Using Gaussian Process Models to Fit an Enhanced Particle Tracking Model to Acoustic Telemetry Data of Juvenile Salmon

Russell W. Perry, Adam C. Pope,
Doug Jackson, and Vamsi Sridharan

U.S. Department of the Interior
U.S. Geological Survey

Bay-Delta Science Conference
15-17 November 2016

Coupled physical-biological models

- Physical models
  - Spatially explicit 1D – 3D hydrodynamics
- Biological models
  - Add “fishy” behaviors to neutrally buoyant particles
    - E.g., swimming velocity, holding during day
- Conduct simulation experiments
  - Water management actions
  - Patterns from fish behaviors and hydrodynamics

How Do We Determine Values of Behavioral Parameters?

- Theory
  - Hypotheses about fish behavior

- Trial and Error
  - “Pattern matching” to observed data

- Problems:
  - No uncertainty in parameter estimates
  - Somewhat subjective

Goals

- Develop methods to fit models to observed data
- Methods should be general
  - Applicable to any model
- Provide parameter estimates + uncertainty
- Allow assessment of different model structures

USGS

Challenges

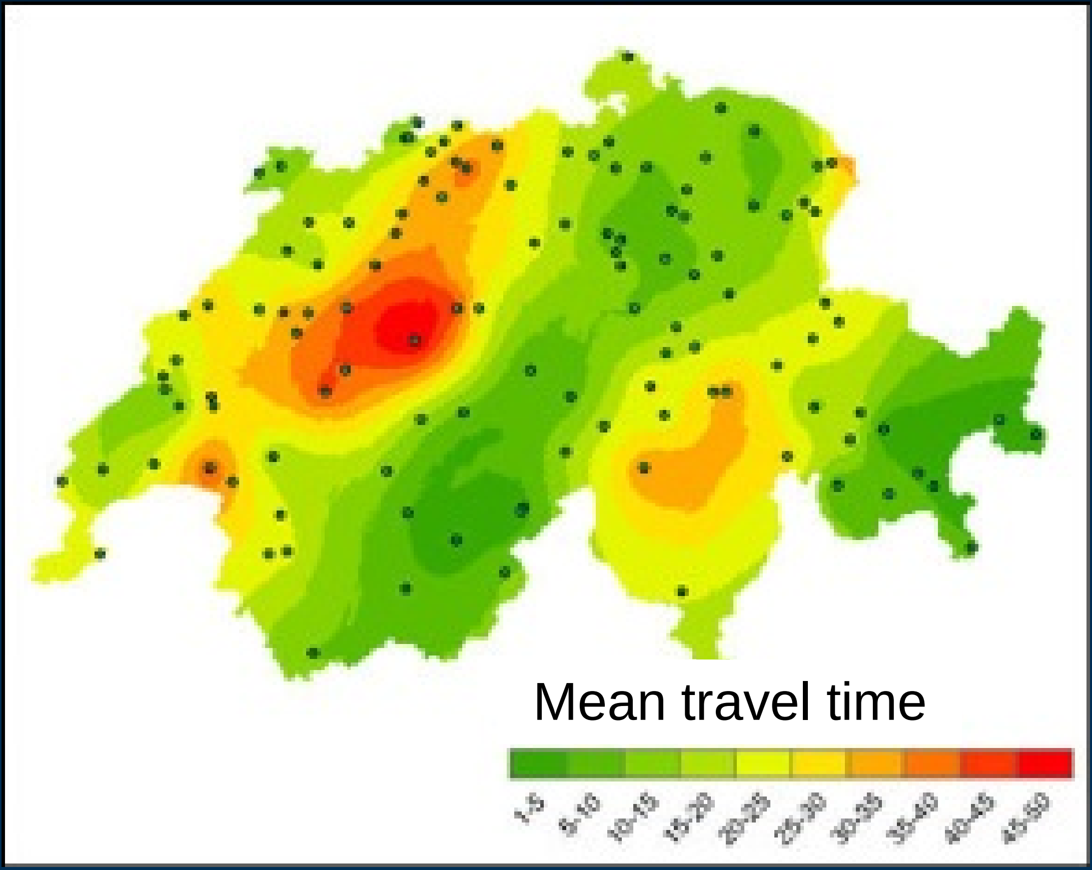
- Models are computationally burdensome
  - Traditional optimization routines take too long
- Models are stochastic
  - Direct search methods won’t work
  - Traditional stochastic methods take too long
- Two potential solutions
  - Gaussian process model + MCMC (this talk)
  - Particle Swarm Optimization (next talk)

USGS

Gaussian Process Models

- Distance-weighted interpolation
- Uses multivariate normal distribution
  - e.g., Kriging

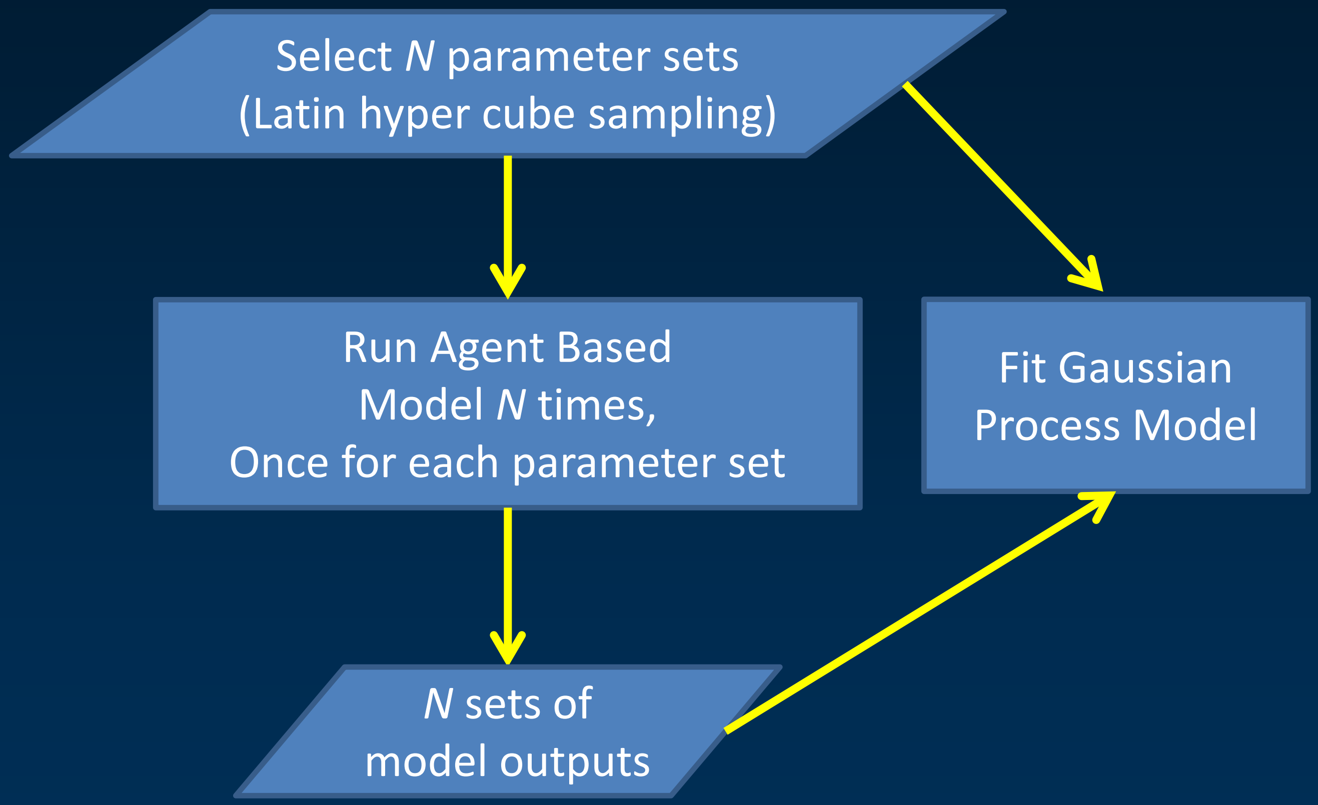
Probability of day-time migration



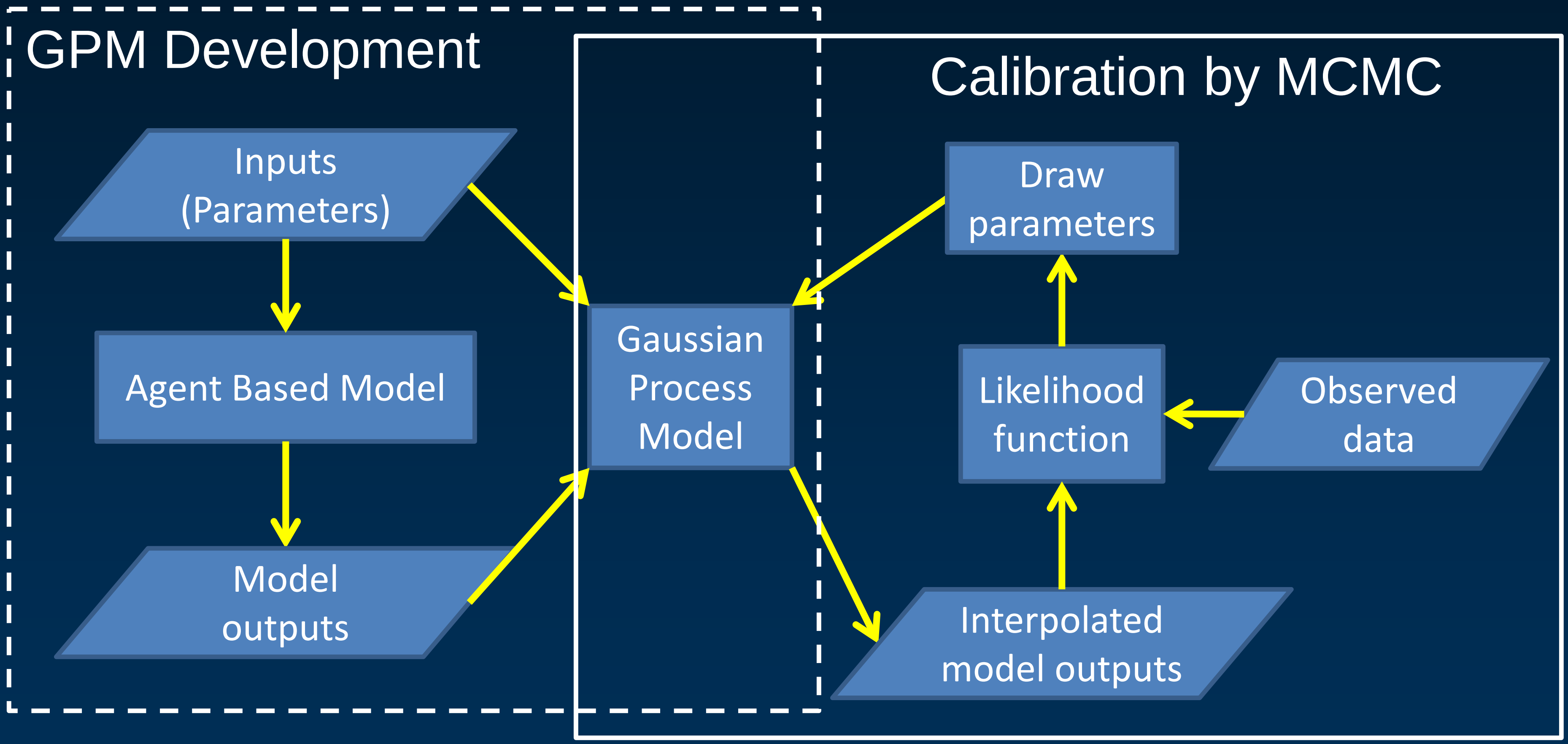
Swimming Velocity

Source: http://www.gitta.info/ContiSpatVar/en/image/kriging.jpg

Developing Gaussian Process Models (GPM)



Calibration using Gaussian Process Models (GPM)



Application to DSM2-ePTM

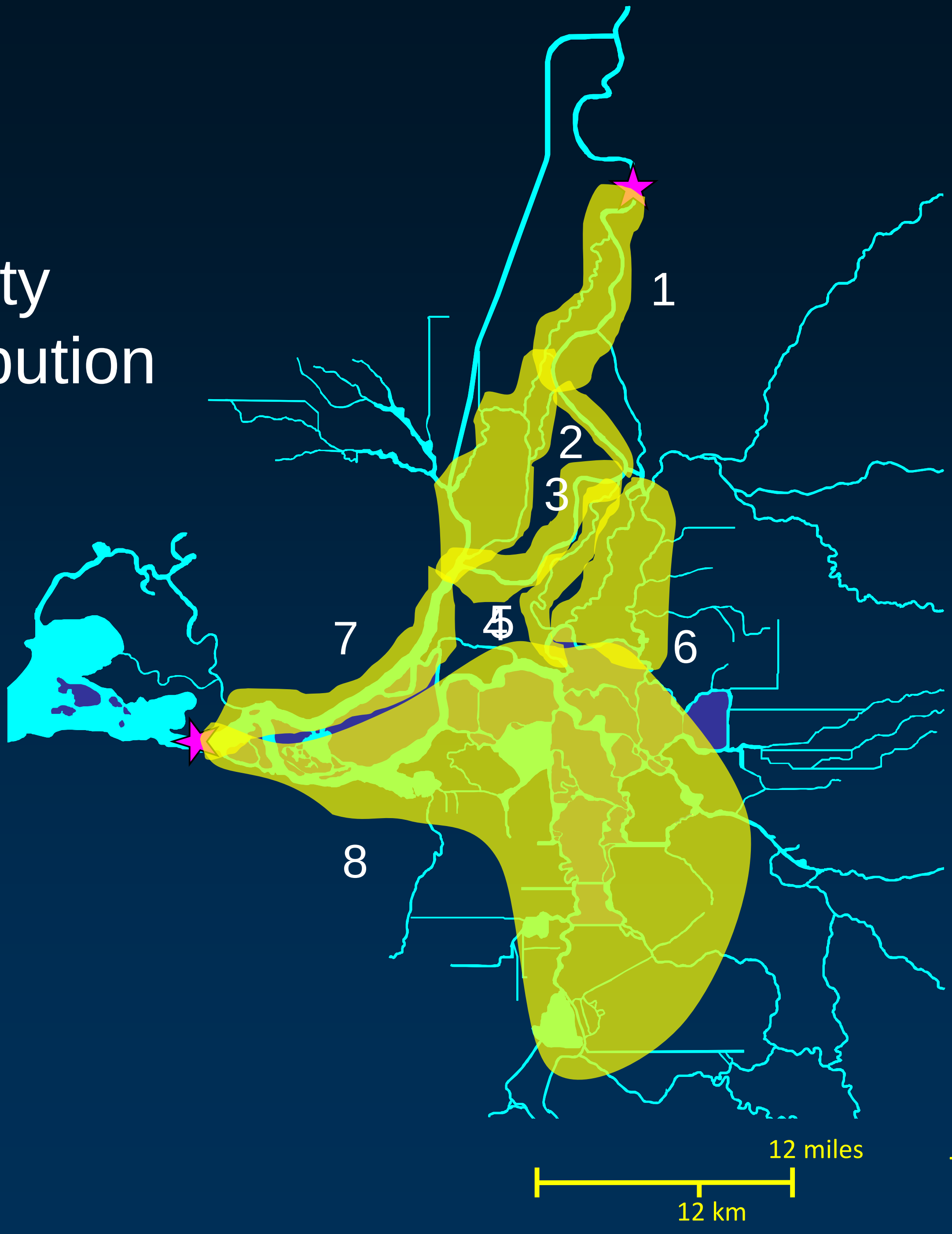
- 7 parameters (per reach)
- Swimming behaviors
  - Swimming velocity (mean + SD)
  - Daytime holding probability
  - Velocity holding threshold
    - Selective tidal stream transport
  - Probability of mis-assessing downstream direction
    - Function of mean velocity relative to SD velocity
- XT Survival model (Anderson et al. 2005)
  - λ, mean distance between predator-prey encounters
  - ω, random encounter velocity

Acoustic Telemetry Data

- USFWS (Delta Action 8 study)
- Late-fall Chinook salmon
- Vemco acoustic telemetry
- 1,583 Acoustic tagged fish
- 4 Years (2007 – 2010)
- 8 unique release groups
- 9 reaches
- Migrated between December and February

Reaches

For each reach:
- Survival probability
- Travel time distribution



USGS

Gaussian Process Model

- 2000 parameter sets
  - – Run for each reach and release
  - – 144,000 ePTM model runs!
    - Ran in parallel on Amazon cloud

- Model outputs for each reach and release
  - – Survival probability
  - – Proportion of fish in 20 travel time bins
  - – Flexible distribution shapes (e.g., bi-modal)

USGS

Likelihood Function

- Multistate mark-recapture model
  - Perry et al. (2010)
  - Survival, detection, routing
- Multinomial distribution for travel times
  - Proportion of fish in 20 travel time bins
  - Observed number in each bin

USGS

Compare Two Fitted Models

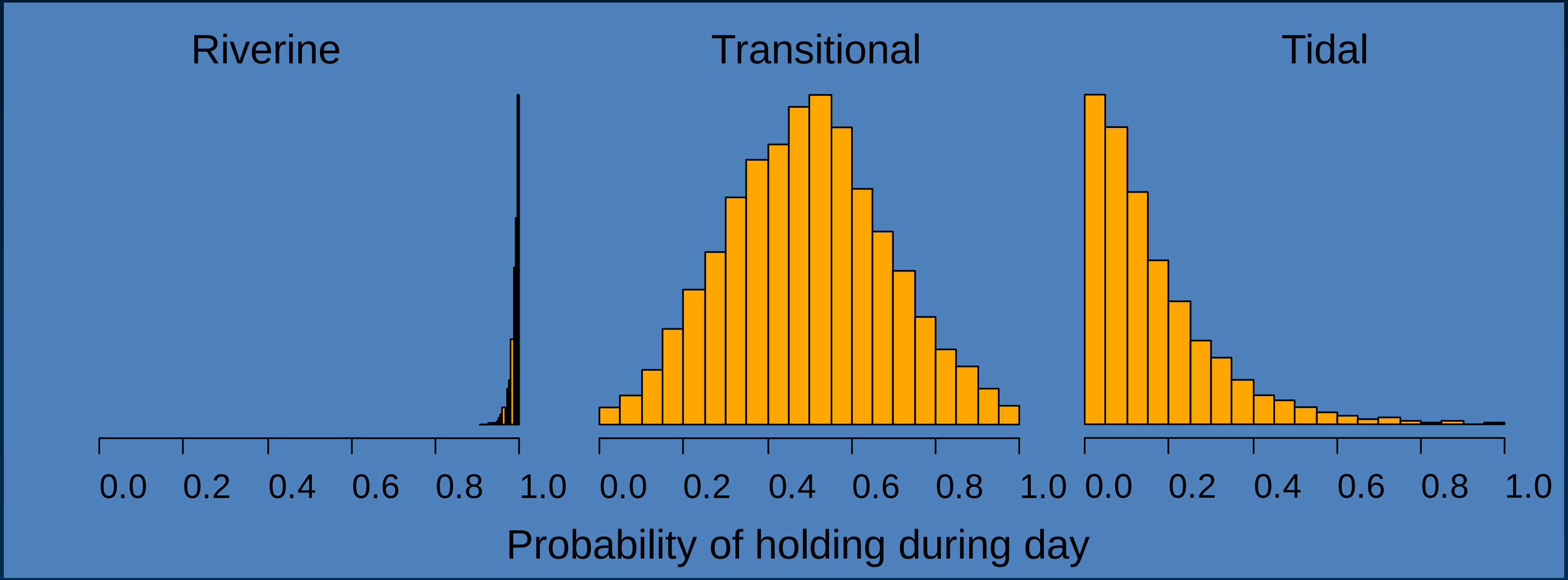
- “Simple” Model
  - Daytime swim probability and Hold Threshold
    - Turned Off
  - All other parameters set equal among reaches
- “Complex” Model
  - Hold threshold turned off
  - Daytime swim probability and ω
    - Different for riverine, transitional, and tidal reaches
  - Probability of mis-assessing direction
  - All other parameters reach-specific
- Compare using WAIC

USGS

Model Selection

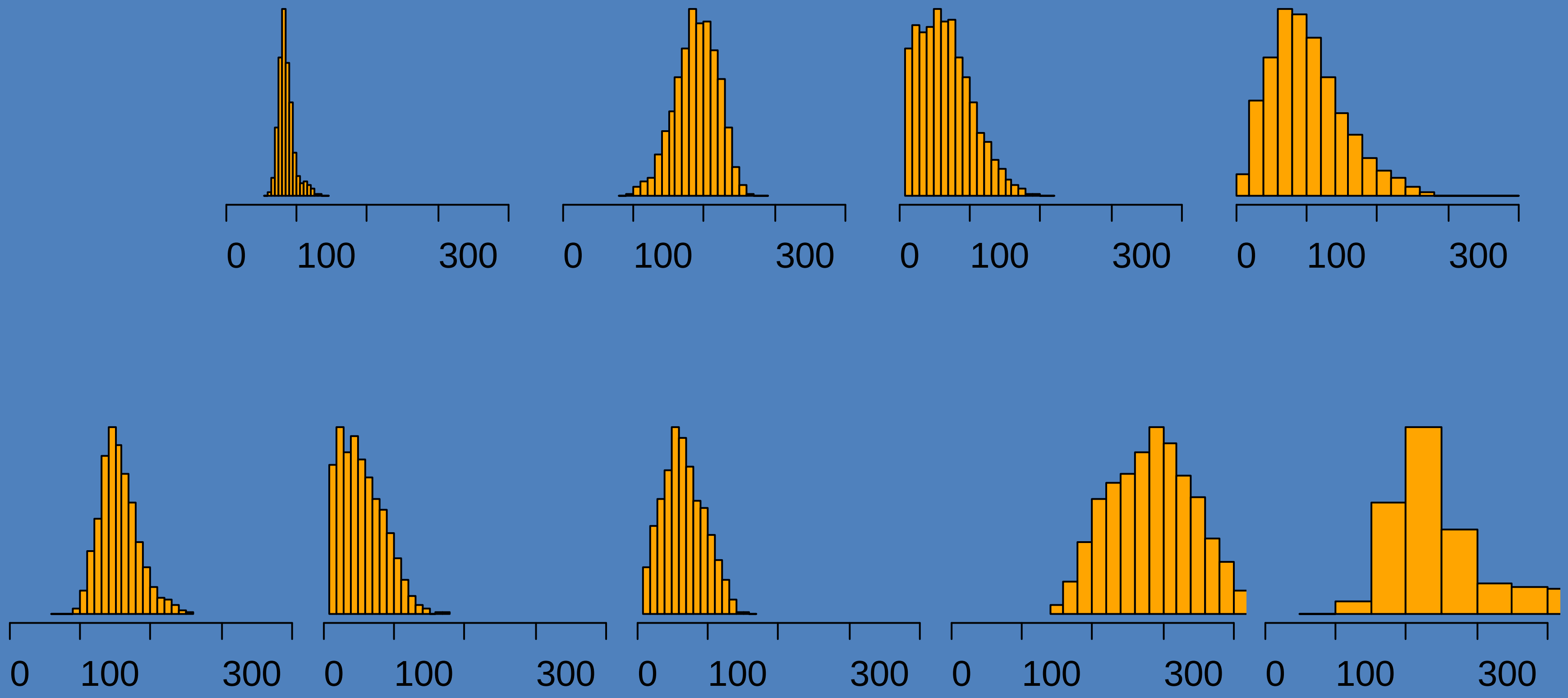
- “Simple” Model
  - 5 parameters
  - WAIC = 505,988
- “Complex” Model
  - 34 parameters
  - WAIC = 423,856
- Difference of 82,131
- Complex model is better fit

Posterior Distributions



Posterior Distributions

λ, Mean distance (km) between predator-prey encounters
Median λ: 52 – 287 km



USGS

Conclusions

- Advantages
  - Fully parametric
  - Posterior distributions of parameters
  - Full accounting of uncertainty due to:
    - PTM stochasticity
    - Error due to GPM interpolation of PTM
    - Sampling uncertainty in observed data

- Disadvantages
  - Many steps in process
  - Not “off the shelf”
  - Not using PTM directly