THE INSTITUTE
OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

DEVELOPMENT OF AN UPPER ECHELON SUBMODEL FOR THE SOUTHERN PINE BEETLE HIERARCHY

BY

WILLIAM D. MAWBY

Biomathematics Series No. 5
Institute of Statistics Mimeo Series No. 1314

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

TABLE OF CONTENTS

CHAPTER I

BIOLOGY OF THE SOUTHERN PINE BEETLE

The southern pine beetle, Dendroctonus frontalis Zimm., is a persistent and effective pest of all southern pines. Ranging from Pennsylvania to the Gulf of Mexico and from eastern Texas to the Atlantic Ocean, this native bark beetle is characterized by periods of dramatic population outbreak. Such epidemic surges occur locally within the range nearly every year. On occasion, when forest and climatic effects are favorable, these local upswings can combine into large scale catastrophes affecting the entire Southeast. This effect appears to be periodic, with peaks separated by eight to ten years; this frequency is not very stable, however, with 1920-1924, 1929-1932, 1949-1963, and 1972-1975 being reported as epidemic periods (King, 1972). Separate areas are quite variable in their infestation patterns. During the most recent explosive period mentioned above, more than two billion board feet of lumber were damaged at an estimated cost of over one hundred seventy million dollars (Price and Doggett, 1978). There is also considerable injury to aesthetic and recreational aspects of the affected areas, some of which has been measured. Forestry and entomology researchers are currently pursuing active investigation of all components of this problem.

In North Carolina, the southern pine beetle undergoes population changes quite rapidly. In 1971 only six of the state's one hundred counties reported active infestations, while a year later, forty-one counties reported activity. In 1975, a maximum for recent years was reached when eight-one counties reported these spots. The number of spots rose from two hundred to two thousand and then to ten thousand over those same three years. Only eleven counties escaped infestations entirely in the eleven year period from 1967 to 1977. The damage from this eleven year period of statewide activity affected nearly four hundred million board feet of timber at an estimated cost of nearly thirty million dollars (price and Doggett, 1978).

The geography of North Carolina is typical of the southeastern United States as a whole. Three major terrains of mountain, piedmont, and coastal plain exist, with the predominant soil being moist utisol. Five species of pine cover the range, loblolly, *Pinus taeda* L., slash, *Pinus elliotti* Engelm., Virginia, *Pinus virginia* Mill., longleaf, *Pinus palustris* Mill., and shortleaf, *Pinus echinata* Mill. Loblolly and shoftleaf pines are the most common hosts for southern pine beetles. The proportion of land forested is approximately sixty percent. There are three major ways in which North Carolina differs from the lower Southeast. One is an eight to ten degree centigrade drop in average January temperature. The second is total precipitation, some 7-8 inches lower on the average than the lower regions. The third is possibly a large degree of heterogeneity, as reflected in patchiness, in the forest environment. An excellent source of such geographical and climatological data is A Forest Atlas of the South (Nelson and Zillgitt, 1969).

Dendroctonus frontalis is a member of the family Scolytidae, beetles which feed on the subcortical region of trees. These insects commonly exhibit behaviors which are difficult to model. Three aspects of this difficulty deserve particular attention; the southern pine beetle demonstrates great variety in its interactions with associates, with host trees, and with members of its own species.

Typically, the beetle carries at least two types of fungus, one being the blue-stain. These fungal species interact in a complicated and as yet incompletely understood manner. Their role in accelerating tree drying is potentially helpful for brood production if begun early enough but may be harmful if started too early. The second type of fungus is the mycangial fungus which plays a role in the nutrition of the bark beetle, but this is not well understood (Barras and Hodges, 1969). Predators include birds, mites, and other insects, with parasites being plentiful. The combined effect of these agents can be to regulate southern pine beetle growth (Moore, 1972). Competitors are primarily other bark beetles, including *Ips avulsus*, *Ips grandicollis*, *Ips calligraphus*, and several Cerambycids.

The sequence of attack and subsequent niche partitioning (Younan, 1979) can play an important role in southern pine beetle spot formation. The difficulty of distinguishing between trees killed by _Ips_ and trees killed by southern pine beetles through casual observation often obscures collected data, but to no great extent in this study. Although _Ips_ damage is commonly found in higher bole levels, such information cannot be gotten from techniques such as aerial survey or photography.

The host trees for southern pine beetle attack can be any species of southern pine. The susceptibility of a species seems to be related to oleoresin properties and site conditions as discussed in Hodges, et. al. (1977), in Fargo, et. al. (1979), and in Leuschner, et. al. (1976). The density of pine in a stand, particularly that of shortleaf pine, seems to be important in North Carolina. Usually, larger and older trees are attacked, and damaged trees seem highly susceptible. Stressed sites are also prime targets of the beetle (Belanger, et. al., 1977). The susceptibility of a tree is also related to the local pine beetle population level, and, under heavy epidemics, almost any pine tree may be attacked.

A third component of diversity in southern pine beetle behavior is the wide fluctuation in individual behavior. A density dependent mechanism is operative during attack and oviposition (Birch, 1978). The complex pheromone system has two attractive components and possibly one repulsive component. Finally, formation and growth of centralized infestations (spots) are heavily dependent upon the proper synchrony of emergence and pheromone production.

The mechanism of beetle attack upon the pines is not completely known, but certain aspects have been observed (Coster, et. al., 1977). First, a small group of females locates a host tree. Often this initial tree is weakened due to injury or water stress. The tree defenses move to block the beetles' entrance by producing resin to encase and eject them. Complete pitchouts in which all attacks are successfully repulsed have been observed. The success of the beetle is determined by tree vigor and species and the intensity of the initial attack. Upon entrance near middle bole, females emit frontalin, which, in combination with alpha-pinene from the host tree, stimulates further attacks above and below the initial holes. The heaviest attack densities are commonly observed 1-3 days after the first observed attack. As males are attracted, the compound verbenone is dispersed by air movement. This also stimulates more attacks but alters the sex ratio of these attacks, finally achieving an even sex distribution along the bole (Payne, et. al., 1978). Attacks continue along the bole until the supply of beetles wanes or the tree area is filled. The shape of the parent adult distribution is discussed in Fargo, et. al. (1979). The inhibition of further attacks seems to be guided by release of higher verbenone concentrations and initiation of endo-brevicomin release.

A nuptial chamber is constructed by the entering female. A male enters and mates with her, and a winding egg gallery is then bored through the phloem with eggs deposited in niches along the sides. The number of eggs in relation to gallery length is discussed in Foltz, et. al. (1976). The adults then feed and emerge from the host, with this re-emergence process playing a vital part of epidemic spot building processes (Coulson, et. al., 1979). These re-emergent adults may stay within the spot or undertake longer migratory flights. Meanwhile, the eggs hatch into voracious larvae that pass through four instars prior to pupation. Finally the brood adult appears and, after developing within the tree, eventually emerges. The generation time depends heavily upon temperature and moisture conditions, but thirty days in summer and fifty days in fall is typical (Kowal, 1960). Attacks in North Carolina occur predominantly in late summer and early fall. Mortality is highest in the free flying adult and larval stages usually. The difficulty of raising these unique insects under laboratory conditions has interfered with detailed understanding of their reproductive processes.

The death of a tree is effectively ensured upon entrance of the initial cohort of attackers and egg gallery initiation. The blue-stain fungus carried by the beetle soon begins growing in the tree tissue, and when the flow of water in the xylem is interrupted, tree drying begins. The progress of this drying is reflected in the crown as it turns from green to orange to brown. The rate of this fade is dependent upon temperature and season and can only provide a rough index to southern pine beetle infestation (Hain, et. al., 1979). The effect of this tree drying on brood development could be important in low level situations when beetle development processes might be delayed in the tree long enough to cause unfavorable brood conditions. The foraging of the beetles and their larvae does cause some tissue damage and girdling but can most commonly be regarded as a secondary effect. The final death of the tree as reflected in crown fade can occur months after the southern pine beetle leaves, a subject upon which Dziadzio (1978) has further details.

Upon emergence, the beetles must fly to another tree, but there seems to be at least two modes of behavior possible at this juncture. Little is known about the migratory mode of southern pine beetles, but new infestations are generally located a considerable distance from older ones. More information is available on the flight of beetles within a growing spot (Coster, 1979). The emerging adults may be attracted to nearby trees, and their pheromones may attract later emergents to facilitate the invasion of the new trees. If this overlapping of pheromone sources and emergents continues, the number of trees invaded may increase at an accelerated rate. This, in qualitative terms, is an epidemic spot. The size and damage of these actively expanding spots can be enormous, with hundreds of trees being involved over several consecutive years. Some areas seem particularly prone to large spots, while others show a positive correlation with smaller ones. The factors causing termination of spots is not well known, but several factors are involved. Tree spacing (Johnson and Coster, 1978), predators, and parasites can decelerate spot growth, and winter mortality can sometimes eliminate spots altogether (McClelland and Hain, 1979). Control procedures such as clear-cutting and pheromone baiting can alter the natural processes, but little is known of the consequences of such actions. Lastly, the natural boundaries of the stand may affect the initial and terminal growth of the spot.

The biology of these local interactions between the southern pine beetle and its hosts is generally well known compared to the processes of migration, spot initialization, and large scale epidemic behavior, which are much less understood. However, they are no less important. Migration must occur throughout much of the spot history but certainly is of greater importance in the initial and terminal phases. Many researchers feel that a background level of southern pine beetle exists throughout the forest, waiting for a proper attraction; yet the life of an aerial beetle is believed to be only 3-4 days (Coulson, 1979). The problem is to coordinate these two facts into a plausible migratory mechanism without assuming strctures that the data cannot support. This paper does not address the full scope of all these facets.

An even deeper biological question lies beyond the migration problem. This involves the characterization of wide-area and long-term behavior of the beetle within the southeastern United States. Maps of infestation patterns reveal that shortleaf pine distribution could be a major determining factor (Price and Doggett, 1978), but no causal relationship has been developed for this conjecture. Weather patterns and wind directions must play an important role, a topic which Coster, et. al., (1978) have discussed. The whole concept of stand susceptibility and indices to measure it is tied into this level of the biology. Daniels, et. al., (1979) and Lorio (1978) discuss this question more fully. Finally, the evolutionary ecology can be examined at this level, and interesting genetic strategies can be developed. Indeed, the practice of any long range control tactics must rest heavily upon answers from these unexplored territories of the southern pine beetle biology.

CHAPTER II

THE SOUTHERN PINE BEETLE HIERARCHY

Science is so successful primarily because it is pursued along a hierarchical course of action. The truth of this statement can be inferred from the commonplace workings of practical science. The scientist is presented with a complicated natural system whose behavior he must explain. So he first breaks the structure of the system into smaller parts, often on the basis of previous experience with similar systems. Each part obtained from the compartmentalization is handed over to a specialist whose task it is to explain that subsystem's behavior. This may still be too hard, and further partitioning may occur. Finally, each piece of the original structure becomes explainable to some degree, usually through the power of mathematics. These submodels are reconnected along the lines of fracture caused by the divisive process with the hope that the patchwork structure will be nearly equivalent in some sense to the original system. The final model is very much a hierarchical model, perhaps not chosen in the best way, but hierarchical nevertheless.

In biological science this hierarchical approach is quite appropriate, perhaps due to an underlying similar structure of living systems themselves (Pattee, 1973). Universal acceptance of this inherent biological hierarchy is not to be found, however, and the rationale of modeling non-hierarchical systems through hierarchical models is under hot debate. However, although many scientists would argue over the reality of hierarchies in nature, it is probably correct to say that few would argue over their utility in past and present scientific inquiry. This paper, in particular, will accept this demonstrated usefulness and make heavy use of the hierarchical structure in what follows.

Mathematical study of hierarchical systems is still at a fundamental stage. Mesarovic, et. al. (1970) give an excellent summary of the more advanced work in this area and form much of the basis for the formal structure presented later. In addition, this fine book suggests that two common features must be considered requisite for any hierarchical order of structures. First, there commonly exists an order of magnitude difference in scale be-between the levels of the hierarchy. Scale may involve size or frequency or importance; the vital concept is that direct comparison of variables on different tiers is considered improper and unimportant. Interaction between levels is allowed in a formal hierarchy if the mechanisms treat the levels as complete entities. Secondly, the constituent units on any particular level in the hierarchy are defined wholly by the interaction processes contained within that level. Conceptually, the within-level bonds between variables must be far stronger than between-level bonds.

The value of creating a partitioned structure to represent a system is multiple, with gains both in control and correction. Data is often limited to subsystems, and previous work is often restricted similarly. If an erroneous submodel has been established, the ability of the hierarchy to isolate its effect makes correction much simpler and cleaner. The explicit modeling of between-level interactions initially saves the researcher time and money when it is time to put the pieces back together again. Indeed, the subsystems do not always fit into any type fo concerted whole special care should be taken to ensure that they do so before submodel structure becomes very detailed. Lastly, hierarchies can be especially well suited for control developments and experiments. Decision hierarchies are presently a major spur in modern hierarchical research.

A particular hierarchical model's validity is closely dependent upon the dual concepts of model adequacy and model purpose. The adequacy criteria state how one detects errors in output. Model purpose reflects how one defines output. The duality of these actions is evident in practice where gross output inadequacies are usually the only method of discovering model-purpose errors. Nothing is ever simple in natural systems, and the problem of validity is a vital consideration at all stages of modeling. The ability to balance tractability with realism is the true "art" of scientific modeling.

Formally, a system may be represented as a relation of abstract sets. Other authors such as Padulo and Arbib (1974) give much more detailed structure, but this definition will satisfy the present needs. A mathematical relation is a subset of the Cartesian product of two sets. One may call the two sets $\underline{I}$ and $\underline{O}$ and the relational subset $\underline{R}$. Symbolically, this may be written as

$$R \subseteq I \times O. \tag{1}$$

The abstract sets $\underline{I}$ and $\underline{O}$ are sets of mathematical objects, possibly with some internal structure such as, say, the group properties. For this paper's elementary use of these concepts, $\underline{I}$ may be thought of as a collection of inputs and $\underline{O}$ as a collection of outputs. The system connecting the two sets, that is, receiving the input and producing the output, is known only by the subset of specific input-output pairs it produces. These pairs are precisely what $\underline{R}$ defines in (1).

Further restriction of the structures of $\underline{R}$, $\underline{I}$, and $\underline{O}$ may occur if needs demand it. In particular, it may be assumed that $\underline{R}$ is a specific type of relation, a functional relation. If $\underline{R}$ is such a functional relation, it may be written as

$$R: \; I \longrightarrow O. \tag{2}$$

The definition of function and examples of use may be found in any calculus book, for example, Sagan (1974).

For purposes of hierarchical modeling, the system must be stratified into subsystems or levels. In terms of the representation in (2), this partitioning requires that I and O be expressible as

$$I = I_1 \times I_2 \times \ldots \times I_n \tag{3}$$

$$O = O_1 \times O_2 \times \ldots \times O_n \tag{4}$$

In terms of levels, then R is consequently representable as

$$R_i: \quad I_i \times U_i \dashrightarrow O_i, \text{ for } i=n \tag{5}$$

$$R_i: \quad I_i \times U_i \times D_i \dashrightarrow O_i, \text{ for } 1< \ i < \ n \tag{6}$$

$$R_i: \quad I_i \times D_i \dashrightarrow O_i, \text{ for } i=1 \tag{7}$$

In this formulation, U is the set of upward connections between levels i-1 and i, and D is the set of downward connections between levels i and i-1. Formally, the U and D variables are defined by mappings h and k in the following manner:

$$h_i: \quad O_i \dashrightarrow U_{i+1} \tag{8}$$

$$k_i: \quad O_i \dashrightarrow D_{i-1} \tag{9}$$

Upward connections are commonly information flows, and downward connections are control processes. Particular applications may demand different names, however.

The model developed in this paper simplifies this specification by adopting a particularly elementary structure, the hierarchy of abstraction. Subsetting is the key to this structure, with each level dealing with larger order combinations of fundamental units in space and time. The h and k functions are considerably simplified by this arrangement. These new functions are:

$$h_i: \ O_i \dashrightarrow U_{i+1}, \text{ for } U_{i+1} \subseteq O_i \tag{10}$$

$$k_i: \ O_i \dashrightarrow D_{i-1}, \text{ for } D_{i-1} \subseteq O_i \tag{11}$$

In the model building phase the k functions are often ignored and reinstituted only at a later stage.

The recently proposed southern pine beetle hierarchical framework (Gold, et. al., 1980) specifies four levels, the tree, the spot, the patch, and the region. Because of the simple type of hierarchy employed in this model, the four levels are defined by three inter-level transitions. The first level dynamics, which would represent the creation of trees, are concerned with stand growth models and will not be considered as part of the present model structure.

Because of the connected nature of the hierarchical model, it is necessary to outline the entire model, including all three transitions, in order to prepare the way for discussion of the upper echelon submodel which is the subject of this thesis. Each level will be characterized as to its inputs, outputs, and level-to-level information flows, but specific details will not be discussed except with respect to the upper echelon.

The basic element of the entire hierarchy is the tree abstracted as a location in space, characterized by certain tree and beetle properties. Within-tree development of the beetle is assumed relatively well known, and most of the emphasis of the model is on between-tree relationships (an epidemiological approach is adopted). The basic qualities of a tree are contained in the number of beetles it admits and the number it emits. The observation that contiguous groups of infested trees can display organized behavior is the basis for the intuitive notion of a spot. The further observation that such organization seems to extend only a short distance around the infested area allows the modeler the luxury of watching only a small number of trees, namely those within some distance of the tree group. Such a neighborhood is a better definition of <u>spot</u>, and this is the one used throughout this thesis. These spots behave like islands of infestation in the "sea" of the forest, and it is the role of the tree-to-spot transition function to classify and describe these islands as they change through time. The classification can be detailed or meager. Whatever the framework, the definitions must be concise and consistent if they are based in the state spaces which describe the spot development through time. These changes, or transitions, from one state to another are influenced by the input weather and biotic variables. The output is identical with the state changes of the system. The connection of this lowest level upwards to the spot-to-patch level consists entirely of providing a classification of the current spots of the system. It is on such information that the dynamics of the second level are built.

The spot structures defined on the lower level of the hierarchy are the raw elements upon which the spot-to-patch transition dynamics work. Just as on the tree level, areas of forest (groups of contiguous trees) have two important properties with respect to the beetle system; these are susceptibility and infectiousness. Spots with emigrating beetles supply infectious agents to susceptible areas of the forest through migration. The length of these migrations is believed to be rather short and sufficiently directed so that susceptible areas have a high probability of being found by the beetles. Susceptibility implies a stressful condition due to either meteorological or biotic components which may occur in a section of forest not covered by the tree-to-spot transition. These new entities are called <u>potential</u> <u>spots</u> and are peculiar to the middle level of the hierarchy. Because of the short range assumption on the migration ability of the beetles and due to the natural heterogeneity of forests, a natural partitioning of the region into patches may occur. A <u>patch</u> is a section of forest which may be described by a simple parametric structure within its interior but which may differ from the parametric structure of any different patch. The concept of patch is left rather undefined with respect to size purposely in order to make the overall modeling effort more effective. For the purpose of this paper, the size of the patch is conveniently specified by the region size and thus appears large. It is important to note that concepts of patch and the more common forestry term stand are not synonymous. The stage of the model development is still too preliminary to be sure that the homogeneous area of forest called a stand will be of any significance to the beetle dynamics. On the other hand, the patch is by definition important to those dynamics.

The input functions include weather and biotic factors which determine how the spots within a patch change in their distributional properties over time. The output of the submodel is again a partitioning of the state space of patches into types. These types are almost certain to include concepts such as epidemic and endemic populations. The names do not matter, but the definitions must be consistent when based on the state space structure. The characterization of the patches by their epidemiological type is the raw information which is passed up to the patch-to-region transition. The information passed down to the lower tier consists of spot initilizations which result from potential spots being infested by migrating beetles. This initilization process cannot be handled on any other level but the middle one.

Now the last level of the hierarchy may be described as in terms of relationships between patches of differing epidemiological type. Beetle movement on this highest level is assumed to be a diffusive process. It reflects the influence of one patch's intensity upon the intensity of nearby patches and is assumed to be smoothly flowing and rather local in effect. Furthermore, the rate of its movement depends only upon the intensity of the source of this motion. A prototype model for this submodel is an irregular mosaic of patches. The manner in which the whole mosaic acts is determined completely by the connection properties of the mosaic and by the fluctuations of the internal intensities over time. The distribution of this flowing intensity over the region yields consistent definitions of the concepts epidemic region or endemic region. The changes in the connectivity properties of the mosaic are effected by the input of weather variables, and the output consists of the new state of the region. Downward flows of information consist of a general background level of infestation intensity which influences the probability of migration from spot to potential spot.

While it is apparent that many details need to be specified before the overall framework of the model can be studied in depth, such a model structure seems consistent with the observed facts while still retaining potential for adequately treating the complexity of the beetle system. This paper serves two purposes with respect to the overall hierarchical model. Firstly, it delineates the basic characteristics of the upper echelon level in a manner which makes the results immediately applicable, taking account of the important interlevel relationships. Secondly, it develops the methodological approach, which will be used at other levels as modeling efforts continue. Especially important in this respect is the use of statistical space-time techniques which should prove useful in all phases of the research effort. A general discussion of these methods is essential to understanding the submodel construction process. This is the topic of the next chapter.

CHAPTER III

EXPLORATORY SPACE-TIME TECHNIQUES

In biological research, space-time patterns are often the only clues that exist as to the internal structure of the system under study. While planned experiments may be possible on some levels, such manipulative studies on ecosystems are difficult to design, dangerous to perform, and often not reproducible. Consequently, such techniques should be, and generally are, approached with great caution, if at all. This inability to freely manipulate the chosen system is limiting, but it is generally not prohibitive. Rather, the researcher must let nature run the experiments for him and rely upon less traditional modes of analysis to abstract results from these behaviors.

The model building process typically utilizes three major steps: exploratory analysis, model translation, and model confirmation (Getis and Boots, 1975). For any complex, i. e. real, system, this process will be iterated and reiterated many times before researcher satisfaction is achieved. Each step serves a fundamental role in the process, but due to the ordered nature of the steps, the exploratory stage takes a primordial position. In this chapter, the emphasis will be placed upon this exploratory step for four reasons:

1. it is less familiar to researchers
2. it will be used in the next two chap-

ters

3. its method should be widely applicable
4. it is interesting in its own right.

Tukey (1977) compares the confirmatory mode of statistics with the exploratory mode by analogy to the judicial and investigative branches of a criminal justice organization. Just as no court case would come to trial before extensive preliminary inquiry, the testing of hypotheses must also be preceded by the discovery of just what hypotheses to consider. This elucidation of relevant hypotheses, or models, is the only first step possible in studying obscure systems.

The exploratory statistical analysis of space-time patterns can be separated into three branches: nonparametric, multivariate, and time-space series techniques. Although theory overlaps in these branches, especially between multivariate methods and the other two types, practice does not. Indeed, most applications deal with only one particular technique or a few closely related ones. The use of several widely different methods on the same problem should be encouraged, however, because of the cross-validation that can result. It is on the basis of their common use in literature, their common mathematical techniques, and their consistencies in applications that the distinction between methods has been drawn.

Nonparametric statistics deals with procedures that give stable results under widely varying assumptions about the underlying populations. This stability is reflected in the distribution-free properties of the statistics. Hollander and Wolfe (1973) argue that nonparametric methods appeal for several reasons:

1. they require less restrictive assumptions
2. they are easy to perform and understand
3. they commonly use ranks, not magnitudes
4. they remain efficient under normality.

Properties one and four are highly desirable for exploratory methods. Property two can be misleading because, although the calculations are simple, the sheer number of calculations can often be prohibitive. Larger samples are almost universally handled through large sample, i. e. asymptotically normal, approximations due to this difficulty.

The fundamental objective of most nonparametric space-time analyses thus far performed is presented by Klauber (1974) as the elucidation of methods which will detect clustering of events as opposed to chance variation but which will be insensitive to clumping in either coordinate frame alone. The typical procedure involves first defining a statistic which acts upon pairs of space-time points.

The sampling distribution of the statistic is then calculated, or estimated, under very loose assumptions, and this calculated distribution forms the basis for future significance testing. The calculation of this theoretical distribution is attempted through U-statistic theory (Randles and Wolfe, 1979), through graph theoretical techniques (Barton and David, 1966), or through simulation (Siemiatycki and McDonald, 1972). Most applications in the literature are epidemiological in focus.

Three particular approaches are of special interest to the study of southern pine beetle dynamics. The use of <u>shape indices</u> is described in Haggett and Chorley (1970) and in Bookstein (1978). Use of <u>join statistics</u> in the study of n-phase mosaics is discussed in Cliff, et. al. (1975) and Pielou (1977). Lastly, a generalized regression approach has been developed in Mantel (1967). Each of these topics will be discussed in greater detail below, along with some indication of their applicability to the beetle modeling problem.

Bookstein's concepts are heavily geometrical in approach. An <u>abstract shape</u> is defined as an equivalence class under the group of similarity transformations. A function to the real line which gives equal results for all members of each equivalence class is a <u>measurement</u> of that shape. For smooth continuous closed shapes, one measurement regime extracts the tangent angle and the arc length at fixed points of the outline for every sample shape. The fixed points are geometrical landmarks and are used to identify particular classes of shapes. The observations obtained from this procedure are then analyzed with multivariate methods to isolate principal or covariant features. The essential element is the similarity of shapes as summarized in the measurements. Although the emphasis of Bookstein's development is morphological, applications in ecology should be realizable. In particular, studies of spot growth and infestation pattern growth should benefit greatly from this approach, but no work has yet appeared.

The basic concept lying behind the mosaic and nearest neighbor statistics is that of link distance measurement. Given a cell within a mosaic and a classification on all the cells of the mosaic, the probability that the nearest similarly classified cell is *r* steps away is the most important feature of interest. Further conditions on the mosaic such as regularity or perfect randomness lead to specifically useful results. The essential point is the introduction of mosaic structure into the measurements.

Mantel's generalized regression has the form

$$Z = \Sigma S_{i,j} * T_{i,j} \tag{1}$$

with $S_{i,j}$ being a space measurement and $T_{i,j}$ a time measurement for the ith and jth observations, and * is a binary relation. For a one-sample situation, the Z-statistic has been shown to behave as a certain fitted Pearsonian distribution by Siemiatycki (1974). Klauber (1974) examines the multiple-sample approach and concludes that the Z-statistic is satisfactory for detecting typical epidemiological space-time clustering under most conditions. A standardized formulation of Mantel's approach is interpretable as an estimate of contagion strength. In general, results are still sketchy, however, and further work must be done before this approach becomes implementable.

Multivariate techniques, the second branch of general exploratory methods, deal with dependent variables and the individual entities upon which the measurements occur (Kendall, 1975). Kendall further remarks that there exist at least four reasons for pursuing such methods:

1. to reduce complexity
2. to group individuals
3. to group variables
4. to characterize the dependencies

Problems in applying classical multivariate analysis include the difficulty of justifying assumptions, the immensity of calculations, and the uncertain interpretation of results. Their great power in space-time research lies in their ability to compare patterns rather than points.

There are three central modes of multivariate analysis: clustering, factor analysis, and discrimination. Each mode has specific advantages and disadvantages for space-time usage. Each type will be discussed briefly below, with factor analysis being emphasized more than the others. A specific form of factor analysis will be an important feature in developments throughout the remainder of this discussion, and this heavy use accounts for the extra interest in those techniques.

The comparative study of objects in the space of their attributes solely on the basis of pair functions between them (Sokal, 1977) is termed clustering analysis. Although the researcher intends to discover inherent relationships between the objects using clustering methods, the results are often highly artificial. Two features of the clustering methods must be specified independently of the problem characteristics: the measure of similarity and rules for comparing similarities. One example of application would use correlation as a similarity measure and would combine highly correlated individuals in an agglomerative procedure. Although much current research is being done in this area, the application of any one method must be made with caution.

Another type of analysis called discrimination analysis is based on the assumption of several populations of individuals and the presence of random samples from each of them. The construction of a decision rule for classifying further individuals with a minimum of errors is the intent of the discrimination analysis. This process is based heavily upon the particular likelihood function of the observations involved. Economic oriented analyses of biological problems, including southern pine beetle problems, often use this approach.

In factor analysis the set of data is regarded as being explainable in terms of common factors. One type of factor analysis, principal components analysis, is particularly popular in ecological (Pielou, 1977) and morphological (Blackith and Reyment, 1971) research. This technique seeks to simplify the dependencies in a set of data by choosing a new coordinate system, each axis of which is a linear combination of the old axes (Klovan, 1975). Through this restructuring, either the variables may be transformed (R-mode), or the individuals may be transformed (Q-mode). The new coordinates are chosen to be orthogonal, but oblique rotation methods often negate this convention. The primary difficulty with this decomposition is the arbitrary nature of the resulting components, which are generally uninterpretable in all but obvious cases.

A specific form of principal components, reference curve analysis, which has proven useful in the construction of the southern pine beetle hierarchical model is given in Sheth (1969). Beginning with a data matrix of N observations (times) and M subjects (areas), the raw N x N cross products matrix is partitioned into the product of two matrices via eigenvalue decomposition. The resulting matrices are then standardized to negate the effect of sample size. The standardized matrices contain the individual parameters (loadings) and the reference curves (scores), respectively. The formation of these reference curves reflects the similarity in temporal pattern between areas. Dominant curves (those with heavy eigenvalues) reflect strongly consistent behavioral modes. Certain hypotheses, which treat area behaviors as linear combinations of standard behaviors, can be treated through axis rotation methods.

The third branch of exploratory space-time methods deals with time series and its extensions. The primary requirement of this approach is an ordered sequence of observations which behaves in a nice, i. e. stationary or intrinsic, manner, at least approximately. Included in this nice behavior is the existence of an autocovariance function which depends only upon the time between observations. Analysis may be directed towards the correlogram (the graph of the autocorrelation function versus distance), towards the spectrum (the Fourier equivalent of the correlogram), or towards both.

In practice, there are several difficulties in applying time series analysis. First, the observations must be identically space, although some work has been done on other cases (Rees, 1970). Secondly, the sample size, i. e. the numbers of times involved, should be as large as possible; often a starting figure of fifty is recommended. This requirement is designed to ensure precision (a sample of size one can be extremely accurate) and to ensure that high frequency effects are not overlooked. A major difficulty with time series analysis is the inability to detect a high frequency if the observations are too widely spaced. The researcher may use these methods freely but must be aware of the difficulties and limitations involved.

A raw time series is usually conceived as consisting of four components (Kendall, 1973):

1. a determinate trend
2. regular fluctuations about the trend
3. seasonal or periodic effects
4. a random effect.

The isolation of the first three components is usually necessary before the random effect, i. e. the true stochastic process, can be studied in full. Often the specification of trends alone is a primary objective of the investigation. The existence of seasonal effects is often unexpected before performing the analysis but is important from a modeling standpoint. The remaining random effect is often assumed to be of an autoregressive nature (or equivalently, of a weighted average) of some order.

The estimation of this autoregressive structure, especially its order, is vital to the modeler.

The definition of a time series does not require that the spaced observations occur along the time axis, and extensions to spatial series have been made. Cliff, et. al. (1975) discuss the basic problems of defining spatial autocorrelation by examining a mosaic of cells. Each cell is defined by its connectance properties and the observed values of some internal variable. The covariance or correlation function must relate a particular cell with its first order, second order, etc., neighbors. In time series there is a unique choice for these neighbors, but in space series, the number of nth order neighbors will often be greater than one. The estimated autocorrelation is therefore an average of pairwise relations summed over all nth order neighbors and over all cells. The choice of representing these relations is arbitrary as is the definition of spatial lags between cells.

The spatial lag concept offers both interesting and disturbing features for the researcher who attempts to use it. The arbitrary nature of connections gives them the properties of weightings on the pairwise relationships. By varying the weights according to some private hypotheses, certain relationships can be emphasized at the expense of others. These weighting schemes can easily be expressed as a product of a connectivity matrix and a weighting matrix, and even the most general structures can be concisely represented in this way. The calculation of nth order lags can then be formulated as powers of the connectivity matrix with non-simple routes eliminated. In general, the elimination of these routes with repeated nodes is complex, and neither the explicit method discussed in Ross and Harary (1952) nor the iterative method discussed in Marshall (1971) seems useful for graphs of diameter greater than seven.

The next stage in time series extensions is to space-time series, but in this area the development is still elementary. Curry (1970) develops an univariate approach but continues no further. Cliff, et. al. (1975) also use an univariate approach based on weighted exponentials. More recent research by Matheron (1975) and others seems to be straying from the traditional ideas and centers on random function theory (Matheron, 1970) instead. At present, a typical analysis would begin with estimation of trends, followed by an examination of the residuals for autocorrelative patterns. Once the lag structure has been specified, estimation of a specific univariate model would be attempted using a simple scheme as described by Cliff, et. al. (1975). A mixture of typical univariate behaviors would then be created in order to describe an entire region.

This paper uses the reference curve technique and the time and space series technique in its development of the upper echelon submodel for the southern pine beetle hierarchical model, although the methods outlined above have general applicability for all areas of ecological research. Especially when used in concert, these exploratory methods could prove invaluable to a researcher faced with a system with which he is unfamiliar. Some of the newer methods, such as the random function approach and shape indices, could be vital components of future ecological modeling work.

CHAPTER IV

MULTIVARIATE ANALYSIS OF THE DATA

Two modes of time-space statistical analysis have been used in constructing an upper echelon submodel for the southern pine beetle hierarchical model. Both techniques (reference curve analysis and space-time series analysis) have been applied to the same set of data. The two analyses are complementary in a sense, in that each looks at the time-space structure of a region on a different level. The multivariate techniques, with their emphasis on the full covariant relationship, tend to exhibit the global structure of the system. The autocorrelative methods utilize individual blocks to build up the regional behavior. These latter techniques may therefore be thought of as local in focus. By hypothesis, these two approaches should be consistent in producing a description of the observed regional dynamics. Thus, the apparent duplication of effort in the two analyses really serves several useful purposes:

1. an illustration of use of the methods
2. a cross-validation procedure
3. a measure of appropriateness of the methods.

The cross-validation property stems from obtaining consistent results after applying dissimilar analytic processes as discussed above. The discussion of the analysis will be separated into two chapters to emphasize this purposeful segregation of investigations; this chapter will deal solely with the multivariate method, while chapter five treats the time-space series analysis.

As described above, this chapter details the exploratory analysis of the data through use of reference curve analysis. The explanation of this analysis process will be presented in four logically consecutive parts:

1. data preparation
2. reference curve analysis
3. blocking patterns
4. secondary analyses.

Part one will appear in the discussion of chapter five as well, but only the present chapter will treat it in full detail. The other parts are peculiar to the multivariate analysis and will only be discussed in this chapter.

All data used in the analysis is based upon the information contained in two sources: _A History of Southern Pine Beetle Outbreaks in the Southeastern United States_ (Price and Doggett, 1978) and _A Forest Atlas of the South_ (Nelson and Zillgitt, 1969). Of these two sets of information, the most important one for the discussion is the Price and Doggett work.

In the Price and Doggett publication, prepared for the Southern Forest Insect Working Group, there are a series of eighteen maps (representing eighteen consecutive years, 1960-1977) of the thirteen southeastern states which make up the typical range of the southern pine beetle. Each state is divided into counties, and some of these counties are shaded to signify the presence of active southern pine beetle infestations within their confines. The shading is binary and ignores all effects of the size and number of spots creating the observed infestation. Theoretically, the masking of the data could cause severe problems in the analysis if the intrinsic and extrinsic measures of intensity are not consistently and simply related. The question of how well the intrinsic measure of infestation intensity compares with the extrinsic measure is not completely answerable at the present time. However, it does not seem unreasonable to assert that epidemic situations are characterized by an increase in size of spots as well as in number of spots. This assumption is at least partially substantiated by the aerial photography work of Hain and DeMars (1979) on declining epidemic beetle populations. Blocking of the data, such as is done in this paper, should also reduce the probability of such erratic behavior. The data may be misleading for several reasons:

1. misclassification of insect agent (<u>Ips</u>)
2. increased surveillance under epidemics

3. precision of spot counting techniques

4. non-homogeneous county sizes.

Blocking should mitigate errors from all of these causes except possibly item two, the increased surveillance problem. This effect results basically from the reduction of between block effects as compared to the more variable natural divisions (the block boundaries smooth the natural discontinuous boundaries). By losing the exactness of the data the misinformation is averaged in with the larger mass of reliable data, and its effects are hidden in the mean effect (assuming a proper random distribution of errors over the region). Error type one may introduce a bias if the Ips-southern pine beetle relationship is not independent (which is a current topic of debate). This suspected relationship, even if it exists, is thought to have its greatest effect at low beetle population levels. It will be assumed in this paper that such errors are negligible.

This original set of data contains approximately seven hundred counties classified as to their southern pine beetle infestation properties for eighteen consecutive years, 1960-1977. In terms of a matrix of seven hundred columns by eighteen rows, this would mean analyzing nearly twelve thousand five hundred numbers. Because of the impracticality of handling this large amount of data, the data were grouped into blocks.

This blocking consisted of drawing a grid of vertical and horizontal axes over the major portion of the maps. On this initial grid, thirty-five squares of approximately one inch by one inch (one hundred eighty miles by one hundred eighty miles) cover the range of infestation patterns given in the eighteen maps. Only twenty-four of these blocks actually contained observed infestations, and thus only these twenty-four areas were used in the majority of analysis procedures (Figure 4.1). However, in some analyses more detail was needed, so a finer partitioning was used, consisting of 96 blocks, 90 miles square. Within each block, the measure of the incidence intensity was defined to be the proportion of infested areas within the block boundaries (the proportion of area covered by the infested counties). A grid cell measure was used for this purpose, resulting in units of approximately one-eighth inch by one-eighth inch (five hundred six square miles). The number of infested grid cells divided by the total number of grid cells within the area provided the estimate of infested area. Thus, a block with an incidence intensity of eight reflects the presence of southern pine beetles in approximately one-eighth of the 32,400 square miles (the area within the larger block) or approximately four thousand fifty square miles of land area. The interpretation here depends upon the relative constancy of county size, which seems to be generally valid for the

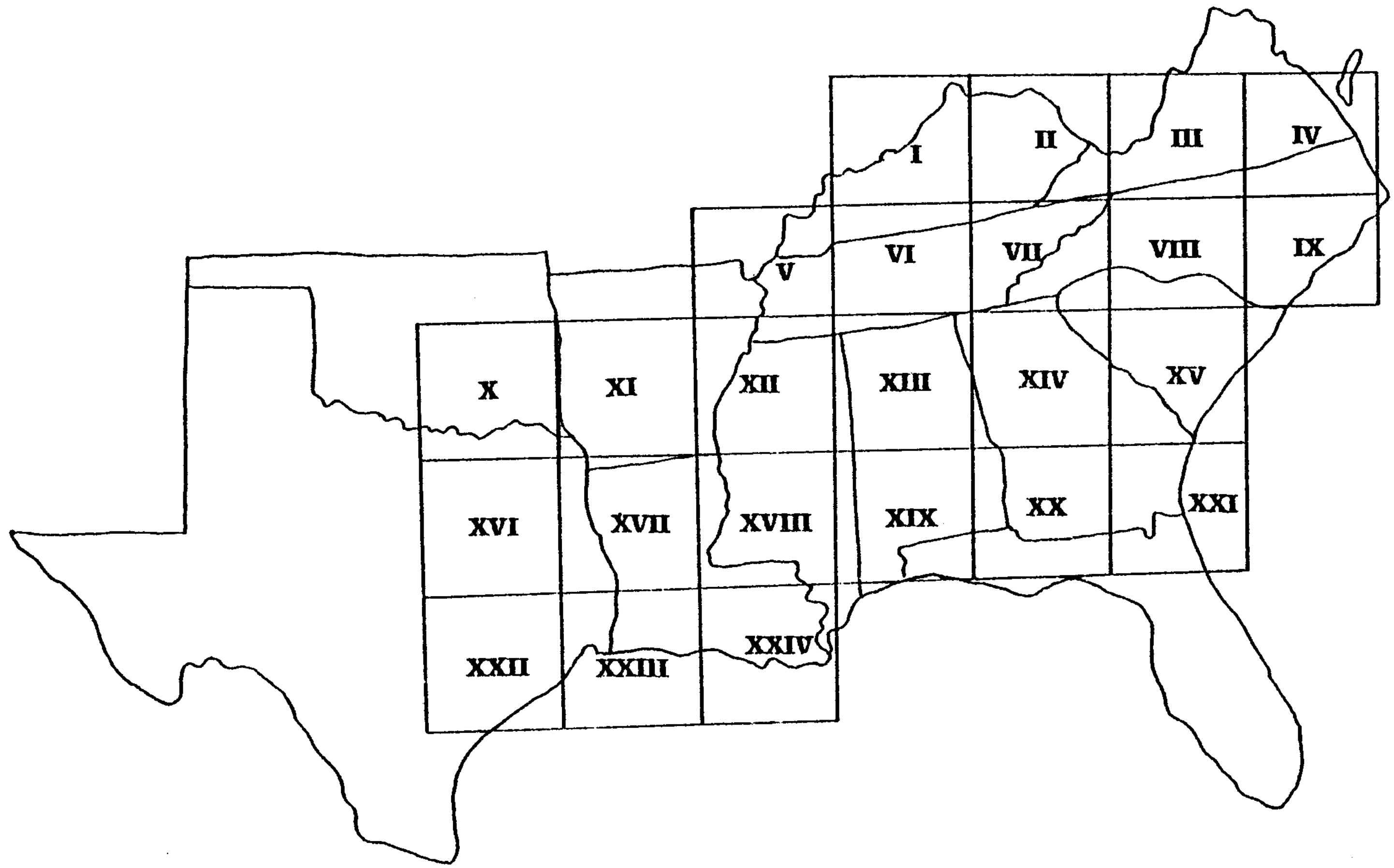

Figure 4.1--Formation of the twenty-four block mosaic from the southeastern United States

area of study. This type of measure is assumed throughout the analyses to adequately reflect the general pattern of southern pine beetle infestation behavior, despite its possible incongruencies. The blocks are treated as being identical with the concept of patch as defined in chapter two from here to the end of the thesis. The necessity of making the patch so large is not addressed in this work, but the sufficiency can and will be considered testable.

The final form of the data then has three forms:

1. incidence intensity on twenty-four blocks
2. incidence intensity on ninety-six sub-blocks (quarterings)
3. binary formulation of items one and two.

In the binary form, all incidence intensity measurements greater than or equal to one are set equal to one. Inherent in the structure of the data is information on the block-to-block connections in the mosaic. These connections are equivalent to those of a regular incomplete lattice where only those blocks sharing a common edge are considered connected. This is a particularly convenient form for space-time analysis but is not necessary if one is willing to struggle with the resulting difficulties.

These sets of raw data were then analyzed under two regimes, ordinary R-mode principal components and reference curve analysis. The ordinary principal components analysis was performed primarily as a check on the reference curve results and will be discussed only briefly. The results garnered from the reference curve technique yield far more important implications for model structuring and will be focused upon more intensely. The fundamental objective of both approaches is a classification of the blocks as to their typical temporal behaviors.

All three sets of data described above were subjected to the Statistical Analysis System's procedure FACTOR (Helwig and Council, 1979) using the METHOD = PRIN option which performs a standard principal components analysis of the data. Table 4.1 summarizes the results of these analyses. This computer algorithm uses the correlation matrix as its starting matrix and chooses enough factors to explain at least ninety-five percent of the variation. There are several interesting facets of this table that will be important later; these facets fall into three categories:

1. the results of binary masking
2. the results of refining the partition
3. the general results of the analyses.

The first category of features involves the consequences of switching the data to binary mode; this switch

TABLE 4.1--Details of the principal components analyses

I. Twenty-four patch, binary data.

Factor	1	2	3	4	5
Eigenvalue	7.508	2.291	2.185	1.594	1.216
Portion of variance explained	.417	.127	.121	.089	.068
Cumulative portion of variance explained	.417	.544	.666	.754	.822

Nine factors were required to explain 95% of the variance.

II. Ninety-six patch data.

Factor	1	2	3	4	5
Eigenvalue	34.247	14.574	9.253	5.051	2.993
Portion of variance explained	.445	.189	.120	.066	.039
Cumulative portion of variance explained	.445	.634	.754	.820	.859

Ten factors were required to explain 95% of the variance.

III. Twenty-four patch data.

Factor	1	2	3	4	5
Eigenvalue	14.053	4.144	2.778	1.147	.719
Portion of variance explained	.586	.173	.116	.048	.030
Cumulative portion of variance explained	.586	.759	.875	.923	.953

Five factors were required to explain 95% of the variance.

increases the number of factors necessary to account for the same amount of variance, more uniformly distributes the portion of variance explained by each eigenvalue, and achieves this uniformity chiefly at the expense of the two dominant eigenvalues, i.e. the largest one and the second largest one. This shift of importance away from the larger eigenvalues is the result of two causes, the required deletion of constantly active blocks (to ensure nonsingularity of the matrix) and the close relationship between this first factor and total intensity as is discussed below. The general blocking achieved by the binary masking is similar to that obtained from the unmasked analysis. Figure 4.2 demonstrates this blocking scheme for the twenty-four block binary scheme. For the uses in this paper, this binary masking results in too great an information loss.

Using the ninety-six non-binary smaller blocks of the finer partition has several apparent effects; it increases the number of factors necessary to explain the same amount of variation, spreads the portion of variance explained more uniformly over the factors, and achieves this uniformity mainly at the expense of the first and third eigenvalues, not the second as above. This finer partition achieves a blocking pattern as shown in Figure 4.3. It emphasizes the second largest eigenvalue and is subtler in its shadings of edges of the region.

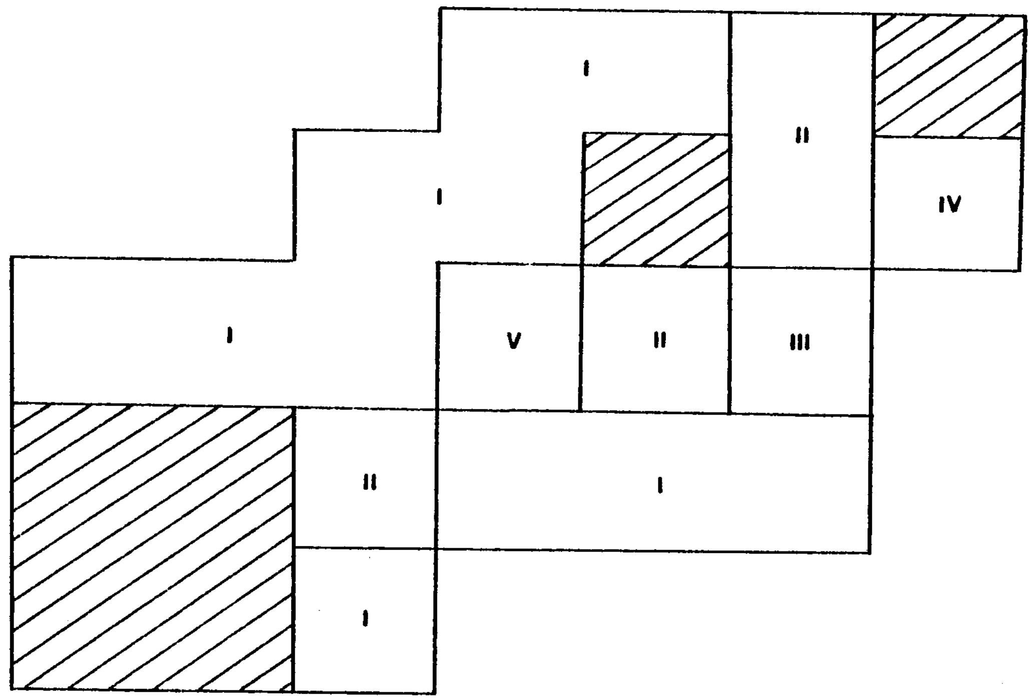

Figure 4.2--Blocking pattern suggested by the principal components analysis of the binary-masked, twenty-four patch data

The roman numeral within the area indicates the component followed.
Hatching shows blocks removed from the analysis due to constant intensity.

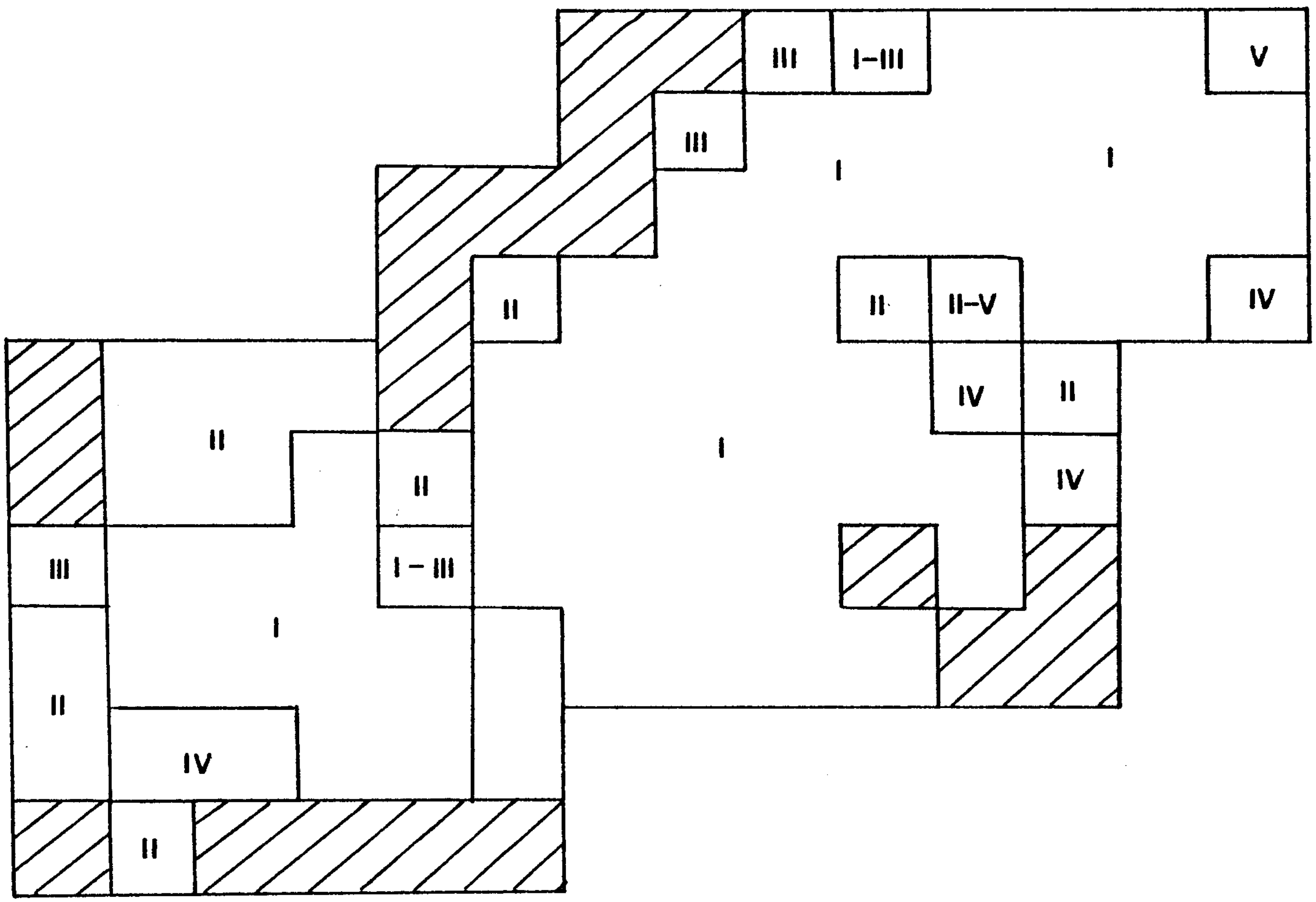

Figure 4.3. Blocking pattern suggested from the principal components analysis of the ninety-six patch data.

The roman numeral within the area indicates the component followed. Areas following several components are designated by two numerals. Hatched areas were removed due to zero infestation level.

Finally, there are certain features of the original twenty-four block and incidence intensity treatments that are important. These features include a typical reduction in axes of about sixfold, a consistent property of the first two factors under all data regimes, a relationship between the dominant eigenvalue and total epidemic size, and the grouping of like-behaving blocks (those following identical reference curves) into geographically contiguous areas. This blocking pattern is illustrated in Figure 4.4. More about the significance of these conclusions will be discussed under the topic of reference curve analysis.

The standard principal components analyses help confirm the general results of the reference curve analyses which form the central thrust of the investigation. The major difference between the two treatments stems from the type of square symmetric matrix used initially. Principal components uses a correlation matrix, while the other method uses the raw cross products matrix. This is a reasonable approach because of the comparable nature of the variables, i. e. they are areas. Using Sheth's notation, the process consists of partitioning the raw data matrix Y into the product of matrices P and V. To facilitate this process, the cross products matrix is formed as YY'. This matrix is square and symmetric and can be written as UGGU', where U contains the eigenvectors of YY' and

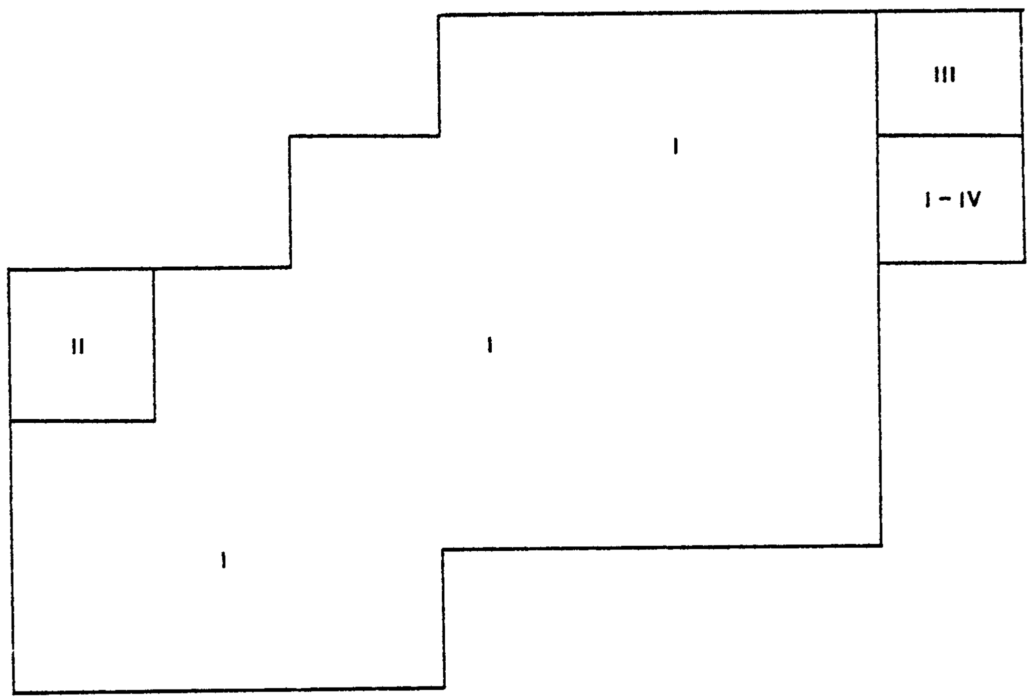

Figure 4.4--Blocking pattern suggested from the principal components analysis of the twenty-four patch data

The roman numeral within the area indicates the component followed. Areas following several modes are represented by several numerals.

GG contains the eigenvalues of YY'. Taking the first r largest eigenvalues and calling U_r and G_r the submatrices of U and G which contain the appropriate eigenvectors or eigenvalues, then the approximate matrix Y_r may be written $Y_r = AS = (U_r * G_r) * (G_r^{-1} * (U_r)' * Y)$. G_r^{-1} is the inverse of G_r. Due to the effect of sample size in this method, however, a rescaling is necessary. With this rescaling, $Y_r = P * V = (A * N^{-1}) * (N * S)$, where N is a diagonal matrix with all elements being the square roots of the sample size and N^{-1} being the inverse of N. In this formulation, P contains the reference curves, and V contains the individual coefficients.

Reference curve analyses were only run on the non-binary data but were run for both twenty-four and ninety-six block partitions. The binary data are excluded from this analysis because several blocks would be of constant value and would have to be eliminated to ensure nonsingularity of the data matrix. Also, the previous analyses suggest that no useful information is forthcoming from running such methods. Results from these analyses may be grouped into three sections:

1. isolation of reference curves
2. blocking patterns derived from the analysis
3. interpretations of these two results.

Each item portrays specific behavioral patterns for the southern pine beetle system which must be included in a competent model. The interpretations section will brush lightly on the uses of models with the observed structure, a topic treated in greater detail in chapter seven.

A reference curve is a standard of behavior which serves as one component in the explanation of an individual area's dynamics. Some blocks follow one specific curve exclusively, while others have behaviors which are amalgamates of several such curves. The partitioning of a sonic pattern into different frequencies is a good analogy of this process, but the reference curves are in general not well-behaved like sine or cosine functions. The reference curve, being a type of principal component, explains some portion of the total activity of the system. The size of this portion reflects the extent of influence of the particular curve in all block behaviors. The strength of this influence on the behavior of any individual block is reflected by the coefficients in the V matrix of the described decomposition. The relative values of these coefficients reflect the strength of the block and curve relationship. The explanatory power of individual curves and the number of curves is a good indicator of the nature of the basic system.

The first four reference curves for the twenty-four block analysis accounted for about ninety-six percent of the total activity of the system. The dominant curve,

i. e. the one corresponding to the largest eigenvalue, explained eighty-five percent by itself and is the major constituent of the system's behavior. Table 4.2 shows the

TABLE 4.2--Details of the twenty-four patch reference curve analysis

Factor	1	2	3	4
Eigenvalue	217755	12679	9226	5794
Portion of total explained	.85	.05	.04	.02
Cumulative portion of total explained	.85	.90	.94	.96

Four factors were required to explain 95% of the total.

activity breakdown of the reference curves, while Figure 4.5 illustrates the shape of the first four curves. The dominant curve (I) will be shown to be closely related to the total epidemic intensity over the entire twenty-four block structure in the correlation studies presented later in this chapter. The existence of secondary reference curves implies that all of the variability in block behavior is not explained by this one dominant curve.

Similar results were apparent after subjecting the ninety-six block mosaic to this standard curve analysis. In this case, the first four eigenvalues accounted for ninety-three percent of the activity with eighty-one percent explained by the dominant vector. Table 4.3 and Figure 4.6 give the details. Again, this reference curve structure reflects the same basic pattern, that of a dom-

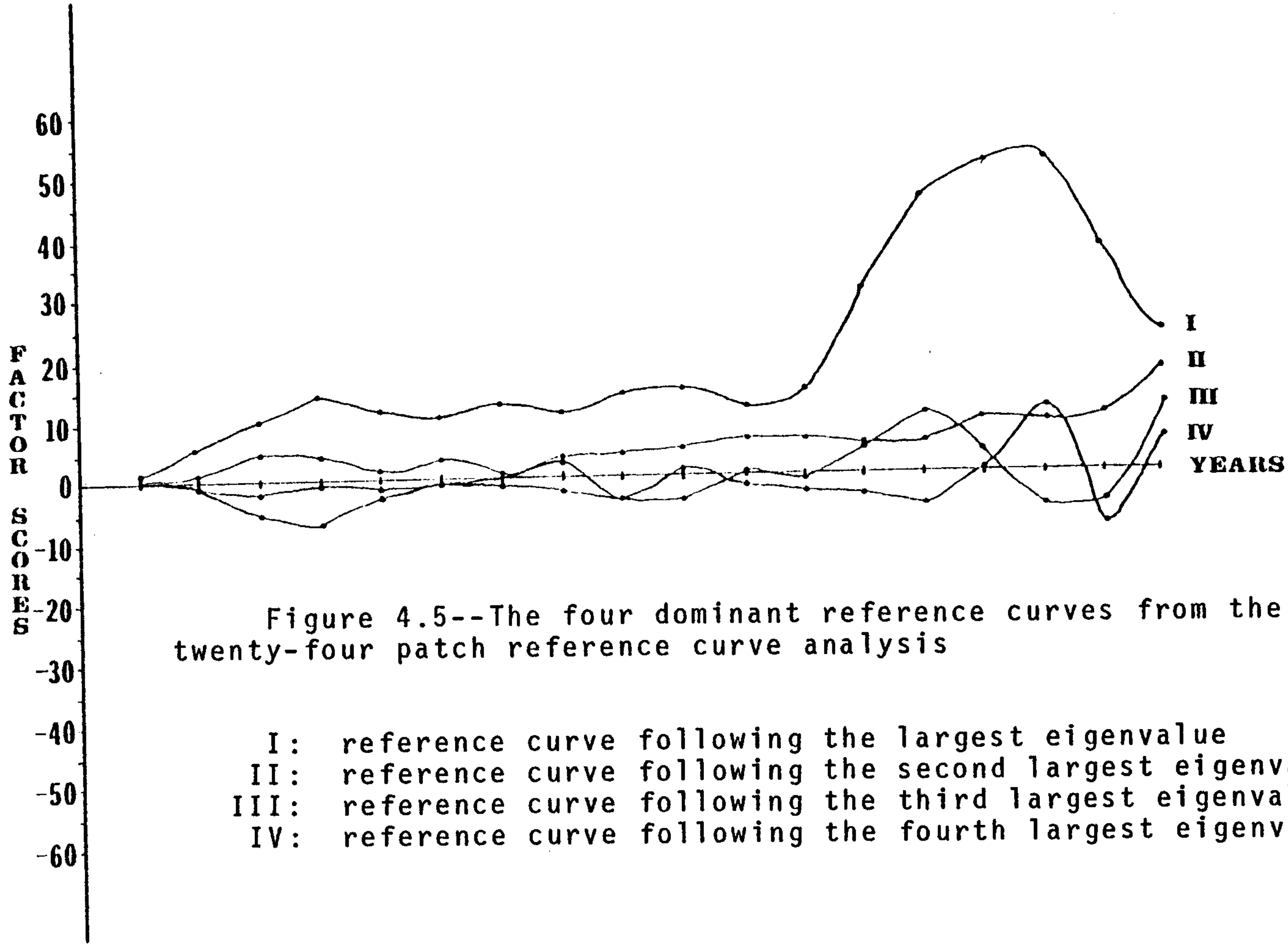

Figure 4.5--The four dominant reference curves from the twenty-four patch reference curve analysis

I: reference curve following the largest eigenvalue
II: reference curve following the second largest eigenvalue
III: reference curve following the third largest eigenvalue
IV: reference curve following the fourth largest eigenvalue

TABLE 4.3--Details of the ninety-six patch reference curve analysis

Factor	1	2	3	4
Eigenvalue	236626	14629	12079	8853
Portion of total explained	.81	.05	.04	.03
Cumulative portion of total explained	.81	.86	.90	.93

Five factors were required to explain 95% of the total.

inant epidemic effect generally obscuring some subtle individual block behaviors. In the ninety-six block partition, the relative importance of these underlying behaviors seems to increase.

The coefficients of the V matrix in the reference curve decomposition can be interpreted as weights, and on the basis of these weights, the blocks may be classified as to their predominant behavior. The classification involves comparing the relative strength of the weights on the four reference curves for a particular block. Often there will not be a clear-cut dominant weight, and this implies a mixture of modes within that block. Such a mixture of modes could occur because of two disparate modes within a block or a mixture of modes. The definition of dominance is not standardized, and the researcher must take care to be consistent in his judgment of this property. In this paper, dominance is assumed to be complete in every case, with the maximum value of each column in the V matrix determining the unique dominant curve

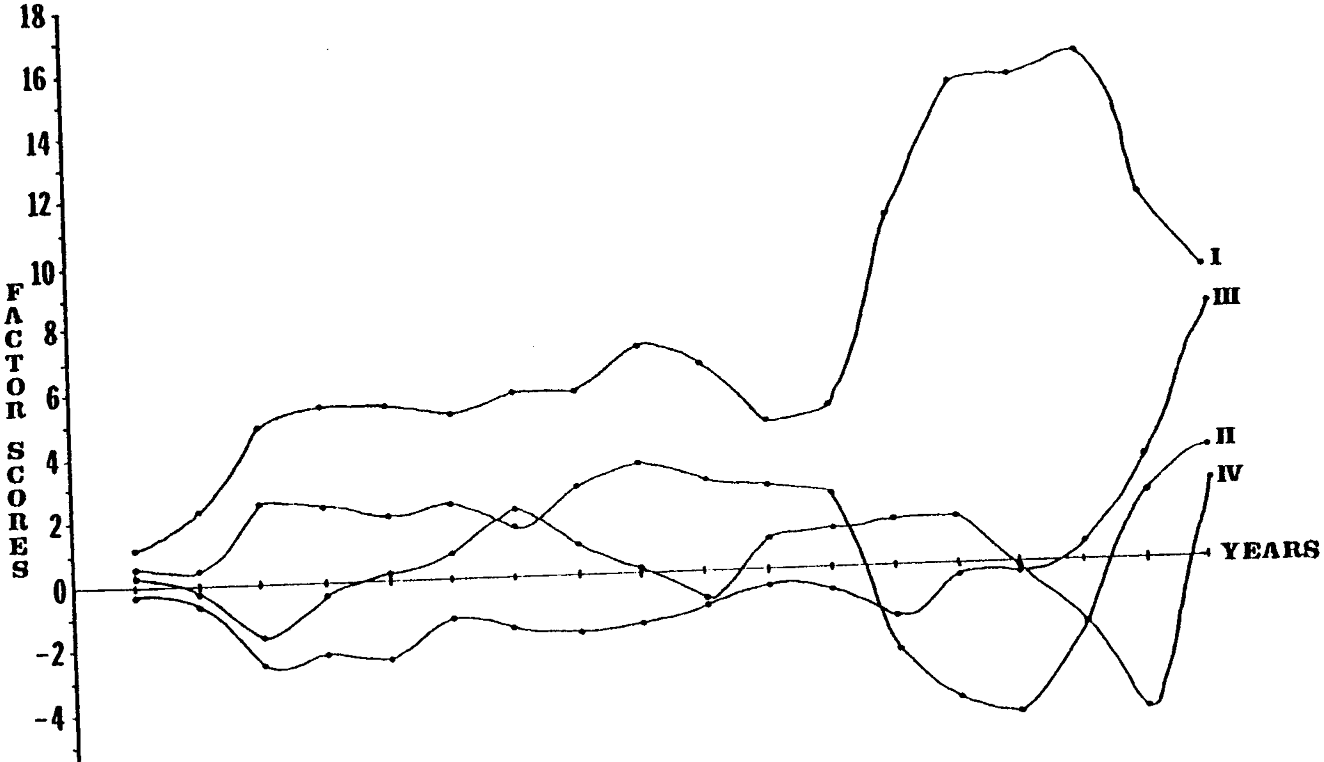

Figure 4.6--The four dominant reference curves from the analysis of the ninety-six patch data

I: reference curve following the largest eigenvalue
II: reference curve following the second largest eigenvalue
III: reference curve following the third largest eigenvalue
IV: reference curve following the fourth largest eigenvalue

whenever comparison between blocking patterns is required.

On the basis of the four reference curves reflecting the behaviors of the twenty-four blocks described above, each block was rated on which curve or set of curves it most clearly imitated. In this manner, a classification of the whole region was made possible, as shown in Figure 4.7. From this regional classification, certain patterns are evident:

1. three blocks follow the dominant curve
2. blocks following the second most dominant curve tend to be located in the western half of the region
3. blocks corresponding to the third largest seem to occur on possible interfaces
4. the fourth largest reference curve reflects sparsely infested areas and probably reflects extreme forest conditions
5. some spatial clustering of similar acting blocks occurs.

An attempt to relate some of these properties to environmental conditions is treated under the discussion of interpretation of results. Also described are some uses of the peculiar structure.

A similar classification scheme was created using

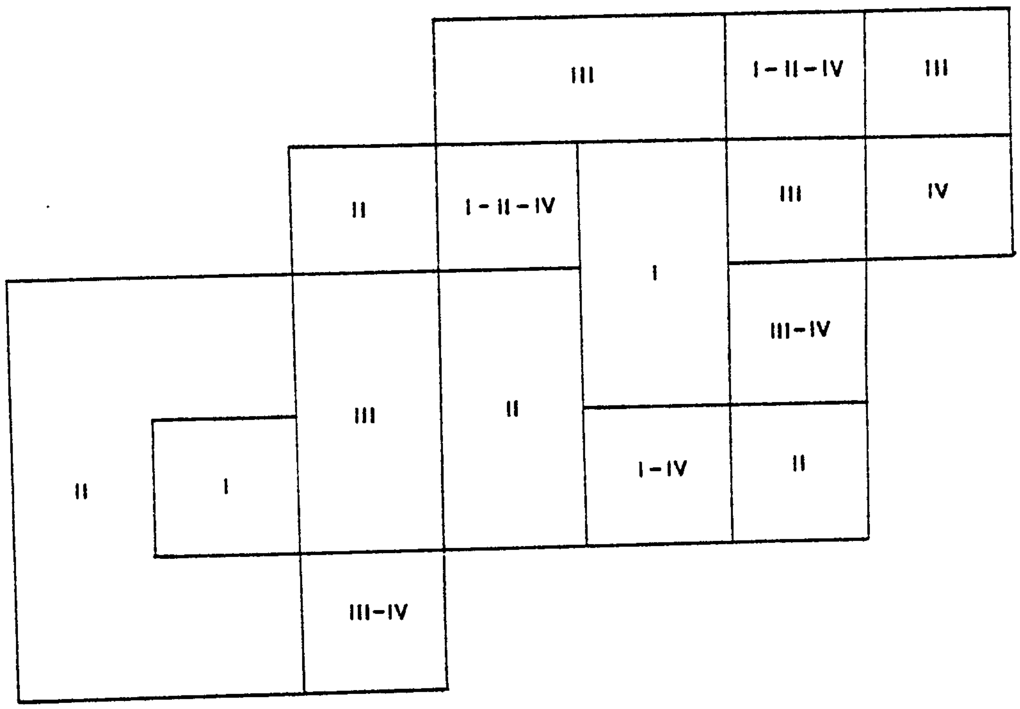

Figure 4.7--Blocking pattern suggested by the twenty-four patch reference curve analysis

The roman numeral within the area indicates the component followed. Areas following mixed modes are designated by several numerals.

the results of the ninety-six block reference curve analysis. The blocking pattern resulting from this process is shown in Figure 4.8. Although the patterns are far more complex and the reference curves do not correspond exactly, there are certain trends that are evident:

1. the dominant acting blocks are still in nearly the same locations and are still only two contiguous areas
2. block behaviors appear to vary widely in mode over short distances
3. blocks which follow type four behaviors may reflect some natural or induced interface
4. four large (over 4 blocks) areas of similar nature occur with two of these following reference curve two, one following reference curve three, and one following curve four
5. ten blocks follow type one behavior, seven located in the northeastern corner
6. coastal areas seem to follow predominantly lower strength reference curves.

The similarities and dissimilarities between the two blocking patterns are evident from the diagrams. The grain of the partition and the origin of the grid used in the par-

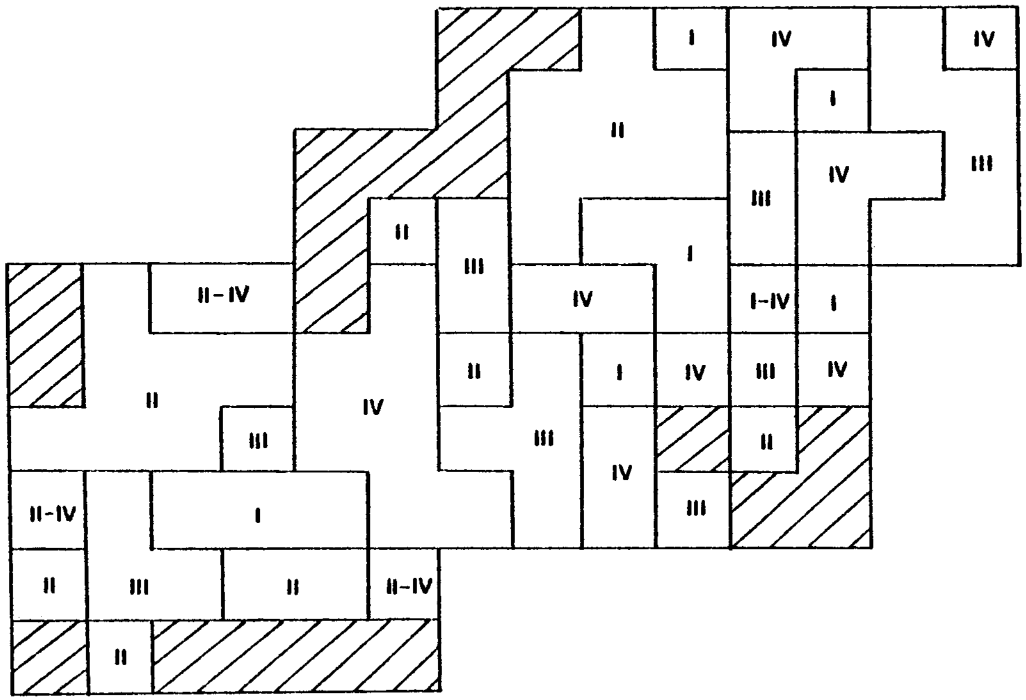

Figure 4.8--Blocking pattern suggested by the reference curve analysis of the ninety-six patch data

The roman numeral within the area indicates the component followed. Areas following several curves are designated by multiple numerals. Hatched areas were removed due to zero infestation level.

titioning process will probably influence the blocking patterns also, but general results should be unaffected.

Many questions are raised by the completed analyses whose answers would require extensive inquiry. However, certain inroads have been made on two of these:

1. the predictability of total intensity from single block intensity
2. the explanation of the blocking patterns in terms of static environmental effects.

The first has implications for surveillance practices, while the second has implications for control. In both cases, the analysis thus far is only correlative.

The first reference curve from the twenty-four block analysis is highly correlated with the total intensity of the infestation pattern (R = .99). Each of the three blocks which follow the primary curve are therefore also highly similar to the total infestation intensity. Table 4.4 shows the correlation of each of the twenty-four blocks with this total intensity, including the three mode one areas. Blocks seven, fourteen, and seventeen on the diagram each have high positive correlation coefficients with total intensity. Simple linear regressions of total on these three blocks yielded R-squares of .84, .80, and .84, respectively. A graph of the predictions from one of these regression equations (TOTAL = 40.169 + 8.592 * BLOCK7) versus the total intensity is presented in Figure 4.9.

TABLE 4.4--Correlations between the twenty-four patches and the total intensity and the four dominant reference curves

Patch	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1	.474	.050	-.427	-.129	.508
2	.540	-.053	-.442	.049	.569
3	.913	-.619	.121	-.189	.898
4	.551	-.165	.836	-.115	.556
5	.463	.321	-.150	.039	.528
6	.773	-.012	-.030	.224	.820
7	.922	-.455	.085	-.089	.914
8	.847	-.509	-.241	-.092	.822
9	.699	-.612	-.134	.580	.685
10	.288	.578	.096	-.062	.368
11	.572	.324	.322	.325	.642
12	.767	.031	.499	-.083	.810
13	.908	-.745	.305	.066	.880
14	.908	-.459	-.027	-.048	.892
15	.722	-.581	-.324	.458	.709
16	.687	.254	.226	.074	.743
17	.890	-.195	.460	.298	.914
18	.899	-.499	.498	.216	.895
19	.877	.905	-.657	.344	.877
20	.908	.897	-.357	.043	.908
21	.574	.501	.361	.023	.575
22	.548	.502	.350	.123	.548
23	.646	.599	.129	.426	.646
24	.697	.640	.096	.501	.697

Multiple regression using combinations of these blocks was not tried due to the high inter-block correlations. The implication of this result is that any of these three blocks could possibly be used as a bellweather for general southern pine beetle activity in the region as a whole, if the correlative effects are stable.

The explanation of the blocking patterns in terms of environmental effects is hindered by four data limitations. The environmental data consists of static measurements

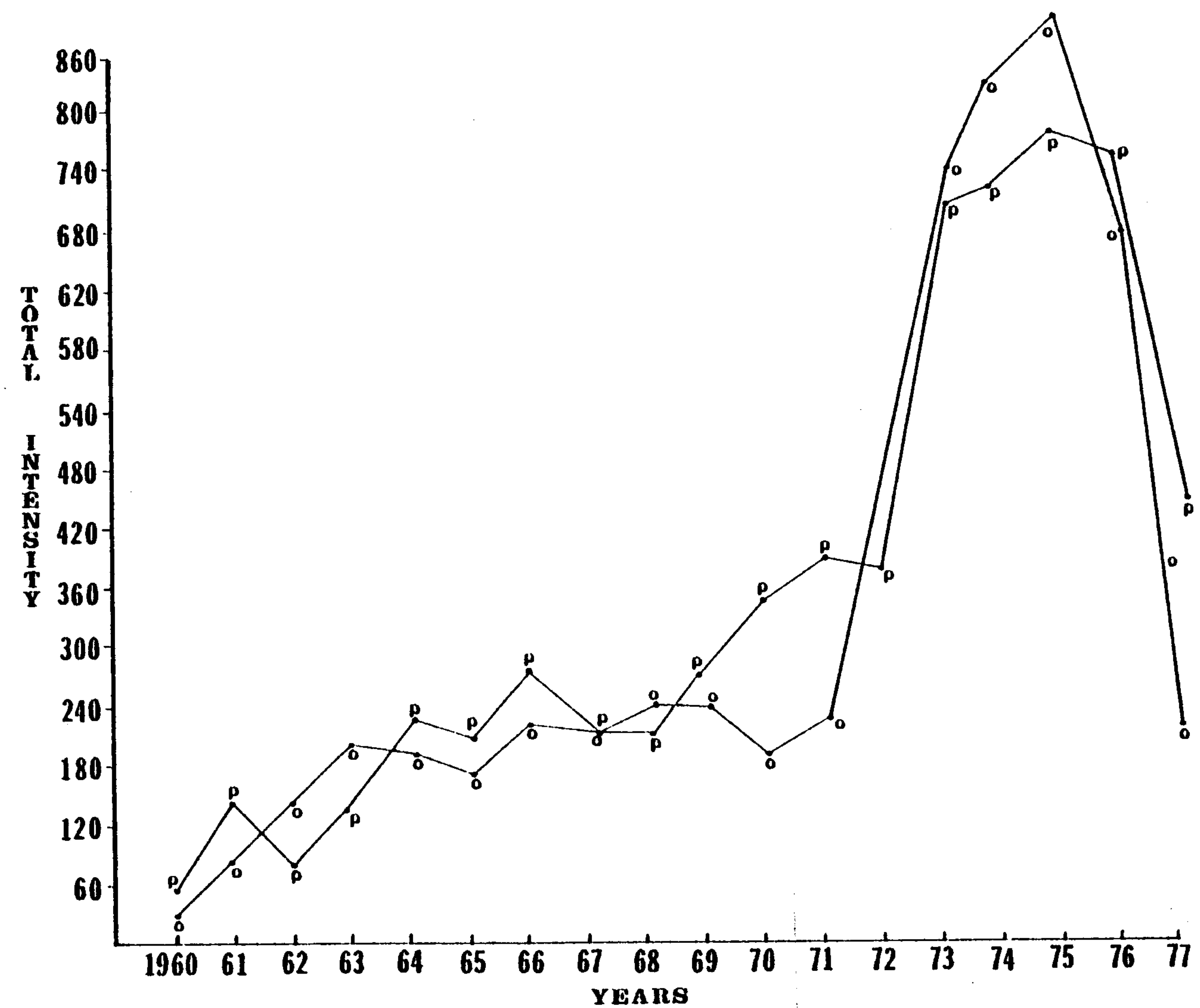

Figure 4.9--Observed total intensities versus the predictions from a linear regression on patch number seven

(or averages) which do not display the intricate dynamics of the forest. The data as given by Nelson and Zillgitt (1969) are categories rather than continuous quantities. Some effects, such as those due to weather, must assume a stochastic nature which the methods of analysis cannot pick out of the data. Finally, no estimate of the beetle effects on the regional forest dynamics can be made based upon this data. Although more detailed data may be available on certain attributes or on certain subsections, this data represents some of the most reliable global information available.

Despite these inadequacies, several results have been established. A standard principal components analysis was run on the twenty-four block system using forest characteristics from Nelson and Zillgitt (1969). Included were loblolly volume, slash volume, longleaf volume, shortleaf volume, and percent forested land. The blocking pattern which results is shown in Figure 4.10. Block-by-block comparison implies that most of the disparity in the west and the deep southeast can be explained by reference to forest differences. The northern areas do not appear explainable in these terms, however. A second principal components analysis incorporating other environmental variables from Nelson and Zillgitt, including physical division, forest type, water deficit, and average monthly temperature for January and July, hints that these north-

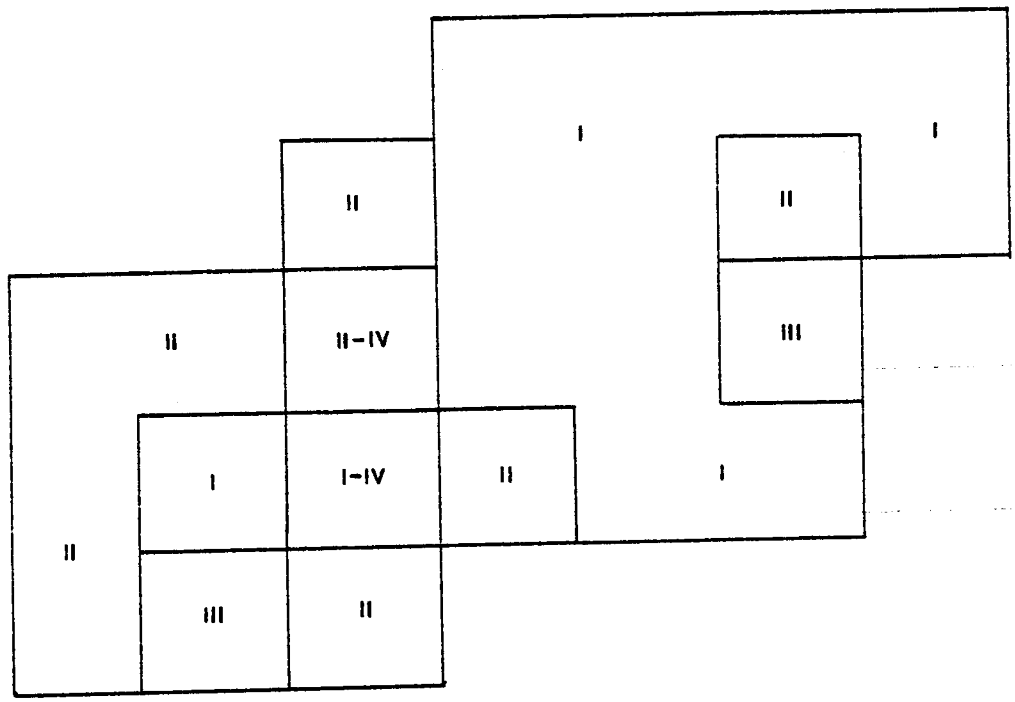

Figure 4.10--Blocking pattern suggested by the principal components analysis of the reduced environmental data

The roman numeral within the area indicates the component followed. Areas of mixed mode are identified by multiple numerals.

ern areas may be disparate in some of these variables. Figure 4.11 shows the results of this second analysis. These analyses are not meant to explain the differences in beetle infestation behavior observed in the other data but rather to illustrate that a blocking pattern based on tree properties, weather effects, water deficits, and temperature drops could mimic those beetle based patterns presented before. Such an approach could form one of the few ways in which the reference curves might be explained via correlated environmental variables. Furthermore, the two analyses described above indicate tree and temperature variables as primary aspects of such an attempted explanation.

A detailed conclusion to this set of analyses fits more appropriately into the initial stages of chapter six and will not be treated here. However, there are some important points that should be listed:

1. reference curve analysis has potential for this type of analysis
2. the possibility of the reduced surveillance is suggested
3. the tremendous variability of the system is possibly stable on this level and requires environmental data for complete description.

The validity of the last point may be questioned at this

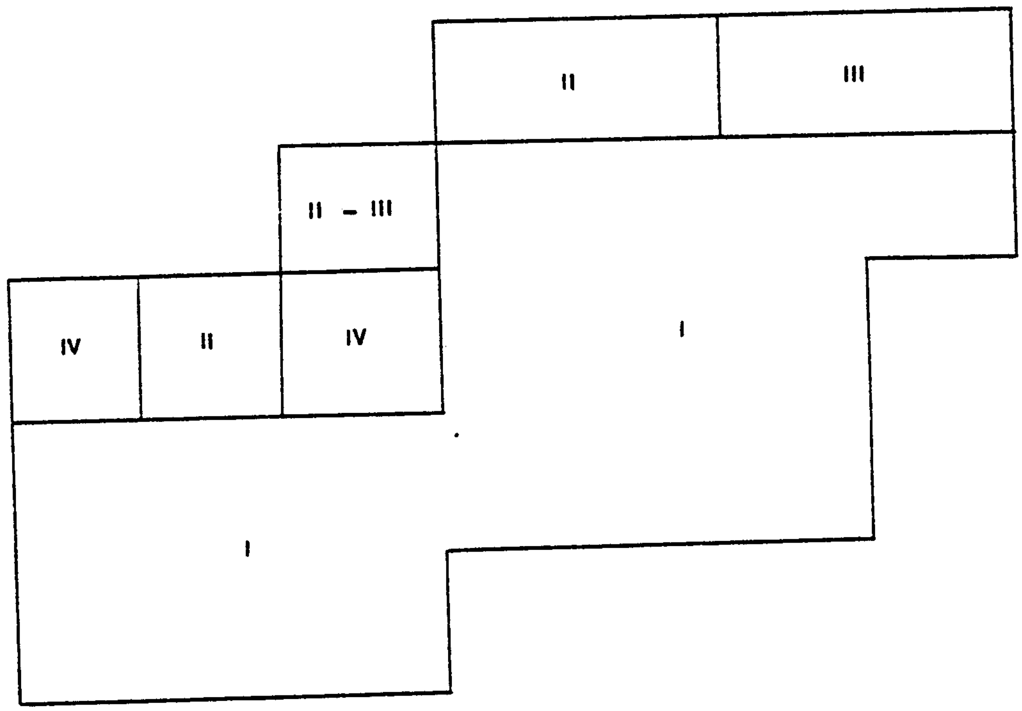

Figure 4.11--Blocking pattern suggested by the principal components analysis of the full environmental data

The roman numeral within the area indicates the component followed.

stage, but later work will help to test it. Results from these studies will be used primarily to develop the surveillance schemes discussed in chapter 7 but certainly are of value in themselves also.

CHAPTER V

SPACE-TIME SERIES ANALYSIS

The space-time series portion of the analysis was based on the same data as used in chapter four, with the same block structure, but without any development of the ninety-six block partition. The reasoning behind this omission will be made clear later when the spatial series analyses are discussed. Thus, the nearly identical sets of data were analyzed by two widely different techniques in an attempt to establish a sense of cross-validity for the results.

The discussion of the space-time series work will be presented in seven parts:

1. initial spatial analysis
2. initial temporal analysis
3. spatial trend removal
4. temporal trend removal
5. secondary spatial analysis
6. secondary temporal analysis
7. other analyses.

Several important details must be discussed in relation to the general framework of autocorrelation structure, and these details will form much of the description of the preliminary items. However, the more important modeling information is contained under parts five, six, and seven. Finally, chapter six will merge the conclusions of the two exploratory chapters into an upper echelon submodel.

The spatial series analysis of the southern pine beetle infestation intensity data proceeds along lines discussed in chapter eight of Cliff, et. al. (1975). Throughout this discussion, the time is assumed fixed. Using this methodology, spatial autocorrelation may be defined as

$$C = \Sigma w_{i,j} * f(x_i, x_j), \quad i \neq j \tag{1}$$

where $w_{i,j}$ represents connection weights, and f represents the pairwise relation which defines similarity, i. e. correlation. If the matrix of weights is called W, and the observation vector is X, then the jth element of WX is the first order lag for element j of the X vector. The nth order lag of a given element of X relates those members of X to which the given element is connected in precisely n steps. In general, n must be less than or equal to the diameter of the graph representing the block to block connection pattern. In this manner, all orders of spatial lags may be defined. However, while powers of the matrix reveal all routes from i to j in a specific number of steps, the necessary elimination of routes with repeated nodes is not so easy. Two methods have been developed, an iterative process and an explicit representation; both prove impractical for lags of order greater than seven or eight. Ross and Harary (1952) and Marshall (1971) give the details. Under simple weighting schemes, such as were used in this paper's analysis, the lag structure can easily be read from a diagram of the mosaic. This simple scheme sets all connection weights of the W matrix to zero except for cell borders sharing a complete edge. Figure 5.1 shows this weighting scheme. Although this paper examines no other weighting schemes because of the large number of lags involved, the use of non-symmetric patterns can isolate important features of the system. The spatial autocorrelation thus takes the form

$$I_k = (n/w) \Sigma (z_i * z_{i,k}) / \Sigma z^2_i \qquad (2)$$

where $z_{i,k} = x_{i,k} - \overline{x}$, $n\overline{x} = \Sigma x_i$, and $w = \Sigma w_{i,j}$, $i \neq j$. The use of $\overline{x}$ rather than $\overline{x}_k$ is recommended by Cliff, et. al. (1975) to ensure lag-to-lag comparability.

Using the measure defined in equation (2) above, the twenty-four blocks were analyzed for spatial autocorrelation properties on both the initial and the binary-masked data. The correlograms, the graphs of I_k versus k, are shown in Figures 5.2 and 5.3 for several typical years of the data. The examination of these correlograms points out two types of results, general features and the effect of the binary-masking.

As a very crude estimate of the significance of the correlations, the variance of I_k can be approximated by Bartlett's formula:

$$Var(I_k) = 1/N(1 + 2 \Sigma I^2_r), k > q \qquad (3)$$

as related in Box and Jenkins (1976). This formula is

PATCHES

PATCHES		2		4		6		8		10		12		14		16		18		20		22		24
	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
8	0	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0
12	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0
14	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
16	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0
18	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0

Figure 5.1--Connectivity matrix for spatial correlation analyses

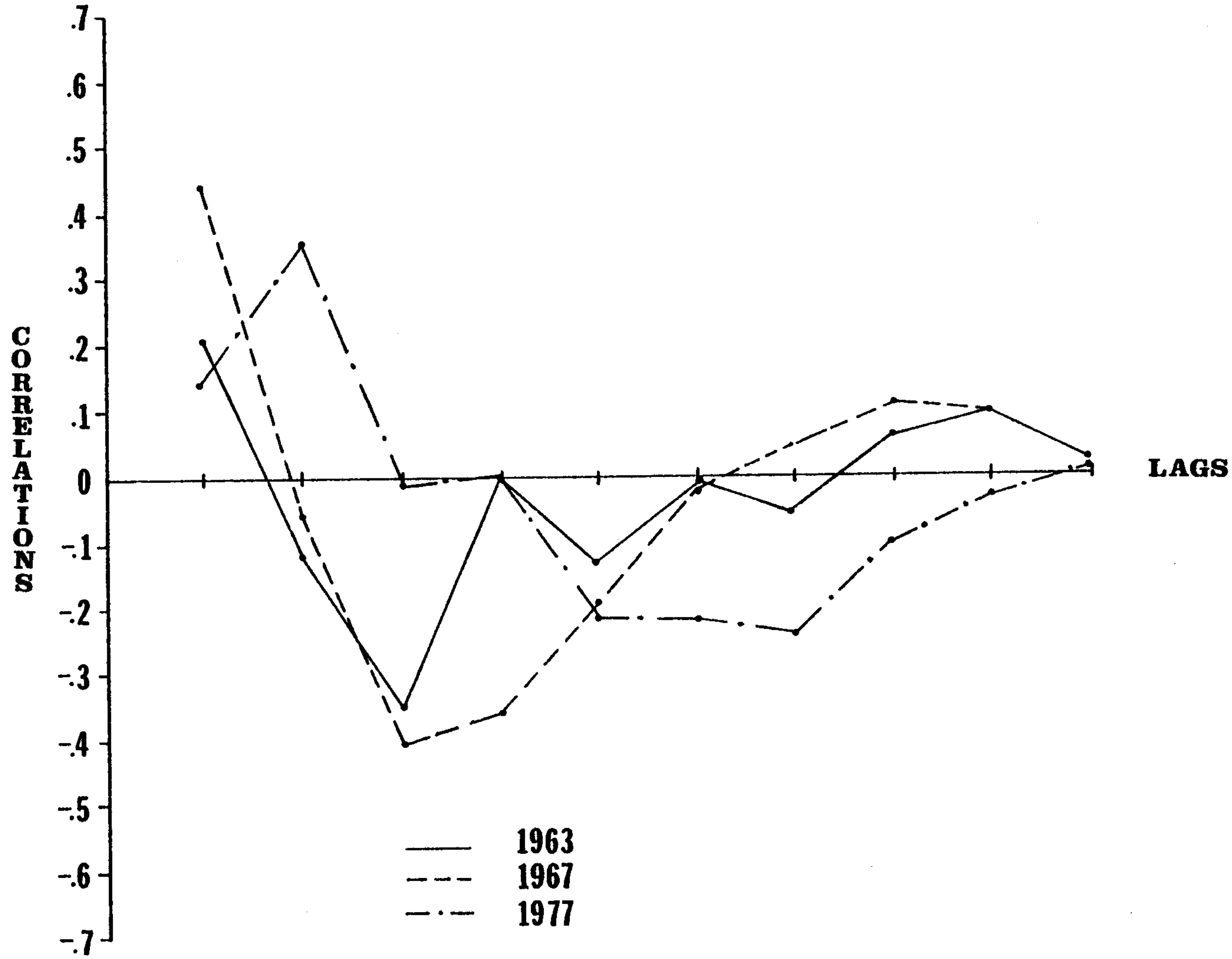

Figure 5.2--Spatial autocorrelation functions for binary data

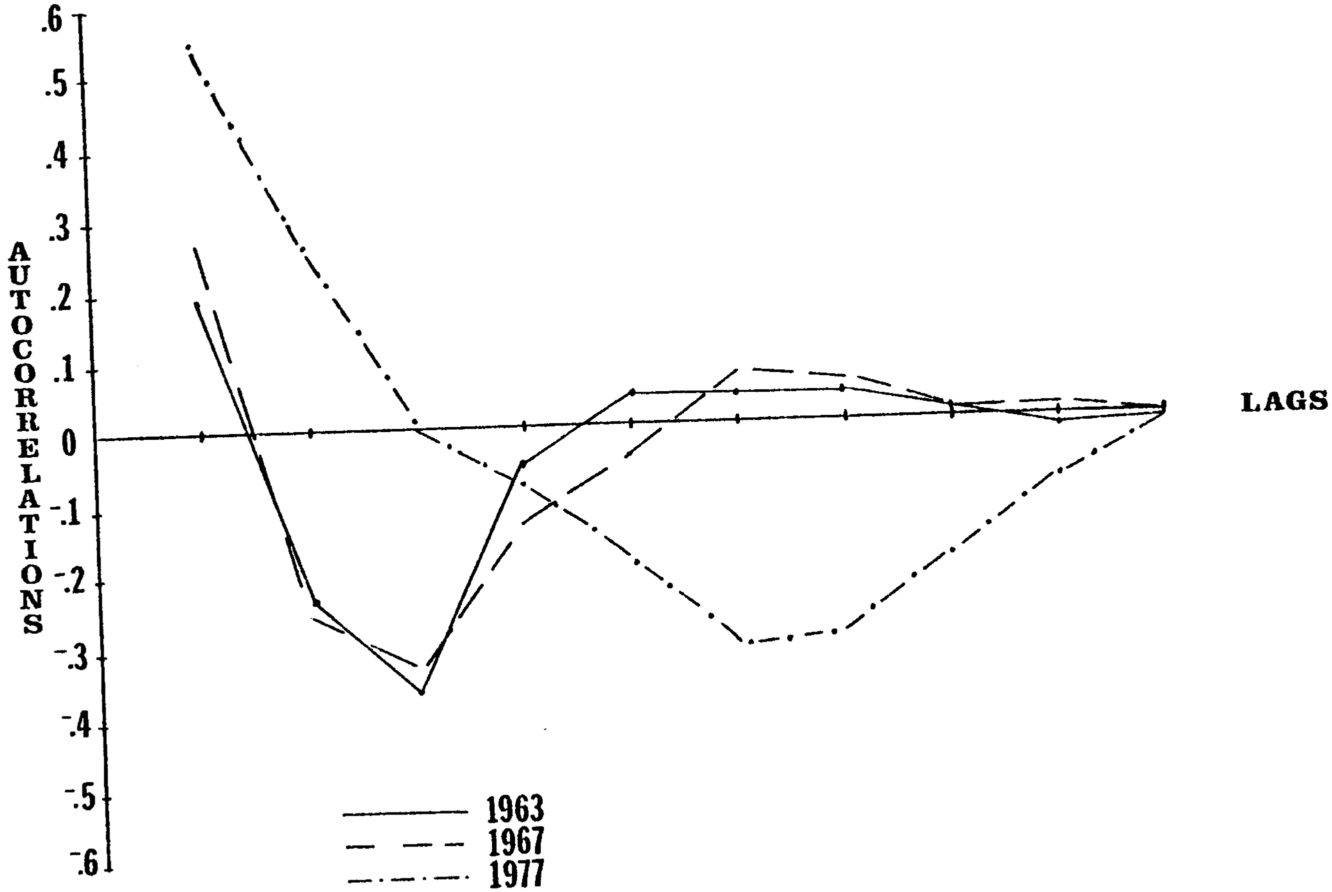

Figure 5.3--Spatial autocorrelation functions for full data

valid for a lag value q at which all further autocorrelations may be assumed to be effectively zero. Using this estimate (and the approximate normality of estimated autocorrelations), a rough check on the order of the process can be performed. Under this scrutiny, nearly all of the correlograms from the raw spatial can be considered merely random, white-noise processes. That is, the neighbor-to-neighbor relationship is not well developed. The conclusion is that at this size partition the spatial correlation effects can be considered as first or zero order processes. However, the autocorrelation measure given in (2) is known to damp the correlogram, and the significance test is very rough indeed.

The binary-masked data exhibits correlograms nearly identical to those from the unmasked data. The only points of interest appear during the years 1972, 1973, and 1974, where the shape of the correlograms differs greatly from the unmasked ones. The lag four autocorrelation especially appears overly large. Under the significance test, however, these years still may be considered random noise.

An analysis of the temporal autocorrelation within each block was also run using the procedure AUTOREG of the Statistical Analysis System package (Helwig and Council, 1979). The estimation of temporal autocorrelations uses the same logic as presented in explanation of spatial autocorrelations but does not meet with the complications of multiple lag elements and the associated repeated node processes. Only unmasked data was run using this method. In discussion of significance, Bartlett's approximation is again used.

As one might expect, the correlograms as a set are quite varied in their shapes and structures. Three categories of behavior are evident:

1. nearly zero correlations
2. non-stationarity
3. stationarity.

Figure 5.4 illustrates these modes for typical blocks. There are six blocks under mode one behavior, thirteen under mode two, and five under mode three, as shown in Figure 5.5. Due to the widespread non-stationarity it was decided to first remove this deterministic part before performing further analyses. The nonstationary behavior of a block is represented by a correlogram which changes at a slow steady rate over most of its lag values.

Time series and space series are often conceptualized as being composed of three components:

1. a deterministic trend
2. a periodic fluctuation about the trend
3. a random element.

Non-stationarity can be a result of a trend, and removal of this trend can enable the true stochastic process, com-

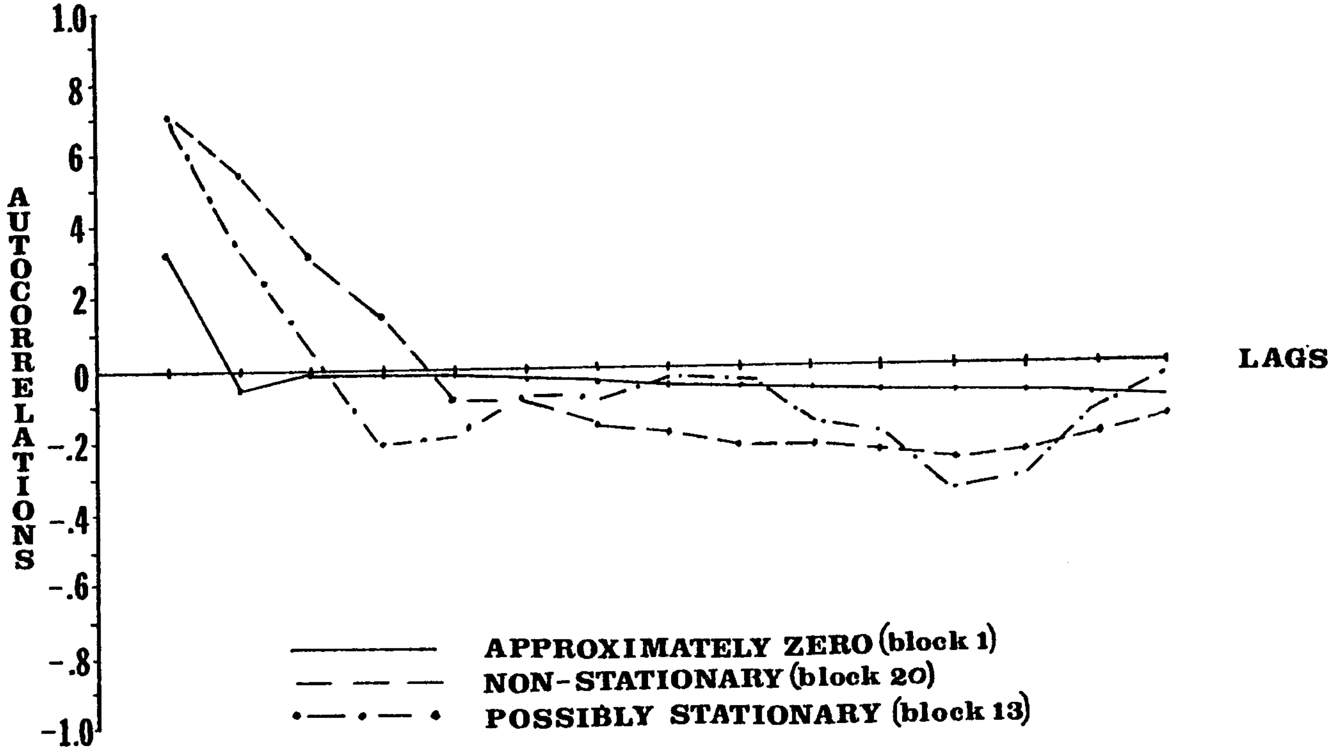

Figure 5.4--Temporal autocorrelation functions for full data

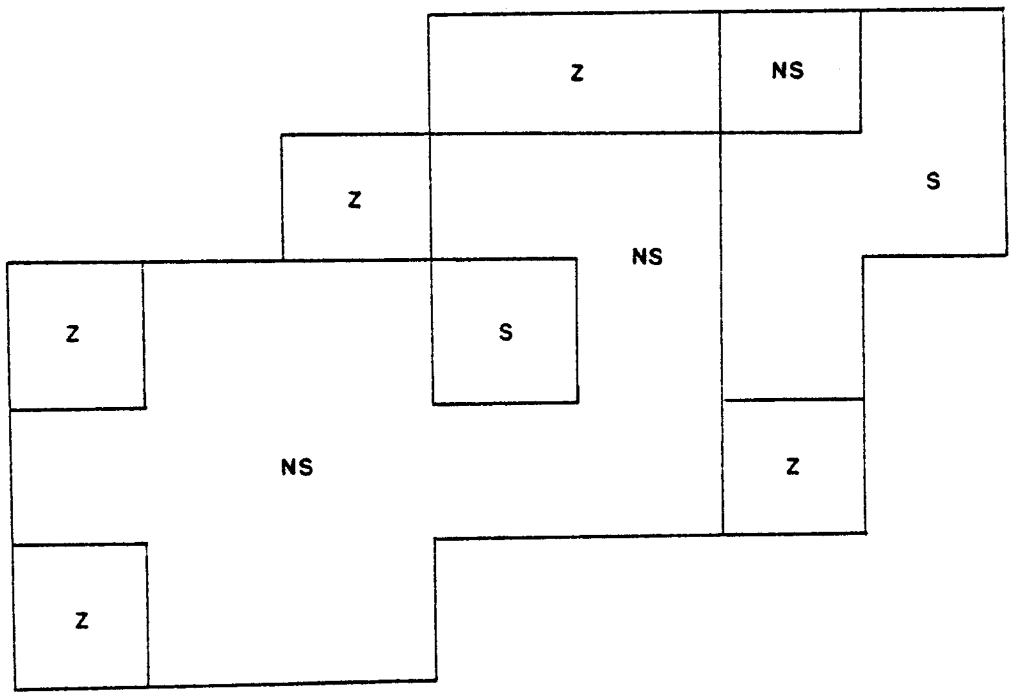

Figure 5.5--Blocking pattern suggested by the temporal autocorrelative structure of the raw data

Z: area resembling a zero order process
S: area resembling a nonzero stationary process
NS: area resembling a non-stationary process

ponent three, to emerge. This is the approach applied to the space-time analysis of the infestation data. Two separate trends can be removed, first a spatial trend from each year and then a temporal trend from each block.

A coordinate system can be placed over the twenty-four block system to yield approximate X and Y values for the midpoint of each block. A multiple linear regression, i. e. a trend surface analysis, is used to model the spatial distribution of beetle infestation intensity as a cubic function of X and Y. This includes X, Y, X^2, Y^2, XY, X^2Y, XY^2, X^3, and Y^3 as independent variables and is done separately for each year of data. The cubic equation was generally more complex than necessary to explain the data (to whatever degree is possible), but it was retained throughout the analyses to ensure a common basis for comparing trends from year to year. Although the regression does not fit well in every year, it does remove a substantial portion of the variance in all cases. Table 5.1 gives a breakdown by year of the R^2 value of each fit. No attempt is made here to relate these surfaces to environmental conditions due to lack of adequate data on forest dynamics, although such an effort might yield valuable results.

The residuals from the spatial regressions can be plotted to show considerable autocorrelation. A temporal trend removal can then be performed on these residuals

TABLE 5.1--Details of the spatial trend fittings by year

Year	R^2	F-value	Significance Level
1960	.5340	1.78	.1602
1961	.3501	.84	.5949
1962	.4424	1.23	.3494
1963	.5353	1.79	.1582
1964	.5296	1.75	.1676
1965	.7112	3.83	.0124
1966	.6332	2.68	.0476
1967	.6734	3.21	.0251
1968	.6654	3.09	.0287
1969	.6064	2.40	.0691
1970	.4410	1.23	.3526
1971	.4886	1.49	.2443
1972	.6945	3.54	.0171
1973	.7378	4.38	.0070
1974	.7711	5.24	.0031
1975	.7566	4.84	.0045
1976	.7011	3.65	.0151
1977	.7193	3.99	.0105

using a cubic function of time. Thus, time, time-squared, and time-cubed are the available variables, with time running from one to eighteen (1960 to 1977). A different regression can be performed for every block in this way. The R^2 value for each trend analysis is given in Table 5.2. The cubic equation was used to ensure enough generality without losing cross block comparability. In general, the fits are quite poor but, again, often can pull out a significant amount of variation from the data. Since the purpose of these manipulations is primarily to eliminate the non-stationarity of the original time series rather than to establish a fit of the data, these fits are considered acceptable.

TABLE 5.2--Details of the temporal trend fittings by block

Block	R^2	F-value	Significance Level
1	.5506	5.72	.0091
2	.4252	3.45	.0458
3	.0493	.24	.8656
4	.0454	.22	.8797
5	.1741	.98	.4286
6	.3962	3.06	.0630
7	.6525	8.76	.0016
8	.1666	.93	.4509
9	.2455	1.52	.2533
10	.3324	2.32	.1194
11	.5924	6.78	.0047
12	.7005	10.92	.0006
13	.2065	1.21	.3409
14	.1815	1.03	.4075
15	.5284	5.23	.0125
16	.5048	4.76	.0173
17	.7758	16.14	.0001
18	.1472	.81	.5113
19	.4493	3.81	.0347
20	.3304	2.30	.1216
21	.3976	3.08	.0620
22	.1588	.88	.4746
23	.4131	3.28	.0525
24	.2791	1.81	.1922

The second order residuals from the temporal trend removal can then be analyzed for spatial autocorrelation structure. The same weighting and lag scheme is used as in the raw analysis, and Bartlett's approximation is again used for lag-order estimation. Each year's correlogram is of nearly identical shape, but the size of the correlograms differs. A year-by-year breakdown of the proposed lag-order is given in Table 5.3. Roughly seven years can be considered of zero-order, while the other eleven years display one-order structure or higher. A

TABLE 5.3--Autocorrelation order estimation for spatial analyses of the detrended data

Year	Order
1960	3
1961	0
1962	0
1963	0
1964	0
1965	1
1966	3
1967	0
1968	5-6
1969	0
1970	0
1971	0
1972	0
1973	1
1974	0
1975	0
1976	2-3
1977	0

typical correlogram from each type is shown in Figure 5.6. The higher order structures seem to be associated with years generally considered to be of increasing epidemic character. Changes in residual direction seem to stem from the trend surface fittings.

A time series analysis of the residuals from trend removal can also be performed using the procedure AUTOREG exactly as before. The results are quite remarkable in their consistency of shape and their implications for model structure. Particularly important is the implication that any model must utilize a memory of the internal block behavior for two to three years at least. But the required length of this memory varies with the block in-

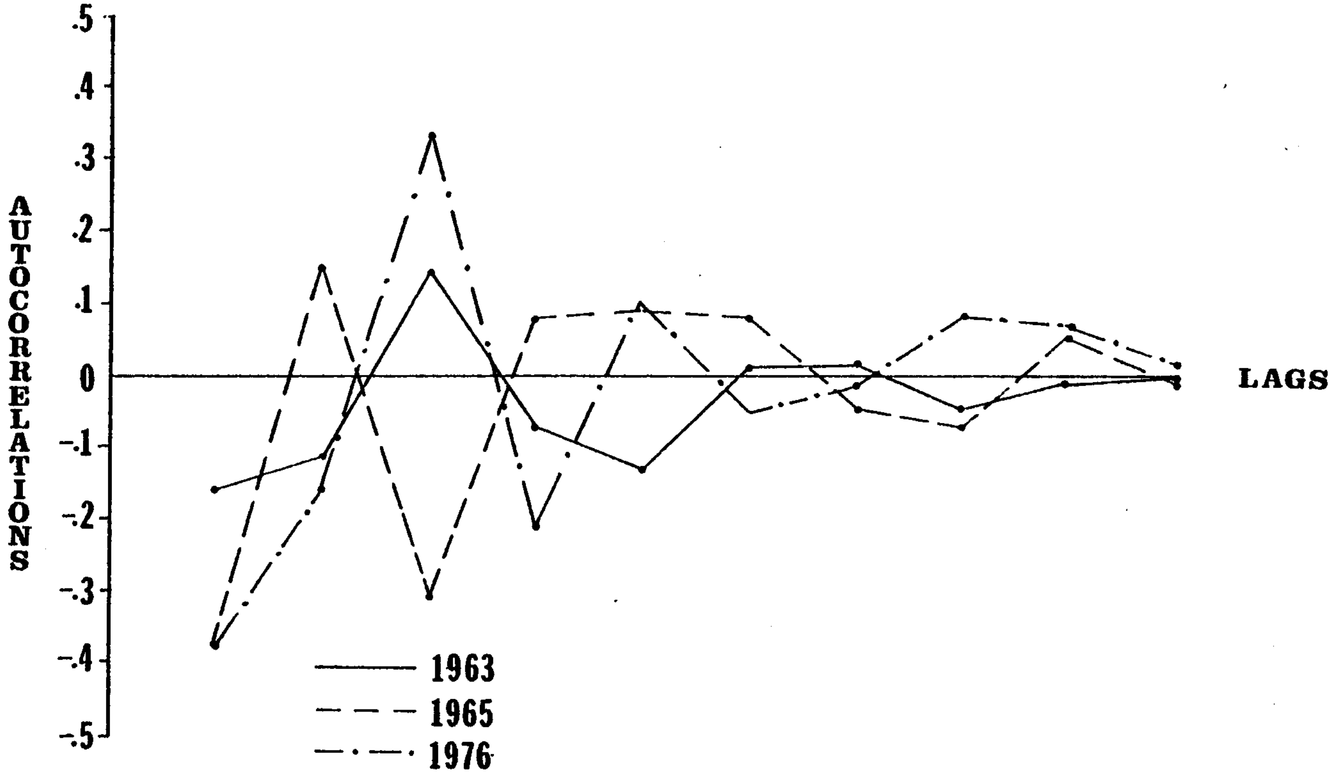

Figure 5.6--Spatial autocorrelation functions from the detrended data

volved, with some of them requiring only neighboring intensity values to determine their own. The smoothly changing shape and the high order of the estimated autoregressive structures imply that a periodicity may be evident which might substantiate historical observations of the same nature. An examination of lag-order is given in Table 5.4 for each block, and some typical correlograms are shown in Figure 5.7.

TABLE 5.4--Autocorrelation order estimation for temporal analyses of the detrended data

Patch	Order
1	5
2	0
3	0
4	2
5	0
6	5
7	0
8	0
9	7
10	2
11	0
12	0
13	4
14	0
15	0
16	5
17	5
18	2
19	0
20	4
21	3
22	3
23	0
24	2

Just as with the raw time series, the classification

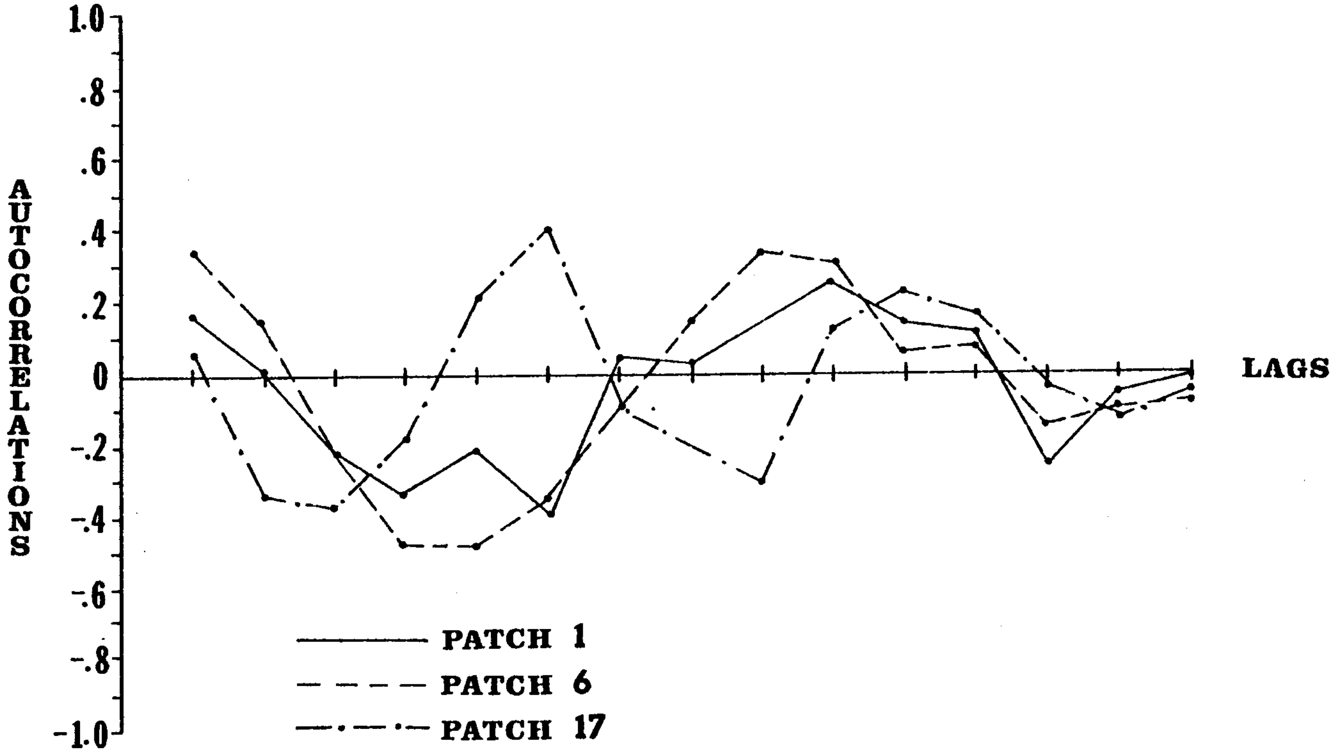

Figure 5.7--Temporal autocorrelation functions for detrended data

of the twenty-four blocks on the basis of the residual time series results in a significant blocking pattern. Intuitively satisfying is the occurrence of zero order processes only at boundary areas. The form of this pattern can be seen in Figure 5.8. The patterns of the raw and residual analyses do not appear equivalent, but the description of either pattern via a simple model does seem feasible. The contiguous grouping of like behaving blocks must be an integral part of any such model.

As in the previous chapter, a detailed conclusion to these studies is more appropriately presented in chapter six, but certain implications should be pointed out. These include general properties of the time series method and of the specific southern pine beetle results:

1. the time-space series results, while requiring a great deal of manipulation and approximation, do seem useful for studying the system
2. the trend surface approach could itself be quite useful in developing large-scale site and stand ratings
3. the non-stationarity of the initial series implies a heavy dependence upon environmental factors
4. the emergence of the high order structure of the final time series implies

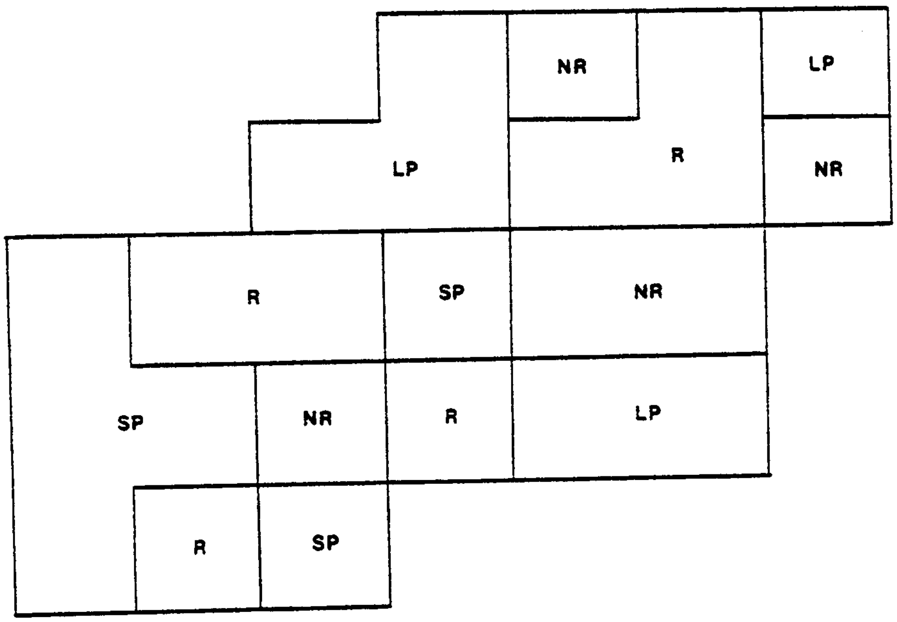

Figure 5.8--Blocking pattern suggested by the temporal autocorrelation behaviors of the patches

R: area of zero order
NR: area of non-zero order
SP: area of possible period of short duration
LP: area of possible period of long duration

a dependence on past activity to a high degree

5. the spatial series approach seems valid and useful, but the system shows little pattern.

The overall conclusion is to portray the time-space behavior of the system as a network of loosely connected oscillators, each with a unique period. Full pursuit of this approach is not made due to the complexity of the required system and the desire to reach some useful conclusions.

CHAPTER VI

FORMULATION AND ANALYSIS OF THE UPPER ECHELON SUBMODEL

The formulation of a flexible and accurate model is a difficult task and one which demands many iterations of the analysis-model-testing sequence. Chapters four and five describe the analysis portion of just one step in the process. The present chapter treats the latter two portions of that same first step. The construction of a general model of the upper echelon submodel for the southern pine beetle hierarchical model based on the results of the exploratory analyses is presented first. The model structure is then tested in some simple ways, by comparing the range of model and system behaviors and by simulating some realizations of the model. Such analyses cannot ensure the necessity of a model, but they can provide a measure of sufficiency. Such a sufficient model provides the basis for further work and becomes a viable component in the overall hierarchical model. Such a model may suggest possible utilizations and extensions, as will be covered in chapter seven.

Several results stemming from the exploratory multivariate analyses must be considered vital for any model of the beetle system. Based upon the two types of analysis (principal components and reference curve analysis), upon three types of data preparation (large-scale, small-scale, and binary), and upon the factor analyses of the environmental conditions, there are at least eight requirements that can be stated. Firstly, there exist at least four basic behavioral modes for individual blocks corresponding to the four reference curves. These modes relate block types through amplitude, timing, and durational differences of the southern pine beetle activity within them. Secondly, a distinct eastern-western separation occurs along blocks 12 and 18 with each side acting somewhat disjointly, although they are probably not uncoordinated. Thirdly, a distinction between center and fringe blocks can be seen, with areas located towards the extremes of the range exhibiting sparse, low activity. Their role in the overall system appears to be as "spillover" blocks from large regional epidemics. Fourthly, of the central blocks, there are three (blocks 7, 14, and 17) which appear to act as centroids of beetle infestation activity. Fifthly, the region acts in a coordinated fashion with the overall epidemic apparently affecting all of the individual block behaviors. Sixthly, the northeastern section (blocks 3, 4, 8, and 9) exhibit erratic patterns, perhaps due to underlying inhomogeneous forest areas. Seventhly, the strong environmental component is evident from both modes of analysis, as reflected in their stability. Eighthly, mode three and mode four blocks as defined in the reference curve analyses appear to differ between themselves only in the timing of their patterns, not in the overall shape.

Any model, while not required to explicitly include these details, must be capable of recreating them under proper conditions.

In a simlar manner, the results of the space-time series analyses offer a list of required components also. While some of the items duplicate those listed above, it should be remembered that the two analyses were done independently and differ in their fundamental approaches. The present analyses emphasize the local behavior of blocks in the overall system. Ten features can be listed, including the existence of:

1. at least three block types
2. a strong environmental trend
3. blocks which appear to be epidemic overflows
4. a spatial contiguity of similar block types
5. a short range spatial flow
6. an underlying cyclic process
7. a large number of types of individual block behaviors
8. a unique section in the northwestern blocks
9. very incoherent behavior in the northeast
10. different cycles for different block types.

The emphasis of this list is on block-to-block relations rather than the block-to-standard curve comparisons above.

These two sets of features must be combined into a simple general framework as specified by the model structure set forth in chapter two. An adequate model should be capable of reproducing nearly all of the observed behaviors as listed above. This is necessary because all the analyses were done with this underlying formulation in mind. Thus, the submodel must have a regular incomplete block mosaic of twenty-four patches. The specification of internal block behavior must be done in terms of environmental qualities which must be completely determined by block identity, not by interblock behavior. Only beetle intensity can be transferred, and this transfer must be completely described by the characteristics of the inter-block connections. And finally, the regional behavior must be merely the collection of these block interactions and cannot depend on any higher hierarchical level. Any model which satisfies the three set of constraints is a candidate for theupper echelon submodel at this point.

Given the 24 block mosaic structure with the limitations imposed by the constraints listed above and the lack of a lower level submodel, there is much that can be accomplished in the way of estimating a sufficient model. The approach followed in this paper is to first estimate the total regional structure through multiple linear regression analysis. If these regression models are stable over time (if they fit well), then reintroduction of the random residual effect coupled with some imposed boundaries should offer a suitable stochastic simulation model for the entire region. The full validation of such a model cannot be done at present due to the absence of another 15 to 18 years of regional data. However, two types of partial validation may be calculated, one on the regression system itself and one on the overall behavior of the simulations. The details and philosophy of these attempts are discussed throughout the chapter. By adopting the regression approach, this paper essentially bases its model on the local level of explanation (identical to the one of the autocorrelative studies). The structure of the overall hierarchical model nearly demands this.

Multiple linear regression assumes a model of the type $Y_i = B_1 X_{1,i} + \dots + B_k X_{k,i} + e_i$. The e_i term is assumed to be normally distributed with constant mean and fixed positive variance. The $X_{j,i}$ values are assumed to be fixed and measured without error. Use of such a model allows (among many other features) estimation of the B_i values, tests of the model fit, and tests of the hypothesis $B_i = 0$ for any i. It is these three items which prove useful in the model building procedure. The computer package SAS was used for all the calculations, employing the PROCEDURE GLM (Helwig and Council, 1979).

For each multiple regression fitting, each block is considered as a separate dependent variable. The independent variables are taken to be all neighboring block values of infestation intensity at one year ago, with the last three time lag values of the block being used as the dependent variable. Certain of the blocks show peculiarities in their data (linear dependencies which introduced singularity into the regression matrix), and so not all models had access to the same number of dependent variables. The largest model possible is $N_{i,t} = a_i N_{i,t-1} + b_i N_{j,t-1} + c_i N_{k,t-1} + d_i N_{l,t-1} + e_i N_{m,t-1} + f_i N_{i,t-2} + g_i N_{i,t-3}$. A full description of the results of the twenty-four regressions in terms of their parameter and independent variable values is given in Tables 6.1 and 6.2. The selection of lag order in space and time issues directly from the autocorrelative studies of Chapter 5, and the model therefore is based on this previous work.

The regression equations estimated in this manner are then judged as to their appropriateness solely on the basis of their R^2 values. All equations fit very well by this measure, with 16 values above .90. The lowest value occurred for block 24 at R^2 equal to .70. Several equations rated above .90 in their values. (See Table 6.3).

Before passing on to the simulation aspects of the model, it is good to examine the patterns of significance

TABLE 6.1--Multiple linear regression equations for blocks one through twelve

$N_1(t) = -2.9858\ N_1(t-1) + .8942\ N_2(t-1) + .0916\ N_6(t-1)$

$N_2(t) = -8.3369\ N_1(t-1) + 2.7441\ N_2(t-1) + .0519\ N_3(t-1)$
$\quad + .0056\ N_7(t-1)$

$N_3(t) = .6441\ N_3(t-1) - .5051\ N_2(t-1) + .3122\ N_4(t-1) +$
$\quad .2679\ N_8(t-1) - .1811\ N_3(t-2) - .2718\ N_3(t-3)$

$N_4(t) = .8001\ N_4(t-1) + .3202\ N_3(t-1) - .2285\ N_9(t-1) -$
$\quad .7525\ N_4(t-2) + .6837\ N_4(t-3)$

$N_5(t) = .9334\ N_5(t-1) + .1089\ N_6(t-1) - .0192\ N_{12}(t-1) -$
$\quad 1.9654\ N_5(t-2)$

$N_6(t) = .9635\ N_6(t-1) - 2.1550\ N_5(t-1) - .0198\ N_7(t-1) +$
$\quad .1114\ N_{13}(t-1) - .2915\ N_6(t-2) + .5013\ N_6(t-3)$

$N_7(t) = .4593\ N_7(t-1) - 7.3232\ N_2(t-1) + .3478\ N_6(t-1) +$
$\quad .0709\ N_8(t-1) + .2602\ N_{14}(t-1) - .0173\ N_7(t-2) +$
$\quad .6526\ N_7(t-3)$

$N_8(t) = .9929\ N_8(t-1) - .2904\ N_3(t-1) + .5800\ N_7(t-1) +$
$\quad .3971\ N_9(t-1) - .2939\ N_{15}(t-1) - .4412\ N_8(t-2) -$
$\quad 2756\ N_8(t-3)$

$N_9(t) = .3349\ N_9(t-1) + .9832\ N_4(t-1) - .0021\ N_8(t-1) -$
$\quad .2233\ N_9(t-2) + .0700\ N_9(t-3)$

$N_{10}(t) = 7.4083\ N_{10}(t-1) + .0941\ N_{11}(t-1) - .0099\ N_{16}(t-1) -$
$\quad 57.0304\ N_{16}(t-2)$

$N_{11}(t) = .8304\ N_{11}(t-1) + .9956\ N_{10}(t-1) + .3267\ N_{12}(t-1) +$
$\quad .0514\ N_{17}(t-1) - .3479\ N_{11}(t-2) + .5959\ N_{11}(t-3)$

$N_{12}(t) = .6928\ N_{12}(t-1) - 1.3831\ N_5(t-1) + .5621\ N_{11}(t-1) +$
$\quad .0861\ N_{13}(t-1) - .0395\ N_{18}(t-1) - .9250\ N_{12}(t-2) +$

TABLE 6.1--Continued

$.1795\ N_{12}(t-3)$

TABLE 6.2--Multiple linear regression equations for blocks thirteen through twenty-four

$$N_{13}(t) = -.3339\ N_{13}(t-1) - 1.5609\ N_6(t-1) - 1.1841\ N_{12}(t-1) + .5188\ N_{14}(t-1) + 1.9092\ N_{19}(t-1) - .2278\ N_{13}(t-2) - .0151\ N_{13}(t-3)$$

$$N_{14}(t) = .7931\ N_{14}(t-1) + .5713\ N_7(t-1) + .4545\ N_{13}(t-1) - .4435\ N_{15}(t-1) - 3.1282\ N_{20}(t-1) - .1251\ N_{14}(t-2) - .1377\ N_{14}(t-3)$$

$$N_{15}(t) = .0396\ N_{15}(t-1) + .4317\ N_8(t-1) + .1689\ N_{14}(t-1) + .3639\ N_{21}(t-1) - .7001\ N_{15}(t-2) + .0435\ N_{15}(t-3)$$

$$N_{16}(t) = .1309\ N_{16}(t-1) + 1.1992\ N_{10}(t-1) + .3927\ N_{17}(t-1) - .0029\ N_{22}(t-1) + .0430\ N_{16}(t-2) + .2405\ N_{16}(t-3)$$

$$N_{17}(t) = .3803\ N_{17}(t-1) - .9727\ N_{11}(t-1) + .9762\ N_{16}(t-1) + .1902\ N_{18}(t-1) + .0303\ N_{23}(t-1) + .0074\ N_{17}(t-2) + .3151\ N_{17}(t-3)$$

$$N_{18}(t) = .2810\ N_{18}(t-1) + .2594\ N_{12}(t-1) + .5369\ N_{17}(t-1) + .2764\ N_{19}(t-1) - 9.1464\ N_{24}(t-1) + .8687\ N_{18}(t-2) - .6347\ N_{18}(t-3)$$

$$N_{19}(t) = .8943\ N_{19}(t-1) - .1234\ N_{13}(t-1) + .5803\ N_{18}(t-1) + .4860\ N_{20}(t-1) - .4378\ N_{19}(t-2) - .3035\ N_{19}(t-3)$$

$$N_{20}(t) = .7276\ N_{20}(t-1) + .0379\ N_{14}(t-1) - .1033\ N_{19}(t-1) - 2.6803\ N_{21}(t-1) + 1.3268\ N_{20}(t-2) + .6265\ N_{20}(t-3)$$

TABLE 6.2--Continued

$$N_{21}(t) = 2.2893\ N_{21}(t-1) + .0047\ N_{15}(t-1) + .1146\ N_{20}(t-1) - 2.2185\ N_{21}(t-2) - 2.3974\ N_{21}(t-3)$$

$$N_{22}(t) = .3308\ N_{22}(t-1) + .2290\ N_{16}(t-1) + .0809\ N_{23}(t-1) + .7256\ N_{22}(t-2) - .2414\ N_{22}(t-3)$$

$$N_{23}(t) = .4289\ N_{23}(t-1) + .0731\ N_{17}(t-1) + .6869\ N_{22}(t-1) - .0432\ N_{24}(t-1) - .2996\ N_{23}(t-1) - .1482\ N_{23}(t-1)$$

$$N_{24}(t) = -.0796\ N_{24}(t-1) + .0416\ N_{18}(t-1) + .0654\ N_{23}(t-1) + .4510\ N_{24}(t-2) - .0400\ N_{24}(t-3)$$

TABLE 6.3--Multiple regression R^2 values and standard deviations for all twenty-four regressions

Block Number	R^2	Standard Deviation
1	.9099	.3875
2	.7510	2.1384
3	.7796	18.2467
4	.7882	12.9293
5	.9278	.3968
6	.9409	3.0630
7	.9448	15.4985
8	.8675	22.1487
9	.8308	13.2707
10	.9968	.3795
11	.9729	3.5860
12	.9350	2.8649
13	.9095	16.3023
14	.9283	16.7310
15	.7383	14.9945
16	.9894	3.5055
17	.9076	19.3612
18	.9177	12.7020
19	.8735	12.6316
20	.8781	2.9945
21	.9916	.4510
22	.9719	4.8806
23	.9765	3.6868
24	.6966	2.3501

for the B_i values in each regression. A full display of each parameter and its significance level (as measured by the probability of obtaining a greater t-test value is given in Tables 6.4 and 6.5. Most blocks show significant

TABLE 6.4--T-test significance levels for B parameters of the regressions

Dependent Variable	Independent Variable	Significance Level
$N_1(t)$	$N_1(t-1)$	.0001
	$N_2(t-1)$	.0001
	$N_6(t-1)$	.0024
$N_2(t)$	$N_1(t-1)$	.0029
	$N_2(t-1)$	.0027
	$N_3(t-1)$	.3110
	$N_7(t-1)$	.8637
$N_3(t)$	$N_3(t-1)$	.2818
	$N_2(t-1)$	.8935
	$N_4(t-1)$	.6937
	$N_8(t-1)$	.3280
	$N_3(t-2)$	.8080
	$N_3(t-3)$	.6953
$N_4(t)$	$N_4(t-1)$	.0659
	$N_3(t-1)$	.4058
	$N_9(t-1)$	.4743
	$N_4(t-2)$	.1231
	$N_4(t-3)$	.1476

TABLE 6.4--Continued

Dependent Variable	Independent Variable	Significance Level
$N_5(t)$	$N_5(t-1)$	.0254
	$N_6(t-1)$	.0111
	$N_{12}(t-1)$	.5202
	$N_5(t-2)$	.0048
$N_6(t)$	$N_6(t-1)$	.0292
	$N_5(t-1)$	.3850
	$N_7(t-1)$	.7854
	$N_{13}(t-1)$	.1103
	$N_6(t-2)$	.5458
	$N_6(t-3)$	.3833
$N_7(t)$	$N_7(t-1)$	.3390
	$N_2(t-1)$	.0649
	$N_6(t-1)$	.8400
	$N_8(t-1)$	.8506
	$N_{14}(t-1)$	.5409
	$N_7(t-2)$	.9715
	$N_7(t-3)$	.3416
$N_8(t)$	$N_8(t-1)$	.0506
	$N_3(t-1)$	.6089
	$N_7(t-1)$	.3077
	$N_9(t-1)$	.5102
	$N_{15}(t-1)$	.7198
	$N_8(t-2)$	.4532

TABLE 6.4--Continued

Dependent Variable	Independent Variable	Significance Level
	$N_8(t-3)$	.5588
$N_9(t)$	$N_9(t-1)$	.1680
	$N_4(t-1)$	.0044
	$N_8(t-1)$	.9912
	$N_9(t-2)$	.5685
	$N_9(t-3)$	.8812
$N_{10}(t)$	$N_{10}(t-1)$	.0001
	$N_{11}(t-1)$	.0074
	$N_{16}(t-1)$	.3208
	$N_{10}(t-2)$	.0001

TABLE 6.5--T-test significance levels for B parameters of the regressions

Dependent Variable	Independent Variable	Significance Level
$N_{11}(t)$	$N_{11}(t-1)$	.0314
	$N_{10}(t-1)$	.1303
	$N_{12}(t-1)$	.3972
	$N_{17}(t-1)$	.4660
	$N_{11}(t-2)$	.5310
	$N_{11}(t-3)$	.3128
$N_{12}(t)$	$N_{12}(t-1)$	.2716
	$N_5(t-1)$	.6994
	$N_{11}(t-1)$	.0840

TABLE 6.5--Continued

Dependent Variable	Independent Variable	Significance Level
	$N_{13}(t-1)$	.4652
	$N_{18}(t-1)$	.7073
	$N_{12}(t-2)$	.0687
	$N_{12}(t-3)$	.8115
$N_{13}(t)$	$N_{13}(t-1)$	.4585
	$N_{6}(t-1)$	.4317
	$N_{12}(t-1)$	.6203
	$N_{14}(t-1)$	.0541
	$N_{19}(t-1)$	.0378
	$N_{13}(t-2)$	.5758
	$N_{13}(t-3)$	.9743
$N_{14}(t)$	$N_{14}(t-1)$	.0515
	$N_{7}(t-1)$	.1786
	$N_{13}(t-1)$	.1419
	$N_{15}(t-1)$	.3126
	$N_{20}(t-1)$	.0772
	$N_{14}(t-2)$	.7560
	$N_{14}(t-3)$	.6983
$N_{15}(t)$	$N_{15}(t-1)$	.9127
	$N_{8}(t-1)$	.4307
	$N_{14}(t-1)$	.7071
	$N_{21}(t-1)$	.9306
	$N_{15}(t-2)$	.3887

TABLE 6.5--Continued

Dependent Variable	Independent Variable	Significance Level
	$N_{15}(t-3)$	.9409
$N_{16}(t)$	$N_{16}(t-1)$	.5473
	$N_{10}(t-1)$	.0135
	$N_{17}(t-1)$	.0008
	$N_{22}(t-1)$	.9832
	$N_{16}(t-2)$	.8312
	$N_{16}(t-3)$	.2436
$N_{17}(t)$	$N_{17}(t-1)$	.6424
	$N_{11}(t-1)$	.7008
	$N_{16}(t-1)$	.3804
	$N_{18}(t-1)$	.7968
	$N_{23}(t-1)$	.9809
	$N_{17}(t-2)$	.9935
	$N_{17}(t-3)$	.7768
$N_{18}(t)$	$N_{18}(t-1)$	.6618
	$N_{12}(t-1)$	.8433
	$N_{17}(t-1)$	.2177
	$N_{19}(t-1)$	.5865
	$N_{24}(t-1)$	.0734
	$N_{18}(t-2)$	.1425
	$N_{18}(t-3)$	.1207
$N_{19}(t)$	$N_{19}(t-1)$	.2464
	$N_{13}(t-1)$	.7851

TABLE 6.5--Continued

Dependent Variable	Independent Variable	Significance Level
	$N_{18}(t-1)$	.2480
	$N_{20}(t-1)$	.8920
	$N_{19}(t-2)$	.6145
	$N_{19}(t-3)$	.7558
$N_{20}(t)$	$N_{20}(t-1)$	.5603
	$N_{14}(t-1)$	.3408
	$N_{19}(t-1)$	.6571
	$N_{21}(t-1)$	.2061
	$N_{20}(t-2)$	.0922
	$N_{20}(t-3)$	.4048
$N_{21}(t)$	$N_{21}(t-1)$	.0001
	$N_{15}(t-1)$	.7237
	$N_{20}(t-1)$	.0242
	$N_{21}(t-2)$	.0226
	$N_{21}(t-3)$	.0908
$N_{22}(t)$	$N_{22}(t-1)$	.3786
	$N_{16}(t-1)$	.1728
	$N_{23}(t-1)$	.8309
	$N_{22}(t-2)$	.2212
	$N_{22}(t-3)$	.5826
$N_{23}(t)$	$N_{23}(t-1)$	.2356
	$N_{17}(t-1)$	.3273
	$N_{22}(t-1)$	.0144

TABLE 6.5--Continued

Dependent Variable	Independent Variable	Significance Level
	$N_{24}(t-1)$	.9637
	$N_{23}(t-2)$	.4502
	$N_{23}(t-3)$	.5990
$N_{24}(t)$	$N_{24}(t-1)$	.9048
	$N_{18}(t-1)$	.4278
	$N_{23}(t-1)$	.3647
	$N_{24}(t-2)$	.4068
	$N_{24}(t-3)$	.9321

values for the parameter weighting their own value one lag back. Also common is the significance of at most two neighboring block values even though some blocks have up to four possibilities. Some oddities occur in blocks having no significant values and in some having all of them significant. Neither of the two other lagged values show much consistency in their significance, but in those regressions that do show this the lag at two steps back clearly outweighs the more removed value. It must be remembered in looking at these significant values that the multicollinearity of the data destroys any clear interpretation. In general, one should not try to infer any causal structure from these sets of coefficients.

The high R^2 values for the regressions argue that such equations might be capable of forming the basis for a more far-reaching model. Just such a step can be taken by combining all the separate equations into a set of 24 simultaneous equations. This system requires the values of all 24 blocks for 3 consecutive years for initialization. The extent to which this model can be utilized directly is limited but not at all minimal, as will be shown below. The problems which beset this approach are those typical of multiple regression studies in general, such as correlated independent variables and linearity. Even worse is the switching of dependent variables into the role of independent variables and back again. There are two arguments against such objections, one based on the effects of collinearity and one based on the overall effectiveness of the approach. Wesolowsky (1976) summarizes his discussion of the matter by claiming that biasing effects are the most dangerous effects of collinearity, but that for predictive purposes such biasing might be a good thing. The overall effectiveness of the approach may be judged from later discussion in this chapter.

As an initial validation test for the full regression system, the following scheme is used. First, the system is initialized with three consecutive years of observed data. Then predictions are made using the equations for one, two or more steps ahead (with a maximum of fifteen steps). These predictions are subtracted from the observed values for the appropriate block and year to obtain residuals. These 24 residuals are stored, and a new three year initial set is chosen for the next run. In this manner all possible initializations can be chosen and used to calculate means and variances for all sets of predictions at one step ahead, two steps, and so on. The variances give an estimate of the stability of the full regression model to different initial values. If the equations are a poor model, then the variances should quickly become large. If the model is stable, it should give reasonable variances for a short number of lags. For any such regression model, the variances should rapidly go to large numbers before many steps are taken due to the continual compounding of estimation errors through time and space.

The 24 equation regression model seems reasonably stable when subjected to the above form of analysis. The variances for one, two, and three step predictions are quite moderate. The four step formulation shows difficulty only in one block, and the five step scheme blows up in only 4 or 5 blocks. For actual prediction purposes, it is recommended to only predict up to three years. For this paper's purposes, the stability of the model over four or five prediction steps is an indication that the model can serve as an adequate base for a simulation model. Table 6.6 shows the means and variances for one, two, and three step prediction validations.

TABLE 6.6--Results of the validation studies on the full regression model

Block	One step		Two step		Three step	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
1	-.0620	.3526	-.1621	1.0882	.5777	1.0684
2	-.2989	1.8699	.3450	3.9160	3.0138	4.3864
3	17.9180	17.7296	31.0543	22.7092	37.4146	26.8512
4	3.3360	10.3672	12.5850	15.3834	18.4646	15.8989
5	-.0567	.3480	-.1250	.5558	-.0315	.5704
6	-.0880	2.4550	.2350	2.7715	.6092	3.1660
7	-.9367	11.6757	2.5714	12.2449	1.6092	36.8856
8	3.9500	16.2352	4.9707	22.3790	2.4277	31.5639
9	.1087	11.2150	3.9314	11.2972	11.8554	19.8979
10	-.0213	.3357	-.2193	2.8522	-.3077	3.3811
11	-.0907	2.8731	-.2807	3.0972	-.3000	4.2336
12	-.1220	2.1617	-.0571	3.4127	-.1023	4.2953
13	2.5533	12.0365	3.3486	23.4428	4.2108	29.8699
14	.9200	12.6129	1.6921	11.3084	2.4377	21.6887
15	-.6820	12.0014	2.1378	13.3765	4.3155	21.3323
16	-1.0047	4.4704	-1.8671	8.6566	-1.6670	5.8090
17	-.2793	14.6324	.3264	17.4017	2.6108	22.2171
18	2.2973	9.3026	7.6485	14.5887	11.7923	16.7083
19	-.4053	10.1191	.1964	16.0135	4.7953	21.1370
20	-.0546	2.4003	.2643	3.0824	.5346	4.0923
21	-.0373	.3799	-.1264	.7259	-.2808	.9503
22	.3100	4.1130	.3457	4.2558	.1269	3.7063
23	.0280	2.9566	.5707	4.1385	.5523	4.2456
24	-.2040	1.9751	-.0636	2.0368	.0100	2.2225

A second assessment uses the individual residuals from the regression as computed above for one, two, and three step prediction schemes. The 24 residuals for each year of prediction (15 in the one step scheme) are examined as a group for certain distributional properties with the goal of developing some region of confidence around the predicted behaviors. An examination of the residuals using PROC UNIVARIATE of SAS (Helwig and Council, 1979) shows two important features. Firstly, although most of the residual sets can be approximated by a symmetric, normal-like distribution about zero, there are certain years in which such a conclusion is not warranted. Tables 6.7, 6.8, and 6.9 give the population characteristics for each of the residual sets and for each of the prediction steps. A second result is displayed in Figure 6.1 which shows the standard deviations of these populations of residuals for each step length over the years. It is worth noting that the three-step variances worsen near the end of the sequence of years, perhaps explaining why the larger step lengths turn out unsatisfactory answers. Finally, a two-sided band is placed around the means of the residual populations to give the 95th and 5th percentiles of the populations. Thus, Figures 6.2, 6.3, and 6.4 yield a type of confidence interval about the predictive abilities of the regression equations. The overall conclusion from all these analyses is that the presented regression model is

TABLE 6.7--Distribution characteristics of the residuals from the one step prediction scheme

Year	N	Mean	Standard Deviation	Skewness	Kurtosis	Maxi-mum	Median	Mini-mum	Range
4	24	.88	6.88	2.22	9.26	27	0.0	-11	38
5	24	-1.88	8.26	-.15	2.89	18	0.0	-24	42
6	24	.83	5.18	.99	2.28	16	0.0	-9	25
7	24	1.58	8.04	1.01	2.32	23	0.0	-15	38
8	24	.25	6.37	-.10	1.20	14	0.0	-14	28
9	24	1.67	6.49	.64	.49	16	0.0	-10	26
10	24	1.71	5.74	.76	.67	15	0.0	-8	23
11	24	-3.08	8.16	-2.10	7.39	9	-1.0	-33	42
12	24	.29	4.76	.32	2.16	13	0.0	-11	24
13	24	7.58	12.83	1.41	1.11	41	0.0	-5	46
14	24	4.75	11.54	1.13	1.15	31	2.0	-16	40
15	24	-.83	7.36	1.30	3.88	22	-1.5	-16	38
16	24	3.50	10.72	1.58	2.84	32	1.0	-12	44
17	24	1.88	16.84	2.04	8.11	64	0.0	-30	94
18	24	-1.83	9.47	-1.58	3.97	14	0.0	-32	46

TABLE 6.8--Distribution characteristics of the residuals from the two step prediction scheme

Year	N	Mean	Standard Deviation	Skewness	Kurtosis	Maxi-mum	Median	Mini-mum	Range
5	24	.67	7.22	1.72	3.32	22	-1.0	-9	31
6	24	-.50	8.31	1.49	3.16	25	-1.5	-13	38
7	24	3.04	9.16	.19	-.18	21	0.0	-14	35
8	24	.96	6.79	.85	4.03	22	0.0	-14	36
9	24	.88	10.69	-.06	2.05	26	0.0	-27	53
10	24	1.54	6.83	.00	1.41	18	0.5	-13	31
11	24	-.63	8.30	-.62	1.32	14	-0.5	-22	36
12	24	-1.00	9.70	.12	1.03	19	-2.0	-21	40
13	24	9.46	14.73	1.04	-.45	42	0.5	-3	45
14	24	14.83	20.08	1.01	-.33	59	6.5	-7	66
15	24	7.08	14.65	1.45	1.97	51	0.0	-7	58
16	24	1.92	13.41	1.92	4.31	43	0.0	-16	59
17	24	.21	19.70	2.58	10.82	78	1.0	-30	108
18	24	2.13	16.47	.06	4.54	49	0.0	-44	93

TABLE 6.9--Distribution characteristics of the residuals from the three step prediction scheme

Year	N	Mean	Standard Deviation	Skewness	Kurtosis	Maxi-mum	Median	Mini-mum	Range
6	23	1.79	8.82	1.10	2.57	28	0.0	-14	42
7	23	2.30	9.24	.16	.18	21	0.0	-18	39
8	23	2.52	10.71	-.42	1.60	25	0.0	-26	51
9	23	3.09	10.52	-.25	2.06	24	1.0	-26	50
10	23	1.27	9.65	-1.20	3.98	18	1.5	-29	47
11	23	-2.22	10.97	-.79	1.00	17	0.0	-29	46
12	23	.65	9.57	1.03	2.08	28	0.0	-15	43
13	23	8.35	20.41	.02	1.53	56	4.0	-44	100
14	23	18.22	22.23	.68	-1.04	60	5.0	-9	69
15	23	16.87	20.69	.95	-.17	67	7.0	-4	71
16	23	13.17	22.37	1.31	.42	65	4.0	-9	74
17	23	-.26	23.40	1.86	7.50	84	2.0	-39	123
18	23	-5.43	31.97	-1.69	3.79	50	1.0	-97	147

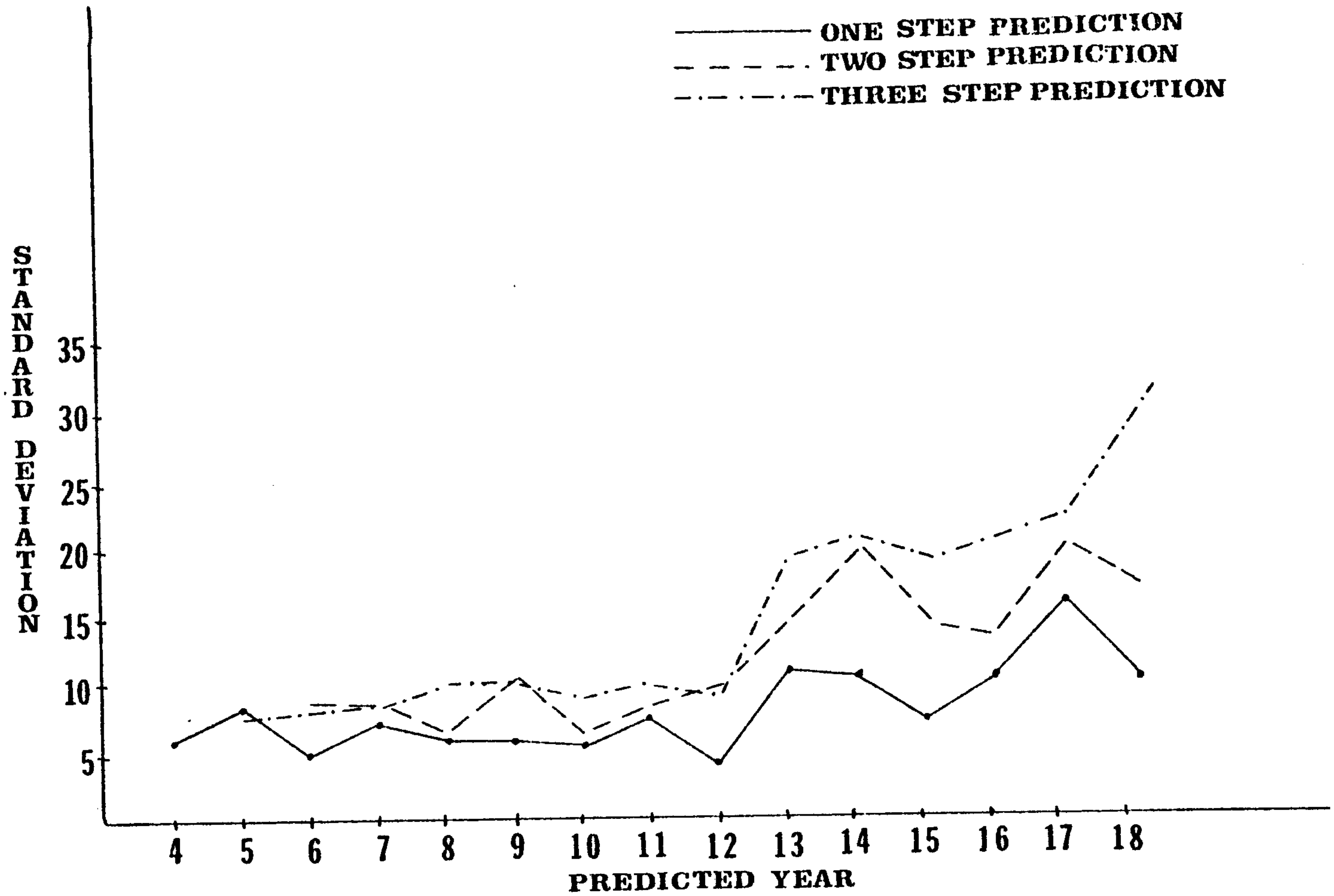

Figure 6.1--Standard deviations of the residuals over time

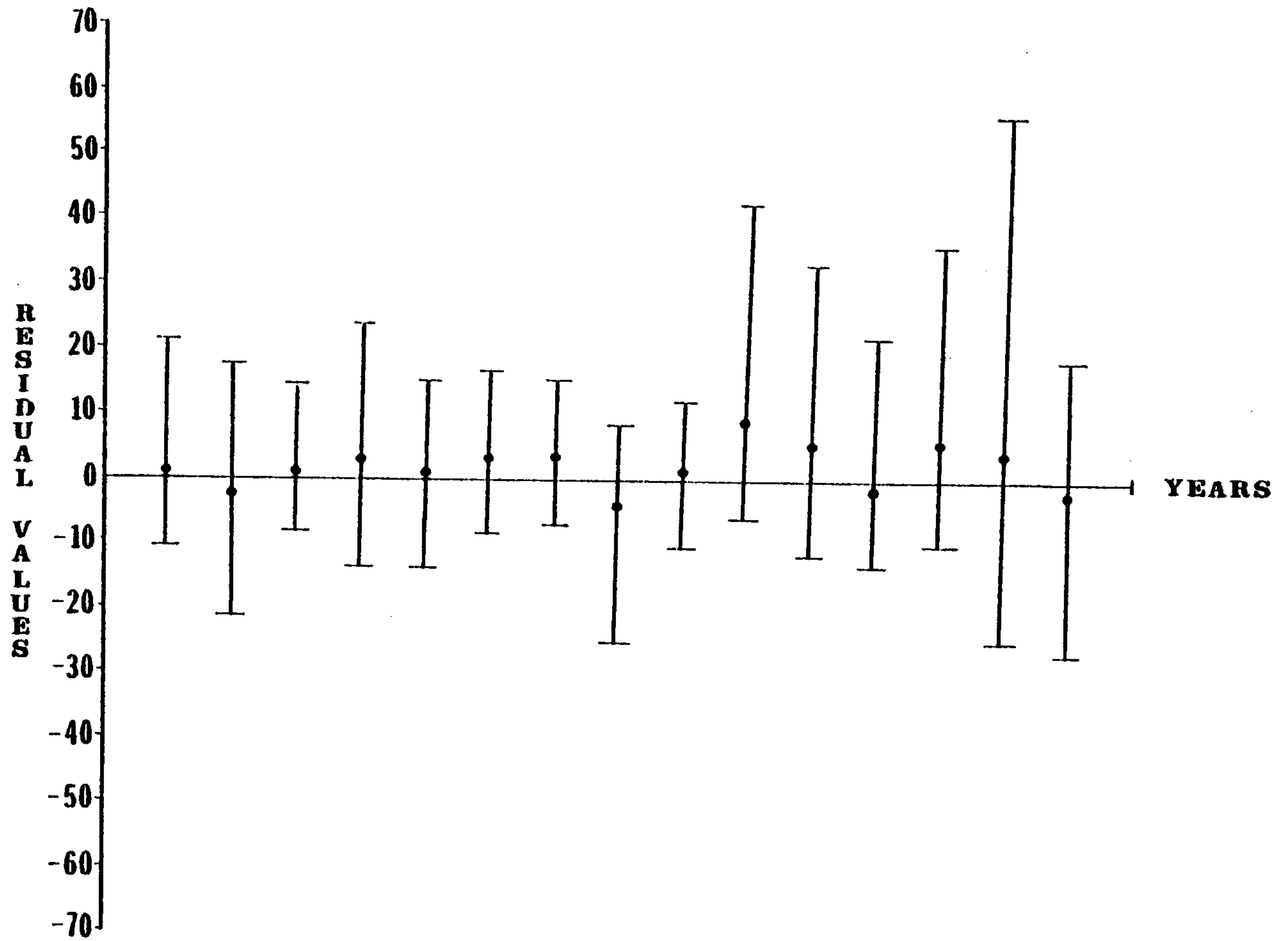

Figure 6.2--Ninety-five percent population bounds about the means of the one step prediction scheme residuals

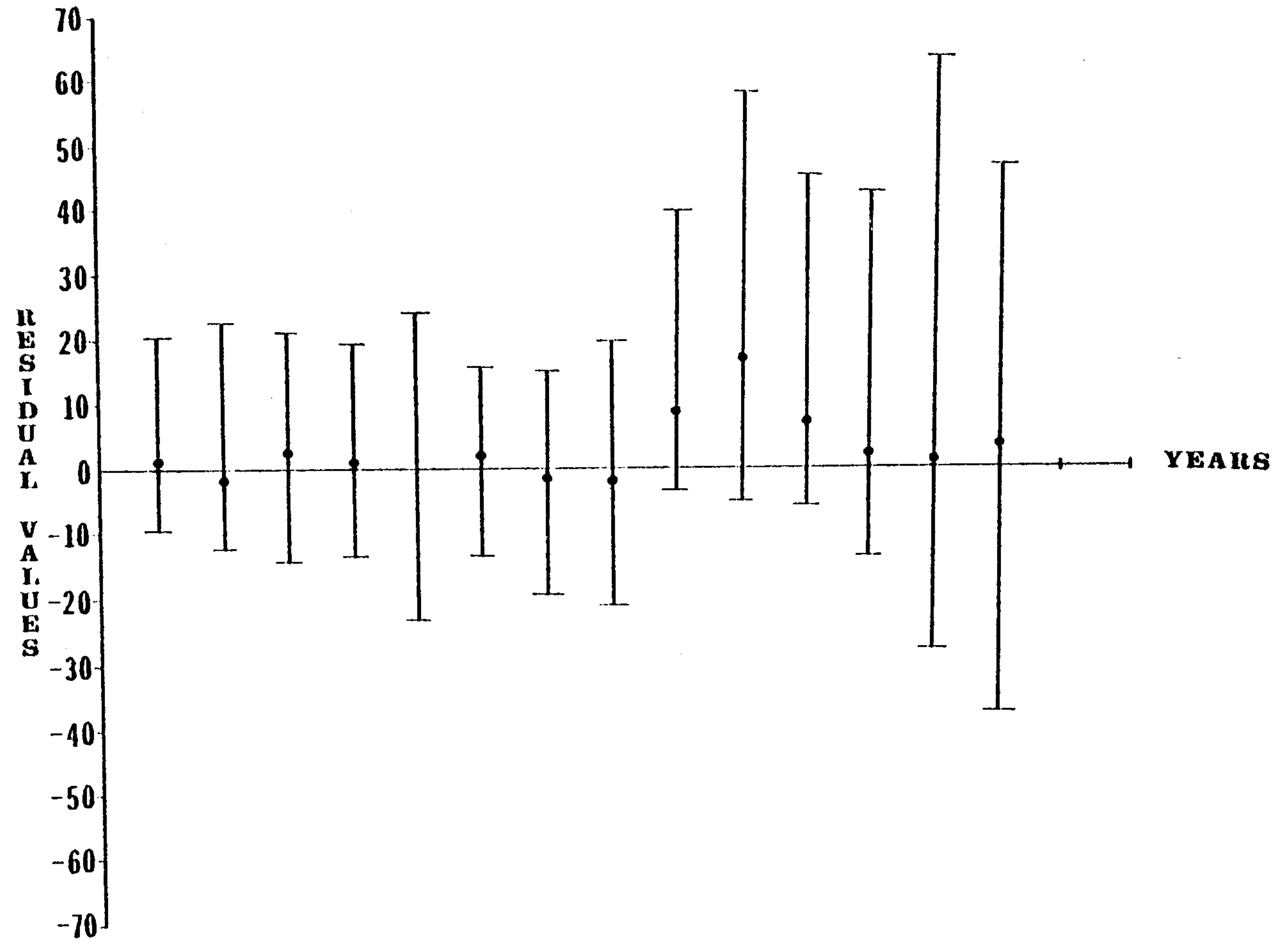

Figure 6.3--Ninety-five percent population bounds about the means of the two step prediction scheme residuals

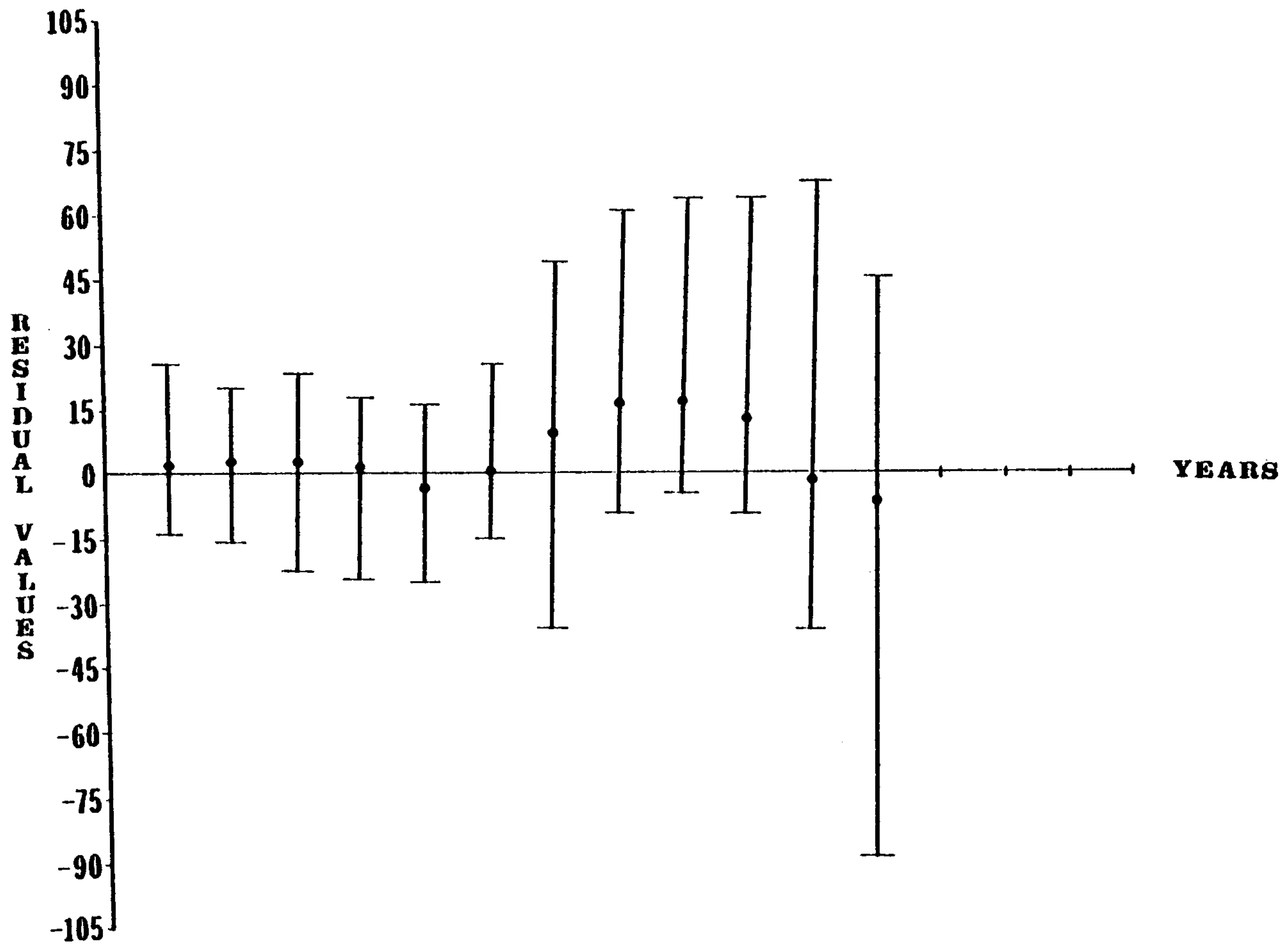

Figure 6.4--Ninety-five percent population bounds about the means of the three step prediction scheme residuals

useful over short periods. More importantly, with proper modification to alleviate the more severe problems, it should serve well as a basic model for further work.

A third evaluation examines the stability of the multiple regression equations. Three years are chosen for deletion from each set of 18 years used in the regressions. Because of the structure of the equations, only four 3-point sets are deletable in this manner (if one demands a full 15 points for further regression work). One of these sets allows no prediction to be made because it deletes years one through three. The remaining three sets are useful and are used to predict one-step ahead predicted values for each equation. Thus deletion of years 1, 2, and 18 allows a prediction of year 18 which may be compared to the observed value at that year. The error in prediction forms the basis for testing the stability of the equations under changes of data. These residuals can be tested using a t-test for testing predicted individual values such as described in Snedecor and Cochran (1978). In this paper, the variance was estimated as s^2 (1 + 1/n-1), since only a rough notion of significance is required. Large numbers of significant values indicates a general instability in the equations, while patterns of significance indicate biasing effects. The results obtained from such an analysis are given in Tables 6.10, 6.11, and 6.12. These tables demonstrate no obvious patterns but do show significant results

TABLE 6.10--Details of omission/prediction test of sufficiency on the regression equations for blocks one through eight

Block #	Year Predicted	Predicted Value	Observed Value	Residual	Standard Deviation	Significance at 5%?
1	18	6.26	0	-6.26	.37	yes
1	17	1.92	4	2.08	.36	yes
2	18	47.35	0	-47.35	2.05	yes
2	17	2.96	11	8.04	1.97	yes
3	18	72.67	0	-72.67	11.86	yes
3	17	-.16	52	52.16	11.26	yes
3	16	27.89	48	20.11	10.88	no
4	18	72.15	40	-32.15	12.08	yes
4	17	48.17	4	-44.17	11.04	yes
4	16	54.26	4	-50.26	9.83	yes
5	18	5.93	4	-1.93	.38	yes
5	17	2.13	2	-.13	.37	no
6	18	14.27	14	-.27	3.06	no
6	17	27.21	23	-4.21	2.91	no
6	16	10.72	21	10.28	2.35	yes
7	18	63.23	20	-43.23	10.87	yes
7	17	119.08	82	-37.08	10.80	yes
7	16	116.23	86	-30.23	10.32	yes
8	18	84.34	11	-73.35	17.63	yes
8	17	92.43	70	-22.43	17.26	no
8	16	99.01	77	-22.01	16.54	no

TABLE 6.11--Details of omission/prediction test of sufficiency on the regression equations for blocks nine through sixteen

Block #	Year Predicted	Predicted Value	Observed Value	Residual	Standard Deviation	Significance at 5%?
9	18	-54.02	0	54.02	11.60	yes
9	17	50.38	4	-46.38	9.04	yes
9	16	65.40	75	9.60	8.86	no
10	18	128.15	14	-114.15	.36	yes
10	17	2.17	17	14.83	.35	yes
11	18	-2.65	48	50.65	3.48	yes
11	17	32.49	28	-4.49	3.30	no
11	16	12.60	27	14.40	2.37	yes
12	18	-31.35	17	48.35	2.40	yes
12	17	54.91	16	-38.91	2.27	yes
12	16	-18.76	8	26.76	2.15	yes
13	18	5.01	11	5.99	16.75	no
13	17	13.92	20	6.08	16.85	no
13	16	236.02	78	-158.02	11.47	yes
14	18	98.48	28	-70.48	15.21	yes
14	17	47.80	72	24.20	17.79	no
14	16	43.09	83	39.91	16.23	yes
15	18	9.29	0	-9.29	14.98	no
15	17	154.20	19	-135.20	8.18	yes
15	16	26.48	34	7.52	7.82	no
16	18	144.08	53	-91.08	6.24	yes
16	17	39.22	52	12.78	6.17	no
16	16	35.12	36	.88	6.29	no

TABLE 6.12--Details of omission/prediction test of sufficiency on the regression equations for blocks seventeen through twenty-four

Block #	Year Predicted	Predicted Value	Observed Value	Residual	Standard Deviation	Significance at 5%?
17	18	-68.93	81	149.93	10.61	yes
17	17	109.99	45	-64.99	6.31	yes
17	16	79.73	81	1.27	6.39	no
18	18	-74.50	39	113.50	7.57	yes
18	17	37.26	16	-21.26	7.13	yes
18	16	132.13	61	-71.13	5.15	yes
19	18	39.46	2	-37.46	12.09	yes
19	17	-59.23	28	87.23	7.48	yes
20	18	26.99	3	-23.99	2.06	yes
20	17	-252.99	16	268.99	.64	yes
20	16	187.04	12	-175.04	.61	yes
21	18	13.80	11	-2.80	.43	yes
21	17	15.44	6	-9.44	.41	yes
21	16	6.00	6	0.00	.95	no
22	18	45.00	34	-11.00	5.28	no
22	17	23.39	39	15.61	3.59	yes
22	16	23.45	22	-1.45	3.63	no
23	18	31.23	30	-1.23	3.70	no
23	17	16.50	22	5.50	3.59	no
23	16	9.79	20	10.20	3.40	yes
24	18	-.35	9	9.35	.90	yes
24	17	3.07	3	-.07	.90	no
24	16	7.15	5	-2.15	.90	no

in about two-thirds of the residuals. This is a weakness but seems to be directly attributable to the larger magnitudes of the observations at these later years. Thus, the problem does not seem to discredit the equations to any great extent but rather warns that caution should be taken when missing data occurs during epidemic periods.

A two-fold modification of the simultaneous regression model forms the stochastic simulation model offered as a submodel for the upper echelon level of the beetle hierarchy. The first modification is the estimation and re-addition of the random residual term removed by the regressions. The second change is the placement of upper and lower boundaries on the values of infestation intensity within each block. These estimates derive from the pine distribution properties of each area. Validation of this simulation model demonstrates the adequacy of the form to explain past beetle behavior and its capability of extension to longer range studies.

The random error term is assumed to be normally distributed by the regression model, and this fact makes it possible to mimic its effects. The specific value of the residual will differ from the actual error in all but rare instances but, on the average, the effects should be equivalent. The International Mathematical and Statistical Libraries subroutine GGNML is used to produce normally distributed pseudo-random variates with mean zero and unit variance. By multiplying each generated residual by the standard deviation estimated from each multiple linear regression, it is possible to achieve the appropriate variance for each normal population. These simulated residuals are then added back into the equation predicting each block infestation intensity. The resulting system is a stochastic simulation model for the beetle regional dynamics.

One more adjustment is necessary to account for the natural limitations placed on intensity by its definition as a proportion. Both upper and lower limits must be chosen, with the obvious choice for the lower limit being zero. If the simulation model achieves or exceeds the upper or lower limit in any block at any time, then the value for that block is set equal to that lower or upper level. The upper limit is less easy to estimate, but it certainly depends on the pine covered area within each block. Using data from the Nelson and Zillgitt work, some estimates can be made. The values obtained from doing this are shown in Table 6.13. However, some blocks achieved observed intensities which exceeded the estimate obtained in this manner when the original data is examined. For these few cases, the maximum value of the observed intensity is used to set an appropriate upper limit. The actual limits used in all subsequent simulations are shown in Table 6.14. Better understanding of forest dynamics could

TABLE 6.13--Estimates of pine tree spatial distribution by percent of land area forested

Block	Percentage
1	6.25
2	31.25
3	50.00
4	68.75
5	12.50
6	25.00
7	81.25
8	50.00
9	56.25
10	43.75
11	81.25
12	62.50
13	93.75
14	81.25
15	75.00
16	50.00
17	93.75
18	81.25
19	87.50
20	87.50
21	56.25
22	25.00
23	31.25
24	18.75

improve this estimate in some cases.

The simulation model as described above was run twenty-five times (using twenty-five different random number seeds) using three initial years (1960 through 1963) for a period representing twenty consecutive years of data. Testing consists of two analyses, one relating total intensity patterns over time between simulated and real data and one utilizing the space-time techniques discussed in chapters four and five. Each of these approaches is treated in more detail below.

TABLE 6.14--Estimates of upper boundary of block infestation intensity as used in the simulations

Block	Percentage
1	10.00
2	30.00
3	60.00
4	70.00
5	10.00
6	30.00
7	90.00
8	80.00
9	80.00
10	40.00
11	80.00
12	60.00
13	90.00
14	90.00
15	80.00
16	60.00
17	90.00
18	80.00
19	90.00
20	90.00
21	60.00
22	40.00
23	30.00
24	20.00

The first evaluation procedure examines the ability of the simulations to track the observed regional epidemic. Figure 6.5 charts five simulations versus the observed total intensity (defined as the sum of the 24 block intensities for a fixed year). These five trials can be considered as typical of the full set of simulations. All of the simulations take on values of the right magnitude but differ in their display of the intensity values. Most of the artificial data shows the proper peak at year 13 (1975), but most also miss tracking the consistent low

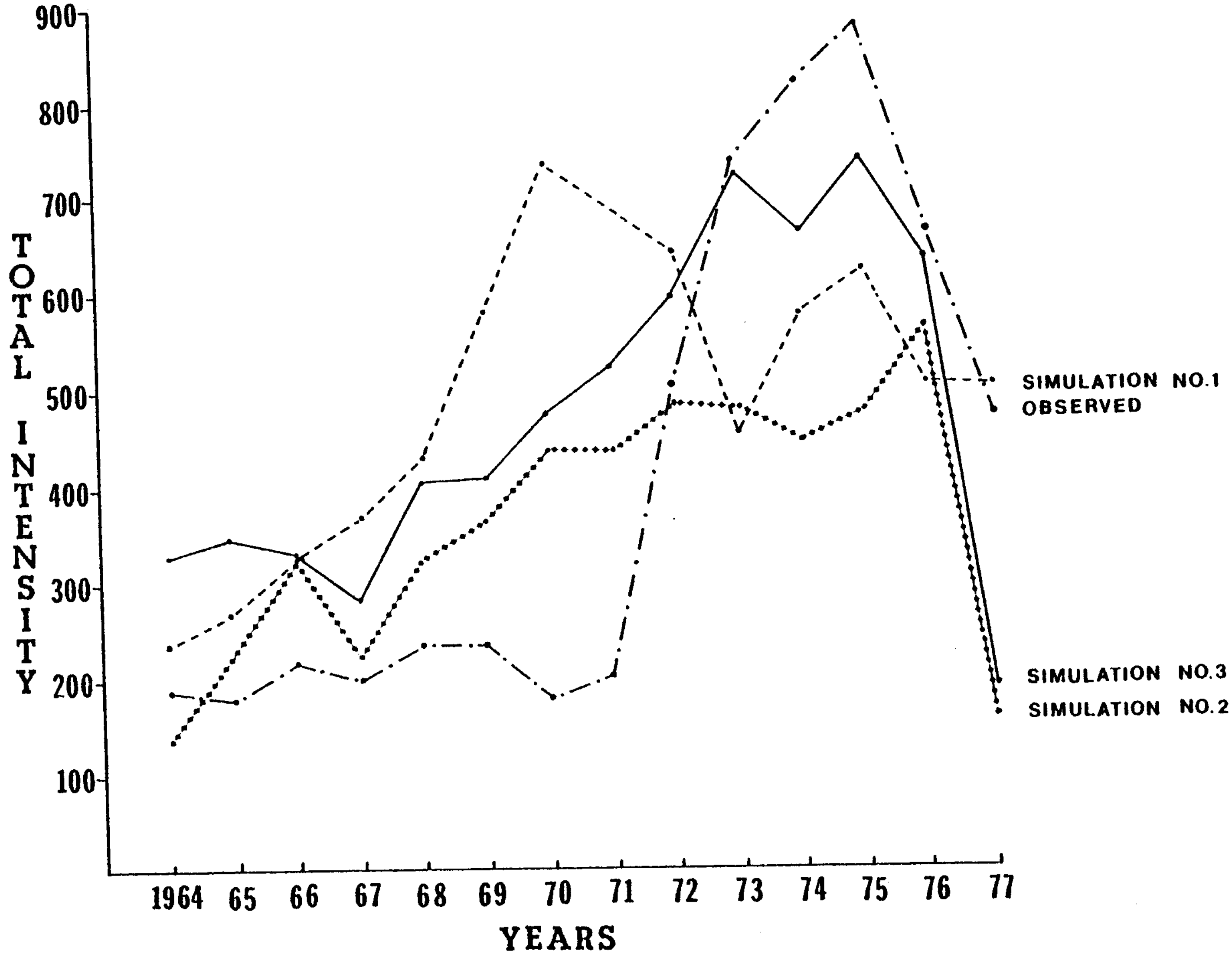

Figure 6.5--Plots of actual and simulated total intensities

nature of the first few years. Considering the innate difficulty of comparing realizations from a complicated random process, there is still the question of whether or not the actual data can be considered as an unsurprising realization from the same stochastic process which generated the simulated data. Such a process would be considered sufficient to account for the observed data.

To test the sufficiency of the simulation process as described above, a statistic based on pairs of realizations is used. This statistic consists of taking the squared differences between all points on the simulation curves for the 15 years corresponding to the actual years 1964 through 1977 (to insure comparability to the observed data). Thus, if a_i represents the value of the regional intensity at year i (i runs from 1 to 15) for simulation a and b_i a similar quantity for simulation run b, then the statistic S equals the sum over i of $(a_i - b_i)^2$. A reasonable estimate of the distribution of S for the simulated populations can be gotten by calculating S for all possible pairs of the twenty-five simulation runs. In particular, the mean, variance, and the shape of the distribution should command attention. The results of running such an analysis (using PROC UNIVARIATE of SAS) are shown in Table 6.15. With this parent distribution estimated, the observed sequence of regional intensities can be paired with all 25 simulations to yield 25 values of S which can be tested for significance. Using

TABLE 6.15--Parent distribution characteristics for the S statistic

Characteristic	Value
N	300
Mean	405,591
Standard Deviation	301,350
Skewness	1.8663
Kurtosis	4.5122
Range	1,761,615
Maximum	1,821,660
Upper Quartile	512,831
Median	312,795
Lower Quartile	195,751
Minimum	60,045
Interquartile Range	317,080

the 95th and 5th percentile bands on the parent distribution of S, a test of the significance of the simulation-observation pairs yields only 5 significant values, which is certainly not enough to reject this sample as being vastly unlike the underlying distribution. (See Table 6.16). In this sense, the simulation is sufficient to explain the data.

Two types of space-time analysis were run on the twenty-five sets of simulated data, reference curve analysis and autocorrelative estimation in time. The five blocking patterns represented in Figures 6.6 through 6.10 are typical examples of the reference curve analyses. A measure of sufficiency in terms of these blocking patterns is developed below. Similarly, the twenty-five autocorrelative analyses can be used to produce blocking patterns like those presented previously. Figures 6.11

TABLE 6.16--Significance testing of the simulations using the S statistic

Simulation #	Value	Significant at 5%?
1	588,544	no
2	757,251	no
3	754,998	no
4	390,158	no
5	1,219,079	yes
6	973,839	no
7	391,313	no
8	1,020,788	yes
9	650,420	no
10	1,412,173	yes
11	510,487	no
12	592,303	no
13	1,122,791	yes
14	831,952	no
15	802,417	no
16	1,287,747	yes
17	496,895	no
18	541,671	no
19	473,278	no
20	568,362	no
21	1,178,938	yes
22	861,050	no
23	925,599	no
24	394,548	no
25	945,371	no

through 6.15 show five examples of these patterns. The measurement of model sufficiency in terms of these analyses is also presented below.

The reference curve blocking patterns involve four modal types being assigned to each of the twenty-four blocks involved in the simulation. A measure of similarity between two blocking patterns can be based on how many blocks are classified into the same mode. Thus, the variable a_i can represent the model class of block i

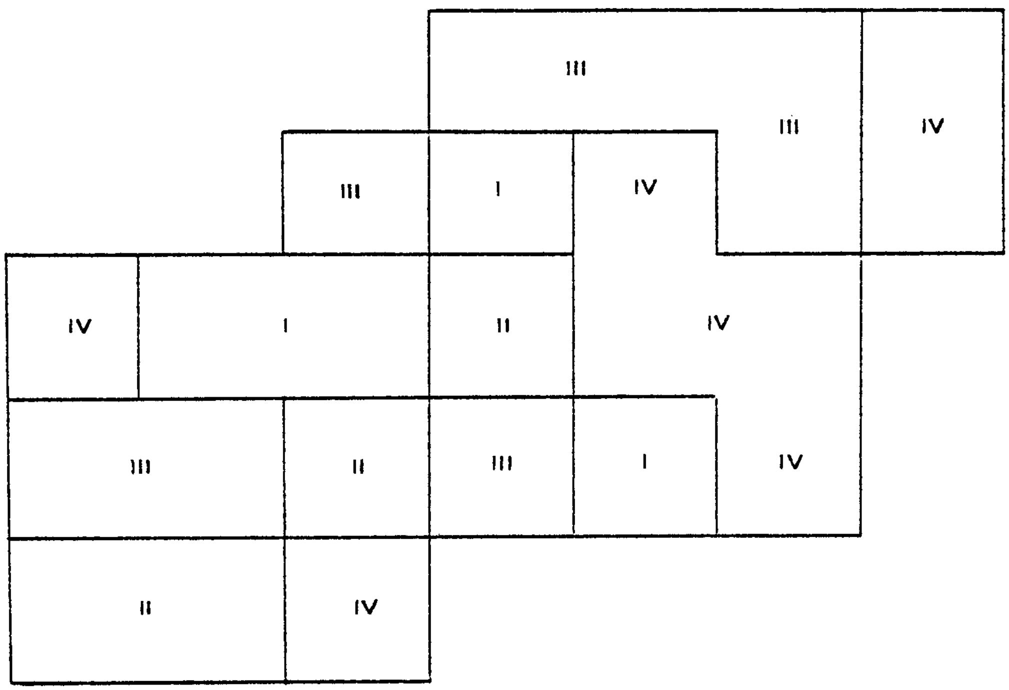

Figure 6.6--Blocking pattern suggested by the reference curve analysis of the first set of simulated data

The roman numeral within the area indicates the component followed.

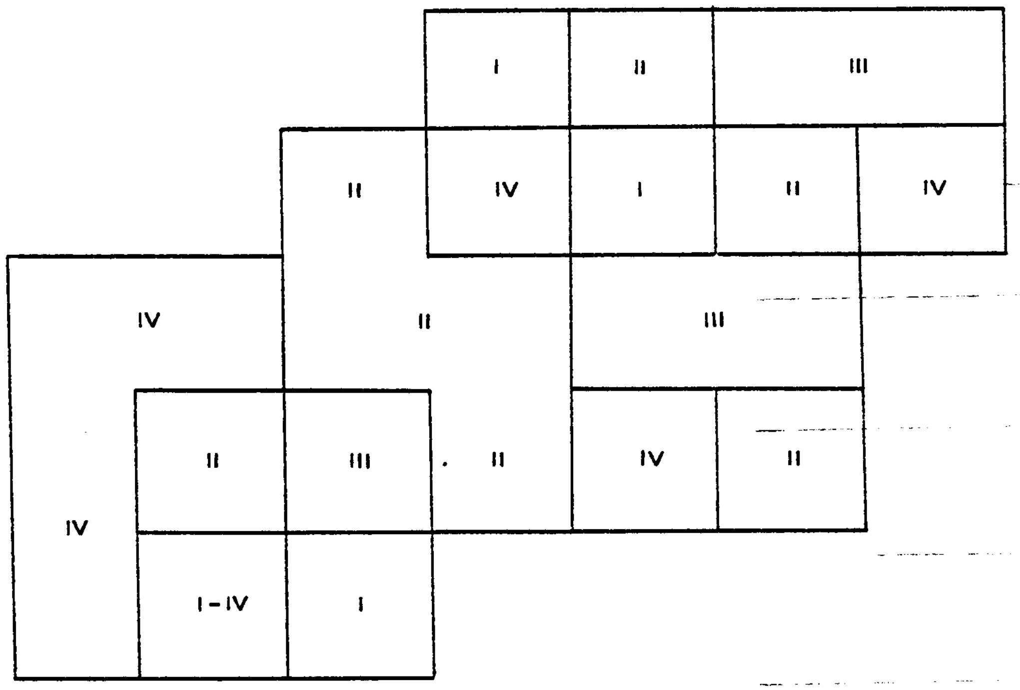

Figure 6.7--Blocking pattern suggested by the reference curve analysis of the second set of simulated data

The roman numeral within the area indicates the component followed.

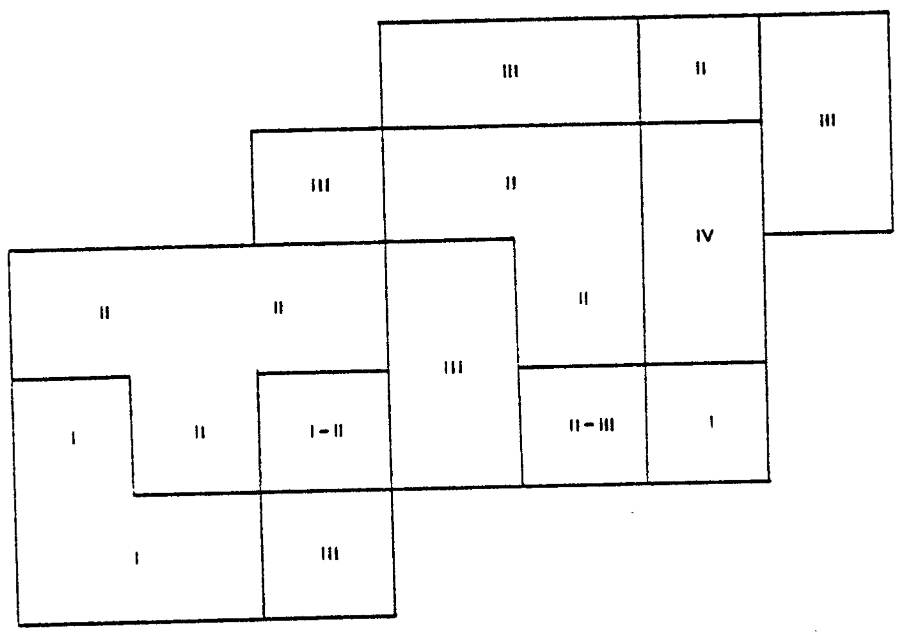

Figure 6.8--Blocking pattern suggested by the reference curve analysis of the third set of simulated data

The roman numeral within each area indicates the curve followed.

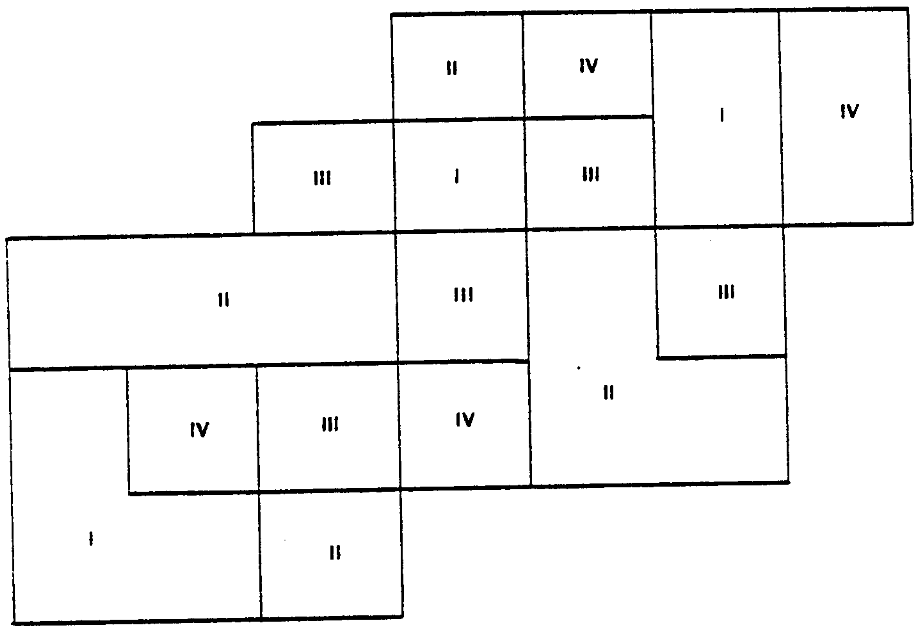

Figure 6.9--Blocking pattern suggested by the reference curve analysis of the fourth set of simulated data

The roman numeral within each area indicates the curve followed.

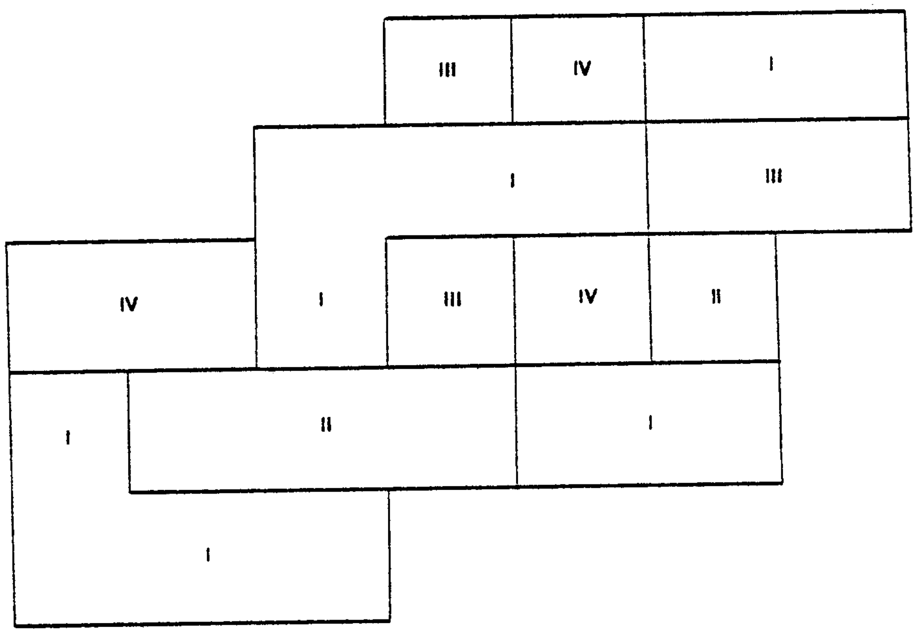

Figure 6.10--Blocking pattern suggested by the reference curve analysis of the fifth set of simulated data

The roman numeral within each area indicates the curve followed.

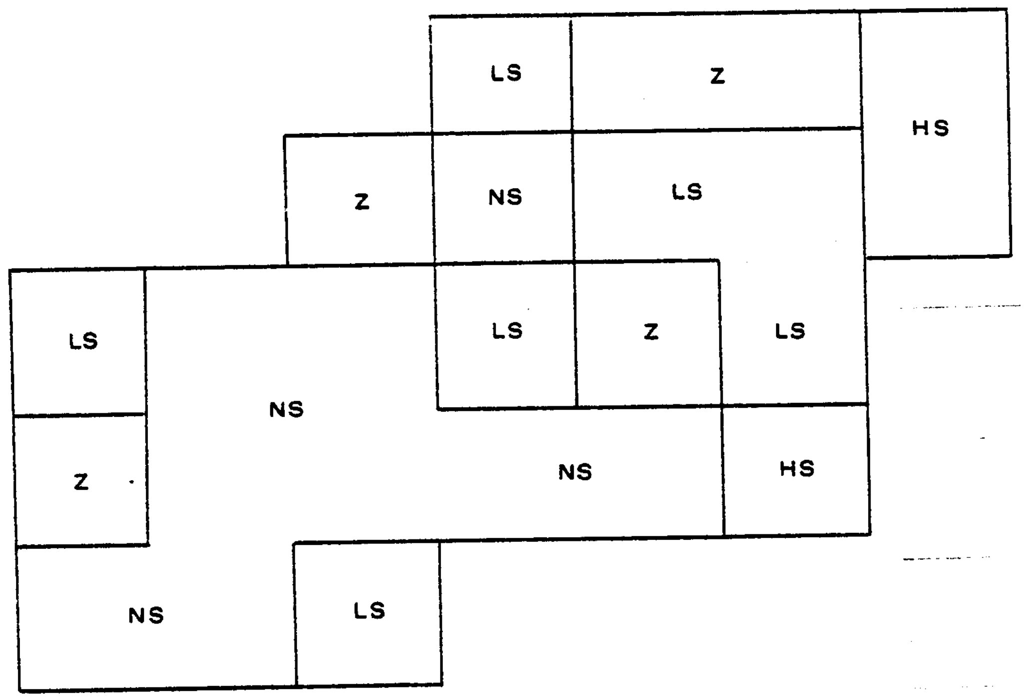

Figure 6.11--Blocking pattern suggested by the temporal autocorrelative structure of the first set of simulated data

Z: area resembling a zero order process
NS: area resembling a non-stationary process
LS: area resembling a stationary low order process
HS: area resembling a stationary high order process

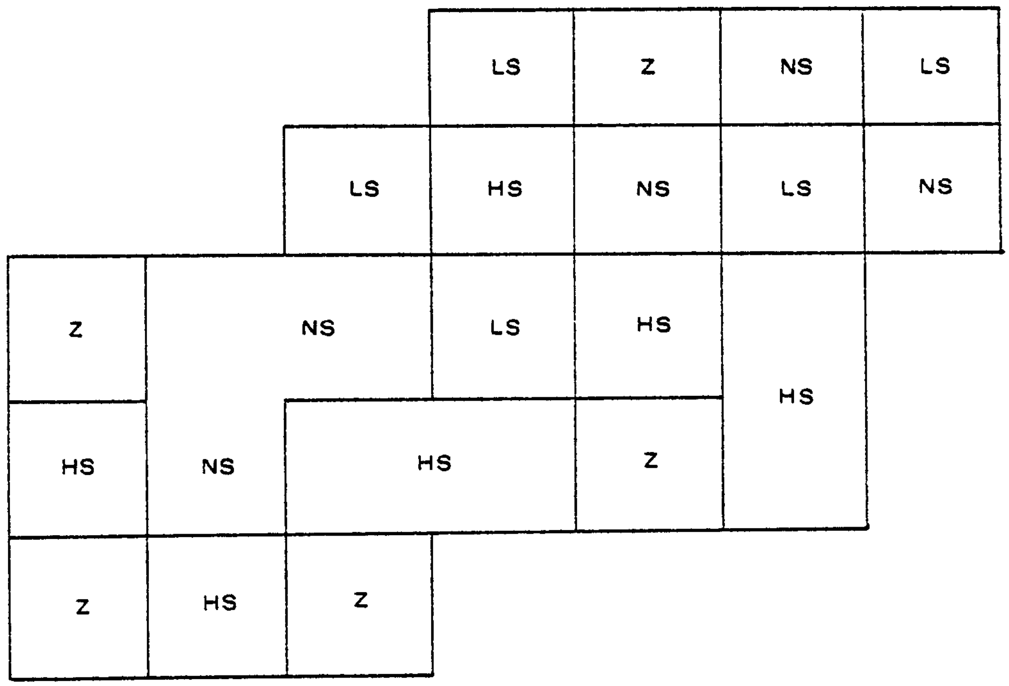

Figure 6.12--Blocking pattern suggested by the temporal autocorrelative structure of the second set of simulated data

Z: area resembling a zero order process
LS: area resembling a stationary low order process
HS: area resembling a stationary high order process
NS: area resembling a non-stationary process

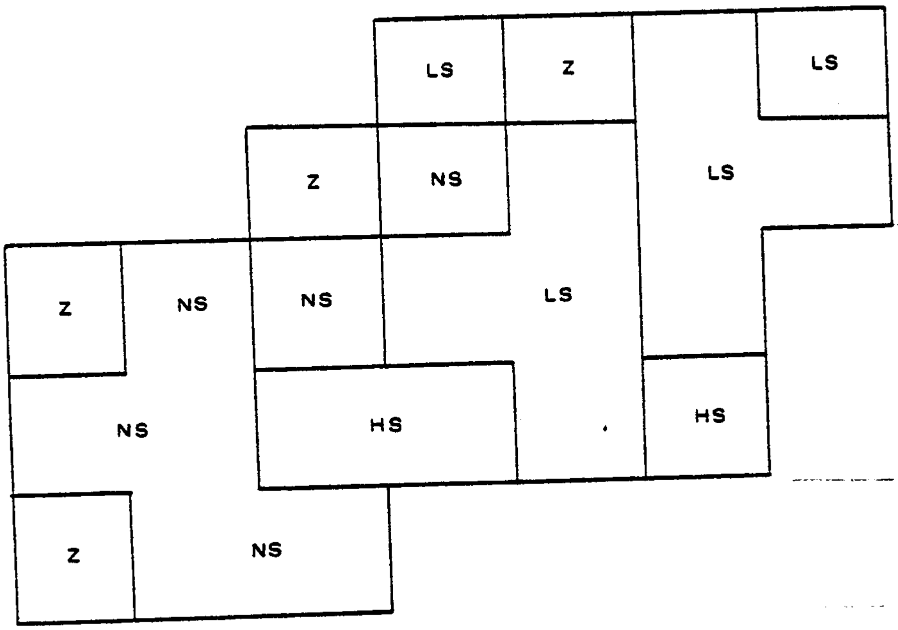

Figure 6.13--Blocking pattern suggested by the temporal autocorrelative structure of the third set of simulated data

Z: area resembling a zero order process
LS: area resembling a stationary low order process
HS: area resembling a stationary high order process
NS: area resembling a nonstationary process

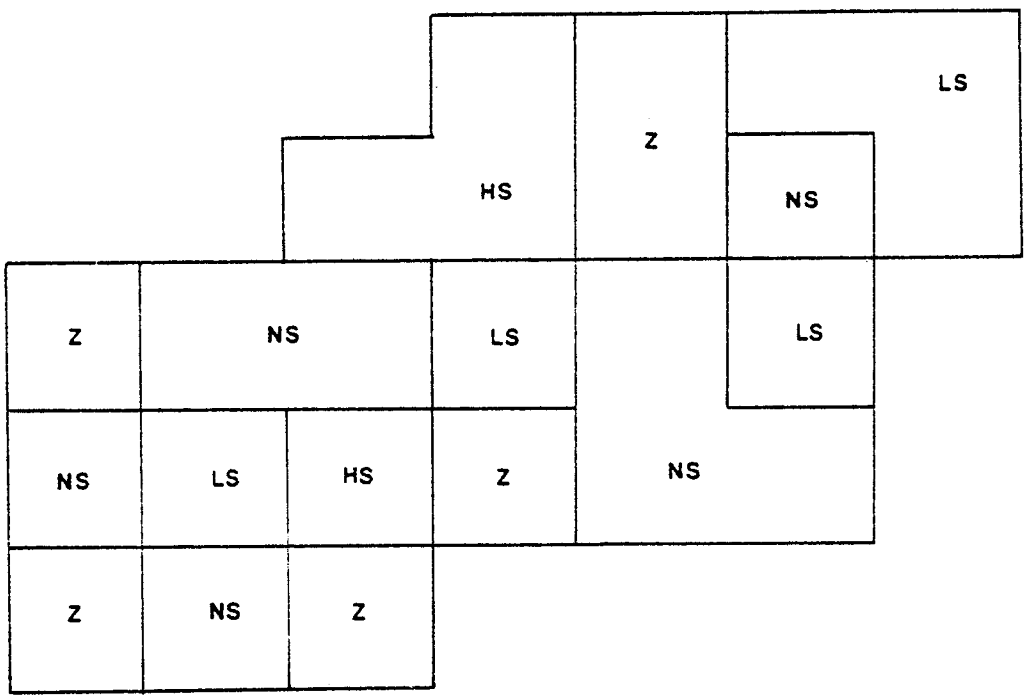

Figure 6.14--Blocking pattern suggested by the temporal autocorrelative structure of the fourth set of simulated data

Z: area resembling a zero order process
LS: area resembling a stationary low order process
HS: area resembling a stationary high order process
NS: area resembling a non-stationary process

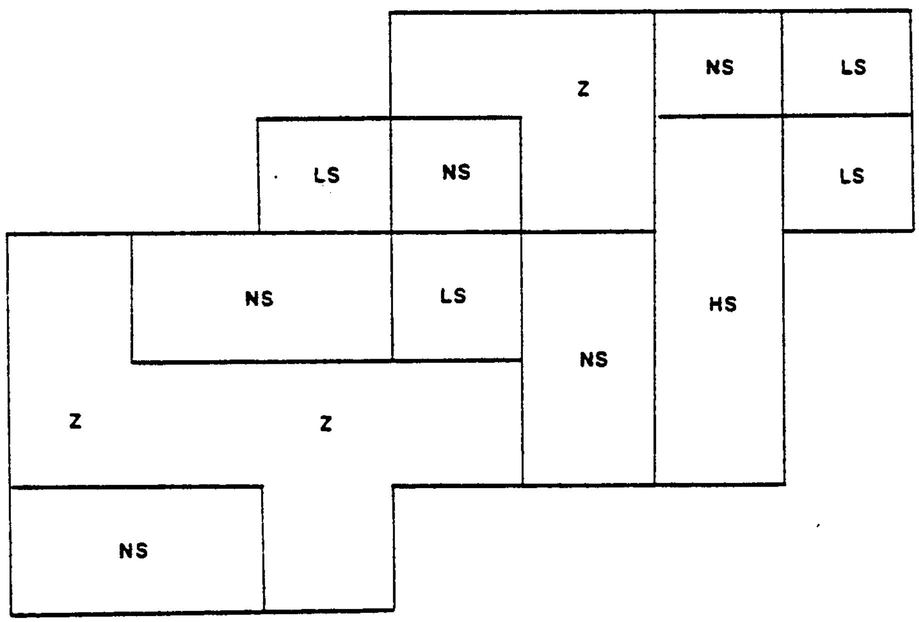

Figure 6.15--Blocking pattern suggested by the temporal autocorrelative structure of the fifth set of simulated data

Z: area resembling a zero order process
LS: area resembling a stationary low order process
HS: area resembling a stationary high order process
NS: area resembling a nonstationary process

(where i runs from 1 to 24) for simulation a. The values of a_i are restricted to one of four values (1, 2, 3, or 4) and likewise for b_i. The statistic T is set equal to one if $a_i = b_i$ and is set to zero otherwise. Then the sum of these <u>T</u> values over the 24 blocks serves as the basic statistic <u>sum(T)</u> which can be computed for all possible pairs of simulations to estimate the parent distribution. The results of this analysis are shown in Table 6.17 and

TABLE 6.17--Parent distribution characteristics of the sum(T) statistic

Characteristic	Value
N	300
Mean	7.7
Standard Deviation	2.31
Skewness	.100
Kurtosis	.028
Range	14
Maximum	15
Upper Quartile	9
Median	8
Lower Quartile	6
Minimum	1
Interquartile range	3

show that the distribution is symmetric and unimodal. The calculation of the sum(T) statistic for the 25 observed versus simulation pairings is the basis for a significance test which shows only one significant value at the 90 percent level. (See Table 6.18). It must be noted, however, that the expected value for the parent distribution is near to that which is expected from a purely random pairing of

TABLE 6.18--Significance testing of the simulations using the sum(T) statistic

Simulation #	Value	Significant at 5%?
1	3	no
2	7	no
3	6	no
4	9	no
5	5	no
6	6	no
7	8	no
8	2	no
9	7	no
10	1	yes
11	5	no
12	6	no
13	6	no
14	6	no
15	5	no
16	5	no
17	6	no
18	4	no
19	5	no
20	5	no
21	8	no
22	9	no
23	6	no
24	5	no
25	5	no

the block values. Thus, although the test is still valid, it is not powerful in testing against this important hypothesis.

A similar approach was taken in analyzing the blocking patterns based on the autocorrelative analyses. In this case, only three possibilities exist for the block values (1, 2, or 3). The T statistic and the sum(T) statistic are calculated as above for all possible pairs of simulations to estimate the parent distribution. This analysis is summarized in Table 6.19. Again, the distri-

TABLE 6.19--Summary of sufficiency testing using the sum(T) statistic on autocorrelative blockings

Characteristic	Value
N	300
Mean	12.83
Standard Deviation	2.31
Skewness	.229
Kurtosis	-.436
Range	12
Maximum	19
Upper Quartile	14
Median	13
Lower Quartile	11
Minimum	7
Interquartile range	3

bution is symmetric and unimodal. The sum(T) values calculated from the observed versus simulation pairs when tested for significance against this distribution show no odd cases. (See Table 6.20). Thus, again, the adequacy of the simulation model is upheld. Here again, however, the case of random sorting seems to be a viable alternative, although the probabilities are much harder to calculate in this case due to the unequal amounts of the three classificatory variables. The problem again is in the test's power.

With the model established in its validity (to some degree at least) by the many types of analysis described above, it remains only to examine the applications and extensions of the model formulation. The concluding chapter

TABLE 6.20--Significance testing of the simulations using the sum(T) statistic on autocorrelative blockings

Simulation #	Value	Significant at 5%?
1	15	no
2	11	no
3	16	no
4	13	no
5	14	no
6	16	no
7	14	no
8	13	no
9	16	no
10	12	no
11	15	no
12	14	no
13	13	no
14	13	no
15	17	no
16	7	yes
17	14	no
18	13	no
19	14	no
20	13	no
21	11	no
22	13	no
23	14	no
24	12	no
25	13	no

of this thesis attempts this task by offering two applications to the areas of surveillance and management.

CHAPTER VII

UTILIZATIONS AND EXTENSIONS

This chapter has three purposes: to illustrate use of the model, to examine its limitations, and to propose extensions to it. The viewpoint adopted is that presented in Holling (1978); i.e., that emphasis should be placed on the development of a rapport between developers and users of models. It is hoped that this chapter is a step in the right direction.

Any pest management problem has two basic needs from a model. They are an efficient surveillance scheme and a way to evaluate control decisions. For the southern pine beetle regional dynamics, the case is no different. The model's role in the creation of such programs is to define variables and to describe their changes in terms of system state changes. External to the model framework itself is the description of what effects perturbations have on the system and how these effects differ with point and strength of application. The perturbations used in surveillance studies are generally small in magnitude and random in point of application. Control problems commonly deal with non-random variations and larger magnitudes. Throughout such studies, the underlying stability of the model structure is taken for granted. However, if the perturbations cause adverse reactions, then even the framework of the model is in doubt. The studies described in this chapter serve both as utilizations of the model and as tests of its structural competence.

In chapter 4 it was mentioned that the reference curve analyses might provide the basis for an efficient and reliable surveillance scheme based on the different block behaviors. Specifically, the blocks following the primary reference curve in the 24 block case are highly correlated with the total intensity of the region (defined as the sum of the 24 block intensities for a given year). This immediately suggests several sampling policies which may be examined. The usefulness of such policies can be tested in terms of their prediction errors, their relative costs, and their validity under different infestation conditions.

Because of the strong inherent link between the mode one blocks and the regional intensity, it seems natural to use such blocks as the primary variables in the proposed schemes. The details of using one such block (block number 7) as a linear regressive predictor of total intensity has been discussed in chapter 4. The significant model and the R^2 value of .84 are considered adequate by most criteria. Using this basic approach, a test of the stability of the formulation under different beetle intensities is possible by an omission, estimation, and prediction procedure. First, the regression is run with the same model on only 17 data pairs. The deleted value of block 7 intensity is then entered into the equation to predict a new total intensity. But the deleted value of total intensity serves as the true value of that predicted quantity and can be used to form a residual value. This residual value should be normally distributed with mean zero and variance equal to the variance estimated in the regression. If all possible deletions are made, then the 18 possible residuals should test as nonsignificant in approximately 95 percent of the cases. A model displaying such residuals would be judged more stable than one with more significant values. Table 7.1 shows the predictions, the residuals, the R-squares, and the significance levels of such an analysis run on the simple linear regressive equation based on the output of block 7. The table also lists the standard deviations of both the set of residuals and the original regressions for comparability. An examination of this table supports the validity of the formulation.

One improvement of the basic equation mimics the reference curve analysis by trying to explain the overall pattern in terms of model types. One block from each of the four modes is chosen as closely following that behavioral form (blocks 7, 10, 2, and 9, respectively) for the purposes of forming a multiple linear regression equation for predicting total intensity. Judging by the relative amounts of variation that each mode explains, the blocks

TABLE 7.1--Summary of the one variable validation of the linear regressive equation using block 7 to predict total intensity

Year	Prediction	Residual	Standard Deviation	Significant at 5%?	R^2
1	62.25	-36.25	113.48	no	.821
2	139.93	-53.93	113.04	no	.828
3	65.42	72.58	112.45	no	.834
4	127.22	76.78	112.24	no	.838
5	223.56	-42.56	113.32	no	.833
6	215.17	-44.18	113.28	no	.833
7	276.12	-66.12	112.60	no	.837
8	195.14	-4.15	113.81	no	.832
9	192.01	36.99	113.44	no	.835
10	274.84	-44.84	113.25	no	.836
11	368.94	-193.94	102.89	no	.862
12	404.31	-202.31	101.90	no	.866
13	367.73	132.27	108.87	no	.846
14	716.14	13.86	113.76	no	.890
15	705.16	106.84	111.14	no	.804
16	749.00	119.00	110.67	no	.792
17	769.68	-111.68	110.94	no	.828
18	192.53	272.47	91.42	yes	.892

Full regression equation---Total = 40.1698 + 8.5922 * Block7

R^2 = .836 Significance level = .0001

Residual mean = 1.713

Residual standard deviation = 114.35

are added to the model in the order 7, 10, 2, and then 9. The three models formed in this manner are compared on the basis of their R^2 values, on their costs, and on their stability under the test mentioned above. Tables 7.2 through 7.4 summarize these analyses. The conclusion based on these examinations is that the mode two and mode three blocks add very little improvement compared to their extra expense and their validity problems. The full 4 block model seems to be a strong basis for further work, and such an extension is considered next.

Based on the above studies, two further regression formulations can be tested. The first is a type of two stage cluster sampling where continual surveillance of blocks 7 and 9 is supplemented with surveillance of blocks 10 and 2 only when the observed intensity in block 7 stays above eighty. Such a scheme is quite reasonable in view of the observed behavior of the beetles. An examination of this approach is considered in Table 7.5 with the usual values given. The second technique is to eliminate mode two and three blocks altogether, using just blocks 7 and 9 for the regression. The analysis of this scheme is portrayed in Table 7.6. Based upon an examination of these tables, it appears that the prime candidates for the predictive system are the one variable (block 7) model, the four variable (blocks 7, 10, 2, and 9) using the two-stage cluster approach, and the two variable (blocks 7 and 9)

TABLE 7.2--Validation summary of the linear regressive model of total intensity on blocks 7 and 10 combined

Year	Prediction	Residual	Standard Deviation	Significant at 5%?	R^2
1	61.47	-35.47	113.49	no	.846
2	136.22	-50.22	113.12	no	.870
3	64.52	73.48	112.32	no	.873
4	123.38	80.62	111.98	no	.860
5	216.80	-35.80	113.46	no	.824
6	208.71	-37.71	113.42	no	.825
7	223.25	-13.25	112.85	no	.811
8	189.19	1.82	113.83	no	.896
9	185.95	43.05	113.30	no	.899
10	266.31	-36.31	113.45	no	.833
11	358.95	-183.94	103.44	no	.539
12	393.53	-191.53	102.58	no	.845
13	353.48	146.52	107.37	no	.849
14	683.56	46.44	113.32	no	.844
15	666.13	145.87	108.75	no	.844
16	675.33	192.67	109.14	no	.847
17	1,082.41	-424.32	89.46	yes	.844
18	270.77	194.23	91.66	no	.846

Full regression equation:

Total = 40.0985 + 8.2091 * Block7 + 5.7353 * Block10

R^2 = .8465 Significance level = .0001

Residual mean = -4.658

Residual standard deviation = 147.51

TABLE 7.3--Validation summary of the linear regressive model of total intensity on blocks 7, 10, and 2 combined

Year	Prediction	Residual	Standard Deviation	Significant at 5%?	R^2
1	30.50	-4.51	108.10	no	.859
2	117.49	-28.49	107.99	no	.864
3	33.17	104.83	104.72	no	.875
4	256.76	-52.76	104.83	no	.877
5	209.88	-28.88	107.83	no	.869
6	200.65	-29.65	107.82	no	.869
7	267.70	-57.69	106.99	no	.872
8	178.64	12.36	108.06	no	.869
9	175.26	53.74	107.16	no	.872
10	266.31	-36.31	107.67	no	.871
11	371.29	-196.29	94.50	no	.899
12	411.48	-209.48	92.61	no	.904
13	366.52	133.48	102.07	no	.883
14	762.44	-32.44	107.87	no	.852
15	730.61	81.39	106.60	no	.843
16	607.96	260.05	92.03	yes	.875
17	1,178.06	-520.06	92.03	yes	.897
18	-95.06	560.06	92.03	yes	.905

Full regression equation:
Total = 11.0593 + 9.3661 * Block7 + 13.7058 * Block10 - 68.4217 * Block2

R^2 = .8715 Significance level = .0001

Residual mean = .520

Residual standard deviation = 208.65

TABLE 7.4--Validation summary of the linear regressive model on total intensity on blocks 7, 10, 2, and 9 combined

Year	Prediction	Residual	Standard Deviation	Significant at 5%?	R^2
1	38.09	-12.09	83.05	no	.923
2	93.64	-7.65	83.08	no	.926
3	25.28	112.72	77.48	no	.937
4	88.89	115.11	76.97	no	.939
5	163.77	17.23	82.97	no	.929
6	200.21	-29.21	82.72	no	.929
7	228.36	-18.36	82.95	no	.929
8	227.13	-36.13	82.52	no	.930
9	282.79	-53.79	81.95	no	.931
10	319.38	-89.38	79.41	no	.935
11	308.67	-133.67	74.95	no	.942
12	334.99	-132.99	75.49	no	.941
13	335.58	164.42	69.59	yes	.950
14	613.81	116.19	79.34	no	.926
15	817.81	-5.81	83.10	no	.912
16	814.03	53.97	82.64	no	.907
17	765.96	-107.96	82.64	no	.924
18	348.73	116.27	82.64	no	.930

Full regression model:
Total = 12.7739 + 7.2725 * Block7 + 21.3397 * Block10 - 80.7305 * Block2
+ 4.4173 * Block9

R^2 = .9299 Significance level = .0001

Residual mean = 3.826

Residual standard deviation = 89.37

TABLE 7.5--Validation summary of the linear regressive model of total intensity on blocks 7, 10, 2, and 9 in a clustering scheme

Year	Prediction	Residual	Standard Deviation	Significant at 5%?	R^2
1	68.86	-42.86	116.61	no	.849
2	129.25	-43.25	116.57	no	.854
3	61.20	76.80	115.35	no	.860
4	121.44	82.56	114.97	no	.863
5	201.75	-20.75	117.03	no	.858
6	220.30	-49.30	116.36	no	.859
7	259.73	-49.73	116.36	no	.861
8	232.20	-41.20	116.65	no	.859
9	270.37	-41.37	116.37	no	.860
10	321.42	-91.42	114.82	no	.865
11	350.80	-175.80	107.36	no	.880
12	388.19	-186.18	106.93	no	.882
13	356.16	143.83	110.19	no	.874
14	646.51	83.49	115.81	no	.842
15	727.42	84.58	116.19	no	.828
16	891.59	-23.59	112.57	no	.814
17	457.47	200.53	112.57	no	.847
18	161.59	303.41	82.64	yes	.930

Full regression model:

Total = 41.8497 + 7.514 * Block7 + 2.987 * Block10 - 15.4253 * Block2 + 2.7309 * Block9, where Block10 and Block9 are zero if Block7 is less than or equal to eighty.

R^2 = .861 Significance level = .0001

Residual mean = 11.65

Residual standard deviation = 121.17

TABLE 7.6--Validation summary of the linear regressive model of total intensity on blocks 7 and 9 combined

Year	Prediction	Residual	Standard Deviation	Significant at 5%?	R^2
1	70.42	-44.42	108.03	no	.849
2	130.51	-44.50	107.98	no	.854
3	64.86	73.13	107.04	no	.859
4	123.47	80.53	106.65	no	.863
5	202.36	-21.36	108.45	no	.858
6	219.97	-48.97	107.85	no	.859
7	258.98	-48.98	107.85	no	.860
8	228.91	-37.91	108.16	no	.859
9	259.15	-30.15	108.35	no	.859
10	312.75	-82.75	106.53	no	.864
11	340.35	-165.35	100.11	no	.878
12	369.91	-167.91	100.11	no	.879
13	357.01	142.99	102.07	no	.874
14	661.10	68.90	107.42	no	.841
15	753.40	58.60	107.76	no	.828
16	893.78	-25.78	108.50	no	.814
17	670.43	-12.43	108.56	no	.846
18	168.29	296.71	77.65	yes	.928

Full regression model:
Total = 44.2817 + 7.4428 * Block7 + 2.5890 * Block9

R^2 = .8603　　Significance level = .0001

Residual mean = -.5361

Residual standard deviation = 106.26

model. Of these three approaches, the worst seems to be the four variable model, and the two variable model should be considered the best, since it achieves a stable, precise result for all amounts of input data.

Two immediate extensions to this surveillance proposal can be suggested at this stage. One uses smaller areas of surveillance but the same model form. Preliminary analysis suggests that this approach is viable down to blocks of one eighth of the size of the 24 block system. The second extension divides the region into smaller subregions for the basis of estimating the predictive equations. In this way, better estimates can be obtained for subregions on the basis of smaller areas of survey which, when recombined, can adequately describe the entire region. Little work has been done on this attack to date, but the potential seems clear.

The discussions above center on the relationship between a current single block value and the overall intensity of the region. It is also possible to predict regional intensities on the basis of past values of individual blocks. The three mode one blocks can be run using PROC STEPWISE of SAS (Helwig and Council, 1979). With the independent variables being the one-lag, two-lag, and three-lag values of the block in question, this procedure selects the best one, two, and three variable linear regression models on the basis of maximum R^2 improvement. The clear winner of the three blocks tested in this manner appears to be block 17 with an R^2 of .814 for the one variable model using only the one-lag value. A full listing of the models and their R^2 values appears in Table 7.7. The three blocks subjected to testing are 7, 14, and 17 (the mode one blocks).

Whereas the surveillance discussion primarily emphasizes only the multivariate analysis of the data, the following work will be based on the simulation model described in chapter 6. This examination of control implications will be performed in two stages; the first examines the effect of small perturbations in the block intensities, and the second discusses sensitivities of the block connections. The two-phase sensitivity analysis of the model in this way reveals the key points at which management intervention is most likely to affect the regional system. The proper manner in which this intervention should proceed must be based on the detailed objectives and capabilities of the situation at hand as well as on these key points.

Due to the number of runs of the simulation model necessary to evaluate these sensitivity experiments, only one perturbation type is considered. This disturbance is a single increase in the value of a single block (or block connection parameter) at the initial year of simulation. Furthermore, the magnitude of this increase is chosen to be 5 percent for each block. In this manner, the 24 blocks

TABLE 7.7--Summary of linear regressive models predicting total from lagged value of single blocks

	Intercept	Lag 1	Lag 2	Lag 3	R^2	Significance Level
Block 7						
	92.4666	7.8237	--	--	.6729	.0002
	93.8368	10.0253	--	-3.4402	.6965	.0008
	92.2081	8.8568	2.3958	-4.6576	.7033	.0031
Block 14						
	79.2075	8.1223	--	--	.6506	.0003
	94.3253	9.1375	--	-1.7802	.6658	.0014
	90.6336	8.6860	0.8570	-2.0820	.6678	.0056
Block 17						
	44.7312	9.8755	--	--	.8142	.0001
	50.9071	10.7539	--	-1.3542	.8203	.0001
	50.7287	10.8367	-0.2372	-1.7710	.8204	.0002

can each be tested separately for sensitivity of system response. A method for evaluating the overall effects of these induced perturbations is difficult since it must judge similarities over 24 blocks over 20 years based on stochastic processes. The method used in the analyses consists of first running a standard (unperturbed) simulation using the same sequence of pseudo-random numbers which will be used in the perturbed run. Total intensity over the twenty-four blocks at each of twenty years is calculated to serve as a base value for comparisons. In this way, a scaling among the simulations (and therefore among the blocks) can be formed on the basis of their similarities to this standard behavior. The comparison of the perturbed and the standard is given in Tables 7.8 through 7.10. Also listed in the tables are the mean squared errors which are used to form the ranking of sensitivities. The ordered values of the mean squared errors breaks into two natural groups, one of which contains the 3 sensitive blocks (blocks 7, 14, and 21). It is these 3 blocks which form the basis for the second stage of the sensitivity analysis as discussed below.

Each of the 10 parameters involving the 3 sensitive blocks are separately altered to 5 percent above their original value. This is accomplished by altering the regression coefficients in the equations of blocks which are connected to the sensitive blocks. The criterion of

TABLE 7.8--Total intensities of the standard and perturbed simulations for blocks one to eight over 20 years

Standard	Blocks Perturbed 1	2	3	4	5	6	7	8
237.50	237.50	237.50	237.71	237.50	237.50	237.50	239.45	238.54
321.83	321.83	321.83	321.84	321.83	321.83	321.83	324.92	323.39
404.91	404.91	404.91	405.00	404.91	404.91	404.91	408.38	405.33
349.44	349.44	349.44	349.48	349.44	349.44	349.44	353.49	348.78
338.37	338.37	338.37	338.44	338.37	338.37	338.37	342.67	337.58
200.94	200.94	200.94	200.98	200.94	200.94	200.94	206.52	201.02
207.16	207.16	207.16	207.25	207.16	207.16	207.16	213.38	207.49
286.46	286.46	286.46	286.57	286.46	286.46	286.46	293.96	287.08
368.82	368.82	368.82	368.87	368.82	368.82	368.82	377.53	369.22
432.36	432.36	432.36	432.40	432.36	432.36	432.36	441.60	431.93
461.34	461.34	461.34	461.43	461.34	461.34	461.34	470.17	460.86
537.16	537.16	537.16	537.21	537.16	537.16	537.16	545.96	536.62
555.99	555.99	555.99	556.02	555.99	555.99	555.99	563.28	555.11
263.55	263.55	263.55	263.57	263.55	263.55	263.55	268.09	262.96
461.29	461.29	461.29	461.31	461.29	461.29	461.29	467.11	460.49
626.74	626.74	626.74	626.76	626.74	626.74	626.74	634.50	625.74
712.26	712.26	712.26	712.28	712.26	712.26	712.26	720.37	711.17
706.96	706.96	706.96	706.96	706.96	706.96	706.96	703.26	707.42
707.46	707.46	707.46	707.42	707.46	707.46	707.46	695.04	709.13
637.49	637.49	637.49	637.47	637.49	637.49	637.49	629.99	638.55

Total squared errors for blocks 1 through 8, respectively, are: 0, 0, .0998, 0, 0, 0, 960.0577, and 14.1926.

TABLE 7.9--Total intensities of the standard and perturbed simulations for blocks nine to sixteen over 20 years

Standard	Blocks Perturbed 1	2	3	4	5	6	7	8
237.50	237.57	237.50	237.50	237.57	239.22	239.38	238.26	238.09
321.83	321.86	321.83	321.83	321.82	322.15	325.17	321.29	322.62
404.91	404.89	404.91	404.91	404.87	405.04	407.66	403.65	405.64
349.44	349.39	349.44	349.44	349.43	349.80	351.08	349.18	350.28
338.37	338.34	338.37	338.37	338.41	338.38	339.61	339.04	339.36
200.94	200.93	200.94	200.94	201.08	201.30	202.17	200.90	201.79
207.16	207.17	207.16	207.16	207.30	207.31	207.51	207.33	208.10
286.46	286.50	286.46	286.46	286.74	286.62	285.72	287.01	287.31
368.82	368.85	368.82	368.82	369.08	369.59	370.81	368.64	370.23
432.36	432.40	432.36	432.36	432.63	435.43	440.24	430.59	433.08
461.34	461.38	461.34	461.34	461.67	464.48	468.44	459.82	462.00
537.16	537.16	537.16	537.16	537.45	540.03	545.16	535.40	537.22
555.99	555.99	555.99	555.99	556.02	559.05	566.68	553.27	555.59
263.55	263.55	263.55	263.55	263.49	265.68	270.57	261.78	262.92
461.29	461.29	461.29	461.29	461.32	464.14	488.00	459.02	460.67
626.74	626.73	626.74	626.74	626.73	630.40	637.27	623.76	625.33
712.26	712.25	712.26	712.26	712.08	716.18	721.63	709.05	711.30
706.96	706.97	706.96	706.96	707.00	705.27	702.06	708.37	707.75
707.46	707.48	707.46	707.46	707.82	701.45	692.47	712.40	709.29
637.49	637.50	637.49	637.49	637.80	633.74	628.29	640.54	638.55

Total squared errors for blocks 9 through 16, respectively, are: .0155, 0, 0, .7601, 135.3231, 995.056, 82.44, and 17.4127.

TABLE 7.10--Total intensities of the standard and perturbed simulations for blocks seventeen to twenty-four over 20 years

	Blocks Perturbed							
Standard	1	2	3	4	5	6	7	8
237.50	238.69	238.03	237.87	237.63	237.51	237.88	238.03	237.54
321.83	322.92	322.21	322.84	321.61	321.84	322.21	322.12	321.83
404.91	405.69	406.39	405.88	404.39	404.94	405.45	404.94	404.93
349.44	350.50	350.68	350.04	348.46	349.64	349.80	349.34	349.43
338.37	339.77	340.30	339.21	336.45	339.55	338.63	338.37	338.40
200.94	202.00	202.47	201.72	198.95	202.60	201.16	201.02	200.95
207.16	208.51	209.43	208.38	203.96	210.35	207.13	207.27	207.19
286.46	287.72	289.53	288.56	281.41	292.09	286.40	286.55	286.48
368.82	371.25	372.14	370.07	365.79	371.32	368.54	368.92	368.91
432.36	433.92	432.85	431.34	434.73	427.07	432.23	432.39	432.42
461.34	462.93	462.17	460.45	463.26	456.01	461.23	461.37	461.38
537.16	538.01	536.64	536.14	539.72	531.36	537.20	537.12	537.13
555.99	555.95	553.20	553.62	562.01	546.54	556.08	555.97	555.97
263.55	262.69	261.20	261.88	267.68	256.82	263.70	263.49	263.50
461.29	460.78	458.67	459.08	466.64	452.58	461.41	461.26	461.25
626.74	624.92	622.39	623.97	633.59	615.45	627.06	626.61	626.65
712.26	710.90	708.42	709.15	719.84	700.00	712.43	712.19	712.18
706.96	707.99	709.08	708.23	703.75	712.23	706.79	707.05	707.01
707.46	710.32	714.00	712.28	695.77	726.41	707.10	707.63	707.59
637.49	639.36	641.43	640.58	630.17	643.35	637.31	637.58	637.57

Total squared errors for blocks 17 through 24, respectively, are: 41.3197, 153.65, 79.134, 456.56, 1,050.29, 1.7807, .5525, .0634.

comparison developed above is then used to rate these sensitivities of connection properties. Tables 7.11 through 7.13 examine these comparisons. Two results are apparent. Firstly, the most sensitive connections appear to be in outgoing connections from sensitive blocks, and, secondly, the connections of blocks 7 to 14, 14 to 20, and 21 to 20 seem highly sensitive.

The implication of this study for pest management is threefold. Firstly, the blocks 7 and 14 play a consistently important role in the system. The reasons for this role are unclear at this point, but their environmental qualities are certainly involved. Secondly, the connections mentioned above as sensitive are vital in determining the flow of the epidemic through the region. If control methods can be designed to affect large areas, then these connection properties will be an important consideration in their application. By deleting or weakening a few connections, there is potential for changing the entire system behavior. Thirdky, the effects of lower level control procedures can be examined for their total effect on the system to insure a global overview of strategies. In the long run, it must be the globally efficient strategies which determine the best management policies. This is true primarily because of the overall stability of the beetle infestation patterns in the region-wide milieu.

Model utilizations are only as good as the model's

TABLE 7.11--Total intensities for standard and perturbed simulations for connections from block seven

Standard	Perturbed Connections 7 to 2	7 to 6	7 to 8	7 to 14
237.50	237.50	237.50	237.62	237.61
321.83	321.83	321.79	323.13	323.15
404.91	404.91	404.95	405.53	408.12
349.44	349.44	349.43	349.24	354.41
338.37	338.38	338.29	338.71	345.14
200.84	200.91	200.79	203.59	207.94
207.16	206.69	206.93	211.18	215.03
286.46	285.29	285.77	292.45	296.37
368.82	367.49	368.31	373.86	381.46
432.36	431.03	431.70	434.14	447.04
461.34	460.06	460.22	462.06	475.52
537.16	536.52	536.31	537.74	550.82
555.99	556.19	555.81	557.66	567.97
263.55	263.53	263.58	263.04	271.66
461.29	461.28	461.09	463.16	468.47
626.74	626.66	627.08	631.85	638.47
712.26	712.20	712.33	713.03	722.14
706.96	707.01	706.56	706.47	701.60
707.46	707.54	707.42	707.16	691.71
637.49	637.56	637.77	641.97	627.75

Total squared errors for connections 7 to 2, 7 to 6, 7 to 8, and 7 to 14, respectively, are: 7.732, 3.6676, 144.3836, and 1,915.484.

TABLE 7.12--Total intensities for standard and perturbed simulations for connections from block fourteen

	Perturbed Connections			
Standard	14 to 7	14 to 13	14 to 15	14 to 20
237.50	237.49	238.39	237.78	237.57
321.83	321.81	322.97	321.94	321.79
404.91	404.85	406.87	404.89	404.64
349.44	349.34	351.27	349.05	348.47
338.37	338.17	340.45	337.68	335.81
200.94	200.66	202.44	200.31	197.53
207.16	206.75	208.21	206.74	200.99
286.46	285.91	287.77	286.71	275.66
368.82	368.18	371.67	368.11	363.90
432.36	431.83	437.65	429.84	442.25
461.34	460.78	466.72	459.26	472.51
537.16	536.66	542.32	535.45	547.46
555.99	555.85	560.90	553.79	572.61
263.55	263.47	266.89	262.00	273.73
461.29	461.20	465.71	459.27	468.71
626.74	626.52	632.58	624.10	639.51
712.26	712.10	718.12	709.41	722.20
706.96	707.04	704.26	708.20	701.10
707.46	707.69	698.40	711.80	689.97
637.49	637.62	632.06	640.17	626.60

Total squared errors for connections 14 to 7, 14 to 13, 14 to 15, and 14 to 20, respectively, are: 1.7103, 404.208, 69.3802, and 1,684.5977.

TABLE 7.13. Total intensities for standard and perturbed simulations for connections from block twenty-one.

	Perturbed Connections	
Standard	21 to 15	21 to 20
237.50	237.50	237.50
321.83	321.84	321.79
404.91	404.94	404.76
349.44	349.57	349.64
338.37	338.33	340.40
200.94	200.85	205.43
207.16	207.07	215.70
286.46	286.39	302.28
368.82	368.75	375.10
432.36	432.41	414.72
461.34	461.47	440.45
537.16	537.35	510.16
555.99	555.87	518.12
263.55	263.50	236.75
461.29	461.24	425.06
626.74	626.66	579.78
712.26	712.36	662.11
706.96	707.62	726.36
707.46	708.57	776.80
637.49	637.75	648.56

Total squared errors for connections 21 to 15 and 21 to 20, respectively, are: 1.8561 and 15,355.74.

ability to mimic the system at hand. The formulation developed in this paper suffers several drawbacks pri-mairly due to data inadequacy. Some of these problems may be ignored for most purposes, while others cannot be ignored. The ones which cannot be neglected demand further observation of the system, and these extensions are discussed below.

The extensions which will be discussed center most directly on what would happen if new data should be avail-able. Three immediate questions to such new acquisitions of data must follow. What new properties does this new data have? Next, how do these properties differ from those of the current data? Lastly, how do these new pro-perties affect the model formulation. Some effects will be negligible in that they will be merely a re-organization of the data. This process would be exemplified by re-running a regression model. Other effects will be negligible by adding simple random disturbances which would average out in the long run. Other effects might not be negligible however. Such effects would directly affect the type of model used. For example, a non-linear regression might replace a linear form. In the event that the entire model formulation seems dubious, then the difficulties of re-forma-tion must be addressed also. This three-pronged attack forms the basis of the discussion below.

Three simple problems exist for the present model, simple in the sense that they are probably negligible for most modeling purposes. One is the lack of sufficient data to determine whether the cycles of the system are periodic or merely regular. The 18 years of data used in this paper's analysis describe the system through a little over one cycle of epidemic behavior and are certainly not enough to ensure accuracy of the model. However, even if the cycles are very regular, the present model can easily be modified to account for this periodicity by proper lagged terms in the regression models. This would entail re-estimation of all the forms but otherwise would cause no deep problems. Solution of this problem can only await further years of observation on the system. Another minor problem is the estimation of the upper limits on the block infestation intensities. In general, the present estimates are probably pretty good but can only be verified by further information on the forest distribution properties over time of the blocks. Such time-varying effects could easily be incorporated into the simulation model by varying the upper limts correspondingly. The third shortcoming of the model is the inter-observation time of one year. Maps such as those in Price and Doggett result from accumulated data over months of observation. How well this type of data conforms to the way the model uses it is unknown but probably not dangerous. Indeed, intuition would lead one to believe that "cleaner" data could only lead to better estimations. Even if this is not true, the model as given does describe the evolution of some quantity related to total intensity.

One of the major unknowns in the model is the unknown effect of the beetle foraging itself on the extent of spatial distribution. This is directly related to the intrinsic versus extrinsic controversy mentioned in chapter 4. At this point, the hierarchical model is built around a variable (infestation intensity) which reflects the full spatial distributional properties of the blocks, but the data analysis is performed on a different variable (infestation incidence intensity).. The relationship between these two quantities has been assumed to be a simple, constant relation of probably linear form, but there is little verification of this assumption. The only indications in the literature thus far (as discussed in chapter 4 also) indicated that the assumption is valid, but more study is definitely required. If this problem exists to a significant degree, then most of the results presented in this paper would need examination.

A second major limitation of the model is the regression type approach which has been employed. Accuracy has been obtained at the expense of direct interpretability of the parameters. The biggest problem involved in this difficulty is the arbitrary definition of patch size and permanence. Although there is no good evidence that the artificial partitioning of the region should cause any real errors, it must be recognized that the true patch system is an irregular mosaic over both space and time. Such irregular mosaics are very hard to deal with in terms of the analyses performed up to this point, and re-coordination of the model would be an arduous task indeed.

The last major problem which will be considered in this discussion is the transience of the mosaic connection properties. It is apparent from the raw data that infestation activity depends upon the strength of the source block intensity in a complex manner. One possible complexity is the existence of a threshold effect, flow of infestation intensity being negligible under low infestation levels. Such a condition is very difficult to estimate but not very hard to implement in the model formulation. Another possible complication is that infestation movement occurs several times during the year and in all directions. Such a system could not be discovered by the present analysis due to the absence of data on the higher frequency movements of the beetles. Only further data collection can fully evaluate just how complex these relations are.

Based on the present level of information about the system and about the underlying levels of the hierarchy, the model formulated in this paper can be considered a good description of the upper echelon submodel. It provides an easily obtainable (due to its intrinsic nature) method of determining wide range effects of the beetle in time and space. Further data is necessary at all levels of the system to properly test and extend the approach with a priority on evaluating forest dynamics and the effects of shorter fluctuations in the internal block behaviors. Several utilizations are directly applicable to forest management and suggest a whole range of further investigations. As a first step into the relatively unknown area of regional southern pine beetle dynamics, it is hoped that use of the information in this paper will facilitate further research into this vital area. Such research is necessary for full evaluation of potential management routines and for understanding the full range of beetle behavior.

LIST OF REFERENCES

Barras, S. J. and Hodges, J. D. 1969. Carbohydrates of inner bark of Pinus taeda as affected by Dendroctonus frontalis and associated microorganisms. Canad. Entomologist. 101: 5-13.

Barton, D. E. and David, F. N. 1966. Random interspection of two graphs. Research Papers in Statistics, F. N. David (ed). New York: Wiley and Sons.

Belanger, R. P.; Hatchell, G. E., and Moore, G. E. 1977. Soil and stand characteristics related to southern pine beetle infestations: a progress report for Georgia and North Carolina. Proceedings of the Sixth Southern Forestry Soils Workshop at Charleston, S. C. held October 19-21, 1976.

Birch, Martin C. 1978. Chemical Communication in Pine Bark Beetles. American Scientist. 66: 409-420.

Blackith, B. E. and Reyment, R. A. 1971. Multivariate Morphometrics. New York: Academic Press.

Bookstein, Fred L. 1978. The measurement of biological shape and shape change. Lecture Notes in Biomathematics 24, S. Levin (ed). New York: Springer-Verlag.

Box, George E. P. and Jenkins, Gwilym M. 1976. Time Series: Forecasting and Control. P. 177. San Francisco: Holden-Day.

Cliff, A. D.; Haggett, P.; Ord, J. K.; Bassett, K. A., and Davies, R. C. 1975. Elements of Spatial Structure. New York: Cambridge University Press.

Coster, J. E. and Johnson, P. C. 1979. Characterizing flight aggregation of the southern pine beetle. Environ. Entomol. 8: 381-387.

Coster, J. E.; Payne, T. L.; Edson, L. J., and Hart, E. R. 1978. Influence of weather on mass aggregation of southern pine beetles at attractive host trees. Southwest Entomol. 3: 14-20.

Coster, J. E.; Payne, T. L.; Hart, E. R., and Edson, L. J. 1977. Aggregation of the southern pine beetle in response to attractive host trees. Environ. Entomol. 6: 725-731.

Coulson, Robert N. 1979. Population dynamics of bark beetles. Ann. Rev. Entomol. 24: 417-447.

Coulson, R. N.; Fargo, W. S.; Pulley, P. E.; Pope, D. N.; Foltz, J. L., and Bunting, A. M. 1979. Spatial and temporal patterns of emergence for within tree populations of Dendroctonus frontalis (Coleoptera: Scolytidae). Canad. Entomol. 111: 273-287.

Curry, L. 1971. Applicability of space-time moving-average forecasting. Regional Forecasting, proceedings of the twenty second symposium of the Colston Research Society at the University of Bristol in April 1970, M. Chisholm, A. Frey, and P. Haggett (eds). London: Butterworths.

Daniels, R. F.; Leuschner, W. A.; Zarnoch, S. J.; Burkhart, H. E., and Hicks, R. R., Jr. 1979. A method for estimating the probability of southern pine beetle outbreaks. For. Sci. 25: 265-269.

Dziadzio, M. F. 1978. Factors related to tree survival after southern pine beetle attack. M. S. Thesis. North Carolina State University at Raleigh. 71p.

Fargo, W. S.;Coulson, R. N.;Gagne, J. A., and Foltz, J. L. 1979. Correlation of southern pine beetle attack density, oviposition and generation survival with host tree characteristics and preceding beetle life stages within the host. Environ. Entomol. 110: 1213-1232.

Foltz, J. L.; Mayyasi, A. M.; Hain, F. P.; Coulson, R. N., and Martin, W. C. 1976. Egg-gallery length relationship and within-tree analyses for the southern pine beetle, Dendroctonus frontalis (Coleoptera: Scolytidae). Canad. Entomol. 108: 341-352.

Getis, Arthur and Boots, Barry. 1975. Models of Spatial Processes. P. 2. New York: Cambridge University Press.

Gold, H. J.; Mawby, W. D., and Hain, F. P. 1980. Proceedings of the Southern Pine Beetle Modeling Symposium at Asheville, North Carolina in February 1980. (In press).

Haggett, Peter and Chorley, R. J. 1970. Network Analysis in Geography 9. New York: St. Martin's Press.

Hain, F. P.; DeMars, C. J.; McClelland, W. T., and Mawby, W. D. 1979. Relating tree mortality to collapsing beetle populations. Symposium Proceedings of Evaluating Control Tactics for the Southern Pine Beetle at Many, Louisiana in January 1980.

Hodges, J. D.; Elam, W. W.; Watson, W. F., and Nebeker, T. E. 1979. Oleoresin characteristics and susceptibility of four southern pines to southern pine beetle attack. Canad. Entomol. 111: 889-896.

Hollander, Myles and Wolfe, Douglas A. 1973. Nonparametric Statistical Methods. P. 1. New York: Wiley and Sons.

Holling, C. S. 1978. Adaptive Environmental Assessment and Management. New York: Wiley and Sons.

Johnson, P. C. and Coster, J. E. 1978. Probability of attack by southern pine beetle in relation to distance from an attractive host tree. For. Sci. 24: 574-580.

Kendall, M. G. 1973. Time-Series. P. 16. London: Griffen.

Kendall, M. G. 1975. Multivariate Analysis. P. 1. London: Griffen.

King, E. 1972. Rainfall and epidemics of the southern pine beetle. Environ. Entomol. 1: 279-285.

Klauber, Melville R. 1974. Space-time clustering analysis: a prospective. A SIMS Conference on Epidemiology at Alta, Utah in 1974, D. Ludwig and K. Cooke (eds). Philadelphia: Society for Industrial and Applied Mathematics.

Klauber, M. R. 1971. Two-sample randomization tests for space-time clustering. Biometrics 24: 541-556.

Klovan, J. E. 1975. R- and Q-mode factor analysis. Concepts in Geostatistics, Richard McCammon (ed). New York: Springer-Verlag.

Kowal, R. J. 1960. The southern pine beetle. USDA Forest Service Forest Pest Leaflet 49. 10p.

Leuschner, W. A.; Burkhart, H. E.; Spittle, G. D.; Ragenovich, I. R., and Coulson, R. N. 1976. A descriptive study of host and site variables associated with the occurrence of Dendroctonus frontalis Zimm. in east Texas. Southwest. Entomol. 1: 141-149.

shortleaf, Virginia, and loblolly pines. Part II. Sequence of arrival of insects associated with bark beetles at severed shortleaf, Virginia, and loblolly pines. M. S. Thesis. North Carolina State University.