Models for Spatially Distributed Populations: The Effect of Within-Patch Variability

PETER L. CHESSON

*Department of Biological Sciences and Marine Science Institute,
University of California, Santa Barbara, California 93106*

This paper studies population models which have the following three ingredients: populations are divided into local subpopulations, local population dynamics are nonlinear and random events occur locally in space. In this setting local stochastic phenomena have a systematic effect on average population density and this effect does not disappear in large populations. This result is an outcome of the interaction of the three ingredients in the models and it says that stochastic models of systems of patches can be expected to give results for average population density that differ systematically from those of deterministic models. The magnitude of these differences is related to the degree of nonlinearity of local dynamics and the magnitude of local variability. These results explain those obtained from a number of previously published models which give conclusions that differ from those of deterministic models. Results are also obtained that show how stochastic models of systems of patches may be simplified to facilitate their study.

1. INTRODUCTION

The chances of survival and reproduction for an individual organism depend on the physical environment and other organisms in the place in which it lives. In self-sustaining natural populations there will be spatial variation in the numbers of individuals. This variation alone leads to spatial variation in the chances of survival and reproduction for any individual.

There is also temporal variability: local populations do not converge towards an equilibrium, but fluctuate; and to the extent to which these fluctuations are independent from place to place, they create an ever-changing spatial pattern in population numbers, which in turn leads to an ever-changing spatial pattern in the chances of survival and reproduction for an individual.

Does this variation somehow average out so that average population density, set wholly by deterministic forces, is all that needs to be considered? Classical population models in ecology make this assumption (e.g., Lotka–Volterra and Nicholson–Bailey) and it has been sustained in the

288

majority of recent developments (see May 1976). On the other hand, Andrewartha and Birch (1954) emphasized spatial and temporal variability as important mechanistic aspects of population development. Since then there has been occasional reemphasis (DenBoer, 1968; Birch, 1970; Levin and Paine, 1974; Wiens, 1976; Connell, 1978; Caswell, 1978; Murdoch, 1979), but the mathematical development of these ideas is relatively recent.

In recognition of the effects of spatial variation in population numbers there is now a number of models for population development in both space and time. (See reviews by Levin, 1976; Mimura *et al.*, 1978; McMurtie, 1978.) However, because these models are deterministic they omit an important temporal aspect of local population development; they do not take account of the randomness involved in births, deaths and migration on the local scale, i.e., they do not take account of within-path variability (Chesson, 1978). As a result these models underestimate both temporal and spatial variation in population numbers; indeed many of these models predict the eventual disappearance of all spatial and temporal variability (Levin, 1974; Allen, 1975; Hastings, 1978b; Conway *et al.*, 1978). When this happens in a model, it loses its usefulness for the study of the effects of the sustained spatial and temporal variability that we see in the real world. Within-patch variability is a factor omitted from these models that maintains spatial and temporal variability.

There is now a growing list of simple models dealing with at least some aspects of within-patch variability (Maynard Smith, 1974; Roff, 1974a, b; Slatkin, 1974; Hastings, 1977, 1978a; Zeigler, 1977; Caswell, 1978; Gurney *et al.*, 1978a, b). Even these simple stochastic models give strikingly different conclusions from deterministic models (Chesson, 1978). Unfortunately there are two difficulties with these models. Firstly, in most cases local populations are described by species presence or absence and this permits only a very limited range of kinds of local population dynamics. Secondly there are no suitable deterministic analogues for most of these models because the description of local populations by presence and absence variables is very unnatural in deterministic models. In deterministic models continuous descriptions of local populations are much more sensible. Thus the two classes of models that we wish to compare differ in ways other than the presence or absence of stochasticity. Consequently the reason for the differences between the predictions of the two sorts of models has remained unclear.

In this paper we present a class of models which overcomes both of these difficulties. Local population sizes are nonnegative integers; local population dynamics can be of any form and the sensible deterministic analogues are easily identified. These models are compared with two different kinds of deterministic models, those of the classical form that do not take account of spatial variation in population numbers and also models that do take

account of such spatial variation. It is shown that in general the stochastic model can be expected to give different results from the deterministic model even though total (but not necessarily local) population sizes are very large. Indeed there is an important systematic difference between the results of stochastic and deterministic models that does not depend greatly on population size and therefore does not disappear in indefinitely large populations. The approximate agreement between deterministic and stochastic models claimed by some authors (e.g., May, 1974a, b) for large populations is based on models that ignore the effect of spatial variation in population numbers.

The results obtained here depend heavily on the assumption that local population dynamics are nonlinear as they are in most real populations. Indeed this paper is best described as a study of the interaction between three properties of real populations: within patch-variability, nonlinearity and the fact that a population spread over a large area is subdivided into local populations (inhomogeneous interaction).

Model Ingredients

Inhomogeneous Interaction

We shall use the term inhomogeneous interaction to mean that the chances of survival and reproduction for an individual can not be determined simply by the spatial averages of population densities for the area but depend on the individual's movements and the spatial arrangement of the populations. The converse situation, homogeneous interaction (Chesson, 1978), is implicit in the classical kinds of models where the only variables in the model are total population sizes for an area or equivalently the spatial averages of population densities for the area.

To model inhomogeneous interaction we consider a system of patches supporting local populations and connected by migration. While an individual is present in a given patch, its chances of survival and reproduction are assumed to be a function of conditions in that patch.

Within-Patch Variability

Within-patch variability (Chesson, 1978) deals with random variation that is simultaneously spatial and temporal. It can be defined as the totality of patch specific random temporal events affecting local population density. Since these events are patch specific, different patches will experience different changes in density and this leads to the spatial aspect of within-patch variability. Within-patch variability is made up of randomness in migration processes, within-individual and between-individual variation and random changes in the local environment.

Within-patch variability is a dynamic property of a patch and this distinguishes it from between-patch variation which refers to those physical differences between patches that do not change with time.

Temporal environmental variability has two components, that which is common to all patches, and that which is specific to a patch and therefore forms part of within-patch variability. The part common to all patches is distinguished from within-patch variability by the fact that it is purely temporal and has no spatial component (its effect is by definition the same everywhere).

Nonlinearity

We assume that population dynamics within a patch ("internal dynamics") are nonlinear because, as we shall see, linearity occurs only under very limited conditions that do not persist in nature for very long periods of time.

Let $\mathbf{z}(t) = (z_1(t),..., z_r(t))$ be the population sizes of r different species. A discrete time deterministic model takes the form

$$\mathbf{z}(t + 1) = \mathbf{g}(\mathbf{z}(t)). \tag{1}$$

Generally $\mathbf{g}$ is said to be linear if $\mathbf{g}(\mathbf{z}) = \mathbf{\alpha} + \mathbf{Az}$, where $\mathbf{\alpha}$ is a constant vector and $\mathbf{A}$ is a constant matrix. For $\mathbf{g}$ to be a sensible growth function $\mathbf{g}(\mathbf{0}) = \mathbf{0}$ and $\mathbf{g}(\mathbf{z}) \geqslant \mathbf{0}$ for all $\mathbf{z} \geqslant \mathbf{0}$. These restrictions imply that $\mathbf{\alpha} = \mathbf{0}$ and $\mathbf{A}$ is non-negative. For the single species case, linear models represent geometric growth. For several species models, the diagonal elements of $\mathbf{A}$ are geometric growth terms while the off-diagonal elements can only represent density independent mutualism or commensalism. Thus exactly linear models are a very restricted class. Linear approximations to nonlinear models can be adequate for some purposes, for a small range of population sizes, and such approximations need not satisfy the strict requirements of exactly linear models. However, in this paper we are interested in the effects of local density variation over large ranges of local population sizes where linear approximation is not adequate.

In our stochastic models, the function $\mathbf{g}$ represents a mean population growth function within a patch. Instead of z we use Z to indicate that it is a random variable and we add a second subscript to give $Z_{ij}(t)$, which is the population size of the ith species in the jth patch at time t.

2. STOCHASTIC MODELS ASSUMING INHOMOGENEOUS INTERACTION

In this section we define stochastic models that incorporate all three of the ingredients listed above. The first, called the finite stochastic model, has a finite number, k, of patches while the second, the infinite stochastic model,

has an infinite number of patches and is defined so that it will approximate the finite model for large k. Large k means that the populations will occupy a large area and this corresponds to large total populations for the area.

The approximation of the finite model by the infinite model leads us to our first conclusion, namely that when k is large, within-patch variability causes little random variability in the spatial average of population density. However, we see in Section 3 that there is a systematic departure from the predictions of deterministic models.

2.1. *The Finite Stochastic Model*

This model is based on the following general assumptions.

1. The physical environment is divided into $k < \infty$ patches. These patches are identical in their physical attributes, i.e., no between-patch variation is involved.

2. Migration of organisms is separated in time from the processes of population growth and intra- and inter-specific interactions. For each time interval $(t, t + 1)$ migration occurs during $(t, t + h)$ and population growth processes take place during $(t + h, t + 1)$, $t = 0, 1, 2,...$. Population changes in any patch during $(t + h, t + 1)$ are therefore assumed to depend only on the population sizes in that patch at time $t + h$, and random events specific to that patch (within-patch variability). Thus the model embodies inhomogeneous interaction.

3. It is assumed that organisms are highly dispersive so that all patches are equally accessible to a migrating orgnism. Moreover there is no interaction during migration—individuals are assumed to migrate independently.

4. The system as a whole is closed. There is no immigration to, nor emigration from, the system as a whole.

Of these assumptions, the least realistic is that all patches are equally accessible to a migrating organism. Although it can be argued that highly dispersive organisms may find all patches equally accessible when the number of patches is small, this is not likely to be so when the number of patches is large, which is the case of most interest here. Fortunately, as discussed in Section 6, our principal conclusions do not depend at all on equal accessibility, but it is assumed here because it leads to far simpler mathematics. Equal accessibility results in a greater degree of population mixing, reducing spatial variation and erring on the side of conservatism.

We now state the model in mathematical terms. For this we define $\mathbf{Z}_j(t) = (Z_{1j}(t),..., Z_{rj}(t))'$, the vector of population variables for patch j at time t. The average population densities are the components of the vector $\bar{\mathbf{Z}}_k(t) = k^{-1} \sum_{j=1}^k \mathbf{Z}_j(t)$, i.e., they are defined as simple spatial averages. The symbol

H_t is used to represent the history of the system up to and including the time t.

Assumption 2 means that $\mathbf{Z}_1(t + 1),..., \mathbf{Z}_k(t + 1)$ are conditionally independent given H_{t+h}, and

$$P(\mathbf{Z}_j(t + 1) = \mathbf{z} \,|\, H_{t+h}) = \gamma(\mathbf{z}, \mathbf{Z}_j(t + h)), \tag{2}$$

where γ is some function, i.e., given the state of the entire system through time $t + h$, the conditional distribution of the population sizes on a patch after the period of growth and interaction is simply a function of the population sizes on that patch at time $t + h$. The conditional expectation of $\mathbf{Z}_j(t + 1)$ given H_{t+h} is

$$E[\mathbf{Z}_j(t + 1) \,|\, H_{t+h}] = \sum \mathbf{z}\gamma(\mathbf{z}, \mathbf{Z}_j(t + h)) = \mathbf{g}(\mathbf{Z}_j(t + h)), \tag{3}$$

where $\mathbf{g}$ is some function.

For assumption 3 we define μ_i as the probability of migration of an individual of species i. Given H_t, an individual of species i present in patch j at time t has probability $\mu_i/(k - 1)$ of migrating to patch $j' \neq j$. Different individuals are assumed to migrate independently and this means that the migration model is of the multinomial kind (Chesson, 1976).

Assumptions 1 and 4 are implicit in the mathematical statements of 2 and 3. In particular, assumption 1 is implicit from the fact that γ and μ_i do not depend on the patch involved. Thus different patches have identical properties and since dispersal is random, the patches are symmetrically related to one another. Although the patches are assumed labelled 1 to k, the order of the labelling is irrelevant to our considerations here and so without loss of generality we can assume that the initial populations $(\mathbf{Z}_1(0),..., \mathbf{Z}_k(0))$ have a "symmetrical" or "exchangeable" distribution. This means that $(\mathbf{Z}_{j_1}(0),..., \mathbf{Z}_{j_k}(0))$ has the same distribution as $(\mathbf{Z}_1(0),..., \mathbf{Z}_k(0))$ for all permutations $(j_1,..., j_k)$ of $(1,..., k)$. By making this additional assumption we preserve the symmetry of the system given by the other assumptions. As a result the probability distribution of $\{\mathbf{Z}_j(t), t \geqslant 0\}$ will be the same for each j. Actual population sizes will of course depend on j and will lead to spatial variation.

It is explained in the Appendix (A1) that the symmetry assumption is perfectly natural and in no way biasses the conclusions that we draw.

EXAMPLE 1.   As an illustration consider a single species model where, given H_{t+h}, $Z_j(t + 1)$ is a Poisson random variable with mean $K(1 - \exp\{-\beta Z_j(t + h)\}) = g(Z_j(t + h))$. This is a simple density dependent growth model where the mean population at time $t + 1$ approaches a maximum value K for large populations at time $t + h$. For this model

$$\gamma(z, Z_j(t + h)) = \frac{[g(Z_j(t + h))]^z \, e^{-g(Z_j(t+h))}}{z!}. \tag{4}$$

The randomness specified by the probabilities γ is a result of random events taking place within patch j during the growth period $(t + h, t + 1)$.

2.2. The Infinite Stochastic Model

This model is very similar to the previous model but has infinitely many patches. Indeed apart from this change the nonmathematical statements of the assumptions are identical. The mathematical statements may seem a little surprising but we shall see their justification in the next subsection.

In the infinite stochastic model there are infinitely many population processes $\{Z_j(t), t \geqslant 0\}$, $j = 1, 2,...$ corresponding to the infinity of patches. These processes are independent and identically distributed. By "independent" we mean, as always, statistically independent. Independence between the population processes does not mean that the patches are isolated and do not affect each other; patches are linked by migration, but the effect of the other patches on the immigrants to any given patch depends only on the limiting average population densities, namely $\lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} Z_j(t)$. This limiting vector is equal to $EZ_j(t)$, by the law of large numbers, and therefore is nonrandom, i.e., has zero variance. Thus patches affect each other in a nonstochastic manner. In contrast, in the finite model the relevant quantities are $k^{-1} \sum_{j=1}^{k} Z_j(t)$, and since $k < \infty$ these are stochastic but may have small variance for large k. The precise model for migration is: Given H_t, different individuals migrate independently. The number of emigrants of species i from patch j is binomial with parameters $(Z_{ij}(t), \mu_i)$ as it is in the finite model. The number of immigrants of species i, to patch j, is Poisson with mean $\mu_i EZ_{ij}(t)$. This immigration model comes from letting $k \to \infty$ in the immigration processes of the finite model.

The model for population change during $(t + h, t + 1)$ is the same as in the finite model, i.e., given H_{t+h}, the conditional distribution of $Z_j(t + 1)$ is $\gamma(\cdot, Z_j(t + h))$. In this model average population density is $\bar{Z}_\infty(t) = \lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} Z_j(t)$. As we have observed, this equals $EZ_j(t)$. Thus, average population densities are deterministic, but as we shall see this does not mean that average population density will be given by the deterministic analogue of the stochastic model.

2.3. Approximation of the Finite Stochastic Model by the Infinite Stochastic Model

Suppose we let $k \to \infty$ in the finite model, i.e., we allow the number of patches to become large, then the probability distribution for the population variables on any fixed collection of patches approaches the corresponding

probability distribution given by the infinite model. Furthermore the average density process $\bar{\mathbf{Z}}_k$ of the finite stochastic model approaches the average density process $\bar{\mathbf{Z}}_\infty$ of the infinite model. These results are proved in the Appendix (Theorem A1) under conditions that are general, but of a technical nature, and need not concern us here because they hold in alomost all cases of practical interest.

These results show that for large k the infinite model is a good approximation to the finite model and is therefore the appropriate infinite patch modification of the finite model. An important consequence of this result is that as $k \to \infty$ the average population densities $\bar{\mathbf{Z}}_k(t)$, $t \geqslant 0$, converge to a nonrandom or deterministic process $\bar{\mathbf{Z}}_\infty(t)$, $t \geqslant 0$. This at first appears to be in accordance with the opinion summarized by May (1974) that deterministic models provide adequate descriptions of the dynamics of large populations. However, we shall see that the correspondence is only partial. Our results do say that the average population density of a large population (living on a large area) will show little random variability, they do not say that the corresponding deterministic model adequately describes the dynamics of average population density.

The approximation of the finite stochastic model by the infinite stochastic model provides a useful tool for determining the behavior of the finite model for large k because as $k \to \infty$ a number of simplifications occur which are expressed in the infinite model. We shall make use of this fact in this paper and it has already been used to advantage to determine conditions for population stability (Chesson, 1981a). These simplifications are as follows.

Our results show that as $k \to \infty$, the different patches become (statistically) independent. The reason for this independence is that communication between patches occurs through the pool of migrating individuals. As $k \to \infty$, the size of this pool, divided by the number of patches, converges to a nonrandom quantity—it is simply a function of $E\mathbf{Z}_j(t)$. Thus for large k the effect of other patches on a given patch is simply a function of $E\mathbf{Z}_j(t)$, the theoretical mean for any patch.

Note however, that immigration and emigration are still stochastic processes. For any finite k immigration is conditionally binomial with parameters $(\sum_{j=1}^{k} Z_{ij}(t), \mu_i/(k-1))$; in the limit it is Poisson with mean $\mu_i EZ_{ij}(t)$. The chief difference is that for finite k one parameter is a random variable, but in the limit the only parameter involved is nonrandom.

Finally for large k, our results say that the dynamics of the system of patches can be studied by restricting attention to the individual patch. In order to determine the amount of immigration into the patch, one does need to know the average population densities for all other patches; however, these average population densities equal the theoretical mean population densities, $EZ_j(t)$, which are the same for all patches and are therefore available from the patch under study.

3. The Deviation between Stochastic and Deterministic Models

In this section we build deterministic analogues to our stochastic models and these are compared with the infinite stochastic model, a model assuming inhomogeneous interaction. The stochastic model is found to give different results for average population density than both homogeneously interacting and inhomogeneously interacting deterministic models. Inhomogeneously interacting deterministic models also have predictions about spatial variation and local population development and in both cases these predictions differ from those of the stochastic model.

A parallel comparison can be done using the finite stochastic model with similar conclusions; but the significant point is that the differences do not disappear as $k \to \infty$, i.e., as the area and total population become large, as testified by the infinite stochastic model.

The deterministic models are built from the deterministic analogues to the internal dynamics and movements components of the stochastic models. Since

$$E[\mathbf{Z}_j(t + 1) \mid H_{t+h}] = \mathbf{g}(\mathbf{Z}_j(t + h)) \tag{5}$$

the deterministic analogue of the internal dynamics process is

$$\mathbf{z}_j(t + 1) = \mathbf{g}(\mathbf{z}_j(t + h)). \tag{6}$$

For movements we have

$$E[\mathbf{Z}_j(t + h) \mid H_t] = \mathbf{Z}_j(t) + \mathbf{M}(\overline{\mathbf{Z}}_k(t) - \mathbf{Z}_j(t)), \tag{7}$$

where $\mathbf{M} = \operatorname{diag}(\mu_1,...,\mu_r)$ is the diagonal matrix with diagonal $(\mu_1,...,\mu_r)$. Hence the deterministic analogue is

$$\mathbf{z}_j(t + h) = \mathbf{z}_j(t) + \mathbf{M}(\overline{\mathbf{z}}(t) - \mathbf{z}_j(t)), \tag{8}$$

where $\overline{\mathbf{z}}(t)$ is the vector of average population densities. These deterministic analogues assume that the population changes will exactly follow the mean whereas in the stochastic models, within-path variability causes random variation about the mean prediction.

3.1. The Simple Deterministic Model

The simple deterministic model assumes homogeneous interaction and in the present deterministic setting this means that average population density behaves just like the local population density of an isolated patch. Thus from Eq. (6) we have

$$\overline{\mathbf{z}}(t + 1) = \mathbf{g}(\overline{\mathbf{z}}(t)). \tag{9}$$

This is a classical deterministic difference equation for average population density and it defines the simple deterministic model. The stochastic model does not satisfy (9), however. To see this we deduce from (5) that

$$\bar{\mathbf{Z}}_{\infty}(t + 1) = E\mathbf{g}(\mathbf{Z}_j(t + h)),\tag{10}$$

which means that $\bar{\mathbf{Z}}_{\infty}$ can only satisfy Eq. (9) if

$$E\mathbf{g}(\mathbf{Z}_j(t + h)) = \mathbf{g}(E\mathbf{Z}_j(t + h))\tag{11}$$

since $E\mathbf{Z}_j(t + h) = \bar{\mathbf{Z}}_{\infty}(t)$. If $\mathbf{g}$ is linear, then Eq. (11) is always satisfied. However, when $\mathbf{g}$ is nonlinear we have good reason to expect that the stochastic and deterministic models will differ. This is best illustrated by Example 1, where

$$g(Z_j(t + h)) = K(1 - \exp\{-\beta Z_j(t + h)\}).\tag{12}$$

Unless $Z_j(t + h)$ has zero variance (an impossibility here) then $E \exp\{-\beta Z_j(t + h)\} > \exp\{-\beta E Z_j(t + h)\}$ by Jensen's inequality, and so

$$\bar{Z}_{\infty}(t + 1) < g(\bar{Z}_{\infty}(t)).\tag{13}$$

Since g is an increasing function it follows for Example 1 that the deterministic model always overestimates average population density.

It is helpful to understand that if $\beta Z_j(t + h)$ is small then

$$\bar{Z}_{\infty}(t + 1) - g(\bar{Z}_{\infty}(t)) \approx -\tfrac{1}{2}\beta^2 K V Z_j(t + h).\tag{14}$$

For larger $\beta Z_j(t + h)$, higher moments become important also.

In Example 1 we can see that the deterministic and stochastic models must always disagree because Eq. (11) is never satisfied, but for some nonlinear $\mathbf{g}$ it is possible to find special probability distributions for $\mathbf{Z}_j(t + h)$ that make (11) true. However, for a given nonlinear $\mathbf{g}$ there are always probability distributions for which (11) is false (A4.1), and the special probability distributions that make (11) true are, in the sense discussed in the Appendix, quite rare (A4.2). Thus we can state with a high degree of confidence that the mean of a nonlinear function is not equal to the nonlinear function of the mean.

Equation (11) has to be true for all t for the deterministic model to correctly predict average population density. Our results indicate that very exceptional circumstances are required for this to be true for nonlinear $\mathbf{g}$.

On the other hand it is clear that Eq. (11) will be approximately true if there is very little variation in $\mathbf{Z}_j(t + h)$ about its mean value $E\mathbf{Z}_j(t + h)$. The results of Section 4 (below) suggest that this can be so if the mean populations on individual patches are always large and there is no local

temporal environmental variability. In other situations there are lower limits to the variation in $\mathbf{Z}_j(t + h)$.

3.2. *The Complex Deterministic Model*

This model assumes inhomogeneous interaction and is therefore of the same general kind as the deterministic models reviewed by Levin (1976), Mimura *et al.* (1978) and McMurtie (1978).

This model is defined by Eqs. (6) and (8) and we assume that there is initial spatial variation, i.e., $\mathbf{z}_j(0) \not\equiv \bar{\mathbf{z}}(0)$. This initial spatial variation can be described by a frequency distribution for patches with given densities, indeed the model can be analyzed quite validly as though the initial pattern were stochastic.

The model remains deterministic in the sense that the mechanisms of change in the spatial pattern, given by Eqs. (6) and (8), are entirely deterministic. However, $\mathbf{z}_j(t)$ is a random vector by virtue of the random initial conditions and therefore has a mean, variance and so on. Thus to complete the definition of this model we assume that $\mathbf{z}_1(0)$, $\mathbf{z}_2(0)$,... is an i.i.d. sequence of random vectors and we define

$$\bar{\mathbf{z}}(t) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} \mathbf{z}_j(t) = E\mathbf{z}_j(t). \qquad (15)$$

Three questions are of relevance:

(a) How does the density on a given patch change as a function of its initial density?

(b) How does spatial variation in patch densities change with time?

(c) How does average population density $\bar{\mathbf{z}}(t)$ change with time?

(a) A stochastic analogue of the first question is: given the initial spatial pattern, how does the conditional mean of $\mathbf{Z}_j(t)$ change with time, i.e., how does $E[\mathbf{Z}_j(t)|H_0]$ change with time? For the stochastic model to give the same answer as the complex deterministic model it is necessary that

$$E[\mathbf{Z}_j(t + 1)|H_0] = \mathbf{g}(E[\mathbf{Z}_j(t + h)|H_0]), \qquad (16)$$

which is equivalent to

$$E[\mathbf{g}(\mathbf{Z}_j(t + h))|H_0] = \mathbf{g}(E[\mathbf{Z}_j(t + h)|H_0]). \qquad (17)$$

It can be expected that Eq. (17) will be false for nonlinear $\mathbf{g}$ in the same way that Eq. (11) is likely to be false. Hence the stochastic and deterministic models can be expected to disagree in their answers to question 1.

(b) In the complex deterministic model, spatial variation can be measured by the variance of $z_j(t)$, since this variance is equal to the expected value of the sample variance calculated in the usual way from a random sample of patches. The same result holds for the infinite stochastic model.

As explained in Section 5 it is easily shown that the variance for the infinite stochastic model satisfies the equations

$$VZ_j(t + h) = (I - M)\, VZ_j(t)(I - M) + \text{diag}(\bar{Z}_\infty(t))\, M(2I - M) \quad (18)$$

and

$$VZ_j(t + 1) = Vg(Z_j(t + h)) + EV(Z_j(t + 1)\,|\,H_{t-h}), \quad (19)$$

while in the complex deterministic model we have

$$Vz_j(t + h) = (I - M)\, Vz_j(t)(I - M) \quad (20)$$

and

$$Vz_j(t + 1) = Vg(z_j(t + h)). \quad (21)$$

Thus the stochastic model has additional terms in the variance equations which will tend to give it a greater amount of spatial variation. These additional terms (due to within-patch variability) always maintain spatial variation above some minimum level set by average population density. In the deterministic model there is no lower bound on spatial variation, in fact it may disappear altogether with increasing t, as it does in those single species models which like Example 1 have strictly concave g (see Appendix 3). Other authors have also observed the disappearance of spatial variation in deterministic models (Levin, 1974; Allen, 1975; Hastings, 1978b; Conway *et al.*, 1978) but it is clear that spatial variation cannot disappear from a stochastic model involving within-patch variability, unless all species become extinct.

(c) It is more difficult to show generally how the stochastic model and the complex deterministic model will differ in their answers to the final question "How does average population density behave?" However, we can show that in a variety of cases the answers do differ and we can demonstrate that in general there is no reason to expect the answers to agree.

In the stochastic model

$$\bar{Z}_\infty(t) = EE[Z_j(t)\,|\,H_0]. \quad (22)$$

For the deterministic model the corresponding equation is true; but because it is deterministic $E[z_j(t)\,|\,H_0] = z_j(t)$ and so

$$\bar{z}(t) = Ez_j(t). \quad (23)$$

As remarked above $z_j(t)$ and $E[\mathbf{Z}_j(t)|H_0]$ will generally differ in nonlinear cases—this is question (a). In some cases, as in Example 1, a strict inequality applies, e.g.,

$$z_j(t) > E[\mathbf{Z}_j(t)|H_0]. \tag{24}$$

This inequality then implies

$$\bar{z}(t) > \bar{\mathbf{Z}}_\infty(t). \tag{25}$$

In other cases inequality (24) (or the reverse inequality) may not hold uniformly for all possible initial conditions; but since a bias in one direction or the other is to be expected in a nonlinear model, only in unusual circumstances can we expect that the average of the cases when $z_j(t) > E[\mathbf{Z}_j(t)|H_0]$ and $z_j(t) < E[\mathbf{Z}_j(t)|H_0]$ will yield the equality $\bar{\mathbf{Z}}_\infty(t) = \bar{z}(t)$.

3.3. Persistence of Differences as $t \to \infty$

Appendix 3 discusses the simple and complex deterministic models in the single species case with g strictly concave, as in Example 1. Although it is apparent that the simple and complex models will give different answers for average population density for all finite t, provided only that there is initial spatial variation, these differences disappear in the limit as $t \to \infty$; i.e. these two deterministic models give the same answer for the ultimate or "equilibrium" average population density. Other authors (Levin, 1974; Allen, 1975; Hastings, 1978b; Conway et al., 1978) have also observed broad situations in which an inhomogeneously interacting deterministic model agrees ultimately with the corresponding simple deterministic model. This raises the possibility that the stochastic model might behave in the same way, i.e., although it differs from the deterministic model (simple or complex) for finite times, in the limit as $t \to \infty$ there is no difference.

Let us suppose that the simple deterministic model has a globally stable equilibrium so that $\bar{z}(t) \to z^*$ and $g(z^*) = z^*$. Is it then possible that $\bar{\mathbf{Z}}_\infty(t) \to z^*$? To answer this question heuristically, suppose that $\mathbf{Z}_j(t + h)$ converges in distribution to some random vector $\mathbf{Z}^*(h)$. In order for $\lim_{t\to\infty} \bar{\mathbf{Z}}_\infty(t) = z^*$ we must have

$$z^* = Eg(\mathbf{Z}^*(h)) \tag{26}$$

in analogy with Eq. (11). But $E\mathbf{Z}^*(h) = z^* = \lim_{t\to\infty} E\mathbf{Z}_j(t + h)$ and so (26) is equivalent to

$$Eg(\mathbf{Z}^*(h)) = g(E\mathbf{Z}^*(h)). \tag{27}$$

Since $\mathbf{g}$ is nonlinear this equation can not be true in general and in broad situations it can be shown that (27) can not be possibly be satisfied; for

example, in the single species case with g strictly concave, we know from Jensen's inequality that

$$Eg(Z^*(h)) < g(EZ^*(h)). \tag{28}$$

Thus if $\bar{Z}_\infty(t)$ converges at all, the limit must be less than z^*. Indeed it is possible to show that $\overline{\lim} \bar{Z}_\infty(t) < z^*$, i.e., the peaks of the fluctuations of $\bar{Z}_\infty(t)$ must converge to some number less than z^*. As we have already noted, in this example the complex deterministic model gives the same answer for the ultimate average population density as the simple deterministic model and this means that the stochastic model continues to disagree in the limit with both kinds of deterministic model. Unlike the complex deterministic model, the stochastic model does not lose spatial variation as t increases and this fact prevents it from agreeing ultimately with the simple deterministic model.

(The heuristics involved in the above demonstration can be justified by the fact that convergence of $\bar{Z}_\infty(t)$ implies tightness of $\{Z_j(t+h)\}$ and thus a subsequence must converge to $Z^*(h)$. A mild condition such as bounded variance is needed to ensure that $\lim_{t\to\infty} EZ_j(t+h) = EZ^*(h)$.)

4. Stochastic Models with Homogeneous Interaction

In contrast to our findings above, Kurtz (1970) and Wang (1976) find essentially no difference between the results of deterministic and stochastic models when populations are large. Kurtz' and Wang's models assume homogeneous interaction while ours assume inhomogeneous interaction. However, Kurtz and Wang also use a somewhat different theoretical framework giving models that are not directly comparable with ours. Thus in order to complete the demonstration that inhomogeneous interaction is necessary and sufficient for the differences that we observe between deterministic and stochastic models, we go on to define our own stochastic models with homogeneous interaction.

In a homogeneous interacting population the chances of survival and reproduction of an individual are simply functions of the average population densities for the whole area. Although there are a number of ways of translating this statement into a statement about how a local population should grow, all of these different ways give the same conclusions regarding average population density. Thus without loss of generality we can define the finite stochastic model with homogeneous interaction by the equation

$$P(Z_j(t+1) = z \mid H_t) = \gamma(z, \bar{Z}_k(t)). \tag{29}$$

The function γ is the same as in the finite stochastic model but now average population density replaces local population density in this conditional probability function. (Movements have been omitted from this model because they are no longer relevant.)

It is shown in the appendix that for large k this finite model is well approximated by the corresponding infinite model. In the latter $Z_1, Z_2,...$ are independent and identically distributed and

$$P(Z_j(t+1) = z | H_t) = \gamma(z, \bar{Z}_\infty(t)). \tag{30}$$

Here $\bar{Z}_\infty(t) = EZ_j(t) = \lim_{n\to\infty} n^{-1} \sum_{j=1}^n Z_j(t)$ and

$$\bar{Z}_\infty(t+1) = g(\bar{Z}_\infty(t)). \tag{31}$$

Thus in the infinite model, average population density $(\bar{Z}_\infty(t))$ is described by the simple deterministic model and this means that average population density in the finite stochastic model with homogeneous interaction is approximately described by the simple deterministic model. Thus we see that within-patch variability has a vanishingly small effect on the average population density in a homogeneously interacting model with large k. Note however that within-patch variability itself does not disappear in homogeneously interacting models because $Z_j(t)$ remains a random variable in the infinite model, but this randomness has neither a random nor a systematic effect on the average population density for the entire infinite collection of patches.

5. THE MAGNITUDE OF LOCAL VARIABILITY

The results of Section 3 above show that within-patch variability does have a systematic effect on the average population density in models with inhomogeneous interaction. This systematic effect appears as a deviation between the results of deterministic and stochastic models and is related to the degree of nonlinearity of the local mean growth function g and the magnitude of the random variability in local population size. Here we explore this relationship to find an appropriate measure of variability. We then go on to show how various factors affect variability and hence affect the deviation between deterministic and stochastic models. For simplicity we consider directly only the infinite stochastic model and the simple deterministic model but the analysis extends to cover other comparisons.

5.1. *The Coefficient of Variation and the Deviation between Deterministic and Stochastic Models*

The relative deviation between the deterministic and stochastic model that appears during a unit of time is measured by the quantity

$$\frac{\bar{Z}_{i\infty}(t+1) - g_i(\bar{\mathbf{Z}}_\infty(t))}{g_i(\bar{\mathbf{Z}}_\infty(t))}, \tag{32}$$

where we note that $g_i(\bar{\mathbf{Z}}_\infty(t))$ is the deterministic prediction for the new average density of species i at time $t + 1$. There are no exact simple or general relationships between the magnitude of (32) and variability, but quadratic approximation of **g** provides approximate relationships. For example consider our Example 1 where $g(z) = K(1 - e^{-\beta z})$. For small β, $g(z) \approx K(\beta z - \tfrac{1}{2}\beta^2 z^2)$ and hence from (10) we have the approximation

$$\bar{Z}_\infty(t+1) \approx K[\beta\bar{Z}_\infty(t) - \tfrac{1}{2}\beta^2\bar{Z}_\infty^2(t) - \tfrac{1}{2}\beta^2 VZ_j(t+h)].$$

Thus an approximation to (32) is given by

$$\psi^2(Z_j(t+h))[1 - 2/\beta\bar{Z}_\infty(t)]^{-1}, \tag{33}$$

where $\psi^2(Z_j(t+h))$ is the coefficient of variation (variance/mean^2) of $Z_j(t+h)$.

The second factor in (33) is the ratio of the quadratic part of g to g and so is a measure of nonlinearity. Thus the deviation between the stochastic and deterministic models is expressed approximately as the product of a measure of variability (the coefficient of variation) and a measure of nonlinearity. This approximate relationship holds in general for the single species case.

In the multispecies case there is a more complicated interaction between variability and nonlinearity and while approximate relationships can be found for any particular model there is no simple general form. Nevertheless the coefficient of variation still provides a useful approximate measure of the effect of stochasticity. For a random vector **Z** the coefficient of variation is the matrix

$$\psi^2(\mathbf{Z}) = (\text{diag } E\mathbf{Z})^{-1} V\mathbf{Z}(\text{diag } E\mathbf{Z})^{-1},$$

which has i–jth element equal to the covariance of Z_i and Z_j divided by the product of their means. The formulae given below are for the single species case only, but it is quickly seen that the multispecies formulae are direct analogues and the conclusions below remain unchanged in the multispecies setting. All of these formulae can be derived easily by repeated application of the following formula or its multivariate analogue:

$$VZ = EV(Z|H) + VE(Z|H).$$

5.2. *The Magnitude of the Coefficient of Variation*

In order to obtain an explicit formula for the contribution of internal dynamics to the coefficient of variation we need to make some assumption about the probability distribution of $Z_j(t + 1)$ given $Z_j(t + h)$. For simplicity we shall assume that this distribution is Poisson with mean $g(Z_j(t + h))$. This is the sort of model that has been used to model growth processes over short periods of time (Bartlett, 1960) and while it is a gross simplification of reality this model embodies the essential feature for our present analysis, viz the variance is of the same order of magnitude as the mean as can be expected in a situation where the within-patch variability is made up of within- and between-individual variability alone.

The coefficient of variation can now be written down as follows:

$$\psi^2(Z_j(t + 1)) = 1/\bar{Z}_\infty(t + 1) + \psi^2(g(Z_j(t + h))), \tag{34}$$

$$\psi^2(Z_j(t + h)) = \mu(2 - \mu)/\bar{Z}_\infty(t) + (1 - \mu)^2\, \psi^2(Z_j(t)). \tag{35}$$

These two expressions are for the coefficient of variation immediately following a growth period and immediately following a migration period. The first term in each represents variability from the immediately preceding growth or migration period, while the second term represents a carry over of variability from earlier periods. In both cases the newly arisen variability is a decreasing function of average density $(\bar{Z}_\infty)$ becoming small for large densities but large for small densities. The reason for this result is that in these models movements and internal dynamics contain only within- and between-individual variation. When local temporal environmental variability is added, different results are obtained.

To add temporal environmental variability to the internal dynamic processes we introduce a nonnegative stochastic process $\rho_j(t + h)$ with mean 1, to represent the state of the local environment, and we now assume that given H_{t+h} and $\rho_j(t + h)$, $Z_j(t + 1)$ is Poisson with mean $g(Z_j(t + h))$ $\rho_j(t + h)$. Note that $E[Z_j(t + 1)|H_{t+h}] = g(Z_j(t + h))$ still, but the variance is different. Local temporal environmental variability in immigration can be handled by replacing the parameter μ with a stochastic process $\xi_j(t)$ having mean μ. We now obtain

$$\psi^2(Z_j(t + 1)) = 1/\bar{Z}_\infty(t + 1) + V\rho_j(t + h) + [V\rho_j(t + h)]\, \psi^2[\, g(Z_j(t + h))]$$
$$+ \psi^2[\, g(Z_j(t + h))] \tag{36}$$

and

$$\psi^2(Z_j(t + h)) = \mu(2 - \mu)/\bar{Z}_\infty(t) + V\xi_j(t) + (1 - \mu)^2\, \psi^2(Z_j(t)). \tag{37}$$

The first two terms in both expressions represent newly arisen variability. The local environmental variance is the second of these terms; it is independent of average population density and hence does not decline to 0 as average population density increases, indeed it puts a strict lower bound on the coefficient of variation. However, the coefficient of variation still becomes arbitraily large for arbitrarily small densities.

In summary, movements and internal dynamics tend to give the same conclusions. By making the population arbitraily sparse, our measure of within-patch variability (the coefficient of variation) can be made arbitrarily large. Conversely, if these is no local random environmental variation, this coefficient of variation will become small at high population densities. However, local temporal environmental variation introduces a term that is independent of density and this prevents the coefficient of variation from becoming small.

5.3. *Temporal Accumulation of Variability*

The analysis above is concerned only with that part of the coefficient of variation representing newly arisen variability. A detailed examination of the behavior of the second part, representing variability accumulated from previous time periods, is beyond the scope of this paper but linear autoregressive models can indicate the likely short term behavior of our models in cases where within-patch variability is small, and so we analyze these greatly simplified models. Models of these kinds have been used in ecology by Roughgarden (1975, 1977).

Let z^* be a positive constant corresponding to the equilibrium point of a deterministic model and let $\varepsilon_1, \varepsilon_2,...$ be an independent and identically distributed sequence of random variables with mean 0 and variance σ_0^2. Let $Z(0) = z^*$ and define

$$Z(t + 1) = z^* + \rho(Z(t) - z^*) + \varepsilon_t. \tag{38}$$

If σ_0^2 were zero then the resulting deterministic difference equation would be globally stable at z^* for $|\rho| < 1$. In the stochastic model $EZ(t) = z^*$, i.e.. is independent of t, and the coefficient of variation is simply proportional to the variance. For $\sigma_0^2 > 0$ we can calculate

$$VZ(t + 1) = \rho^2 VZ(t) + \sigma_0^2 \tag{39}$$

and so

$$VZ(t) = \sigma_0^2 \sum_{j=0}^{t} \rho^{2j} = \sigma_0^2(1 - \rho^{2t})/(1 - \rho^2), \qquad |\rho| \neq 1$$

$$= t\sigma_0^2, \qquad |\rho| = 1. \tag{40}$$

It is clear that $VZ(t)$ is always an increasing function of t for nonzero ρ– variability is virtually always accumulating. For the cases where the deterministic model is stable, the variability has a limiting value $\sigma_0^2/(1 - \rho^2)$ and we see that the ultimate magnitude of variability tends to ∞ as $|\rho| \to 1$, for all values of σ_0^2. A value of $|\rho|$ near 1 indicates a weakly stable deterministic model and in this case there will be a large accumulation of variability in the stochastic model.

In the neutrally stable case ($|\rho| = 1$), the accumulation of variability is linear and unbounded, but in the unstable case ($|\rho| > 1$) the accumulation is exponential and unbounded. In a sufficiently realistic population model there is likely to be an upper bound to variability, but these results may indicate the kind of behavior to be expected before that upper bound is reached and instabilities in the model may mean that the upper bound is high.

5.4. *Local Instability*

The results above indicate that instabilities in the deterministic analogue of a stochastic model facilitate the accumulation of variability. This conclusion is corroborated by the results of Becker (1973) and also by well-known results for branching processes (Jagers, 1975). This suggests that instabilities in the local dynamics of our models may lead to a large accumulation of within-patch variability and hence lead to large differences between the results of the deterministic and stochastic models. There is evidence for this in published models incorporating within-patch variability (Maynard Smith, 1974; Hastings, 1977; Zeigler, 1977; Caswell, 1978). In these models internal dynamics are deterministic and unstable (except for one of Hasting's models where internal dynamics are also stochastic) and within-patch variability is entirely due to migration. Despite these instabilities the stochastic versions are stable, i.e., allow species persistence for broad ranges of parameters. Thus variability has produced large qualitative differences between the results of the deterministic and stochastic models.

6. GENERALIZATIONS

There are many directions in which the work described in this paper can be generalized. The most important and most difficult direction is the consideration of more general types of migration. Other directions are the inclusion of other types of variability and non-Markov components. These latter directions amount to no more than modification of our framework and will not be discussed.

To include more general types of migration, consider a system of patches in which migration between patches and growth and interaction within patches are separated in time as in our previous models. Assume that

internal dynamics are the same as in our previous stochastic models but we make no assumptions about movements between patches. The simple deterministic model, which ignores all variability, needs no modification to apply here; it is simply $\bar{z}(t + 1) = g(\bar{z}(t))$. For this model to agree with the full stochastic model's answer to average population density then the following equation must hold:

$$EZ_j(t + 1) = Eg(Z_j(t + h)),\qquad(41)$$

that is, Eq. (11) of Section 3 must hold here as well. Because of nonlinearity and within-patch variability this equation will generally be false and the stochastic and deterministic models will give different answers for average population size. In a similar manner one can show deviations between $E[Z_j(t)|H_0]$ and the variable $z_j(t)$ in the modification of the complex deterministic model to the present setting. Thus our major conclusions about deviations between deterministic and stochastic models do not depend on the specific form of the migration model chosen.

Some generalizations of migration remain within the framework established in this paper, for example emigration can be made density dependent by dividing each population class into migrant and nonmigrant individuals. The migrants all emigrate during $(t, t + h)$ and the nonmigrants do not emigrate. Internal dynamic processes formally determine how many individuals belong to the migrating class and arbitrary density dependence can be included here. Death during migration, which is ignored in our models, can be accommodated in a similar way.

7. DISCUSSION

The stochastic effects discussed in this paper are often ignored by population modellers. The justification for doing so is the claim that such stochasticity, even though present in all real systems, has little effect on the population density of large populations. May (1974a) states this claim clearly and the most general demonstrations of the claim are to be found in Kurtz (1970) and Wang (1976). Kurtz's and Wang's results are based on models that make the unrealistic homogeneous interaction assumption, i.e., the chances of survival and reproduction for any individual depend on average population density as opposed to local population density. Their results concern the behavior of average population density over a large area supporting a correspondingly large population. They find that average population density will be near the prediction of the analogous deterministic model (the model corresponding to our "simple deterministic model"). Furthermore, the difference between average population density in the two models disappears completely as the area is increased.

Our results are to the contrary. We make the realistic assumptions that within-patch variability is present; there is inhomogeneous interaction and local dynamics are nonlinear. The conclusion is that average population density differs systematically from the prediction of the simple deterministic model. Moreover, this difference persists as the total area (number of patches) is increased to infinity. We do find that average population density will be close to its theoretical mean value when the number of patches is large, but this theoretical mean value is not given by the simple deterministic model. The differences observed here are not simply an artifact of our theoretical framework because when inhomogeneous interaction is unrealistically omitted from our models we are able to reproduce the results of Kurtz and Wang.

If within-patch variability is taken out of our models leaving inhomogeneous interaction and nonlinear dynamics we obtain our "complex deterministic model". It incorporates migration and local population growth in a deterministic way and is thus analogous to the models reviewed by Levin (1976), Mimura *et al.* (1978) and McMurtie (1978). The complex deterministic model gives different answers for average population density than both the stochastic model and the simple deterministic model. However, if there is no initial spatial variation it agrees exactly with the simple deterministic model. Moreover the complex deterministic model tends to underestimate spatial variation because it omits within-patch variability. In some cases this model predicts that all spatial variation disappears as t (time) $\to \infty$, contrary to the predictions of the stochastic model. When this happens the simple and complex deterministic models agree in the limit as $t \to \infty$ and both continue to differ from the stochastic model in their answers for averge population density. This disappearance of spatial variation in deterministic models is not unique to our models for it has been observed in other settings as well (Levin, 1974; Allen, 1975; Hastings, 1978b, Conway *et al.*, 1978).

We have seen that the complex deterministic model fails to describe accurately the average population and spatial variation. In addition it gives a different answer than the stochastic model for the mean population on a particular patch as a function of the initial populations on all patches.

The reasons for the differences between average population density in the stochastic models and the simple deterministic model can be outlined as follows: The mean of a nonlinear function of a random variable is generally different from the nonlinear function of the mean. Consequently nonlinear stochastic models will generally have a mean that differs from the value given by the analogous deterministic model. This is found in our models but is also a feature of most other models (Feller, 1939; Kendall, 1949; Becker, 1973; Getz, 1976). However, the difference between the two is small if the relevant variability is small. In models that make the homogeneous

interaction assumption the relevant variability is in averge population density, and this variability will be small whenever the area is large. However, for inhomogeneously interacting models, the relevant variability is in local population density and this variability does not decrease as the area is made large, and so the difference between the theoretical mean local density and the deterministic prediction does not disappear as the area is made large. Finally, in our models, there is sufficient independence between patches for a law of large numbers to be applicable which says that average population density for the entire area will be close to the theoretical mean density for a randomly chosen patch. As we have observed, this theoretical mean density differs from the prediction of the simple deterministic model. Hence average population density in the two models shows a systematic difference.

Essentially the same reasons are involved in the failure of the complex deterministic model to correctly predict mean local density and average population density of the stochastic model.

The magnitude of the difference that we observe between deterministic and stochastic models depends on the degree of nonlinearity of local dynamics and the magnitude of within-patch variability. If local dynamics are linear then there will be no difference between the deterministic and stochastic answers for average population density. Exact linearity is quite unrealistic but sometimes approximate linearity occurs and then the stochastic and deterministic models will give similar answers for average population density.

If within-patch variability is large, the stochastic and deterministic models can be expected to show large differences. If within-patch variability is small the stochastic models will be similar to the complex deterministic model. However, initial spatial variation must also be small if these models are to give similar results to the simple deterministic model except, perhaps, in the limit as time tend to ∞, when the initial spatial variation *may not* matter.

Within-patch variability will be large if local populations are small or if local temporal environmental variability is large. Within-patch variability can only be small if local populations are large and local temporal environmental variability is small. There is also some evidence that unstable local dynamics can interact with wihin-patch variability significantly enhancing the effect of the latter.

Population Models

We have shown that average population density in our more realistic stochastic models differs from the predictions of the usual kinds of deterministic models. However, we do know that average population density will not show very much random variation due to within-patch variability, when the number of patches is large, and so it is possible that some other kind of

deterministic model might usefully describe average population density. In fact we can find such deterministic models in some cases, for example, Chesson (1981b) shows that in the special case where all organisms migrate in every time period, it is possible to find a first order difference equation for average population density. However, the stochastic model is required for the deduction of this difference equation because it is not the deterministic analogue of the stochastic model. In other cases it may be possible to approximate the behavior of average population density using a difference equation of some order, probably higher than one.

Thus it appears that average population density may possibly be described by models of the classical difference equation kind. One could forget about within-patch variability, inhomogeneous interaction and the like, and concentrate on describing average population density in the usual deterministic manner. If some temporal environmental variability is added to this it would undoubtedly be possible to reproduce most behaviors exhibited by average population densities in the real world. However, this is mere curve fitting, it does not amount to explanation in terms of the biology of the organisms, i.e., the mechanisms. For example, Hastings (1977) gives two simple predator–prey models with inhomogeneous interaction and within-patch variability. In this system the predator always causes local extinction of the prey; but inhomogeneous interaction and within-patch variability make coexistence possible—they are the biological explanation for coexistence in such circumstances. Average population density in Hastings' models could reasonably be described by the usual sorts of simple deterministic models. However, the use of such descriptions, without reference to the basic stochastic mechanisms, does not help explain the observed phenomena; it does no more than describe them.

By showing how stochastic models give different results from their deterministic analogues, this paper indicates the possibility of stochastic explanations for observable population phenomena in a wide variety of circumstances. More work on stochastic mechanisms is needed. Deterministic models will remain useful descriptions of population phenomena whenever within-patch variability is small. In other cases stochastic models are required. Unfortunately stochastic models are generally more complex and therefore more difficult to analyze than deterministic models. However, there are methods of simplifying stochastic models by shifting the emphasis from deterministic to stochastic mechanisms as has been done in the many successful models involving within-patch variability mentioned in the introduction; this paper provides other methods of simplification and Chesson (1981a) gives necessary and sufficient conditions for the stability of the stochastic models introduced here. Stochastic models will become easier to study as different approaches are explored.

Most of the successful previous approaches to stochastic modelling of

inhomogeneously interacting populations have dealt with the situation where only presence and absence are recorded for the local populations. Caswell (1978) gives a review. Models which Caswell terms "open and equilibrial" correspond to our systems with infinitely many patches and "open nonequilibrial" correspond to our systems with finitely many patches. However, the distinction between the two is not as great as Caswell tends to suggest because the infinite sytems provide good approximations to the finite systems when the number of patches is large. This fact has been discussed for Caswell's model by Crowley (1979). Infinite systems are often easier to analyze than finite systems because they tend to be simpler and indeed many models of infinite systems have yielded to analytical methods (e.g.. Slatkin. 1974; Hastings, 1977, 1978a; Crowley, 1979).

Infinite systems that deal only with species presence and absence on patches may superficially appear to be deterministic even though they involve within-patch variability. It is true that in these models the relative frequency of path occupancy by a given species is a non-random or deterministic process because it corresponds to the spatial average of population density in our models. However, this relative frequency is also the probability that the species is present on any given patch, and at the local population level these models are highly stochastic. Moreover these models can generally be derived as limits of very recognizably stochastic models with a finite number of patches. As the number of patches becomes large the relative frequencies of patch occupancy become nonrandom but stochasticity at the level of a patch persists and the relative frequencies of patch occupancy depend in a systematic manner on this local stochasticity just as within-patch variability has systematic effects on average population density in our models.

A cautionary remark is necessary in respect of published models of infinite systems. Some of these are not genuine stochastic models and are not limits of genuine finite stochastic systems because the appropriate probability rules are violated. For example, this is so in the models of Horn and MacArthur (1972) and Vandermeer (1973), where the occupancy of patches by different species is assumed independent despite interaction between species on individual patches.

APPENDIX

A1. *Symmetric Distributions*

In the text it is assumed that $(\mathbf{Z}_1(0),..., \mathbf{Z}_k(0))$ has a symmetric distribution in keeping with the general symmetry of the model. This assumption does not restrict the application of the model at all because if $(\mathbf{Z}_1(0),..., \mathbf{Z}_k(0))$

does not have a symmetric distribution, and $(\rho_1,...,\rho_k)$ is a random permutation of $(1,...,k)$ then $(\mathbf{Z}_{\rho_1}(0),..., \mathbf{Z}_{\rho_k}(0))$ has a symmetric distribution. Thus a symmetric initial distribution is always created by simply randomly labelling the patches. The labelling of the patches is quite irrelevant to any of our considerations here. We are interested in the properties of the spatial average of the densities of patches, spatial variation, the development of spatial pattern given an initial pattern and the properties of a representative patch (one chosen at random). None of these properties depend on the labelling of the patches and so we are free to choose the simplest labelling, which is the random one.

A2. *Proof of the Approximation of Finite by Infinite Models*

In this section we stated and prove the main theorems in the paper which show how the finite space models are approximated by the infinite space models. It is necessary to distinguish between the population variables in the finite and infinite models and to indicate the dependence on k for the finite models. This will be done by adding a superscript k to the population variables in the finite models.

The first theorem concerns models with inhomogeneous ineraction.

THEOREM A1. *Under regularity conditions (discussed below) and for every n and every t, the joint distribution of $(\overline{\mathbf{Z}}_k(s), \mathbf{Z}_1^k(s),...,\mathbf{Z}_n^k(s), 0 \leqslant s \leqslant t)$ for the finite stochastic model converges to the joint distribution of $(\overline{\mathbf{Z}}_\infty(s), \mathbf{Z}_1(s),..., \mathbf{Z}_n(s), 0 \leqslant s \leqslant t)$ in the infinite stochastic model.*

As it is stated, the theorem concerns only finite dimensional distributions; however, convergence of finite dimensional distributions is sufficient to imply convergence of the infinite dimensional distributions on the countable product topology (Billingsley, 1968, p. 22). Therefore we can infer weak convergence of the distribution of $(\overline{\mathbf{Z}}_k(t), \mathbf{Z}_1^k(t),..., \mathbf{Z}_k^k(t), 0 \leqslant t < \infty)$.

The second theorem concerns models with homogeneous interaction.

THEOREM A2. *Under regularity conditions (discussed below) and for every n and every t, the joint distribution of $(\overline{\mathbf{Z}}_k(s), \mathbf{Z}_1^k(s),..., \mathbf{Z}_n^k(s), 0 \leqslant s \leqslant t)$ in the finite stochastic model with homogeneous interaction, converges to the joint distribution of $(\overline{\mathbf{Z}}_\infty(s), \mathbf{Z}_1(s),..., \mathbf{Z}_n(s), 0 \leqslant s \leqslant t)$ in the corresponding infinite model.*

Regularity Conditions

For Theorem A1 to be true it must at least be true for $t = 0$. The behavior at $t = 0$ cannot be deduced from the dynamic parts of the model and so must be assumed. However, these assumptions simply amount to saying that it is

meaningful to talk about an initial distribution of values for the population variables. They are:

1a. There is a probability distribution π_0 such that, for every nonnegative integer vector z, the quantity

$$k^{-1} \sum_{j=1}^{k} 1_{\{z\}}(\mathbf{Z}_j^k(0)), \tag{A1}$$

converges in probability to $\pi_0(\mathbf{z})$. Note that expression (A1) is the relative frequency of patches with population variables having the value z.

1b. $$E\mathbf{Z}_j^k(0) \to \int \mathbf{z} \, d\pi_0(\mathbf{z})$$

This second condition simply insists that the convergence in 1a is sufficiently regular for the mean to converge to the value given by the limiting distribution.

Some conditions need to be placed on the dynamic parts of the model to preclude irregular behavior. Four *alternative* conditions are given below. They are all conditions which ensure that when $\mathbf{Z}_j^k(t)$ converges in distribution to a random vector $\mathbf{Z}$, then $E\mathbf{Z}_j^k(t)$ will converge to $E\mathbf{Z}$.

2a. There is a vector $\mathbf{\alpha}$ and a matrix $\mathbf{A}$ such that $\mathbf{g}(\mathbf{z}) \leqslant \mathbf{\alpha} + \mathbf{A}\mathbf{z}$ for nonnegative z.

2b. There are constants $M_{ijn} < \infty$ such that $E \, |Z_{ij}^k(0)|^n \leqslant M_{ijn}$ for all k and the moments of the distribution $\gamma(\cdot, \mathbf{z})$ are bounded by polynomials in z.

2c. $$\lim_{M \to \infty} \sup_{k} E \, |\mathbf{Z}_j^k(t)| \, 1_{\{|\mathbf{Z}_j^k(t)| > M\}} = 0 \qquad \text{for every } t.$$

2d. There are functions $\rho(t) > 0$ and $K(t) < \infty$ such that

$$E \, |\mathbf{Z}_j^k(t)|^{1 + \rho(t)} \leqslant K(t) \qquad \text{for every } k.$$

Theorem A1 is true if 1a, 1b and one of 2a, 2b, 2c or 2d are true.
The regularity conditions for Theorem A2 are the same as those for Theorem A1 with the addition of the condition.

3. $\gamma(\cdot, \mathbf{z})$ is a weakly continuous function of z.
Theorem A2 is true if 1a, 1b, 3 and one of 2a, 2b, 2c, or 2d are true.
The proofs of the theorems proceed by proving that they are true for Assumption 2c and then showing that 2a, 2b and 2d each imply 2c.

Assumption 2c involves the concept of uniform integrability (Billingsley, 1971; Chung, 1974). Random variables $X_1, X_2,...,$ are a uniformly integrable sequence if

$$\lim_{M \to \infty} \sup_{k} E \, |X_k| \, 1_{\{|X_k| > M\}} = 0.$$

If $E|X_k|^{1+\rho} < K$ for all k, where K and ρ are positive constants, then $X_1, X_2, \ldots$ is uniformly integrable because

$$K \geqslant E|X_k|^{1+\rho} \geqslant M^\rho E|X_k| 1_{\{|X_k|>M\}}$$

and so

$$\sup_k E|X_k| 1_{\{|X_k|>M\}} \leqslant K/M^\rho$$

We see from this that Assumption 2d implies 2c, and is therefore a special case of 2c. In addition, if $Y_1, Y_2, \ldots$ is uniformly integrable and $|X_j| \leqslant |Y_j|$ then clearly $X_1, X_2, \ldots$ is uniformly integrable.

Interest in uniform integrability stems from the fact that if X_k converges in distribution to X and $X_1, X_2, \ldots$ is uniformly integrable then $EX_k \to EX$. Conversely, if $E|X_k| \to E|X| < \infty$ then $X_1, X_2, \ldots$ is uniformly integrable (Billingsley, 1971; Chung, 1974). These ideas are used in the proofs.

Preliminary Results

To prove theorems A1 and A2 we need to discuss some aspects of the theory of weak convergence of probability measures (Billingsley, 1968, 1971) and some results about random probability measures (Kallenberg, 1973, 1975; Chesson, 1976). We also need some lemmas for the proofs of the theorems.

Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_k$ be random vectors in R^m (real m dimensional space). Their empirical distribution is defined to be the random probability measure $p(B) = k^{-1} \sum_{j=1}^k 1_B(\mathbf{Z}_j)$, where B is a Borel set in R^m. The random probability measure p has a probability distribution on the space of probability measures endowed with the weak topology. The probability measure Ep is defined by the equation $(Ep)(B) = Ep(B)$ for Borel sets B. If f is a nonnegative measurable function then $E \int f \, dp = \int f \, dEp$.

To prove the theorems we need to relate convergence in distribution for $\mathbf{Z}_1, \ldots, \mathbf{Z}_k$ to convergence in distribution for their empirical distribution. This is done in the following theorem proved elsewhere (Kallenberg, 1973; Chesson, 1976). (The symbol "$\Rightarrow$" means "converges weakly in distribution.")

THEOREM A3. *Let $\mathbf{Z}_1^k, \ldots, \mathbf{Z}_k^k$ be jointly symmetrically distributed random vectors and p_k their empirical distribution. Then p_k converges weakly in distribution to a non random probability measure π if and only if $(\mathbf{Z}_1^k, \ldots, \mathbf{Z}_n^k) \Rightarrow (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$, for every n, where $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ are an independent and identically distributed sequence of random vectors with common distribution π.*

We also have to identify conditions under which convergence of p_k will imply convergence of its integrals $\int f \, dp_k$. This is the purpose of Lemma A4.

LEMMA A4. *Let p_k be a random probability measure on R^m such that $p_k \to \pi$, where π is non random. Let f be a non negative continuous function. If $\int f \, dEp_k \to \int f \, d\pi$ as $k \to \infty$ then $\int f \, dp_k \to \int f \, d\pi$.*

Proof. Define $f_M = \min(f, M)$ where M is some positive constant. The function f_M is bounded and continuous and so $\int f_M \, dp_k \Rightarrow \int f_M \, d\pi$.

Let $\varepsilon > 0$. Since $f_M \leqslant f$

$$P\left(\int f \, dp_k > \int f_M \, d\pi - \frac{\varepsilon}{2}\right) \geqslant P\left(\int f_M \, dp_k > \int f_M \, d\pi - \frac{\varepsilon}{2}\right).$$

Hence

$$\lim_{k \to \infty} P\left(\int f \, dp_k > \int f_M \, d\pi - \frac{\varepsilon}{2}\right) = 1. \tag{A2}$$

If $\int f \, d\pi = \infty$ then $\int f_M \, d\pi \to \infty$ as $M \to \infty$ and the result is proved. Now assume $\int f \, d\pi < \infty$.

Choose M so that $\int f \, d\pi - \int f_M \, d\pi < \varepsilon/2$ combining this with Eq. (A2) we get

$$\lim_{k \to \infty} P\left(\int f \, dp_k > \int f \, d\pi - \varepsilon\right) = 1 \tag{A3}$$

for all $\varepsilon > 0$. To complete the proof we must show that

$$\lim_{k \to \infty} P\left(\int f \, dp_k < \int f \, d\pi + \varepsilon\right) = 1 \tag{A4}$$

for all $\varepsilon > 0$. Let $c = \int f \, d\pi$ and define

$$\alpha(k, \varepsilon) = P\left(\int f \, dp_k > c - \varepsilon\right) \quad \text{and} \quad \beta(k, \varepsilon) = P\left(\int f \, dp_k \geqslant c + \varepsilon\right).$$

Let $\varepsilon, \varepsilon' > 0$ then

$$E\int f \, dp_k \geqslant (c - \varepsilon')[\alpha(k, \varepsilon') - \beta(k, \varepsilon)] + (c + \varepsilon)\beta(k, \varepsilon)$$
$$= (c - \varepsilon')\,\alpha(k, \varepsilon') + (\varepsilon + \varepsilon')\beta(k, \varepsilon).$$

Hence $c \geqslant (c - \varepsilon') + (\varepsilon' + \varepsilon)\,\overline{\lim}\,\beta(k, \varepsilon)$.

Letting $\varepsilon' \to 0$ we see that $\overline{\lim}\,\beta(k, \varepsilon) = 0$ for all positive ε proving (A4) and hence the required result.

The next lemma is needed in order to prove convergence inductively, viz., given convergence occurs at time t or $t + h$ to prove that convergence also occurs at time $t + h$ or $t + 1$.

LEMMA A5. *Let* $\mathbf{U}_k$, $\mathbf{V}_k$ *be sequences of random vectors in* R^α *and* R^β, *respectively. Suppose* $\mathbf{U}_k \Rightarrow \mathbf{U}$ *and* $P(\mathbf{V}_k \in B | \mathbf{U}_k) = \tau(B; \mathbf{U}_k)$, *where* $\tau(\cdot, \mathbf{u})$ *is a weakly continuous probability measure valued function of* $\mathbf{u}$. *Then* $(\mathbf{U}_k, \mathbf{V}_k) \Rightarrow (\mathbf{U}, \mathbf{V})$, *where* $\mathbf{V}$ *is a random vector such that* $P(\mathbf{V} \in B | \mathbf{U}) = \tau(B; \mathbf{U})$.

Proof. Let f be a bounded continuous function on R^β then $\int f(\mathbf{v}) \tau(d\mathbf{v}; \mathbf{u})$ is a bounded continuous function of $\mathbf{u}$. If g is a bounded continuous function on R^α then $Eg(\mathbf{U}_k) f(\mathbf{V}_k) = Eg(\mathbf{U}_k) \int f(\mathbf{v}) \tau(d\mathbf{v}; \mathbf{U}_k) \to Eg(\mathbf{U}) \int f(\mathbf{v}) \tau(d\mathbf{v}; \mathbf{U}) = Eg(\mathbf{U}) f(\mathbf{V})$, as $k \to \infty$, which proves the lemma.

The last result needed before proving the theorem is associated with condition 2b. It shows that the migration processes always satisfy a condition similar to that satisfied by the internal dynamics.

LEMMA A6. *In the finite stochastic model the conditional moments of* $Z_{ij}^k(t + h)$, *given* H_t *are bounded by polynomials in* $Z_{ij}^k(t)$ *and* $(k - 1)^{-1} \sum_{l \neq j} Z_{il}^k(t)$.

Proof. Let X be binomial with parameters (n, μ). The cumulant generating function of X is thus $n \log\{\mu(e^t - 1) + 1\}$, which means that the cumulants of X are proportional to n. It follows that the moments of X are polynomials in n.

If Y is binomial with parameters $(m, \mu/(k - 1))$ then the cumulant generating function of Y is

$$m \log\{\mu(e^t - 1)/(k - 1) + 1\} = \frac{m}{k - 1} \sum_{j=1}^{\infty} (-1)^{j+1} \frac{\mu^j (e^t - 1)^j}{j(k - 1)^{j-1}}.$$

It follows that the lth cumulant of Y is of the form

$$\frac{m}{k - 1} \mu \left[a_1 + a_2 \frac{\mu}{k - 1} + \cdots + a_l \left(\frac{\mu}{k - 1} \right)^{l-1} \right].$$

Hence the lth cumulant is bounded by $(m/(k - 1)) \mu [|a_1| + |a_2| \mu + \cdots + |a_l| \mu^{l-1}]$.

If X and Y are independent then the lth cumulant of $X + Y$ is bounded by a function of the form $K_1 n + K_2 m/(k - 1)$ and it follows that the moments of $X + Y$ are bounded by polynomials in n and $m/(k - 1)$.

The conditional distribution of $Z_{ij}^k(t + h)$ given H_t is a convolution of a binomial with parameters $(Z_{ij}^k(t), 1 - \mu_i)$ and a binomial with parameters

$((k-1)^{-1}\sum_{l\neq j}Z_{il}^k(t),\ \mu_i/(k-1))$ and it follows that the moments of the conditional distribution of $Z_{ij}^k(t+h)$ are bounded by polynomials in $Z_{ij}^k(t)$ and $(k-1)^{-1}\sum_{l\neq j}Z_{il}^k(t)$.

We now prove that Theorem A1 is true when conditions 1a, 1b, and 2c apply. The proof is by induction. At the tth stage assume that the joint distribution of $(\bar{Z}_k(s),Z_1^k(s),...,Z_n^k(s))$, $s\leqslant t$, converges to the joint distribution of $(\bar{Z}_\infty(s),Z_1(s),...,Z_n(s))$, $s\leqslant t$, for every n. Using this we prove that it is true for $t+1$ replacing t. That the induction hypothesis is true for $t=0$ is a consequence of 1a, 1b, Theorem A3 and Lemma A4.

The conditional probability generating function for $(Z_{i1}^k(t+h),...,Z_{in}^k(t+h))$, given H_t, is

$$E[\theta_1^{Z_{i1}^k(t+h)}\cdots\theta_n^{Z_{in}^k(t+h)}\,|\,H_t]$$
$$=\prod_{j=1}^n\left(\frac{k-n}{k-1}\mu_i+\theta_j[1-\mu_i]+\frac{\mu_i}{k-1}\left[\sum_{\substack{l=1}}^n\theta_l-\theta_j\right]\right)^{Z_{ij}^k(t)}$$
$$\times\prod_{j=n+1}^k\left(\frac{\mu_i}{k-1}\sum_{l=1}^n\theta_l+1-\frac{n\mu_i}{k-1}\right)^{Z_{ij}^k(t)}.\tag{A5}$$

Each factor is the generating function of a multinomial. The jth factor represents migration from patch j to patches 1 to n. (Chesson, 1976, discusses multinomial migration.)

Observing that

$$\prod_{j=n-1}^k\left(\frac{\mu_i}{k-1}\sum_{l=1}^n\theta_l+1-\frac{n\mu_i}{k-1}\right)^{Z_{ij}^k(t)}$$
$$=\left(\frac{\mu_i}{k-1}\sum_{l=1}^n\theta_l+1-\frac{n\mu_i}{k-1}\right)^{(k-1)\bar{Z}_{ik}(t)+\bar{Z}_{ik}(t)-Z_{i1}^k(t)-\cdots-Z_{in}^k(t)}\tag{A6}$$

we see that expression (A5) is a continuous function of $(\bar{Z}_{ik}(t),Z_{i1}^k(t),...,Z_{in}^k(t),(k-1)^{-1})$. This random vector $\Rightarrow(\bar{Z}_{i\infty}(t),Z_{i1}(t),...,Z_{in}(t),0)$, hence expression (A5) $\Rightarrow$

$$\prod_{j=1}^n(\mu_i+\theta_j[1-\mu_i])^{Z_{ij}(t)}\exp\left\{\mu_i\bar{Z}_{i\infty}(t)\sum_{j=1}^n(\theta_j-1)\right\}$$
$$=E[\theta_1^{Z_{i1}(t+h)}\cdots\theta_n^{Z_{in}(t+h)}\,|\,H_t].$$

Recalling that $\bar{Z}_k(t+h)=\bar{Z}_k(t)$, $\bar{Z}_\infty(t+h)=\bar{Z}_\infty(t)$ and $(Z_{i1}^k(t+h),...,Z_{in}^k(t+h))$, $i=1,...,r$ are conditionally independent given H_t, it is now clear that the conditional distribution of $(\bar{Z}_k(t+h),Z_1^k(t+h),...,Z_n^k(t+h))$, given H_t, converges in distribution to the conditional distribution of $(\bar{Z}_\infty(t+h),Z_1(t+h),...,Z_n(t+h))$ given H_t. The method of Lemma A5 now shows that $(\bar{Z}_k(s),Z_1^k(s),...,Z_n^k(s))$, $s\leqslant t+h\Rightarrow(\bar{Z}_\infty(s),Z_1(s),...,Z_n(s))$, $s\leqslant t+h$.

To prove the induction hypothesis for $t + 1$ we note that the conditional distribution of $\mathbf{Z}_1^k(t + 1),..., \mathbf{Z}_n^k(t + 1)$, given H_{t+h}, is $\gamma(\cdot, \mathbf{Z}_1^k(t + h)) \times \cdots \times \gamma(\cdot, \mathbf{Z}_n^k(t + h))$, which is a continuous function of $\mathbf{Z}_1^k(t + h),..., \mathbf{Z}_n^k(t + h)$ since the random vectors are discrete valued. By Lemma A5 $(\mathbf{Z}_1^k(s),..., \mathbf{Z}_n^k(s), s \leqslant t + 1) \to (\mathbf{Z}_1(s),..., \mathbf{Z}_n(s), s \leqslant t + 1)$. To complete the proof we need to show that $\bar{\mathbf{Z}}_k(t + 1) \Rightarrow \bar{\mathbf{Z}}_\infty(t + 1)$ and since the latter is non random the induction hypothesis will be proved for $t + 1$.

Define p_k to be the empirical distribution of $\mathbf{Z}_1^k(t + 1),..., \mathbf{Z}_k^k(t + 1)$ and π the distribution of $\mathbf{Z}_j(t + 1)$. By Theorem A3, $p_k \Rightarrow \pi$. From 2c, $E\mathbf{Z}_j^k(t + 1) \to E\mathbf{Z}_j(t + 1) = \bar{\mathbf{Z}}_\infty(t + 1)$ and so $E\bar{\mathbf{Z}}_k(t + 1) \to \bar{\mathbf{Z}}_\infty(t + 1)$. Lemma A4 now gives $\bar{\mathbf{Z}}_k(t + 1) \Rightarrow \bar{\mathbf{Z}}_\infty(t + 1)$ and Theorem A1 is proved under assumption 2c.

As remarked above, assumption 2d implies assumption 2c and so the theorem is also proved under assumption 2d. To prove Theorem A1 using 2a or 2b we can use all the steps in the above proof leading to $p_k \Rightarrow \pi$.

Assume 2a. Now $E\bar{\mathbf{Z}}_k(t + 1) = E\mathbf{Z}_j^k(t + 1) = E\mathbf{g}(\mathbf{Z}_j^k(t + h))$ and $\mathbf{g}(\mathbf{Z}_j^k(t + h)) \leqslant \alpha + A\mathbf{Z}_j^k(t + h)$. $E\mathbf{Z}_j^k(t + h) = E\mathbf{Z}_j^k(t) \to E\mathbf{Z}_j(t)$, by hypothesis, and so $E\mathbf{Z}_j^k(t + h) \to E\mathbf{Z}_j(t + h)$. From this we conclude that $\mathbf{Z}_j^k(t + h)$ is uniformly integrable. Therefore $\alpha + A\mathbf{Z}_j^k(t + h)$ is uniformly integrable and, as a consequence, so is $\mathbf{g}(\mathbf{Z}_j^k(t + h))$. Hence $E\mathbf{g}(\mathbf{Z}_j^k(t + h)) \to E\mathbf{g}(\mathbf{Z}_j(t + h)) = \bar{\mathbf{Z}}_\infty(t + 1)$, that is $E\bar{\mathbf{Z}}_k(t + 1) \to \bar{\mathbf{Z}}_\infty(t + 1)$ and by Lemma A4 $\bar{\mathbf{Z}}_k(t + 1) \Rightarrow \bar{\mathbf{Z}}_\infty(t + 1)$. Theorem A1 is now proved under assumption 2a.

Assume 2b. We prove Theorem A1 by proving that 2c holds. Since $E|Z_{ij}^k(0)|^{n+1} \leqslant M_{ijn+1}$ for all k and n it follows that $|Z_{ij}^k(0)|^n$ is uniformly integrable for all n and hence $E|Z_{ij}^k(0)|^n \to E|Z_{ij}(0)|^n$ for all n. If

$$E|Z_{ij}^k(s)|^n \to E|Z_{ij}(s)|^n < \infty \tag{A7}$$

for all n, then 2c holds. We prove statement (A7) by induction. If statement (A7) holds for all $s \leqslant t$ then $\mathbf{Z}_j^k(t + h) \Rightarrow \mathbf{Z}_j(t + h)$. Now $E[|Z_{ij}^k(t + h)|^n | H_t]$ is bounded by a polynomial in $1/(k - 1) \sum_{l \neq j}^k Z_{il}^k(t)$ and $Z_{ij}^k(t)$ (Lemma A6). Using statement (A7) and the inequalities

$$\left| \frac{1}{k - 1} \sum_{l \neq j}^k Z_{il}^k(t) \right|^n \leqslant \frac{1}{k - 1} \sum_{l \neq j}^k |Z_{il}^k(t)|^n,$$

$$E \left| \frac{1}{k - 1} \sum_{l \neq j}^k Z_{il}^k(t) \right|^n |Z_{ij}(t)|^m$$

$$\leqslant \left(E \left| \frac{1}{k - 1} \sum_{l \neq j}^k Z_{il}^k(t) \right|^{2n} \right)^{1/2} (E|Z_{ij}(t)|^{2m})^{1/2}$$

it follows that $E|Z_{ij}^k(t + h)|^n$ is bounded as $k \to \infty$, for all n, which means $|Z_{ij}^k(t + h)|^n$ is uniformly integrable for all n, and so statement (A7) holds for $s = t + h$. Similar reasoning, using the fact that the moments of $\mathbf{Z}_j^k(t + 1)$

given H_{t-h} are bounded by polynomials in $\mathbf{Z}_j^k(t+h)$, implies statement (A7) for $s = t + 1$. This completes the proof.

To prove Theorem A2 we prove that

$$(\overline{\mathbf{Z}}_k(s), \mathbf{Z}_1^k(s),..., \mathbf{Z}_n^k(s)),\, s \leqslant t \Rightarrow (\overline{\mathbf{Z}}_\infty(s), \mathbf{Z}_1(s),..., \mathbf{Z}_n(s)),\, s \leqslant t.$$

for all n, using induction on t. The $t = 0$ case is dealt with in the proof of Theorem A1.

The conditional distribution of $\mathbf{Z}_1^k(t+1),..., \mathbf{Z}_n^k(t+1)$ given H_t is

$$\gamma(\cdot, \overline{\mathbf{Z}}_k(t)) \times \cdots \times \gamma(\cdot, \overline{\mathbf{Z}}_k(t)),$$

which is a weakly continuous function of $\overline{\mathbf{Z}}_k(t)$. By Lemma A5

$$(\mathbf{Z}_1^k(s),..., \mathbf{Z}_n^k(s)),\, s \leqslant t + 1 \Rightarrow (\mathbf{Z}_1(s),..., \mathbf{Z}_n(s)),\, s \leqslant t + 1.$$

Theorem A3 and Lemma A4 imply $\overline{\mathbf{Z}}_k(t+1) \Rightarrow \overline{\mathbf{Z}}_\infty(t+1)$ provided assumption 2c holds. Thus Theorem A2 is proved for assumptions 2c and 2d. The proofs for 2a and 2b are simplified parallels of the corresponding proofs for Theorem A1 and are therefore omitted.

A3. One Dimensional Deterministic Models

Consider the difference equation

$$z_{t+1} = g(z_t), \tag{A8}$$

where g is a strictly concave increasing function on $[0, \infty)$ with $g(0) = 0$. The function

$$g(z) = K(1 - e^{-\beta z}) \tag{A9}$$

from Example 1 is of this kind. If Eq. (A8) has a positive equilibrium point z^* then this equilibrium is globally stable, that is, for $z_0 \in (0, \infty)$, $z_t \to z^*$ as $t \to \infty$. If there is no positive equilibrium point then $z_t \to 0$ or $z_t \to \infty$. The function (A9) has a positive equilibrium if $g'(0) = K\beta > 1$ because $g(z) > z$ for z sufficiently small and $g(z) < z$ for z sufficiently large. If $K\beta \leqslant 1$, $z_t \to 0$ for this g. We have proved:

LEMMA A3.1. *For example* 1, *the simple deterministic model says*

$$\lim_{t \to \infty} \bar{z}(t) = 0 \qquad if \quad K\beta \leqslant 1.$$

but if $K\beta > 1$ *there is a positive number* z^* *such that*

$$z^* = (1 - e^{-\beta z^*}) \qquad (A10)$$

and $\lim_{t \to \infty} \bar{z}(t) = z^*$ for $\bar{z}(0) > 0$.

We now consider one dimensional versions of the complex deterministic model with increasing strictly concave g. We assume $\mu > 0$.

THEOREM A3.2. *If* $0 < \bar{z}(0) < \infty$ *and Eq. (A8) has a positive equilibrium* z^* *then*

$$\lim_{t \to \infty} z_j(t) = \lim_{t \to \infty} \bar{z}(t) = z^* \qquad \text{for all } j.$$

Proof. First of all we note that if $z_j(t) \geqslant a$, for all j, where a is a constant, then $z_j(t + 1) \geqslant g(a)$, for all j, because $\bar{z}(t) \geqslant a$, $z_j(t + h) = (1 - \mu)$ $z_j(t) + \mu\bar{z}(t) \geqslant a$, $z_j(t + 1) = g(z_j(t + h))$ and $g(z_j(t + h)) \geqslant g(a)$, since g is monotonic.

Now $z_j(h) \geqslant \mu\bar{z}(0) > 0$ and so $z_j(1) \geqslant g(\mu\bar{z}(0))$. It follows that $z_j(t) \geqslant g_{t-1}(\mu\bar{z}(0))$, where g_{t-1} is g composed with itself $t - 1$ times. It follows that

$$\underline{\lim} \, z_j(t) \geqslant z^*. \qquad (A11)$$

Now $\bar{z}(t + 1) = Eg(z_j(t + h)) \leqslant g(Ez_j(t + h)) = g(\bar{z}(t))$, the inequality being a consequence of concavity of g. Since g is strictly increasing we have

$$\overline{\lim} \, \bar{z}(t) \leqslant \overline{\lim} \, g_t(\bar{z}(0)) = z^* \qquad (A12)$$

Combining (A11) and (A12) and using Fatou's lemma we get

$$z^* \leqslant \underline{\lim} \, \bar{z}(t) \leqslant \overline{\lim} \, \bar{z}(t) \leqslant z^*,$$

i.e.,

$$\lim \bar{z}(t) = z^*.$$

From (A11) we have $\overline{\lim} \, z_j(t) \geqslant z^*$. Also, since $z_j(t + 1) \leqslant z_j(t)$ if $z_j(t) \geqslant \max(z^*, \bar{z}(t))$, we have $\overline{\lim} \, z_j(t) < \infty$. Now $z_j(t + 1) = g((1 - \mu) z_j(t) + \mu\bar{z}(t))$ and so $z^* \leqslant \overline{\lim} \, z_j(t) = g((1 - \mu) \overline{\lim} \, z_j(t) + \mu z^*)$ which means $\overline{\lim} \, z_j(t) = z^*$. Combining this with (A11) we find $\lim z_j(t) = z^*$.

A4. *Means of Non Linear Functions*

THEOREM A4.1. *Let* g *be a function on* R^r *satisfying*

$$\int g(\mathbf{z}) \, dP(\mathbf{z}) = g \left(\int \mathbf{z} \, dP(\mathbf{z}) \right) \qquad (A13)$$

*for every probability measure P on the non negative integers for which either
side of* (A13) *is defined. Then g must be of the form* $g(z) = a + b'z$ *for* $z \geq 0$.

Proof. Consider the $r = 1$ case first. Let $a = g(0)$, $b = g(1) - g(0)$ and
define $P(\{0\}) = 1 - 1/n$, $P(\{n\}) = 1/n$. Equation (A14) gives $g(1) = (1 - 1/n)$
$g(0) + (1/n) g(n)$, which rearranges to $g(n) = a + bn$. Defining
$P(\{n\}) = 1 - \alpha$, $P(\{n + 1\}) = \alpha$ we get $g(n + \alpha) = a + b(n + \alpha)$ proving that
g is linear on R^+.

This one dimensional demonstration immediately extends to show that, in
general, g must be linear in each variable when the others are held constant.
In particular $g(z_1, \mathbf{z}_2) = a(\mathbf{z}_2) + b(\mathbf{z}_2) z_1$.

Setting $z_1 = 0$ shows that a is linear in each component of $\mathbf{z}_2$. Setting
$z_1 = 1$ and subtracting a shows that the same must be true of b. We can
write $b(z_2, \mathbf{z}_3) = \alpha(\mathbf{z}_3) + \beta(\mathbf{z}_3) z_2$ and

$$g(z_1, z_2, \mathbf{z}_3) = a(z_2, \mathbf{z}_3) + \alpha(\mathbf{z}_3) z_1 + \beta(\mathbf{z}_3) z_1 z_2$$

Thus $g(n, n, \mathbf{z}_3) = a(n, \mathbf{z}_3) + \alpha(\mathbf{z}_3) n + \beta(\mathbf{z}_3) n^2$. Defining $P(\{0, 0, \mathbf{z}_3\}) = \frac{1}{2}$,
$P(\{2n, 2n, \mathbf{z}\}) = \frac{1}{2}$, using (A14) and the linearity of a in z_2 we get

$$g(n, n, \mathbf{z}_3) = a(n, \mathbf{z}_3) + \alpha(\mathbf{z}_3) n + 2\beta(\mathbf{z}_3) n^2,$$

which implies $\beta \equiv 0$. Hence

$$g(z_1, z_2, \mathbf{z}_3) = a(z_2, \mathbf{z}_3) + \alpha(\mathbf{z}_3) z_1.$$

Chasing through variables z_3 to z_r in this manner one sees that b is a
constant b_1, say. Subtracting $b_1 z_1$ from $g(\mathbf{z})$ and repeating the procedure it is
clear that $g(\mathbf{z}) = a + \sum_{i=1}^{r} b_i z_i$ where a and $b_1, ..., b_r$ are constants.

This theorem shows that for a given nonlinear function g there will always
be a probability measure P for which (A13) fails to hold. For some
nonlinear functions there will be probability measures for which (A14) does
hold. We now go on to discuss the sense in which such probability measures
are rare. For simplicity we consider only the one dimensional case.

A probability measure on $\{0, 1, 2, ...\}$ is defined by a sequence of non
negative numbers $p_1, p_2, ...$ with sum ≤ 1 ($p_0 = 1 - \sum_1^\infty p_i$). We ask the
following question. If $p_1, p_2, ...$ are generated by a random mechanism (i.e.,
are random variables) what is the probability that $\sum g(i) p_i = g(\sum i p_i)$ for
given nonlinear g? The theorem below shows that this probability is 0 when
the distribution of $(p_1, p_2, ...)$ belongs to a general class of infinite dimen-
sional continuous distributions. Before stating the theorem we need to
discuss the space on which these distributions are defined.

Let g be nonlinear and continuously differentiable. Without loss of
generality we can assume $g(0) = 0$. Define $\Omega = \{\mathbf{p} = (p_1, p_2, ...) | p_i \geq 0$, for
all i, $\sum p_i \leq 1$, $\sum (i + |g(i)|) p_i < \infty\}$. Thus Ω is the space of probabilities

for which both $\sum ip_i$ and $\sum g(i)p_i$ exist as finite numbers. It is easy to see that Ω is a complete separable metric space with metric

$$\rho(\mathbf{p}, \mathbf{q}) = \sum (i + |g(i)|) |p_i - q_i|.$$

For $\mathbf{p} \in \Omega$ define $f(\mathbf{p}) = \sum g(i)p_i - g(\sum ip_i)$. For many g the function f will never be 0 in which case we need not proceed further. Otherwise let $\mathbf{p}_0 = (p_{10}, p_{20},...)$ be a point such that $f(\mathbf{p}_0) = 0$.

To apply the implicit function theorem to f we calculate

$$\frac{\partial}{\partial p_i} f(\mathbf{p}) = g(i) - g'\left(\sum jp_j\right) i.$$

At $\mathbf{p}_0$ this is $g(i) - g'(\sum jp_{j0}) i$. The derivative can only be 0 for all i if g is linear on the integers—a case specifically excluded. Therefore there is an i for which this derivative is non zero. By the implicit function theorem there is a continuously differentiable function h^* of 2 variables defined on an open neighbourhood S of $(\sum_{j \neq i} jp_{j0}, \sum_{j \neq i} g(j)p_{j0})$ such that $p_i = h^*(\sum_{j \neq i} jp_j, \sum_{j \neq i} g(j)p_j)$ in a neighbourhood of $(p_{i0}, \sum_{j \neq i} jp_{j0}, \sum_{j \neq i} g(j)p_{j0})$. Since p_i, $\sum_{j \neq i} jp_j$, $\sum_{j \neq i} g(j)p_j$ are continuous function on Ω it follows that we can write

$$p_i = h^*\left(\sum_{j \neq i} jp_j, \sum_{j \neq i} g(j)p_j\right) = h(\mathbf{p}), \tag{A14}$$

where h is a continuous function on Ω, not depending on p_i, and the relation (A14) is true for an open neighbourhood of $\mathbf{p}_0$ in Ω.

We are now ready to prove the following:

THEOREM A4.2. *Suppose that g is continuously differentiable and nonlinear on the non negative integers. Let $(p_1, p_2,...)$ have a probability distribution on Ω such that the conditional distribution of p_i, given $\{p_j, j \neq i\}$, is absolutely continuous with respect to Lebesgue measure, for all i. Then $P(\sum g(i)p_i = g(\sum ip_i)) = 0$.*

Proof. If $f(\mathbf{p}) \neq 0$ for all $\mathbf{p} \in \Omega$ the theorem is proved. Otherwise let $\mathbf{p}_0$, h and S be as above. Now

$$P(S \cap \{f(\mathbf{p}) = 0\}) = P(S \cap \{p_i = h(\mathbf{p})\})$$
$$= EP(S \cap \{p_i = h(\mathbf{p})\} | \{p_j, j \neq i\})$$
$$= E0 = 0$$

because the conditional distribution of p_i, given $\{p_j, j \neq i\}$, is absolutely continuous.

Since the space is separable, the set $\{f(\mathbf{p}) = 0\}$ is contained in a countable union of open neighbourhoods S on which $p_i = h(\mathbf{p})$, for some i and some continuous function h not depending on p_i. It follows that

$$P\{f(\mathbf{p}) = 0\} = 0,$$

proving the theorem.

ACKNOWLEDGMENTS

REFERENCES

ALLEN, J. C. 1975. Mathematical models of species interactions in time and space, *Amer.Natur.* **109**, 319–341.

ANDREWARTHA, H. G.AND BIRCH, L. C. 1954. "The Distribution and Abundance of Animals," Univ. Chicago Press, Chicago.

BARBOUR, A. D. 1976. Quasi-stationary distributions in Markov population processes, *Adv.Appl. Probab.* **8**, 296–314.

BARTLETT, M. S. 1960. "Stochastic Population Models in Ecology and Epidemiology," Methuen, London.

BECKER, N. G. 1970. Interactions between species: Some comparisons between deterministic and stochastic models, *Rocky Mount. J. Math.* **3**, 53–68.

BILLINGSLEY, P. 1968. "Convergence of Probability Measures," Wiley, New York, 253 pp.

BILLINGSLEY, P. 1971. "Weak Convergence of Measures," SIAM, Philadelphia.

BIRCH, L. C. 1970. The role of environmental heterogeneity and genetical heterogeneity in determining distribution and abundance, *in* "Proceedings, Advanced Study Institute on Dynamics of Numbers in Populations," Oosterbeek, The Netherlands, 7–18 September 1970.

CASWELL, H. 1978. Predator- mediated coexistence: a non equilibrium model, *Amer. Natur.* **112**, 127–154.

CHESSON, P. L. 1976. "Models for Animal Movements," Ph.D. thesis, University of Adelaide, South Australia.

CHESSON, P. L. 1978. Predator–prey theory and variability, *Ann. Rev.Ecol. Syst.* **9**, 323–347.

CHESSON, P. L. 1981a. The stability of a spatially distributed population, manuscript.

CHESSON, P. L. 1981b. Are population fluctuations dampened by with-patch variability? to appear.

CHUNG, K. L. 1974. "A Course in Probability Theory," 2nd ed. Academic Press, New York.

CONNELL, J. H. 1978. Diversity in tropical rain forests and coral reefs, *Science* **199**, 1302–1310.

CONWAY, E., HOFF, D., AND SMOLLER, J. 1978. Large time behavior of solutions of systems of nonlinear reaction-diffusion equations, *SIAM J. Appl. Math.* **35**, 1–16.

CROWLEY, P. H. 1977. Spatially distributed stochasticity and the constancy of ecosystems, *Bull. Math. Biol.* **39**, 157–166.

DEN BOER, P. J. 1968. Spreading of risk and stabilization of animal numbers, *Acta Biotheoret.* **18**, 165–194.

FELLER, W. 1939. Die Grundlagen der volterraschen Theorie des Kampes ums Dasein in Wahrscheinlichkeitstheorie der Behandlung, *Acta Biotheoret.* **5**, 11–40.

GETZ, W. M. 1976. Stochastic equivalents of linear and Lotka–Volterra systems of equations—a general birth-and-death process formulation, *Math. Biosci.* **29**, 235–258.

GURNEY, W. S. C. AND NISBET, R. M. 1978a. Single species population fluctuations in patchy environments, *Amer. Natur.* **112**, 1075–1090.

GURNEY, W. S. C. AND NISBET, R. M. 1978b. Predator–prey fluctuations in patchy environments, *J. Anim. Ecol.* **47**, 85–102.

HASTINGS, A. 1977. Spatial heterogeneity and the stability of predator–prey systems, *Theoret. Pop. Biol.* **12**, 37–48.

HASTINGS, A. 1978a. Spatial heterogeneity and the stability of predator–prey systems: Predator mediated coexistence, *Theoret. Pop. Biol.* **14**, 380–395.

HASTINGS, A. 1978b. Global stability in Lotka–Volterra systems with diffusion, *J. Math. Biol.* **6**, 163–168.

HORN, H. S. AND MACARTHUR, R. H. 1972. Competition among fugitive species in a harlequin environment, *Ecology* **53**, 749–752.

HUTCHINSON, G. E. Copepodology for the ornithologist, *Ecology* **32**, 571–577.

JAGERS, P. 1975. "Branching Processes with Biological Applications," Wiley, London.

KALLENBERG, O. 1973. Canonical representations and convergence criteria for processes with interchangeable increments, *Z. Wahrsch. Verw. Gebiete* **27**, 23–36.

KALLENBERG, O. 1975. Random measures, *Schriftenr. Zentralinstitut Math. Mech.* **23**, 1–104.

KENDALL, D. G. 1949. Stochastic processes and population growth, *J. Roy. Statist. Soc. B* **11**, 230–265.

KURTZ, T. G. 1970. Solutions of ordinary differential equations as limits of pure jump Markov processes, *J. Appl. Probab.* **7**, 49–58.

LEVIN, S. A. 1974. Dispersion and population interactions, *Amer. Natur.* **108**, 207–228.

LEVIN, S. A. 1976. Population dynamic models in heterogeneous environments, *Ann. Rev. Ecol. Syst.* **7**, 287–310.

LEVIN, S. A. AND PAINE, R. T. 1974. Disturbance, patch formation and community structure, *Proc. Nat. Acad. Sci. U.S.A.* **71**, 2744–2747.

MCMURTIE, R. 1978. Persistence and stability of single-species predator–prey systems in spatially heterogeneous environments, *Math. Biosc.* **39**, 11–51.

MAY, R. M. 1974a. "Stability and Complexity in Model Ecosystems," 2nd ed., Princeton Univ. Press, Princeton, N.J.

MAY, R. M. 1974b. Ecosystem patterns in randomly fluctuating environments, *in* "Progress in Theoretical Biology" (R. Rosen and F. Snell, Eds.), Academic Press, New York.

MAY, R. M. 1976 ed. "Theoretical Ecology, Principles and Applications," Saunders, Philadelphia.

MAY, R. M. and Oster, G. 1976. Bifurcations and dynamic complexity in simple ecological models, *Amer. Natur.* **110**, 573–599.

MAYNARD SMITH, J. 1974. "Models in Ecology," Cambridge Univ. Press, Cambridge.

MIMURA, M. AND MURRAY, J. D. 1978. On a diffusive prey–predator model which exhibits patchiness, *J. Theoret. Biol.* **75**, 249–262.

MURDOCH, W. W. 1979. Predation and the dynamics of prey populations, *Fortschr. Zool.* **25**, 296–310.

ROFF, D. A. 1974a. Spatial heterogeneity and the persistence of populations, *Oecologia* **15**, 245–258.

ROFF, D. A. 1974b. The analysis of a population model demonstrating the importance of dispersal in a heterogeneous environment, *Oecologia* **15**, 259–275.

ROUGHGARDEN, J. 1975. A simple model for population dynamics in a stochastic environment, *Amer. Natur.* **109**, 713–736.

ROUGHGARDEN, J. 1977. Patchiness in the spatial distribution of a population caused by stochastic fluctuations in resources, *Oikos* **29**, 52–59.

SLATKIN, M. 1974. Competition and regional coexistence, *Ecology* **55**, 128–134.

VANDERMEER, J. H. 1973. On the regional stabilization of locally unstable predator–prey relationships, *J. Theoret. Biol.* **41**, 161–170.

WANG, F. J. S. 1975. Limit theorems for age and density dependent stochastic population models. *J. Math. Biol.* **2**, 373–400.

WIENS, J. A. 1976. Population responses to patchy environments, *Ann. Rev. Ecol. Syst.* **7**, 81–120.

ZEIGLER, D. P. 1977. Persistence and patchiness of predator–prey systems induced by discrete event population exchange mechanisms, *J.Theoret. Biol.* **67**, 687–713.