Bayesian State Space Models

Models for Socio-Environmental Data

Chris Che-Castaldo, Mary B. Collins, and N. Thompson Hobbs

August 23, 2017

SESYNC

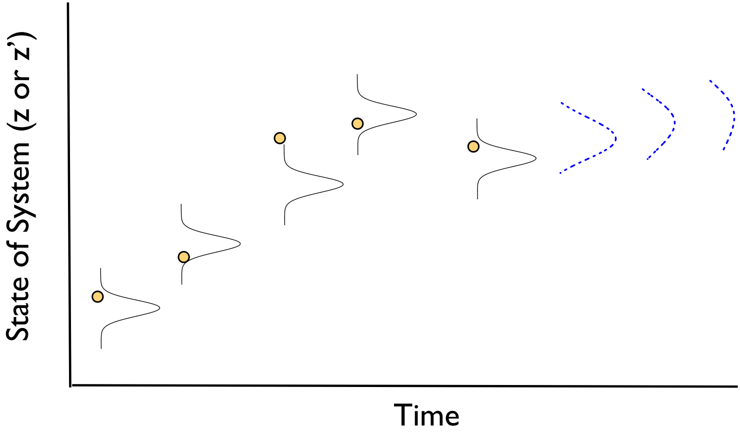
Roadmap

- Overview
- Model types with examples
  - discrete time, single state
  - continuous time (briefly)
- Forecasting
- Coding tips
- Discrete time, multiple states (for those interested, Tamara)

What are state space models?

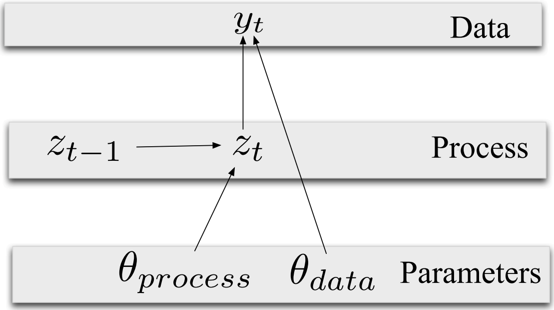
$$[y_t | \boldsymbol{\theta}_d, z_t]$$
$$[z_t | \boldsymbol{\theta}_p, z_{t-1}]$$

The idea is simple. We have a stochastic model of an unobserved, true state (z_t) and a stochastic model that relates our observations (y_t) to the true state.


State of System (z or z')
Time

A broadly applicable approach to modeling dynamic processes in ecology

$$[\mathbf{z}, \theta_{process}, \theta_{data} | \mathbf{y}] \propto \prod_{t=2}^{T} [y_t | \theta_{data}, z_t] [z_t | \theta_{process}, z_{t-1}] [\theta_{process}, \theta_{data}, z_1]$$



Sources of uncertainty in state space models

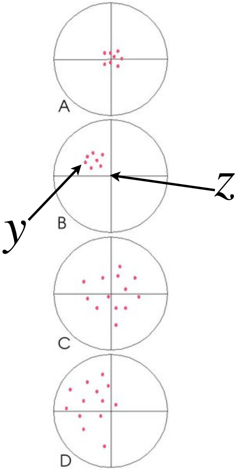
Process uncertainty

- Failure to perfectly represent process
- Propagates in time
- Decreases with model improvement
- Estimation allows forecasting

Observation uncertainty

- Failure to perfectly observe process
- Does not propagate
- Sampling uncertainty decreases with increased sampling effort.
- Measurement uncertainly decreases with improved instrumentation, calibration, etc.

Components of observation uncertainty



- Measurement $\left[y|h\left(z,\boldsymbol{\theta}_d\right),\boldsymbol{\sigma}^2_{measurement}\right]$
- Sampling $\left[y|z,\boldsymbol{\sigma}^2_{sampling}\right]$

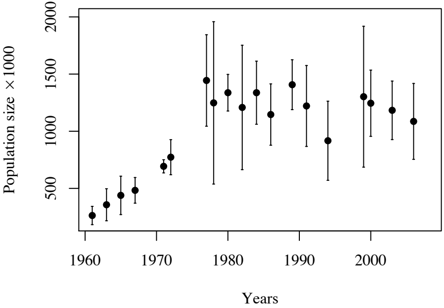
When can we separate process variance from observation variance?

- Replication of the observation for the latent state with sufficient n
- Calibration model with properly estimate prediction variance
- Strongly differing "structure" in process and observation models
- We may not need to separate them–sometimes the observed state and the true state are the same.

General joint and posterior distribution for single state model

$$
\begin{aligned}
\text{Deterministic model} &= g(\boldsymbol{\theta}_{process}, z_{t-1}, \mathbf{x}_{t-1}) \\
\left[\mathbf{z}, \boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data}, \boldsymbol{\sigma}_p^2, \boldsymbol{\sigma}_d^2 | \mathbf{y}\right] &\propto \prod_{t=2}^{T} \left[y_t | \boldsymbol{\theta}_{data}, z_t, \boldsymbol{\sigma}_o^2\right] \\
&\quad \times \left[z_t | g(\boldsymbol{\theta}_{process}, z_{t-1}, \mathbf{x}_{t-1}), \boldsymbol{\sigma}_p^2\right] \\
&\quad \times \left[\boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data}, \boldsymbol{\sigma}_p^2, \boldsymbol{\sigma}_o^2, z_1\right]
\end{aligned}
$$

Modeling the Serengeti wildebeest population



- 48 year time series
- Annual means and standard deviations of population size for 19 years
- Spatially replicated census
- Annual data on dry season rainfall

How does rainfall influence density dependence?

$$g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right) = z_{t-1} e^{\left(\beta_0 + \beta_1 z_{t-1} + \beta_2 x_{t-1} + \beta_3 z_{t-1} x_{t-1}\right)\Delta t}$$

- $z_t =$ true population size
- $x_{t-1} =$ standardized, annual dry season rainfall during time $t-1$ to t.
- $\beta_0 = r_{max} =$ intrinsic, per-capita rate of increase at average rainfall
- $\beta_1 =$ strength of density dependence, $\frac{r}{K}$ at average rainfall.
- $\beta_2 =$ change in rate of increase per standard deviation change in rainfall
- $\beta_3 =$ effect of rainfall on strength of density dependence

Process model predicting medians

$$z_t \sim \text{lognormal}\left(\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right), \sigma_p^2\right)$$

- $\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right)$, the centrality parameter, the mean of z_t on the log scale
- σ_p^2, the scale parameter, the variance of z_t on the log scale
- What does the deterministic model predict?
  - define centrality parameter $= \alpha_t$
  - median$(z_t) = e^{\alpha_t}$
  - $\alpha_t = \log(\text{median}(z_t))$
  - median$\left(z_t\right) = g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)$

Process model predicting medians

$$z_t \sim \text{lognormal}\left(\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right), \boldsymbol{\sigma}_p^2\right)$$

- $\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right)$, the centrality parameter, the mean of z_t on the log scale
- $\boldsymbol{\sigma}_p^2$, the scale parameter, the variance of z_t on the log scale
- What does the deterministic model predict?
  - define centrality parameter $= \alpha_t$
  - median$(z_t) = e^{\alpha_t}$
  - $\alpha_t = \log(\text{median}(z_t))$
  - median$\left(z_t\right) = g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)$

Process model predicting medians

$$z_t \sim \text{lognormal}\left(\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right), \sigma_p^2\right)$$

- $\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right)$, the centrality parameter, the mean of z_t on the log scale
- σ_p^2, the scale parameter, the variance of z_t on the log scale
- What does the deterministic model predict?
  - define centrality parameter $= \alpha_t$
  - median$(z_t) = e^{\alpha_t}$
  - $\alpha_t = \log(\text{median}(z_t))$
  - median$\left(z_t\right) = g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)$

Process model predicting medians

$$z_t \sim \text{lognormal}\left(\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right), \sigma_p^2\right)$$

- $\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right)$, the centrality parameter, the mean of z_t on the log scale
- σ_p^2, the scale parameter, the variance of z_t on the log scale
- What does the deterministic model predict?
  - define centrality parameter $= \alpha_t$
  - median$(z_t) = e^{\alpha_t}$
  - $\alpha_t = \log(\text{median}(z_t))$
  - median$\left(z_t\right) = g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)$

Process model predicting medians

$$z_t \quad \sim \quad \text{lognormal}\left(\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right), \sigma_p^2\right)$$

- $\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right)$, the centrality parameter, the mean of z_t on the log scale
- σ_p^2, the scale parameter, the variance of z_t on the log scale
- What does the deterministic model predict?
  - define centrality parameter $= \alpha_t$
  - median$(z_t) = e^{\alpha_t}$
  - $\alpha_t = \log(\text{median}(z_t))$
  - median$\left(z_t\right) = g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)$

Review of relationships between normal and lognormal

1. $z_t = g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right) \exp\left(\varepsilon_t\right),\ \varepsilon_t \sim \text{normal}\left(0, \sigma_p^2\right)$
2. $\log\left(z_t\right) = \log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right) + \varepsilon_t,\ \varepsilon_t \sim \text{normal}\left(0, \sigma_p^2\right)$
3. $\log\left(z_t\right) \sim \text{normal}\left(\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right), \sigma_p^2\right)$
4. $z_t \sim \text{lognormal}\left(\underbrace{\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right)}_{\text{centrality parameter}}, \underbrace{\sigma_p^2}_{\text{scale parameter}}\right)$

Review of relationships between normal and lognormal

1. $z_t = g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}) \exp(\varepsilon_t)$, $\varepsilon_t \sim \text{normal}(0, \sigma_p^2)$
2. $\log(z_t) = \log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1})) + \varepsilon_t$, $\varepsilon_t \sim \text{normal}(0, \sigma_p^2)$
3. $\log(z_t) \sim \text{normal}(\log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1})), \sigma_p^2)$
4. $z_t \sim \text{lognormal}\left(\underbrace{\log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}))}_{\text{centrality parameter}}, \underbrace{\sigma_p^2}_{\text{scale parameter}}\right)$

Review of relationships between normal and lognormal

1. $z_t = g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}) \exp(\varepsilon_t)$, $\varepsilon_t \sim$ normal $(0, \sigma_p^2)$
2. $\log(z_t) = \log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1})) + \varepsilon_t$, $\varepsilon_t \sim$ normal $(0, \sigma_p^2)$
3. $\log(z_t) \sim \text{normal}\left(\log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1})), \sigma_p^2\right)$
4. $z_t \sim$ lognormal $\left(\underbrace{\log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}))}_{\text{centrality parameter}}, \underbrace{\sigma_p^2}_{\text{scale parameter}}\right)$

Review of relationships between normal and lognormal

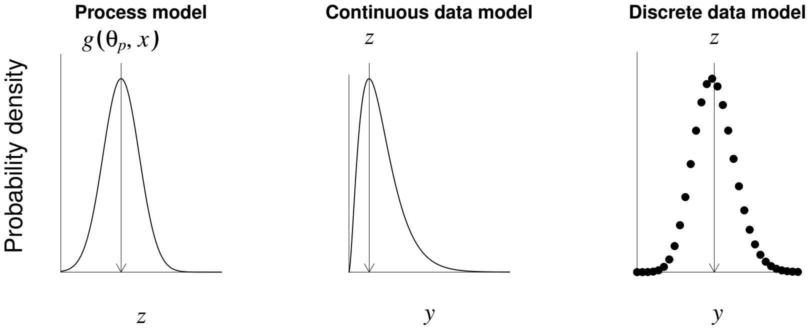
1. $z_t = g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\exp\left(\varepsilon_t\right),\ \varepsilon_t \sim \text{normal}\left(0, \sigma_p^2\right)$
2. $\log\left(z_t\right) = \log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right) + \varepsilon_t,\ \varepsilon_t \sim \text{normal}\left(0, \sigma_p^2\right)$
3. $\log\left(z_t\right) \sim \text{normal}\left(\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right), \sigma_p^2\right)$
4. $z_t \sim \text{lognormal}\left(\underbrace{\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right)}_{\text{centrality parameter}}, \underbrace{\sigma_p^2}_{\text{scale parameter}}\right)$

It is also possible to moment match the mean

$$
\begin{aligned}
\boldsymbol{\mu}_t &= g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right) && (1)\\
\boldsymbol{\alpha}_t &= \log(\boldsymbol{\mu}_t) - \frac{1}{2}\log\left(\frac{\boldsymbol{\mu}_t^2 + \boldsymbol{\sigma}^2}{\boldsymbol{\mu}_t^2}\right) && (2)\\
z_t &\sim \text{lognormal}(\boldsymbol{\alpha}_t, \boldsymbol{\sigma}^2) && (3)
\end{aligned}
$$

You should do it this way if you have derived quantities computed as sums of the z_t, for example when modeling a total population from subpopulations in different sites.

Why a continuous distribution for a “discrete state”?



The data

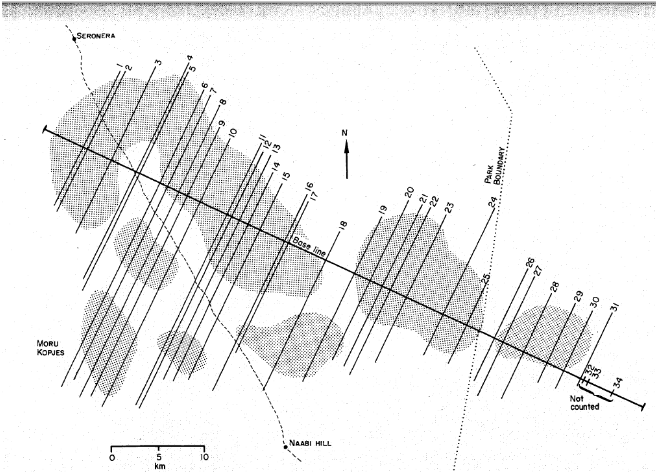


Fig. 2. The orientation of the base-line and of the random transects in the May 1971 sample count. Shading shows approximate positions of the main wildebeest herds.

Observation model

$$y_t \sim \mathrm{normal}\left(z_t, y.sd_t\right)$$

- y_t is the observed mean number of animals across all transects
- $y.sd_t$ is the observed standard deviation across transects
- z_t is the unobserved, true state, the *mean of the data distribution*

We choose a normal distribution for the likelihood because the y_t are the annual mean of means of densities of wildebeest on many transects. For now, we ignore the potential for spatial autocorrelation among transects.

Posterior and joint distributions

$$
\left[\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\sigma}_p^2 | \mathbf{y}\right] \propto \underbrace{\prod_{\forall t \in \mathbf{y}.i} \left[y_t \,\middle|\, z_t, y.sd_t \right]}_{\text{data model}}
$$

$$
\times \underbrace{\prod_{t=2}^{48} \left[z_t | g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right), \boldsymbol{\sigma}_p^2 \right]}_{\text{process model}} \times \underbrace{[\beta_0]\,[\beta_1]\,[\beta_2]\,[\beta_3]\,\left[\boldsymbol{\sigma}_p^2\right][z_1]}_{\text{parameter models}}
$$

- $\mathbf{y}.i$ is a vector of years with non-missing census data
- $y_t \sim$ normal$(z_t, y.sd_t)$
- $z_t \sim$ lognormal$\left(\log\left(g\left(\boldsymbol{\beta}, z_{t-1}, x_{t-1}\right)\right), \boldsymbol{\sigma}_p^2\right)$
- $\beta_0 \sim$ normal$\left(.234, .136^2\right)$ informative prior
- $\beta_{i \in 1,2,3} \sim$ normal$(0, 1000)$
- $\boldsymbol{\sigma}_p^2 \sim$ gamma$\left(.01, .01\right)$
- $z_1 \sim$ normal$(y_1, y.sd_1)$

Continuous time models

$$\frac{dz_1}{dt} = k_1 z_1 - k_2 z_1 z_2 \tag{4}$$

$$\frac{dz_2}{dt} = - k_3 z_1 + \alpha k_2 z_1 z_2 \tag{5}$$

$$\frac{dz_3}{dt} = \frac{k_4 z_3}{k_5 + z_3} \tag{6}$$

$\left[\mathbf{z_t} | g\left((\mathbf{k}, \mathbf{z}_{t-1}, x_t\right), \boldsymbol{\sigma}_p^2\right]$

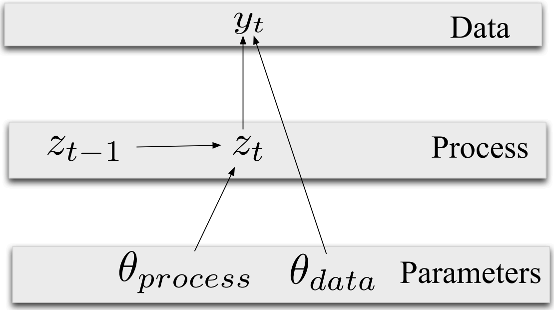
Implementing the process model may need a numerical solver to align the states with the data.

Continuous time models

- Must deterministically update states at discrete intervals to match with data
- To estimate states:
  - Use analytical solutions to ODE system if available.
  - For models without analytical solutions:
    - OpenBUGS and STAN have ODE solvers.
    - Euler's or Runge-Kutta IV can be embedded in JAGS or OpenBUGS for simple models.
    - Best: Write your own MCMC sampler with embedded numerical solver.
    - See: Campbell, E. E., W. J. Parton, J. L. Soong, K. Paustian, N. T. Hobbs, and M. F. Cotrufo. 2016. Using litter chemistry controls on microbial processes to partition litter carbon fluxes with the Litter Decomposition and Leaching (LIDEL) model. Soil Biology & Biochemistry 100:160-174.

A broadly applicable approach to modeling dynamic processes in ecology

$$[\mathbf{z}, \theta_{process}, \theta_{data} | \mathbf{y}] \propto \prod_{t=2}^{T} [y_t | \theta_{data}, z_t] \, [z_t | \theta_{process}, z_{t-1}] \, [\theta_{process}, \theta_{data}, z_1]$$



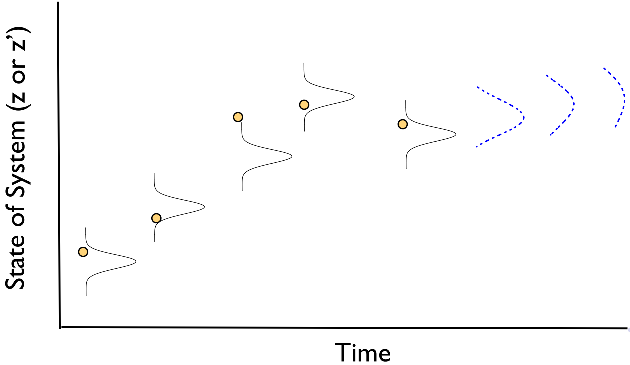
Roadmap

- Overview
- Model types with examples
  - discrete time, single state
  - continuous time (briefly)
- Forecasting
- Coding tips

Bayesian forecasting future states z'

$$\underbrace{\left[z'_{T+1}|\mathbf{y}\right]}_{\text{predictive process distribution}} =$$

$$\int_{\theta_1...\theta_P}\int_{z_1...}\int_{z_T}\left[z'_{T+1}|\mathbf{z},\boldsymbol{\theta}_{process}\right]\underbrace{\left[\mathbf{z},\boldsymbol{\theta}_{process},\boldsymbol{\theta}_{data}|\mathbf{y}\right]}_{\text{posterior distribution}}dz...dz_t\,d\boldsymbol{\theta}_1...d\boldsymbol{\theta}_P$$



Predictive process distribution

The MCMC output:

i	θ_1	θ_1	θ_3								
1	.42	3.3	20.3	$z_{1,1}$	$z_{1,2}$	$\cdots$	$z_{1,T}$	$z'_{1,T+1}$	$z'_{1,T+2}$	$\cdots$	$z'_{1,T+F}$
2	.41	2.3	18.5	$z_{2,1}$	$z_{2,2}$	$\cdots$	$z_{2,T}$	$z'_{2,T+1}$	$z'_{,2T+2}$	$\cdots$	$z'_{2,T+F}$
3	.46	3.1	16.6	$z_{3,1}$	$z_{3,2}$	$\cdots$	$z_{3,T}$	$z'_{3,T+1}$	$z'_{3,T+2}$	$\cdots$	$z'_{3T+,F}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	.39	3.4	22.1	$z_{n,1}$	$z_{n,2}$	$\cdots$	$z_{n,T}$	$z'_{n,T+1}$	$z'_{n,T+2}$	$\cdots$	$z'_{n,T+F}$

n = number of iterations

T = final time with data

F = number of forecasts beyond data

Posterior and joint distribution with forecasts

$$\mu_t = g(\boldsymbol{\theta}_{process}, z_{t-1}, \mathbf{x}_{t-1})$$

$$[\mathbf{z}, \boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data} | \mathbf{y}] \propto$$
$$\prod_{t=2}^{T} [y_t | \boldsymbol{\theta}_{data}, z_t] \prod_{t=2}^{T+F} [z_t | \mu_t] [\boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data}, z_1]$$

Posterior and joint distribution with missing data

$$\mu_t = g(\boldsymbol{\theta}_{process}, z_{t-1}, \mathbf{x}_{t-1})$$

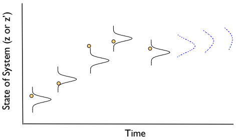
$$[\mathbf{z}, \boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data} | \mathbf{y}] \propto \prod_{\forall t \in \mathbf{y}.i}^{T} [y_t | \boldsymbol{\theta}_{data}, z_t] \prod_{t=2}^{T} [z_t | \mu_t] [\boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data}, z_1]$$

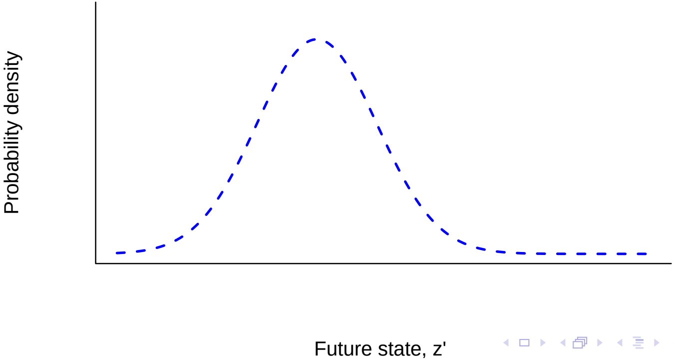
Forecasting

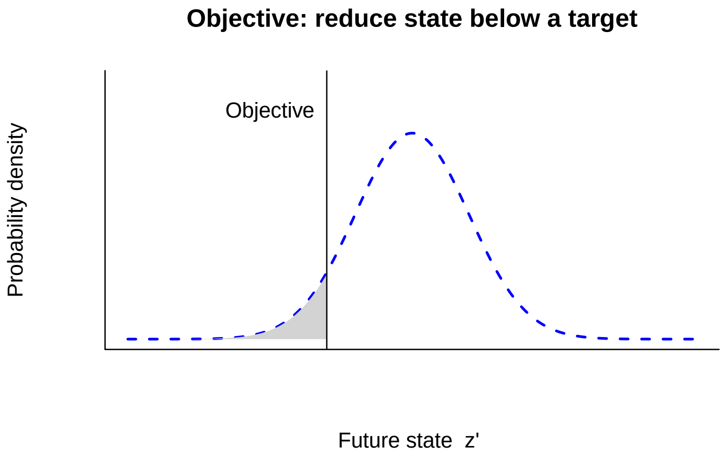
The fundamental problem of management:

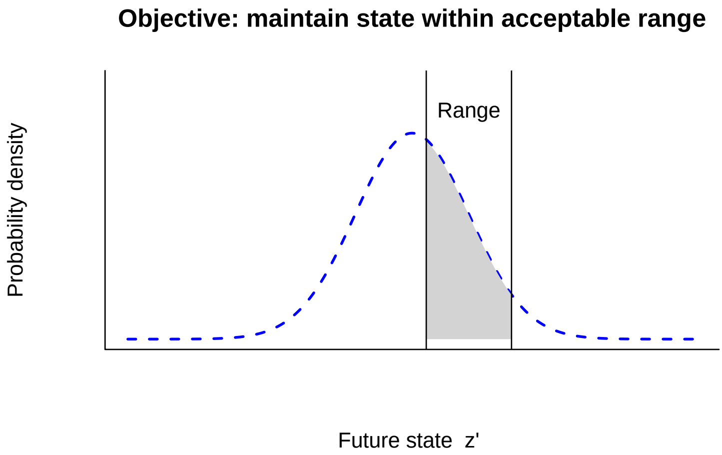
What actions can we take today that will allow us to meet goals for the future?

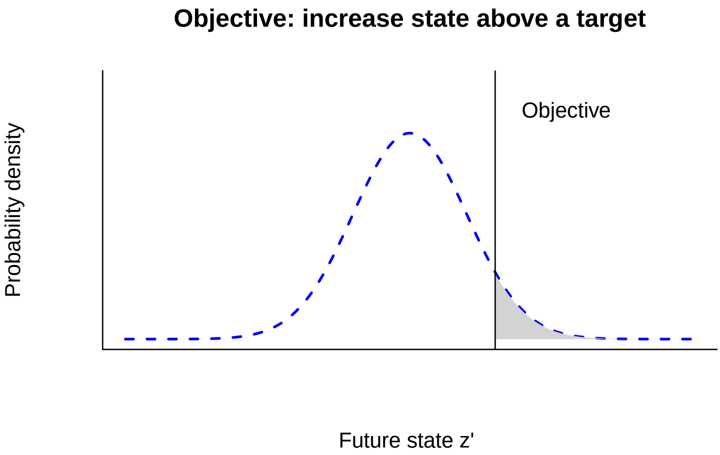
Predictive process distribution of z'

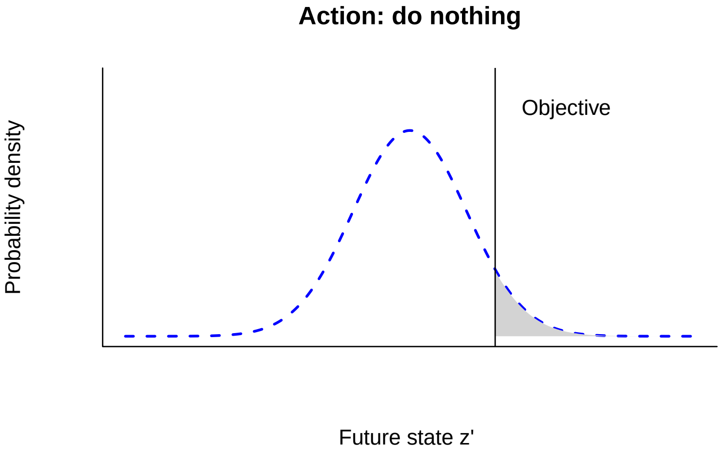


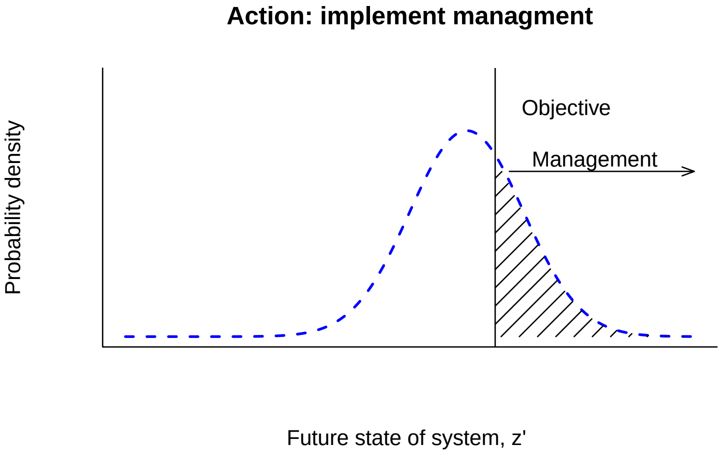


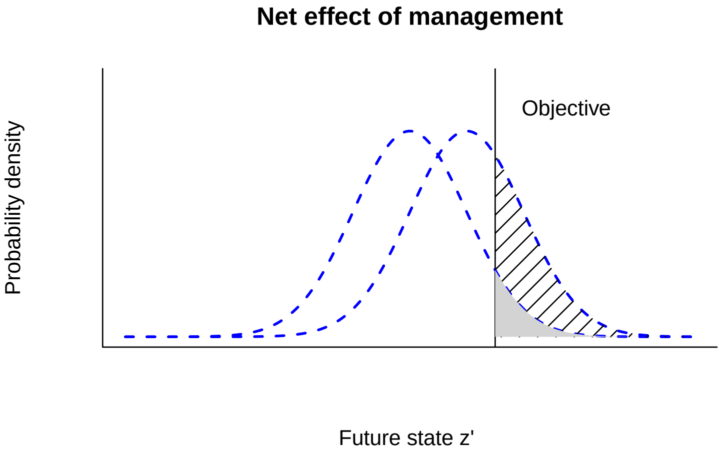

Objective: reduce state below a target
Objective
Probability density
Future state z'


Objective: maintain state within acceptable range
Range
Probability density
Future state z'

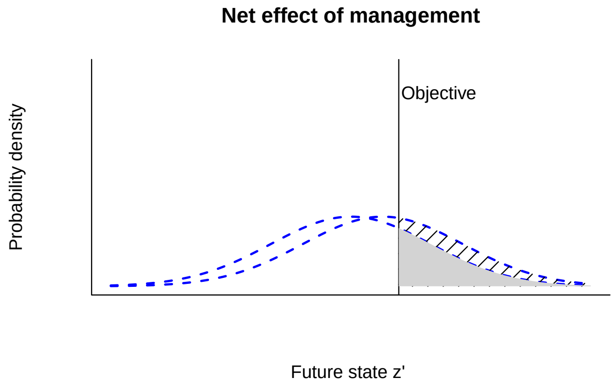

Objective: increase state above a target
Objective
Probability density
Future state z'


Action: do nothing
Objective
Probability density
Future state z'


Action: implement managment
Objective
Management
Probability density
Future state of system, z'



Hobbs, N. T., C. Geremia, J. Treanor, R. Wallen, P. J. White, M. B. Hooten, and J. C. Rhyan. 2015. State-space modeling to support management of brucellosis in the Yellowstone bison population. Ecological Monographs 85:3-28.


Net effect of management
Objective
Probability density
Future state z'

JAGS code for posterior and joint distributions

$$\left[\mathbf{z},\boldsymbol{\beta},\sigma_p^2|\mathbf{y}\right] \propto \underbrace{\prod_{\forall t\in y.i}\left[y_t\,\middle|\,z_t, y.sd_t\right]}_{\text{data model}}$$

$$\times \underbrace{\prod_{t=2}^{48}\left[z_t|g\left(\boldsymbol{\beta},z_{t-1},x_{t-1}\right),\sigma_p^2\right]}_{\text{process model}} \times \underbrace{[\beta_0][\beta_1][\beta_2][\beta_3]\left[\sigma_p^2\right][z_1]}_{\text{parameter models}}$$

```
model{
#Priors
b[1] ~ dnorm(.234,1/.136^2)
for(j in 2:n.coef){
b[j] ~ dnorm(0,.0001)
}
tau.p ~ dgamma(.01,.01)
sigma.p <- 1/sqrt(tau.p)
z[1]  ~ dnorm(N.obs[1],tau.obs[1])  #this must be treated as prior so that you have z[t-
1]
##Process model
for(t in 2:(T+F)){
mu[t] <- log(z[t-1]*exp(b[1] + b[2]*z[t-1] + b[3]*x[t] +b[4]*x[t]*z[t-1]))
z[t] ~ dlnorm(mu[t], tau.p)
}

#Data model
for(j in 2:n.obs){
N.obs[j] ~ dnorm(z[index[j]],tau.obs[j]) #index to match z[t] with data
}
}#end of model
```

Posterior predictive checks for time series data

Test statistic:

$$\frac{1}{T-1}\sum_{t=2}^{T} \mid y_t - y_{t-1} \mid \tag{7}$$

Conventional statistics are also used (mean, CV, discrepancy statistic for the y_t.

Reilly, C., A. Gelman, and J. Katz, 2001. Poststratification without Population Level Information 731 on the Poststratifying Variable, with Application to Political Polling. Journal of the American 732 Statistical Association 96:1–11.

Posterior predictive checks and test for autocorrelation

```
#Derived quantities for model evaluation

for(i in 1:n.obs){

    #for autocorrelation test

epsilon.obs[i] <- N.obs[i] - z[index[i]]

 # simulate new data

         N.new[i] ~ dnorm(z[index[i]],tau.obs[i])

sq[i] <- (N.obs[i] - z[index[i]] )^2
sq.new[i] <-(N.new[i] - z[index[i]]) ^2
}
fit  <- sum(sq[])
fit.new <- sum(sq.new[])
pvalue <-step(fit.new-fit)
```

General joint and posterior distribution for multi-state model

$$\boldsymbol{\mu}_t = \mathbf{A}\mathbf{z}_t, \text{ process parameters are elements of matrix } \mathbf{A}$$

$$[\mathbf{z}, \boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data} | \mathbf{Y}] \propto \prod_{t=2}^{T} [\mathbf{y}_t | \boldsymbol{\theta}_{data}, \mathbf{z}_t] [\mathbf{z}_t | \boldsymbol{\mu}_t] [\boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data}, \mathbf{z}_1]$$

Multiple states: Ann Raiho's matrix model[1]

- Problem: Evaluate management alternatives for managing overabundant deer in national parks.
- Data
  - Annual census, corrected for uncounted animals using distance sampling
  - Annual classification counts

[1] Raiho, A. M., M. B. Hooten, S. Bates, and N. T. Hobbs. 2015. Forecasting the effects of fertility control on overabundant ungulates: white-tailed deer in the National Capital Region. PLoS ONE 10. 10.1371/journal.pone.0143122

States

state	definition
n_1	The number of juvenile deer, aged 6 months on their first census
n_2	The number of adult female deer, aged 18 months and older
n_3	The number of adult male deer, aged 18 months and older

Deterministic Model

f	number of recruits per female surviving to census
ϕ_j	probability that a juvenile (aged 6 months) survives to 18 months
ϕ_d	annual survival probabilty of adult females
ϕ_b	annual survival probability of adult males
m	proportion of juveniles surviving to adults that are female

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} 0 & \phi_d^{\frac{1}{2}} f & 0 \\ m\phi_j & \phi_d & 0 \\ (1-m)\,\phi_j & 0 & \phi_b \end{pmatrix} \\ \mathbf{n}_t &= \mathbf{A}\mathbf{n}_{t-1}. \end{aligned}$$

The posterior and joint distribution

$$\left[\boldsymbol{\phi}, m, f, \mathbf{N}, \underbrace{\boldsymbol{\sigma}_p, \boldsymbol{\rho}}_{\text{elements of } \boldsymbol{\Sigma}} \middle| \mathbf{y}^{\text{census}}, \mathbf{y}^{\text{census.sd}}, \mathbf{Y}^{\text{class}}\right] \propto$$

$$\underbrace{\prod_{t=2}^{T} \text{multivariate normal}\left(\log(\mathbf{n}_t) \middle| \log\left(\mathbf{A}_t \mathbf{n}_{t-1}\right), \boldsymbol{\Sigma}\right)}_{\text{process model}}$$

$$\times \text{data models} \times \text{priors}$$

The posterior and joint distribution

$$\left[\boldsymbol{\phi}, m, f, \mathbf{N}, \underbrace{\boldsymbol{\sigma}_p, \boldsymbol{\rho}}_{\text{elements of } \boldsymbol{\Sigma}} \mid \mathbf{y}^{\text{census}}, \mathbf{y}^{\text{census.sd}}, \mathbf{Y}^{\text{class}}\right] \propto$$

$$\underbrace{\prod_{t=2}^{T} \text{multivariate normal}\left(\log(\mathbf{n}_t) \mid \log\left(\mathbf{A}_t \mathbf{n}_{t-1}\right), \boldsymbol{\Sigma}\right)}_{\text{process model}}$$

$$\prod_{t=2}^{T} \underbrace{\text{normal}\left(y_t^{\text{census}} \mid \sum_{i=1}^{3} n_{i,t}, y_t^{\text{census.sd}}\right)}_{\text{data model 1}}$$