Journal of Coastal Research	14	1	331–336	Royal Palm Beach, Florida	Winter 1998

Particle Size Differentiation of Some Coastal Sands: A Multinomial Logit Regression Approach

Peter Vincent

Department of Geography
University of Lancaster
Lancaster, LA1 4YB, England, U,K.

ABSTRACT

VINCENT, P., 1998. Particle size differentiation of some coastal sands: a multinomial logit regression approach. *Journal of Coastal Research*, 14(1), 331–336. Royal Palm Beach (Florida), ISSN 0749-0208.

This paper describes the application of multinomial logit regression to the discrimination of three groups of coastal sediments at Drigg, Cumbria, UK. Moment measures of particle size were obtained for 150 sand samples taken from the present-day beach, erosion scarps in Holocene dunes, and from glacial-fluvial horizons in eroding Devensian drumlins. Particle size was measured using a Coulter laser diffraction instrument. A multinomial regression model including kurtosis, skewness and standard deviation correctly assigned 86 per cent of the sand samples. Multinomial logit regression is easy to use, provides probabilistic assignments and uses the independent variables in a direct manner, unlike many alternative statistical methods.

ADDITIONAL INDEX WORDS: *Multinomial logit regression, discrimination, coastal sediments.*

INTRODUCTION

A number of researchers over the last 30 years have sought methods and criteria to distinguish between grain-size distributions of sands from different sedimentary environments (FRIEDMAN, 1961, 1979; KLOVAN, 1966; GREENWOOD, 1969; VISHER, 1969; MIOLA and WEISER, 1968; CHAMBERS and UPCHURCH, 1979; VINCENT, 1986, SUTHERLAND and LEE, 1994a). The statistical techniques employed have become more and more sophisticated but they have been concerned with classifying new objects into existing groups, rather than identifying any possible tendency for data to clump together to form groups.

The stimulus for the present paper comes from Sutherland and Lee's recent study of coastal environments at Oahu, Hawaii (SUTHERLAND and LEE, 1994b) At this site, they showed how it was possible to discriminate reasonably successfully between modern beach sub-environments using non-parametric discriminant analysis applied to sieved, beach sand data. A non-parametric approach is advantageous since discriminant function estimators are sensitive to the assumption of normality. In particular, the estimators of the coefficients for non-normally distributed variables are biased away from zero when the coefficient is, in fact, different from zero. As HOSMER and LEMESHOW (1989, p.20) also note, for dichotomous independent variables, the discriminant function estimators will overestimate the magnitude of the association.

In the present paper an alternative procedure, multinomial logit regression, is discussed. This is a more direct, and intuitively more appealing approach. In particular, as SUTHERLAND and LEE (1994b) note, the general problem with the discriminant function, as with many multivariate procedures, is that new variables are created and their meaning, which can sometimes be quite complex, often has to be interpreted from eigenvector scores (JAMES, 1985). This is not the case with the multinomial logit regression (MLR) which can be thought of in the same way as Ordinary Least Squares multiple regression. That is to say, the identity of the independent variables remains throughout the analysis and transformations and interactions can be easily modelled. The particular advantage of the MLR approach is that predicted values are probabilities, and the assignment of individual samples to known sub-environments is itself probabilistic.

In this paper I illustrate the MLR approach by reference to an on-going study of coastal sediments in north west England.

STUDY AREA

The sediments analysed in this paper were sampled on the Cumbrian coast at Drigg (Figure 1). The sediments on this stretch of the Cumbrian coast are all derived from Devensian glacial deposits which thickly mantle the coastal lowlands. Drumlins are well developed in the area and there are good sections exposed along the coast between Seascale and Drigg. These sections reveal the drumlins to be composed of a basal lodgement till, overlain by thick, homogenous, glacio-fluvial sands up to 6 metres thick, which are themselves capped by a thin meltout till (BARNE *et al.*, 1996). The drumlins are presently being eroded by winter storms and glacio-fluvial sands drape the upper beach zone. Between the drumlins there is an extensive, vegetated, coastal dune system. The dunes are up to 10 m high and covered by marram grass (*Ammophila arenaria*) except where they abut the beach.

In this zone there are active erosion scarps and parabolic

96001 received 19 January 1996; accepted in revision 6 September 1996.

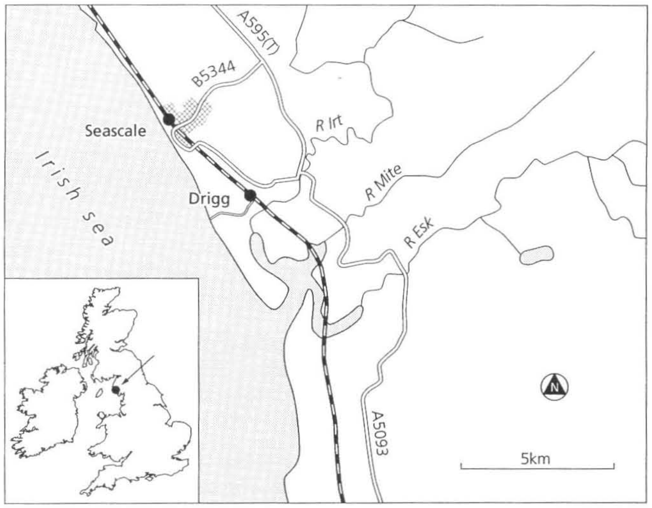

Figure 1. Location of Drigg sampling site, Cumberland, UK.

blow-out basins. The beaches along this stretch of the Irish Sea are a complex of large gravel lags (locally called skaers) derived from glacial tills and drumlins, and ridge and runnel sands. At the back of the beach, sands derived from dune scarps and drumlin erosion partially bury beach sands and gravels.

FIELD AND LABORATORY METHODS

Fifty random samples from each of the dune, beach and glacio-fluvial environments were collected for granulometry. Beach samples were collected from the uppermost few laminae using a flat spatula; sand samples from active dune scarps and glacial-fluvial erosion scars were both collected as small channel samples.

In the laboratory all samples were washed with fresh tap water to remove salt. Particle size measurements were then obtained using Coulter LS laser diffraction granulometer. This instrument is capable of measuring particle size in the range 2000 μ–0.4 μ. Suitable diffraction conditions usually required about 5 g of sample. Moment measures (GRIFFITHS, 1967) based on particle size by per cent volume are one of several different types of statistical output available from the Coulter apparatus. In the present paper the following moment-based statistics are used.

$$\text{Mean } (\bar{x}) = \frac{\Sigma\, fm_{\mu}}{100} \tag{1}$$

where f is percentage in each size class, and m_{μ} is the midpoint of each size class.

$$\text{Standard Deviation } (\sigma) = \left[\frac{\Sigma\, f(m_{\mu} - \bar{x})^2}{100}\right]^{1/2} \tag{2}$$

$$\text{Skewness } (\alpha_3) = 1/100\sigma^{-3}\, \Sigma\, f(m_{\mu} - \bar{x})^3 \tag{3}$$

$$\text{Kurtosis } (\alpha_4) = 1/100\sigma^{-4}\, \Sigma\, f(m_{\mu} - \bar{x})^4 \tag{4}$$

SAND GRANULOMETRY RESULTS

Figure 2 illustrates three representative particle size curves for sands at Drigg obtained with the Coulter diffraction instrument.

There are several differences in the grain size curves in terms of the moment statistics and these are brought out further in Figures 3 and 4 in which pairs of moments for all 150 samples have been plotted. It can be observed in Figure 3, that kurtosis seems to be quite important in discriminating glacial-fluvial sand from beach and dune sand.

And in Figure 4 there is a good deal separation of sands from the three environments in terms of their standard deviation, but less so in terms of their skewness.

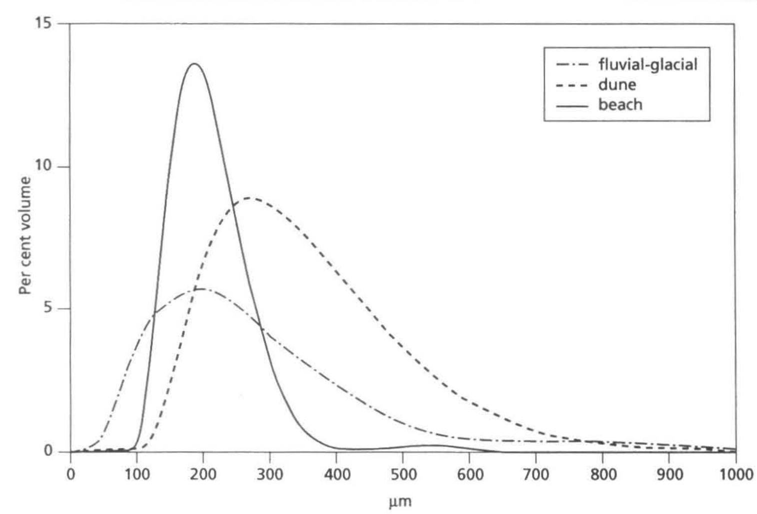

Figure 2. Representative grain size curves for the three sediment groups.

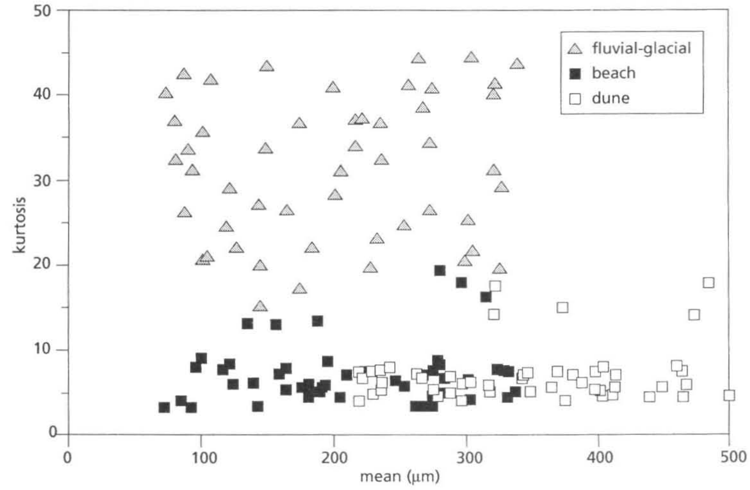

Figure 3. Scatter plot of kurtosis versus mean particle size.

MULTINOMIAL LOGIT REGRESSION

MLR is a generalisation to $k > 2$ categories of the binomial logit model fitted by VINCENT (1986). For that particular model, the dichotomous, dependent, response variable was coded '1' for dune sand samples and '0' for beach samples. For the three site MLR in the present study the response variable has 3 categories, coded as follows: beach sands, '1'; glacial-fluvial sands, '2'; dune sands, '3'. The ordering of the coding in this type of polytomous logit model is unimportant (HOSMER and LEMESHOW, 1989, p.216). The independent variables are the four moment-based statistics described in the previous section.

In the present paper, the multinomial regression model with continuous covariates was fitted using the statistical package GLIM4 (1992) using the method described by FRANCIS *et al.* (1992). A slower method is also possible in GLIM3.77 using the macros described by AITKIN and FRANCIS (1992).

For $n = 50$ observations from $j = 3$ environments, the multinomial logit regression model can be written as:

$$\log\left(\frac{p_{ji}}{p_{1i}}\right) = \boldsymbol{\beta}_j'\mathbf{x}_i; \qquad i = 1, \ldots, n;$$

$$j = 1, \ldots, 3; \qquad \beta_1 = 0 \qquad (5)$$

Where P_{ji} is the probability of the n'th sample belonging to the j'th sub-environment and $\mathbf{x}$ is a vector of explanatory variables.

After some re-arrangement of Equation 5 the following 3 multinomial logit regressions are obtained (see— Appendix for the derivation):

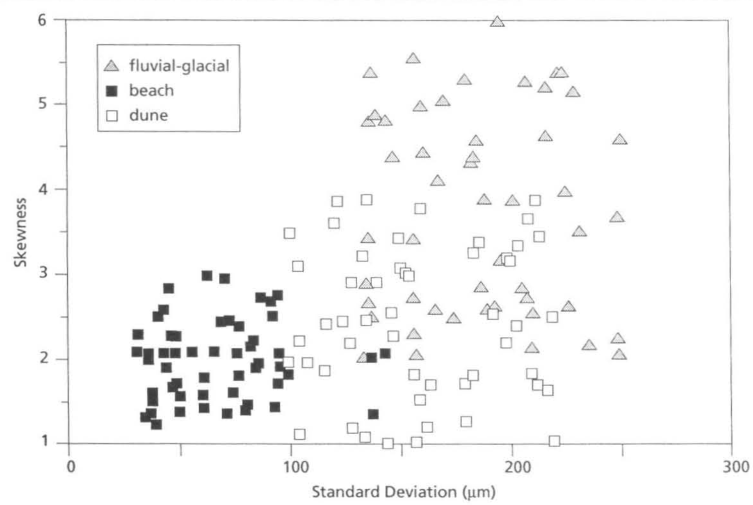

Figure 4. Scatter plot of skewness versus standard deviation.

Table 1. *Goodness-of-fit statistics for all single independent variable models (GM = grand mean; ME = mean; ST = standard deviation; SK = skewness; KU = kurtosis).*

Model	Scaled Deviance	DF	Reduction in Scaled Deviance
GM	329.58	298	
GM + ME	257.30	296	72.28
GM + ST	167.54	296	162.04
GM + SK	256.64	296	72.94
GM + KU	155.68	296	173.90

$$p_{1i} = \frac{1}{1 + e^{\beta'_2 x_i} + e^{\beta'_3 x_i}}$$

$$p_{2i} = \frac{e^{\beta'_2 x_i}}{1 + e^{\beta'_2 x_i} + e^{\beta'_3 x_i}}$$

$$p_{3i} = \frac{e^{\beta'_3 x_i}}{1 + e^{\beta'_2 x_i} + e^{\beta'_3 x_i}}$$

where $p_{1i} + p_{2i} + p_{3i} = 1$

Output from this multinomial regression model comprises a set of probability triples. From these we can assign samples to one of the three environments in the study. If, for example, a predicted triple was {0.2, 0.7, 0.1} we would assign the sample to the glacio-fluvial environment (coding 2) since this has the highest probability. Goodness-of-fit for a sequence of multinomial regression models fitted in GLIM is commonly assessed by examining the changes in a log-likelihood measure called the scaled deviance. Differences in the scaled deviances are approximated by a χ^2 distribution. AITKIN *et al.* (1989) provide a full explanation of the procedure. The fit of the final regression model can also be judged in terms of the proportion of samples which are correctly assigned to sites. This method is illustrated in the present paper.

RESULTS

In Table 1 the effect of adding a single explanatory variable to the intercept term or grand mean (GM) can be observed. In particular the drop in the scaled deviance due to the explanatory power of the variable in question should be noted.

The results of an analysis of deviance for a sequence multinomial regression models is show in Table 2. In this table it can be observed that there is no significant drop in the scaled deviance when ME (mean) is added to the regression model. The scaled deviance drops by 3.50 which is less than $\chi^2_{0.05,2} = 5.99$. It can, therefore, be concluded that the mean particle size has no explanatory power in terms of distinguishing between the three sets of sand samples from the Drigg site.

The logit regression equations for the final model shown in Table 3 are as follows, and since the probabilities must sum to 1 only two equations need to be presented in full for the n = 3 case:

Table 2. *Sequential MLR models and goodness-of-fit (model codes as in Table 1).*

Model	Scaled Deviance	DF	Reduction in Scaled Deviance
GM	329.58	298	—
GM + KU	155.68	296	173.90
GM + KU + ST	45.30	294	110.38
GM + KU + ST + SK	17.14	292	28.16
GM + KU + ST + SK + ME	13.64	290	3.50

$$\hat{p}_{1i} = 1 - \hat{p}_{2i} - \hat{p}_{3i}$$

$$\hat{p}_{2i} = e^{35.96 \;\; 1.554SK_i \;\; 0.1171ST_i \;\; 1.1144KU_i}$$

$$\div\; 1 + e^{35.96 \;\; 1.554SK_i \;\; 0.1171ST_i \;\; 1.1144KU_i}$$

$$+\; e^{21.87 \;\; 0.4724SK_i \;\; 0.0217ST_i \;\; 0.9620KU_i}$$

$$\hat{p}_{3i} = e^{21.87 \;\; 0.4724SK_i \;\; 0.0217ST_i \;\; 0.9620KU_i}$$

$$\div\; 1 + e^{35.96 \;\; 1.554SK_i \;\; 0.1171ST_i \;\; 1.1144KU_i}$$

$$+\; e^{21.87 \;\; 0.4724SK_i \;\; 0.0217ST_i \;\; 0.9620KU_i}$$

Using the above logit equations it is then possible to construct a table of assignments from the probability triples. The predicted environmental assignments are shown in Table 3. It can be observed that, of the 150 samples analysed in the study, 23 were misclassified using moment statistics.

DISCUSSION

The multinomial logit approach to sediment classification is intuitively appealing given its similarity to multiple regression. Like multiple regression it retains the independent variables in their original form thus making model development and interpretation easy. And, unlike alternative multivariate procedures, such as discriminant functions, there is no problematic interpretation of eigenvector scores.

Additionally, as with multiple regression, complex transformations of the independent variables are easy in MLR allowing very sophisticated models to be developed. Furthermore, since the predicted values from the multinomial logit regression are probabilities the approach has the advantage of predicting the odds in favour of group membership. At Drigg, only three populations of sand have been studied but there is no theoretical limit to the number of populations that can be analysed using the MLR approach.

When interpreting the results presented in this paper it is important to emphasise that the three parent sand bodies do not form a contemporaneous, decoupled system (BAUER, 1991). That is to say, they have not been sorted into three subpopulations from a single source deposit. If this were the

Table 3. *Classification table based on the final MLR model.*

Observed	Beach	Dune	Glacio-fluvial
Beach	39	7	4
Dune	6	43	1
Glacio-Fluvial	5	—	45

case it would be very difficult to explain why the dune sands were coarser than those sampled from the modern beach and the glacio-fluvial scarps. At Drigg, as with much of the glaciated coastline in northern Britian, sediment inputs into the mordern beach system are derived from a wide variety of Quaternary and Holocene sediments, which may differ greatly in age. It is thought that the initiation of the dunes at Drigg dates from about 5000 BP with the deflation of Holocene storm beaches, themselves derived from tills disturbed by rising post-glacial sea-levels (KING, 1976). At Drigg, therefore, it is not surprising that sediments eroded from vegetated, Holocene dunes have not been decoupled from the present-day beach even though their spatial proximity might suggest otherwise.

Of note is the fact that the kurtosis proved to be the most important explanatory variable in the multinomial logit model. At first glance, this may appear to run counter to the commonly held view that the kurtosis is usually a weak discriminant in the differentiation of dune and other environments (PYE and TSOAR, 1990) and it is readily observed in Figure 3 that there is, in fact, no difference in the kurtosis of the Drigg beach and dune sands. Kurtosis enters the regression model because of the very clear separation of glacio-fluvial sands from the other two sediment groups.

Figure 3 also shows that there is a good deal of overlap in the range of the mean particle size which confirms the poor explanatory power of this variable.

CONCLUSIONS

Multinomial logit regression modelling provides a straightforward way to approach the environmental assignment problem using grain size data. The original explanatory variables are retained and complicated interpretations of eigenvectors are done away with. At Drigg, a logit regression involving the kurtosis, skewness and standard deviation produced a model which successfully assigned 85 per cent of the samples to their environments.

ACKNOWLEDGEMENTS

I am grateful to Brian Francis and Damian Berridge, Centre for Applied Statistics, Lancaster University, UK, for help with the GLIM implementation of the multinomial logit model. Paul Williams, Department of Geography, Lancaster University helped with the laser diffraction measurements and Nicki Higgit constructed the figures.

APPENDIX

For n observations from j = 3 environments, we can write the multinomial logit regression model as:

$$\log\left(\frac{p_{ji}}{p_{1i}}\right) = \boldsymbol{\beta}_j'\mathbf{x}_i; \qquad i = 1, \ldots, n;$$
$$j = 1, \ldots, 3; \qquad \beta_1 = 0 \qquad (1)$$

Where is the probability of the i'th sample belonging to the j'th sub-environment and is a vector of explanatory variables. For the i-th observation at each of the three environments we have

$$\text{Environment 1: } \log\left(\frac{p_{1i}}{p_{1i}}\right) = 0 \rightarrow \frac{p_{1i}}{p_{1i}} = 1 \qquad (2)$$

$$\text{Environment 2: } \log\left(\frac{p_{2i}}{p_{1i}}\right) = \beta_2'x_i \rightarrow \frac{p_{2i}}{p_{1i}} = e^{\beta_2'x_i} \qquad (3)$$

$$\text{Environment 3: } \log\left(\frac{p_{3i}}{p_{1i}}\right) = \beta_3'x_i \rightarrow \frac{p_{3i}}{p_{1i}} = e^{\beta_3'x_i} \qquad (4)$$

Gathering the L.H.S.

$$(1) + (2) + (3): \rightarrow \frac{p_{1i}}{p_{1i}} + \frac{p_{2i}}{p_{1i}} + \frac{p_{3i}}{p_{1i}} = \frac{p_{1i} + p_{2i} + p_{3i}}{p_{1i}}$$
$$= \frac{1}{p_{1i}} \qquad (5)$$

Gathering the R.H.S.

$$(1) + (2) + (3): \rightarrow 1 + e^{\beta_2'x_i} + e^{\beta_3'x_i} \qquad (6)$$

Equating (5) and (6)

$$\rightarrow \frac{1}{p_{1i}} = 1 + e^{\beta_2'x_i} + e^{\beta_3'x_i}$$

So that,

$$p_{1i} = \frac{1}{1 + e^{\beta_2'x_i} + e^{\beta_3'x_i}}$$

$$p_{2i} = \frac{e^{\beta_2'x_i}}{1 + e^{\beta_2'x_i} + e^{\beta_3'x_i}}$$

$$p_{3i} = \frac{e^{\beta_3'x_i}}{1 + e^{\beta_2'x_i} + e^{\beta_3'x_i}}$$

LITERATURE CITED

AITKIN, M.; ANDERSON, D.; FRANCIS, B., and HINDE, D., 1989. *Statistical Modelling in GLIM.* Oxford, Clarendon Press.

AITKIN, M. and FRANCIS, B., 1992. Fitting the multinomial logit model with continuous covariates in GLIM. *Computational Statistics and Data Analysis, 14,* 89–97.

BARNE, J.H.; ROBSON, C.F.; KAZNOWSKA, S.S.; DOODY, J.P., and DAVIDSON, N.C., 1996. *Coasts and Seas of the United Kingdom. Region 13 Northern Irish Sea Colwyn Bay to Stranraer, including the Isle of man.* Peterborough: Joint Nature Conservation Committee.

BAUER, B.O., 1991. Aeolian decoupling of beach sediments. *Annals of the Asssociation of American Geographers,* 81, 290–303.

CHAMBERS, R.L. and UPCHURCH, S.B., 1979. Multivariate analyses of sedimentary environments using grain size frequency distributions. *Journal of Mathematical Geology, 11,* 27–43.

FRANCIS, B.; GREEN, M., and CLARKE, M., 1992. Model fitting applications in GLIM4. *In* Proceedings of GLIM92 and the 7th International Workshop on Statistical Modelling. *Lecture Notes in Statistics, 78,* 6–12. Berlin: Springer-Verlag.

FRIEDMAN, G.M., 1961. Distinction between dune, beach and river sands from their textural characteristics. *Journal of Sedimentary Petrology, 31,* 514–529.

FRIEDMAN, G.M., 1979. Differences in size distributions of populations of particles among sands of various origins. *Sedimentology, 26,* 1–30.

GREENWOOD, B., 1969. Sediment parameters and environmental discrimination: an application of multivariate statistics. *Canadian Journal of Earth Science, 6,* 1347–58.

GRIFFITHS, I.C., 1967. *Scientific Method in the Analysis of Sediments. New York:* McGraw-Hill.

HOSMER, D.W. and LEMESHOW, S., 1989. *Applied Logit Regression*. New York: Wiley.

JAMES, M., 1985. *Classification Algorithms*. London: Collins.

KING, C.A.M., 1976. *Northern England*. London: Methuen.

KLOVAN, J.E., 1966. The use of factor analysis in determining depositional environments from grain-size distributions. *Journal of Sedimentary Petrology, 36*, 115–125.

MIOLA, R.J. and WEISER, D., 1968. Textural parameters: an evaluation. *Journal of Sedimentary Petrology, 38*,57–69.

PYE, K. and TSOAR, H., 1990. *Aeolian Sand and Sand Dunes*. London: Unwin and Hyman.

SUTHERLAND, R.A. and LEE, C-T., 1994a) Application of the log-hyperbolic distribution to Hawai'ian beach sands. *Journal of Coastal Research, 10*, 251–262.

SUTHERLAND, R.A. and LEE, C-T., 1994b. Discrimination between coastal subenvironments using textural characteristics. *Sedimentology, 41*, 1133–43.

VINCENT, P., 1986. Differentiation of modern beach and coastal dune sands—a logit regression approach using the parameters of the hyperbolic distribution. *Sedimentary Geology, 49*: 167–76.

VISHER, G.S. 1969. Grain size distributions and depositional processes. *Journal of Sedimentary Petrology, 39*, 1074–106.