclear Cu^+ sites along the b axis are progressively converted to binuclear Cu^{2+} sites with a superoxide or a peroxide bridge. Experimental work on the uptake of Co (or NO) by $YBa_2Cu_3O_{7-x}$, where x is in the range 0.0 to 0.6, could shed additional light on the unique behaviour of Cu 1 sites in such defect perovskites.

1 Fee, J. A., *Struct Bonding (Berlin)*, 1975, 23, 1
2. Solomon, E. I., Penfield, K. W. and Wilcox, D. E., *Struct Bonding (Berlin)*, 1983, 53, 1
3 Solomon, E. I. *et al*, *J Am. Chem Soc*, 1976, 98, 1029
4 Reinhammer, B., *Adv Inorg Biochem*, 1979, 1, 102
5 Report by Ad hoc Research Group for the New Superconductors, *J Elect Engg (Jpn)*, Oct 1987, 63
6 Goodenough, J. B., *Mat Res Bull*, 1988, 23, 401
7 Sarma, D. D., Sreedhar, K., Ganguly, P. and Rao, C. N. R., *Phys Rev*, 1988, B36, 2371
8 Sarma, D. D. and Rao, C. N. R., *Solid State Commun*, 1988, 65, 47
9 Brown, J. M., Powers, L., Kincaid, B., Larrabee, J. A. and Spiro, T., *J Am Chem Soc.*, 1980, 102, 4210
10 Larrabee, J. A. and Spiro, T., *J Am. Chem Soc.*, 1980, 102, 4217
11 Iwazumi, T. *et al*, *Solid State Commun*, 1988, 68, 213
12 Antonini, G. M., Calandra, C., Corin, F., Matacotta, F. C. and Sacchi, M., *Eur. Phys Lett*, 1987, 4, 851
13 Santro, A. *et al*, *Mat Res Bull*, 1987 22, 1007
14 Macfarlane, R. M., Rosen, H. and Saki, H., *Solid State Commun.*, 1987, 68, 631
15. Czerwosz, E. *et al*, *Solid State Commun*, 1987, 64, 535
16 Mcginnety, J. A., *Int. Rev Sci Inorg Chem., Series I* 1972, 5, 233

Received 14 June 1990, accepted 22 March 1991

A parsimonious model for prediction of monsoon rainfall in India

S. A. Paranjpe and A. P. Gore
Department of Statistics, University of Poona, Pune 411 007, India

Recently Gowariker *et al.* have used multiple and power regression involving 15 independent variables for long-range forecasting of monsoon rainfall in India. They have also argued that, when most of the independent variables are 'favourable', almost invariably the monsoon rainfall is normal. In this note we formalize this approach using a parsimonious logistic regression model. The probability of a normal rainfall can be assessed in most cases using only five of the 15 variables. Of these, three are related to temperature and two to wind. This gives us correct results for 36 (out of 38) years, including the exceptional year 1957, when all but one independent variables were unfavourable and yet rainfall was normal.

OF late, successful prediction of monsoon rainfall in India by meteorologists has attracted considerable attention. In a recent paper Gowariker *et al.*[1] explained their approach to long-range forecasting. Gopinathan and Shastri[2] expressed reservations about the model used by Gowariker *et al.* This model is

$$R = c_0 + \sum_{i=1}^{15} c_i x_i^{p_i},$$

where c_i and p_i are model constants, $x_1, x_2, \ldots, x_{15}$ are values of independent variables and R is the rainfall. The rainfall values predicted using this model had a very high correlation (multiple correlation coefficient 0.94) with observed values.

One statistical reservation about this model might be the fact that it involves too many parameters (16 c_i values and 15 p_i values) for the size of the data. This can sometimes lead to unstable estimates. Perhaps because of this, Gowariker *et al.*[1] indicated an alternative way of looking at the data. They classified the value of each independent variable for each year as either favourable (F) to normal rainfall or unfavourable (U) (coded in our analysis as F = 1, U = −1; a zero is used when data are not available). They note that, when most independent variables are favourable, a normal rainfall seems to result. The year 1957 is odd in that all variables but one were unfavourable and yet the rainfall was normal. Their conclusion is that favourableness of the independent variables seems to be a sufficient (though not a necessary) condition for a normal rainfall.

If rainfall is classified as normal or deficient, the multiple regression model cannot be used since it requires a continuous-valued dependent variable. However, the tecnnique of logistic regression (see Hosmer and Lemeshow[3]) can be applied even when the dependent variable is dichotomous. We have subjected the data in Gowariker *et al.*[1] to such an analysis and have found that a smaller number of independent variables can provide a good indication of rainfall.

If $p(x)$ is the probability that the dependent variable will assume the value 1 (in our case, rainfall is normal), the logistic regression model assumes the following relationship between $p(x)$ and x, the vector $(x_1, x_2, \ldots, x_k)$ of independent variables:

$$\ln \frac{p(x)}{1-p(x)} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i = g(x), \text{ say.}$$

Here $\beta_0, \beta_1, \ldots, \beta_k$ are unknown regression coefficients. This can be rewritten as $p(x) = e^{g(x)}/(1+e^{g(x)})$. If $y_1, y_2, \ldots, y_n$ are the observed values of the dependent variable on n occasions, the likelihood of the data can be written as

$$L(y_1, y_2, \ldots, y_n) = \prod_{j=1}^{n} p(x_j) = \prod_{j=1}^{n} \left[\frac{e^{g(x_j)}}{1+e^{g(x_j)}} \right]_j^{y},$$

where x_j is the set of values of k independent variables corresponding to y_j.

The parameters β_i are estimated by maximizing the likelihood (using Gauss–Newton method or Hooke–Jeeves method (see Walsh[4]) if the former fails). Tests of significance of β's are also based on maximized likelihoods. Let L_r represent the maximized likelihood when the model contains r independent variables and L_{r+s} the maximized likelihood when the model contains s extra independent variables. Then $T = -2 \ln (L_r/L_{r+s})$ follows asymptotically (for large n) a χ^2 distribution with s degrees of freedom. If the statistic T is not large enough we conclude that the s extra independent variables are not necessary.

To select sequentially from a set of independent variables, the following procedure due to Hosmer and Lemeshow[3] is used. To select the first variable, one independent variable is tried at a time and maximized likelihood is calculated. The variable that yields the largest value for the maximized likelihood is included. If there are r variables already chosen to be in the model, and the $(r+1)$th variable is to be selected, a similar procedure is followed. The inclusion process is terminated when the additional variable gives an insignificant value for the T statistic described above. Whenever a new variable is included, terms representing its possible interaction with previously included

Table 1. Sequential inclusion of independent variables in the logistic regression model.

Step no. (i)		Variable included	Maximized log likelihood ($\ln L_i$)	T ($-2\ln (L_{i-1}/L_i)$)	P
0		Only intercept	−13.20	—	—
1	(x_1)	500-hPa ridge	−8 32	9.76	<0 01
2	(x_2)	Temperature in northern hemisphere	−5 41	5 82	<0.025
3	(x_3)	10-hPa wind and its interaction with temperature in northern hemisphere	−2 51	5 80	<0 025
4	(x_4)	Temperature in north India,	-8×10^{-4}	5 00	<0 05
	(x_5)	Temperature on east coast and interaction between temperature in north India and 10-hPa wind			

The last column indicates the result of the χ^2 test to judge whether the variable deserves to be included in the model.

variables (e.g. $\beta_{ij} x_i x_j$) are also considered.

In analysing the data, we have tried to follow closely the strategy in Gowariker *et al.*[1] Thus we have used the data for 1958–1980 to fit the model and the data for 1951–1957 and 1981–1988 to test it.

Table 1 gives the sequence of variable selection and related computation. The 500 hPa ridge emerges as the most important single-variable. This is somewhat expected since, in 1957, favourableness of this variable alone was associated with normal rainfall. The step 4 is an example of a case where inclusion of one variable does not improve the model but simultaneous inclusion of more than one variable is useful.

The final model is

$$\ln \frac{p(x)}{1-p(x)} = g(x) = 10.78 + 11.33x_1 + 7.47x_2 + 11.08x_3 - 4.81x_2 x_3 + 6.05x_4 + 7.03x_3 x_4 - 10.61x_5.$$

Two points may be noted about the coefficients here. The interaction between x_2 and x_3 is negative. When the two factors work in the same direction the interaction reduces the chance of normal rainfall while it increases the chance when they work in the opposite directions. Secondly, the effect of x_5 is negative when other variables are present.

The goodness of fit of the model is seen by comparing the estimated values of probability of normal rainfall and actual occurrence. For every one of the 23 years from 1958 to 1980 the two match well. For all 17 years of normal rainfall, the fitted value of probability of normal rainfall was above 0.9999 and for the 6 years of low rainfall it was below 10^{-4}. The fit is quite good. This model was therefore used to 'predict' rainfall for 15 other years, namely 1951–1957 and 1981–1988. The results are given in Table 2.

In this exercise the model was successful in predicting normal rainfall in ten years and deficit in three years.

Table 2. Performance of the logistic model in prediction of monsoon rainfall.

Year	P (estimated probability of normal rainfall)	Actual rainfall
1951	0.03	Deficient
1952	1	Normal
1953	1	Normal
1954	1	Normal
1955	1	Normal
1956	1	Normal
1957	1	Normal
1981	1	Normal
1982	-5×10^{-7}	Deficient
1983	34×10^{-5}	Normal*
1984	1	Normal
1985	1	Normal
1886	2×10^{-5}	Deficient
1987	1	Deficient*
1988	1	Normal

*Cases of wrong prediction.

For one deficient and one normal year, the predictions were wrong.

The performance of the logistic regression model is worthy of note because (i) it uses only five out of 15 independent variables, (ii) it fitted perfectly to data used for estimation, and (iii) it predicted normal rainfall correctly in 1957 when most indicators suggested otherwise. The model by itself is, of course, not always right in prediction. However, it deserves a place among the set of statistical tools used in meteorology.

1. Gowariker, V. *et al.*, *Mausam*, 1988, **40**, 115.
2. Gopinathan, C. K. and Sastry, J. S., *Curr. Sci.*, 1990, **59**, 394.
3. Hosmer, D. W. and Lemeshow, S., *Applied Logistic Regression*, Wiley, 1989.
4. Walsh, G. R., *Methods of Optimization*, John-Wiley and Sons, 1975.

Received 13 December 1990, accepted 19 December 1990

MEETINGS/SYMPOSIA/SEMINARS

National Symposium on Recent Advances in Drought Research

Place: Rubber Research Institute of India, Kottayam
Date: 10–13 December 1991
Sessions on agroclimatology and crop modelling, screening for drought resistance, genetic and molecular basis of drought-resistance traits, relevance of drought-avoidance mechanisms in crop improvement, physiological approaches for drought management, interaction between drought and other stresses.
Contact: Dr K. R. Vijayakumar
Drought Research Symposium
Plant Physiology Division
Rubber Research Institute of India
Kottayam 686 009
Phone: 481 8311 Telex. 888 285 RRII IN
Fax: 481 8317

Second Congress of Toxicology in Developing Countries

Place: New Delhi, India
Date: 24–28 November 1991
Theme is 'Development without destruction'.
Contact: Dr P. K. Seth
Secretary-General, Second CTDC
Industrial Toxicology Research Centre
P. B. No. 80
Lucknow 226 001, India
Phone: (91) 522 248227, (91) 522 232592, (91) 522 242787, (91) 522 240107
Telex: 535 456 ITRC IN