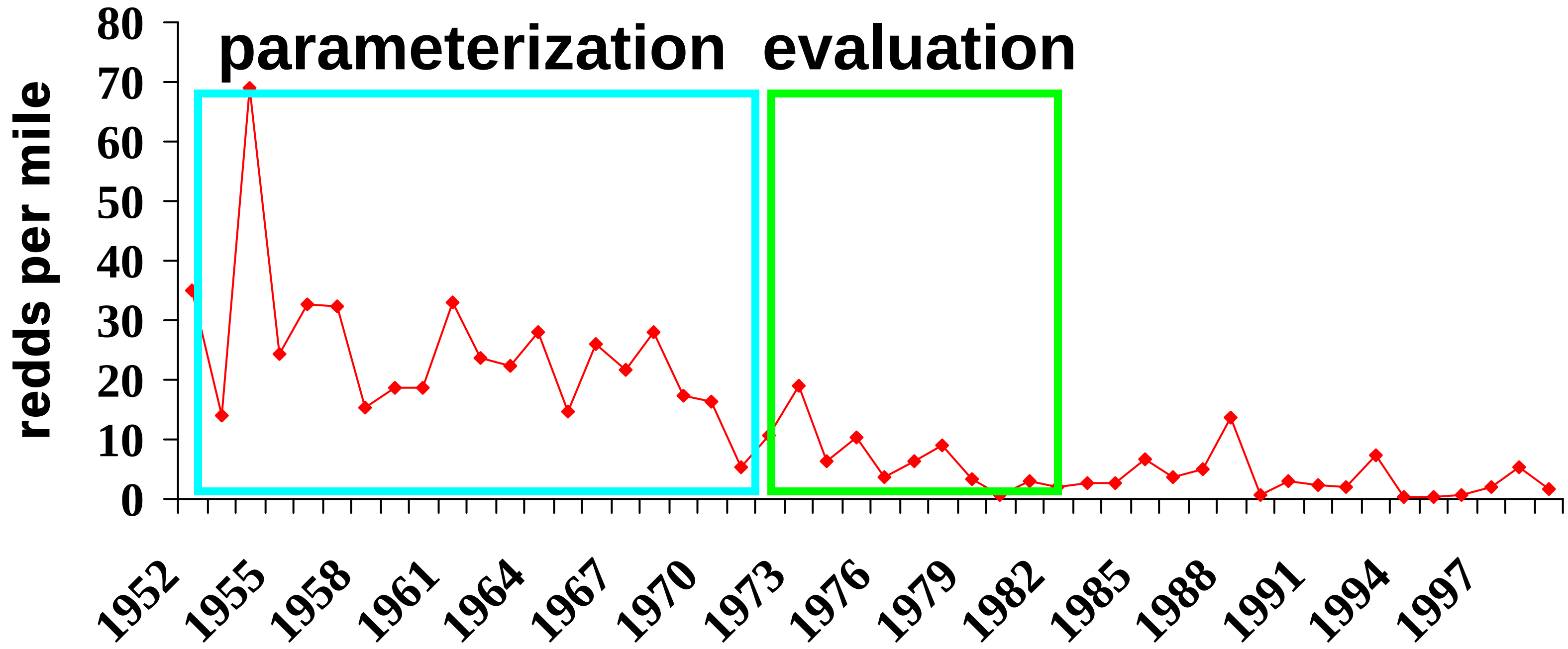
Patterns of first-passage probabilities in population monitoring data

Confronting the theory with data (Holmes & Fagan 2002)

- 141 chinook and 41 steelhead 30-70 year time series from ESUs in WA, OR, and CA



Corrupted Diffusion Approximation (CDA) really a ‘Corrupted Random Walk Model’

$$\log(N_{t+1}) = b\log(N_t) + \mu + \varepsilon_{t,p}$$

$$\log(y_{t+1}) = \log(N_{t+1}) + \varepsilon_{t+1,np}$$

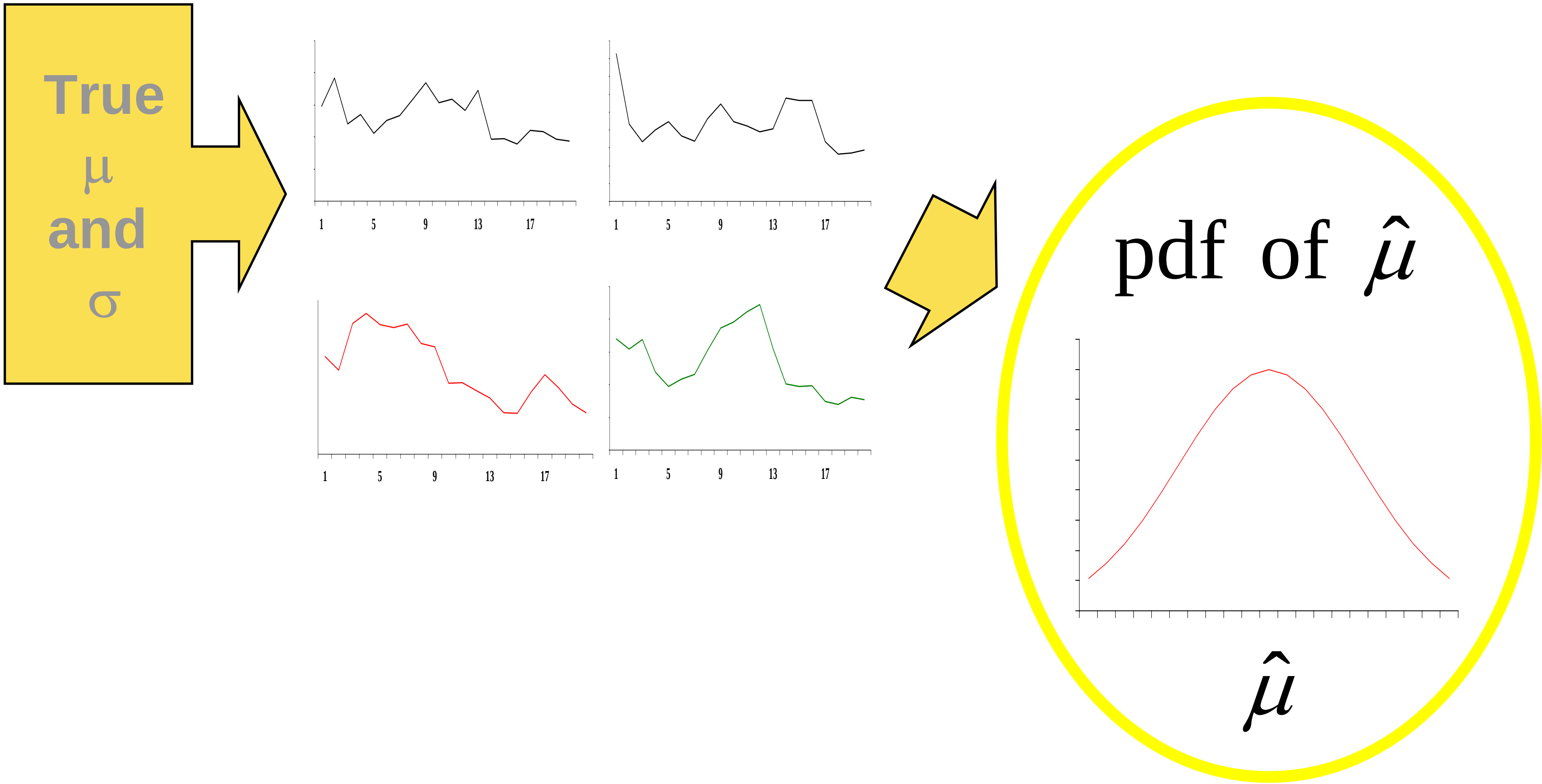
$$\varepsilon_{t,p} \sim Normal(0,\sigma_p^2)$$

$$\varepsilon_{t,np} \sim f(\beta,\sigma_a^2)$$

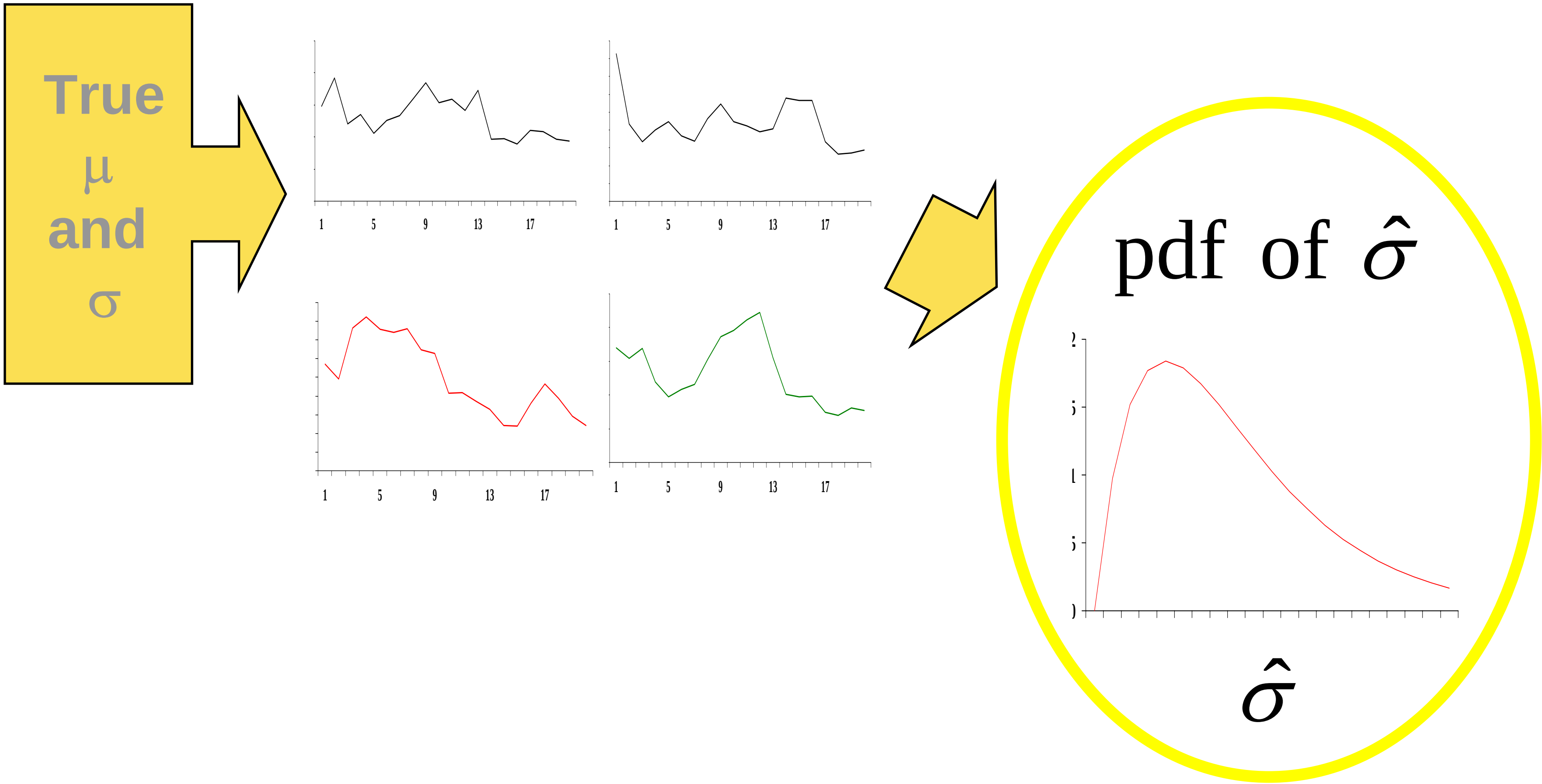
$$b = 1$$

No density-dependence
i.i.d. errors -> no auto-correlations
Holmes (2001)

Theory makes a prediction about the distribution of mu_hat
*random walk
*N_t+1/N_t variance is related in a particular way to mu_hat variance



Similarly theory makes a prediction about the specific distribution of s_hat
that we should observe



But…

- **Problem: don’t view the same population process over and over**
- **Actual data: many different processes with different underlying parameters (growth rates and variability)**
- **Solution: transform data to a standardized metric that has the same normalized statistical distribution for all processes**

Standardized μ distribution

$$\frac{(\hat{\mu}_p - \hat{\mu}_e)}{\sqrt{\frac{df_{slp}\hat{\sigma}_p^2 + df_{slp}\hat{\sigma}_e^2}{2df_{slp}}\left(\frac{1}{n_p - L} + \frac{1}{n_e - L}\right)}} \sim \frac{1}{\sqrt{\gamma}} t_{2df_{slp}} - 0$$

no trend in environment (world isn't getting worse)

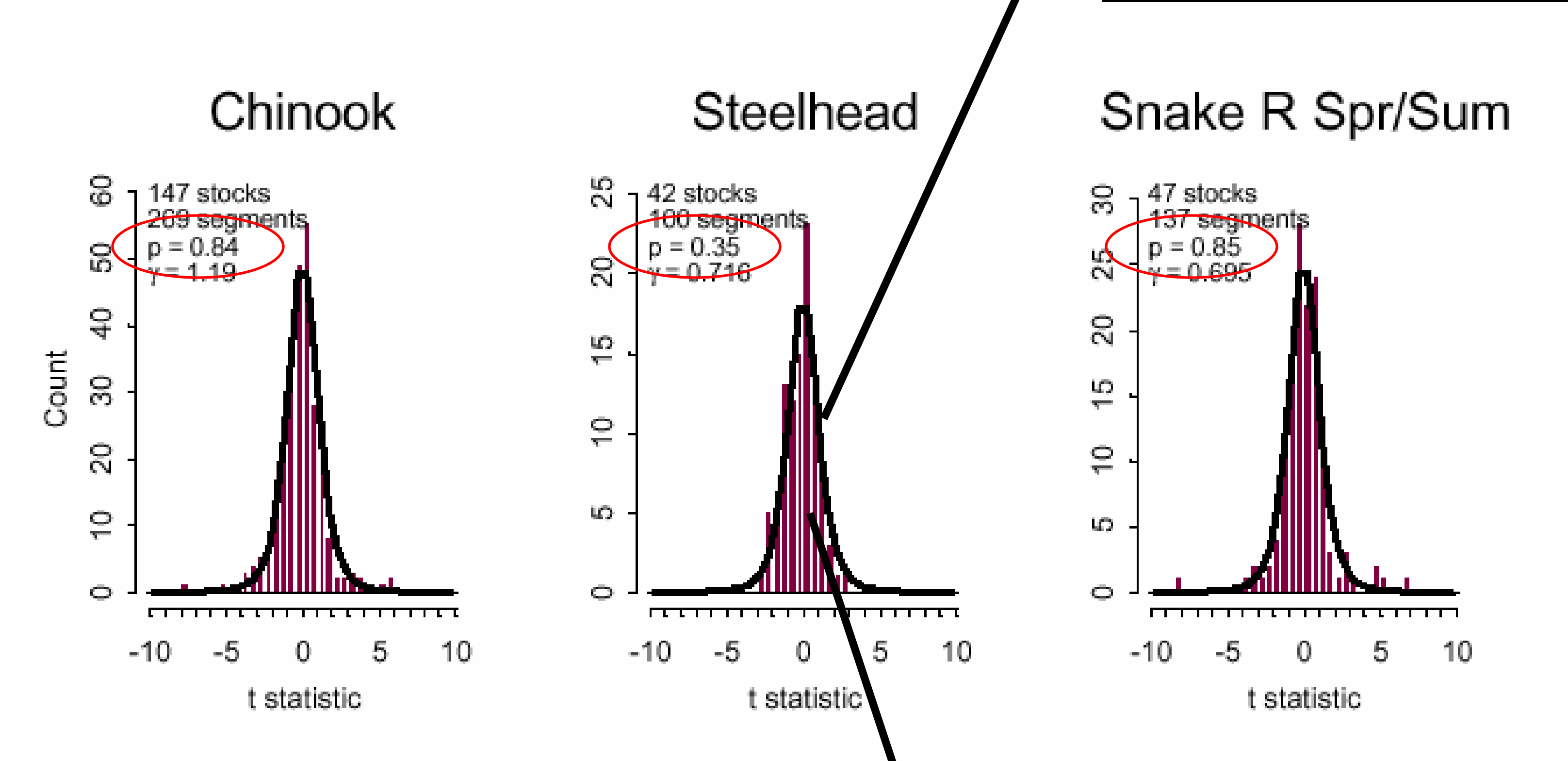
Standardized σ distribution

$$\left(\hat{\sigma}_e^2 / \hat{\sigma}_p^2\right) \sim F(df_{slp}, df_{slp}) \quad \text{X 1}$$

no trend in variance (variance is not a function of N)

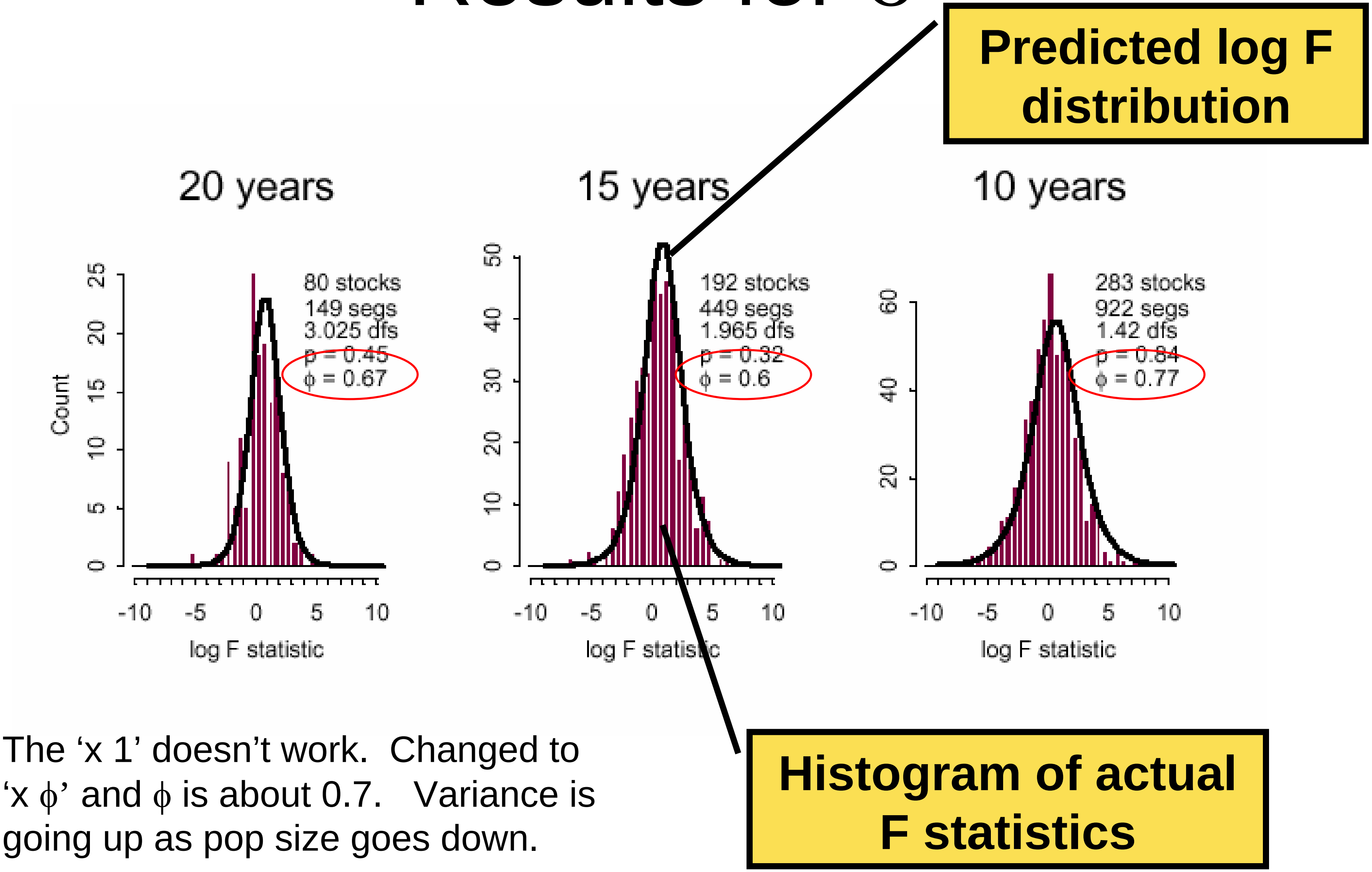
Results for μ

Predicted t distribution



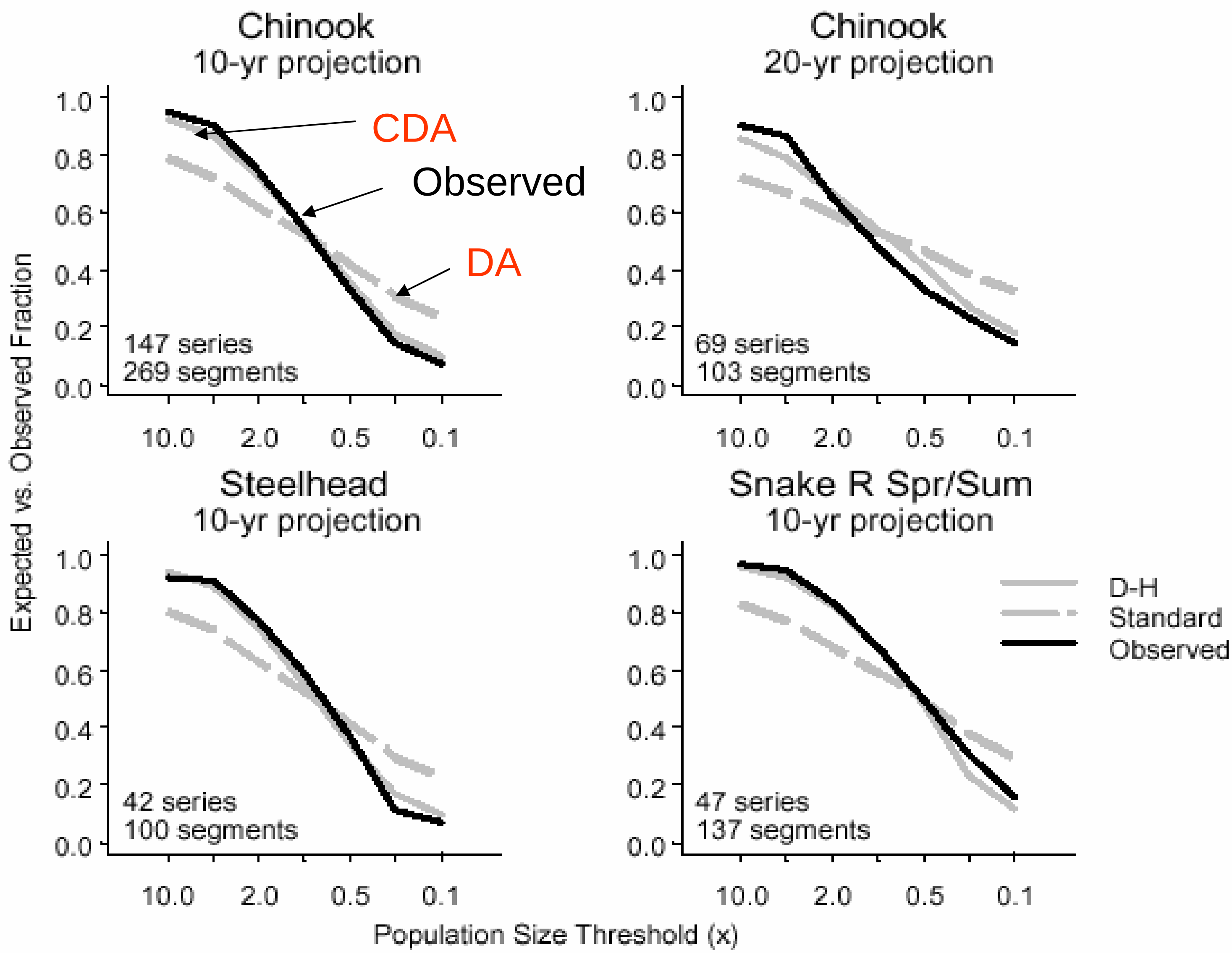
Histogram of actual t statistics

Results for σ



The 'x 1' doesn't work. Changed to 'x ϕ' and ϕ is about 0.7. Variance is going up as pop size goes down.

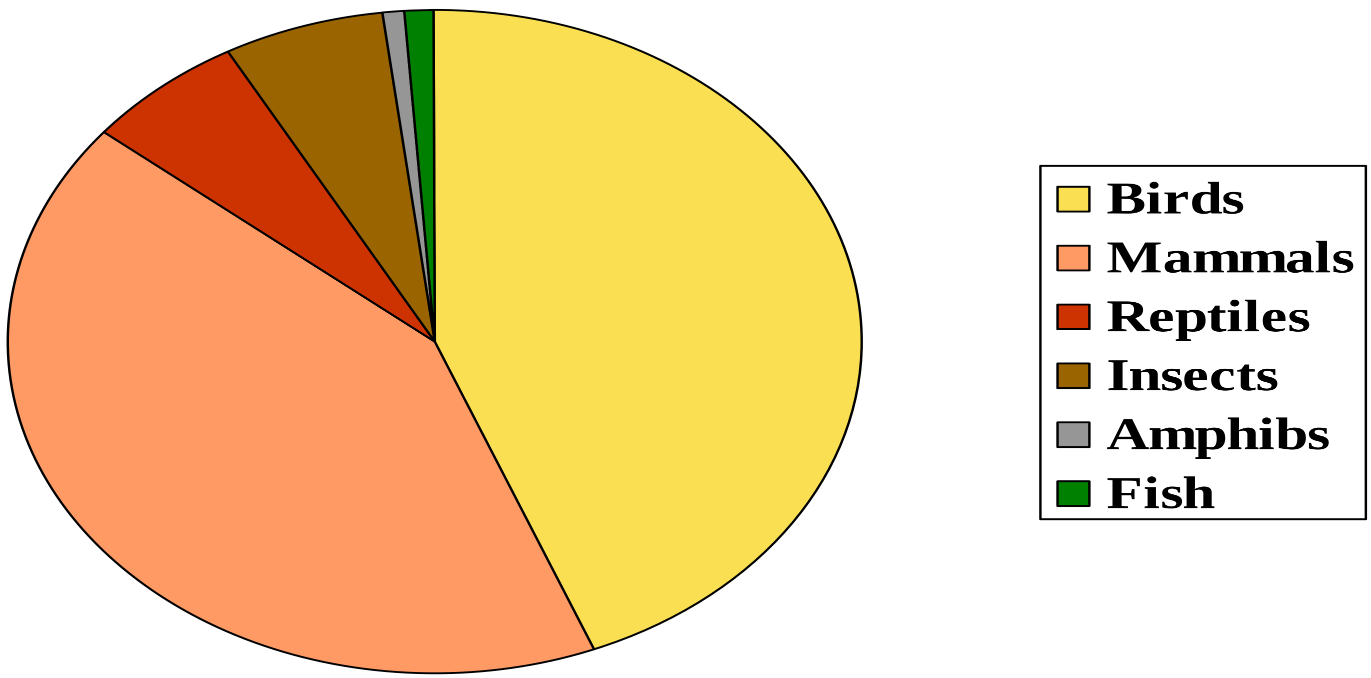
First passage patterns: predicted versus observed



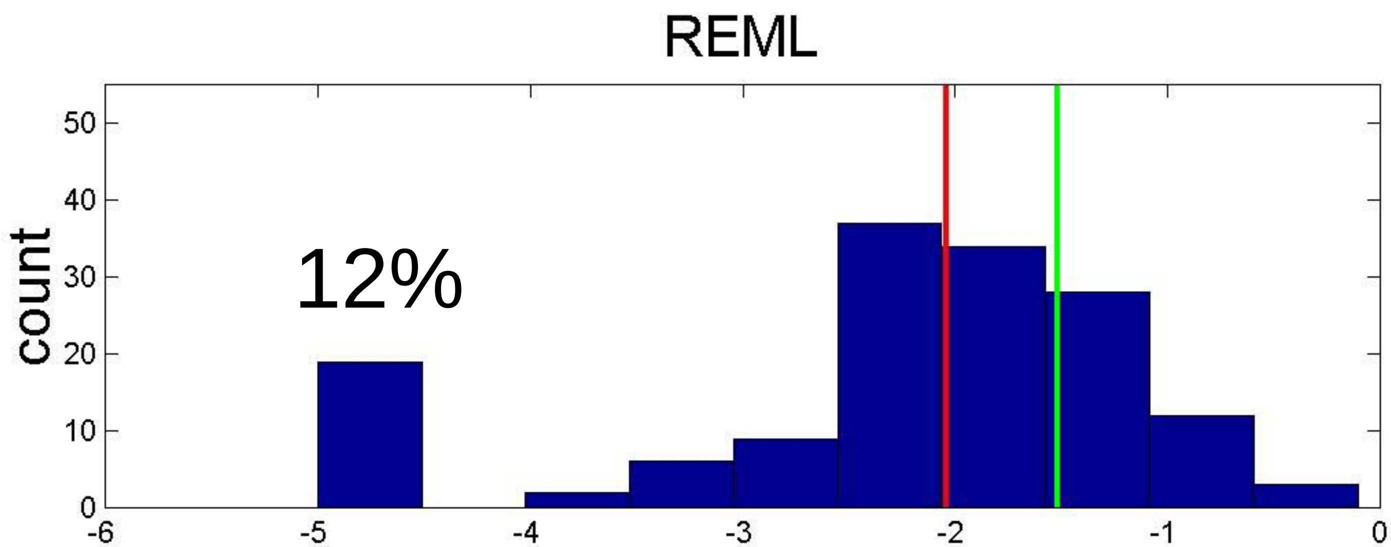
Let’s look at some different data and a different study…

This data has a lot less non-process error

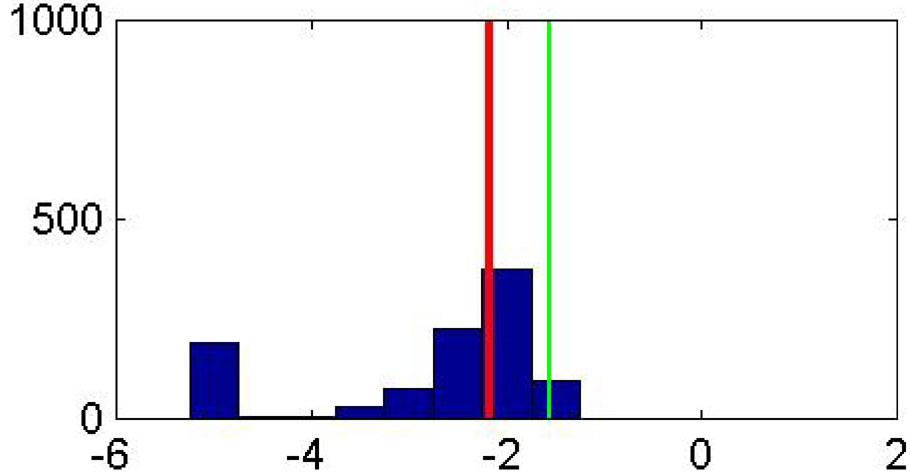
117 Time series 20-50 yrs long
72 are listed species



Distribution of process error estimates



DATA

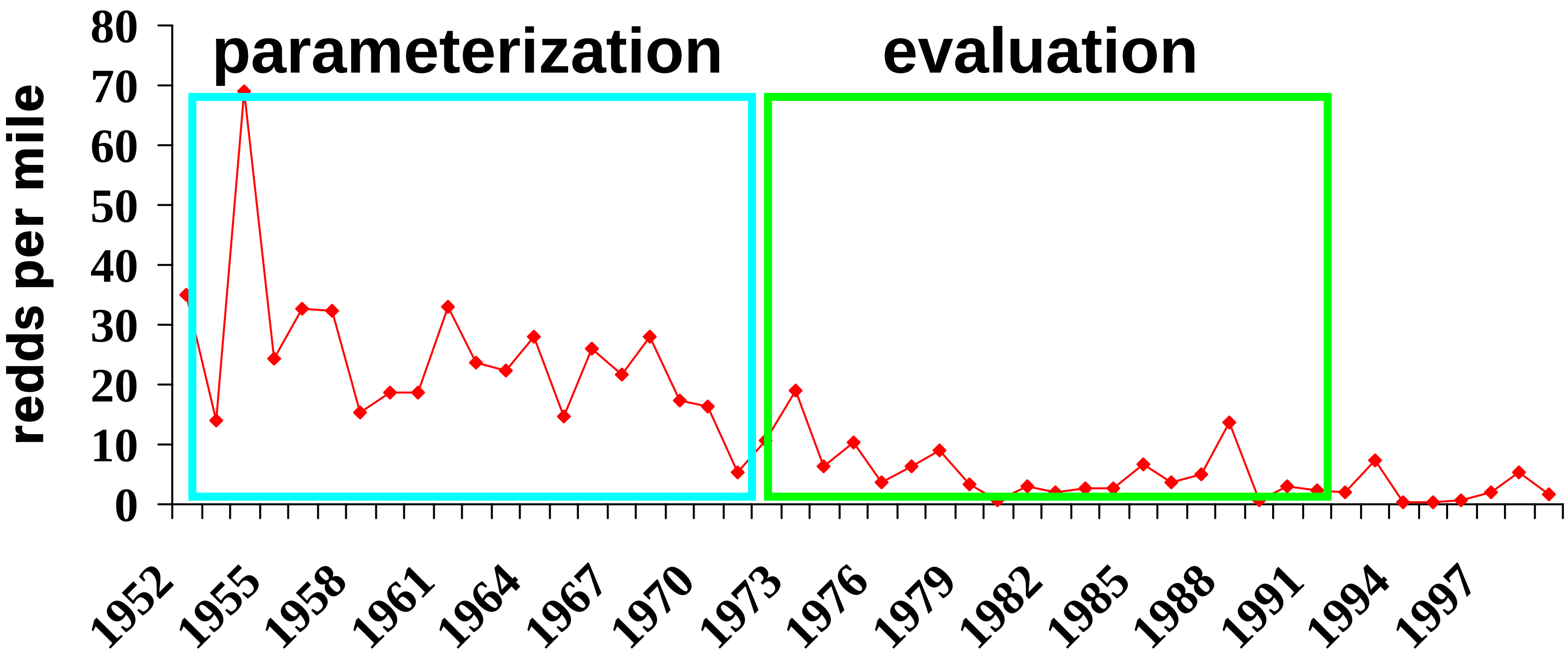


PETREL SIMS with low non-process error

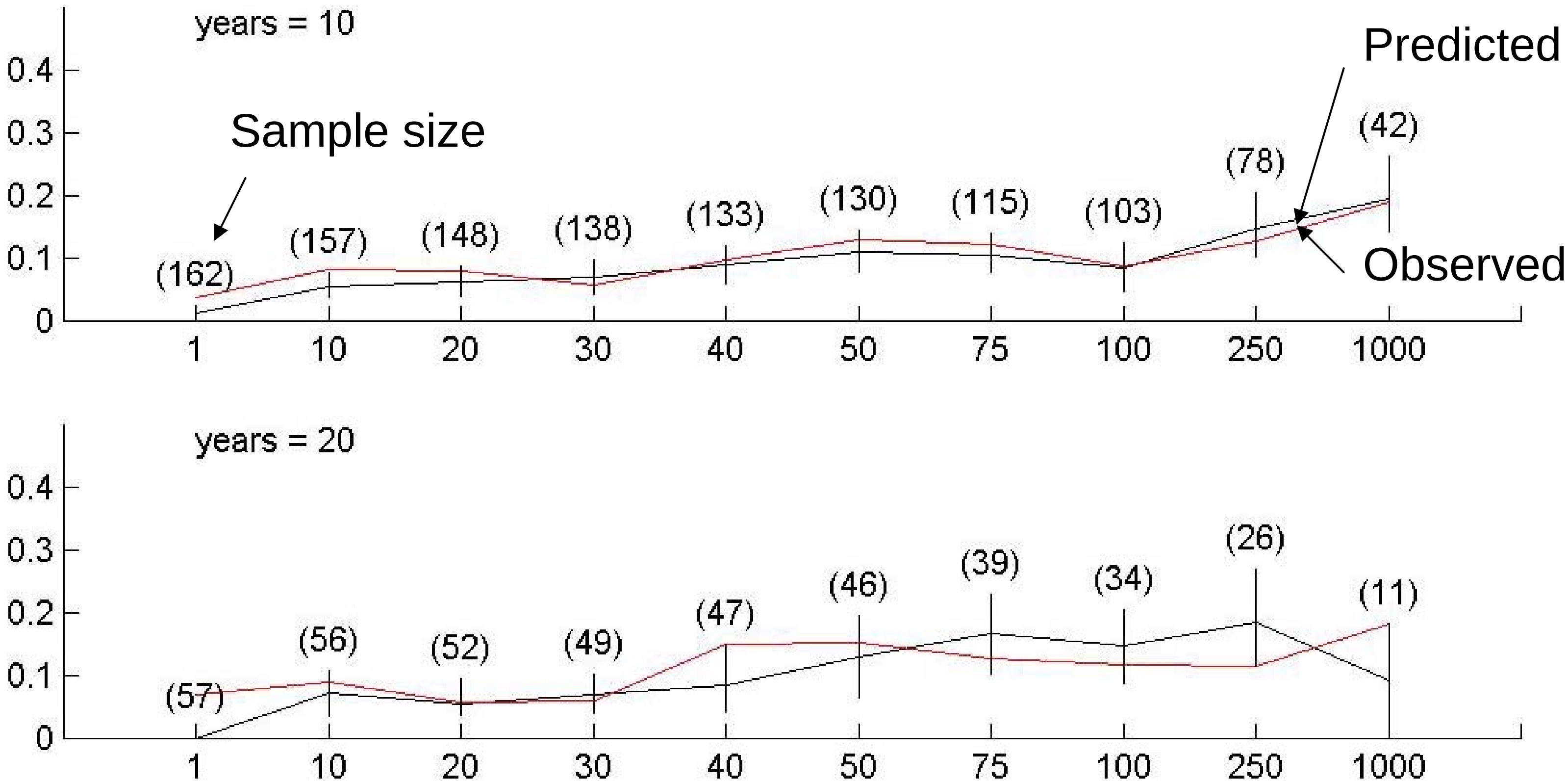
IUCN Red List Criteria

- Criteria A2: “A reduction of at least xx%, projected or suspected to be met within the next xx years….”
- Criteria C1: “Population estimated to number less than xx and an estimated continuing decline of at least xx% within xx years....”
- Criteria E: “Quantitative analysis showing the probability of extinction in the wild is at least xx% within xx years...”

Cross-validation



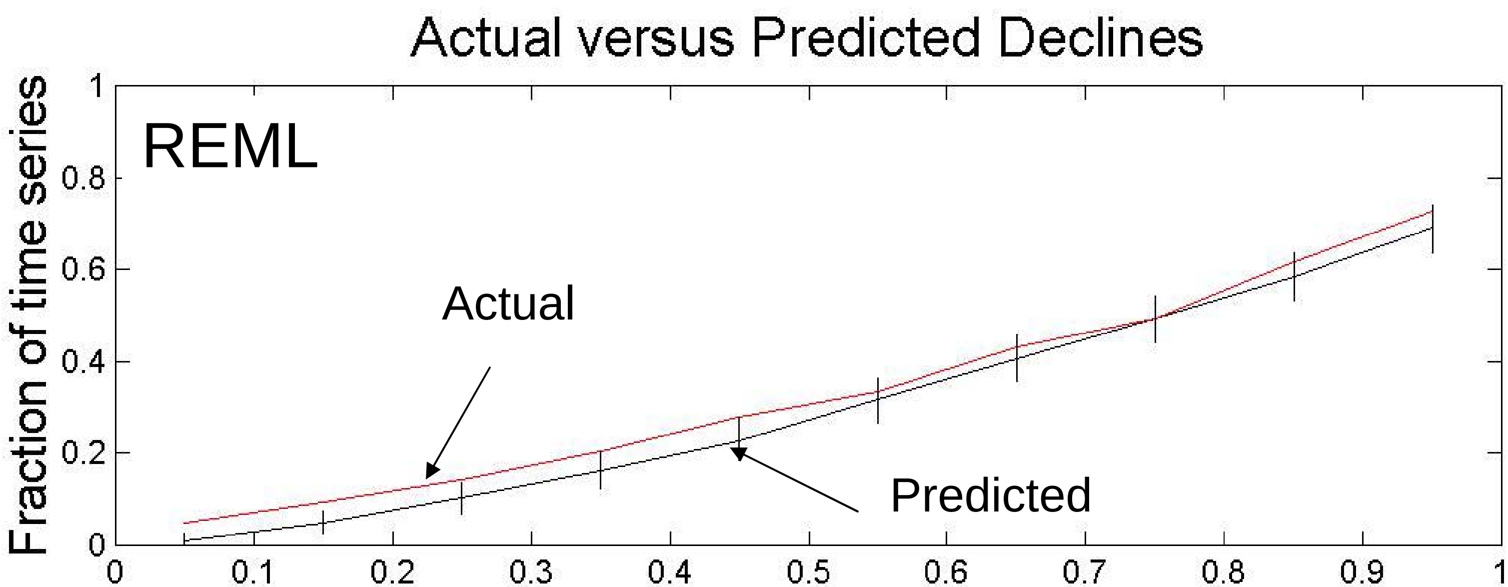
Expected vs Observed Freq. of Hitting Particular Thresholds

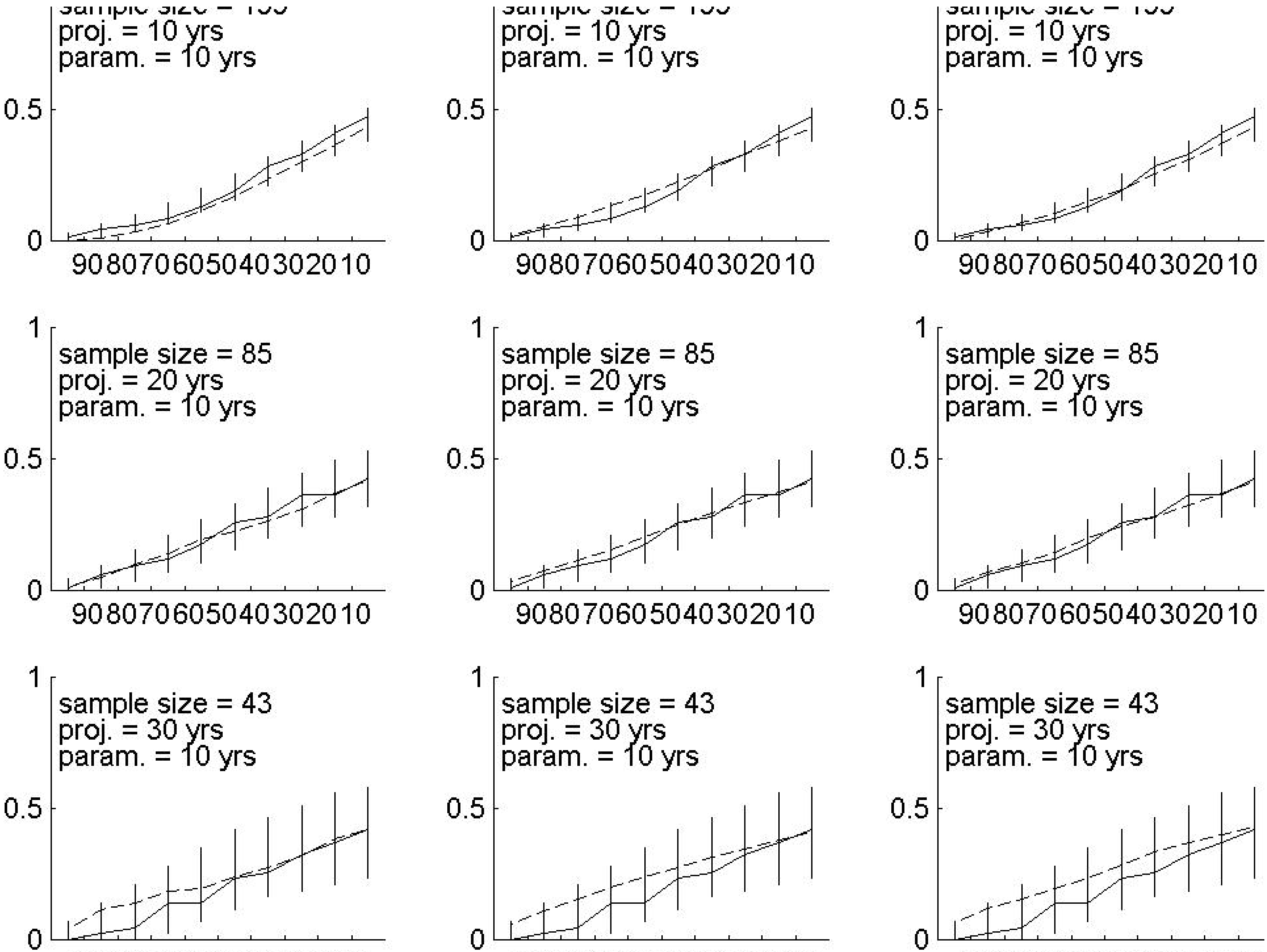


Quasi-extinction level

REML

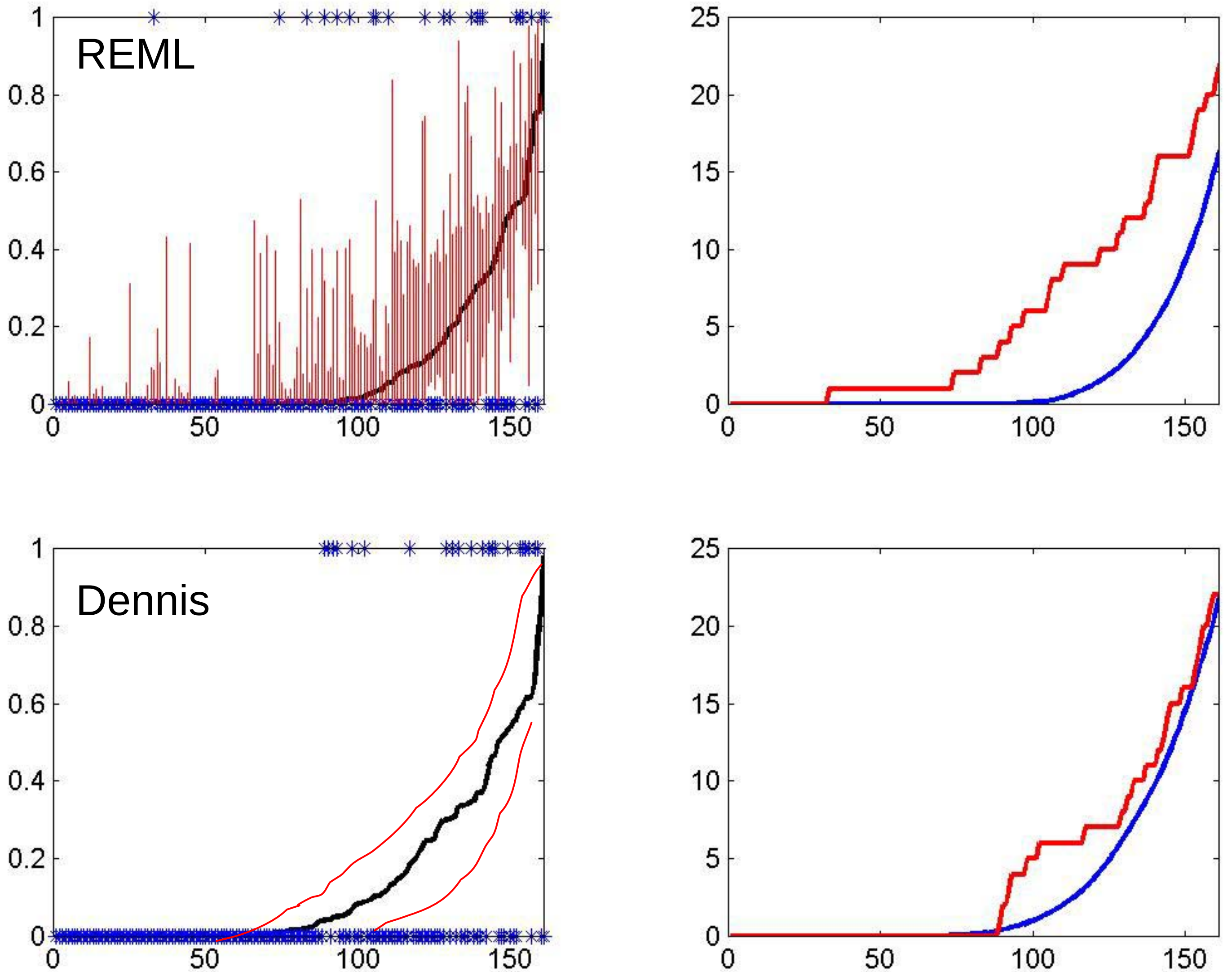
Proportional Declines




proj. = 10 yrs
param. = 10 yrs
0.5
0
908070605040302010
proj. = 10 yrs
param. = 10 yrs
0.5
0
908070605040302010
proj. = 10 yrs
param. = 10 yrs
0.5
0
908070605040302010
1
sample size = 85
proj. = 20 yrs
param. = 10 yrs
0.5
0
908070605040302010
1
sample size = 85
proj. = 20 yrs
param. = 10 yrs
0.5
0
908070605040302010
1
sample size = 85
proj. = 20 yrs
param. = 10 yrs
0.5
0
908070605040302010
1
sample size = 43
proj. = 30 yrs
param. = 10 yrs
0.5
0
1
sample size = 43
proj. = 30 yrs
param. = 10 yrs
0.5
0
1
sample size = 43
proj. = 30 yrs
param. = 10 yrs
0.5
0

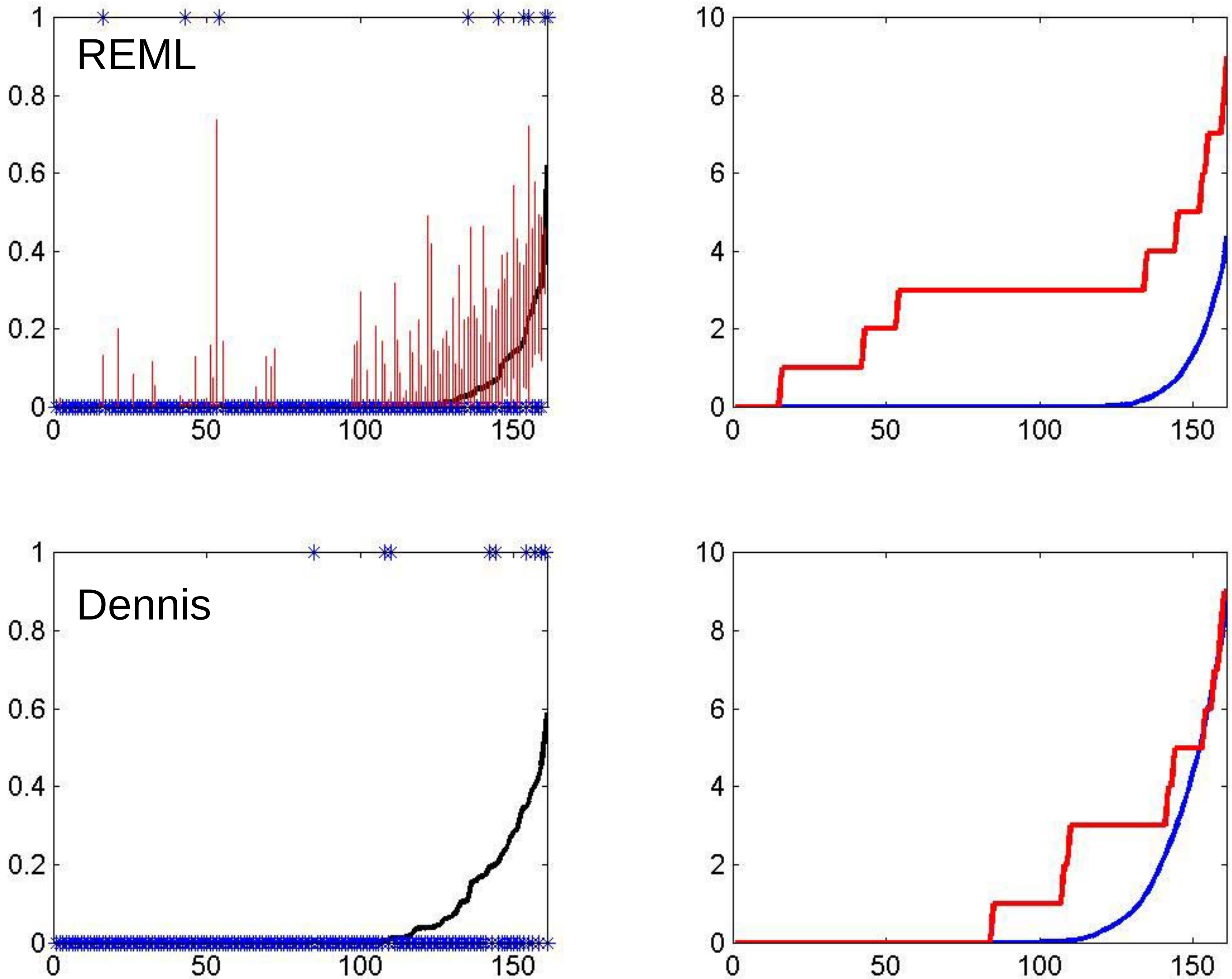
Estimated 75% decline risk vs actual 75% decline

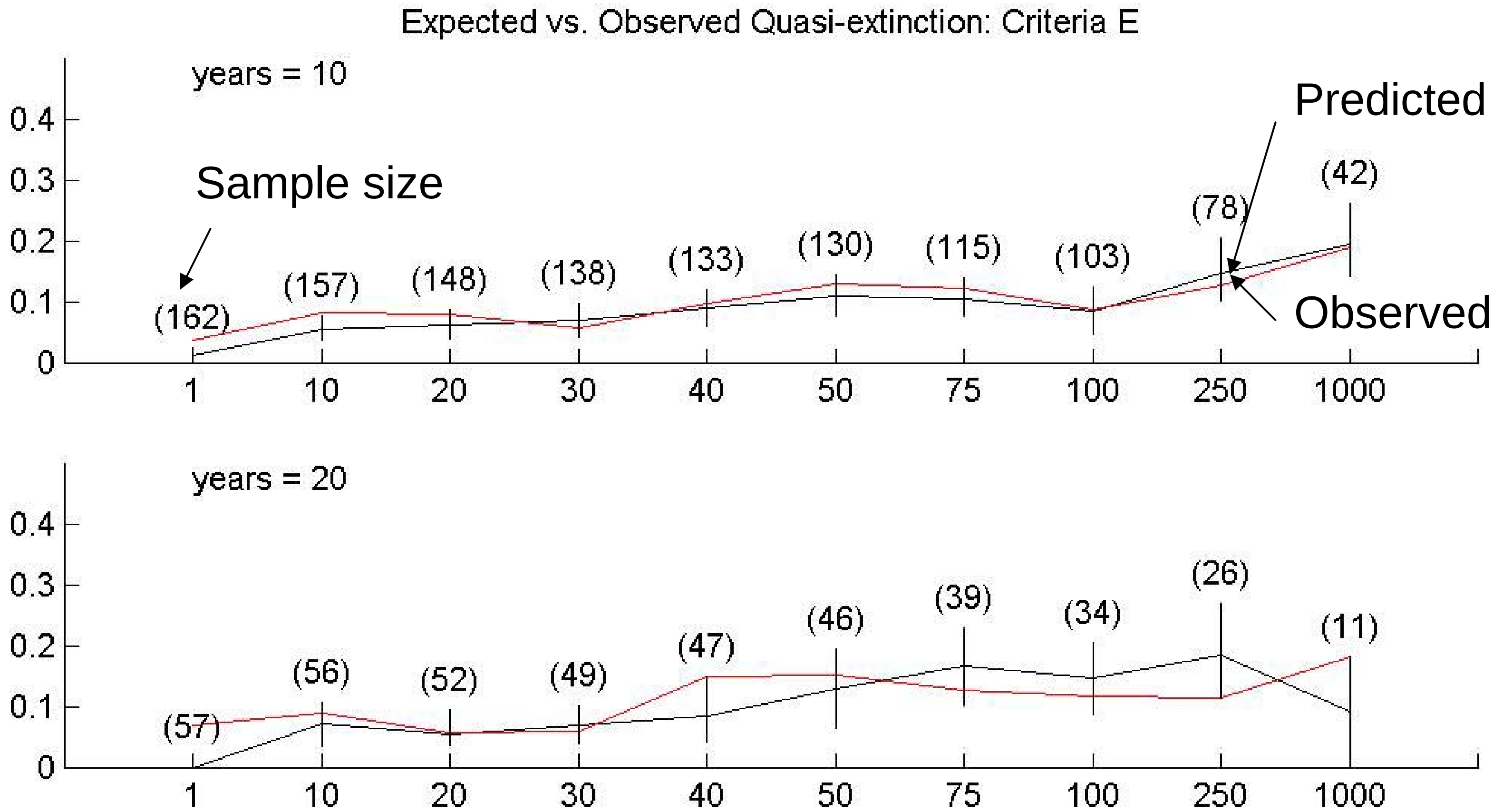
Predicted accumulated 75% declines versus actual



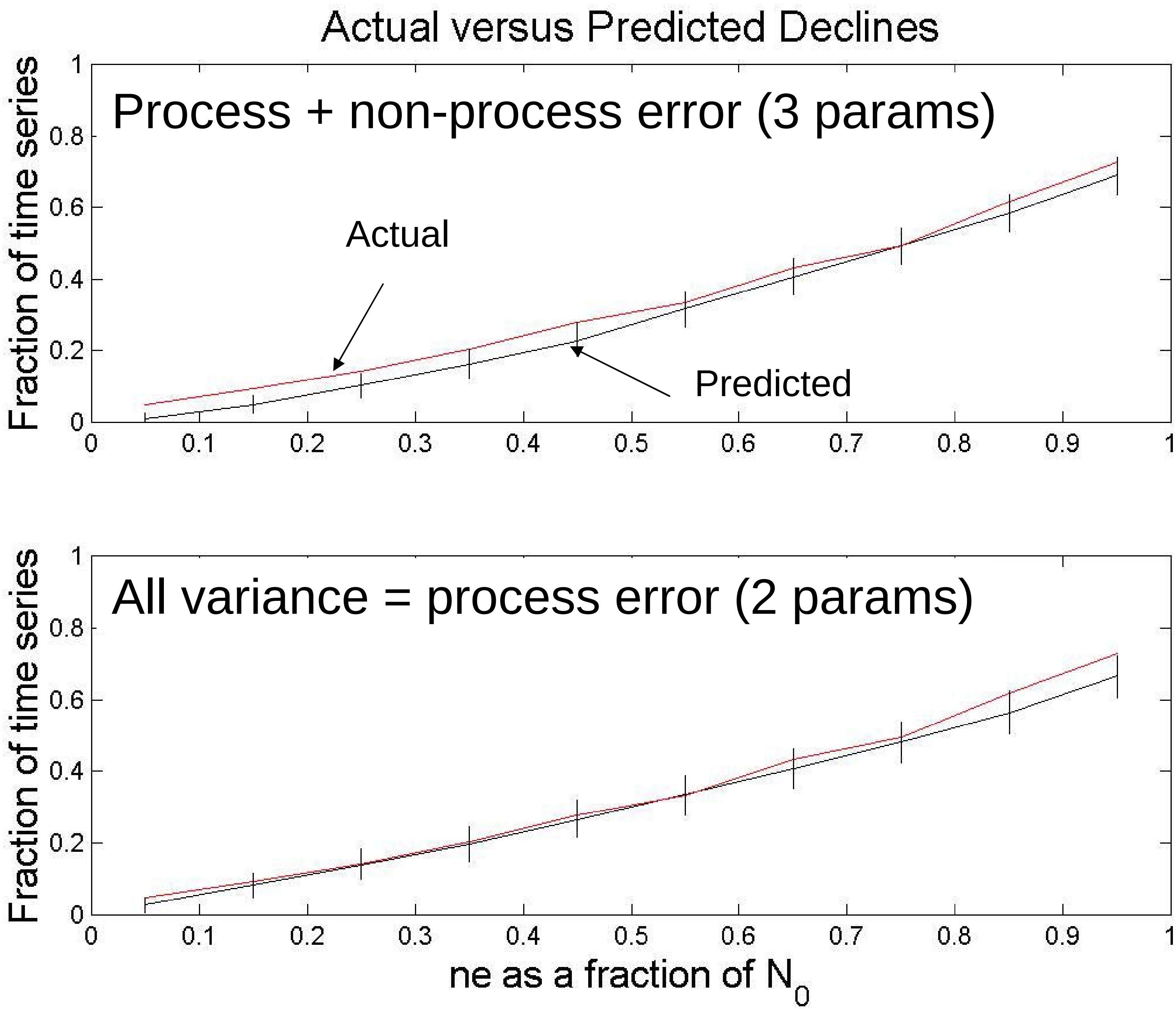
Estimated 90% decline risk vs actual 90% decline

Predicted accumulated 90% declines versus actual



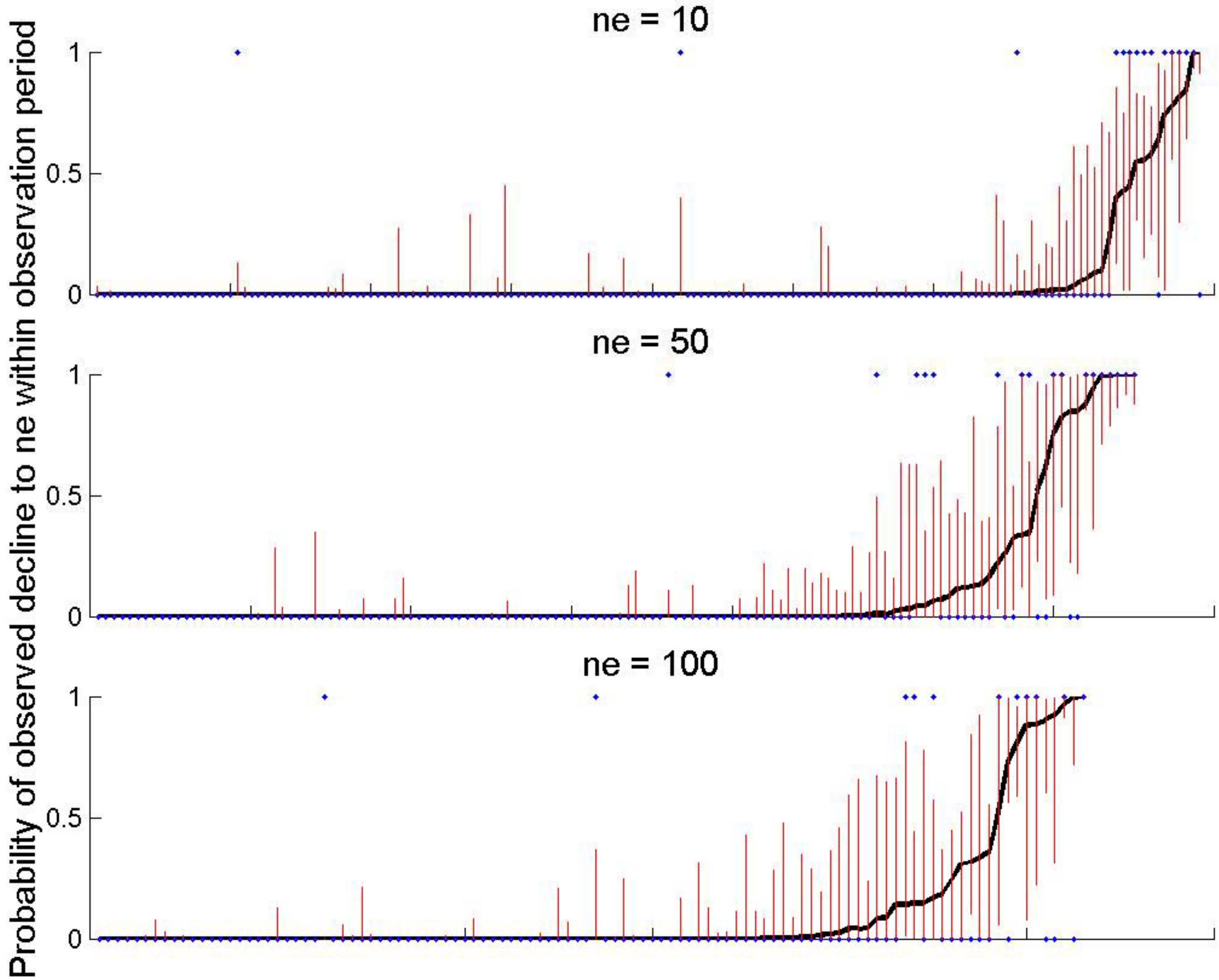


Expected vs. Observed Quasi-extinction: Criteria E
years = 10
Predicted
Sample size
Observed
(162)
(157)
(148)
(138)
(133)
(130)
(115)
(103)
(78)
(42)
0.4
0.3
0.2
0.1
0
1
10
20
30
40
50
75
100
250
1000
years = 20
(57)
(56)
(52)
(49)
(47)
(46)
(39)
(34)
(26)
(11)


Actual versus Predicted Declines
Process + non-process error (3 params)
Actual
Predicted
Fraction of time series
0
0.2
0.4
0.6
0.8
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
All variance = process error (2 params)
Fraction of time series
ne as a fraction of N0

precision versus bias

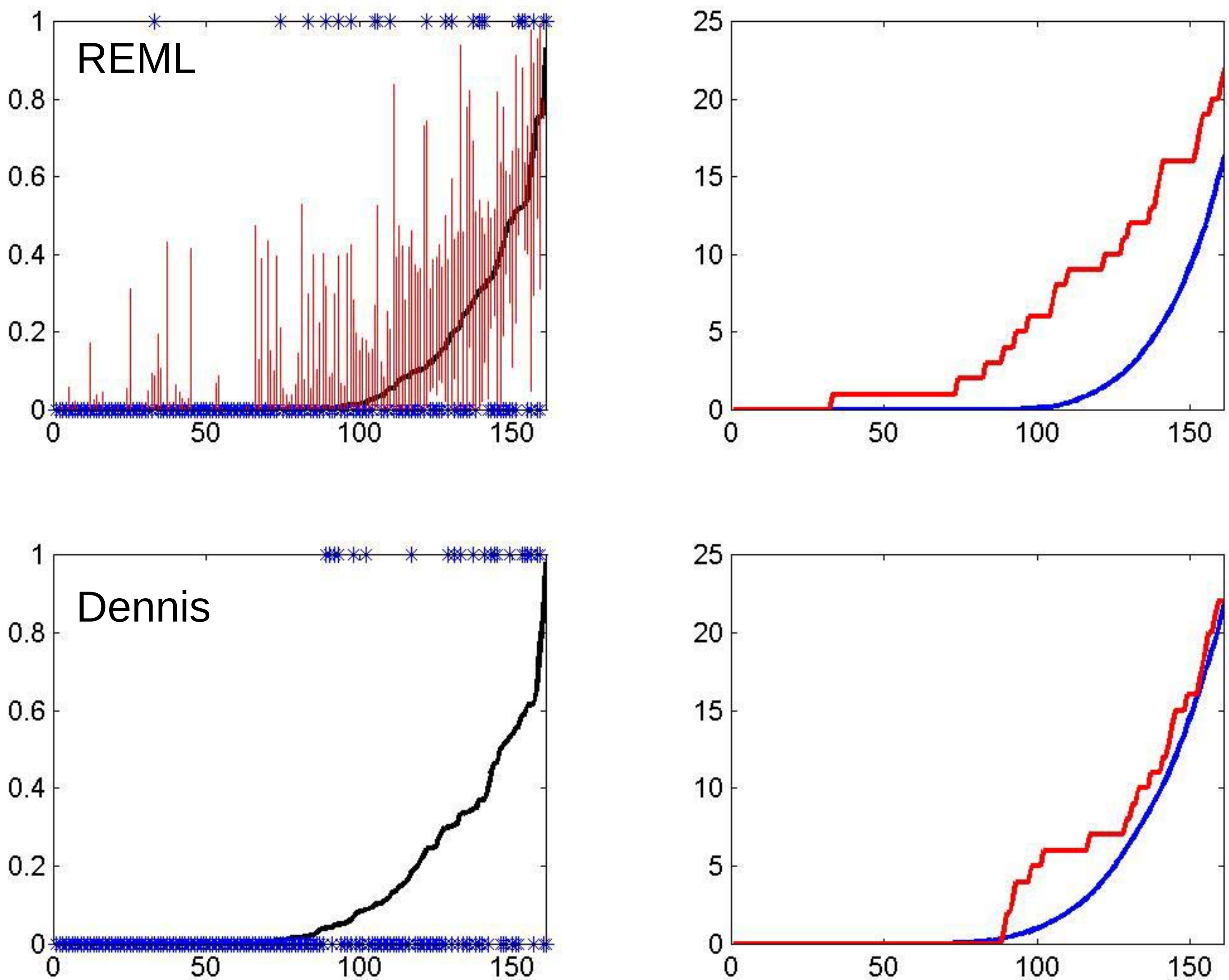
- Once again 3 parameter model is successful at capturing number of extinctions observed in dataset = low bias….
- But maybe it’s over-estimating low risks and under-estimating high risks (or visa-versa) = precise??



ne = 10
ne = 50
ne = 100
Probability of observed decline to ne within observation period
1
0.5
0

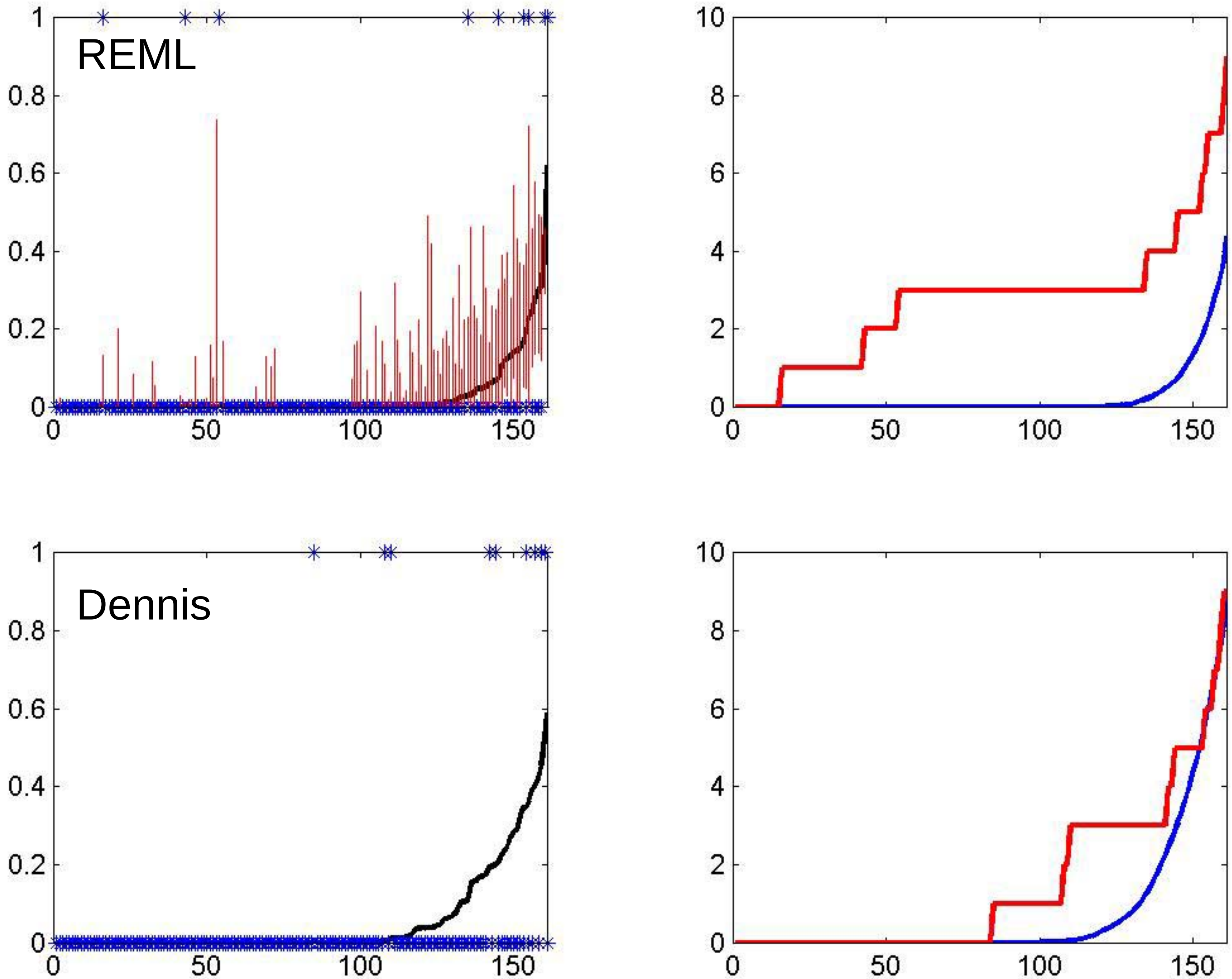
Estimated 75% decline risk vs actual 75% decline

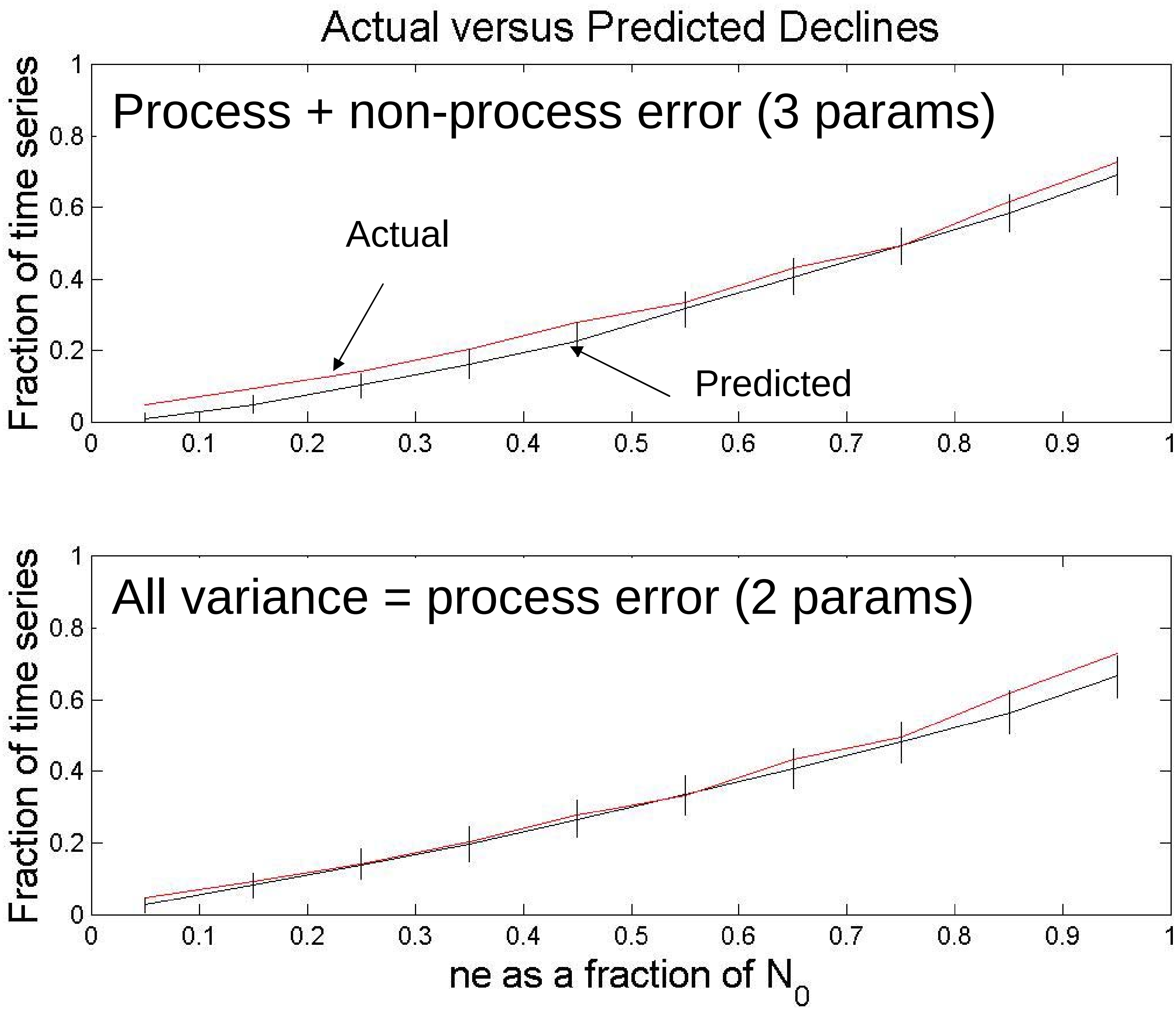
Predicted accumulated 75% declines versus actual



Estimated 90% decline risk vs actual 90% decline

Predicted accumulated 90% declines versus actual




Actual versus Predicted Declines
Process + non-process error (3 params)
Actual
Predicted
Fraction of time series
0
0.2
0.4
0.6
0.8
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
All variance = process error (2 params)
Fraction of time series
ne as a fraction of N0

These studies raise a larger question…

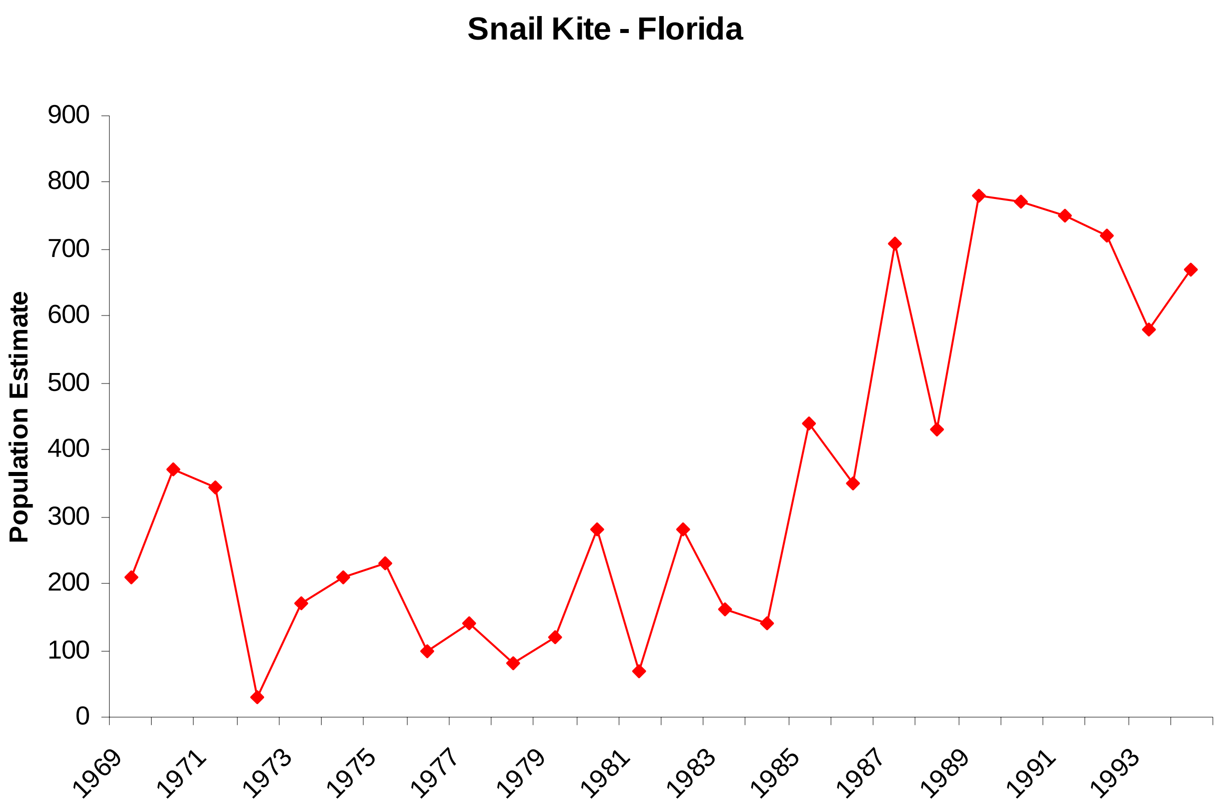
- Does there exist a general stochastic approximation to a very broad class of population trajectories?
  - Something like the stochastic Gompertz model in form
- How do we go about discovering that? And demonstrating that that form is indeed seen in real population trajectories?

Searching for general approximations for stochastic population trajectories…

- Building a theoretical foundation
  - Does there exist a general equivalent of (Tuljapurkar, Orzack, Heyde, Cohen)'s results but for population time series with density-dependence within a community web? Ives et al. 2004's result for derivation of Gompertz model from community models seems to be a start.
  - Can it be shown that the 'order' of this approximation is time dependent? Conjecture: short time (high order) → medium time (CDA) → long-time (DA)

Searching for general approximations for stochastic population trajectories…

- Building a statistical foundation
  - Need something akin to a sufficient statistic (parameter-free metrics) so that we can combine data from many different populations and study the distribution of those statistics.
  - Need to properly condition on observed data.
  - Cross-validation involves ‘distribution of suff. statistic’ times ‘distribution of the estimates’. Theoretical pdfs of estimates are based on approximations, and on the CDA model. The CDA is merely an approximation for the real process. Do these approximations hold up with real data?
  - Can we delineate the set of equally plausible alternative explanations? Can we set up tests that can reject alternatives?


Snail Kite - Florida
Population Estimate
900
800
700
600
500
400
300
200
100
0
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993