Introduction to multivariate state-space models
using multi-site population data as the example

Eli Holmes

FISH 507 – Applied Time Series Analysis

19 January 2017
Topics

Lecture

- Short example of multivariate observations
- Examples of multivariate structure in population data
- How to express these structures mathematically
- Adding a multivariate observation process

Computer Labs

- Analysis of population structure using multi-site data
- Combining diverse data sources to estimate an underlying model
Other examples

Combine multiple station data into a single metric

Multiple individuals measured over time
Imagine we have 3 sampling locations for a population:

- Location 1: mark-recapture
- Location 2: mark-recapture
- Location 3: line transect
Mathematically we can write

\[ x_t = x_{t-1} + u + w_t, \quad w_t \sim N(0, q) \]

\[ y_{1,t} = x_t + v_{1,t}, \quad v_{1,t} \sim N(a_1, r_1) \]

\[ y_{2,t} = x_t + v_{2,t}, \quad v_{2,t} \sim N(a_2, r_2) \]

\[ y_{3,t} = x_t + v_{3,t}, \quad v_{3,t} \sim N(a_3, r_3) \]
The observation part can be rewritten as:

\[
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
\end{bmatrix}_t = \begin{bmatrix}
1 \\
1 \\
1 \\
\end{bmatrix} x_t + \begin{bmatrix}
a_1 \\
a_2 \\
a_3 \\
\end{bmatrix} + \begin{bmatrix}
v_1 \\
v_2 \\
v_3 \\
\end{bmatrix}_t
\]

We need to fix one of the a’s. Traditionally we fix to the first to 0.
The model with one a fixed to zero

\[
\begin{bmatrix}
    y_1 \\
    y_2 \\
    y_3 \\
\end{bmatrix}_t =
\begin{bmatrix}
    1 \\
    1 \\
    1 \\
\end{bmatrix} x_t +
\begin{bmatrix}
    0 \\
    a_2 \\
    a_3 \\
\end{bmatrix} +
\begin{bmatrix}
    v_1 \\
    v_2 \\
    v_3 \\
\end{bmatrix}_t
\]

- observations
- Z matrix
- population size
- bias
- noise
The observation errors are multivariate. For now, let’s assume Normality.

The variance-covariance matrix tells you how the observation errors are related. Are they independent? Or do they covary? Do they have the same variance or different variances?

\[
\begin{bmatrix}
\nu_1 \\
\nu_2 \\
\nu_3 \\
\end{bmatrix}_t 
\sim MVN
\begin{pmatrix}
0, \\
\begin{bmatrix}
\eta_1^2 & \eta_{1,2} & \eta_{1,3} \\
\eta_{1,2} & \eta_2^2 & \eta_{3,2} \\
\eta_{1,3} & \eta_{3,2} & \eta_3^2 \\
\end{bmatrix}
\end{pmatrix}
\]
Example observation error var-cov matrices

<table>
<thead>
<tr>
<th></th>
<th>unconstrained</th>
<th>“equal varcov”</th>
</tr>
</thead>
</table>
|               | \[
| \begin{bmatrix}
| \eta_1^2 & \eta_{1,2} & \eta_{1,3} \\
| \eta_{1,2} & \eta_2^2 & \eta_{3,2} \\
| \eta_{1,3} & \eta_{3,2} & \eta_3^2 \\
| \end{bmatrix}
| \]
|               | \[
| \begin{bmatrix}
| \eta^2 & \alpha & \alpha \\
| \alpha & \eta^2 & \alpha \\
| \alpha & \alpha & \eta^2 \\
| \end{bmatrix}
| \]
|               | unique variances and  | identical variances   |
|               | uncorrelated errors   | and uncorrelated      |
|               | \[
| \begin{bmatrix}
| \eta_1^2 & 0 & 0 \\
| 0 & \eta_2^2 & 0 \\
| 0 & 0 & \eta_3^2 \\
| \end{bmatrix}
| \]
|               | \[
| \begin{bmatrix}
| \eta^2 & 0 & 0 \\
| 0 & \eta^2 & 0 \\
| 0 & 0 & \eta^2 \\
| \end{bmatrix}
| \]
Example of errors coming from these variance-covariance matrices

error = how much the pop. growth rate is above or below average

mvrnorm(10,rep(0,3),diag(.1,3))
Fitting MARSS models using the MARSS R Package

- Fits MARSS models
- *Model specification is 1-to-1 with the equation for the model*
- General, fits any MARSS model with Gaussian errors.
- **BUT**
- Maximum likelihood
- Slow for large data sets. Students working with large data sets have gotten huge speed improvements by coding their models in TMB (or ADMB).
MARSS R Package Model Syntax

\[ X(t) = B \times X(t-1) + U + w(t), \quad w(t) \sim N(0, Q) \]
\[ Y(t) = Z \times X(t) + A + v(t), \quad v(t) \sim N(0,R) \]

- \texttt{fit2=MARSS(y,model=mod.list)}

- \texttt{y} is data; \texttt{model} tells MARSS what the parameters are
- The parameters are MATRICES
- You write matrices just like they appear in your model on paper. Matrices must be MATRICES (not scalars, not vectors)
- You pass \texttt{model} to MARSS as a list
\[ X(t) = B \, X(t-1) + U + w(t), \, w(t) \sim \mathcal{N}(0, Q) \]
\[ Y(t) = Z \, X(t) + A + v(t), \, v(t) \sim \mathcal{N}(0, R) \]

Let's say we want to fit this model:

\[
\begin{align*}
  x_t &= x_{t-1} + u + w_t, \, w_t \sim \mathcal{N}(0, 0.1) \\
  y_t &= x_t + v_t, \, v_t \sim \mathcal{N}(0, r) \\
  x_0 &= 0
\end{align*}
\]

Write in matrix form:

\[
\begin{align*}
  [x]_t &= [1][x]_{t-1} + [u] + [w]_t, \, [w]_t \sim \mathcal{MVN}(0, [0.1]) \\
  [y]_t &= [1][x]_t + [v]_t, \, [v]_t \sim \mathcal{MVN}(0, [r]) \\
  x_0 &= [0]
\end{align*}
\]
\[
X(t) = B \cdot X(t-1) + U + w(t), \quad w(t) \sim N(0, Q)
\]
\[
Y(t) = Z \cdot X(t) + A + v(t), \quad v(t) \sim N(0, R)
\]

```
mod.list=list(
  U=matrix("u"),
  x0=matrix(0),
  B=matrix(1),
  Q=matrix(0.1),
  Z=matrix(1,2,1),
  A= matrix(list(0,"a2"),2,1),
  R= matrix(list("r",0,0,"r"),
  tinitx=0)
```

Let's say we want to fit a model where two sites are sampling temperature x in a lake:

**Our temperature model:**

\[
[x]_t = [1][x]_{t-1} + [u] + [w]_t, \quad [w]_t \sim MVN(0, [0.1])
\]

\[
[x]_0 = [0]
\]

**Our two temperature sensors:**

\[
\begin{bmatrix}
y_1 \\
y_2
\end{bmatrix}_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x \\
0
\end{bmatrix}_t + \begin{bmatrix} 0 \\
a_2
\end{bmatrix} + \begin{bmatrix} v_1 \\
v_2
\end{bmatrix}_t
\]

\[
\begin{bmatrix} v_1 \\
v_2
\end{bmatrix}_t \sim MVN\left(\begin{bmatrix} 0 \\
0
\end{bmatrix}, \begin{bmatrix} r & 0 \\
0 & r
\end{bmatrix}\right)
\]
Some short examples

• MARSS_example_1.R
• MARSS_example_2.R
• MARSS_example_3.R
Multi-site data (Pacific harbor seals)
An example: modeling the population dynamics of harbor seals in Puget Sound, WA
Let’s hypothesize (and model) that the population has 3 subpopulations.

Three subpopulations that are independent but have correlated population dynamics (dispersal, similar environment, etc.)
A multivariate model for the population (not the observations but the actual population)

Multivariate stochastic exponential growth

\[
\begin{bmatrix}
    x_{JF,t} \\
    x_{N,t} \\
    x_{S,t}
\end{bmatrix}
= \begin{bmatrix}
    x_{JF,t-1} \\
    x_{N,t-1} \\
    x_{S,t-1}
\end{bmatrix}
+ \begin{bmatrix}
    u_{JF} \\
    u_{N} \\
    u_{S}
\end{bmatrix}
+ \begin{bmatrix}
    w_{JF,t} \\
    w_{N,t} \\
    w_{S,t}
\end{bmatrix}
\]

3 different x’s, one for each subpopulation

3 mean population growth rate terms

3 different process errors

\[e \sim MVN(0,Q)\]
The population model in matrix form

Exponential population growth with drift (tendency to increase or decline)

\[ x_t = x_{t-1} + u + w_t \]

\[ w_t \sim \text{MVN}(0, Q) \]

Each parameter has “structure”. Different structures imply different population structure.
The mean population growth rates ($u$) can have spatial structure

\[
\begin{bmatrix}
  u_{JF} \\
  u_N \\
  u_S
\end{bmatrix}
\]

unconstrained (all different)

\[
\begin{bmatrix}
  u \\
  u \\
  u
\end{bmatrix}
\]

all the same

\[
\begin{bmatrix}
  u_{JF} \\
  u_{N&S} \\
  u_{N&S}
\end{bmatrix}
\]

Strait of Juan de Fuca different
North and South same
The process error var-cov matrix can have structure: $w_t \sim \text{MVN}(0, Q)$

<table>
<thead>
<tr>
<th>Unconstrained</th>
<th>Diagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variances all different and year-to-year population changes covary</td>
<td>Unique variances and year-to-year population changes are uncorrelated</td>
</tr>
</tbody>
</table>

Unconstrained matrix:

$$
\begin{bmatrix}
\sigma_{JF}^2 & \sigma_{JF,N} & \sigma_{JF,S} \\
\sigma_{JF,N} & \sigma_N^2 & \sigma_{N,S} \\
\sigma_{JF,S} & \sigma_{N,S} & \sigma_S^2
\end{bmatrix}
$$

Diagonal matrix for JF:

$$
\begin{bmatrix}
\sigma_{JF}^2 & 0 & 0 \\
0 & \sigma_N^2 & 0 \\
0 & 0 & \sigma_S^2
\end{bmatrix}
$$

Diagonal matrix for N & S sharing the same variance:

$$
\begin{bmatrix}
\sigma^2 & \alpha & \alpha \\
\alpha & \sigma^2 & \alpha \\
\alpha & \alpha & \sigma^2
\end{bmatrix}
$$

Same variances and year-to-year population changes are uncorrelated.

JF has unique variance; N & S share the same variance; yr-to-yr changes have equal covariance.
\[ x_t = x_{t-1} + u + w_t \]

The structure of the \( U \) and \( Q \) specify our hypotheses (or assumptions) about how the environment is shared.
We observe our subpopulations and those observations have error.
The obs. err. model specifies how the observed time series are related to the true subpopulation sizes.

The observation model

\[
\begin{bmatrix}
\log_{\text{counts}}
y_{1,t} \\
y_{2,t} \\
y_{3,t} \\
y_{4,t} \\
y_{5,t}
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x_{JF,t} \\
x_{N,t} \\
x_{S,t}
\end{bmatrix}
+ \begin{bmatrix}
a_1 \\
a_2 \\
a_3 \\
a_4 \\
a_5
\end{bmatrix}
+ \begin{bmatrix}
\eta_{1,t} \\
\eta_{2,t} \\
\eta_{3,t} \\
\eta_{4,t} \\
\eta_{5,t}
\end{bmatrix}
\]

- \text{observations}
- \text{Z matrix}
- \text{relates each observation time series to a different state process}
- \text{true population “hidden”}
- \text{measurement errors}
- \text{observation biases}
- \text{observation biases}
- \text{observation biases}
The observation errors have a var-cov matrix

\[
\begin{bmatrix}
\eta_1^2 & \eta_{1,2} & \eta_{1,3} & \eta_{1,4} & \eta_{1,5} \\
\eta_{1,2} & \eta_2^2 & \eta_{3,2} & \eta_{2,4} & \eta_{2,5} \\
\eta_{1,3} & \eta_{3,2} & \eta_3^2 & \eta_{3,4} & \eta_{3,5} \\
\eta_{1,4} & \eta_{2,4} & \eta_{3,4} & \eta_4^2 & \eta_{4,5} \\
\eta_{1,5} & \eta_{2,5} & \eta_{3,5} & \eta_{4,5} & \eta_5^2 \\
\end{bmatrix}
\]

unconstrained

unique
variances and
uncorrelated
errors
diagonal

identical
variances and
uncorrelated
errors
diagonal
The harbor seal multivariate state-space model in matrix form

\[ x_t = B x_{t-1} + u + w_t \quad \text{where} \quad w_t \sim MVN(0, Q) \]

\[ y_t = Z x_t + a + v_t \quad \text{where} \quad v_t \sim MVN(0, R) \]
Instead of N, S, Str. J subpopulations, we could have other combinations and numbers of subpopulations.

Str. JF
San Isl.
E. Bays
P.S.
Hood C.

\[
\begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

Str of Juan de Fuca & San Juan Is sites = 1st subpop
Eastern bays = 2nd subpop
Hood C. & S. Puget S. = 3rd subpop

Strait of Juan de Fuca = 1st sub pop
San Juan Is sites = 2nd sub pop
Eastern bays, Hood Canal & S. Puget Sound = 3rd sub pop

One Puget Sound population and all sites are sampling it
One population
The same model can capture many different underlying population structures and observation structures.

\[ \mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{u} + \mathbf{w}_t \quad \text{where} \quad \mathbf{w}_t \sim \text{MVN}(0, \mathbf{Q}) \]

\[ \mathbf{y}_t = \mathbf{Zx}_t + \mathbf{a} + \mathbf{v}_t \quad \text{where} \quad \mathbf{v}_t \sim \text{MVN}(0, \mathbf{R}) \]
Inferring spatial structure from time-series data: using multivariate state-space models to detect metapopulation structure of California sea lions in the Gulf of California, Mexico

Hypotheses about the population structure:
Diet, Disease, DNA, Distance
(2 null models: no structure and fully structured)
Table 1. Model performance, given by Akaike’s Information Criterion (AIC) b-value, across the six hypotheses for the subpopulation configuration

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Hypotheses (m = no. subpopulations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>Q</td>
</tr>
<tr>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>Unique</td>
<td>Same</td>
</tr>
<tr>
<td>Same</td>
<td>Unique</td>
</tr>
<tr>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>Unique</td>
<td>Same</td>
</tr>
<tr>
<td>Unique</td>
<td>Unique</td>
</tr>
<tr>
<td>Same</td>
<td>Unique</td>
</tr>
<tr>
<td>Unique</td>
<td>Unique</td>
</tr>
<tr>
<td>Same</td>
<td>Correlated</td>
</tr>
<tr>
<td>Unique</td>
<td>Correlated</td>
</tr>
<tr>
<td>Same</td>
<td>Correlated</td>
</tr>
<tr>
<td>Unique</td>
<td>Correlated</td>
</tr>
</tbody>
</table>

Process errors (Q) may be independent (a diagonal matrix) with variances that are the same magnitude across subpopulations (same), independent with unequal variances across subpopulations (unique) or may be temporally correlated, meaning an unconstrained Q matrix (correlated). The growth rate (u) and observation error matrix (R) parameters may also be equal (same) or unique across subpopulations. The model best supported by the data is shown in bold; complex models that did not fully converge are not applicable.
The MARSS manual has two chapters using MARSS models to analyze spatial count data harbor seal count data from the west coast of the USA.
Chapter 7: Combining multi-site data to estimate regional population trends
Chapter 8 in HWS 2014

Identifying spatial structure and covariance in harbor seals on the west coast of the USA

2000km
Shortcut for the Z matrix

\[
\begin{array}{c}
\text{Coastal Estuaries} \\
\text{Olympic Peninsula} \\
\text{Str. Juan de Fuca} \\
\text{San Juan Islands} \\
\text{Eastern Bays} \\
\text{Puget Sound} \\
\text{CA.Mainland} \\
\text{CA.Channel Islands} \\
\text{OR North Coast} \\
\text{OR South Coast} \\
\text{Georgia Strait}
\end{array}
\begin{pmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{pmatrix}
\]

\[=\]

\[+ a + v\]

\[\text{factor(c("or.wa","or.wa","ps","ps","ps","ps","ca","ca","or.wa","or.wa","ps"))}\]