Deterministic models, stochastic processes

Confronting population models with time series data

Benjamin Rosenbaum

iDiv German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig

EcoNetLab

Choice overload

- Modeling vs. Fitting
- Linear vs. nonlinear models
- Discrete-time vs. continuous-time models
- Deterministic vs. stochastic models
- Population-based vs. individual-based models
- Statistical methods (e.g. frequentist vs. Bayesian)
- Statistical software packages

→ Choice depends on the problem (data, questions)

but also personal preference

In this session

- **Modeling AND Fitting**
- Linear vs. **nonlinear** models
- **Discrete-time** vs. continuous-time models
- **Deterministic** vs. stochastic models
- **Population-based** vs. individual-based models
- Statistical methods (e.g. frequentist vs. **Bayesian**)
- Statistical software packages: **rstan**

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Concepts & Synthesis | Open Access

Integrating the underlying structure of stochasticity into community ecology

Lauren G. Shoemaker, Lauren L. Sullivan, Ian Donohue, Juliano S. Cabral, Ryan J. Williams, Margaret M. Mayfield, Jonathan M. Chase, Chengjin Chu, W. Stanley Harpole, Andreas Huth ... See all authors

First published: 25 October 2019 | https://doi.org/10.1002/ecy.2922

ECOLOGICAL MONOGRAPHS
ECOLOGICAL SOCIETY OF AMERICA

Review | Open Access

A guide to state–space modeling of ecological time series

Marie Auger-Méthé, Ken Newman, Diana Cole, Fanny Empacher, Rowenna Gryba, Aaron A. King, Vianey Leos-Barajas, Joanna Mills Flemming, Anders Nielsen, Giovanni Petris, Len Thomas

First published: 14 June 2021 | https://doi.org/10.1002/ecm.1470

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

REVIEW | Open Access

State-space models for ecological time-series data: Practical model-fitting

Ken Newman, Ruth King, Víctor Elvira, Perry de Valpine, Rachel S. McCrea, Byron J. T. Morgan

First published: 21 February 2022 | https://doi.org/10.1111/2041-210X.13833

ECOLOGY LETTERS

Review and Synthesis | Full Access

From noise to knowledge: how randomness generates novel phenomena and reveals information

Carl Boettiger

First published: 22 May 2018 | https://doi.org/10.1111/ele.13085

ECOLOGY LETTERS

Review and Synthesis | Open Access

Uncovering ecological state dynamics with hidden Markov models

Brett T. McClintock, Roland Langrock, Olivier Gimenez, Emmanuelle Cam, David L. Borchers, Richard Glennie, Toby A. Patterson

First published: 19 October 2020 | https://doi.org/10.1111/ele.13610

ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

ARTICLE | Open Access

Confronting population models with experimental microcosm data: from trajectory matching to state-space models

Benjamin Rosenbaum, Emanuel A. Fronhofer

First published: 23 April 2023 | https://doi.org/10.1002/ecs2.4503

Modeling / Simulation

Our toy model for this session

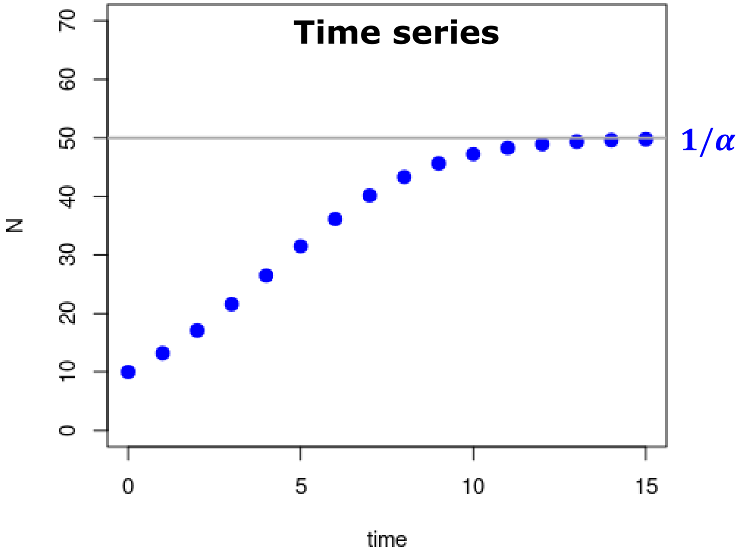
The Ricker model

- Discrete-time logistic growth for a single population

$$N_{t+1} = N_t * e^{r*(1-\alpha N_t)}$$ states

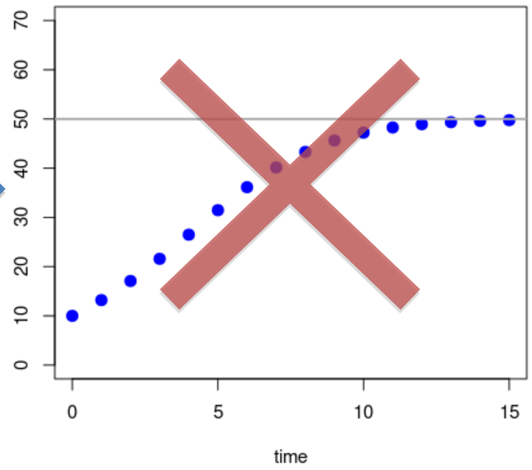
$\left.\begin{array}{l} r \text{ growth rate} \\ \alpha \text{ competition} \end{array}\right\}$ parameters

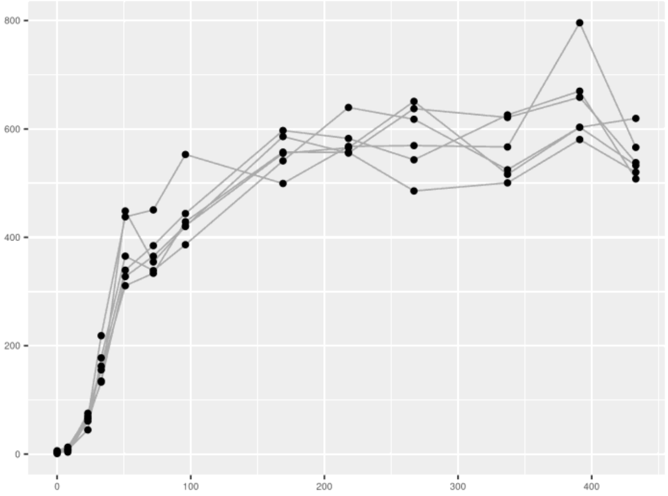
- Multispecies model extensions available
- Other examples: Gompertz model, Beverton-Holt model



Reality check

But time series from the lab or from the field don‘t look like that (deterministic, smooth)



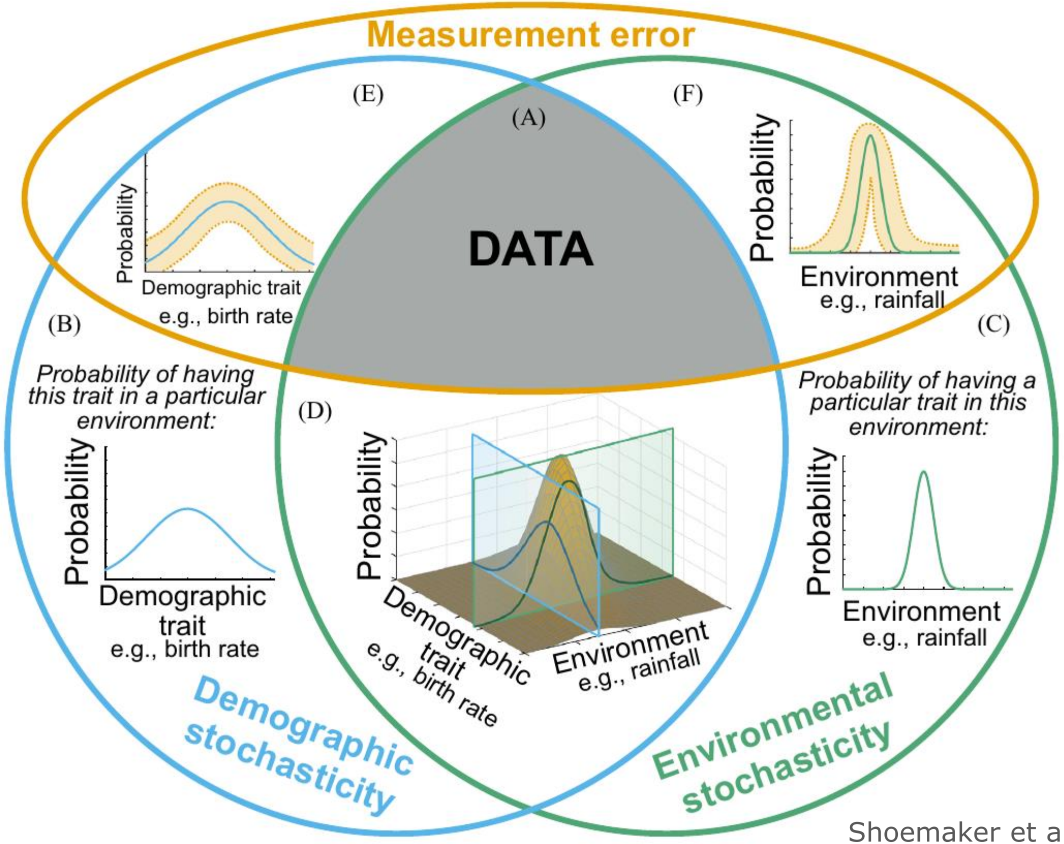


Rosenbaum & Fronhofer 2022

These are microcosm time-series.
Data from the field is even more messy.
Each single time series noisy,
variation among replicates

Sources of variability

Process error

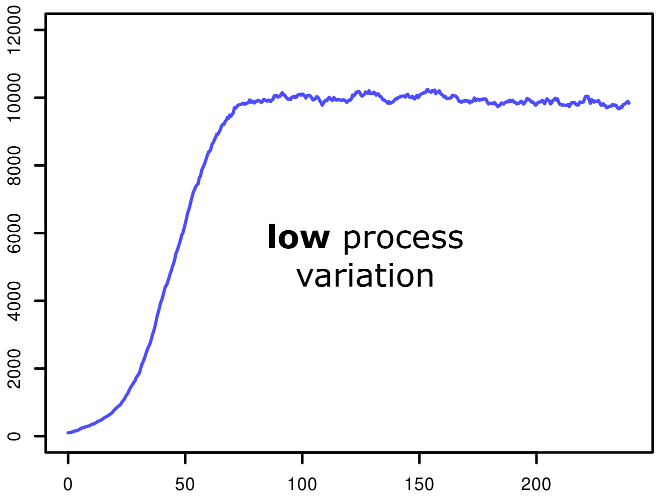


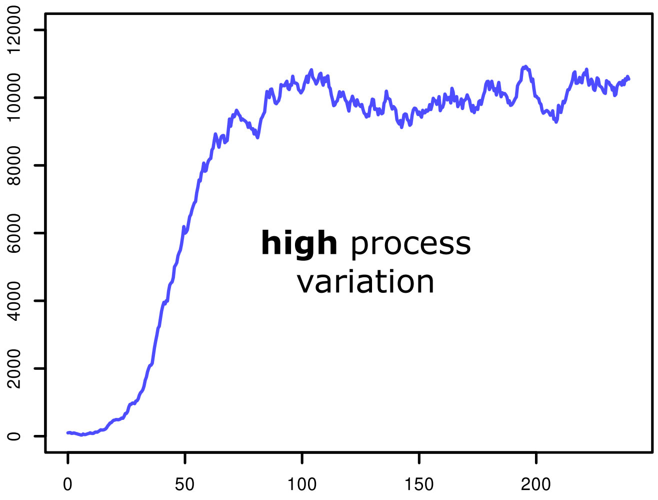
Shoemaker et al. 2020 Ecology

Process error

- Environmental stochasticity, e.g. temperature
- Demographic stochasticity: births and deaths random events

Variation in N_t affects later states N_s $(s > t)$





Modeling process error

- Here piecewise deterministic model
  with random variation in each discrete timestep

- Random variation in growth rates, e.g. by **environmental variation**

$$N_{t+1} = N_t * e^{r*(1-\alpha N_t)+\boldsymbol{\epsilon_t}}$$
$$= N_t * e^{r*(1-\alpha N_t)} * \boldsymbol{e^{\epsilon_t}}$$

- independent, or temporally autocorrelated process errors

$$\epsilon_t \sim \mathrm{Normal}(0, \sigma_{\mathrm{proc}})$$

→ Variance in N_t scales with $\sigma_{\mathrm{proc}}^2 N^2$

Modeling process error

- Or: Random birth and death events, **demographic variation**

$$N_{t+1} \sim \text{Poisson}\left(N_t * e^{r*(1-\alpha N_t)}\right)$$

→ Variance in N_t scales with N

- Or: environmental and demographic variation combined

$$\begin{aligned} N_{t+1} &\sim \text{Poisson}\left(N_t * e^{r*(1-\alpha N_t)+\epsilon_t}\right), \\ \epsilon_t &\sim \text{Normal}\left(0, \sigma_{\text{proc}}\right) \end{aligned}$$

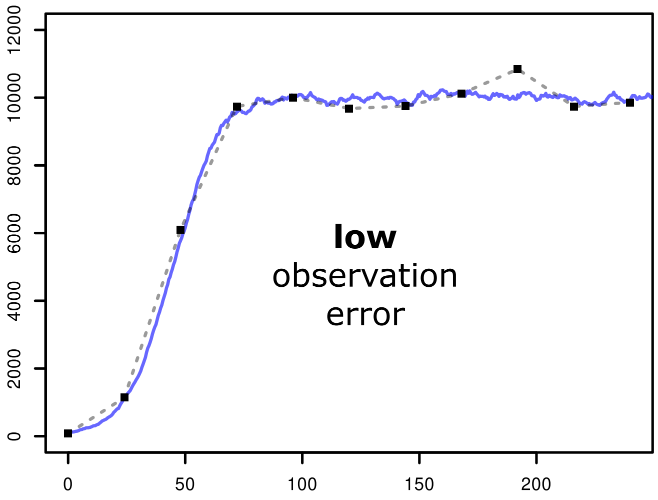
→ Variance in N_t scales with $N + \sigma_{\text{proc}}^2 N^2$

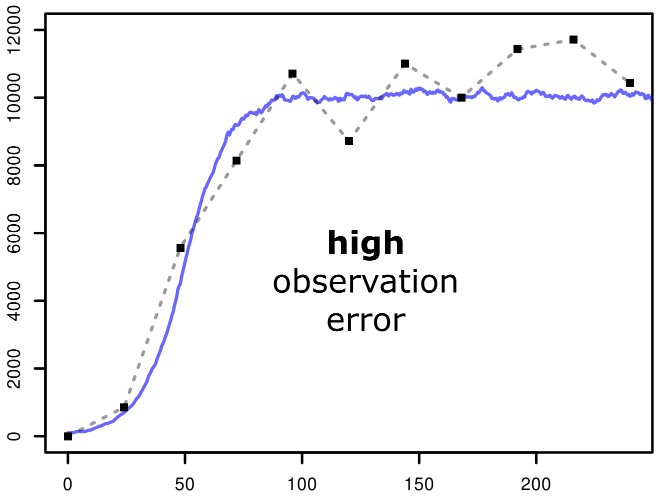
(Shoemaker et al. 2020 Ecology, Boettiger 2018 ELE)

Observation error

- Imprecise measurements Y_t of true abundances N_t
- Incomplete sampling, e.g. abundances counted in fraction p of total area

Observation error in Y_t independent from error in Y_s (at different times s, t)





Modeling observation error

Observe abundance N_{t_i} in times $t_1, \ldots, t_{\text{total}}$

Add independent errors, e.g. $Y_{t_i} \sim \text{Normal}\left(N_{t_i}, \sigma_{\text{obs}}\right)$

Be aware of variance scaling (both for process and observation error modeling)

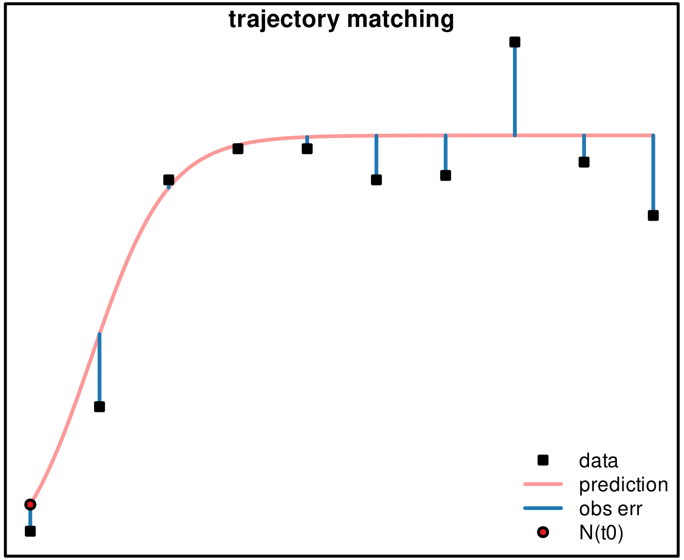
Error distribution	**Variance scaling**
$Y_t \sim \text{Normal}(N_t, \sigma)$	independent of N
$Y_t \sim \text{logNormal}(\log(N_t)\,, \sigma)$	with N^2
$Y_t \sim \text{Poisson}(N_t)$	with N
$Y_t \sim \text{quasi} - \text{Poisson}(N_t, \tau)$	with $N + \frac{N^2}{\tau}$
$Y_t \sim \text{Binomial}(N_t, p)$	with $N \cdot p$

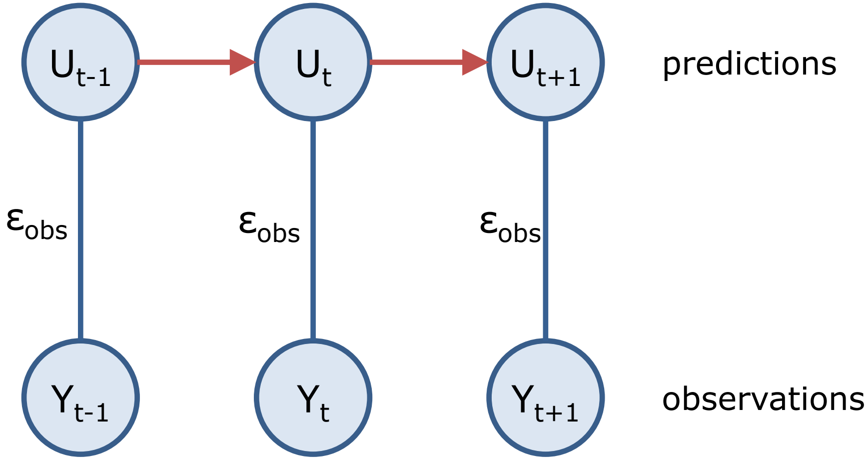
Fitting / Statistical modeling

(1) Fitting: observation error only

Neglect process error in data → Model is completely deterministic

Find „best" parameters for **deterministic** prediction model





(1) Fitting: observation error only

Neglect process error in data → Model is completely deterministic

Find „best" parameters for **deterministic** prediction model

Predictions are computed from previous predictions:

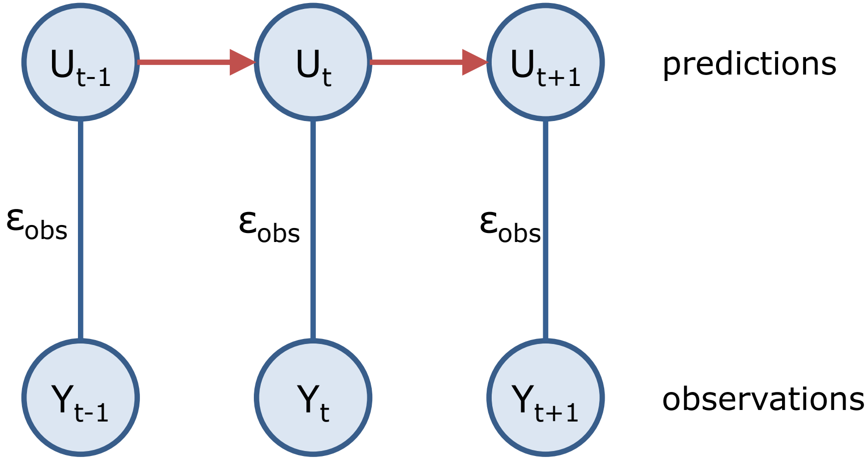
$$U_t = f(U_{t-1}, \theta)$$

Data has observation error only

$$Y_t \sim \mathrm{Normal}(U_t, \sigma_{\mathrm{obs}})$$

→ Residuals are independent

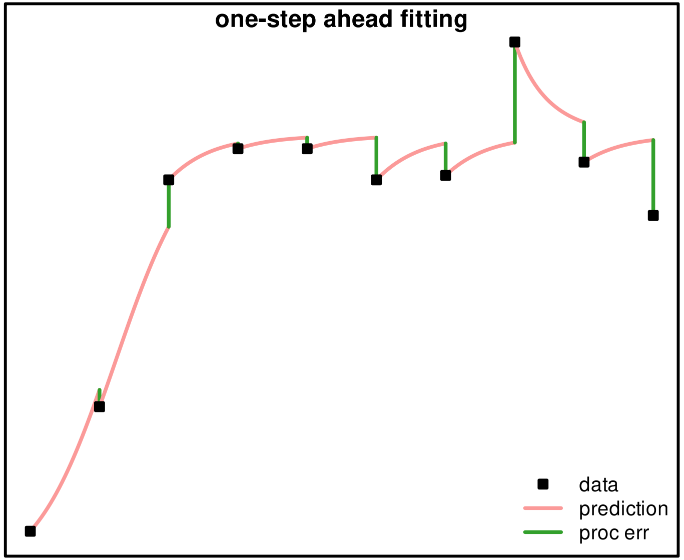
→ Nonlinear regression problem

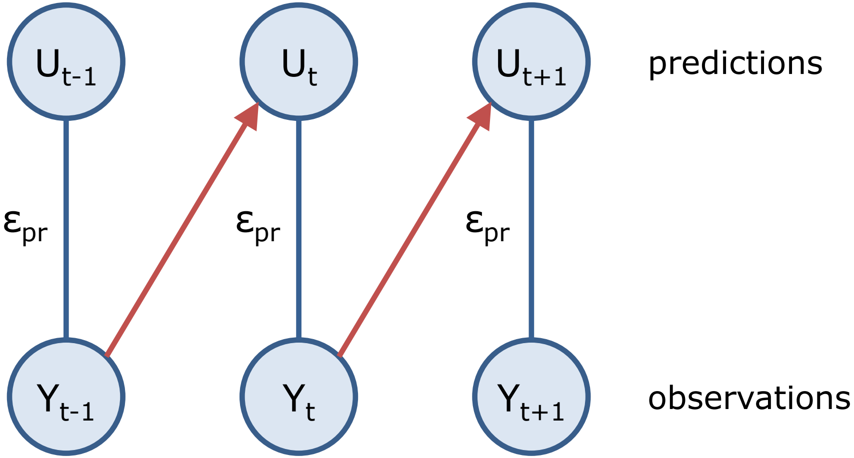


(2) Fitting: process error only

Neglect observation error in data → Observations = true abundances

Find „best“ parameters for **piecewise** prediction model





(2) Fitting: process error only

Neglect observation error in data → Observations = true abundances

Find „best" parameters for **piecewise** prediction model

Predictions are computed from previous observations:

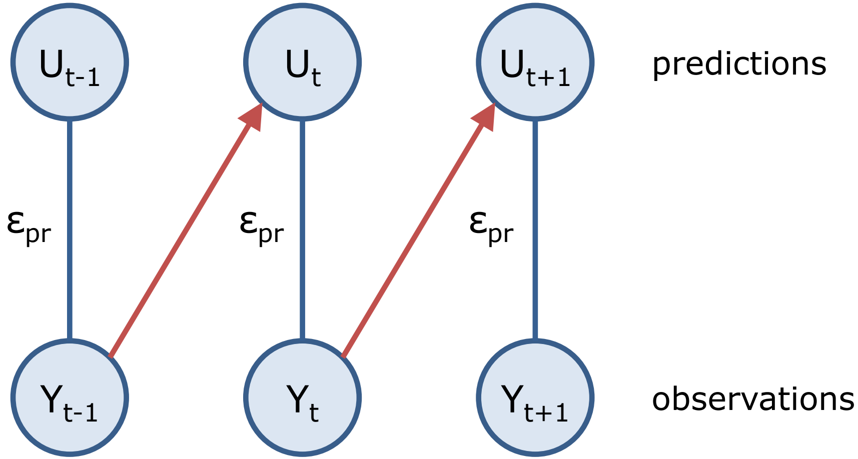
$$U_t = f(Y_{t-1}, \theta)$$

Data has process error only

$$Y_t \sim \text{Normal}(U_t, \sigma_{\text{pr}})$$

→ Residuals are independent

→ Nonlinear autoregressive problem



(2) Fitting: process error only

Some (discrete-time) models can be **linearized**

$$N_{t+1} = N_t * e^{r*(1-\alpha N_t)}$$

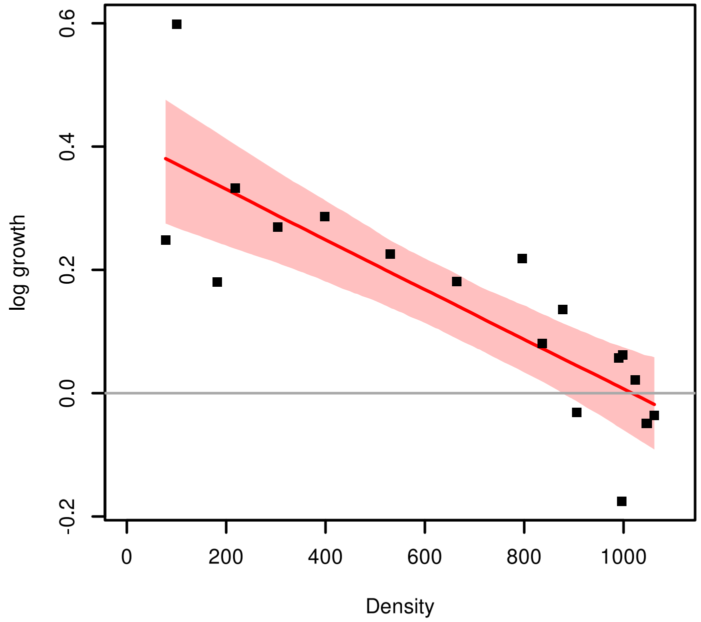
(Ricker model)

$$\log(N_{t+1}) = \log(N_t) + r * (1 - \alpha N_t)$$

$$\log\left(\frac{N_{t+1}}{N_t}\right) = r * (1 - \alpha N_t)$$

$$\log\left(\frac{N_{t+1}}{N_t}\right) = \boldsymbol{r} - \frac{\boldsymbol{\alpha}}{\boldsymbol{r}} N_t$$

intercept **slope**

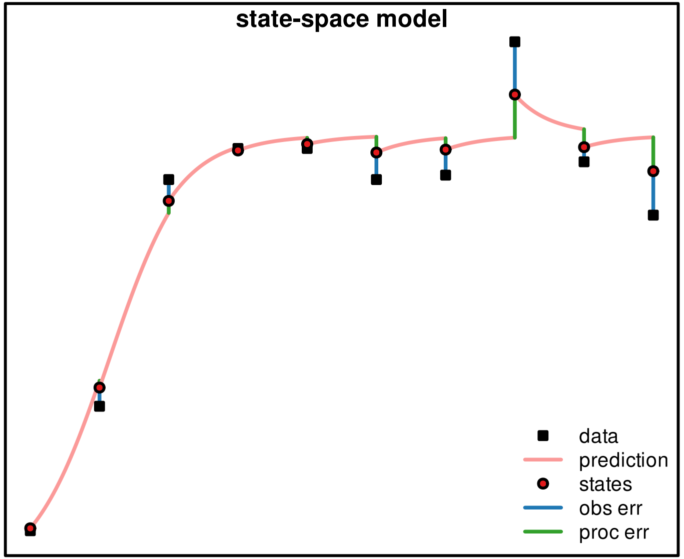


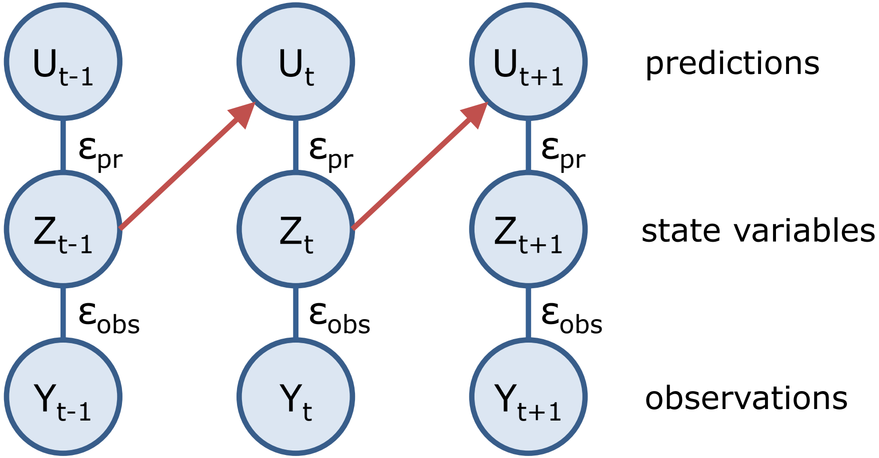
→ Can be solved with standard `lm( growth ~ N )`

(3) Fitting: state-space models

Account for both errors → model unknown true states as separate time series

Find „best" **parameters and states** for piecewise prediction model





(3) Fitting: state-space models

Account for both errors → model unknown true states as separate time series

Find „best" **parameters and states** for piecewise prediction model

Predictions are computed from previous states:

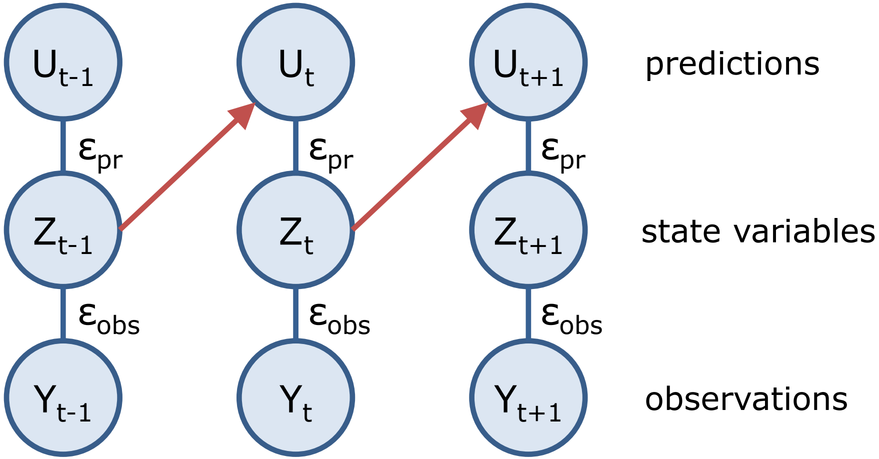
$$U_t = f(Z_{t-1}, \theta)$$

States time series with process error

$$Z_t \sim \mathrm{Normal}(U_t, \sigma_{\mathrm{pr}})$$

Observed time series with obs. error

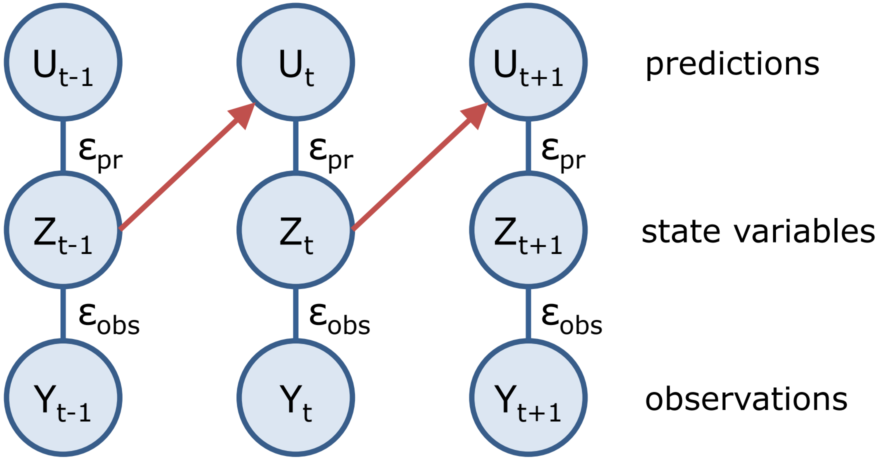
$$Y_t \sim \mathrm{Normal}(Z_t, \sigma_{\mathrm{obs}})$$



→ Autocorrelation of residuals is accounted for

(3) Fitting: state-space models

- All ecological time series feature proc and obs error!
- But SSM fitting can be quite complex
  (coding, runtime, number of parameters)
- Special case: „Kalman filter"
  linear model and Gaussian errors
  direct solution exists
- R-packages for specific problems:
  MARSS, pomp, TMB
  moveHMM, bsam (movement ecology)
- Bayesian methods:
  JAGS, Stan, Nimble



Want to learn Bayesian statistics?

Introduction to Bayesian statistics in R & brms

Check my course materials (including video recordings):
https://github.com/benjamin-rosenbaum/bayesian-intro

Curriculum

	Lecture	Practical
(1) Statistical modeling	pdf	pdf html Rcode
(2) Bayesian principles	pdf	pdf html Rcode
(3) Priors and posteriors	pdf	pdf html Rcode
(4) Linear models	pdf	pdf html Rcode
(5) Generalized linear models	pdf	pdf html Rcode
(6) Mixed effects models	pdf	pdf html Rcode
(7) Stan introduction	pdf	pdf html Rcode
(8) Conclusions	pdf	pdf html Rcode

Take-home messages

- Many options available
  deterministic prediction model + stochastic model
- Stochastic models correspond to noise in the data
  observation / process / both
- Highly controlled lab data (microcosm)
  proc-only model performed worse
  (Rosenbaum & Fronhofer 2023)
- Field data: usually too noisy for obs-only model
- Think of variance scaling (N or N^2)
- Some models can be linearized (Ricker, Gompertz)
  and quick solutions exist (LM, MARSS)
- Nonlinear model fitting with MCMC
  (Stan, Nimble)

R coding ...