BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-86/26

SPATIAL MODELING OF REGIONAL VARIABLES*

by

Noel Cressie and Ngai H. Chan
Professor Assistant Professor
Department of Statistics Department of Mathematics
Iowa State University Indiana University
Ames, IA 50011 Bloomington, IN 47405

*This paper was written when both authors were American Statistical Association Fellow and Associate, respectively, at the U.S. Bureau of the Census, 1986.

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended by: Kirk M. Wolter

Report completed: November 25, 1986

Report issued: November 25, 1986

Spatial Modeling of Regional Variables

by

Noel Cressie* and Ngai H. Chan*

Abstract

In developing the statistics of spatial data there is a need for methods in both the areas of data analysis and statistical modeling. Here we analyze a data set of Sudden Infant Deaths, 1974 - 1978, in the counties of North Carolina, using a Markov-random-field approach to spatial modeling. We model the spatial trend with what we call large-scale-variation parameters, and the variance and spatial dependence with small-scale variation parameters. We show that a combination of resistant trend fitting to the data, and simple spatial auto Gaussian fitting to the residuals, is an effective way to analyze the (transformed) data.

Keywords: auto Gaussian, auto Poisson, large-scale variation, Markov random field, small-scale variation, sudden infant death

*Noel Cressie is Professor, Department of Statistics, Iowa State University, Ames IA 50011, and Ngai H. Chan is Assistant Professor, Department of Mathematics, Indiana University, Bloomington, IN 47405. This research was conducted while Cressie and Chan were ASA/NSF/Census Fellow and Associate respectively, at the U.S. Bureau of the Census. It was supported by NSF grant SES 84-01460, "On-Site Research to Improve Government-Generated Social Science Data Base". The authors would like to thank M. Symons and D. Atkinson for kindly showing them the 1979-1984 North Carolina SIDS data.

1 Introduction

Statistical analyses of data sets often assume, even at the exploratory stage, that the data are a random sample from a possibly multivariate distribution; i.e. data $Z_1,\cdots,Z_n$ are independent and identically distributed (i.i.d.) as F say. At the model-building stage F is often assumed to be a Gaussian distribution function (perhaps after a transformation $Y_i = g(Z_i)$).

More often than not, a large amount of "data preparation" is needed before the above assumptions are realistic. The not necessarily disjoint areas of robustness, (heteroscedastic) regression, and time series, all attempt to deal with departures from the i.i.d. Gaussian paradigm. Closest of interest to us in this article is the time series approach that looks for dependence in data received over time, where often the data occur at equally-spaced time points: $\{Z(t): t = 1,\cdots,n\}$.

However our concern here is with data occurring at a finite number of spatial locations that are not necessarily regularly spaced: $\{Z(\underset{\sim}{s}_i): i = 1,\cdots,n\}$, where $\{\underset{\sim}{s}_i: i = 1,\cdots,n\}$ are the data locations in d-dimensional Euclidean space $\mathbb{R}^d$; usually d = 2 or 3. Modeling of spatio-temporal data: $\{Z(\underset{\sim}{s}_i,t): i = 1,\cdots,n_t; t = 1,\cdots,m\}$, will not be considered here.

Our emphasis in this article is on fitting Markov-random-field models to data which are spatially "labeled". We have chosen a data set which is representative of many problems encountered in the health and social sciences. Counts of individuals from a known base occur in epidemiological studies (e.g. cancer incidence in a particular year from the base of

population-years-at-risk, for the counties of the USA), Census surveys (e.g. the dual-system-estimate of uncounted people in a decennial census from the base of total number of people, for the states of the USA), etc. Here we analyze the number of sudden infant deaths (SIDs) from the base of number of live births, for the counties of North Carolina, 1974-1978.

Sudden infant death syndrome (SIDS) is currently a leading category of postneonatal death yet its cause is still a mystery. Goldberg and Stein (1978) show there to be excess mortality in the winter months, which is consistent with, say, a genetic defect in the respiratory system; Fogerty et al. (1984) present evidence that poor nutrition in the form of high liver fatty acids has an effect; Symons, Grimson, and Yuan (1983) (hereafter, SGY) find geographical clustering in the North Carolina data set, which could be due to some surrogate variable such as race; and Cressie and Read (1986) (hereafter, CR) find both geographical clustering (large-scale variation) and spatial dependence (small-scale variation) from a spatial exploratory data analysis. Atkinson (1978) provides a review of the SIDS literature.

We shall present spatial models of the North Carolina data given by Atkinson (1978), and augmented by SGY and CR. The latter's Table 1 gives the complete data set; our Figure 1 shows a map of the counties (the numbers refer to their alphabetical ordering). Section 2 gives a summary of the spatial exploratory data analysis in CR and discusses other possible plots that look for association of SID rates with other variables. Section 3

develops the Markov-random-field models for both discrete and continuous data, featuring in particular the auto Poisson and the auto Gaussian models. Section 4 analyzes the North Carolina SIDS data using these models: when the mean structure is assumed constant, significant spatial association is detected, when a spatial trend is removed (by an additive two-way structure), the spatial association is no longer significant. Conclusions are drawn in Section 5.

Figure 1 here

---

2 Spatial exploratory data analysis of SIDS

The very use of the word "sudden" to describe this category of postneonatal death implies that very little is known about it. A recent "Request for Applications" from the National Institutes of Health (NIH Guide for Grants and Contracts, vol. 15, no. 2, January 31, 1986) gives the following background:

> ... SIDS accounts for about 7000 deaths a year, taking the lives of two infants per 1000 live births. The syndrome is defined as the sudden death of an infant that is unexpected by life history and where the death remains inexplicable after post-mortem examination. The [National Institute of Child Health and Human Development] Cooperative Epidemiological Study has identified some features found more frequently in SIDS victims than in age-matched control infants. The peak incidence of SIDS is between 2 and 4 months of age. It is more common in male infants, low birth-weight infants, black infants, infants of teenage mothers, and infants of mothers who smoked during pregnancy. SIDS infants also were likely to have received less postnatal pediatric care. Research has also revealed that SIDS victims as a group, tend

> to have more serious neonatal and early infant medical problems of various kinds. Pathologic and physiologic studies suggest that some SIDS victims have had chronic problems of respiratory control, which may make these infants vulnerable, especially in the event of mild upper respiratory infections.
>
> Despite the fact that various aspects of development physiology of young infants have been studied extensively over the past fifteen years, no specific biological markers for SIDS have been discovered, and the cause or causes of the syndrome remain unknown. ...

The number of SIDs and the number of live births from July 1, 1974 to June 30, 1978, for each of the counties of North Carolina and their county-seat locations, are found in Table 1 of CR. In that article, the authors went through the steps of a spatial exploratory data analysis to gain some understanding of how the data were behaving and to see whether a spatial statistical model would be appropriate. In the absence of any known cause for SIDS it is sensible to take this "univariate" look at the data: much as time series modeling looks to see how a particular value is influenced by its past values, spatial process modeling looks to see how a particular value is influenced by its "neighboring" values. The construction of spatial models is more complicated than temporal models however, because the space index does not possess the natural ordering that the time index does.

Exploratory data analysis is mostly distribution free, but we believe it is important when doing it to always have a vague statistical model in mind. CR were guided by this in their spatial data analysis of the North Carolina, 1974-1978, SIDS data. We summarize their techniques and findings here.

• Thinking of the sudden death of an infant or not, as being a Bernoulli "experiment", then the number of SIDs S_i out of n_i live births in county i is the sum of n_i (spatially dependent) Bernoulli random variables. Independence between the Bernoulli experiments means $S_i \sim \text{Bin}(n_i, p_i)$; $i = 1, \cdots, 100$, where it is of interest to determine the heterogeneity (or not) of $E(S_i/n_i) = p_i$. Note that the parameter of interest is also contained in $\text{var}(S_i/n_i) = p_i(1-p_i)/n_i$. Even when there is local spatial dependence between the Bernoulli experiments, $E(S_i/N_i) \equiv p_i$ is still the parameter of interest but now $\text{var}(S_i/N_i) = \{p_i(1-p_i)/n_i\}\{1+k_i\}$, where k_i accounts for the local spatial covariation. It makes sense then to look for a transformation of the data which removes the (gross) dependence of the variance on the mean. The presence of the unequal n_i's causes some difficulty, but by partitioning the set of 100 counties into six similarly sized subsets with roughly equal n_i values, CR are able to conclude that the Freeman-Tukey square root transformation

$$Y_i = (1000S_i/n_i)^{1/2} + (1000(S_i+1)/n_i)^{1/2}, \qquad (2.1)$$

is a variance-stabilizing transformation; i.e.

$$\text{var}(Y_i) \approx \tau^2/n_i. \qquad (2.2)$$

A comparison of stem-and-leaf plots of $\{1000(S_i+1)/n_i : i = 1, \cdots, 100\}$ and $\{Y_i : i = 1, \cdots, 100\}$ shows the

transformation (2.1) also to symmetrize the data.

- Thinking of the data as being made up of large-scale variation (spatial trend) plus small-scale variation (spatially dependent noise), each component could then be isolated and interpreted. The principle of transforming to achieve additivity over all scales (Cressie, 1985) and the two-dimensional (spatial) coordinates representation of the counties' locations, leads naturally to a two-way additive decomposition for the spatial trend. CR overlay a rectangular grid with 20 mile x 20 mile grid spacing onto North Carolina. This yields 9 east-west transects and 24 north-south transects, and 9x24 nodes of the grid; a county is identified with the node that is closest to its county seat. Let Y_i be written as $Y(u(i),v(i))$; $i = 1,\cdots,100$, where $(u(i),v(i)) \in \{(u,v): u = 1,\cdots,9, v = 1,\cdots,24\}$. Decompose

$$Y(u(i),v(i)) = m + r_{u(i)} + c_{v(i)} + \nu(x_i,y_i), \qquad (2.3)$$

where (x_i,y_i) are the Cartesian coordinates (given in Table 1 of CR) of the i-th county seat. Then $(m+r_{u(i)}+c_{v(i)})$ is a candidate for the large-scale variation component, and $\nu(x_i,y_i)$ for the small-scale variation component, with $var(\nu(x_i,y_i)) = \tau^2/n_i$. Because of the exploratory nature of the analysis, a <u>resistant</u> two-way decomposition is preferable; CR carry out a

weighted median polish, weighting each y_i proportional to its $(n_i)^{1/2}$. They check for the presence of a cross-product spatial trend using a diagnostic plot described in Cressie (1986), and do not find it. They map the fitted values $\{\tilde{m} + \tilde{r}_{u(i)} + \tilde{c}_{v(i)}: \ i = 1,\cdots,100\}$, now free from the unequal variation due to unequal n_i's, and note its essential agreement with a Poisson model-based probability map.

- Thinking of the residuals

$$R(x_i,y_i) = Y_i - \tilde{m} - \tilde{r}_{u(i)} - \tilde{c}_{v(i)} \tag{2.4}$$

as estimating the small-scale variation, stationary error models could be fit to the detrended and standardized data

$$\{\sqrt{n_i}\ R(x_i,y_i): \ i = 1,\cdots,100\}. \tag{2.5}$$

CR use its stem-and-leaf plot to estimate the stationary density, and to show that the standardized residual of Anson county (county 4) is unacceptably high. It is removed from subsequent analysis in CR and from our model-based approach in Section 4. An estimate of the variogram (a measure of spatial dependence) is computed from the standardized residuals (2.5), and positive spatial dependence is found up to 30 miles (i.e. local dependence).

Therefore, from CR's data analysis we conclude that county 4 is unusual and should be removed, and that Freeman-Tukey transformed SID rates can be sensibly modeled as two-way additive spatial trend plus stationary error that is positively correlated over short distances.

This "univariate" data analysis should in general be followed up with various scatter plots of Y on potential explanatory variables X. Any dependence observed could be modeled as large-scale variation, helping to explain the spatial trend. With an additive model, the coefficients of both explanatory variables and trend variables can then be estimated efficiently by maximum likelihood, based on a model that includes the spatial correlations. An explanatory variable whose regression coefficient is significantly different from zero, offers an exciting possible advance towards explaining the mystery of SIDS. We show here in Figure 2 a plot of Y versus X = arcsin(% black), where each point is (Y_i, X_i), X_i being the (arcsin transformed) percentage of blacks in the i-th county of North Carolina (source: County and City Data Book, 1977, Bureau of the Census, U.S. Department of Commerce, Washington D.C. 20233). This shows a very weak positive relationship, although it is expected from some of the introductory remarks in this section. Unfortunately this relationship is even weaker for more recent values, 1979 to 1984. Other plots (not shown here) drawn were: Y versus X = population density, Y versus X = $\sin^{-1}$(% urban), Y versus X = log(no. hospital beds + 1); none showed any relationship between (transformed) SID rate Y, and variable X.

Therefore we look for a model for the $\{Y_i: \ i = 1,\cdots,n\}$ based purely on spatial considerations.

Figure 2 here

---

3 Markov random fields and spatial models

If the data set $\{Z(i): \ i = 1,\cdots,n\}$ is in fact indexed by time; i.e. $\{1,2,\cdots,n\}$ are the times of occurrences of the data; then it is often not appropriate to build a statistical model that assumes the $Z(i)$'s are independent. Assuming that the data are generated by a Markov process, is many times a very effective way of introducing dependence. Simply stated, the Markov property says that given the whole history of the process up to the present, the distribution of the present value depends only on the immediate past value observed (this property is sometimes generalized to include the immediate past two values, etc.).

It is not obvious how to generalize this type of temporal dependence to a spatial dependence. Recall that the data are

$$\{Z(\underset{\sim}{s}_i): \ i = 1,\cdots,n\}, \tag{3.1}$$

where

$$\{\underset{\sim}{s}_i: \ i = 1,\cdots,n\}, \tag{3.2}$$

are the spatial locations of the Z's. For the purposes of this article, these locations will determine entirely the spatial dependence structure between the data, although more general

formulations are possible. For example, the same type of modeling can be carried out when the regions $1,\cdots,n$ can be represented as the vertices of an undirected graph (Kindermann and Snell, 1980). Fix consideration on county i and consider the conditional distribution of $Z(\underset{\sim}{s}_i)$ given all other values $\{Z(\underset{\sim}{s}_j): j \neq i\}$. Then the spatial analogue to the Markov property is that this conditional distribution in fact depends only on a subset $\{Z(\underset{\sim}{s}_j): j\varepsilon N_i\}$. The set N_i is called the neighborhood of region i, and for the purposes of this article will be defined according to the distance $d(\underset{\sim}{s}_i,\underset{\sim}{s}_j)$ between region i and region j; $j \neq i$. More generally, the edges of an undirected graph with vertices $1,2,\cdots,n$, can be used to define

$$\{N_i: \ i = 1,\cdots,n\}. \tag{3.3}$$

The Hammersley-Clifford theorem (Besag, 1974) gives the conditions needed for the conditional probability distributions, viz.

$$\Pr(Z(\underset{\sim}{s}_i)\,|\,\{Z(\underset{\sim}{s}_j): j\varepsilon N_i\};\ i = 1,\cdots,n, \tag{3.4}$$

to yield a well-defined joint probability distribution

$$\Pr(Z(\underset{\sim}{s}_1),\cdots,Z(\underset{\sim}{s}_n)). \tag{3.5}$$

Expressions (3.4) and (3.5) are examples of the notation $\Pr(\{Z(\underset{\sim}{s}_i): i\varepsilon E_1\}\,|\,\{Z(\underset{\sim}{s}_j): j\varepsilon E_2\})$, for E_1, E_2 disjoint subsets

of $\{1,2,\cdots,n\}$, which is shorthand for the conditional probability measure of $\{Z(\underset{\sim}{s}_i): i\in E_1\}$, conditioned on the σ-algebra generated by $\{Z(\underset{\sim}{s}_j): j\in E_2\}$. When expressions (3.4) (i.e. when the neighborhoods (3.3)) define (3.5), the joint probability distribution is called a Markov random field.

One consequence of the Hammersley-Clifford theorem is that

$$Q(\underset{\sim}{z}) \equiv \log(P(\underset{\sim}{z})/P(\underset{\sim}{0})), \tag{3.6}$$

where $P(\cdot)$ is the joint density or probability mass function of $\{Z(\underset{\sim}{s}_i): i = 1,\cdots,n\}$, can be represented as a sum of functions of subsets of $\{z_1,\cdots,z_n\}$, those subsets being indexed by the cliques. A clique is a single region, or a set of regions for which any pair in the set are neighbors; it is defined uniquely by the neighborhoods (3.3). By assuming further that only pairwise interactions are nonzero, i.e.

$$Q(\underset{\sim}{z}) = \sum_{i=1}^{n} z_i G_i(z_i) + \sum_{i<j=1}^{n}\sum z_i z_j G_{ij}(z_i,z_j), \tag{3.7}$$

and (3.4) is of the exponential family form, Besag (1974) shows the spatial models to take on an easily interpretable form.

Suppose the density or probability mass function of (3.4) can be written as

$$P(z_i|\{z_j: j\in N_i\}) = \exp[A_i(\{z_j: j\in N_i\})B_i(z_i) + C_i(z_i) + D_i(\{z_j: j\in N_i\})], \tag{3.8}$$

and $Q(\underset{\sim}{z})$ is of the form (3.7). Then Besag (1974) shows that

$$A_i(\{z_j: j\varepsilon N_i\}) = \alpha_i + \sum_{j\varepsilon N_i} \alpha_{ij} B_j(z_j), \tag{3.9}$$

where in the terminology of Section 2, α_i is a parameter that can be used to model the large-scale variations, and $\{\alpha_{ij}: j\varepsilon N_i\}$ are small-scale variation parameters that model the spatial dependence (when $\alpha_{ij} = 0$; $i,j = 1,\cdots,n$ in (3.8), the joint independence model results). It is a consequence of the model that $\alpha_{ji} = \alpha_{ij}$; further define $\alpha_{ij} = 0$; $j \notin N_i$.

We shall give two examples, both of which will be used in the next section to model the SIDS data. The first is a discrete model, whose conditional probability mass functions are

$$P(z_i|\{z_j: j\varepsilon N_i\}) = \exp(-\theta_i)\theta_i^{z_i}/z_i!;\ i = 1,\cdots,n, \tag{3.10}$$

where $\theta_i = \theta_i(\{z_j: j\varepsilon N_i\})$ is a function of values observed for the neighboring regions. Then (3.8) and (3.9) together yield

$$\theta_i(\{z_j: j\varepsilon N_i\}) = E(Z(\underset{\sim}{s}_i)|\{z_j: j\varepsilon N_i\}) = \exp(\alpha_i + \sum_{j\varepsilon N_i} \alpha_{ij} z_j) \tag{3.11}$$

which is of log linear form. This spatial model is often called the auto Poisson.

The second example is a continuous model, whose conditional density is

$$P(z_i|\{z_j: j\varepsilon N_i\}) = (2\pi\tau_i^2)^{-1/2}\exp\{-(z_i-\theta_i)^2/2\tau_i^2\}, \tag{3.12}$$

where $\theta_i = \theta_i(\{z_j: j\varepsilon N_i\})$. Then (3.8) and (3.9) together yield

$$\theta_i(\{z_j: j\in N_i\}) = E(Z(\underset{\sim}{s}_i)|\{z_j: j\in N_i\}) = \tau_i^2\{\alpha_i + \sum_{j\in N_i} \alpha_{ij} z_j\}, \quad (3.13)$$

which is of linear form. This spatial model is often called the auto Gaussian.

With the parameters $\{\alpha_i: i = 1,\cdots,n\}$ and $\{\alpha_{ij}: j\in N_i, i = 1,\cdots,n\}$ specified through the conditional model, but estimated by maximum likelihood, there is a need to reconstruct the joint probability distribution (3.5). The auto Poisson and auto Gaussian cases will illustrate some of the difficulties involved. The auto Poisson defined by (3.10) has a joint distribution which can be obtained from

$$Q(\underset{\sim}{z}) = \sum_{i=1}^{n} \alpha_i z_i + \sum_{i=1}^{n} \sum_{j\in N_i} \alpha_{ij} z_i z_j - \sum_{i=1}^{n} \log(z_i!); \quad (3.14)$$

specifically from (3.6),

$$P(\underset{\sim}{z}) = \exp(Q(\underset{\sim}{z}))/\sum_{\underset{\sim}{z}} \exp(Q(\underset{\sim}{z})), \quad (3.15)$$

provided the normalizing constant $\sum_{\underset{\sim}{z}} \exp(Q(\underset{\sim}{z})) < \infty$. This latter condition means that the parameter space is restricted to be $\{(\alpha_1,\cdots,\alpha_n), (\alpha_{ij}): \sum_{\underset{\sim}{z}} \exp(Q(\underset{\sim}{z})) < \infty\}$, which in turn restricts $\alpha_{ij} \leq 0$, for all i,j (Besag, 1974). Thus the auto Poisson can only model negative spatial dependence, a consequence we shall discuss in Section 4. By substituting (3.14) into (3.15) and interpreting it as a function of (the parameters) $\underset{\sim}{\alpha}$ and (α_{ij}), we obtain the likelihood. In practice its maximization is not possible because the normalizing constant in (3.15), involving n-

fold infinite summations, is intractable. This led Besag (1975) to propose maximizing the pseudo likelihood function

$$\prod_{i=1}^{n} \exp(-\theta_i)\theta_i^{z_i}/z_i!, \tag{3.16}$$

which from (3.11) can be seen to be an easily calculable function of the parameters. In special cases (Besag, 1977) the loss of efficiency due to using maximum pseudo likelihood estimation, is found to be acceptable over most of the parameter space. There are no general results available for the behavior of these estimators.

For the auto Gaussian case, we rewrite (3.12) and (3.13) equivalently as

$$E(Z(\underset{\sim}{s}_i)|\{z(\underset{\sim}{s}_j): j\in N_i\}) = \mu_i + \sum_{j=1}^{n} c_{ij}(z(\underset{\sim}{s}_j)-\mu_j), \tag{3.17}$$

$$\text{var}(Z(\underset{\sim}{s}_i)|\{z(\underset{\sim}{s}_j): j\in N_i\}) = \tau_i^2,$$

and the conditional distribution of $Z(\underset{\sim}{s}_i)$ is Gaussian; $i = 1,\cdots,n$. Besag (1974) shows that provided $M^{-1}(I-C)$ is symmetric, positive definite and invertible, $\underset{\sim}{Z} = (Z(\underset{\sim}{s}_1),\cdots,Z(\underset{\sim}{s}_n))$ has joint Gaussian distribution:

$$\underset{\sim}{Z} \sim \text{Gau}(\underset{\sim}{\mu},(I-C)^{-1}M), \tag{3.18}$$

where $\underset{\sim}{\mu} = \{\mu_1,\cdots,\mu_n)$, $M = \text{diag}(\tau_1^2,\cdots,\tau_n^2)$, and $C = (c_{ij})$; $c_{ij} = 0$, $j \notin N_i$; $c_{ij}\tau_j^2 = c_{ji}\tau_i^2$. Notice that $\underset{\sim}{\mu}$ captures the large-scale

variation and C models the spatial dependence. Hence the likelihood is:

$$\ell(\underset{\sim}{\mu},M,C) = (2\pi)^{-n/2}|M^{-1}(I-C)|^{1/2}\exp\{-(1/2)(\underset{\sim}{z}-\underset{\sim}{\mu})'M^{-1}(I-C)(\underset{\sim}{z}-\underset{\sim}{\mu})\}, \tag{3.19}$$

which is to be maximized with respect to $\underset{\sim}{\mu}$, M, and C. Equivalently, the negative log likelihood:

$$L(\underset{\sim}{\mu},M,C) = (n/2)\log(2\pi) + (1/2)\log|(I-C)^{-1}M| + (1/2)(\underset{\sim}{z}-\underset{\sim}{\mu})'M^{-1}(I-C)(\underset{\sim}{z}-\underset{\sim}{\mu}), \tag{3.20}$$

is to be minimized with respect to $\underset{\sim}{\mu}$, M, and C. There are a number of exact and approximate ways to do this minimization depending on how the model is further parametrized; see Cressie (1987), Section 7.2 for the details.

In general, these models are over-parametrized; some reduction is necessary to allow accurate estimation of variances of estimators, and good predictive power of the model. In classical statistics we have a lot of experience with parametrizations like $\underset{\sim}{\mu} = X\underset{\sim}{\beta}$, but very little experience with how to parametrize the spatial dependence matrix C. In fact it is easy to show that the partial autocorrelation satisfies

$$\text{corr}(Z(\underset{\sim}{s}_i),Z(\underset{\sim}{s}_k)|\{Z(\underset{\sim}{s}_j): \ j \neq i,k\}) = \text{sgn}(c_{ik})(c_{ik}c_{ki})^{1/2}, \tag{3.21}$$

which is symmetric in i and k since $c_{ik}\tau_k^2 = c_{ki}\tau_i^2$. Thus the c_{ij}'s are essentially (conditional) correlation coefficients; in particular (3.21) implies that $0 < c_{ik}c_{ki} < 1$, or $c_{ik}^2(\tau_k^2/\tau_i^2) < 1$.

In the next section, various suggestions for further parametrizing the large-scale and small-scale variation parameters will be made.

4 Spatial modeling of SIDS data

We shall try two approaches to the analysis of these data: one is to recognize that $\{S_i: i = 1,\cdots,n\}$ are counts and to fit a discrete model, in particular the auto Poisson given by (3.10) and (3.11); the other is to transform the data (here the Freeman-Tukey transformation will be used) so that they follow a (approximate) Gaussian distribution and to fit the auto Gaussian given by (3.17). Both are approximations. The auto Poisson implicitly assumes that the rate p_i is small and the number of live births n_i is large, so that $n_i p_i = E(S_i) = E(E(S_i|\{S_j: j\epsilon N_i\})) = E(\text{var}(S_i|\{S_j: j\epsilon N_i\}))$. This seems reasonable since the overall SID rate is $\hat{p} = \sum_{i=1}^{100} S_i / \sum_{i=1}^{100} n_i = .0020214$, and $248 \leq n_i \leq 21{,}588$; $i = 1,\cdots,100$. The auto Gaussian is a model for continuous data fit to transformed counts; Figures 5 and 8 of CR (stem-and-leaf plots of $\{Y_i\}$ and $\{\sqrt{n_i}R_i\}$) show this approximation to be reasonable.

The heterogeneous live-birth counts $\{n_i: i = 1,\cdots,100\}$, necessitate caution when parametrizing both scales of variation in the models. We start with the large-scale variation.

4.1 Modeling large-scale variation

In the absence of any obvious explanatory variable for SIDS, we resort to fitting spatial trend to account for the large-scale

variation. For the auto Poisson model (3.10) and (3.11), where the SID counts S_i play the role of the $Z(\underset{\sim}{s}_i)$, we fit

$$\begin{aligned} \text{I:} \quad & \exp(\alpha_i) = n_i\rho \quad ; i = 1,\cdots,100 \\ \text{II:} \quad & \exp(\alpha_i) = n_i\rho_{R(i)}; i = 1,\cdots,100 \\ \text{III:} \quad & \exp(\alpha_i) = n_i\rho_{T(i)}; i = 1,\cdots,100, \end{aligned}$$

where model I fits only one large-scale-variation parameter, and corresponds to a homogeneous rate model. Model II uses 12 parameters, each corresponding to one of 12 artificially created groups of counties (see Figure 1) which partition the counties of the state into contiguous parcels. Thus,

$$R:\{1,2,\cdots,100\} \to \{1,2,\cdots,12\}. \tag{4.1}$$

Model III uses 32 parameters, which are obtained by overlaying the same 9x24 grid referred to in Section 2 onto the county map of North Carolina. For $i = 1,\cdots,100$, county i is assigned to the nearest node $(u(i),v(i))$ of the grid and its rate is modeled via

$$\begin{aligned} \exp(\alpha_i) &= n_i\rho_{T(i)} \\ &= n_i\exp(m+r_{u(i)}+c_{v(i)}), \end{aligned} \tag{4.2}$$

a multiplicative two-way fit to the SID rates. With restrictions on the parameter to make them identifiable, there are 9+24-1 = 32

large-scale variation parameters to be estimated.

Now consider the auto Gaussian model (3.17), where the Freeman-Tukey transformed Y_i given by (2.1) play the role of the $Z(\underset{\sim}{s}_i)$. We fit $\underset{\sim}{\mu}_{n\times 1} = X_{n\times p}\underset{\sim}{\beta}_{p\times 1}$ as follows:

$$\text{I:} \qquad \mu_i = m \quad ; \; i = 1,\cdots,100; \qquad (4.3)$$
$$\text{i.e.} \quad \underset{\sim}{\beta} = m.$$

$$\text{II:} \qquad \mu_i = m_{R(i)}; \; i = 1,\cdots,100; \qquad (4.4)$$
$$\text{i.e.} \quad \underset{\sim}{\beta} = (m_1,\cdots,m_{12}).$$

$$\text{III:} \qquad \mu_i = m_{T(i)}; \; i = 1,\cdots,100; \qquad (4.5)$$
$$\text{i.e.} \quad \underset{\sim}{\beta} = (m,r_1,\cdots,r_8,c_1,\cdots,c_{23}).$$

In model II, $R(\cdot)$ is defined by (4.1) where there are $p = 12$ large-scale-variation parameters to be estimated. In model III, $m_{T(i)} = m + r_{u(i)} + c_{v(i)}$; $i = 1,\cdots,100$; with identifiability restrictions on the parameters, there are $p = 32$ of them to be estimated.

4.2 Modeling small-scale variation

To define the Markov random field we used the same neighborhood system as CR, namely

$$j\in N_i \quad \text{iff} \quad d_{ij} \equiv \{(x_i-x_j)^2 + (y_i-y_j)^2\}^{1/2} \leq 30 \text{ miles}, \qquad (4.6)$$

where recall that (x_i,y_i) is the location of the i-th county

seat. Now any $j \notin N_i$ has $\alpha_{ij} = 0$, and for those nonzero α_{ij}'s left, we tried to capture the spatial association with just one parameter ϕ, and model different dependencies between county i and its neighbors as a function of distance (Cliff and Ord, 1981, p. 144):

$$\alpha_{ij} \propto \phi d_{ij}^{-k};\ j \in N_i;\ k = 0,1,2,$$

for the auto Poisson, and

$$c_{ij} \propto \phi d_{ij}^{-k}(n_j/n_i)^{1/2};\ j \in N_i;\ k = 0,1,2,$$

for the auto Gaussian. For comparability across different values of k, we chose the constant of proportionality (it has to be in the same units as d_{ij}^{k}) to be

$$C(k) = (\min\{d_{ij}: j \in N_i,\ i = 1,\cdots,100\})^{k}. \qquad (4.7)$$

Thus

$$\alpha_{ij} = \phi.\{C(k)d_{ij}^{-k}\};\ j \in N_i;\ k = 0,1,2, \qquad (4.8)$$

for the auto Poisson, and

$$c_{ij} = \phi.\{C(k)d_{ij}^{-k}\}(n_j/n_i)^{1/2};\ j \in N_i;\ k = 0,1,2, \qquad (4.9)$$

for the auto Gaussian; (4.9) is compatible with

$$\text{var}(Y_i) = \tau^2/n_i, \tag{4.10}$$

the variance relationship alluded to in Section 2. Note that $c_{ij}\tau_j^2 = c_{ji}\tau_i^2$, and $c_{ij}c_{ji} = \phi^2\{C(k)d_{ij}^{-k}\}^2$, independent of n_i and n_j.

Putting these small-scale variation parameters together with the large-scale ones, we see that the auto Poisson model I, II, III has respectively 2, 13, and 33 parameters to be fit, and the auto Gaussian model I, II, III has respectively, 3, 14, and 34 parameters (τ^2 in (4.10) is the extra parameter) to be fit.

4.3 Pseudo likelihood-based fitting of the auto Poisson model

In all of the analysis to follow, Anson County (county 4) is omitted. Its SID count of 15 out of 1570 live births is very unlike the other 99 counties. This unusually high rate was not repeated from 1979 to 1984 when only 4 SIDs out of 1875 live births were reported for that county.

We used the pseudo likelihood (3.16), to fit the auto Poisson models I, II, and III with spatial dependence modeled by (4.8). The pseudo likelihood was maximized by the IMSL subroutine ZXMIN, which is based on a quasi-Newton optimization method. Starting with the very simple two-parameter auto Poisson model I, and putting $k = 1$ in (4.8), we obtained $\hat{\rho} = 1.785 \times 10^{-3}$, and $\hat{\phi} = 8.216 \times 10^{-3}$. The first thing to note is that $\hat{\phi} > 0$; i.e. positive spatial dependence is obtained, which as has been shown in Section 3 does not correspond to a well-defined auto Poisson model. This pattern is repeated when (unconstrained)

maximum pseudo likelihood fitting of the data to auto Poisson models II and III is carried out for k = 1 and 2. Negative values of ϕ were obtained for k = 0, however the minimized value of the pseudo likelihood in this case was always larger than the corresponding minimized values for nonzero k's.

Even should a well-defined auto Poisson model result, we still need to make inferences on the parameters, based on their estimators. For example, an important (spatial) hypothesis to test is H_0: $\phi = 0$, versus H_1: $\phi < 0$. Finding (asymptotic) distribution theory for maximum pseudo likelihood estimators is at present an open problem. Clayton and Kaldor (1984) model lip-cancer incidence by assuming that their joint distribution, conditional on the county rates $\{p_i: i = 1,\cdots,n\}$, are independent Poisson random variables with means $\lambda_i = n_i p_i$. They then assume that the rates themselves are random and follow a Markov random field. In particular they assume $\{-\log p_i: i = 1,\cdots,n\}$ is auto Gaussian with large-scale variation (mean) modeled using an age variable, and small-scale variation (variance matrix) modeled by essentially the k = 0 case in (4.9). These parameters are estimated by empirical Bayes methods.

4.4 Likelihood-based fitting of the auto Gaussian model.

We now proceed with fitting an auto Gaussian model to the Freeman-Tukey transformed SID rates for 99 of the 100 counties of North Carolina, 1974-1978. The negative log likelihood is given by $L(\underset{\sim}{\mu},M,C)$ in (3.20), which after the parametrizations $\underset{\sim}{\mu} = X\underset{\sim}{\beta}$ in (4.3), (4.4), (4.5), $M = \tau^2 \mathrm{diag}(n_1^{-1},\cdots,n_{100}^{-1})$ in (4.10), and c_{ij}

in (4.9), we shall write as

$$L(\underset{\sim}{\beta},\tau^2,\phi) = (n/2)\log(2\pi) + (n/2)\log\tau^2 - (1/2)\log|D^{-1}(I-\phi H)| + (1/2)(\underset{\sim}{z}-X\underset{\sim}{\beta})'D^{-1}(I-\phi H)(\underset{\sim}{z}-X\underset{\sim}{\beta})/\tau^2, \tag{4.11}$$

where $D = \text{diag}(n_1^{-1},\cdots,n_{100}^{-1})$ and $H = (h_{ij})$; $h_{ij} = C(k)d_{ij}^{-k}(n_j/n_i)^{1/2}$ if $j\varepsilon N_i$, $= 0$ otherwise.

To minimize (4.11), assume for the moment that ϕ is fixed. The maximum likelihood estimators (mle's) of $\underset{\sim}{\beta}$ and τ^2 are

$$\hat{\underset{\sim}{\beta}}(\phi) = (X'D^{-1}(I-\phi H)X)^{-1}X'D^{-1}(I-\phi H)\underset{\sim}{z} \tag{4.12}$$

$$\hat{\tau}^2(\phi) = (\underset{\sim}{z}-X\hat{\underset{\sim}{\beta}})'D^{-1}(I-\phi H)(\underset{\sim}{z}-X\hat{\underset{\sim}{\beta}})/n$$

$$= \underset{\sim}{z}'D^{-1}(I-\phi H)\{I-X(X'D^{-1}(I-\phi H)X)^{-1}X'D^{-1}(I-\phi H)\}\underset{\sim}{z}/n. \tag{4.13}$$

Substituting (4.12) and (4.13) back into (4.11), the mle of ϕ can be obtained by minimizing $L(\phi) = L(\hat{\underset{\sim}{\beta}}(\phi),\hat{\tau}^2(\phi),\phi)$ given by

$$L(\phi) = (99/2)\log(2\pi) + (99/2) - (1/2)\sum_{i\neq 4}\log(n_i) + (99/2)\log\hat{\tau}^2(\phi) - (1/2)\sum_{i=1}^{99}\log(1-\phi\delta_i), \tag{4.14}$$

where $\{\delta_i: i = 1,\cdots,99\}$ are the eigenvalues of the symmetric matrix $D^{-1/2}HD^{1/2}$. We would like to remind the reader here that county 4 has been omitted from the analysis, and hence there are

$n = 99$ pieces of data to analyze. Figure 3 shows a plot of $L(\phi)$ against ϕ for auto Gaussian model I given by (4.3), (4.9) and (4.10), and for $k = 1$ (i.e. $c_{ij} \propto d_{ij}^{-1}$). This enables a $100(1-\alpha)\%$ confidence interval for ϕ to be determined.

Figure 3 here

From Whittle (1954),

$$\Pr\{\phi: \; L(\phi) \leq L(\hat{\phi}) + (n/(n-p-2))\chi_1^2(\alpha)/2\} \simeq 1 - \alpha, \qquad (4.15)$$

where $L(\phi)$ is given by (4.14), $\hat{\phi}$ is the mle obtained by minimizing $L(\cdot)$, p is the number of large-scale parameters fitted, and $\chi_1^2(\alpha)$ is the upper $100(1-\alpha)\%$ point of the chi-square distribution on one degree of freedom. For $n = 99$, $p = 1$, $\alpha = .05$, the 95% confidence interval becomes $\{\phi: L(\phi) \leq L(\hat{\phi}) + 1.98\}$. Now for auto Gaussian model I, $\hat{\phi} = 0.833$, $L(\hat{\phi}) = 124.87$, and hence $\{\phi: L(\phi) \leq 126.85\}$ is a 95% confidence interval. Since $L(0) = 130.26$, we reject the null hypothesis H_0 in

$$H_0: \; \phi = 0, \quad \text{versus} \quad H_1: \; \phi \neq 0. \qquad (4.16)$$

Thus the spatial interaction <u>is</u> significant. We shall see in Table 1 what happens to this spatial interaction as more and more parameters are used to model the large-scale variation.

To see the effect on the models for varying values of k, we

show in Figure 4 a plot of $L(\phi)$ against ϕ for auto Gaussian model II (given by (4.4), (4.9) and (4.10)) for $k = 0$ ($c_{ij} \propto 1$), $k = 1$ ($c_{ij} \propto d_{ij}^{-1}$), and $k = 2$($c_{ij} \propto d_{ij}^{-2}$) together. The likelihoods are considerably flatter, and it can be seen from (4.15) and Table 1 that $\phi = 0$ is in the 95% confidence interval for ϕ (for $k = 1$, and incidentally for $k = 0$, and 2 as well).

Figure 4 here

Table 1 here

To see the effect on $\hat{\phi}$ and the confidence intervals for ϕ, of fitting more and more large-scale parameters, we show these statistics in Table 1 for auto Gaussian models I, II, III given by (4.3), (4.4), (4.5) respectively. For model III, it looks as though the 32 large-scale parameters in $\underset{\sim}{\beta}$ remove all spatial dependence (as measured by $\hat{\phi}$). Mardia and Marshall (1984) and Cressie and Glonek (1984) note that these maximum likelihood (linear) estimates for spatial trend cause problematic negative bias in the estimation of the covariance parameter ϕ. Cressie (1984) suggests a resistant, median-based (nonlinear) fit to the spatial trend, namely median polish (see Section 2). The residuals $\{R_i: i = 1,\cdots,n\}$ from this fit, defined by (2.4), can be analyzed then according to auto Gaussian model I, viz. (4.3),

(4.9) and (4.10). To maintain comparability we chose k = 1, and plot simultaneously the graph of $L(\phi)$ against ϕ, for Model III on the Y_i's and for Model I on the R_i's; see Figure 5. It is clear that the resistant method of fitting has the anticipated effect, namely to revive spatial correlation in the data, although it is not significant; see Table 1. This agrees with the findings of CR, who observe spatial correlation in the R_i's up to thirty miles.

Figure 5 here

To see the effect of the different models on the correlations between pairs of counties, we use (3.18) to conclude that for the auto Gaussian model defined by its conditional distributions,

$$\text{var}(\underset{\sim}{Y}) = (I-C)^{-1}D\tau^2,$$

and hence

$$\hat{\text{corr}}(Y_i, Y_j) = a_{ij}/(a_{ii}a_{jj})^{1/2}, \tag{4.17}$$

where

$$A = (a_{ij}) = (I-\hat{\phi}D^{-1/2}HD^{1/2})^{-1}. \tag{4.18}$$

Figure 6 shows a plot of $\text{corr}(Y_{80},Y_j)$ versus $d_{80,j} = \{(x_{80}-x_j)^2 + (y_{80}-y_j)^2\}^{1/2}$ for model I, model II, and model I on the R_i's, all for $k = 1$. The effect of fitting more and more large-scale variation parameters is apparent: the more parameters, the smaller the ϕ, the weaker the intercounty correlation.

Figure 6 here

We turn now to the mle's $\hat{\underset{\sim}{\beta}}$ in $X\underset{\sim}{\beta}$, defined by (4.3), (4.4), and (4.5). Table 2 presents the estimates for $k = 1$. Its variance matrix is estimated by

$$\hat{\text{var}}(\hat{\underset{\sim}{\beta}}) = (X'D^{-1}(I-\hat{\phi}H)X)^{-1}\hat{\tau}^2, \qquad (4.19)$$

but not shown here since for model II, III this is a 12x12, 32x32 matrix respectively. For model I, $\hat{m} = 2.838$ with $\hat{\text{var}}(\hat{m}) = 0.006218$. To give some idea of $\hat{\underset{\sim}{\mu}}$ under the different models, let us choose say county 80 and find its corresponding entry in the 99 x 1 vector $X\hat{\underset{\sim}{\beta}}$: For model I, $\hat{\mu}_{80} = \hat{m} = 2.84$; for model II, $\hat{\mu}_{80} = \hat{m}_5 = 2.15$ (reflecting its presence in a group of counties with low rates); for model III, $\hat{\mu}_{80} = \hat{m} + \hat{r}_4 + \hat{c}_{11} = 1.96$; for model I on the R_i's, $\hat{\mu}_{80} = \tilde{m} + \tilde{r}_4 + \tilde{c}_{11} + \hat{m}(R) = 2.13$ (cf model II). Its actual observed value is $Y_{80} = 1.74$.

Table 2 here

By conditioning on a county's neighboring values, we can use the model to predict that county's value. Under the auto Gaussian model the joint distribution is Gaussian, given by (3.18), and hence the optimal predictor (minimizes mean-squared error) is

$$P_i = E(Y_i | \{Y_j: j \in N_i\}) = \mu_i + \sum_{j \in N_i} c_{ij}(Y_j - \mu_j); \; i = 1, \cdots, n.$$

Therefore

$$\underset{\sim}{P} = \underset{\sim}{\mu} + C(\underset{\sim}{Y} - \underset{\sim}{\mu}),$$

which is estimated by

$$\hat{\underset{\sim}{P}} = X\hat{\underset{\sim}{\beta}} + (\hat{\phi}H)(\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}). \tag{4.21}$$

Ignoring the variation in $\hat{\phi}$, the prediction variance is $\hat{var}(\hat{\underset{\sim}{P}}) = F\hat{\Sigma}F'$, where $F = (I - \hat{\phi}H)(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1} + \hat{\phi}H$, and $\hat{\Sigma} = (I - \hat{\phi}H)^{-1}D\hat{\tau}^2$. Since county 4 was omitted, we might predict it from its neighbors, viz. counties 62, 77, 84, and 90. Fix k = 1. For model I, $\hat{Y}_4 = 2.82$ ($\hat{m} = 2.84$); for model II, $\hat{Y}_4 = 2.68$ ($\hat{m}_9 = 2.68$); for model III, $\hat{Y}_4 = 2.17$ ($\hat{m} + \hat{r}_6 + \hat{c}_{12} = 2.17$). For model I on the R_i's, $\hat{Y}_4 = \tilde{m} + \tilde{r}_6 + \tilde{c}_{12} + \hat{R}_4 = 1.81$ ($\tilde{m} + \tilde{r}_6 + \tilde{c}_{12} + \hat{m}(R) = 1.75$). The actual observed value is $Y_4 = 6.28$.

Finally, we need to interpret the estimates of τ^2. For model I, $\hat{\tau}^2 = 1443.17$; for model II, $\hat{\tau}^2 = 864.61$; for model III, $\hat{\tau}^2 = 681.91$; for model I on the R_i's, $\hat{\tau}^2 = 841.37$. These mle's

are biased estimates of τ^2 for the same reason the $\hat{\phi}$ are biased estimates of ϕ (Mardia and Marshall, 1984; Cressie and Glonek, 1984). Between $\hat{\tau}^2 = 681.91$ and $\hat{\tau}^2 = 841.37$, least biased will be the resistance-based $\hat{\tau}^2 = 841.37$. Thus $2\hat{\tau}^2 = 1683$, which is roughly the "sill" value of the variogram estimator given in CR's Figure 11, as it should be.

Recall in Section 2 we addressed the question of spatial dependence between Bernoulli experiments, which can be characterized by the parameter k_i in

$$\text{var}(S_i/n_i) = \{p_i(1-p_i)/n_i\}\{1+k_i\};\ i = 1,\cdots,n. \qquad (4.22)$$

Now although the k_i's are parameters of the local spatial dependence, we can estimate them from the more global county data. It can be shown that

$$\hat{k}_i = a_{ii} - 1, \qquad (4.23)$$

where a_{ii} is given by (4.18). For example, consider county 80: for model I, $\hat{k}_{80} = 0.0584$; for model II, $\hat{k}_{80} = .0405$; for model III, $\hat{k}_{80} = .0005$; for model I on the R_i's, $\hat{k}_{80} = .0299$. The same pattern is present: as more and more large-scale variation parameters are fit, the local spatial dependence decreases, revived somewhat by (weighted) median polish fitting of trend followed by fitting auto Gaussian model I to its residuals.

5 Discussion and Conclusions

Sudden infant death data $\{(S_i, n_i): i = 1, \cdots, 100\}$ of the counties of North Carolina, 1974-1978, have been analyzed as follows:

- Transform the data to

$$Y_i = (1000S_i/n_i)^{1/2} + (1000(S_i+1)/n_i)^{1/2},$$

the Freeman-Tukey-transformed death rates per thousand.

- Delete county 4 from the analysis as an outlier.

- Using the Markov-random-field approach, fit by maximum likelihood, the spatial model

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{5.1}$$

where the large-scale-variation parameter vector $\underset{\sim}{\beta}$ has either one component m (model I), 12 components $m_1, \ldots, m_{12}$ (model II), or 32 components $m, r_1, \ldots, r_8, c_1, \ldots, c_{23}$ (model III), and

$$\underset{\sim}{\varepsilon} \sim \text{Gau}(\underset{\sim}{0}, (I-\phi H)^{-1} D\tau^2); \tag{5.2}$$

$D = \text{diag}(n_1^{-1}, \cdots, n_{100}^{-1})$, $H = (d_{ij}^{-k} C(k) I(j \varepsilon N_i))$. Small-scale-variation parameters of the model are τ^2 and ϕ (respectively variance and spatial dependence parameter).

- An adhoc but nevertheless less-biased (in estimation of small-scale-variation parameters) approach was to fit by maximum likelihood

$$\underset{\sim}{R} = m\underset{\sim}{1} + \underset{\sim}{\varepsilon}, \tag{5.3}$$

where the R_i's are the residuals from a weighted median polish on these data, performed by Cressie and Read (1986).

Transforming the data is usually for the statistician's "convenience". It would be remiss of us not to interpret our results back on the original scale. Recall from Section 2, it is

$$1000p_i \equiv E(1000S_i/n_i);\ i = 1,\cdots,100, \tag{5.4}$$

that are the parameters of interest. In the context of spatial variation, it is naive and statistically inefficient to estimate p_i by S_i/n_i. Besides, what confidence can one put in (i.e. what is the standard error of) such an estimator?

To interpret the estimation of (5.4) in terms of the Y_i's, we see that

$$\begin{aligned} 1000p_i &= E(Y_i^2/4) \\ &= (1/4)\{var(Y_i) + (E(Y_i))^2\}. \end{aligned} \tag{5.5}$$

The quantities in (5.5) are each available from the model (5.1): to estimate p_i, simply substitute the maximum likelihood

estimates of $\underset{\sim}{\beta}$, τ^2, ϕ into the right hand side of (5.5), to yield a maximum likelihood estimator of $1000p_i$, the SID rate per thousand in county i. For example, consider county 80, and fix $k = 1$ (i.e. $h_{ij} \propto d_{ij}^{-1}$). We obtain,

$$\begin{aligned}
&\text{model I :} && 1000\hat{p}_{80} = 2.04 \ (\hat{se} = 0.11),\\
&\text{model II :} && 1000\hat{p}_{80} = 1.17 \ (\hat{se} = 0.21),\\
&\text{model III:} && 1000\hat{p}_{80} = 0.97 \ (\hat{se} = 0.24),\\
&\text{model I on } \underset{\sim}{R}\text{:} && 1000\hat{p}_{80} = 1.14 \ (\hat{se} = 0.06),
\end{aligned}$$

Standard error estimates ($\hat{se}$) are obtained from,

$$\hat{se} \simeq (1/2)\{\text{var}(\hat{\mu}_{80})(\hat{\mu}_{80})^{-2}\}^{1/2}. \qquad (5.6)$$

The naive estimator is $1000S_i/n_i = 0.65$; an even more naive binomial assumption yields $\hat{se} = 0.38$. The value of 0.65 is clearly too low.

Which is the most appropriate model for these data? Model I is clearly a gross over-simplification; to capture the clusters of low and high counties seen in the probability map of Cressie and Read (1986) with just one spatial dependence parameter is asking too much. Model I on the weighted-median-polish residuals is our choice although it is always comparable in its fitting with model II. The $\hat{\phi}$ in model III is small, but because of bias in the mle's it is probably too small; this negative bias gets worse as the number of large-scale variation parameters increases. The weighted median polish solves these bias problems, and still has the flexibility to fit two-way spatial

trend (large-scale variation) plus spatially correlated error (small-scale variation). Cressie and Read (1986) perform an exploratory investigation of the spatial nature of the weighted median polish residuals and show them to have no trend and spatial correlation up to thirty miles. To check the "trend" conclusion we ran model II on the R_i's and found their estimated means to be roughly the same; to check the "spatial correlation" conclusion, we made plots like Figure 6 for a number of the counties.

Although we have taken the Markov-random-field approach to building the spatial statistical models, there is another approach which constructs the spatial dependence simultaneously. Instead of (5.1), write

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\chi}, \tag{5.7}$$

where $(I-B)D^{-1/2}\underset{\sim}{\chi} \sim N(\underset{\sim}{0}, I\sigma^2)$; i.e. $\underset{\sim}{\chi} \sim N(\underset{\sim}{0}(I-B)^{-1}D(I-B')^{-1}\sigma^2)$, which is often called a spatial autoregressive process. Besag (1974) and Cliff and Ord (1981) discuss the differences between (5.7) and (5.1), and Ord (1975) shows how (5.7) can be fit to regional variables.

In conclusion, we have built a spatial statistical model which allows estimation of the sudden-infant-death rate of the counties of North Carolina, 1974-1978. M. Symons and D. Atkinson kindly sent us the 1979-1984 data, and we repeated our analysis. Broadly speaking, we saw the same pattern emerging although the spatial dependence tended not to be quite as strong.

References

Atkinson, D. (1978). Epidemiology of sudden infant death in North Carolina: Do cases tend to cluster? PHSB Studies, No. 16. Raleigh, North Carolina: N.C. Department of Human Resources, Division of Health Services, Public Health Statistics Branch.

Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). The Journal of the Royal Statistical Society Series B, 36, 192-236.

Besag, J.E. (1975). Statistical analysis of non-lattice data. The Statistician, 24, 179-195.

Besag, J.E. (1977). Efficiency of pseudo likelihood estimators for simple Gaussian fields. Biometrika, 64, 616-618.

Clayton, D., and Kaldor, J. (1984). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Preprint, Department of Community Health, Leicester Royal Infirmary.

Cliff, A.D., and Ord, J.K. (1981). Spatial Processes: Models and Applications. Pion: London.

Cressie, N. (1985). When are relative variograms useful in geostatistics? Journal of the International Association of Mathematical Geology, 17, 693-702.

Cressie, N. (1986). Kriging nonstationary data. Journal of the American Statistical Association, 81, 625-634.

Cressie, N. (1987). The Statistics of Spatial Data. Wiley: New York, in preparation.

Cressie, N., and Glonek, G.F. (1984). Median based covariogram estimators reduce bias. Statistics and Probability Letters, 2, 299-304.

Cressie, N., and Read, T.R.C. (1986). Spatial data analysis of regional counts. The American Statistican, submitted.

Fogerty, A.C., Ford, G.L., Willcox, M.E., and Clancy, S.L. (1984). Liver fatty acids and the sudden infant death syndrome. American Journal of Clinical Nutrition, 39, 201-208.

Goldberg, J., and Stein, R. (1978). Seasonal variations in SIDS. Lancet, 2, 107.

Kindermann, R., and Snell, J.L. (1980). Markov Random Fields and their Applications. American Mathematical Society: Providence, RI.

Mardia, K.V., and Marshall, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika, 71, 135-146.

Ord, J.K. (1975). Estimation methods for models of spatial interaction. Journal of the American Statistical Association, 70, 120-126.

Whittle, P. (1954). On stationary processes on the plane. Biometrika, 41, 434-449.

Table 1

Maximum likelihood estimates ($\hat{\phi}$) and 95% confidence intervals (ci) for spatial dependence parameter ϕ. Upper (ϕ_u) and lower (ϕ_ℓ) bounds on ϕ are determined so that $(I-\phi D^{-1/2}HD^{1/2})$ is positive definite.

	k		
	0	1	2
(ϕ_ℓ, ϕ_u)	(-0.328,0.190)	(-0.997,0.902)	(-0.999,0.998)
model I			
$\hat{\phi}$	0.173	0.833	0.596
ci	[0.13,0.18*]	[0.53,0.90*]	[-0.73,0.99*]
model II			
$\hat{\phi}$	0.0792	0.710	0.810
ci	[-0.06,0.17]	[-0.1,0.89]	[-0.22,0.99*]
model III			
$\hat{\phi}$	0.117	0.081	0.0336
ci	[-0.2,0.17]	[-0.98,0.89]	[-0.99*,0.99*]
model I on median-polish residuals			
$\hat{\phi}$	0.0811	0.620	0.405
ci	[-0.33,0.16]	[-0.11,0.89]	[-0.90,0.99*]

*These limits of the confidence interval are rounded inwards to avoid being outside the permissible range (ϕ_ℓ, ϕ_u).

Table 2

Estimates ($\hat{\underset{\sim}{\beta}}$) of $\underset{\sim}{\beta}$ in $E(\underset{\sim}{Y}) = X\underset{\sim}{\beta}$, for k = 1 in (4.9).

model I

$\hat{m} = 2.8378$

model II

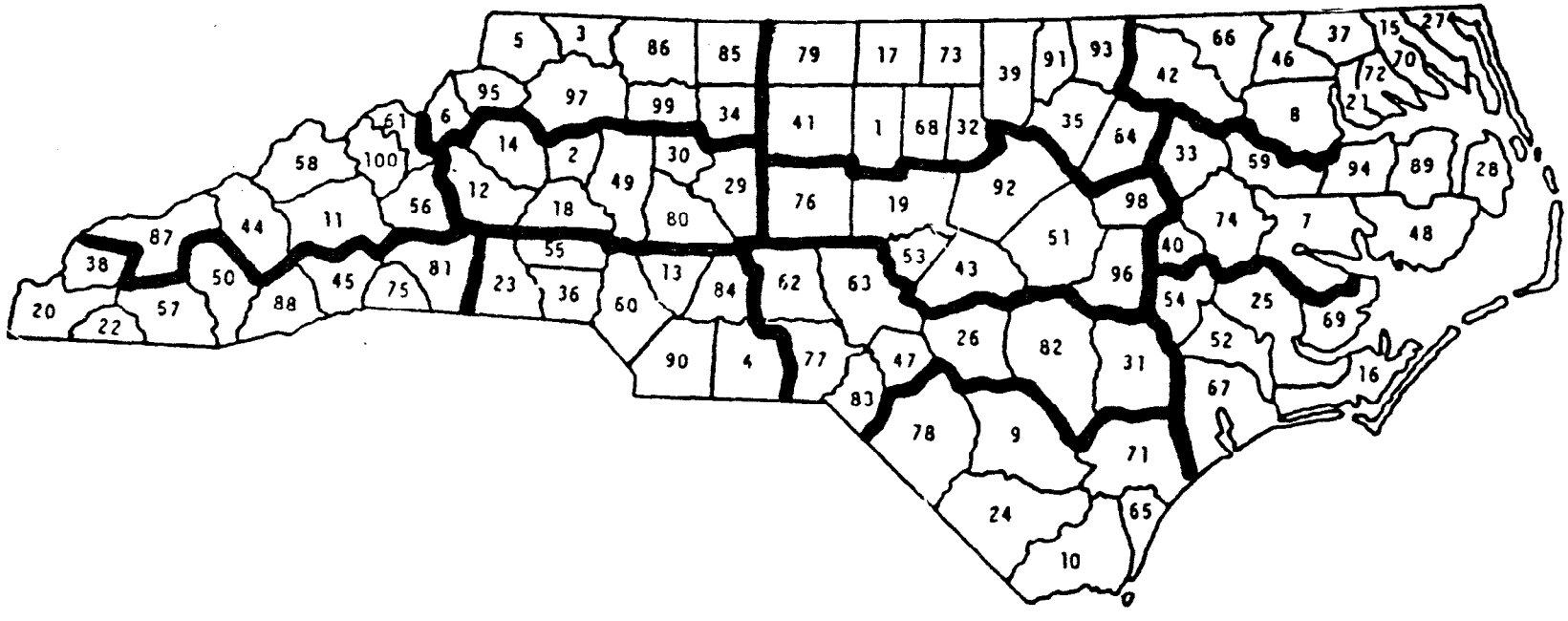
$\hat{m}_1 = 2.0559$	$\hat{m}_2 = 2.8704$	$\hat{m}_3 = 4.2564$
$\hat{m}_4 = 2.4778$	$\hat{m}_5 = 2.1505$	$\hat{m}_6 = 2.6388$
$\hat{m}_7 = 3.2788$	$\hat{m}_8 = 3.1126$	$\hat{m}_9 = 2.6793$
$\hat{m}_{10} = 2.8369$	$\hat{m}_{11} = 3.1786$	$\hat{m}_{12} = 3.6866$

model III

$\hat{m} = 2.8256$		
$\hat{r}_1 = 0.5332$	$\hat{r}_2 = -0.4364$	$\hat{r}_3 = -0.1554$
$\hat{r}_4 = -0.0551$	$\hat{r}_5 = 0.1798$	$\hat{r}_6 = -0.2261$
$\hat{r}_7 = 0.3464$	$\hat{r}_8 = 1.7900$	$\hat{r}_9 = -0.1751^*$
$\hat{c}_1 = 0.0994$	$\hat{c}_2 = -1.1723$	$\hat{c}_3 = -0.0449$
$\hat{c}_4 = -0.6183$	$\hat{c}_5 = -0.2885$	$\hat{c}_6 = -0.3209$
$\hat{c}_7 = 0.4522$	$\hat{c}_8 = -0.1693$	$\hat{c}_9 = -0.5585$
$\hat{c}_{10} = -0.2663$	$\hat{c}_{11} = -0.8086$	$\hat{c}_{12} = -0.4334$
$\hat{c}_{13} = 0.0699$	$\hat{c}_{14} = 0.7044$	$\hat{c}_{15} = 0.2574$
$\hat{c}_{16} = -0.3301$	$\hat{c}_{17} = -0.1000$	$\hat{c}_{18} = 0.4031$
$\hat{c}_{19} = 0.5548$	$\hat{c}_{20} = 0.3764$	$\hat{c}_{21} = 0.7087$
$\hat{c}_{22} = 0.2892$	$\hat{c}_{23} = -0.4697$	$\hat{c}_{24} = -1.2848^*$

*There are 32 independent parameters in model III; $\hat{r}_9$ and $\hat{c}_{24}$ are obtained from the linear constraints, $\sum_{i=1}^{9} q_i \hat{r}_i = 0$ and $\sum_{j=1}^{24} w_j \hat{c}_j = 0$, where $q_i = \sum_{i=u(k)} n_k$ and $w_j = \sum_{j=v(k)} n_k$.

FIGURE 1


5
3
86
85
79
17
73
39
91
93
66
46
37
15
27
95
97
99
34
41
1
68
32
35
64
42
72
70
21
8
61
6
14
2
30
49
29
76
19
92
98
33
59
94
89
28
58
100
56
12
18
80
51
74
7
48
87
44
11
55
40
38
50
45
75
81
23
36
60
13
84
62
63
53
43
96
54
25
69
20
22
57
88
90
4
77
47
26
82
31
52
16
83
78
9
71
67
24
65
10

FREEMAN-TUKEY TRANSFORMED SID VS ARCSINE(BLACK/100)

FIGURE 2

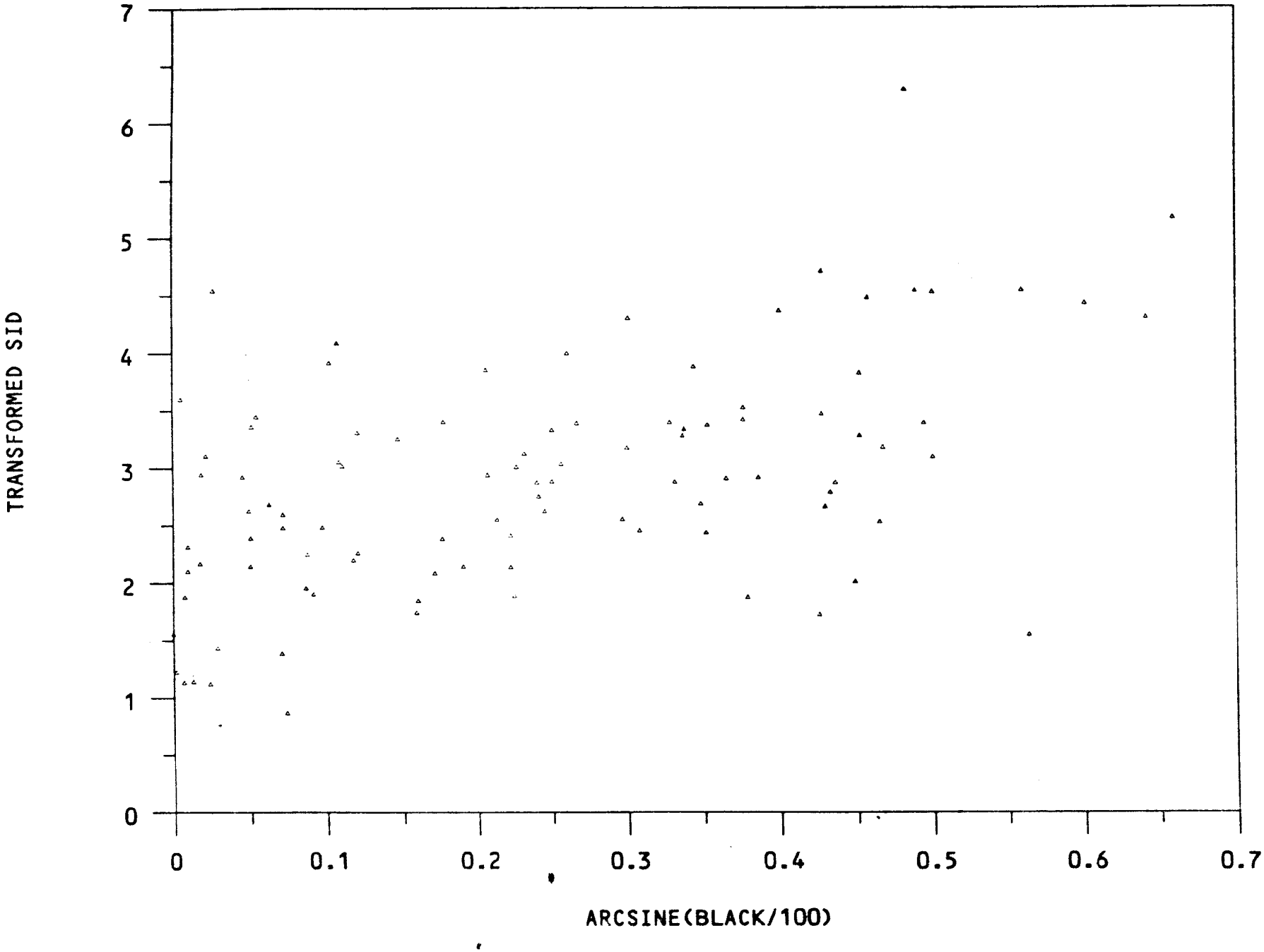


FIGURE 3

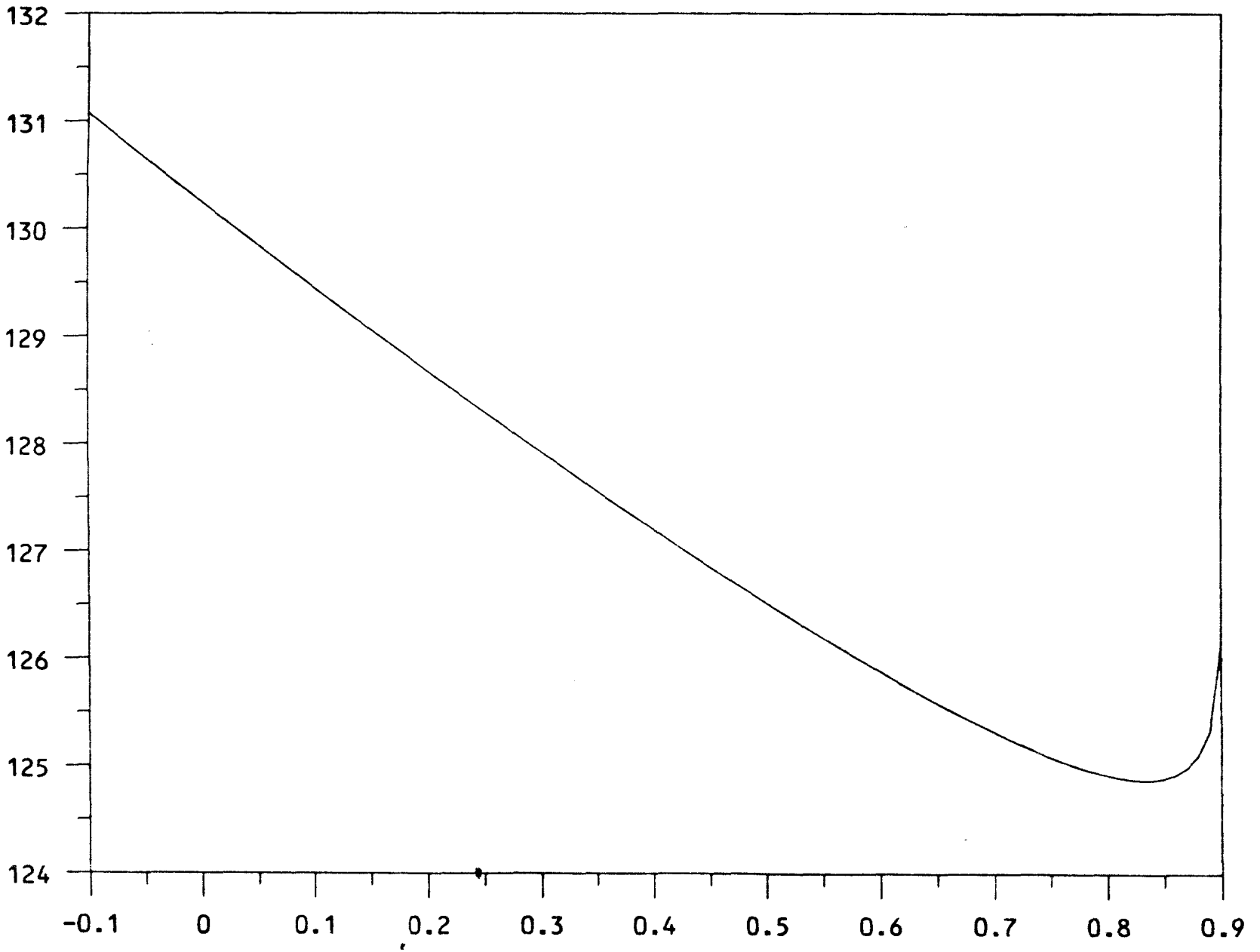


FIGURE 4

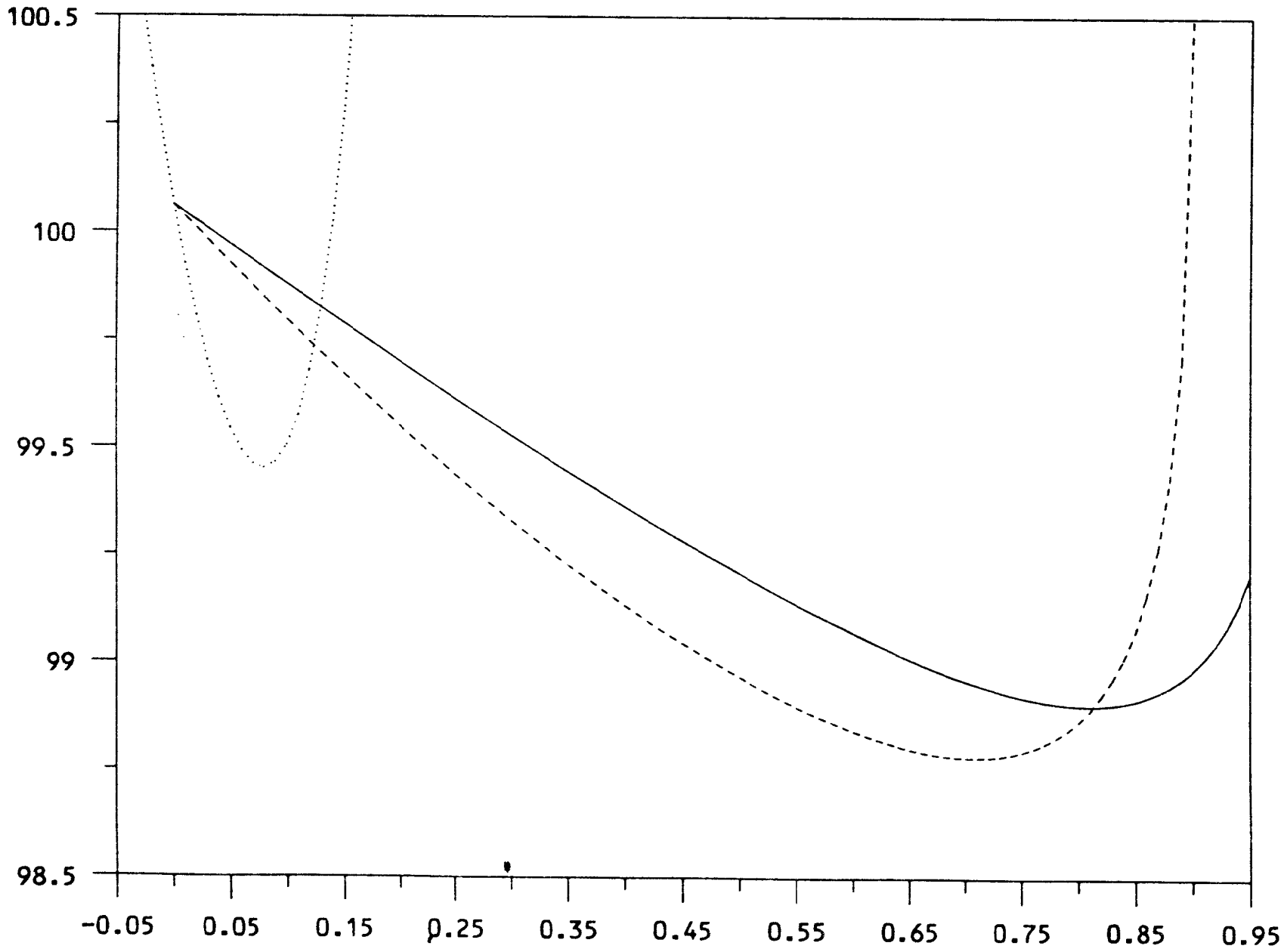


FIGURE 5

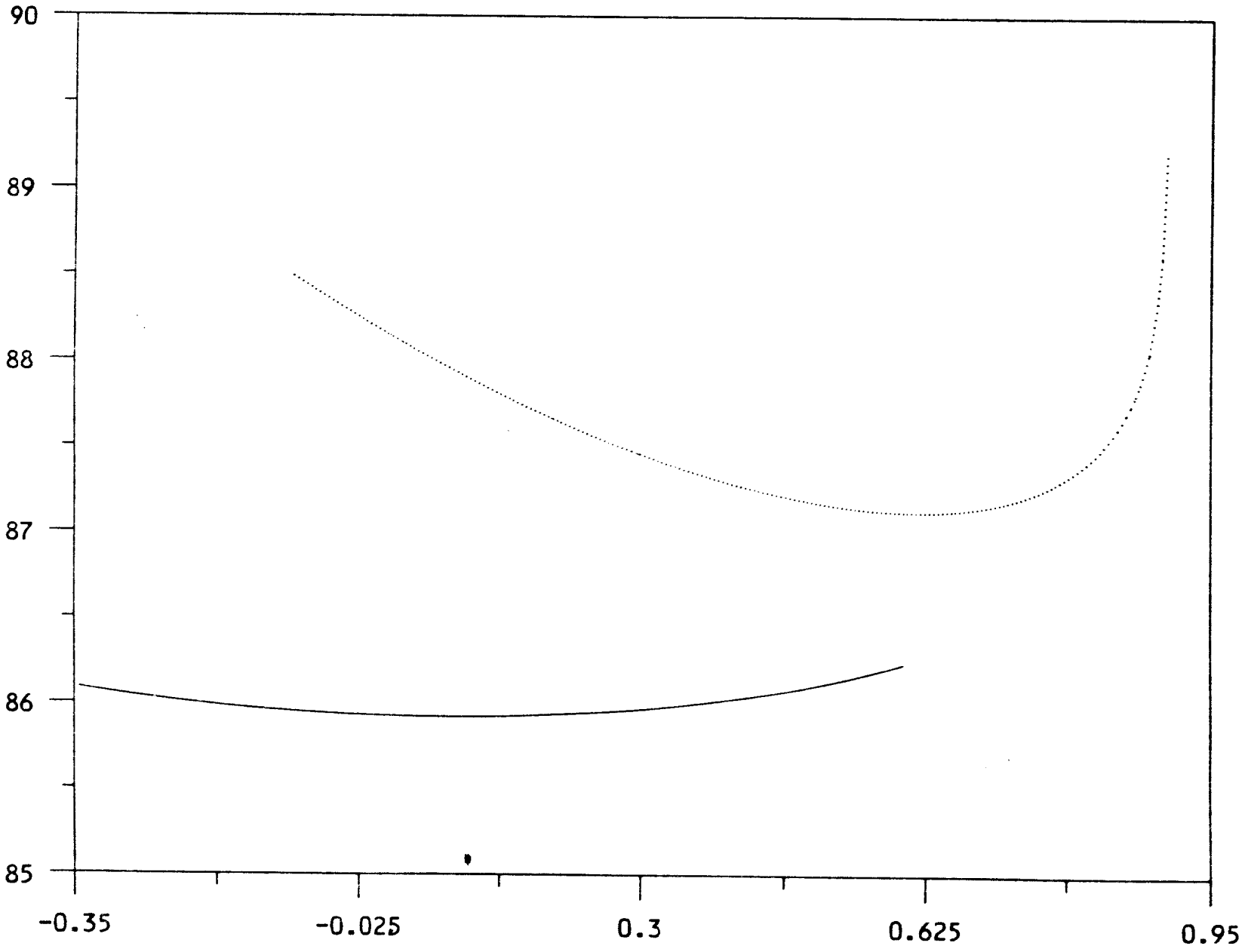


FIGURE 6

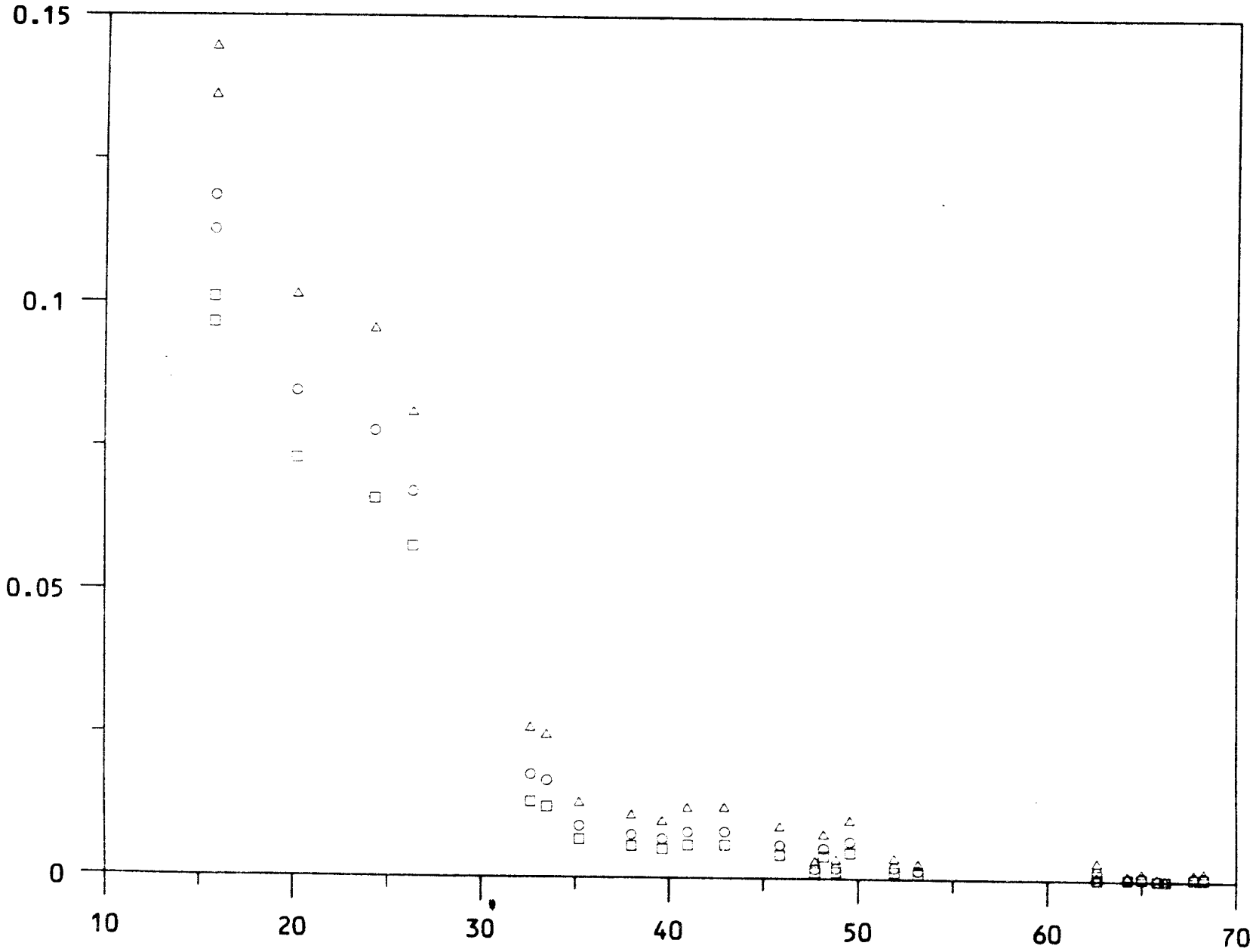


Figure Captions

Figure 1: Counties of North Carolina numbered alphabetically, featuring the 12 groups of counties used in defining model II.

Figure 2: Plot of Freeman-Tukey transformed SIDs versus arcsin(percent black) for the counties of North Carolina, 1974-1978.

Figure 3: Plot of likelihood as a function of spatial dependence parameter ϕ; auto Gaussian model I, and k = 1.

Figure 4: Same for Figure 3, except auto Gaussian model II is shown for k = 0 (dotted line), k = 1 (dashed line), k = 2 (solid line).

Figure 5: Same for Figure 3, except model III (solid line; scale on the left) is shown with model I on weighted-median-polish residuals (dotted line; scale on the right), for k = 1.

Figure 6: Plot of $\text{corr}(Y_{80}, Y_j)$ versus Euclidean distance $d_{80,j}$, as j varies.