State-space models

Christopher L. Cahill

Goals for 27 November 2019

1. Questions or discussion from last time
2. Learn what state-space models are, why they can be useful, why we might use them, and where they can go wrong
3. Simulate fake time-series data, then estimate using GAM and TMB
4. Assignment for the brave

Credit: Andrew Muir

State space models: what are they and why?

- A state-space model basically does the following:
  - Given some observations y_t in each year t, let's reconstruct an unobserved (i.e., latent) process of interest x_t
  - Remember, what you see is not always what you want to get
  - Can think of these models as "filters"
  - Original derivation quite mathematical, but can reframe this important class of problems in terms of mixed-effects models
- Why do people use state-space models:
  - Best predictor for unobserved state
  - Allows you to explicitly separate process and measurement error, at least in theory
  - One can often interpolate or extrapolate beyond the measurements
  - Not accounting for these errors can have implications for both basic and applied work
- Approximately 7 trillion resources on time-series modeling

Rudolf Emil Kálmán

De Valpine and Hastings 2002; Royle and Dorazio 2008; Kery and Schaub 2011

State space models: one of the most useful statistical tools of the 20th century?

- Nonlinear trajectory estimation (Apollo program)
- Rocketry (guided missiles)
- Guidance systems (aircraft, cars, boats)
- Dynamic positioning systems (feedback control loops)
- Econometrics (time-series)
- Ecology (time-series, density-dependence)
- Control theory (various engineering applications)
- Adaptive management of renewable resources
- Dynamics of uncertainty
- Decision analysis (making wise decisions given error)
- Markovian processes (i.e., autoregressive)
- Basically any situation where there is observation error that one wants to filter out of the problem

Rudolf Emil Kálmán

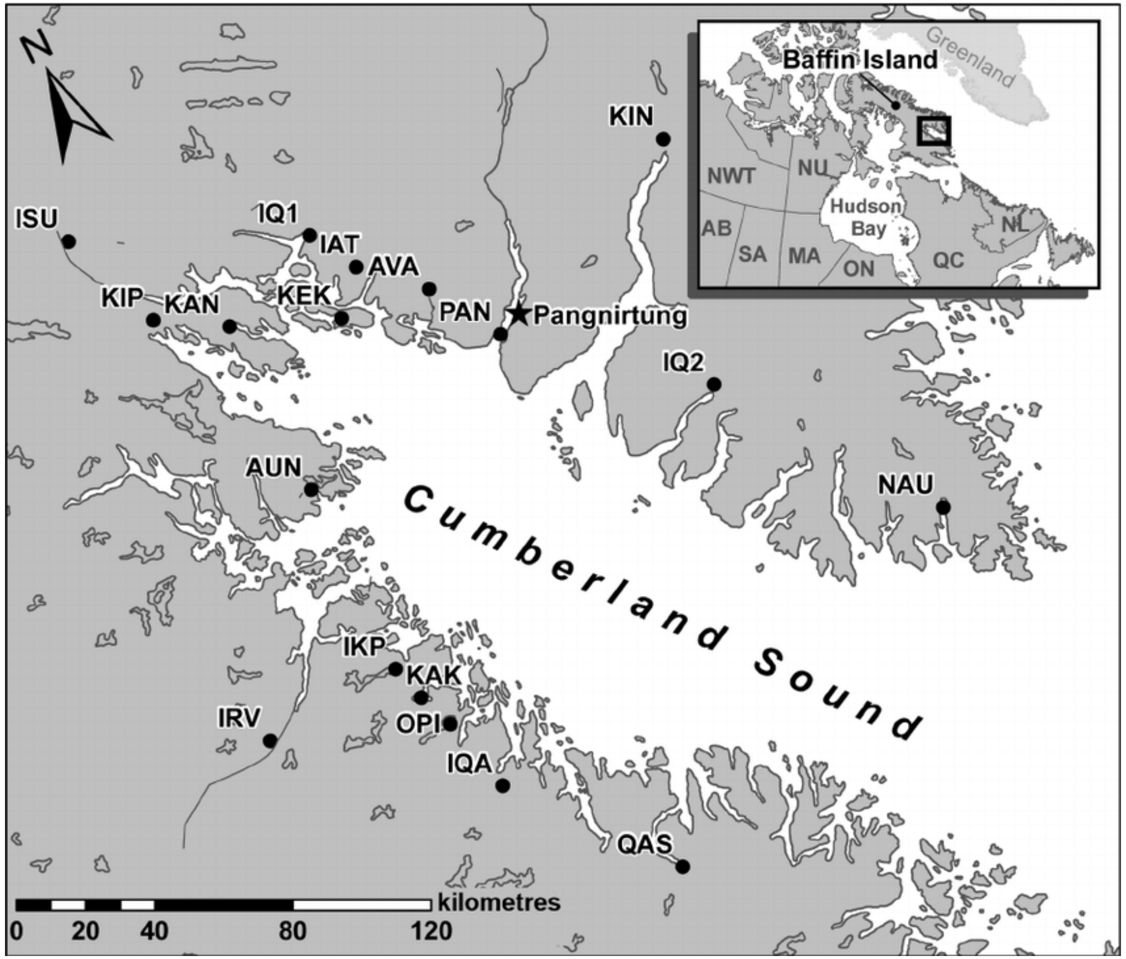
Kalman 1960; Walters 1986; De Valpine and Hastings 2002; Royle and Dorazio 2008; Kery and Schaub 2011

A common problem

- What is the probability distribution for critter counts in 2020 given you've sampled during 1990-2019?
  - Can be really hard (or even impossible) to write these joint probability statements out
- However, we can factor a complex joint probability distribution into a series of simpler pieces (remember the law of total probability)
  - We often want to use the joint distribution for inference, but will use the simpler conditional specification to get there
- In terms of the models we will be dealing with, this means fixing or estimating the starting conditions, and then "sweeping downstream" through the remainder of the time-series
- Basically we will condition on the initial conditions, i.e., reconstruct the rest of the time-series given these conditions

Let's build a (simple) mechanistic ecological model from first principles, and then use this to motivate a state-space model

- A (semi) made-up example to show how some of the concepts from previous lessons start to blend together
- Beluga *Delphinapterus leucas*in Cumberland Sound, Nunavut have been reduced to extremely low abundance following extensive harvest
  - i.e., population is likely far from carrying capacity
- Let's say the data you have come from aerial counts where whale people guess the number and size of the whales, and this gives you a time-series of biomass estimates
- You seek to build a model that explores Beluga population dynamics—in particular you seek to explore how population growth rate varies through time
  - These values are often used to parameterize population viability analyses
- Let's build a Berliner-style hierarchical model



Map from Moore et al. 2013

Exponential growth model

- The simplest population dynamics model specifies population growth rate as a function of population size (we'll assume biomass here):

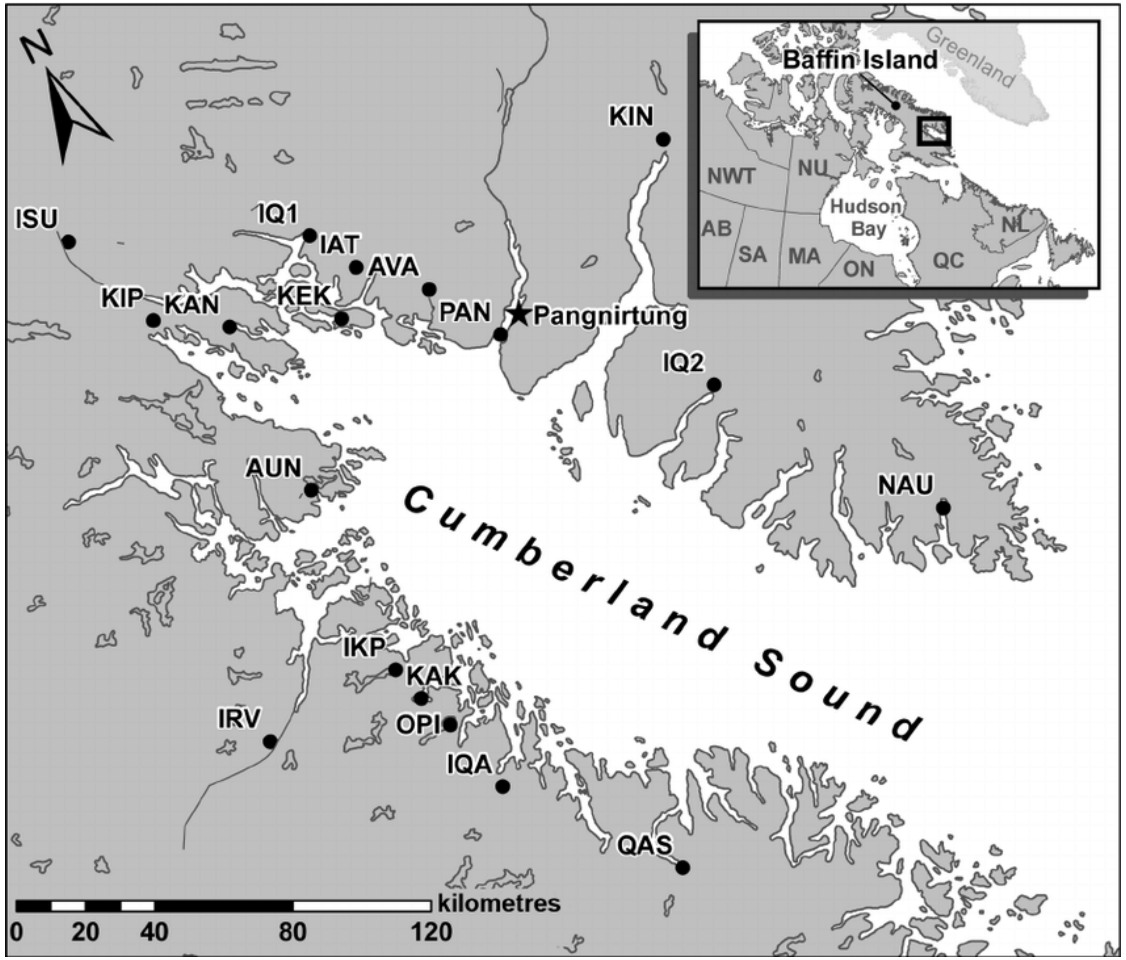
$$\frac{db}{dt} = rb$$

- What does this thing mean?
- What is r?
- We have annual data so need to discretize this and turn it into a difference equation to fit the data we have:

$$\frac{db}{dt} \approx \Delta b_t = b_{t+1} - b_t =$$

$$b_{t+1} = \lambda_t b_t$$

- While simple, this is an autoregressive (AR-1) nonlinear model
- What is λ_t?
- These discrete-time parameters are analogous to the mechanistic differential equation parameter interpretation (cool)



Map from Moore et al. 2013

Let's hierarchicalize this exponential growth model

Model the true process:

$$b_{t+1} = \lambda_t b_t$$

$$\lambda_t \sim N(\bar{\lambda}, \sigma^2_{process})$$

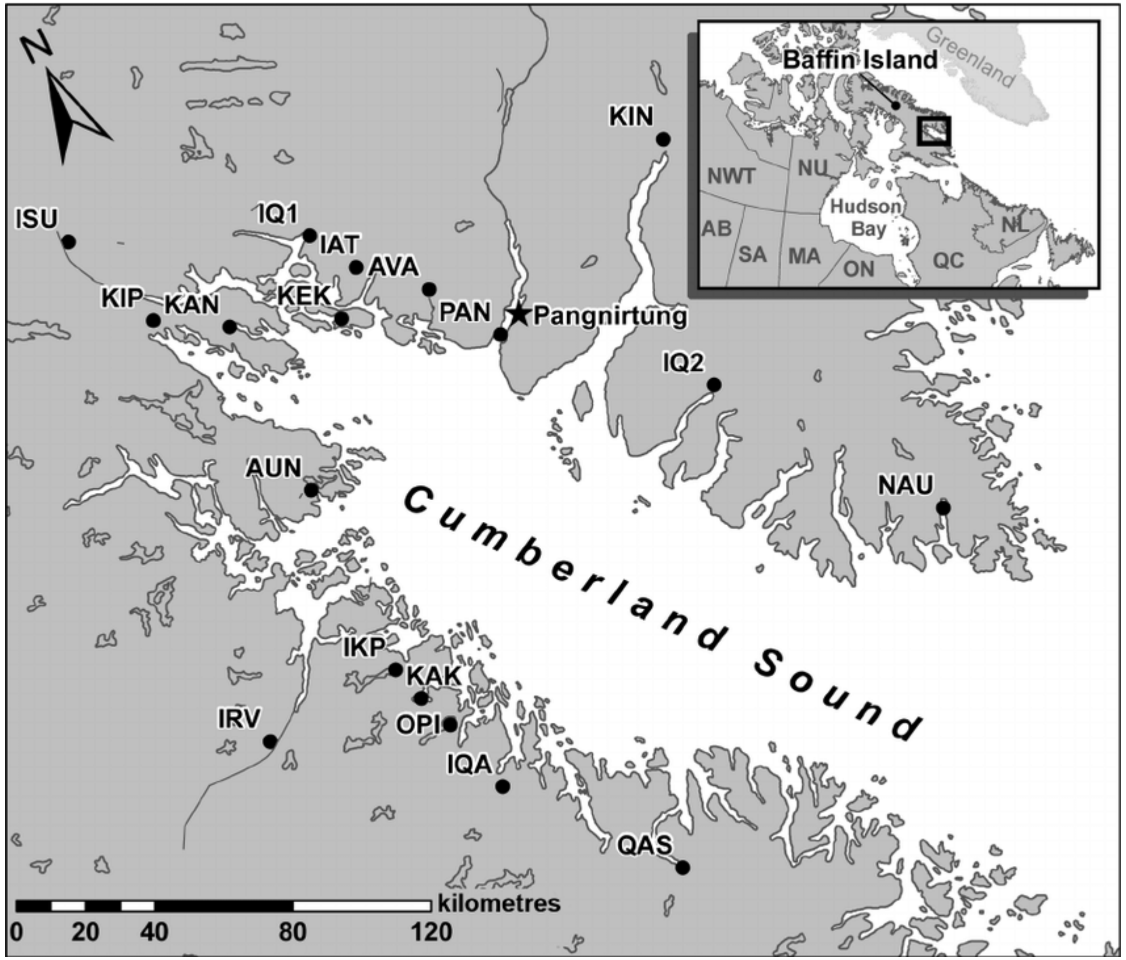
- Need to define the initial state of the model somehow--b_0

Map the true process to the observed data (i.e., observed data | true process):

$$y_{obs_t} = b_t + \epsilon_t$$

$$\epsilon_t \sim N(0, \sigma^2_{observation})$$

- What are the parameters? What are the data? What are the random effects? Why do we need to specify the initial conditions?
- Think "condition upstream and sweep downstream"



Map from Moore et al. 2013

Important things to keep in mind

- State-space models are still rarely used in ecology, and there are some things you should be aware of if you use them
- Can provide biased estimates, particularly if there is systematic bias in the observation process.
    - Say you always have underestimated/overestimated observations. Probably an underappreciated aspect of these models, especially for ecologists.
- Can have estimation problems--see Auger-Methe et al. 2014 “State space models’ dirty little secrets”
    - Can (should) profile parameters to make sure things are behaving.
    - When in doubt, use simulation to help safeguard against nefarious model behaviour

Auger-Methe et al. 2014; Kery and Schaub 2011