RRB 89-35

A PROBABILITY MODEL FOR CANADIAN LAKES ACID RAIN DATA

by

A.H. El-Shaarawi and A. Naderisamani

Rivers Research Branch
National Water Research Institute
Canada Centre for Inland Waters
Burlington, Ontario, L7R 4A6

May 1989
NWRI Contribution #89-120

ABSTRACT

The pH-alkalinity relationship developed by Small and Sutton is fitted to the Canadian lakes alkalinity data, and is it shown that the relationship adequately fits the data when the water colours are less than 30 Hazen units. Further, the model developed by Small and Sutton to predict the regional lake alkalinity distribution is modified to maintain the same distributional form throughout the changes in acid deposition rates. Moreover, a three parameter lognormal distribution is fitted to the Canadian lakes alkalinity data, and the corresponding distribution for pH is derived directly using the pH-alkalinity relationship. However, when the pH-alkalinity relationship does not hold, a distribution is required to fit the pH data. It is shown that a mixture of two normal distributions with four parameters fits the pH-data adequately.

Keywords: alkalinity, pH, pH-alkalinity relationship, lake acidification, probability distributions, statistical inference

RÉSUMÉ

Le rapport pH/alcalinité mis au point par Small et Sutton est incorporé aux données sur l'alcalinité des lacs canadiens, et il semble que le rapport correspond adéquatement aux données lorsque les couleurs de l'eau sont inférieures à 30 unités Hazen. De plus, le modèle mis au point par Small et Sutton dans le but de prévoir la distribution régionale de l'alcalinité des lacs est modifié de façon à maintenir la même forme de distribution en présence de tous les changements des niveaux de dépôt acide. De plus, une distribution lognormale à trois paramètres est incorporée aux données sur l'alcalinité des lacs canadiens, et la distribution correspondante pour le pH est dérivée directement à l'aide du rapport pH/alcalinité. Cependant, lorsque ce rapport ne tient pas, une distribution est alors nécessaire pour accommoder les données sur le pH. Le rapport indique qu'un mélange de deux distributions normales, avec quatre paramètres, s'adapte adéquatement aux données sur le pH.

Mots-clés : alcalinité, pH, rapport pH/alcalinité, acidification des lacs, distributions des probabilités, inférence statistique

MANAGEMENT PERSPECTIVE

Fitting stochastic models to regional data is important not only for characterizing the observed pattern of variability, but also for comparing patterns with those observed in other regions or at different time points in the same region. This paper is concerned with developing and testing the adequacy of regional acidification models using data consisting of alkalinity and pH measurements for lakes in New Brunswick, Newfoundland, Nova Scotia and Quebec. It is shown that a lognormal distribution adequately fitted the alkalinity data while for pH, a mixture of two normal distributions was better.

PERSPECTIVE-GESTION

L'ajustement des modèles stochastiques aux données régionales est important, non seulement pour caractériser les modèles observés de la variabilité, mais également pour comparer les modèles avec ceux observés dans d'autres régions, ou à des moments différents dans une même région. Le présent rapport traite de l'élaboration et de la vérification de la précision des modèles d'acidification régionale à l'aide de données consistant en des mesures de l'alcalinité et du pH dans les lacs au Nouveau-Brunswick, à Terre-Neuve, en Nouvelle-Écosse et au Québec. On peut voir que la distribution lognormale correspondait de façon adéquate aux données sur l'alcalinité tandis que dans le cas du pH, il était préférable d'utiliser des distributions à deux normales.

INTRODUCTION

Acidification models permit the prediction of changes in lake or stream chemistry that correspond to changes in acid deposition and allow the identification of important acidification factors, thus giving guidance to decision makers in the formulation of policy for combatting the deleterious effects of acid rain. Although the prediction of pH is almost always required for evaluating the biological impacts, most models predict the changes in alkalinity which then are used to predict the corresponding changes in pH. Recently, Small and Sutton (1986a) derived an explicit equation relating pH and alkalinity of natural waters and tested its applicability using both regional and temporal lakes acidification data. In a second paper, the authors (1986b) used a three-parameter lognormal to represent the regional distribution (cumulative distribution function abbreviated to CDF) of lake alkalinity and developed a model for predicting the impacts of lake acidification on fish survival. Since the lognormality of the regional distribution of alkalinity and the pH-alkalinity relationship are among the critical assumptions involved in generating the predictions of the model, it is then natural to examine their adequacy to populations of lakes other than those considered by Small and Sutton. Hence, the main objective of the paper is to evaluate the adequacy of those assumptions using alkalinity and pH data for Canadian lakes from New Brunswick, Newfoundland, Nova Scotia and Quebec.

Consider a population of N lakes and let pH_i and Alk_i be the pH and alkalinity values for the ith lake (i=1,2,...,N). Small and Sutton (1986a) expressed the relationship between pH and Alk by the equation:

$$pH = a + \frac{1}{\ln 10} \operatorname{arcsinh} \left(\frac{Alk - d}{c}\right), \tag{1}$$

where a, c and d are unknown constants. This equation was fitted to the Canadian lakes data and the results indicated its adequacy for lakes with water colours less than 30 Hazen units. By assuming that the alkalinity values from the N lakes represent a random sample from a lognormal population, equation (1) can be used to derive the regional distribution for pH.

The model developed by Small and Sutton (1986b) to predict the regional lake alkalinity distribution has the form: $Alk_c = Alk_o - c(1-F)$, where Alk_c and Alk_o are respectively, the current and zero-deposition lake alkalinity, c is a deposition concentration factor and F is a random variable representing weathering rates in different watersheds. The distribution of both Alk_c and Alk_o in the above model are assumed to be lognormal. The difficulty with such a model is that the distribution of Alk_c must be the convolution of Alk_o and $-c(1-F)$ and, in general, Alk_c and Alk_o both can not have the same distribution. To maintain the same distributional form for alkalinity, we have replaced F in the model by the mean μ_F of F. As a result, Alk_c and Alk_0 both have the same distribution G, not necessarily lognormal, and c(1-F) is a location parameter. This model predicts the effect of changes in acid deposition rates on the mean of G.

When relation (1) does not hold for a pH-alkalinity data set, the distribution of pH can not be derived directly from the distribution of alkalinity. In such cases, a distribution is required to fit the pH data. We found that a mixture of two normal distributions fitted the pH data from New Brunswick, Newfoundland, Nova Scotia, and Quebec adequately.

pH-Alkalinity Relationship

Equation (1) was fitted to the pH-alkalinity data from Canadian lakes in New Brunswick, Newfoundland, Nova Scotia, and Quebec using the IMSL subroutines for nonlinear regression methods. The estimates of a, c and d are given in Table 1, and the fitted and observed pH values are compared in Figures 1-4. The results show that the fit is poor for the coloured lakes In Nova Scotia. Howell (1987) indicates that when dissolved organic carbon concentrations increase above 5 mg/L and/or water colours exceed 30 Hazen units, organic acidity will become more important to the overall acidification process. In order to determine whether the lack of fit is due to the mixed population of coloured and noncoloured lakes, equation (1) was fitted to the pH-alkalinity data from lakes in Newfoundland and Nova Scotia with water colours less than 30 Hazen units. The results are given in Table 1 and Figures 5 and 6. These figures indicate good agreement between the observed and fitted values. Also, Table 1 shows that, as a result of removing the coloured lakes, the mean square error has been reduced from 0.057 to 0.043, and from 0.278 to 0.072 for Newfoundland and Nova Scotia lakes respectively.

A Probability Model for Lake Alkalinity Data

Consider a probability model of the form:

$$Alk_c = Alk_o + A,$$

where $A = -c\ (1 - \mu_F)$. This model is a modification of the model developed by Small and Sutton (1986b). Note that, by definition, A is a location parameter, and thus the same distributional form for lake alkalinity is maintained. The following relationships between the means and variances of Alk_c and Alk_o are clear:

$$\mu_{Alk_c} = \mu_{Alk_o} + A,$$
$$V_{Alk_c} = V_{Alk_o}. \qquad (2)$$

As the amount of acid deposition decreases, the value of A increases. As a result, the mean of the lake alkalinity distribtuion increases while its variance remains constant.

Distribution of Lake Alkalinity

The lognormal distribution has been frequently used by many authors for statistical analysis of environmental data. Recently, Small and Sutton (1986b) have considered a three parameter lognormal to represent the regional distribution of lake alkalinity. In order to fit a distribution to the Canadian lakes alkalinity data, we tried several distributions such as a four parameter generalized gamma, a mixture of two normal distributions with four parameters and a three parameter lognormal. The results indicated that a three parameter lognormal was in good agreement with the observations.

A three parameter lognormal probability density function (Pdf), g, is defined by:

$$g(x) = \frac{1}{(x-\theta)\sqrt{2\pi}\gamma} \exp\{-\frac{[\ln(x-\theta) - \beta]^2}{2\gamma^2}\}. \qquad (3)$$

The estimates of θ, β and γ are obtained (Stedinger, 1980) as follows:

$$\hat{\theta} = \frac{X_{(1)} X_{(n)} - X_{0.5}^2}{X_{(1)} + X_{(n)} - 2X_{0.5}},$$

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \ln(X-\hat{\theta}),$$

$$\hat{\gamma}^2 = \frac{1}{n} \sum_{i=1}^{n} [\ln(X-\hat{\theta}) - \hat{\beta}]^2.$$

Here $X_{(1)}$ is the minimum, $X_{(n)}$ the maximum, and $X_{0.5}$ the median of the n observations. Estimates of the parameters of g for the lake alkalinity data from New Brunswick, Newfoundland, Nova Scotia, and Quebec are shown in Table 2. The goodness of fit was examined using the Kolmogorov-Smirnov statistic defined by $D_n = \sup|F_n(x)-F(x)|$, where $F_n(x)$ and $F(x)$ are respectively, the sample and assumed distributions, and n is the number of observations. The results are summarized in Table 2. These results suggest that the three parameter lognormal g fits the alkalinity data from Quebec and Newfoundland adequatly. The fitted and sample distributions of Quebec alkalinity data are compared in Figure 7.

The mean of the current alkalinity distribution, μ_{Alk_c}, is obtained from the relation:

$$\mu_{Alk_c} = \hat{\theta} + \exp(\hat{\beta} + \hat{\gamma}^2/2).$$

Let μ_{Alk_o} be the mean of the zero-deposition alkalinity, and let $\mu_{Alk_{.5}}$ be the mean of the predicted alkalinity resulting from 50% reduction in acid deposition. Then it follows from (2) that

$$\mu_{Alk_o} = \mu_{Alk_c} - A,$$

$$\mu_{Alk_{.5}} = \mu_{Alk_o} + \frac{A}{2}.$$

Estimates of the parameter θ of g for the zero-deposition alkalinity as well as the predicted alkalinity resulting from 50% reduction in acid deposition are obtained as follows:

$$\hat{\theta}_o = \mu_{Alk_o} - \exp(\hat{\beta} + \hat{\gamma}^2/2),$$

$$\hat{\theta}_{.5} = \mu_{Alk_{.5}} - \exp(\hat{\beta} + \hat{\gamma}^2/2).$$

These estimates in addition to $\hat{\theta}$, $\hat{\beta}$ and $\hat{\gamma}^2$ completely determine the distributions of the current and predicted alkalinity.

Since we do not have the measurements to compute A, we shall assume A = 10, and show the predicted effect of 50% and 100% acid deposition reductions on the pdf of Newfoundland alkalinity data in Figure 8.

Distribution of pH

Assuming (1), the pH distribution can be derived directly from the corresponding alkalinity distribution:

$$F_{pH}(x) = \text{prob}(pH \leq x) = \text{prob}(Alk \leq d + c \sinh(\tfrac{x-a}{b})) = F_{Alk}(d + c \sinh(\tfrac{x-a}{b})).$$

The corresponding pdf is:

$$f_{pH}(x) = \frac{c}{b} \cosh(\tfrac{x-a}{b})\, f_{Alk}\,(d + c \sinh(\tfrac{x-a}{b})). \qquad (4)$$

Assuming that (3) represents the pdf of lake alkalinity, (4) becomes

$$f_{pH}(x) = \frac{c \cosh (\frac{x-a}{b})}{b(d-\theta + c \sinh(\frac{x-a}{b}))\ \sqrt{2\pi}\ \gamma}$$

$$\exp \{ - \frac{[\ln(d-\theta + c \sinh (\frac{x-a}{b})) - \beta]^2}{2\gamma^2}\}.$$

However, when equation (1) does not hold for a pH-alkalinity data set, a distribution is required to fit the pH values. We found that a mixture of two normal distributions fitted the Canadian lakes pH data adequately.

A mixture, f, of two normal distributions with probability density functions f_1 and f_2 is defined by:

$$f = pf_1 + q\ f_2$$

where p = 1-q is a parameter which lies between zero and one. f_1 and f_2 have different means u_{f_1} and μ_{f_2}, but the same variance ν, totaling the number of parameters of f to four. The method of moments was used to estimate the parameters of f for pH data from New Brunswick, Newfoundland, Nova Scotia, and Quebec. The estimates are the solutions of the following four equations:

$$P\mu_{f_1} + q\mu_{f_2} = M(1),$$
$$P\mu^2{}_{f_1} + q\mu^2{}_{f_2} + \nu = M(2),$$
$$P\mu^3{}_{f_1} + q\mu^3{}_{f_2} + 3\nu\ (P\mu_{f_1} + q\mu_{f_2}) = M(3),$$
$$P\mu^4{}_{f_1} + q\mu^4{}_{f_2} + 6\nu\ (P\mu^2{}_{f_1} + q\mu^2{}_{f_2}) + 3\nu^2 = M(4).$$

Here M(k) denotes the kth sample moment. The estimates, as well as the values of the Kolmogorov-Smirnov statistic are shown in Table 3. The results indicate that the mixture fits the pH data adequately. Comparisons of the fitted, sampled and derived distributions of the pH data from Nova Scotia and New Brunswick are respectively shown in Figures 9 and 10. These figures indicate that the mixture fits the pH data better than the derived distribution even in the case of the New Brunswick data where there is a good agreement between the pH-alkalinity relationship and the observations.

CONCLUSIONS

The pH-alkalinity relationship developed by Small and Sutton is fitted to the Canadian lakes alkalinity data. It is shown that the relationship adequately fits the data when the water colours are less than 30 Hazen units. Further, the model developed by Small and Sutton to predict the regional lake alkalinity distribution is modified in a way that it preserves the shape of the distribution throughout the changes in acid deposition rates. It is shown that the mean of the regional lake alkalinity distribution increases as the amount of acid deposition decreases. Moreover, by assuming the pH-alkalinity relationship, and a distribution for the regional lake alkalinity, the corresponding distribution for pH can be derived directly. However, when the pH-alkalinity relationship does not hold, a distribution is required to fit the pH values. It is shown that a mixture of two normal distributions with four parameters adequately fits the pH data from New Brunswick, Newfoundland, Nova Scotia and Quebec.

ACKNOWLEDGEMENTS

The authors would like to thank G.D. Howell who provided the data, D.S. Jeffries for his helpful discussion, F. Norouzian for his help in preparing the graphs, and to C. Minnie for her help in typing the manuscript.

REFERENCES

Howell, G.D. and Brooksbank, P. (1987). An assessment of LRTAP acidification of surface waters in Atlantic Canada. Environment Canada, IW/L-AR-WQB-87-121.

IMSL (1987). MATH/LIBRARY and STAT/LIBRARY, IMSL, Inc., Houston, Texas.

Small, M.J. and Sutton, M.C. (1986a). A regional pH-alkalinity relationship. Water Resour. Res., 20:335-343.

Small, M.J. and Sutton, M.C. (1986b). A direct distribution model for regional aquatic acidification. Water Resour. Res., 22:1749-1758.

Stedinger, J.R. (1980). Fitting lognormal distributions to hydrologic data. Water Resour. Res., 16:481-490.

Table 1. Estimates of the unknown constants a, d and c of the pH-alkalinity relationship.

pH-alkalinity Data	Number of Observations	$\hat{a}$	$\hat{d}$	$\hat{c}$	Degrees of Freedom for Error	Mean Square Error
New Brunswick	160	5.378	0.346	0.515	157	0.064
Newfoundland	334	5.329	0.081	0.521	331	0.057
Newfoundland (water colours <30 Hazen units)	168	5.329	-0.056	0.459	165	0.043
Nova Scotia	326	5.293	0.009	1.637	323	0.278
Nova Scotia (water colours <30 Hazen units)	97	5.295	0.019	0.509	94	0.072
Quebec	177	5.014	0.039	0.219	174	0.313

Table 2. Estimates of the parameters of g and the Kolmogorov-Smirnov statistic.

Alkalinity data	$\hat{\beta}$	$\hat{\gamma}^2$	$\hat{\theta}$	The Kolmogorov-Smirnov Statistic
New Brunswick	1.735	0.913	-2.642	0.157
Newfoundland	1.306	0.757	-0.891	0.092
Nova Scotia	3.772	0.407	-35.872	0.281
Quebec	0.090	1.482	-0.042	0.049

Table 3. Estimates of the parameters of f and the Kolmogorov-Smirnov statistic.

pH Data	$\hat{\mu}_{f_1}$	$\hat{\mu}_{f_2}$	$\hat{\nu}$	$\hat{p}$	The Kolmogorov-Smirnov Statistic
New Brunswick	5.585	6.924	0.266	0.601	0.033
Newfoundland	6.301	4.163	0.295	0.998	0.066
Nova Scotia	6.514	4.998	0.204	0.490	0.065
Quebec	0.243	6.027	0.404	0.006	0.052

FIGURES

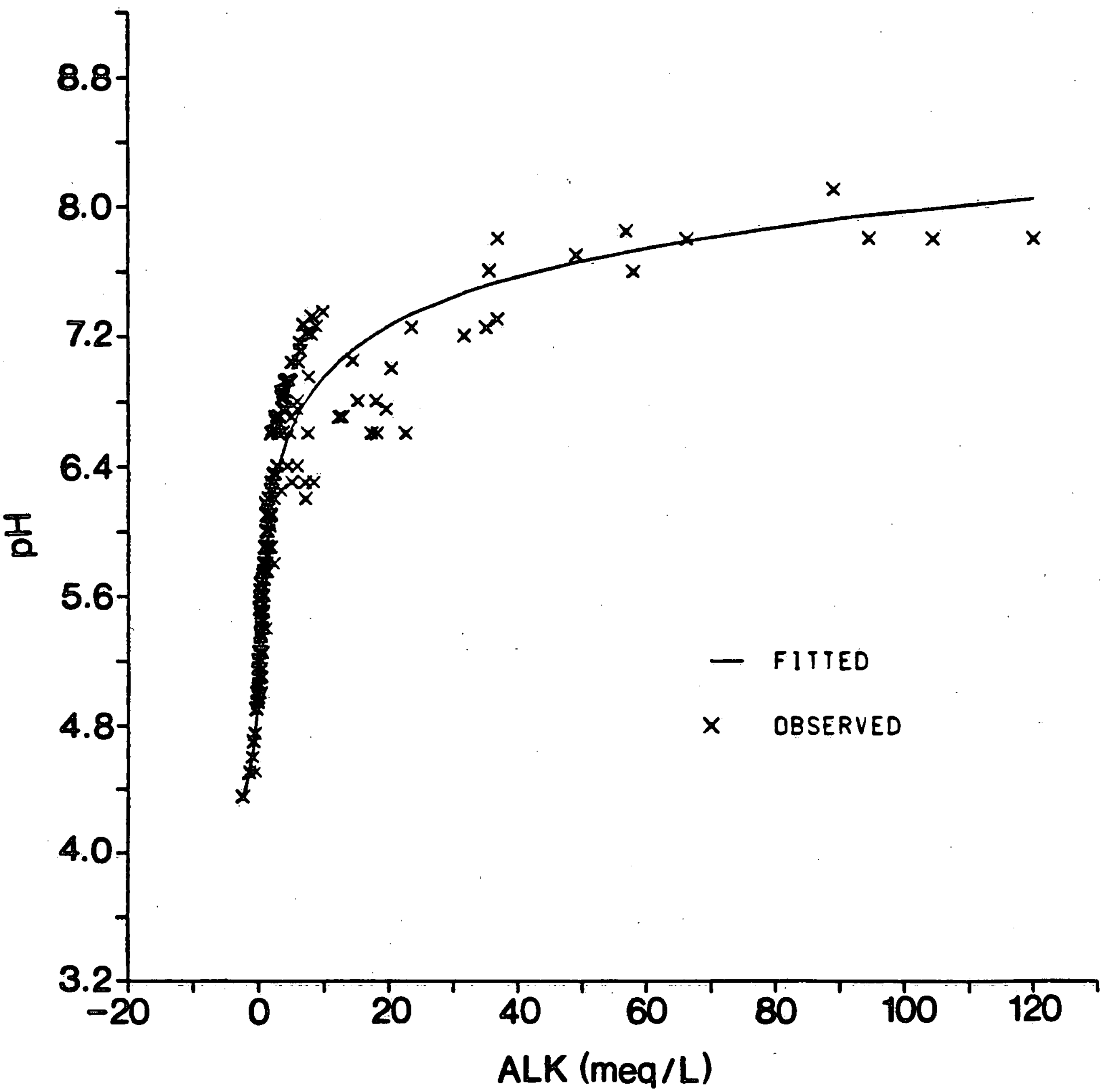

FIGURE 1. OBSERVED AND FITTED PH FOR NEW BRUNSWICK DATA

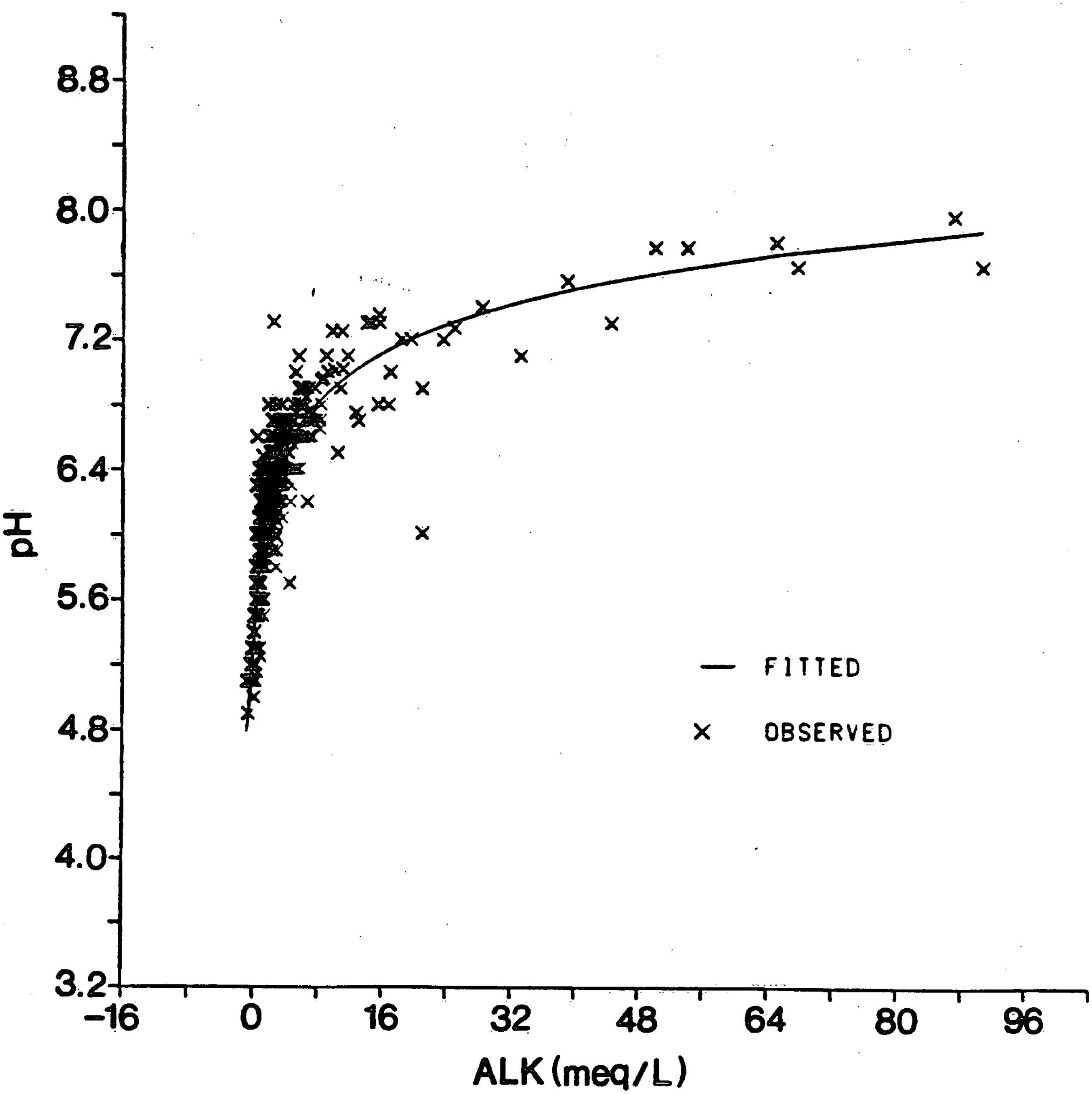

FIGURE 2. OBSERVED AND FITTED PH FOR NEWFOUNDLAND DATA

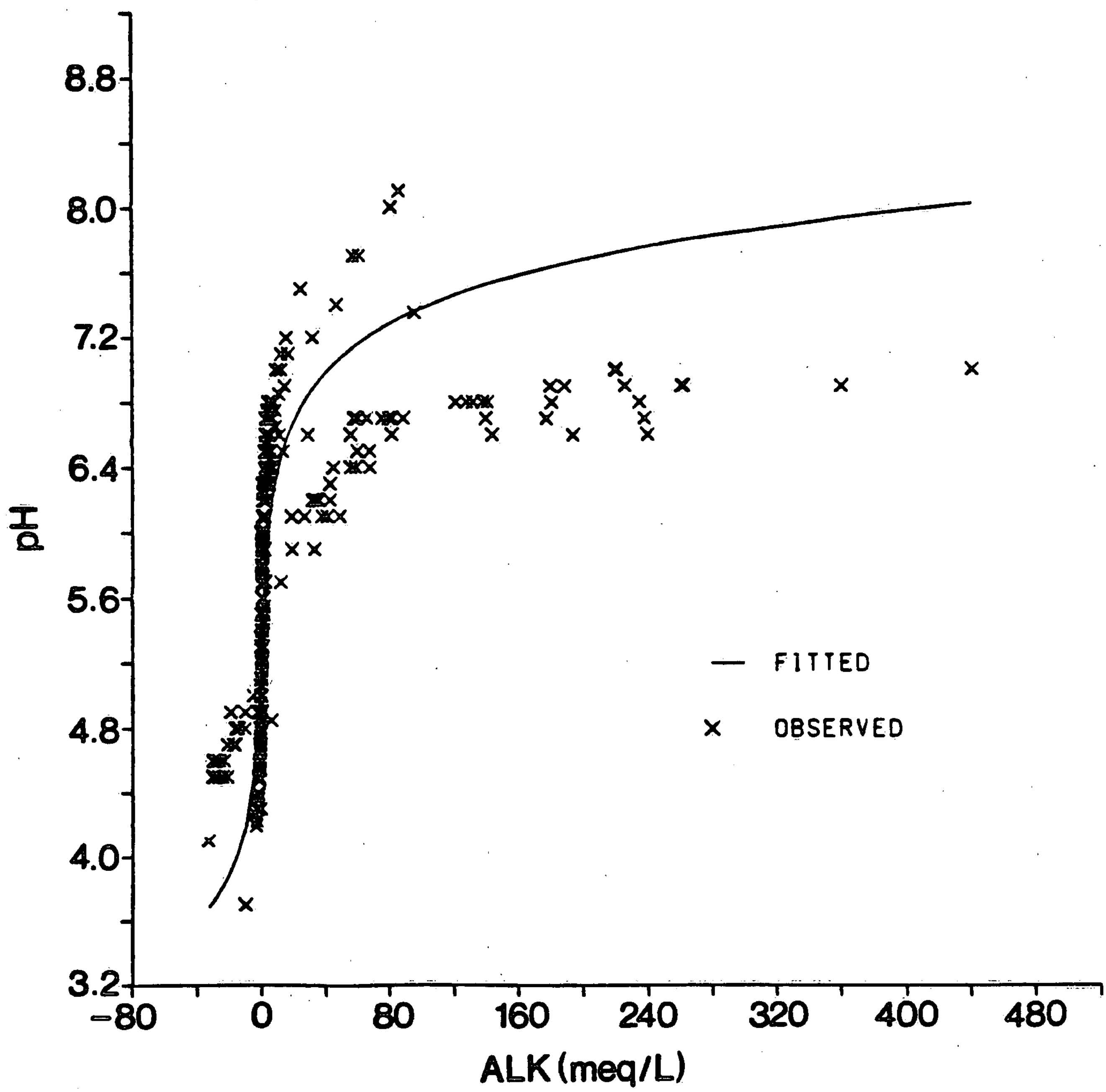

FIGURE 3. OBSERVED AND FITTED PH FOR NOVA SCOTIA DATA

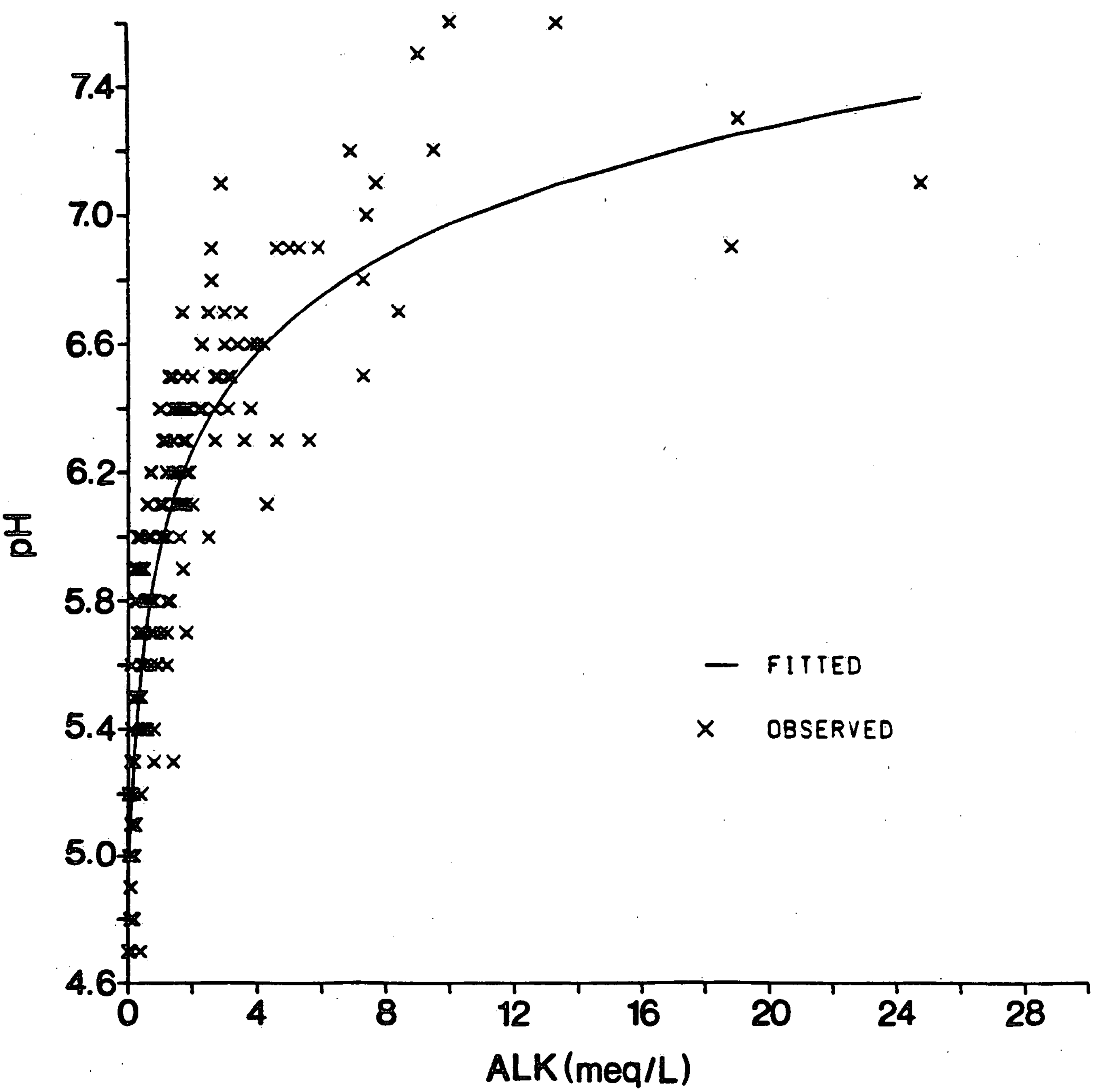

FIGURE 4. OBSERVED AND FITTED PH FOR QUEBEC DATA

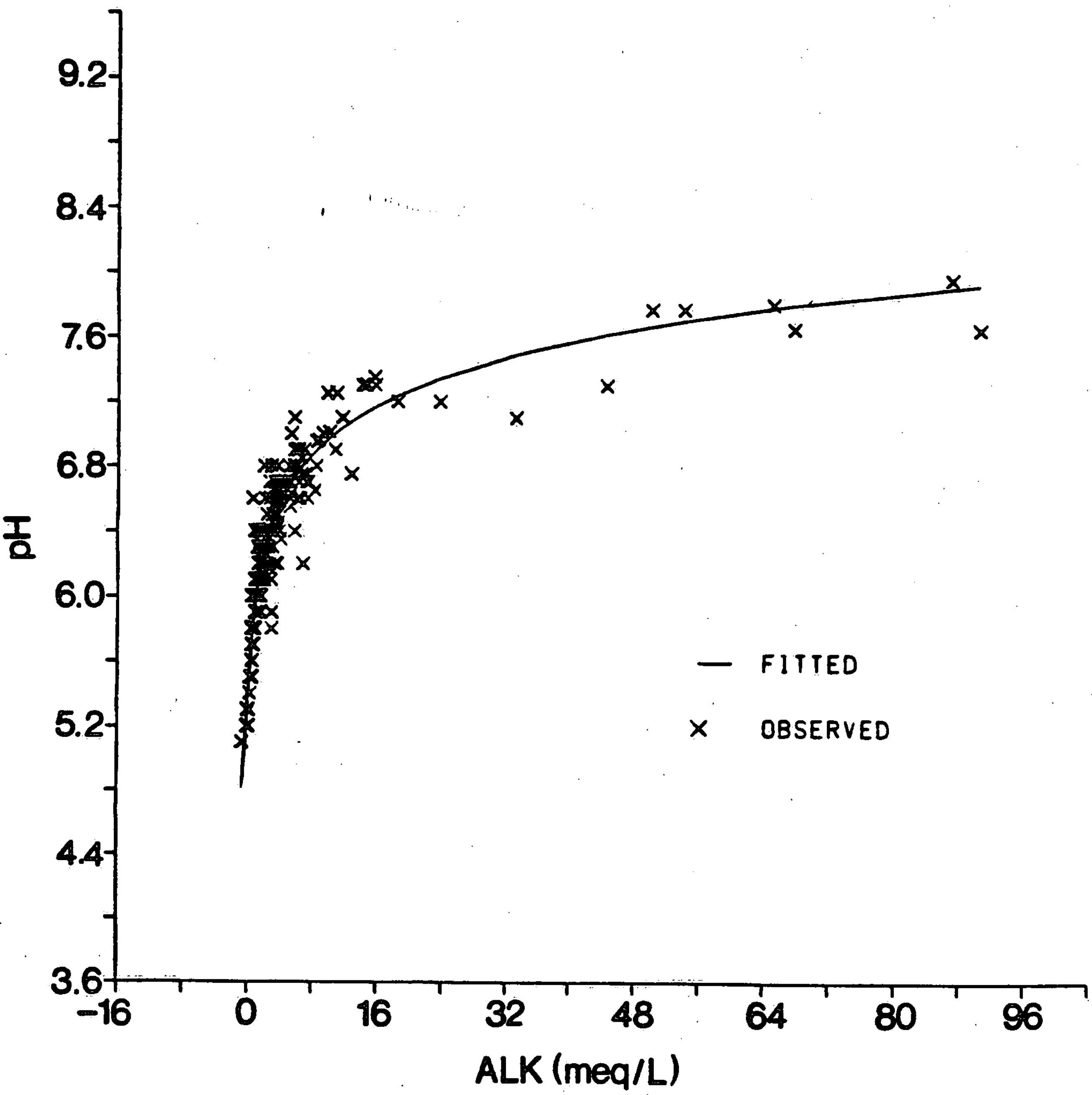

FIGURE 5. OBSERVED AND FITTED PH FOR NEWFOUNDLAND DATA WITH WATER COLOURS LESS THAN 30. HAZEN UNITS

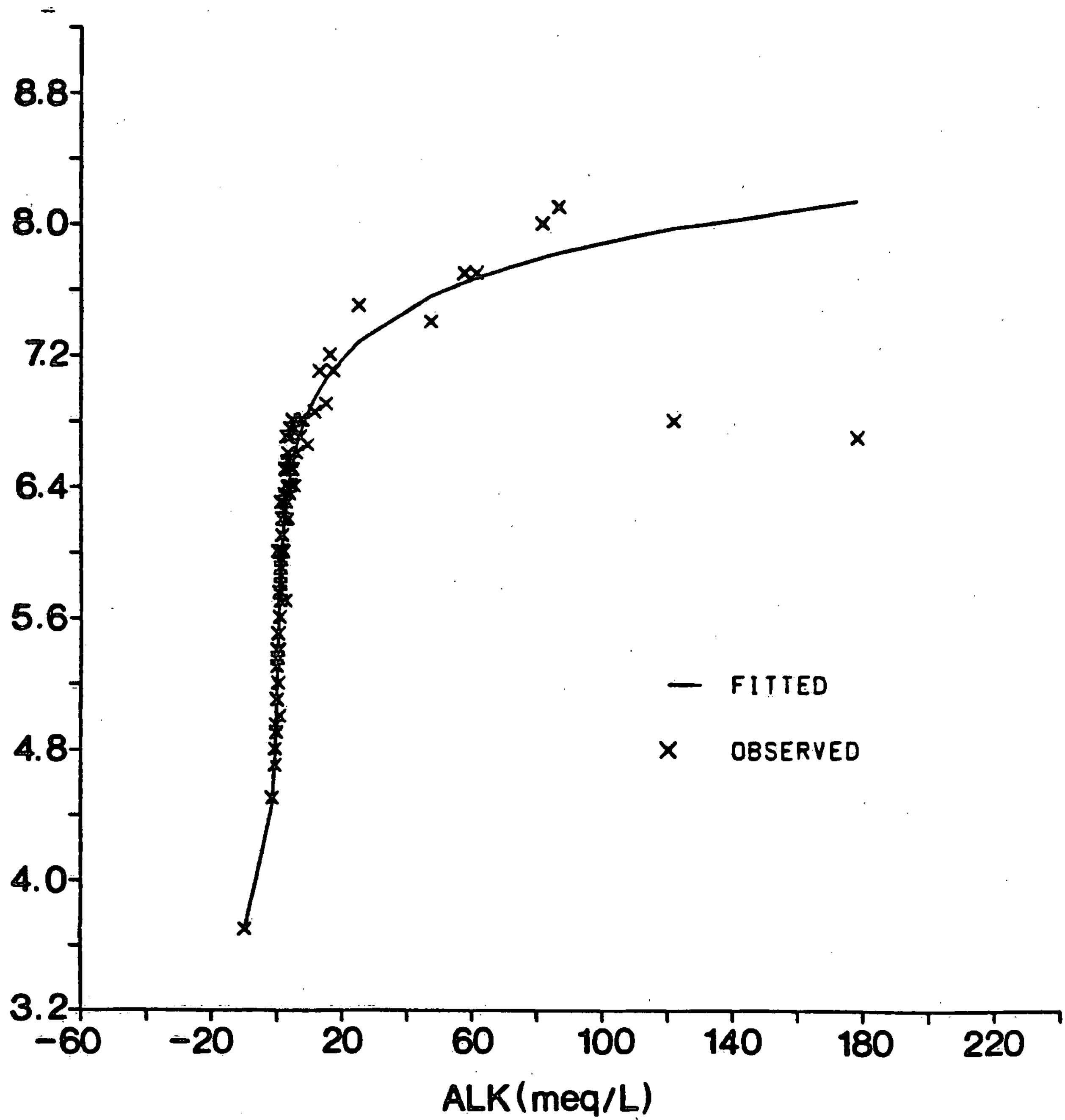

FIGURE 6. OBSERVED AND FITTED PH FOR NOVA SCOTIA DATA WITH WATER COLOURS LESS THAN 30. HAZEN UNITS

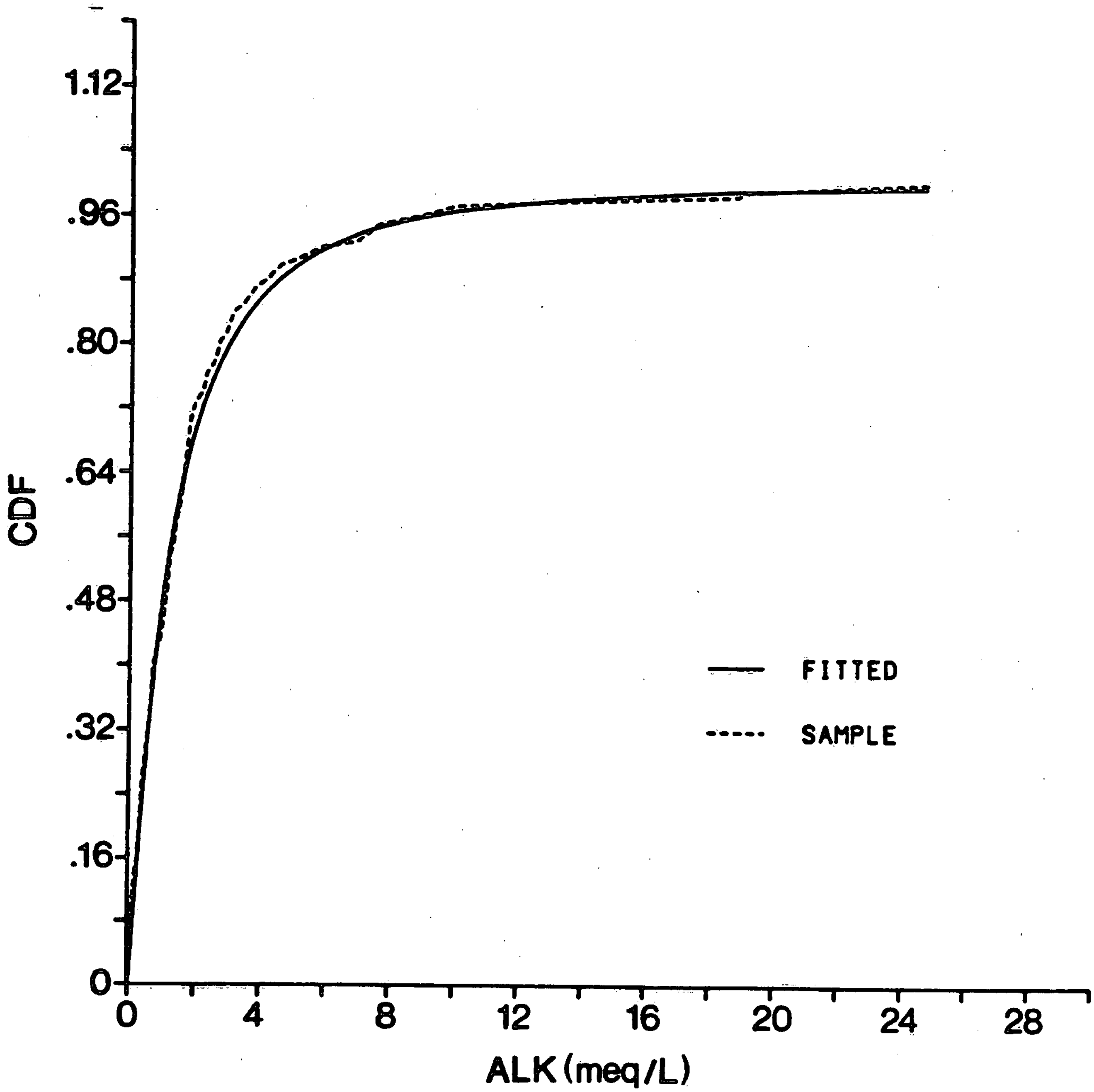

FIGURE 7. FITTED AND SAMPLE DISTRIBUTION FUNCTIONS OF QUEBEC ALKALINITY DATA

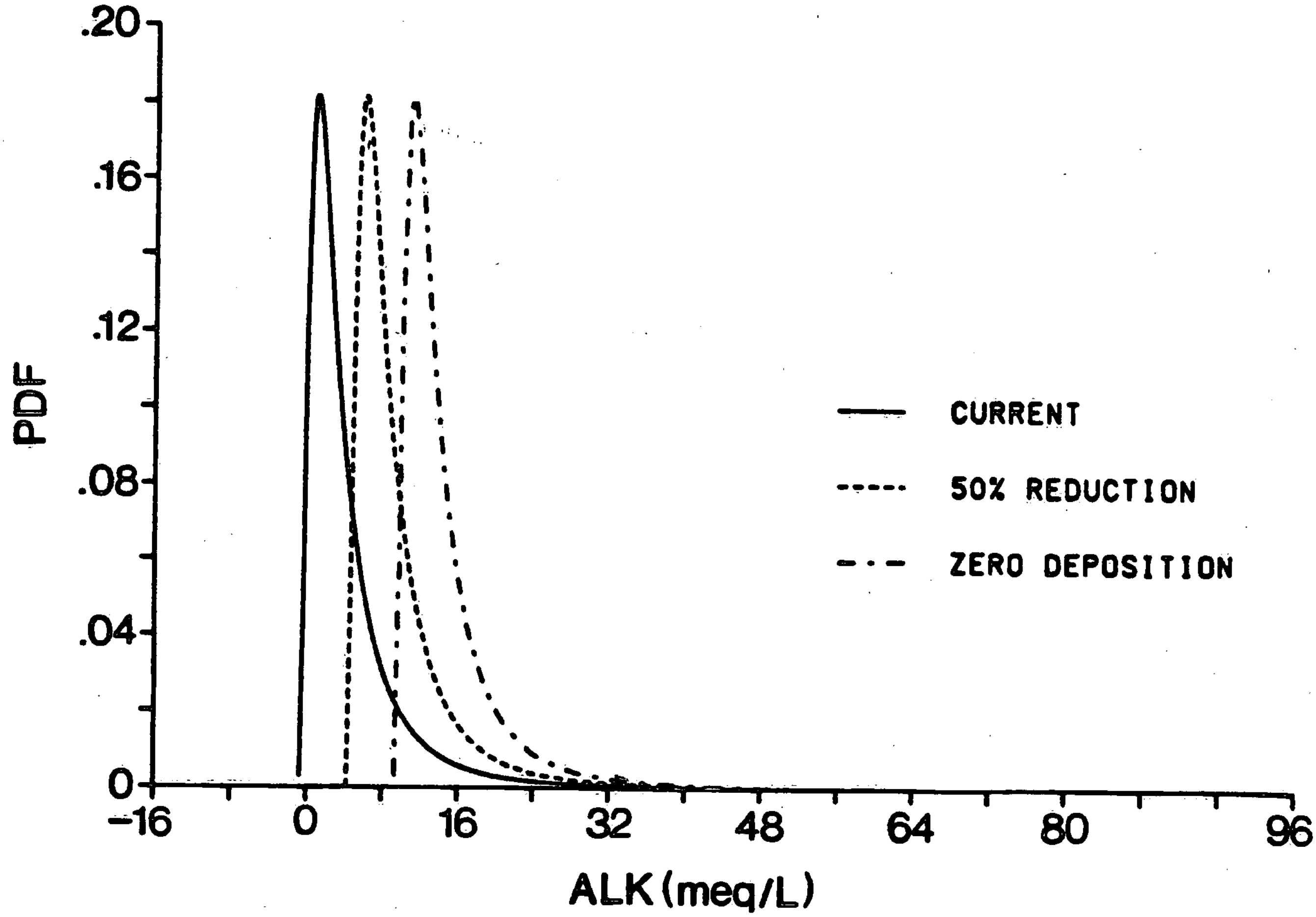

FIGURE 8. PREDICTION OF 50% AND 100% ACID DEPOSION REDUCTIONS ON THE PDF OF NEWFOUNDLAND ALKALINITY DATA

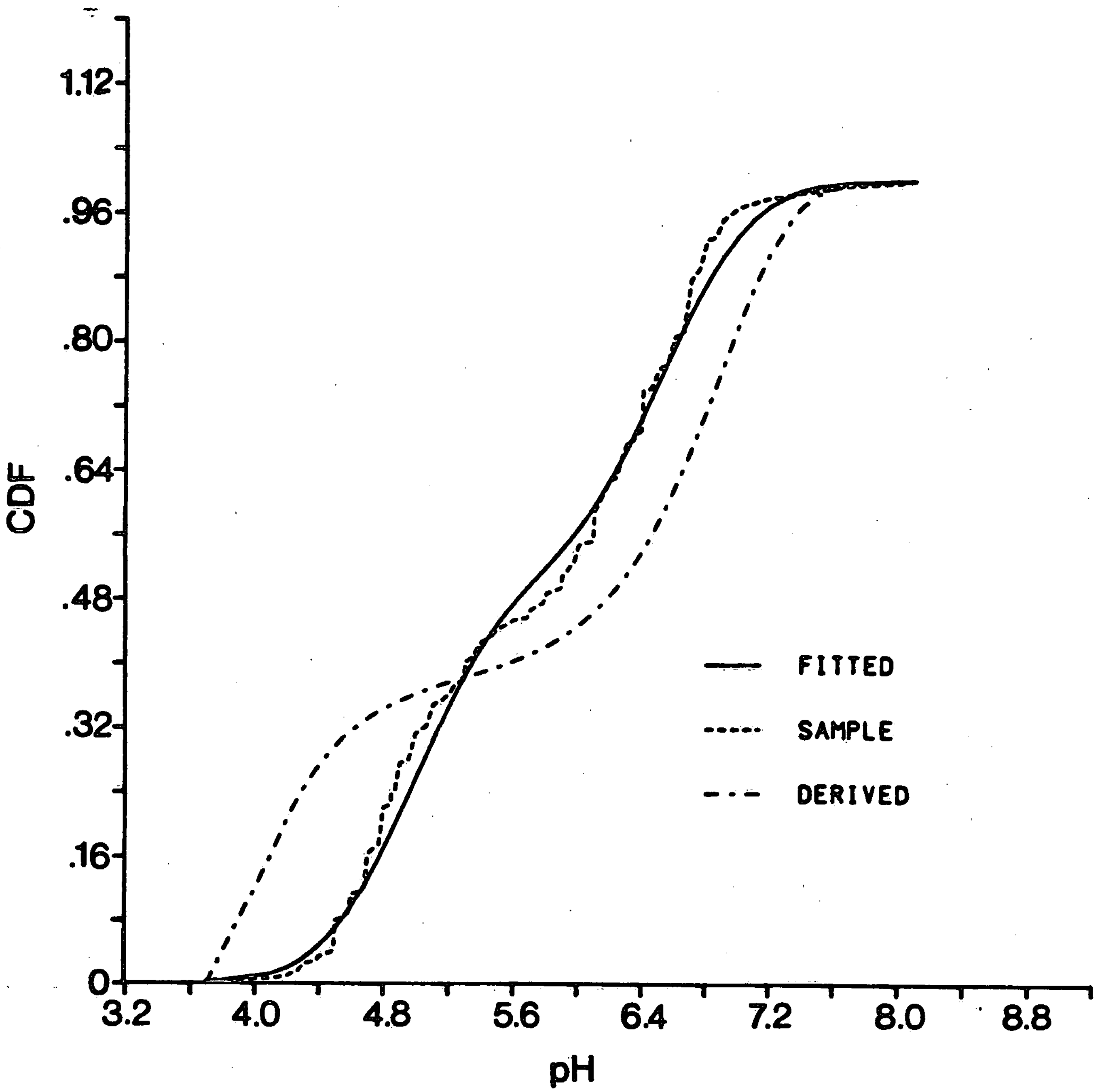

FIGURE 9. COMPARISONS OF THE FITTED, SAMPLE, AND DERIVED DISTRIBUTION FUNCTIONS OF NOVA SCOTIA DATA

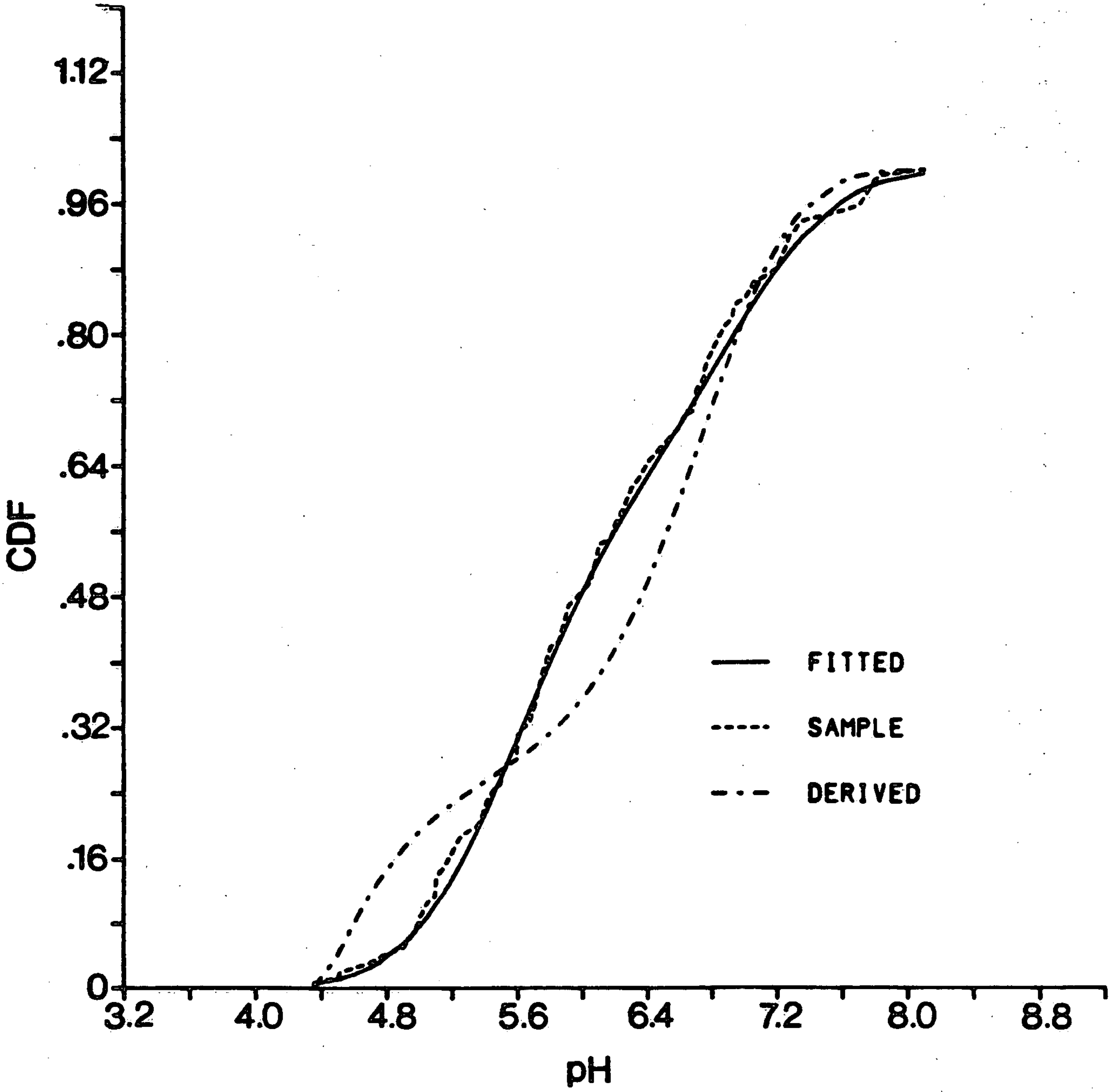

FIGURE 10. COMPARISONS OF THE FITTED, SAMPLE, AND DERIVED DISTRIBUTION FUNCTIONS OF NEW BRUNSWICK DATA