

Bayesian State Space Models

Models for Socio-Environmental Data

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

August 23, 2017



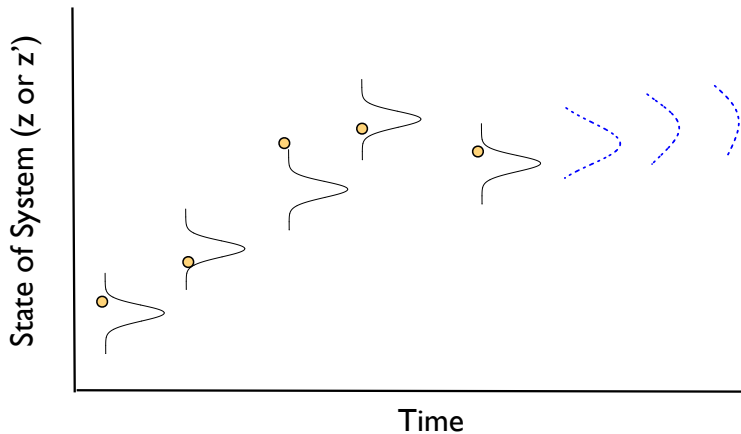
Roadmap

- ▶ 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?

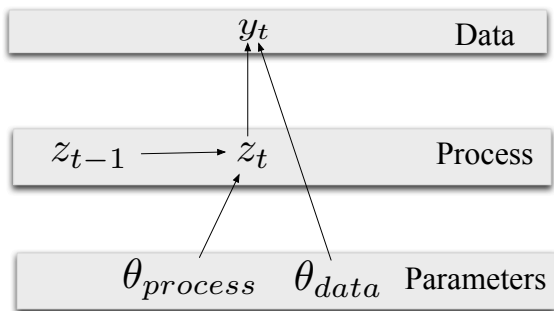
$$\begin{aligned} &[y_t | \boldsymbol{\theta}_d, z_t] \\ &[z_t | \boldsymbol{\theta}_p, z_{t-1}] \end{aligned}$$

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.



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

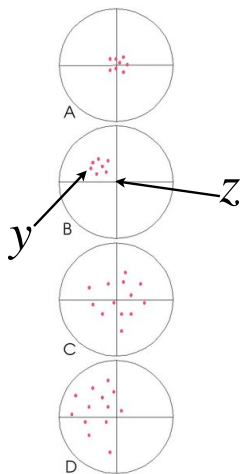
Process uncertainty

- ▶ 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 uncertainty decreases with improved instrumentation, calibration, etc.

Components of observation uncertainty



- ▶ Measurement $[y|h(z, \theta_d), \sigma_{\text{measurement}}^2]$
- ▶ Sampling $[y|z, \sigma_{\text{sampling}}^2]$

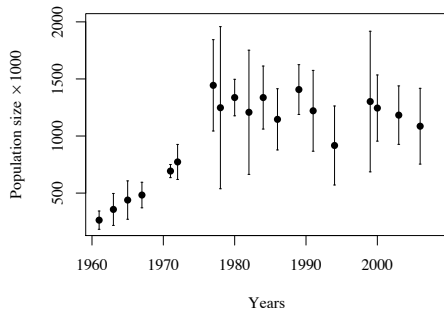
When can we separate process variance from observation variance?

- ▶ 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

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

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?

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

- ▶ 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
- ▶ β_1 = strength of density dependence, $\frac{r}{K}$ at average rainfall.
- ▶ β_2 = change in rate of increase per standard deviation change in rainfall
- ▶ β_3 = effect of rainfall on strength of density dependence

Process model predicting medians

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

- ▶ $\log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}))$, 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 $= \alpha_t$
 - ▶ $\text{median}(z_t) = e^{\alpha_t}$
 - ▶ $\alpha_t = \log(\text{median}(z_t))$
 - ▶ $\text{median}(z_t) = g(\boldsymbol{\beta}, z_{t-1}, x_{t-1})$

Process model predicting medians

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

- ▶ $\log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}))$, 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 $= \alpha_t$
 - ▶ $\text{median}(z_t) = e^{\alpha_t}$
 - ▶ $\alpha_t = \log(\text{median}(z_t))$
 - ▶ $\text{median}(z_t) = g(\boldsymbol{\beta}, z_{t-1}, x_{t-1})$

Process model predicting medians

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

- ▶ $\log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}))$, 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 $= \alpha_t$
 - ▶ $\text{median}(z_t) = e^{\alpha_t}$
 - ▶ $\alpha_t = \log(\text{median}(z_t))$
 - ▶ $\text{median}(z_t) = g(\boldsymbol{\beta}, z_{t-1}, x_{t-1})$

Process model predicting medians

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

- ▶ $\log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}))$, 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 $= \alpha_t$
 - ▶ $\text{median}(z_t) = e^{\alpha_t}$
 - ▶ $\alpha_t = \log(\text{median}(z_t))$
 - ▶ $\text{median}(z_t) = g(\boldsymbol{\beta}, z_{t-1}, x_{t-1})$

Process model predicting medians

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

- ▶ $\log(g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}))$, 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 $= \alpha_t$
 - ▶ $\text{median}(z_t) = e^{\alpha_t}$
 - ▶ $\alpha_t = \log(\text{median}(z_t))$
 - ▶ $\text{median}(z_t) = g(\boldsymbol{\beta}, z_{t-1}, x_{t-1})$

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

It is also possible to moment match the mean

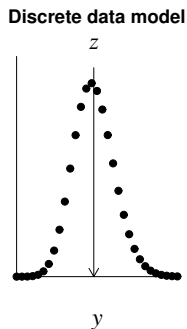
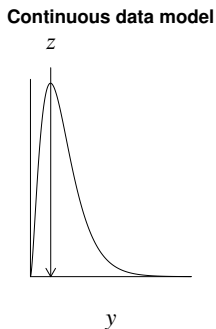
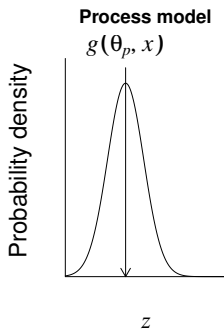
$$\mu_t = g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}) \quad (1)$$

$$\alpha_t = \log(\mu_t) - \frac{1}{2} \log\left(\frac{\mu_t^2 + \sigma^2}{\mu_t^2}\right) \quad (2)$$

$$z_t \sim \text{lognormal}(\alpha_t, \sigma^2) \quad (3)$$

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

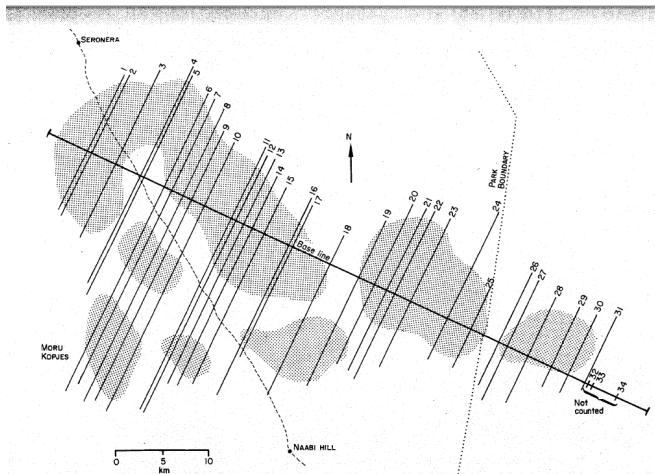


Fig. 2. The orientation of the base-line and of the random transects in the May 1971 sample count. Shading shows approximate positions of the main wildebeest herds.

Observation model

$$y_t \sim \text{normal}(z_t, y.sd_t)$$

- ▶ 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

$$\begin{aligned}
 [\mathbf{z}, \boldsymbol{\beta}, \sigma_p^2 | \mathbf{y}] &\propto \underbrace{\prod_{\forall t \in \mathbf{y}.i} \left[y_t \mid z_t, y.sd_t \right]}_{\text{data model}} \\
 &\times \underbrace{\prod_{t=2}^{48} \left[z_t | g(\boldsymbol{\beta}, z_{t-1}, x_{t-1}), \sigma_p^2 \right]}_{\text{process model}} \times \underbrace{[\beta_0] [\beta_1] [\beta_2] [\beta_3] [\sigma_p^2] [z_1]}_{\text{parameter models}}
 \end{aligned}$$

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

Continuous time models

$$\frac{dz_1}{dt} = k_1 z_1 - k_2 z_1 z_2 \quad (4)$$

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

$$\frac{dz_3}{dt} = \frac{k_4 z_3}{k_5 + z_3} \quad (6)$$

$$[\mathbf{z}_t | g((\mathbf{k}, \mathbf{z}_{t-1}, x_t), \sigma_p^2)]$$

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

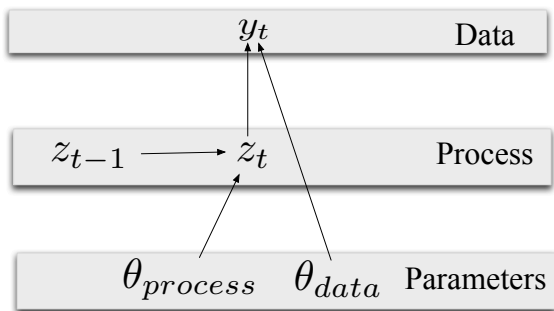
Continuous time models

- ▶ 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

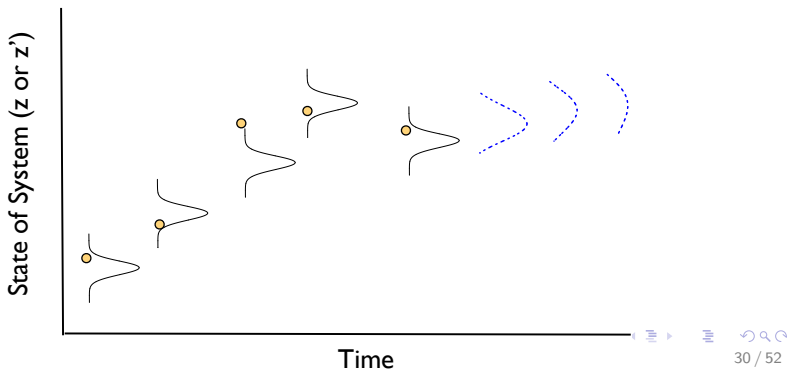
- ▶ Overview
- ▶ Model types with examples
 - ▶ discrete time, single state
 - ▶ continuous time (briefly)
- ▶ Forecasting
- ▶ Coding tips

Bayesian forecasting future states z'

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

predictive process distribution

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



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	$z_{2,T}$	$z'_{2,T+1}$	$z'_{2,T+2}$	\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	$z'_{3,T+F}$
\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

$$\begin{aligned}\mu_t &= g(\boldsymbol{\theta}_{process}, z_{t-1}, \mathbf{x}_{t-1}) \\ [\mathbf{z}, \boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data} | \mathbf{y}] &\propto \\ \prod_{t=2}^T [y_t | \boldsymbol{\theta}_{data}, z_t] &\prod_{t=2}^{T+F} [z_t | \mu_t] [\boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data}, z_1]\end{aligned}$$

Posterior and joint distribution with missing data

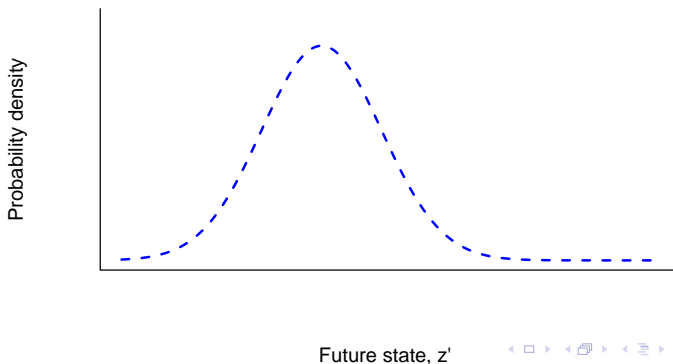
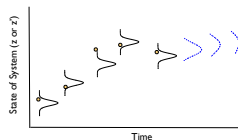
$$\begin{aligned}\mu_t &= g(\boldsymbol{\theta}_{process}, z_{t-1}, \mathbf{x}_{t-1}) \\ [\mathbf{z}, \boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data} | \mathbf{y}] &\propto \\ \prod_{\forall t \in \mathbf{y}.i}^T [y_t | \boldsymbol{\theta}_{data}, z_t] &\prod_{t=2}^T [z_t | \mu_t] [\boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data}, z_1]\end{aligned}$$

Forecasting

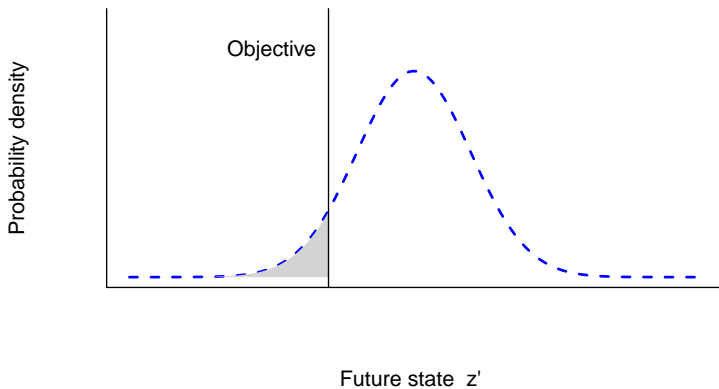
The fundamental problem of management:

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

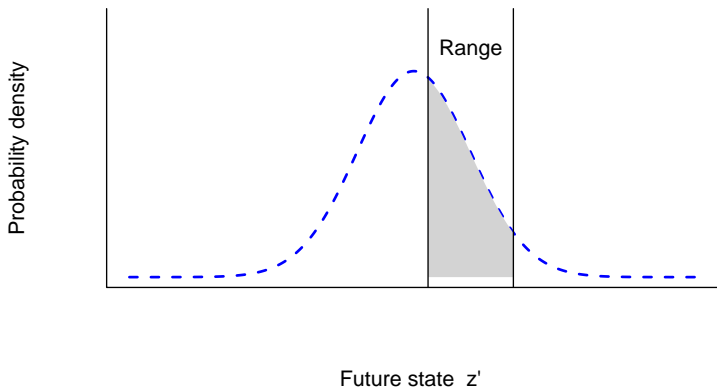
Predictive process distribution of z'



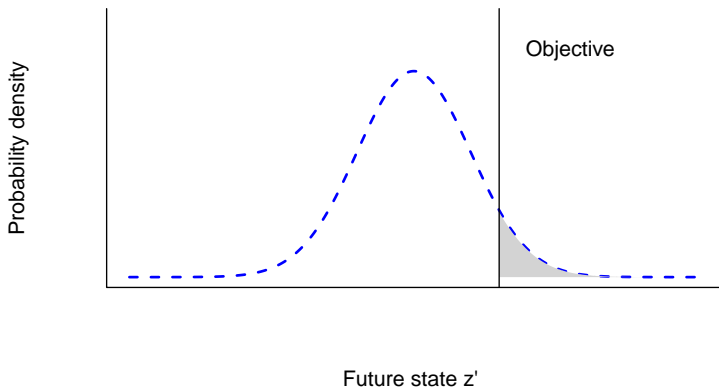
Objective: reduce state below a target



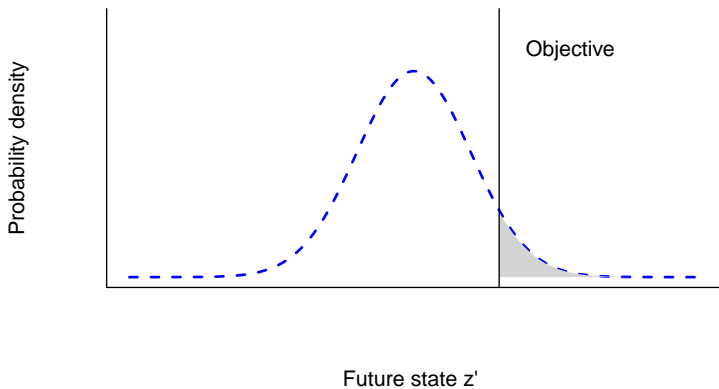
Objective: maintain state within acceptable range



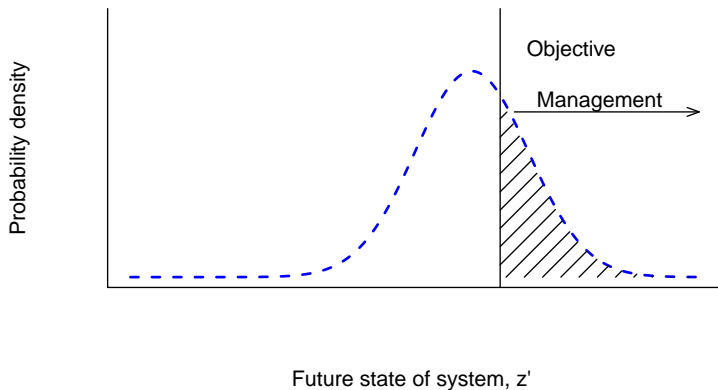
Objective: increase state above a target



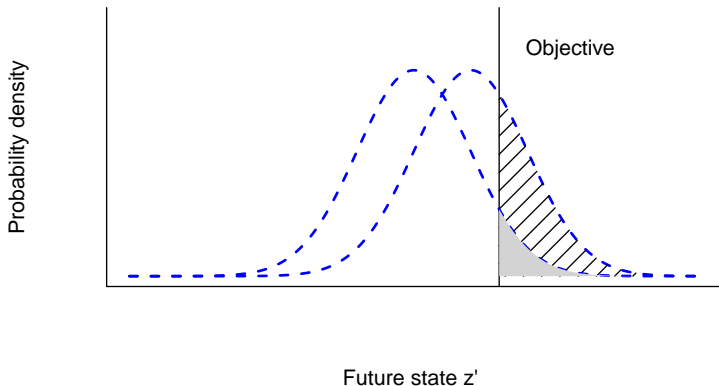
Action: do nothing



Action: implement management

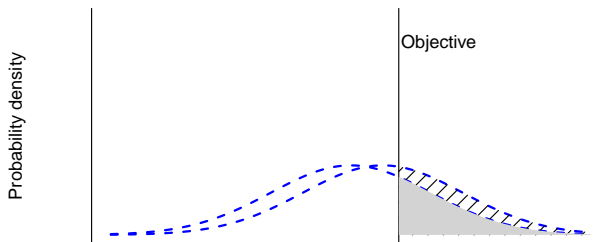


Net effect of management



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

Net effect of management



re state z'

JAGS code for posterior and joint distributions

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

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

```

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:

$$\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}| \quad (7)$$

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

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

$$[\mathbf{z}, \boldsymbol{\theta}_{process}, \boldsymbol{\theta}_{data} | \mathbf{Y}] \propto$$

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

Multiple states: Ann Raiho's matrix model¹



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

¹Raiho, 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 probability of adult females
- ϕ_b annual survival probability of adult males
- m proportion of juveniles surviving to adults that are female

$$\mathbf{A} = \begin{pmatrix} 0 & \phi_d^{\frac{1}{2}} f & 0 \\ m\phi_j & \phi_d & 0 \\ (1-m)\phi_j & 0 & \phi_b \end{pmatrix}$$

$$\mathbf{n}_t = \mathbf{A}\mathbf{n}_{t-1}.$$

The posterior and joint distribution

$$\begin{aligned}
 & \left[\boldsymbol{\phi}, m, f, \mathbf{N}, \underbrace{\boldsymbol{\sigma}_p, \boldsymbol{\rho}}_{\text{elements of } \boldsymbol{\Sigma}} \mid \mathbf{y}^{\text{census}}, \mathbf{y}^{\text{census.sd}}, \mathbf{Y}^{\text{class}} \right] \propto \\
 & \underbrace{\prod_{t=2}^T \text{multivariate normal}(\log(\mathbf{n}_t) \mid \log(\mathbf{A}_t \mathbf{n}_{t-1}), \boldsymbol{\Sigma})}_{\text{process model}} \\
 & \qquad \times \text{data models} \times \text{priors}
 \end{aligned}$$

The posterior and joint distribution

$$\begin{aligned}
 & \left[\phi, m, f, \mathbf{N}, \underbrace{\boldsymbol{\sigma}_p, \boldsymbol{\rho}}_{\text{elements of } \boldsymbol{\Sigma}} \mid \mathbf{y}^{\text{census}}, \mathbf{y}^{\text{census.sd}}, \mathbf{Y}^{\text{class}} \right] \propto \\
 & \underbrace{\prod_{t=2}^T \text{multivariate normal}(\log(\mathbf{n}_t) \mid \log(\mathbf{A}_t \mathbf{n}_{t-1}), \boldsymbol{\Sigma})}_{\text{process model}} \\
 & \underbrace{\prod_{t=2}^T \text{normal} \left(y_t^{\text{census}} \mid \sum_{i=1}^3 n_{i,t}, y_t^{\text{census.sd}} \right)}_{\text{data model 1}}
 \end{aligned}$$