

## West African rivers as biogeographic islands: species richness of fish communities

Bernard Hugueny

Laboratoire d'Ichtyologie générale et appliquée, Museum National d'Histoire Naturelle, 43 rue Cuvier, 75231 Paris cedex 05, France



13 NOV. 1995

O.R.S.T.O.M. Fonds Documentaire

N° 42798

Cote B E1 M

**Summary.** Some factors influencing the species richness of West African fish communities were studied in a sample of 26 rivers using four habitat and hydrologic variables. Analysis of a larger sample of 39 rivers showed that species richness was positively related to area. A power function with an exponent of 0.32 gave the best fit. As the surface area used was that of the catchment area and not that (unknown) of the river, the biological significance of this relationship and the possibilities of comparison were limited. Ridge regression analysis and forward stepwise selection indicated that a model that explained  $\ln(\text{species richness})$  as a function of  $\ln(\text{mean annual discharge})$  and  $\ln(\text{catchment surface area})$  was best, accounting for 90% of the variance of the dependent variable. The combination of surface area and discharge was presumed to act through the volume of water available for the fishes and habitat productivity. Habitat diversity, measured by the diversity of the terrestrial vegetation covering the catchment area, had no significant positive effect when surface area was used in the regression. Rivers ("islands") should have fewer species than tributaries of similar size since, for fishes within a river system ("continent"), there is free circulation between all its branches. The model derived from the river data underestimated the species richness of a sample of 11 tributaries. This was compatible with the hypothesis of higher population extinction rates in insular biotopes. The residuals of the linear model did not show random geographical distribution; the rivers in some areas had more species than expected. The possibility that historical factors, especially Quaternary climatic variations, might cause this distribution is discussed.

**Key words:** Species richness – Insular biogeography – Freshwater fish communities – West Africa – Forest refuge hypothesis

---

Since the work of MacArthur and Wilson (1963, 1967), it has been customary in community ecology to consider environments surrounded by habitats unfavourable to the biological group studied as biogeographic islands. Some examples are mountaintops for boreal species (Brown 1971), caves for cave-dwelling species (Vuilleumier 1973) and forest woodlots for forest species (Martin 1980).

Insofar as rivers are separated from each other by barriers which strictly aquatic animals cannot pass, it is justifi-

able to consider them as biogeographic islands but, up to now, only a few attempts have been made (Daget and Ittis 1965; Daget 1968; Sepkoski and Rex 1974; Daget and Economidis 1975; Welcomme 1979; Livingstone et al. 1982; Eadie et al. 1986). Freshwater fish communities in Africa have been analysed three times in this context and species-area relationships have been emphasized. Daget and Ittis (1965) tried to check Preston's (1962) predictions of the value of the exponent of Arrhenius's law, where the species richness of a community increases in proportion to a power function of the surface area. As the water surface of a river cannot be measured, Daget and Ittis, like the following authors, used the catchment surface area. Their study included 13 rivers, mostly in the Ivory Coast. Welcomme (1979) described species-area relationships for 25 rivers throughout Africa. The most complete study was that of Livingstone et al. (1982): data on 26 rivers were analysed by multiple regression with species richness as the dependent variable and the surface area of the drainage, length of the main branch, and discharge at the mouth as predictor variables. The principal results of their study was that discharge was a better predictor of species richness than surface area.

For about 10 years, ichthyologists of ORSTOM (Institut Français de Recherche Scientifique pour le Développement en Coopération) have participated in a program studying water courses in West Africa to evaluate the impact on the fish communities of anti-simuliid insecticides used by the World Health Organisation in the prevention of onchocerciasis (Lévêque et al. 1988). This program has greatly enlarged our knowledge of West African ichthyofauna, and it appears that some species richness values used in the studies cited above are underestimated. In this study we analyse species richness of West African rivers again in the light of these new data. Species richness is compared in rivers and in a group of 11 tributaries that can be considered, for aquatic animals, as equivalent to "continental" situations since, within a river system, species circulate freely between all the branches. The aim of this comparison is to test whether insularity affects species richness of West African rivers.

## Methods

### Data

Of the 39 rivers studied, all except in Nile, which empties into the Mediterranean, discharge into the Atlantic ocean

along the West African coast from the mouth of the Senegal to that of the Zaïre. A sample of 11 tributaries was studied in the drainage areas of the Niger, Volta, Sassandra, and Bandama.

The following variables were measured for each river:

*SR: species richness.* The species lists were compiled from existing publications, unpublished reports, and collections of the Museum National d'Histoire Naturelle de Paris (Paris, France), the Musée Royal de l'Afrique Centrale (Tervuren, Belgium) and the British Museum (London, UK). Species capable of covering long distances in the sea were excluded from the analysis, even if present in fresh water, because for these species rivers are not equivalent to islands. When catchment areas included one or more lakes, the lacustrine species were not included. These species clearly show different speciation processes from riverine species (Lowe McConnell 1975). Many lakes have a high percentage of endemic fish, which makes it difficult to compare the species richness of a catchment area containing lacustrine ichthyofauna with that of catchments without lakes. For some rivers, the values of species richness thus obtained were much higher than those previously used by Livingstone et al. (1982).

*S: surface area* (in km<sup>2</sup>). The surface area was that of the drainage area found in the literature or measured on maps at 1/1 000 000.

For 26 of the 39 rivers we also studied:

*DV: diversity of terrestrial vegetation types* present in the drainage area. The proportion of the drainage surface area occupied by each type of vegetation was used to determine Shannon's diversity index using Napierian logarithms. The vegetation types follow White (1983).

*PF: percentage of the drainage area covered by lowland rain forest* and associated transition types. These are forest types 1-4, 8, 9, and 11-13 of White's vegetation map (1983).

*AD: annual mean discharge* (in m<sup>3</sup>·s<sup>-1</sup>). This was determined from the station closest to the river mouth (54%-100% of total drainage area depending of the river) and for the longest possible period of time (1-66 years, depending on the river).

The values of the variables for each river are shown in Table 1.

#### Statistical methods

The contribution of hydrology and habitat factors to between-river variation in species richness was analysed by multiple regression, as is common in this type of study.

When colinearity between the predictor variables is high, an estimate of regression coefficients by the least-squares method may lead to erratic values whose sign may even be reversed compared to the real values. One way of estimating colinearity between variables is to calculate the maximum variance inflation factor (*VIF*, see Appendix 2). If this is higher than 10, more suitable methods should be used, such as ridge regression (Hoerl and Kennard 1970a, b). The multiple regression equation is modified by adding an extra parameter, *k*, which limits the length of the regression coefficient vector. The analysis is based on

Table 1. Untransformed data for the 39 rivers studied

Rivers	Species richness	Surface (km) <sup>2</sup>	Mean annual discharge (m <sup>3</sup> ·s <sup>-1</sup> )	Vegetation diversity	Forest (%)
Nile	127	3349000	2640	2.62	8
Senegal	110	441000	687	0.97	3
Gambia	89	77000	170	0.50	20
Tomine	55	23200	-	-	-
Konkouré	74	16470	353	0.00	100
Kolonté	58	7540	-	-	-
Jong	65	7750	-	-	-
Sewa	61	19050	-	-	-
Moa	44	18760	-	-	-
Mano	45	8260	-	-	-
Loffa	35	13190	-	-	-
St. Paul	76	18180	-	-	-
Nipoué	63	11920	-	-	-
Cavally	62	28850	384	0.00	100
Dodo	15	850	-	-	-
Nero	18	985	16	0.00	100
San Pedro	22	3310	31	0.00	100
Sassandra	71	75000	513	0.88	91
Boubo	38	4690	32	0.00	100
Bandama	91	97000	392	0.95	76
Agnébi	56	8520	50	0.00	100
Mé	39	3920	32	0.00	100
Comoé	90	78000	206	0.97	44
Bia	37	9730	81	0.00	100
Volta	147	398371	1260	1.08	19
Mono	53	22000	104	0.63	33
Oueme	92	50000	220	0.37	12
Ogun	67	22370	-	-	-
Niger	207	1125000	6100	1.88	25
Cross	45	48000	-	-	-
Mungo	29	4570	-	-	-
Wouri	52	11500	308	0.00	100
Sanaga	126	135000	2060	0.64	100
Nyong	79	27800	443	0.00	100
Lobé	32	2305	102	0.00	100
Ntem	100	31000	348	0.00	100
Ogowe	181	205000	4758	0.68	100
Niari	83	56000	913	0.69	100
Zaïre	690	3457000	40487	2.08	65

Except the Nile, rivers are ordered according to the position of their mouth along the Atlantic coast

the change in coefficient values as a function of *k* (the ridge trace, see Appendix 1).

Variables were selected from the results of ridge regression. We also used forward stepwise selection by the least-squares method. At each iteration, the variable showing the highest partial correlation with the dependent variable was determined and then included in the model only if the correlation was significant at the 5% level. When no variable could be added to the model, the procedure was finished.

Model quality was determined by studying residuals (observed values minus the values predicted by the regression model): their distribution was analysed as a function of the dependent variable and their possible spatial autocorrelation was determined.

The independence of residual values as a function of species richness value was checked by analysing the sequence of residual signs using a test proposed by Draper

Table 2. Species-area relationships

Variables	Intercept	Slope	<i>r</i>	<i>I</i>	<i>Z</i>
SR-S	60.58	$1.07 \times 10^{-4}$	0.764**	0.047	-5.19**
SR-lnS	-302.46	38.07	0.683**	0.032	-2.05*
lnSR-S	4.03	$5.63 \times 10^{-7}$	0.619**	0.264*	-5.18*
lnSR-lnS	0.84	0.32	0.899**	0.044	0.65

SR = species richness; S = surface area; *r* = correlation coefficient; *I* = Moran's *I*; *Z* = sign test of the residuals

\*  $P < 0.05$ , \*\*  $P < 0.001$

and Smith (1966). The aim of this test is to detect whether, once the residuals are arranged as a function of dependent variable value, those of the same sign tend to be grouped more than would occur at random; if they do group together, it may be assumed that there is no linearity.

To evaluate to what degree the multiple regression models we used would account for possible geographic trends, we tested for spatial autocorrelation of the residuals using Moran's *I* (Sokal and Oden 1978a, b), which tests whether the value of a variable at one locality is independent of values of the variable at neighbouring localities. Positive autocorrelation shows that high values tend to be spatially grouped; in our case, this indicated that we might have failed to include a factor related to geography in our model. To use Moran's *I* it was necessary to construct a proximity matrix for the drainage areas. We decided to consider two drainage areas as connected if they were contiguous at at least one point. As only one positive spatial autocorrelation could be interpreted, a one-tailed test was used (*I*-test for small samples; Sokal and Oden 1978a, b).

## Results

### Species-area relationships

Table 2 summarizes the different species-area relationships tested in the sample of 39 rivers.

In terms of the coefficient of determination, the best fit was obtained by the power relation (after double transformation of the data by Napierian logarithms) with an exponent of 0.32 (Fig. 1). This also gave the most satisfactory distribution of the residuals, since residual sign was distributed at random in relation to the dependent variable.

There was no spatial autocorrelation of the residuals, indicating that the geographic factor in the variation in species richness as a function of surface area was well described by the model.

### Multiple regressions

Knowing that discharge was proportional to drainage surface area multiplied by precipitation in the drainage area, the surface area and discharge variables were integrated into the analysis after transformation to Napierian logarithms to ensure a linear relationship between them. The relationship between species richness and each of the two transformed variables was also more linear when species richness was replaced by its logarithm. Logarithmic transformation of the percentage of forests and vegetation diversity index did not significantly change the results, so these two variables were retained untransformed in the analysis.

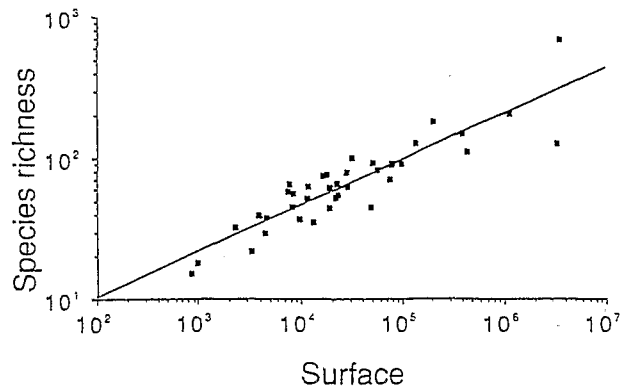


Fig. 1. The relationship between species richness and catchment surface area ( $\text{km}^2$ ) in 39 West African rivers

Table 3. Correlation matrix of the variables used in regression analysis

	ln S	ln AD	DV	PF
ln SR	0.91	0.93	0.73	-0.43
ln S		0.90	0.90	-0.63
ln AD			0.75	-0.31
DV				-0.64

SR = species richness; S = surface area; AD = annual mean discharge; DV = diversity of terrestrial vegetation types; PF = percentage forest cover (see text)

Table 4. Multiple regression of ln(species richness) versus the four predictor variables. Coefficients are estimated by the least-squares method

Coefficients						
ln S	ln AD	DV	PF	Intercept	VIF	<i>R</i>
0.229	0.229	-0.305	-0.001	0.771	23.53	0.955
<i>P</i> = 0.048	<i>P</i> = 0.017	<i>P</i> = 0.065	<i>P</i> = 0.778			<i>P</i> < 0.001

VIF = maximum variance inflation factor; *R* = multiple correlation coefficient; variable abbreviations as in Table 3

Table 3 shows the correlation matrix between variables. Some predictor variables were highly intercorrelated, such as ln(surface area) with ln(discharge) and with the vegetation diversity index. The result of multiple regression with all the predictor variables is presented in Table 4. The maximum variance inflation factor was 23, indicating high collinearity between the predictive variables and suggesting prudence when examining the results obtained by classic multiple regression. These indicated that two variables had positive coefficients. ln(surface area) and ln(discharge), the coefficients being significantly different from zero in both cases. The two other variables had negative coefficients, one of which (vegetation diversity) approached statistical significance.

The multiple correlation coefficient was high (0.955); this value was highly significant.

The results obtained after ridge regression for *k* values from 0 to 1 are summarized in Fig. 2, which shows changes in the coefficients of the regression equation as a function of *k*. Since the variables were standardized beforehand, co-

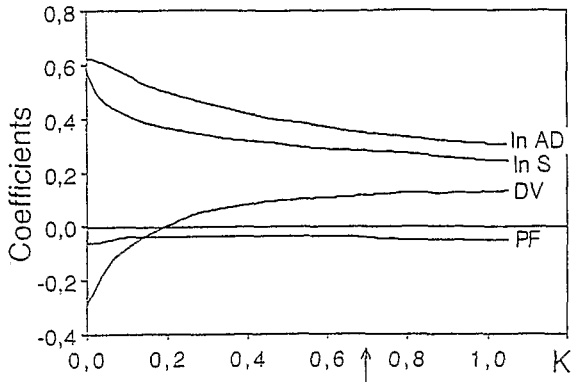


Fig. 2. The ridge trace for the rivers data. Arrow shows the value of  $k$  for which  $VIF=1$ . See text for abbreviations

efficient amplitude could be compared directly. The maximum variance inflation factor was 1 at about  $k=0.7$ . It is interesting to note that the sign of one of the coefficients changed after coefficient value stabilized at about  $k=0.5$ . The standardized coefficient of vegetation diversity index increased from  $-0.29$  to  $0.10$ . A positive value agreed better with the effect of habitat diversity on species richness usually found, but it could not be determined whether this was statistically significant, since the tests usually employed cannot be applied in the case of ridge regression (Hocking 1976). No matter what the  $k$  value was, the forest percentage coefficient remained near to zero, while the coefficients of surface area and discharge remained high.

If this curve is used to select model predictor variables, Hocking (1976), following Hoerl and Kennard (1970 a, b), proposes that variables with a coefficient near zero or varying rapidly with  $k$  should be eliminated. This led us to eliminate forest percentage and the vegetation diversity index, keeping only surface area and discharge.

Forward stepwise selection led us to integrate  $\ln(\text{discharge})$ , then  $\ln(\text{surface area})$  and to stop there. We thus kept the same variables by this method as those retained after a study of the coefficient changes as a function of  $k$ .

We therefore kept the model integrating  $\ln(\text{discharge})$  and  $\ln(\text{surface area})$  as a predictor of  $\ln(\text{species richness})$ . The results are given in Table 5. The maximum variance inflation factor was not more than 5 and coefficient estimation using the least-squares method was acceptable. The multiple correlation coefficient ( $0.947$ ) was highly significant.

The sign test showed that the residuals were distributed at random as a function of dependent variable value. However, the spatial autocorrelation of the residuals was significant. Residual positive values were grouped in three areas: (1) Gambia and Guinea (the Gambia and the Konkouré rivers); (2) a region including coastal rivers between the Boubo (eastern Ivory Coast) and the Oueme (Benin); (3) Cameroon, Gaboon and Zaïre (the Nyong, Ntem, Ogowe, and Zaïre).

#### Species richness in tributaries

The characteristics of the 11 tributaries used in this study are shown in Table 6. Figure 3 compares the observed values of species richness in tributaries with those calculated from the predictive model for rivers. The predicted values

Table 5. Multiple regression of  $\ln(\text{species richness})$  versus  $\ln(\text{catchment surface area})$  and  $\ln(\text{mean annual discharge})$ . Coefficients are estimated by the least-squares method

Coefficients						
$\ln S$	$\ln AD$	Intercept	$VIF$	$I$	$Z$	$R$
0.135	0.245	1.504	5.15	0.239	$-0.811$	0.947
$P=$	$P=$			$P<$	$P>$	$P<$
0.021	0.001			0.05	0.05	0.001

Abbreviations as in Tables 3 and 4

Table 6. Untransformed data for the 11 tributaries studied

Tributaries	Farthest down-stream locality	Species richness	Surface (km <sup>2</sup> )	Mean annual discharge (m <sup>3</sup> ·s <sup>-1</sup> )
Niandan (Niger)	Baro	81	12770	252
Bénoué (Niger)	Lokoja	120	64000	382
Black Volta (Volta)	Bamboi	96	134200	247
Nasia (Volta)	Nasia	36	5175	30
Daka (Volta)	Yendi	28	1214	7
Yani (Bandama)	Seguela	39	3000	18
Marahoué (Bandama)	Mankono	65	6700	27
N'Zi (Bandama)	Dimbokro	68	24100	64
Kan (Bandama)	Tiebissou	36	1200	3
N'Zo (Sassandra)	Guiglo	40	6410	84
Sassandra (Sassandra)	Guessabo	62	35000	308

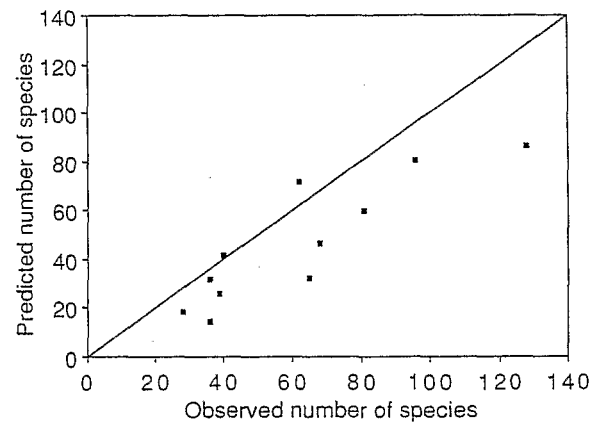


Fig. 3. The relationship between the number of species observed in 11 tributaries and the number of species predicted by the following model:  $\ln(\text{species richness}) = (\ln(\text{mean annual discharge}) \times 0.245) + (\ln(\text{catchment surface area}) \times 0.135) + 1.504$ . The straight line represents points for which predicted value equals observed value

tended to be lower than the observed ones, as shown in 9 out of the 11 tributaries. This proportion was statistically different from that expected using the null hypothesis of 50% of the observed values lower than the predicted ones ( $\chi^2=4.45$ ,  $P<0.05$ ).

#### Discussion

The most striking results of this study can be summarized as follows. The sample of 39 African rivers studied shows a positive species-area relationship described by a power

function of the type  $N=cS^z$  ( $N$ =number of species;  $S$ =drainage surface area);  $z=0.32$ . In a sample of 26 rivers, ridge regression and forward stepwise selection led us to use a predictive model of  $\ln(\text{species richness})$  as a function of  $\ln(\text{surface area})$  and  $\ln(\text{discharge})$ . The residuals obtained by this model showed positive spatial autocorrelation. The model predicting the number of riverine species was applied to 11 tributaries. The predicted numbers of species were lower than the observed values; species richness of the tributaries (continental) thus appeared higher than that of rivers (insular).

We here discuss these results in the context of insular biogeography, and judge whether this provides a good model when applied to African rivers. Historical factors which might explain the geographical distribution of spatially correlated residuals are also discussed.

#### *Significance of the exponential value of the power relationship*

Preston (1962) proposed that the expected value of  $z$  in conditions of insularity should be 0.25. His demonstration rests on the following assumptions: (1) species abundance distribution on each island is described by the same relationship (in this case, canonical log-normal); (2) the total number of individuals (all species pooled) is proportional to island surface area. May (1975), using the same conditions but considering several log-normal richness distributions, not only the canonical one, obtained  $z$  values between 0.15 and 0.39: this range includes the known  $z$  values of different insular systems, and our value of 0.32 falls within this range.

In view of many studies (Connor and McCoy 1979; Martin 1981; Coleman et al. 1982; Abbott 1983), prudence is called for in attributing a biological significance to the  $z$  value. In studies such as those carried out on rivers where the surface area is not measured directly but estimated by that of the drainage area, the proportionality between the number of individuals and the surface area estimator must be checked to justify applying the power relationship. While the proportionality between surface area and number of individuals seems logical, that between real river surface and drainage surface area has to be demonstrated. Daget and Iltis (1965) used Preston's relationship to determine the theoretical species richness of Ivory Coast rivers. Comparing the calculated values with the observed ones and noting that the observed values were lower, they concluded that the rivers had not been sufficiently investigated. However, for the reasons mentioned above, we thought it preferable to determine the empirical relationship *a priori*.

While the value of  $z$  itself certainly has no significant biological implication, at least in our study, it could be a useful criterion when comparing the results of studies of different habitats. The possibilities of comparison are unfortunately limited by the small number of studies on species-area relationships in a fluvial environment. The  $z$  value computed in our study is quite different from that of Welcomme (0.43; 1979) who used a wider geographic sample of African rivers. However, Welcomme mixed tributaries and rivers and did not report the values used to calculate  $z$ , so it is difficult to discuss his results further, and especially to estimate the part played by insularity. The  $z$  values obtained for European rivers are much lower (0.236, Welcomme 1979; 0.24, Daget and Economidis 1975;

0.19, Daget 1968) than those found in Africa or South America (0.552, Welcomme 1979). These results indicate that species richness tends to increase more rapidly with surface area in tropics than in temperate regions. This is usually interpreted as due to greater inter-habitat diversity in the tropics (Welcomme 1979). This conclusion, however, is only plausible if the relationship linking drainage surface area and the surface area of the water available to the fish, is the same in the two regions.

#### *Hypothesis of habitat diversity*

The hypothesis of habitat diversity proposed by Williams (1964) is the following: the larger the surface is, the more probable that a certain type of habitat will be represented there; consequently, the probability that the associated fauna will be present also increases. If this hypothesis is valid, in its extreme form, surface area would only play a role via habitat diversity. In other words, once the effect of habitat diversity is taken into account, surface area would have no effect. In its loosest form, the effect of habitat diversity is added to that of surface area. The main difficulty in validating this hypothesis is to determine in what way the group studied 'perceives' habitat diversity and to define a pertinent index. The relationships between habitat diversity and fish communities have been analysed by Gorman and Karr (1978) in temperate and tropical zones. According to their study, the diversity of riverine fish communities is determined locally by factors such as the diversity of the current, depth and substrata, but it is difficult to quantify these factors on a wider scale.

In West Africa, Daget and Iltis (1965) distinguish "sudanian" fishes living in the waters of the extensive savanna-covered peneplain, and "guinean" species inhabiting forested streams. Logically, then, a drainage area with both types of vegetation should have a greater number of species than another with only one vegetation type. Moreover, vegetation integrates many climatic factors and so could act as an index for those which may affect river hydrology. Also, water chemical composition as well as river productivity are affected by the supply of exogenous organic elements, which differs with the type of vegetation crossed. Therefore, the diversity of terrestrial vegetation could be a measure of habitat diversity with significance for fish.

When vegetation diversity was integrated into our regression models, it was negatively related to species richness when the coefficients were calculated by the least-squares method. However, when they were estimated using ridge regression, the effect was positive from  $k=0.2$ . Coefficient values obtained by this method were more reliable and habitat diversity would thus follow expectation. However, the coefficient value was not high and this effect might be non-significant.

The data did not confirm the effect of habitat diversity; additional analyses are needed as well as measures of diversity which are more suitable than those we used.

#### *Habitat productivity and the hypothesis of Livingstone et al. (1982)*

Except for the hypothesis of habitat diversity, the hypotheses discussed above suppose an island system homogeneous enough that surface area is the main factor affecting the number of species; in other words, density is considered

to be constant. To improve already-existent hypotheses, attempts have been made to determine factors which might cause density to vary from one island to another. Wright (1983) proposed considering the total primary productivity of an island instead of its surface area because he noted better fit of the data with this single variable.

A similar approach was proposed by Livingstone et al. (1982) for African rivers. Those authors reported that in a sample of 26 rivers distributed throughout Africa, log(mean annual discharge) was the only significant predictor of log(species number), even if log(surface area) was integrated into the analysis. They emphasized that

discharge = constant  $\times$  precipitation  $\times$  surface area

and

total primary productivity = constant  $\times$  precipitation  $\times$  surface area

The results is that total primary productivity and discharge are very closely related. The correlation of discharge with species richness would then probably result from a correlation between productivity and species richness. It should be noted that by "primary productivity" of the drainage area, Livingstone et al. meant terrestrial and not aquatic productivity. As a great deal of the organic matter in tropical rivers is allochthonous, this is reasonable.

Using the Napierian logarithm of each variable (for additivity), a path analysis (Sokal and Rohlf 1981) led us to affirm that, if this hypothesis was correct, the partial correlation coefficient to  $\ln(\text{surface area})$  with  $\ln(\text{species number})$ , with  $\ln(\text{discharge})$  remaining constant, should be zero since surface area does not act directly but via discharge. The data of Livingstone et al. agree with this hypothesis. However, many species richness values cited by those authors are clearly lower than those now known, and the results must be examined with caution. Our study, including 12 of the 26 rivers analysed by Livingstone et al., but with new values, does not confirm their hypothesis on the scale of West Africa. Both  $\ln(\text{discharge})$  and  $\ln(\text{surface area})$  had a statistically significant effect on  $\ln(\text{species number})$ .

The relationship between surface area, discharge, and species richness is thus more complex than the results of Livingstone et al. indicate. It is very probable that discharge does not act solely through primary productivity but is also linked to the volume of water available to the fishes.

However, we agree with Livingstone et al. that drainage surface area is not a satisfactory predictor of the number of riverine species, especially on a large geographical scale. Rather than replace it by discharge, we propose to use a combination of these variables. We consider that this is the minimum basis on which to make valid comparisons between different geographical areas.

#### *Area-dependent extinction rates*

The hypothesis of area-dependent extinction rates is one element of the so-called dynamic equilibrium theory (McArthur and Wilson 1963, 1967) but it can be considered separately. Since the probability of the extinction of a population increases as its size declines, small islands should have higher extinction rates because the populations cannot reach great sizes. Therefore, small islands should have fewer species than large islands. It is impossible to check this

hypothesis on the basis of a positive species-area relationship (Gilbert 1980). However, area-dependent extinction rates are one reason why islands should have fewer species than comparable continental areas. On continents, local extinction can be compensated quickly by the immigration of new individuals from neighbouring populations, while on islands, where the immigration rate is low, this process is slower.

Following this logic, there should be a larger number of species in a tributary (continental) than in a river (insular) of comparable size and characteristics; even more so since the natural immigration of new fish species in a river is almost zero, on the ecological time-scale, because there are no aquatic connections between most rivers. The data analysed here agree with this hypothesis, since our river model underestimated the species richness in a sample of 11 tributaries.

It should be noted that this result was obtained on a small sample of tributaries and that other predictive models might have led to different conclusions.

If the hypothesis of surface-area-dependent extinction rate is correct, species able to reach high densities would be greatly over-represented in small rivers, since they would be less subject to extinction. A study is in progress to test this hypothesis.

#### *Historical factors*

It is tempting to try to explain some differences from the model by the particular characteristics of the rivers concerned or by the intervention of historical factors not accounted for by the analysis.

It is interesting to note that the Nile river is one in which the number of species is clearly lower than expected. Consequently, the paradox raised by Greenwood (1976) still stands; he noted that, for a comparable surface area, the Nile always had considerably fewer species than the Zaïre, even if this difference was lessened by taking into account the differences in discharge between the two. An historical explanation, in this particular case, is rather convincing. During the Quaternary, dry and humid periods alternated in Africa, and the discharge of the Nile was highly affected by these climatic vicissitudes (Beadle 1981; Hamilton 1982). A reduction in the discharge of the Nile during the last dry period would have caused a considerable decline in the number of fish species in that river. Fish being poor immigrants, this river would only have recovered part of its original richness when the climate became more humid. What is true of the Nile is also true of Sahelian rivers and the Senegal and Niger have fewer species than expected, while the reverse is true of Central African rivers.

In general, the present species richness of rivers would be determined by the hydrological characteristics of the last dry period rather than by present conditions. If this mechanism is really an important factor in explaining residuals, then residuals should show a non-random spatial distribution because the intensity of climatic perturbation certainly has not been random. This is seen in the model integrating surface area and discharge.

A study of the distribution of birds, butterflies, and mammals indicates three probable refuge zones for humid forests during maximum dry periods: (1) Liberia and Sierra Leone, (2) Cameroon and Gaboon, (3) east of Zaïre (Endler 1982; Mayr and O'Hara 1986). It is supposed that in these

regions river discharges were not so greatly reduced and that, today, the rivers should harbour a greater number of fish species than those with similar characteristics but located outside the forest refuge zones. To what extent does the pattern of residual distribution conform to the location of the refuge areas generally admitted to be in West and Central Africa? The Cameroon-Gaboon-Zaire region is, on the whole, a zone where the species richness of the rivers exceeds that of model predictions, this agrees with the supposed location of refuge zones 2 and 3. The Gambia and the Konkoure, which are the only rivers in our study close to refuge zone 1, also have more species than expected. To discuss this further, we would need data on river discharges in Liberia and Sierra Leone. The third area of high species richness shown by our analysis, between eastern Ivory Coast and Benin, cannot be explained by the data now available on the climatic history of West Africa, and it is not evident what factor could act in the east of the Ivory Coast and not in the west. This pattern is probably artificial and results from a poor knowledge of the fauna of some small rivers to the west. Our data can thus only confirm the presence of refuge areas in Central Africa. There is also the possibility that non-random spatial distribution of residuals is due to the fact that a varying geographical factor has not been integrated into the model. Moreover, the fact that no spatial autocorrelation is perceptible when only surface area is considered should invite caution. The validity of the Pleistocene forest-refuge hypothesis, when applied to fish communities of Africa, should be evaluated considering the phylogeny of related groups, distribution, and endemism, and not only the species richness of the rivers.

*Acknowledgements.* I should like to thank Christian Lévêque for his comments on this manuscript. Christian Lévêque, Didier Paugy, Guy Teugels made available unpublished lists of the ichthyofauna of some West African rivers. This paper is a contribution to the PEDALO program, supported by ORSTOM and PIREN.

#### Appendix 1 Ridge regression

A near colinearity between predictor variables may be expressed by an artificially long regression coefficient vector when the vector is calculated by the classic least-squares method, i.e. the coefficient variance is too high. The principle of ridge regression is to fix a maximum value of the regression coefficient vector length, that is, to transform the normal equation permitting an estimate of coefficient value by the least-squares method:

$$b = (X'X)^{-1} X'Y$$

into:

$$b = (X'X + kI)^{-1} X'Y$$

where  $I$  is the identity matrix (Hoerl and Kennard 1970a). The variables should be transformed so that  $X'X$  is in the form of a correlation matrix. It is evident that if  $k=0$ , we have arrived at the least-squares estimation.

As  $k$  increases, the variance of the coefficients of regression equation decreases, but their estimation is progressively more biased and the coefficients are all zero for infinite  $k$ . Therefore,  $k$  must be chosen so that the precision gained due to the lowest variance will not be negated by a great

bias. The coefficient of determination also decreases as  $k$  increases but for small  $k$  values it is hoped to obtain a more predictive model (Marquardt and Snee 1975; Snee 1977).

The main difficulty with this method is to select the optimal  $k$  value; there are several numerical ways of doing this but the simplest is to use a curve, called the ridge trace (Hoerl and Kennard 1970b; Marquardt and Snee 1975). If the evolution of the standardized coefficients is analysed as a function of  $k$ , it is seen that, after considerable change for the low  $k$  values, the coefficients stabilize rapidly and that  $k$  must be selected in that range. Finally, it is preferable not to choose  $k > 1$ , and the maximum variance inflation factor should be near 1.

#### Appendix 2 Variance inflation factor

The diagonal elements of the following matrix are called variance inflation factors:

$$(X'X)^{-1}$$

or, in ridge regression, diagonal elements of the following matrix (Marquardt 1970):

$$(X'X + kI)^{-1} (X'X) (X'X + kI)^{-1}$$

In both cases,  $X'X$  must be a correlation matrix.

The variance inflation factor gives an estimate of the increase of regression coefficient variance of one of the predictor variables due to near colinearity with the other variables. If all the variables are orthogonal among themselves, all the variance inflation factors will be 1. If the maximum variance inflation factor is higher than 10, the estimation of coefficients by the least-squares method can be expected to be inaccurate.

#### References

- Abbott I (1983) The meaning of  $z$  in species-area regressions and the study of species turnover in island biogeography. *Oikos* 41:385-390
- Beadle LC (1981) The inland waters of tropical Africa, 2nd edn. Longman, London
- Brown JH (1971) Mammals on mountain tops: non equilibrium insular biogeography. *Am Nat* 105:467-478
- Coleman BD, Mares MA, Willig MR, Hsieh YH (1982) Randomness, area, and species richness. *Ecology* 63:1121-1133
- Connor EF, Mc Coy ED (1979) The statistics and biology of the species-area relationship. *Am Nat* 113:791-833
- Daget J (1968) Diversité des faunes de poissons dans les cours d'eau du Portugal. *Arq Mus Bocage* 2:21-26
- Daget J, Economidis PS (1975) Richesse spécifique de l'ichtyofaune de Macédoine orientale et de Thrace occidentale (Grèce). *Bull Mus Natl Hist Nat Paris Ecol Gén* 346:81-84
- Daget J, Iltis A (1965) Les poissons de Côte d'Ivoire (eaux douces et eaux saumâtres). *Mém Inst Fr Afr Noire* 74:1-385
- Draper NR, Smith H (1966) Applied regression analysis. Wiley, New York
- Eadie J McA, Hurly TA, Montgomerie RD, Teather KL (1986) Lakes and rivers as islands: species-area relationships in the fish faunas of Ontario. *Environ Biol Fishes* 15:81-89
- Endler JA (1982) Pleistocene forest refuges: fact or fancy? In: Prance GT (ed) Biological diversification in the tropics. Columbia University Press, New York, pp 641-657
- Gilbert FS (1980) The equilibrium theory of island biogeography: fact or fiction? *J Biogeogr* 7:209-235



- Gorman OT, Karr JR (1978) Habitat structure and stream fish communities. *Ecology* 59: 507-515
- Greenwood PH (1976) Fish fauna of the Nile. In: Rzoska J (ed) *The Nile, biology of an ancient river*. Dr. W. Junk, The Hague, pp 127-141
- Hamilton AC (1982) *Environmental history of East Africa, a study of the Quaternary*. Academic Press, London
- Hocking RR (1976) The analysis and selection of variables in linear regression. *Biometrics* 32: 1-49
- Hoerl AE, Kennard RW (1970a) Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* 12: 55-67
- Hoerl AE, Kennard RW (1970b) Ridge regression: applications to non-orthogonal problems. *Technometrics* 12: 69-82
- Lévêque C, Fairhust CP, Abban K, Paugy D, Curtis MS, Traore K (1988) Onchocerciasis control programme in West Africa: ten years monitoring of fish populations. *Chemosphere* 17: 421-440
- Livingstone DA, Rowland M, Bailey PE (1982) On the size of African riverine fish faunas. *Am Zool* 22: 361-369
- Lowe Mc Connell RH (1975) *Fish communities in tropical freshwaters*. Longman, London
- MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution* 17: 373-387
- MacArthur RH, Wilson EO (1967) *The theory of island biogeography*. Princeton University Press, New Jersey
- Marquardt DW (1970) Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* 12: 591-612
- Marquardt DW, Snee RD (1975) Ridge regression in practice. *American Statist* 29: 3-20
- Martin TE (1980) Diversity and abundance of spring migratory birds using habitat islands on the great plains. *Condor* 82: 430-439
- Martin TE (1981) Species-area slopes and coefficients: a caution on their interpretation. *Am Nat* 118: 823-837
- May RM (1975) Patterns of species abundance and diversity. In: Cody ML, Diamond JM (eds) *Ecology and evolution of communities*. Harvard University Press, Cambridge, pp 81-120
- Mayr E, O'Hara RJ (1986) The biogeographic evidences supporting the pleistocene forest refuge hypothesis. *Evolution* 40: 55-67
- Preston FW (1962) The canonical distribution of commonness and rarity. *Ecology* 43: 185-215, 410-437
- Sepkoski JJ, Rex MA (1974) Distribution of freshwater mussels: coastal rivers as biogeographic islands. *Syst Zool* 23: 165-188
- Snee RD (1977) Validation of regression models: methods and examples. *Technometrics* 19: 415-428
- Sokal RR, Oden NL (1978a) Spatial autocorrelation in biology. I. Methodology. *Biol J Linn Soc* 10: 199-228
- Sokal RR, Oden NL (1978b) Spatial autocorrelation in biology. II. Some biological implications and four applications of evolutionary and ecological interest. *Biol J Linn Soc* 10: 229-249
- Sokal RR, Rohlf FJ (1981) *Biometry - the principles and practice of statistics in biological research*, 2nd edn. Freeman, San Francisco
- Vuilleumier F (1973) Insular biogeography in continental regions. II. Cave faunas from Tesin, Southern Switzerland. *Syst Zool* 22: 64-76
- Welcomme RL (1979) *Fisheries ecology of floodplain rivers*. Longman, London
- White F (1983) *The vegetation of Africa*. UNESCO
- Williams CB (1964) *Patterns in the balance of nature and related problems in quantitative ecology*. Academic Press, New York
- Wright DH (1983) Species energy theory: an extension of species-area theory. *Oikos* 41: 496-506

Received June 29, 1988