CHAPTER 1

Using multivariate state-space models to study spatial structure and dynamics

Richard A. Hinrichsen and Elizabeth E. Holmes

1.1 Introduction

Populations in nature are rarely unstructured, that is acting as a single, well-mixed, and random-mating unit. Instead populations are structured by various mechanisms. One ubiquitous mechanism that structures populations is spatial subdivision—spread out across multiple sites, populations naturally form subpopulations that covary to a restricted degree.

In our work as population analysts, we are primarily concerned with risk assessment and forecasting of imperiled populations. Understanding the spatial structure within a population is important in this context, because spatial structure has a strong effect on extinction risk and the degree to which a population is buffered from environmental fluctuations. Monitoring data collected for populations of concern reflect this ubiquitous spatial structure. Such data are typically collected from multiple census sites, sites which are often thought to represent subpopulations that are at least partly independent. Modeling multi-site data, however, presents serious challenges for population analysts. In many cases, monitoring data are limited to simple abundance counts, and data necessary to specify spatial structure—movement patterns or common environmental drivers—are missing. This hinders the use of mechanistic spatial models which require knowledge of the movement and the covariance of environmental drivers throughout a landscape (such as the models used in Lahaye *et al.* 1994, Dunning *et al.* 1995, and Schumaker *et al.* 2004).

Recently, statistical approaches based on time-series analysis and maximum-likelihood estimation[‡] have been developed to analyze population count data and infer underlying dynamics (Lindley 2002, Holmes and Fagan 2002, Holmes 2004, Staples *et al.* 2004, Dennis *et al.* 2006). These methods are based on research concerning the asymptotic distributions of abundance that evolve from stochastic population

[‡] Bayesian state-space approaches have also been developed, however, this chapter focuses exclusively on frequentist approaches and research.

processes (Tuljapurkar and Orzack 1980, Dennis *et al.* 1991, Holmes and Semmens 2004, Holmes *et al.* 2007). To date, this research has focused on the analysis of single population time series and how to deal with multiple sources of variability, specifically variability from environmental fluctuations and from measurement errors. These approaches use univariate state-space models that incorporate both variance in population growth due to process error and variance in the observations due to measurement error.

In this chapter, we extend this theory and present an analytical framework for the analysis of multi-site count data based on *multivariate* state-space models. This work has two related objectives. The first objective is to estimate the stochastic growth rates and variances that drive the dynamics of the population given a *known* or hypothesized spatial structure within the population. The second objective is to infer the spatial pattern of synchrony and correlation across sites. The latter objective allows us to make statistical inferences about which groups of sites act as independent subpopulations with uncorrelated changes in population abundance, which groups of sites act as independent but correlated subpopulations, and whether the sites appear to be independent observations of a single population. Figure 1.1 illustrates some of the different structures that a group of five sites might have: independent with different growth rates, independent with a shared growth rate and uncorrelated or correlated variability, and fully synchronized such that they appear to be observations of a single population. By synchronized, we mean that the sites not only have correlated changes in abundance but the sites also track each other over time (without diverging).

The methods described in this chapter are designed to infer the spatial patterns of synchrony and correlation by disentangling the variability due to measurement error, which causes the appearance of asynchrony and uncorrelation, from the underlying variability in population counts due to temporal variability in growth rates. However, we do not model the mechanisms causing these patterns explicitly—rather the methods look for the consequences: synchrony and correlation. For example, dispersal is one mechanism that can synchronize population dynamics. We do not model movement rather we model the resulting synchrony. Similarly common abiotic environmental drivers, common exposure to diseases and a common prey base can cause correlated growth rates, but we do not model drivers explicitly only the resultant correlation. Determining what mechanisms drive patterns of synchronization and correlation revealed by the analysis would require separate studies and different types of data, however the patterns that are revealed may suggest which mechanisms are more likely and guide further data collection and analysis.

1.2 Multivariate state-space models for multi-site population processes

The stochastic exponential model with Gaussian errors is the asymptotic approximation for a wide-variety of density-independent population processes, including complex age-structured and spatially-structured processes (Holmes *et al.* 2007). As

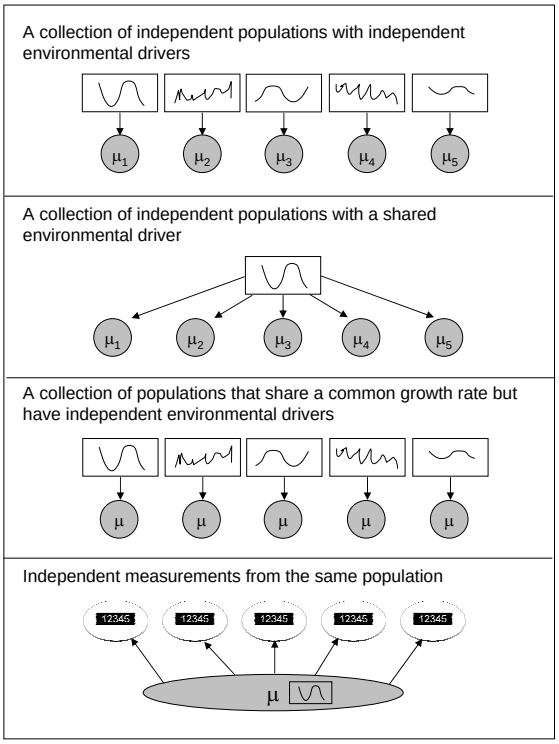

Figure 1.1 *Some of the different spatial structures possible for multi-site data. The μ in the figures refers to the subpopulation's stochastic growth.*

such, this model is the foundation of much research on stochastic population dynamics. Written in log space, this model is

$$x_t = x_{t-1} + \mu + e_t, \tag{1.1}$$

where x_t represents the log-population abundance at time t, and μ is the mean rate of population growth per time step. The process-error term, e_t, represents the stochastic deviations in population-growth rate though time. The process errors are assumed to have a normal distribution† with a mean of zero and constant variance. The stochastic exponential model is closely related to the stochastic Gompertz model, $x_t = bx_{t-1} + \mu + e_t$, which is the stochastic approximation for a variety of density-dependent processes (Ives *et al.* 2003, Dennis *et al.* 2006). Although we use the stochastic exponential process in this chapter, the framework we present can be used for a stochastic Gompertz process also.

† The normality assumption arises not from convenience but from the multiplicative nature of population growth. As a result, the error terms in a process become normal (in log space) over multiple time steps.

Suppose that instead of a single population, there are m subpopulations, which together comprise the total population. We can model the dynamics of this type of population using a multivariate stochastic exponential model:

$$\mathbf{X}_t = \mathbf{X}_{t-1} + \mathbf{B} + \mathbf{E}_t, \tag{1.2}$$

where $\mathbf{X}_t$ is an $m \times 1$ vector of log abundance in each of the m subpopulations at time t. $\mathbf{B}$ is an $m \times 1$ vector of the underlying stochastic growth rates, $\mu_1, \mu_2, \ldots, \mu_m$, in each of the m subpopulations. The process-error term, $\mathbf{E}_t$, is an $m \times 1$ vector of the serially uncorrelated stochastic deviations in each subpopulation's growth rate at time t. We assume that the process errors can be correlated between subpopulations by specifying that $\mathbf{E}_t$ has a multivariate normal distribution with a mean of zero and an $m \times m$ covariance matrix $\mathbf{Q}$.

Monitoring data also contain variance due to measurement error, and this will confound the estimation of $\mathbf{Q}$—the variance due to process error. Recent methods for addressing measurement error in population data use state-space models, which combine a model for the hidden true abundances with a model for the observations of the true abundance (deValpine and Hastings 2002, Lindley 2002, Dennis *et al.* 2006, Holmes *et al.* 2007). We use the same approach to address measurement error in multi-site data by using a multivariate state-space model. This is achieved by combining equation 1.2 with a measurement equation that relates the observed values of log abundance at time t to the true abundances at time t:

$$\mathbf{Y}_t = \mathbf{Z}\mathbf{X}_t + \mathbf{D} + \boldsymbol{\Upsilon}_t. \tag{1.3}$$

$\mathbf{Z}$ is an $n \times m$ matrix that defines how the n observations relate to the m true abundances; in general, the n observations could be any additive combination of the m true abundances. The $n \times 1$ vector $\mathbf{D}$ specifies the bias between the observations and the true abundances. The measurement errors at time t are denoted by $\boldsymbol{\Upsilon}_t$, which is an $n \times 1$ vector of serially uncorrelated disturbances with a mean of zero and an $n \times n$ covariance matrix $\mathbf{R}$. It is important to note that $\mathbf{R}$ and $\mathbf{D}$ are not the same as the variance and bias in the sampling process—for example, the errors resulting from counting animals from, say, a plane or the errors resulting from only sampling along a transect. The sampling process is but one source, and probably a minor source, of measurement variability and bias in many population data sets. Bigger sources of measurement variability come from temporal changes in sightability due to effects of age-structure, effects of environmental conditions, and changes in the fraction of the population contained in a site-specific census. These other sources of measurement variance are usually unknown and unknowable (for all practical purposes).

Equation 1.3 permits many different relationships between the measurements and the true abundances. For this chapter, we consider only two cases. In case 1, there are m subpopulations, and each is measured once. In this case, m equals n, $\mathbf{Z}$ is an $m \times m$ identity matrix, and $\mathbf{D}$ equals 0[‡]. In case 2, there is one subpopulation, $m = 1$, that has been measured at n different sites. In this case, $\mathbf{Z}$ is an $n \times 1$ vector of ones,

[‡] $\mathbf{D}$ cannot be estimated in this case because it is a scaling factor that drops out during estimation.

and $\mathbf{X}_t$ is a scalar (since $m = 1$). In this case, we allow $\mathbf{D}$ to be an $n \times 1$ vector with the first element equal to zero and the other elements estimated. This allows for the possibility that differences in the mean log abundance between sites are due to different biases in the measurement errors at each site.

Equation 1.2 and 1.3 form the multi-site state-space model. The objective is to use this model to estimate the parameters $\{\mathbf{B}, \mathbf{Q}, \mathbf{R}, \mathbf{D}\}$: $\mathbf{B}$ gives the mean population growth rate in each subpopulation, $\mathbf{Q}$ gives the variance in the population growth between time steps, and $\mathbf{R}$ and $\mathbf{D}$ give the measurement-error variance and bias for each site. We assume for this chapter that the process errors and measurement errors are Gaussian and uncorrelated. These assumptions allow us to use estimation methods designed for linear Gaussian state-space models. The assumption of uncorrelated errors can easily be relaxed (see for example Shumway and Stoffer 2000). The assumption of Gaussian errors can also be relaxed (Durbin and Koopman 2000), however parameter estimation would be considerably more involved.

1.3 Specification of the spatial structure among the subpopulations

In its unconstrained form, the multi-site state-space model allows each subpopulation to have its own population growth rate, its own process-error variance, and any level of correlation in the process errors between subpopulations. It allows a similar level of flexibility in the measurement errors. We can incorporate spatial structure by imposing constraints on the $\mathbf{B}$, $\mathbf{Q}$, and $\mathbf{R}$ terms. For example, we might specify that the process-error variances are the same across subpopulations, or that the measurement errors are independent.

We will denote the alternative model structures by the triplet $\{f_B, f_Q, f_R\}$, where f_B denotes the constraint used for $\mathbf{B}$, f_Q denotes the constraint used for $\mathbf{Q}$, and f_R denotes the constraint used for $\mathbf{R}$. In all cases, $f = 1$ will denote the unconstrained form.

1.3.1 Structure of the population growth rates (f_B)

1. $f_B = 1$ *Each subpopulation has an independent and different mean growth rate*. In this case, $\mathbf{B} = \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is an $m \times 1$ vector of subpopulation-specific stochastic growth rates μ_i. This is the unconstrained form for $\mathbf{B}$.
2. $f_B = 2$ *Each subpopulation has the same mean growth rate*. $\mathbf{B} = \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is an $m \times 1$ vector of μ's, all of which are equal.
3. $f_B = 3$ *There is one population ($m = 1$) that has been measured at n different sites*. Consequently, there is only one population growth rate. In this case, $\mathbf{B}$, $\mathbf{Q}$ and $\mathbf{X}_t$ are scalars: $\mathbf{B} = \mu$, $\mathbf{Q} = \sigma_{var}$ and $\mathbf{X}_t = x_t$. This case also affects the structure in the measurement errors: $\mathbf{Z}$ is an $n \times 1$ vector of ones and $\mathbf{D}$ is an $n \times 1$ vector of biases, the first of which is set equal to 0[†]

[†] Effectively we are estimating the biases relative to the bias for the first measurement time series. This should be kept in mind when interpreting the estimated true abundances.

1.3.2 Structure of the process-error variances (f_Q)

1. $f_Q = 1$ *An unconstrained covariance matrix*. Each subpopulation has a different level of process-error variance, and each pair of subpopulations has a different level of covariance between their process errors. In this case, $\mathbf{Q}$ is an $m \times m$ covariance matrix with terms on the diagonal and off-diagonals.
2. $f_Q = 2$ *A diagonal covariance matrix with unequal diagonal entries*. In this case, each subpopulation has a different level of process-error variance, but the process errors between subpopulations are independent. Thus, the off-diagonal terms in the covariance matrix are 0. $\mathbf{Q}$ is an $m \times m$ covariance matrix with terms on the diagonal and zeros on the off-diagonals.
3. $f_Q = 3$ *A diagonal covariance matrix with equal diagonal entries*. Each subpopulation has the same level of process-error variance, but the errors are independent. $\mathbf{Q} = \sigma_{var}\mathbf{I}$, where σ_{var} is the common process-error variance term and $\mathbf{I}$ is an $m \times m$ identity matrix. This gives a covariance matrix with all terms on the diagonal equal to σ_{var} and the off-diagonal terms equal to 0.
4. $f_Q = 4$ *A covariance matrix with equal variances and covariances*. Each subpopulation has the same level of process-error variance, and the covariances between process errors are equal between any two subpopulations. $\mathbf{Q} = \sigma_{var}\mathbf{I} + \sigma_{cov}(\mathbf{U} - \mathbf{I})$, where σ_{var} is the common variance term and σ_{cov} is the common covariance term. $\mathbf{I}$ is an $m \times m$ identity matrix and $\mathbf{U}$ is an $m \times m$ unit matrix. This gives a covariance matrix with all terms on the diagonal equal to σ_{var} and the off-diagonal terms equal to σ_{cov}.

1.3.3 Structure of the measurement errors (f_R)

The constraints on the measurement-error variances are the same as for the process-error variance. The different measurement-error models are denoted, $f_R = 1$, 2, 3 or 4, where the constraints are defined as in section 1.3.2 with references to 'process-error' replaced with 'measurement-error' and with references to $\mathbf{Q}$ replaced with $\mathbf{R}$.

1.4 Estimation of the population parameters using maximum likelihood

Equation 1.2 and 1.3, along with the model constraints specified by $\{f_B, f_Q, f_R\}$, form the constrained model for the multi-site data. Using this model, we can estimate the parameters that describe the population dynamics, $\mathbf{B}$ and $\mathbf{Q}$, and the parameters that describe the measurement error, $\mathbf{R}$ and $\mathbf{D}$. There are two main approaches to parameter estimation: maximum-likelihood estimation and Bayesian estimation. In this chapter, we focus on maximum-likelihood estimation. However, the likelihood functions specified in this chapter are also used in Bayesian estimation. Thus, this chapter provides the building blocks needed for a Bayesian approach as well.

1.4.1 The likelihood function

The first step of maximum-likelihood estimation is to specify the likelihood of the parameters, $\Theta = \{\mathbf{B}, \mathbf{Q}, \mathbf{R}, \mathbf{D}\}$, given the observed data. In our case, the data are n time series of observations for time 1 to T. We denote the n observations at time t as $\mathbf{Y}_t$, and the set of all observations for time t to T as $\mathbf{Y}_1^T \equiv \mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T$.

The observations at time t, $\mathbf{Y}_t$, are dependent on the past observations, $\mathbf{Y}_1^{t-1}$. Thus we cannot write the likelihood simply as $L(\Theta) = \prod_{t=1}^{T} p(\mathbf{Y}_t)$. Instead we write the likelihood as a product of the conditional probabilities[†]:

$$L(\Theta|\mathbf{Y}_1^T) = p(\mathbf{Y}_1) \prod_{t=2}^{T} p(\mathbf{Y}_t|\mathbf{Y}_1^{t-1}), \tag{1.4}$$

where $p(\mathbf{Y}_t|\mathbf{Y}_1^{t-1})$ is the probability density function for $\mathbf{Y}_t$ given all of the observations up to time $t-1$. Note that $\mathbf{Y}$ is not Markov (only $\mathbf{X}$ is), thus we must condition on $\mathbf{Y}_1^{t-1}$ rather than $\mathbf{Y}_{t-1}$. The distribution of $p(\mathbf{Y}_t|\mathbf{Y}_1^{t-1})$ is multivariate normal, and we denote the mean of this distribution as $\tilde{\mathbf{Y}}_{t|t-1}$ and its covariance matrix as $\mathbf{F}_t$. $\tilde{\mathbf{Y}}_{t|t-1}$ is defined as $\mathrm{E}(\mathbf{Y}_t|\mathbf{Y}_1^{t-1})$, the expected value of $\mathbf{Y}_t$ conditioned on $\mathbf{Y}_1^{t-1}$. $\mathbf{F}_t$ is defined as $\mathrm{E}((\mathbf{Y}_t - \tilde{\mathbf{Y}}_{t|t-1})(\mathbf{Y}_t - \tilde{\mathbf{Y}}_{t|t-1})')$, also conditioned on $\mathbf{Y}_1^{t-1}$. The initial conditions are specified by $p(\mathbf{Y}_1)$, and will be treated as either an estimated or a nuisance parameter (see section 1.4.2).

Using the probability density for a multivariate normal, we can write out the likelihood function given in equation 1.4 as

$$L(\Theta|\mathbf{Y}_1^T) = \prod_{t=1}^{T} \frac{\exp\left\{-\frac{1}{2}(\mathbf{Y}_t - \tilde{\mathbf{Y}}_{t|t-1})'\mathbf{F}_t^{-1}(\mathbf{Y}_t - \tilde{\mathbf{Y}}_{t|t-1})\right\}}{\left((2\pi)^n|\mathbf{F}_t|\right)^{1/2}}. \tag{1.5}$$

To calculate the likelihood, we need estimates of $\tilde{\mathbf{Y}}_{t|t-1}$ and $\mathbf{F}_t$. We do not have these directly, but we can solve for them indirectly by rewriting them in terms of $\mathbf{X}_t$ and the deviations in $\mathbf{X}_t$ from the predicted values. First, we define $\mathbf{x}_{t|t-1} \equiv \mathrm{E}(\mathbf{X}_t|\mathbf{Y}_1^{t-1})$. Then, using the measurement equation (equation 1.3), we have:

$$\begin{aligned} \tilde{\mathbf{Y}}_{t|t-1} &= \mathrm{E}(\mathbf{Y}_t|\mathbf{Y}_1^{t-1}) = \mathrm{E}(\mathbf{Z}\mathbf{X}_t + \mathbf{D} + \boldsymbol{\Upsilon}_t|\mathbf{Y}_1^{t-1}) \\ &= \mathbf{Z}\mathbf{x}_{t|t-1} + \mathbf{D}. \end{aligned} \tag{1.6}$$

Next, we define $\mathbf{P}_{t|t-1} \equiv \mathrm{E}((\mathbf{X}_t - \mathbf{x}_{t|t-1})(\mathbf{X}_t - \mathbf{x}_{t|t-1})')$. Then, using the measurement equation (equation 1.3) again, we have:

$$\begin{aligned} \mathbf{F}_t &= \mathrm{E}\left(\left[\mathbf{Y}_t - \tilde{\mathbf{Y}}_{t|t-1}\right]\left[\mathbf{Y}_t - \tilde{\mathbf{Y}}_{t|t-1}\right]'\right) \\ &= \mathrm{E}\left(\left[\mathbf{Z}(\mathbf{X}_t - \mathbf{x}_{t|t-1}) + \boldsymbol{\Upsilon}_t\right]\left[\mathbf{Z}(\mathbf{X}_t - \mathbf{x}_{t|t-1}) + \boldsymbol{\Upsilon}_t\right]'\right) \\ &= \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}' + \mathbf{R}. \end{aligned} \tag{1.7}$$

[†] For more background on the derivation of the likelihood see, for example, Harvey (1989) section 3.4.

Thus using equations 1.6 and 1.7, we can solve for the likelihood if we have estimates of $\mathbf{x}_{t|t-1}$, the expected value of $\mathbf{X}_t$ given the observed data up to time $t-1$, and $\mathbf{P}_{t|t-1}$, the deviations between $\mathbf{X}_t$ and $\mathbf{x}_{t|t-1}$.

1.4.2 Estimation of $\mathbf{x}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ using the Kalman filter

The multi-site state-space model is a linear dynamical system with discrete time and Gaussian errors. This type of problem is extremely important in many engineering fields. In 1960, Rudolf Kalman published an algorithm that solves for the optimal (lowest mean square error) estimate of the hidden $\mathbf{X}_t$ based on the observed data up to time t for this class of linear dynamical system. This algorithm, now known as the Kalman filter, gives an estimate of $\mathrm{E}(\mathbf{X}_t|\mathbf{Y}_1^t)$, which we will denote as $\mathbf{x}_{t|t}$, and the covariance, $\mathrm{E}((\mathbf{X}_t - \mathbf{x}_{t|t})(\mathbf{X}_t - \mathbf{x}_{t|t})')$, which we will denote as $\mathbf{P}_{t|t}$. The Kalman filter also provides the optimal estimates of $\mathbf{X}_t$ conditioned on the data up to time $t-1$, i.e., $\mathbf{x}_{t|t-1}$ and its covariance, $\mathbf{P}_{t|t-1}$. These are the estimates that are needed to calculate $\tilde{\mathbf{Y}}_{t|t-1}$ and $\mathbf{F}_t$ in equations 1.6 and 1.7. These in turn are used to calculate the likelihood. The Kalman filter is widely used in time-series analysis, and there are many textbooks covering it and its applications. The books by Harvey (1989) and Shumway and Stoffer (2000) are particularly useful for ecologists because they are geared towards physical, biological and economics applications.

The Kalman filter is a recursion that consists of a set of prediction equations followed by a set of updating equations. The prediction equations are so named because they predict the states at time t given information up to and including time $t-1$:

$$\mathbf{x}_{t|t-1} = \mathbf{x}_{t-1|t-1} + \mathbf{B} \tag{1.8}$$

$$\mathbf{P}_{t|t-1} = \mathbf{P}_{t-1|t-1} + \mathbf{Q}. \tag{1.9}$$

Using the output from the prediction equations, new estimates conditioned on the data up to time t are calculated using the updating equations:

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{F}_t^{-1}(\mathbf{Y}_t - \mathbf{Z}\mathbf{x}_{t|t-1} - \mathbf{D}) \tag{1.10}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{F}_t^{-1}\mathbf{Z}\mathbf{P}_{t|t-1}. \tag{1.11}$$

This recursive algorithm is started with initial values $\mathbf{x}_{0|0}$ and $\mathbf{P}_{0|0}$, which are the mean and variance the population abundance at time $t = 0$. Using those initial values, one iterates through the prediction and updating equations for $t = 1, 2, 3, \ldots, T$. This provides the time series, $\mathbf{x}_{t|t-1}$ and $\mathbf{F}_t$, that are needed to calculate the likelihood.

Typically, there is no prior information for the abundances at time $t = 0$. One solution is to estimate $\mathbf{x}_{0|0}$ and $\mathbf{P}_{0|0}$ as extra free parameters. Alternatively, the initial conditions can be specified using a diffuse prior distribution for $\mathbf{X}_0$. This is done by setting $\mathbf{P}_{0|0} = \kappa\mathbf{I}$ (where $\mathbf{I}$ is an $m \times m$ identity matrix), substituting this into the Kalman filter equations and allowing κ to grow arbitrarily large. Since $\mathbf{X}_0$ is defined as normal with a mean of $\mathbf{x}_{0|0}$ and variance $\mathbf{P}_{0|0}$, this has the effect of setting a diffuse prior on $\mathbf{X}_0$. When $f_B = 1$ or 2, this diffuse prior leads to $\mathbf{x}_{1|1} = \mathbf{Y}_1$ and $\mathbf{P}_{1|1} = \mathbf{R}$. When $f_B = 3$, $\mathbf{x}_{1|1}$ and $\mathbf{P}_{1|1}$ are scalars, and the diffuse prior leads to

$\mathbf{x}_{1|1} = \left(\mathbf{O}'\mathbf{R}^{-1}\mathbf{Y}_1\right) / \left(\mathbf{O}'\mathbf{R}^{-1}\mathbf{O}\right)$ and $\mathbf{P}_{1|1} = 1/\left(\mathbf{O}'\mathbf{R}^{-1}\mathbf{O}\right)$, where $\mathbf{O}$ indicates an $m \times 1$ vector of ones.

For simplicity, we presented the likelihood calculation as if there were no missing values in the data. However, one of the strengths of state-space approaches is that missing values are easy to accommodate; if some values within $\mathbf{Y}_t$ are missing, those values become a place-holder that will be filled with the optimal estimate for the missing data point. Harvey (1989), section 3.4, shows how the Kalman filter equations are modified when there are missing values.

1.4.3 Maximization of the likelihood function

The Kalman filter provides estimates of $\mathbf{x}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ that together with equations 1.6, 1.7, and 1.5 allow us to calculate the likelihood of the parameters, Θ. Our objective is to find the Θ that maximizes the likelihood. There are a variety of approaches to the maximization problem. One standard approach is a Nelder-Mead algorithm, which is available as a pre-packaged routine for most computing software. However, for the multi-site state-space model, we found that this algorithm did not always converge. Another approach, which we found to always converge, is the estimation-measurement (EM) algorithm presented in Shumway and Stoffer (1982) and Shumway and Stoffer (2000, section 4.3). The EM algorithm involves iteratively estimating the true, hidden, abundances conditioned on all of the data, using that to re-estimate the parameters and then using the updated parameters to re-estimate the true abundances. This is repeated until the likelihood converges[‡]. Another wrinkle that can be added is restricted maximum likelihood (REML). Because of the measurement errors, there is a negative temporal correlation in the data. This negative correlation provides additional information which can be used to improve the estimates (Staples *et al.* 2004, Dennis *et al.* 2006). In section 1.6, we will discuss bootstrap methods for specifying the confidence intervals, standard errors and bias of the maximum-likelihood parameters.

One issue of concern for all of these maximization methods is that when the time series are short (T is small) or contain many missing values, the likelihood surface can become multimodal. The problem in this case is that the likelihood surface has its largest peak with either the $\mathbf{Q}$ or $\mathbf{R}$ diagonal terms are set at zero, and there is a smaller peak at the correct value where all $\mathbf{Q}$ and $\mathbf{R}$ diagonal terms are non-zero. The result is that all of the variance in the data is put into process-error or measurement-error variance. Intuitively, what is happening is that there is not enough information in the data to partition the variance. If this is discovered to be a problem, which will be apparent by either of the $\mathbf{Q}$ or $\mathbf{R}$ diagonal terms going to zero, there are two general solutions. First, the size of the model can be constrained such that it is commiserate with the information in the data. For example, the population structure can be constrained (e.g., by setting $f_B = 3$ or $f_Q = 4$) so that there are fewer

[‡] The EM algorithm is a hill-climbing algorithm. Thus steps must be taken to ensure that it does not get stuck on local maxima (Biernacki *et al.* 2003)

parameters to estimate. The second general approach is add an informative prior on the variance parameters using a Bayesian approach. In this case, the prior will affect the posterior estimates. This is the objective in this case, since the data do not contain enough information in and of themselves to partition the variance. Obviously, the use of an informative prior should be done with caution, but there are situations where researchers have external information on the plausible range of measurement-error or process-error variance.

1.5 Investigation of the population structure using model-selection criteria

In section 1.4, we specified a particular population structure by putting constraints on $\mathbf{B}$, $\mathbf{Q}$, and $\mathbf{R}$. We can also use the multi-site state-space framework to measure the data support for different population structures (Figure 1.1) rather than specifying a structure *a priori*. The different structures are denoted by the triplet $\{f_B, f_Q, f_R\}$ presented in section 1.3. These form a nested set of models varying from unstructured (a single population but measured with multiple time series) to fully structured (different stochastic growth rates and process-error variances in each subpopulation and correlations in the process errors between subpopulations).

Using model-selection criteria (Burnham and Anderson 2002, Johnson and Omland 2004), we can measure the data support for the different models. The basic idea is that different models are fit to the data, the fit of the model to the data is measured using the likelihood function, and the fit is penalized for the number of parameters estimated by the model. The latter corrects for the fact that more complex models will tend to fit data better, simply because there is more flexibility in the model. The function that specifies how the likelihood is penalized for complexity is the model-selection criterion, and it gives a relative measure of data support. There are a variety of different model-selection criteria used in model selection. The most commonly used are Akaike's information criterion (AIC) (Akaike 1973, Burnham and Anderson 2002), Bayesian or Schwarz information criterion (BIC) (Mcquarrie and Tsai 1998), and deviance information criterion (DIC) (Spiegelhalter *et al.* 2002). Mcquarrie and Tsai (1998) is a good reference for model selection approaches specific to time-series data, and Burnham and Anderson (2002) is a good reference for model-selection approaches for the ecological sciences. In the example below, we illustrate the use of AIC for measuring the data support for different structures within a group of chinook salmon subpopulations.

1.6 Analysis of Snake River chinook salmon dynamics and structure

The Snake River is one of the major tributaries of the Columbia River, and historically it produced a large proportion of the chinook salmon within the Columbia basin. However, anthropogenic impacts such as the construction of hydropower dams on the Columbia and Snake Rivers, habitat destruction, and over-fishing led to large declines in the chinook populations within the Snake River and its tributaries. In

1992, the Snake River spring/summer chinook Evolutionary Significant Unit† (ESU) was listed as threatened under the U.S. Endangered Species Act, along with other salmonid ESUs in the Columbia River basin. This Snake River ESU includes all wild (not hatchery-released) chinook salmon that spawn in the spring and summer in the Snake River and its tributaries: the Tucannon, Grande Ronde, Imnaha, and Salmon Rivers (Figure 1.2). Chinook that spawn in the spring and summer spend their first year in freshwater near their natal streams and migrate to the ocean as yearlings.

Figure 1.2 *Map of the Snake River spring/summer chinook ESU. The location of the subpopulations are shown with the gray ovals: 1) Bear Valley/Elk Creek , 2) Sulphur Creek, 3) Marsh Creek, 4) Upper Valley Creek, 5) Big Creek, 6) Lemhi River. To reach these spawning areas, fish must pass through (or be barged around) eleven hydropower dams.*

To illustrate the use of the multi-site state-space model, we analyzed time series from six distinct chinook subpopulations in the upper reaches of the Snake River basin (Figure 1.2). Chinook salmon show strong fidelity to their natal streams, thus the fish spawning within a specific stream are most likely to have been spawned in that stream or nearby. The six subpopulations we analyzed were Bear Valley/Elk Creek, Sulphur Creek, Marsh Creek, Valley Creek, Big Creek, and Lemhi River (Figure 1.2). The time-series data for each subpopulation represent estimates of the spawning salmon abundances within each subpopulation from 1980 to 2001 (Figure 1.3). Our analysis focused on two questions: 1) Given a particular structure for the six subpopulations,

† Evolutionary Significant Unit is the term for a population segment that is considered distinct for the purpose of conservation under the U.S. Endangered Species Act (Waples 1991).

what are the maximum-likelihood estimates of the stochastic growth rates and the true abundances? and 2) What population structures are most supported by the data?

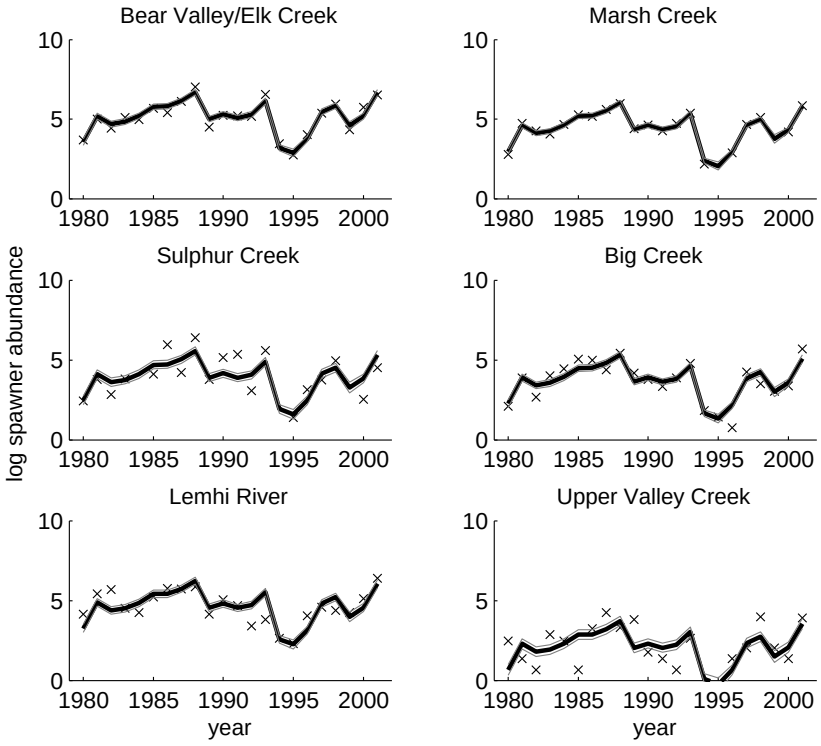

Figure 1.3 *The 22-year time series of spawning abundance and smoothed estimates for 1) Bear Valley/Elk Creek , 2) Sulphur Creek, 3) Marsh Creek, 4) Upper Valley Creek, 5) Big Creek, and 6) Lemhi River. The $\times$'s are the actual spawner counts, the black line is the smoothed estimates of spawner abundance from the Kalman smoother, and the gray lines give the upper and lower 95% CIs for the smoothed estimates (the estimates with the measurment errors removed).*

1.6.1 Estimation of the stochastic growth rates and true abundances

To separate out the measurement errors and provide estimates of the true abundances within each spawning site, we used the Kalman smoother. The Kalman smoother provides the optimal estimates of $\mathbf{X}_t$ given all the data, $\mathbf{Y}_1^T$. The Kalman smoother

starts with the $\mathbf{x}_{T|T}$ estimates from the Kalman filter[‡] and works backwards from T to 1 using the following updating equations:

$$\mathbf{x}_{t-1|T} = \mathbf{x}_{t-1|t-1} + \mathbf{J}_{t-1}\left(\mathbf{x}_{t|T} - \mathbf{x}_{t|t-1}\right) \quad (1.12)$$

$$\mathbf{P}_{t-1|T} = \mathbf{P}_{t-1|t-1} + \mathbf{J}_{t-1}\left(\mathbf{P}_{t|T} - \mathbf{P}_{t|t-1}\right)\mathbf{J}_{t-1}^{'}, \quad (1.13)$$

where $\mathbf{J}_{t-1} \equiv \mathbf{P}_{t-1|t-1}\mathbf{P}_{t|t-1}^{-1}$. At the end of the recursion, we have the smoothed estimates of $\mathbf{X}_t$ conditioned on all the data. The smoothed estimates are denoted $\mathbf{x}_{t|T}$.

The smoothed estimates have a simple relationship to the maximum-likelihood estimates for the stochastic growth rates:

$$\mathbf{B} = \frac{1}{T-1}\left(\mathbf{x}_{T|T} - \mathbf{x}_{1|T}\right), \quad \text{if } f_B = 1 \text{ or } 3, \quad (1.14)$$

$$\mathbf{B} = \frac{1}{T-1}\frac{\mathbf{O}'\mathbf{Q}^{-1}\left(\mathbf{x}_{T|T} - \mathbf{x}_{1|T}\right)}{\mathbf{O}'\mathbf{Q}^{-1}\mathbf{O}}, \quad \text{if } f_B = 2, \quad (1.15)$$

where $\mathbf{O}$ is an $m \times 1$ matrix of ones.

1.6.2 Investigation of the subpopulation structure using AIC and AICb

To determine which population structure was best supported by the data, we used Akaike's Information Criteria (AIC) (Akaike 1973, Burnham and Anderson 2002) and the bootstrap AICb (Cavanaugh and Shumway 1997). In particular, we were interested in whether the six sites should be treated as one population sampled with six independent time series or as six separate subpopulations with correlated process errors. AIC and AICb measure the data support for models with different population structure. Models with lower AIC and AICb scores have better data support relative to models with higher AIC or AICb scores.

For model q, where q specifies a unique $\{f_B, f_Q, f_R\}$ triplet, the AIC is defined as:

$$\mathrm{AIC}_q = -2\log L(\hat{\Theta}_q) + 2p, \quad (1.16)$$

where $\hat{\Theta}_q$ is the parameter set $\{\mathbf{B}, \mathbf{Q}, \mathbf{R}, \mathbf{D}\}$ that maximizes the likelihood of the observed data $\mathbf{Y}_1^T$ given model q, and p is the number of effective parameters in model q. We calculated the maximum-likelihood estimates $\hat{\Theta}_q$ using the EM algorithm (Shumway and Stoffer 1982, 2000).

AICb is a variant of AIC that corrects for AIC's bias towards overly complex models when the sample size is small. AICb has the same objective as the more familiar AICc—the small sample-size corrected AIC (Burnham and Anderson 2002)—but

[‡] The Kalman filter is a forward recursion and provides an optimal prediction of $\mathbf{X}_t$ given the past, $\mathbf{Y}_1^{t-1}$. The Kalman smoother is a backwards recursion that provides optimal estimates of the past given the future, in this case $\mathbf{X}_t$ given $\mathbf{Y}_1^T$.

AICb is designed for state-space models. AICb for model q is defined as:

$$\mathrm{AICb}_q = -2\log L(\hat{\Theta}_q) + 2\left\{\frac{1}{N}\sum_{b=1}^{N} -2\log\frac{L(\hat{\Theta}_q(b))}{L(\hat{\Theta}_q)}\right\}, \qquad (1.17)$$

where $\hat{\Theta}_q(b)$, $b = 1, \ldots, N$, represents a set of N bootstrap replicates of $\hat{\Theta}_q$. The bootstrap replicates are generated using the following procedure (Stoffer and Wall 1991; Shumway and Stoffer 2000, section 2.6). Using the parameters $\hat{\Theta}_q$, the model q is fit to the data. This provides a time series of the innovations $\boldsymbol{\epsilon}$ for $t = 1$ to T, where $\boldsymbol{\epsilon}_t \equiv \mathbf{Y}_t - \tilde{\mathbf{Y}}_{t|t-1}$. The bootstrap replicates of the innovations time series, $\boldsymbol{\epsilon}(b)$, are generated by taking T samples with replacement from $\boldsymbol{\epsilon}$. The bootstrap-generated $\boldsymbol{\epsilon}(b)$ are then used in what is termed the innovations form of the state-space model to generate a bootstrapped $\mathbf{Y}(b)$ time series. This process is repeated N times to produce N bootstrapped $\mathbf{Y}(b)$ time series. For each of these $\mathbf{Y}(b)$ time series, the parameters that maximize the likelihood of the bootstrapped data $\mathbf{Y}(b)$ are found. This produces the N bootstrapped estimates: $\hat{\Theta}_q(b)$, $b = 1, \ldots, N$.

1.6.3 Confidence intervals and diagnostics

To determine the accuracy of the parameter estimates, we used a parametric bootstrap approach (Shumway and Stoffer 2000, section 2.6). The N bootstrap replicates of $\hat{\Theta}_q(b)$ were used to estimate confidence intervals, standard errors and bias for each of the estimated model parameters. To construct 95% confidence intervals for the k-th parameter within $\hat{\Theta}_q$, the 2.5 and 97.5 percentiles for the k-th parameter in the bootstrap replicates $\hat{\Theta}_q(b)$ were used as the lower and upper confidence limits.

The bootstrap standard error for the k-th parameter was defined as the mean squared difference between the k-th parameter in the bootstrapped samples and the mean value of the k-th parameter in the bootstrapped samples:

$$\mathrm{SE} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{k}_q(b) - \frac{1}{N}\sum_{b=1}^{N}\hat{k}_q(b)\right)^2, \qquad (1.18)$$

where $\hat{k}_q(b)$ denotes the k-th parameter in $\hat{\Theta}_q(b)$. The bootstrap bias was calculated as:

$$\mathrm{bias} = \frac{1}{N}\sum_{b=1}^{N}\hat{k}_q(b) - \hat{k}_q, \qquad (1.19)$$

where $\hat{k}_q$ is the maximum-likelihood estimate of the k-th parameter in $\hat{\Theta}_q$. As a rule of thumb, bias is considered a potential problem when it exceeds 5% of the SE.

Diagnostics were applied to the fitted state-space models to determine whether they were appropriate for analyzing the salmon data. When running model diagnostics, prediction errors in state-space models play a similar role to that of residuals in an ordinary least-squares regression. Like residuals, the prediction errors are assumed to be independent and normally distributed. The normality assumption was examined

using quantile-quantile (QQ) plots (Chambers *et al.* 1983) of the standardized prediction errors. Jarque-Bera tests for normality were also applied to the standardized prediction errors (Cromwell *et al.* 1994). The assumption of serially independent prediction errors was examined using autocorrelation-function plots and the Box-Pierce test for independence (Box and Pierce 1970).

1.6.4 Results

The estimates of the mean stochastic growth rate were positive across all the top models. This indicates a population that is increasing. The stochastic growth rate estimate for the best model was 0.14, suggesting a robust mean growth rate of 14% per year. However, the estimated growth rates had large standard errors and correspondingly wide confidence intervals that included negative values (Table 1.1). This indicates that even with six 22-year time series, the data are insufficient for confidently estimating whether the population is increasing or decreasing.

Results on the population structure however are more informative. Out of the 36 state-space models considered, the model with $f_B = 3$, $f_Q = 1^{\ddagger}$,and $f_R = 2$ had the lowest AICb and AIC values (Table 1.1). This is the model with a single population that is measured with six different time series, each with independent and different measurement errors. This model had considerably more support (ΔAIC > 10) than the next competitor. The model equivalent to six independent salmon subpopulations each measured independently, $\{f_B, f_Q, f_R\} = \{1, 2, 2\}$, fit extremely poorly. This model ranked 30 out of the 36 models with a ΔAICb score of 170.7 compared to the best model (Table 1.1).

The result that the model with a single population fit the data best indicates that the six subpopulations were highly correlated. Surporting this, we also found that the model $\{f_B, f_Q, f_R\} = \{1, 1, 2\}$, which has an unrestricted process-error covariance matrix, also indicated that the process-error correlations were high. The correlation coefficients for this model ranged from 0.86 to 1.0 (Table 1.2). In contrast, the measurement errors were found to be uncorrelated. The best-fitting model had a diagonal measurement-error covariance matrix, $\mathbf{R}$, with unequal variances (f_R = 2) and zero correlation between all subpopulations. The measurement-error variances differed greatly between the six subpopulations (Table 1.3). Marsh Creek had an estimated measurement-error variance of 0.03 (SE = 0.03) compared to Upper Valley Creek, which had an estimated measurement-error variance of 1.14 (SE = 0.36). Changing from a diagonal $\mathbf{R}$ matrix with unequal variances to a diagonal matrix with equal variances produced the second best model with a ΔAICb of 14.5 above the best model. Bias in the measurement-error estimates ranged from 4% to 63% of SE.

The methods are designed to look for correlation and synchrony across sites. In the case of these salmon time series, we see both strong correlation and strong synchrony.

‡ Note that when $f_B = 3$, $m = 1$ so $\mathbf{Q}$ is a scalar.

Table 1.1 *Results of the model-selection analyses which fit models with different population structures to the salmon data. Models are by ranked by ΔAICb scores. Lower ΔAICb indicates more data support for that model. Generally, a ΔAICb > 10 indicates low data support. p indicates the number of parameters in each model.*

	Model form			Mean	Lower 95%	Upper 95%			
Rank	f_B	f_Q	f_R	μ	limit	limit	p	ΔAIC	ΔAICb
1	3	1	2	0.14	-0.32	0.59	14	0	0
2	3	1	3	0.11	-0.27	0.54	9	19	14.5
3	3	1	4	0.09	-0.23	0.44	10	20.2	18.9
4	2	4	2	0.14	-0.37	0.60	15	22.9	31.6
5	1	4	2	0.13	-0.34	0.61	20	28.5	36
6	2	4	3	0.11	-0.27	0.52	10	44.4	46.4
7	2	2	4	0.00	-0.07	0.08	15	49.2	50.5
8	2	3	4	-0.01	-0.08	0.07	10	47	51.5
9	1	4	3	0.11	-0.31	0.51	15	52.5	53.2
10	2	4	4	0.08	-0.20	0.34	11	45.5	57.5
11	1	3	4	-0.01	-0.08	0.06	15	55.3	58.8
12	1	2	4	-0.02	-0.13	0.09	20	57.8	73
13	1	4	4	0.08	-0.18	0.35	16	53.6	73.2
14	3	1	1	0.10	-0.19	0.40	29	8.3	88.8
15	2	3	3	0.06	-0.05	0.16	9	109.1	104.3
16	2	3	2	0.07	-0.04	0.17	14	113.1	110.1
17	2	2	3	0.06	-0.07	0.17	14	117.9	118
18	2	2	2	0.05	-0.06	0.16	19	119.9	121.8
19	2	3	1	-0.01	-0.07	0.06	29	41.7	122.4
20	2	4	1	0.10	-0.19	0.39	30	28.3	123.7
21	2	1	2	0.14	-0.20	0.48	34	44.5	124.2
22	2	2	1	0.01	-0.06	0.08	34	37.1	125.5
23	1	3	1	-0.01	-0.08	0.05	34	45.6	132.8
24	1	3	3	0.06	-0.05	0.15	14	119	135.2
25	1	3	2	0.07	-0.05	0.18	19	123	144.2
26	1	1	2	0.12	-0.29	0.53	39	52.6	145.2
27	1	4	1	0.11	-0.18	0.38	35	32.6	150.2
28	2	1	3	0.17	-0.11	0.47	29	53.1	157.1
29	1	2	3	0.06	-0.04	0.17	19	127.8	161.5
30	1	2	2	0.06	-0.05	0.16	24	129.4	170.7
31	1	2	1	0.00	-0.07	0.07	39	42.4	172.3
32	1	1	3	0.12	-0.32	0.57	34	61.7	180.2
33	2	1	4	0.17	-0.12	0.47	30	54.9	213.2
34	1	1	4	0.12	-0.33	0.54	35	63.6	231.5
35	2	1	1	0.15	-0.10	0.40	49	52.3	274.9
36	1	1	1	0.09	-0.15	0.33	54	61.1	294.7

Table 1.2 *Estimated correlation coefficients for the process errors. Model* $\{f_B, f_Q, f_R\} = \{1, 1, 2\}$ *was used to estimate the unconstrained covariance matrix* $\mathbf{Q}$ *which was then used to calculate the correlation matrix.*

	Bear Valley/ Elk Cr.	Marsh Cr.	Sulphur Cr.	Big Cr.	Lemhi R.
Marsh Cr.	1.00				
Sulphur Cr.	0.99	0.99			
Big Cr.	1.00	1.00	0.99		
Lemhi R.	0.86	0.87	0.79	0.85	
Up. Valley Cr.	0.98	0.99	0.95	0.98	0.93

Table 1.3 *Measurement-error variances using the best-fitting model with lowest AICb.*

	Estimate	SE	Lower 95% CI	Upper 95% CI	Bias
Bear Valley/Elk Cr.	0.13	0.05	0.04	0.23	-0.20
Marsh Cr.	0.03	0.03	0.00	0.10	0.14
Sulphur Cr.	0.66	0.22	0.26	1.11	-0.17
Big Cr.	0.28	0.09	0.11	0.47	-0.22
Lemhi R.	0.53	0.16	0.24	0.87	-0.21
Up. Valley Cr.	1.14	0.36	0.49	1.83	-0.20

From the time-series data alone, we cannot infer what mechanism is driving this pattern in the salmon data. These six salmon stocks are exposed to a similar ocean environment and river-migration environment, and this would lead to correlated process errors. However, the process errors would need to be perfectly correlated in order to produce synchrony because without perfect correlation, the time series across the six sites would eventually diverge. This suggests that there is another mechanism that is causing synchrony. Dispersal, in this case straying of spawners to non-natal streams, is known to occur and is a possible mechanism for the synchrony.

Diagnostics were run on the model with the lowest AICb score. The QQ plots indicated no deviation from normality in the prediction errors except for the Marsh Creek subpopulation. For Marsh Creek, the normal QQ plot showed large deviations from a straight line in the tails of the prediction errors, and the Jarque-Bera test indicated that the distribution was not normal (p-value = 0.003). This deviation from normality, however, was driven by a single prediction error (from the year 1994), which was 2.74 standard deviations below zero. When this prediction error was deleted, the Jarque-Bera test indicated no significant deviation from normality (p-value = 0.59). The Box-Pierce tests, based on lags up to five years, indicated that the prediction errors were serially uncorrelated. The autocorrelation functions, however, did indicate

a relatively large negative lag-1 autocorrelation for Sulphur Creek (r = -0.53) and a relatively large positive lag-5 autocorrelation (r = 0.45) for the Bear Valley/Elk Creek subpopulation.

1.7 Discussion

The analysis of the salmon data suggests that the population dynamics within the upper Snake River basin are highly synchronized. The best fitting models indicated very high correlations in the year-to-year fluctuations in subpopulation growth rates and a common stochastic growth rate for all of the six subpopulations. This implies that these subpopulations tend to act as a single population. Biologically, this is not surprising; a certain amount of straying of spawners into non-natal streams is known to occur and in addition, the salmon from the different spawning sites are exposed to a similar environment after they leave their spawning stream. They migrate down the same river corridor to the ocean and then spend two to four years in the ocean. In contrast, the modeling suggests that measurement errors are uncorrelated among the six subpopulations, with variances that differ. This is not surprising given that site differences can greatly affect the accuracy of spawning-abundance counts and given that counts at different subpopulations are made on different days.

Aside from revealing these important patterns in the data, does the multivariate technique improve the accuracy of the stochastic growth rate estimates—relative to simply fitting a univariate model to each subpopulation time series independently, then taking the average? At first glance, the answer appears to be no. The univariate stochastic growth rates can be obtained by using the model $\{f_B, f_Q, f_R\} = \{1, 2, 2\}$. This is the model that specifies an independent stochastic growth rate and variance for each subpopulation, and treats the data as if there is no correlation between subpopulations or measurements. This model gives an SE of 0.052 for the average stochastic growth rate, while the best multivariate model, $\{f_B, f_Q, f_R\} = \{3, 1, 2\}$, gives a much larger SE of 0.23. Shouldn't we expect the model with the lowest AICb to produce lower standard errors? The answer is no, because standard error estimates of models with poor AICb are unreliable. The standard errors are largely a function of the estimated variance matrices (Harvey 1989). Therefore, poor estimates of the variance matrices mean poor standard error estimates and poor confidence intervals. The model $\{f_B, f_Q, f_R\} = \{1, 2, 2\}$, which gives low standard errors, has one of the worst ΔAICb scores (170.7 in Table 1.1), and therefore inferences on precision are not as reliable as those from the top model.

Bias is another part of accuracy that must be considered. The stochastic growth rate estimates were not biased, but variance estimates were. For example, the model with smallest AICb, had a process-error variance that was biased downward by 20% SE. It also had biases in the measurement-error variance that ranged from 14% to 22%. Lindley (2003) found that when time series are short, the Kalman filter (used in this chapter) tends to lead to underestimates of the true process error. This suggests that some bias correction procedure ought to be investigated for the variance estimates.

Another possibility is using restricted maximum likelihood, which was found to generate unbiased estimates of process- and measurement-error variance in a univariate setting (Staples *et al.* 2004). Currently, however, this method does not handle multivariate data or missing values, it sometimes fails to converge, and it can generate negative estimates of measurement-error variance. The slope method (Holmes 2001), can also reduce process-error bias, but it also does not handle multivariate data and may generate negative variance estimates.

Multivariate state-space modeling has a long, rich history in the engineering and economics literature and has proved a powerful tool for modeling and forecasting dynamical systems. This approach allows analysts to deal with data from multiple sites simultaneously, handle missing values, and impose different assumptions concerning the spatial structure within the population dynamics and within the measurement process. Although we have assumed a linear model with Gaussian and uncorrelated errors, these assumptions can be relaxed and the same framework could be used but with the parameters estimated via alternate estimation algorithms. In summary, the multivariate state-space approach provides a formal framework for incorporating spatial structure into the analysis of multi-site time series data and can reveal important relationships among subpopulations—relationships that would remain concealed with a single-site or non-spatial approach.

1.8 References

H. Akaike, "Information theory and an extension of the maximum likelihood principle", B.N. Petrov and F. Csaki, editors, Second international symposium on information theory, Akademiai Kiado, Budapest, Hungary, pp. 267–281, 1973.

C. Biernacki, G. Celeux, and G. Govaert, "Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models", Comp. Stat. Data Anal., Vol. 41, pp. 561–575, 2003.

G. E. P. Box, and D. A. Pierce, "Distribution of residual correlations in autoregressive-integrated moving average time series models", J Am. Stat. Assoc., Vol. 65, pp. 1509–1526, 1970.

K. P. Burnham, and D. R. Anderson, "Model selection and multi-model inference", New York, NY, Springer Publishing, 2002.

J. E. Cavanaugh, and R. H. Shumway, "A bootstrap variant of AIC for state-space model selection", Stat. Sinica, Vol. 7, pp. 473–496, 1997.

J. Chambers, W. Cleveland, B. Kleiner, and P. Tukey, "Graphical methods of data analysis", The Wadsworth statistics/probability series, Belmont, Calif, Wadsworth International Group, 1983.

J. B. Cromwell, W. C. Labys, and M. Terraza, "Univariate tests for time series models", Sage Publications, Inc., Thousand Oaks, CA, 1994.

B. Dennis, P. L. Munholland, and J. M. Scott, "Estimation of growth and extinction parameters for endangered species", Ecol. Monogr., Vol. 61, pp. 115–143, 1991.

B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples, "Estimating density dependence, process noise, and observation error", Ecol. Monogr., Vol. 76, pp. 323-341, 2006.

P. deValpine and A. Hastings, "Fitting population models incorporating process noise and observation error", Ecol. Monog., Vol. 72, pp. 57–76, 2002.

J. B. Dunning, Jr., D. J. Stewart, B. J. Danielson, B. R. Noon, T. L. Root, R. H. Lamberson, and E. E. Stevens, "Spatially explicit population models: current forms and future uses", Ecol. Appl., Vol. 5, pp. 3–11, 1995.

J. Durbin and S. J. Koopman, "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives", J. R. Statist. Soc. B, Vol. 62, pp. 3–56, 2000.

A. C. Harvey, "Forecasting, structural time series models and the Kalman filter", Cambridge University Press, Cambridge, UK, 1989.

E. E. Holmes, "Estimation risks in declining populations with poor data", P Natl. Acad. Sci. USA, Vol. 98, pp. 5072–5077, 2001.

E. E. Holmes, "Beyond theory to application and evaluation: diffusion approximations for population viability analysis", Ecol. Appl., Vol. 14, pp. 1272–1293, 2004.

E. E. Holmes, and W. F. Fagan, "Validating population viability analysis for corrupted data sets", Ecology, Vol. 83, pp. 2379–2386, 2002.

E. E. Holmes, J. L. Sabo, S. V. Viscido, and W. Fagan, "A statistical approach to quasi-extinction forecasting", Ecol. Lett., Vol. 10, pp. 1182–1198, 2007.

E. E. Holmes, and B. Semmens, "Population viability analysis for metapopulations: a diffusion approximation approach", pp. 565–598 in Ecology, Genetics, and Evolution of Metapopulations, editors Illka Hanski and Oscar E. Gaggiotti, Elsevier Press, 2004.

A. R. Ives, B. Dennis, K. L. Cottingham, and S. R. Carpenter, "Estimating community stability and ecological interactions from time-series data", Ecol. Monogr., Vol. 73, pp. 301–330, 2003.

J. B. Johnson, and K. S. Omland, "Model selection in ecology and evolution", Trends Ecol. Evol., Vol. 19, pp. 101–108, 2004.

R. E. Kalman, "A new approach to linear filtering and prediction problems", Transactions of the ASME Journal of Basic Engineering, D, Vol. 82, pp. 35-45, 1960.

W. S. Lahaye, R. J. Gutierrez, and H. R. Akcakaya, "Spotted owl metapopulation dynamics in southern California", J Anim. Ecol., Vol. 63, pp. 775–785, 1994.

S. T. Lindley, "Estimation of population growth and extinction parameters from noisy data", Ecol. Appl., Vol. 13, pp. 806–813, 2003.

A. D. R. Mcquarrie, and C.-L. Tsai, "Regression and time series model selection", Singapore: World Scientific Publishing Company, 1998.

N. H. Schumaker, T. Ernst, D. White, J. Baker, and P. Haggerty, "Projecting wildlife responses to alternative future landscapes in Oregon's Willamette basin", Ecol. Appl., Vol. 14, pp. 381–400, 2004.

R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm", J Time Ser. Anal., Vol. 3, pp. 253–264, 1982.

R. H. Shumway, and D. S. Stoffer, "Time series analysis and its applications", Springer-Verlag, New York, New York, USA, 2000.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit," J Roy. Stat. Soc. B, Vol. 64, pp. 583–639, 2002.

D. F. Staples, M. L. Taper, and B. Dennis, "Estimating population trend and process variation for PVA in the presence of sampling error", Ecology, Vol. 85, pp. 923–929, 2004.

D. S. Stoffer, and K. D. Wall, "Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter", J Am. Stat. Assoc., Vol. 86, pp. 1024–1033, 1991.

S. D. Tuljapurkar, and S.H. Orzack, "Population dynamics in variable environments. I. Long-run growth rates and extinction", Theor. Popul. Biol., Vol. 18, pp. 314–342, 1980.

R. S. Waples, "Pacific salmon, *Oncorhynchus* spp., and the definition of species under the Endangered Species Act", Mar. Fish. Rev., Vol. 53, pp. 11–22, 1991.