Challenges in Dynamic Optimization in Natural Resource Management

Michael J. Conroy

Topics

- Background and motivation (brief)
- ASDP and other approaches for optimal harvest management
- Use of heuristic methods for harvest optimization
- Some thoughts on the future

Background and motivation

- Most NR decision problems involve dynamic, stochastic systems with sequential controls
- Attractiveness H-J-B (DP)
- Adaptation/ Adaptive management
- Some downsides

Background and motivation

- Most NR decision problems involve dynamic, stochastic systems with sequential controls
- Attractiveness of H-J-B (DP)
- Adaptation/ Adaptive management
- Some downsides

Examples

- Forest harvest scheduling
- Optimal wildlife and fisheries harvest
- Stocking, translocations, re-introductions
- Regulations of dams on rivers
- Impoundment management

Background and motivation

- Most NR decision problems involve dynamic, stochastic systems with sequential controls
- Attractiveness of H-J-B (DP)
- Adaptation/ Adaptive management
- Some downsides

The general problem

Value function

Terminal value

$$\max_{a(t)\in A} J = \sum_{t=t_0}^{t_f} I(x, a, z, t) + F[x(t_f)]$$

states

actions

random var.

Subject to:

$$x(t+1) = x(t) + f(x, z, a, t)$$ Dynamics

$$x(t_0) = x_0$$ Initial conditions

Leads to recursive solution (dynamic programming):

Present value

$$V^*[x(t),t] = \max_{a \in a(t)} \{R(x,a,t) + V^*[x(t+1),t+1]\}$$

Optimal decision at t

[Expected]
Future value

Sustainability

- Objective (harvest) is defined over infinite time
- To maximize objective requires sustaining population

Dynamic programming

- Guarantees a globally optimal strategy for control

- Provides closed-loop feedback
  - Future resource opportunities "anticipated"

Background and motivation

- Most NR decision problems involve dynamic, stochastic systems with sequential controls
- Attractiveness of H-J-B (DP)
- Adaptation/ Adaptive management
- Some downsides

Sources of uncertainty

- Environmental stochasticity
- Partial controllability
- Partial observability
- Structural uncertainty

Active adaptation

- Accounts for structural uncertainty in DM
  - Model-specific transitions
  - Model-specific information weights (model probabilities)

- Explicitly treats information weights as another system state

- Current decision making "anticipates" future reward to objective of learning

Dual Control

Expected value of decision

$$\overline{V}[a(t),x(t),t] = \{R_i(x,a,t) + \sum_i \sum_x p_i(t)\, p_i(x_{t+1} \mid x_t, a_t) V_i[a(t+1), x(t+1), t+1]\}$$

Model probability

Transition probability (model-specific)

Background and motivation

- Most NR decision problems involve dynamic, stochastic systems with sequential controls
- Attractiveness of H-J-B (DP)
- Adaptation/ Adaptive management
- Some downsides

Some issues

- The Curse of Dimensionality
  - High-dimensioned problems difficult or intractable to solve with DP
- In our community
  - Issues of software accessibility and support
  - Relative complexity for the end users
  - Still a relatively small user group

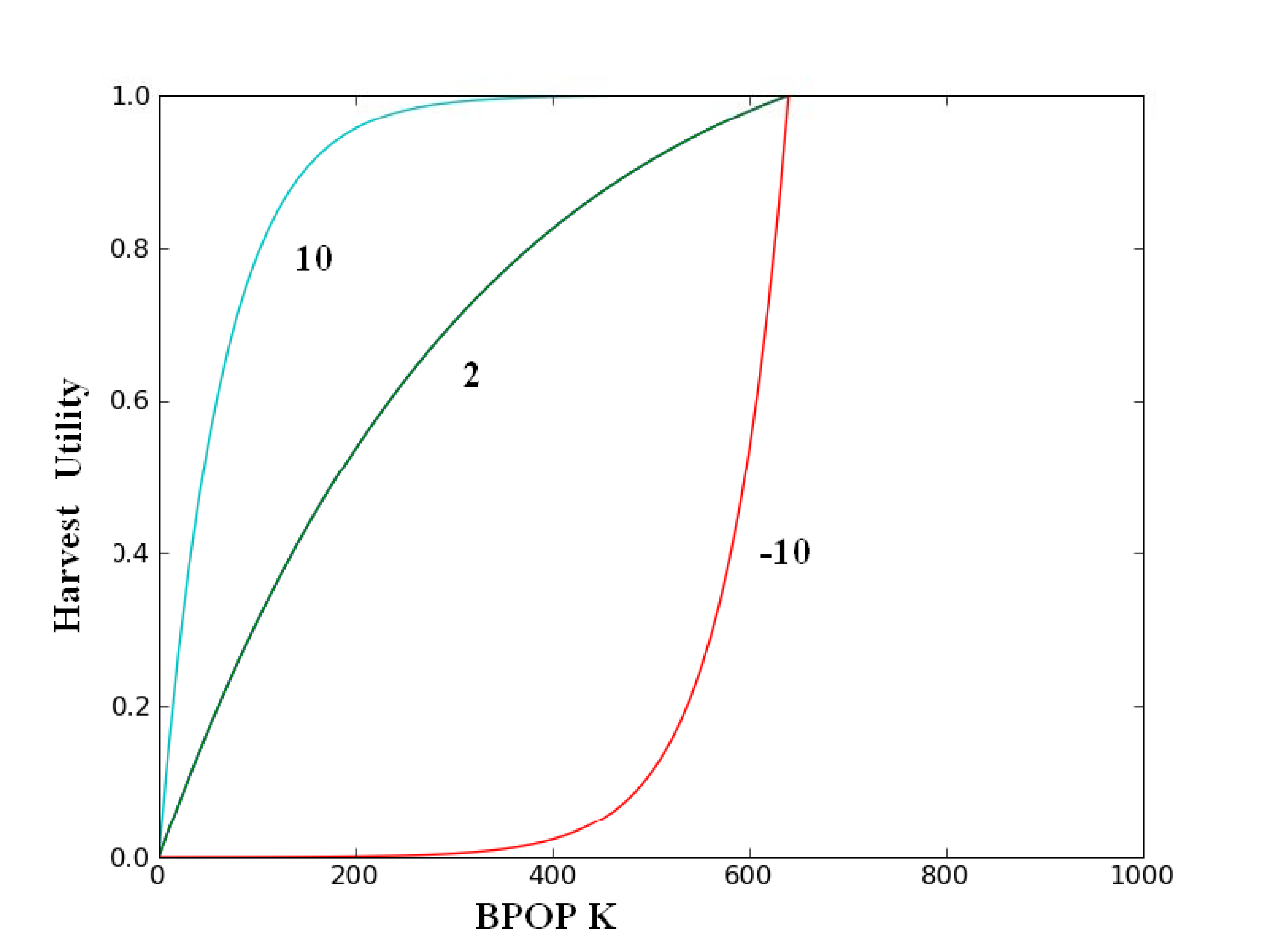
ASDP for Optimal Harvest Management

ADAPTIVE HARVEST MANAGEMENT FOR AMERICAN BLACK DUCKS

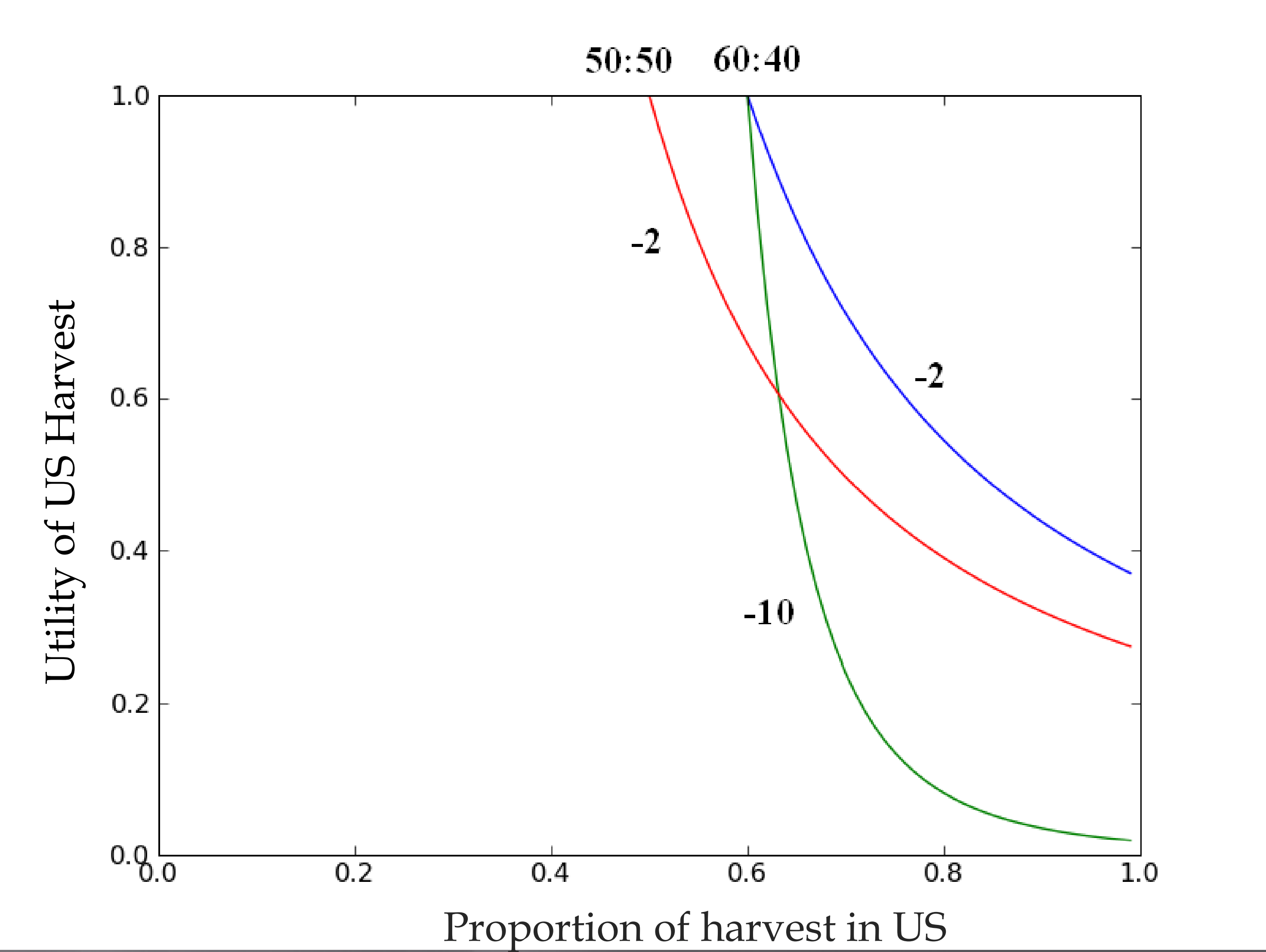
Objectives

- Maximum long-term total harvest … but
- Constraints for achieving population goals
- Allocation (parity) sub-objective
  - Canada vs. US

Population constraint



Parity constraint

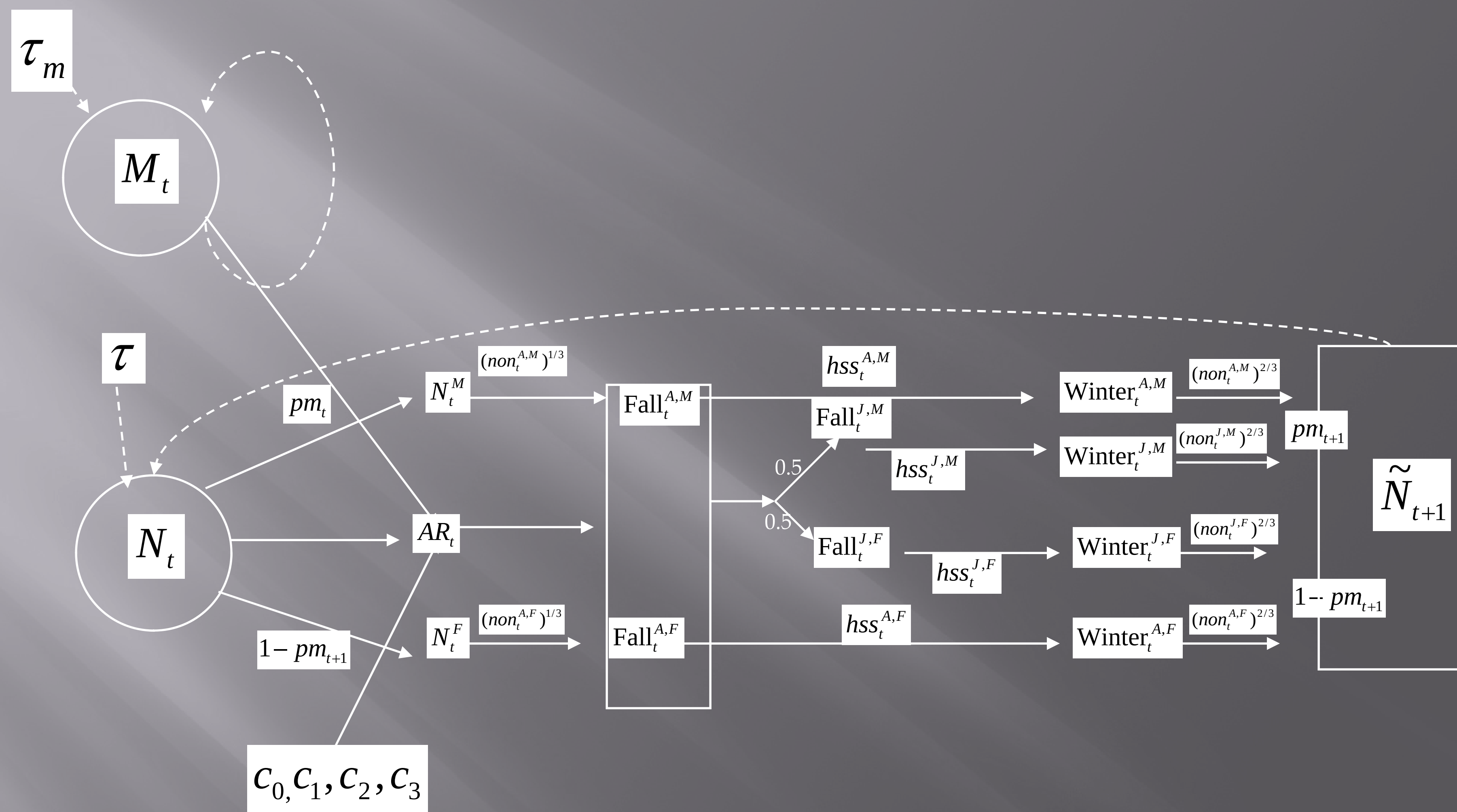


Decision alternatives

- Harvest regulations
  - Canada and US set these independently at present
  - Regulations in US can differ by flyways or portions of flyways
  - Can result in up to 6 combinations of spatially-stratified regulations
    - 3 zones in Canada
    - 3 in US
    - 7^6 = 117,649 decision combinations
  - For now assuming regulations are homogenous within US and Canada
  - For now assuming fixed harvest rate levels
    - Regulations perfectly control harvest rates
    - 7 harvest rate levels/ nation = 49 decision combinations

System states /Dynamics

- State variables
  - Spring population size of black ducks (60 discrete levels)
  - Spring population size of mallards (a competitor; 60 discrete levels)
- Dynamics
  - Black ducks
    - Density impacts on reproduction (presumed resource limitation)
    - Competition impacts from mallards (absent under alternative H)
    - Survival impacts from harvest (absent under alternative H)
    - Generalized stochastic effects (estimated)
  - Mallards
    - Simply Markovian growth (stationary)
    - Generalized stochastic effects (estimated



τ_m
M_t
τ
N_t
pm_t
$1-pm_{t+1}$
N_t^M
$(non_t^{A,M})^{1/3}$
AR_t
N_t^F
$(non_t^{A,F})^{1/3}$
$c_{0,}c_1,c_2,c_3$
$\text{Fall}_t^{A,M}$
$\text{Fall}_t^{A,F}$
$hss_t^{A,M}$
$\text{Fall}_t^{J,M}$
0.5
0.5
$hss_t^{J,M}$
$\text{Fall}_t^{J,F}$
$hss_t^{J,F}$
$hss_t^{A,F}$
$\text{Winter}_t^{A,M}$
$\text{Winter}_t^{J,M}$
$\text{Winter}_t^{J,F}$
$\text{Winter}_t^{A,F}$
$(non_t^{A,M})^{2/3}$
$(non_t^{J,M})^{2/3}$
$(non_t^{J,F})^{2/3}$
$(non_t^{A,F})^{2/3}$
pm_{t+1}
$1-pm_{t+1}$
$\tilde{N}_{t+1}$

Uncertainty

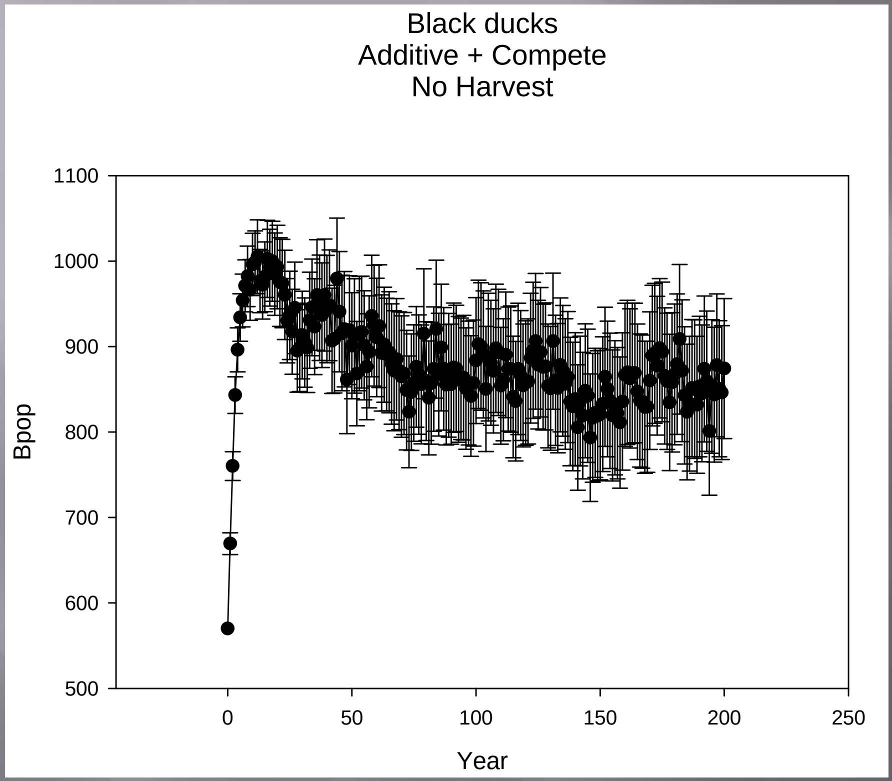
- Environmental stochasticity
  - Represented by estimated random effect on black duck and mallard dynamics
  - Discrete lognormal distribution (14 levels)
- Partial controllability
  - Assume for now that specific harvest rates can be achieved
    - Further work needed to characterize stochastic relationship of regulations to harvest outcomes
- Partial observability
  - Incorporated into state-space mode
  - Ignored in optimization
- Structural uncertainty
  - 4 alternative process models
    - Harvest effects X Mallard competition

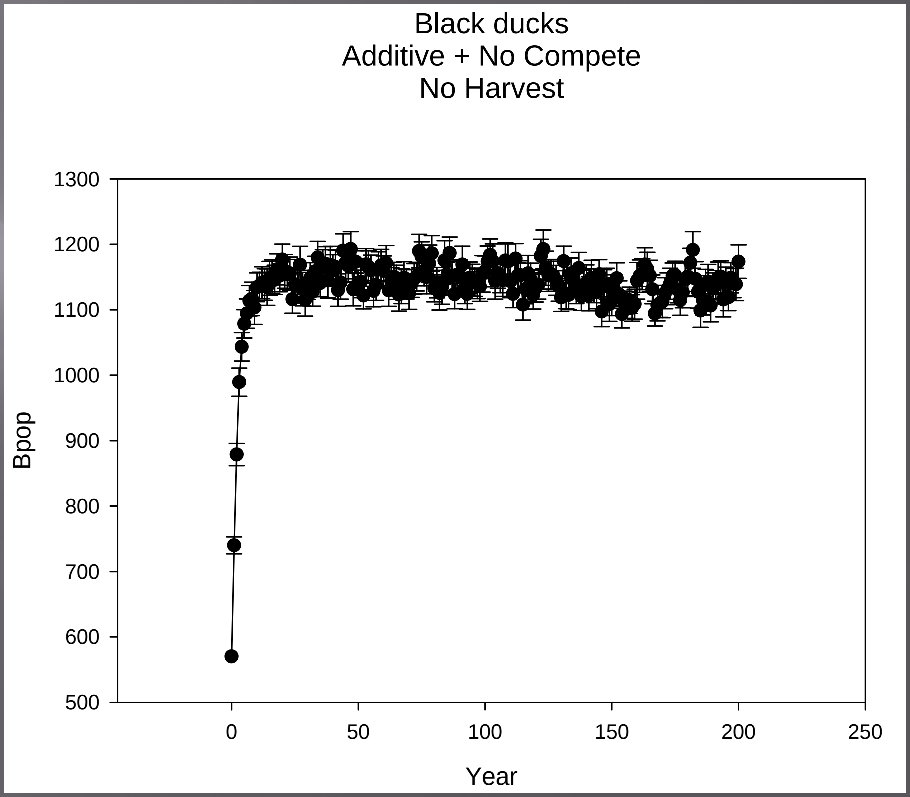
Casting in ASDP

- State-decision- RV space
  - $60^2 \text{ X } 7^2 \text{ X } 14^2 = 3.5 \text{ X } 10^7$
- Stationarity issues
  - Most model/ objective scenario combinations did not converge on stationary solution in 200 iterations
  - Reported stationary state-specific strategy (if found) or iteration 200 strategy
- Simulation of "optimal" strategies
  - Initial conditions 570K black ducks 470L mallards
  - 100 simulations of 200 years

Typical results

No harvest, simulated trajectory (2 models)

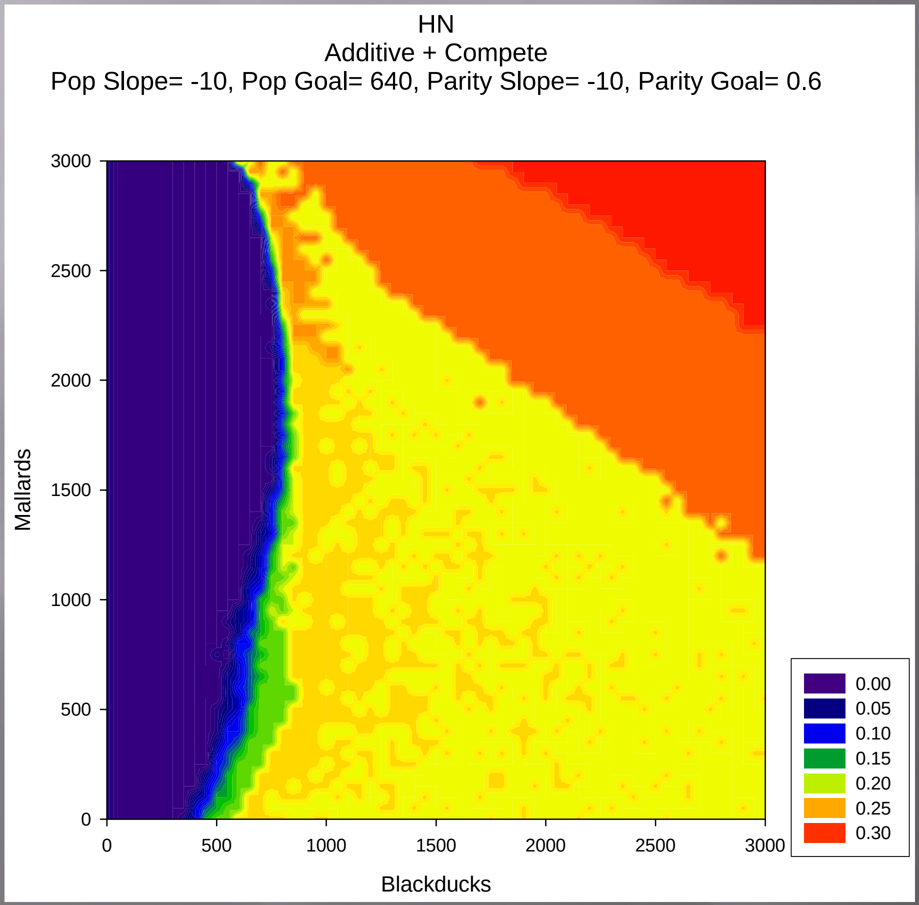




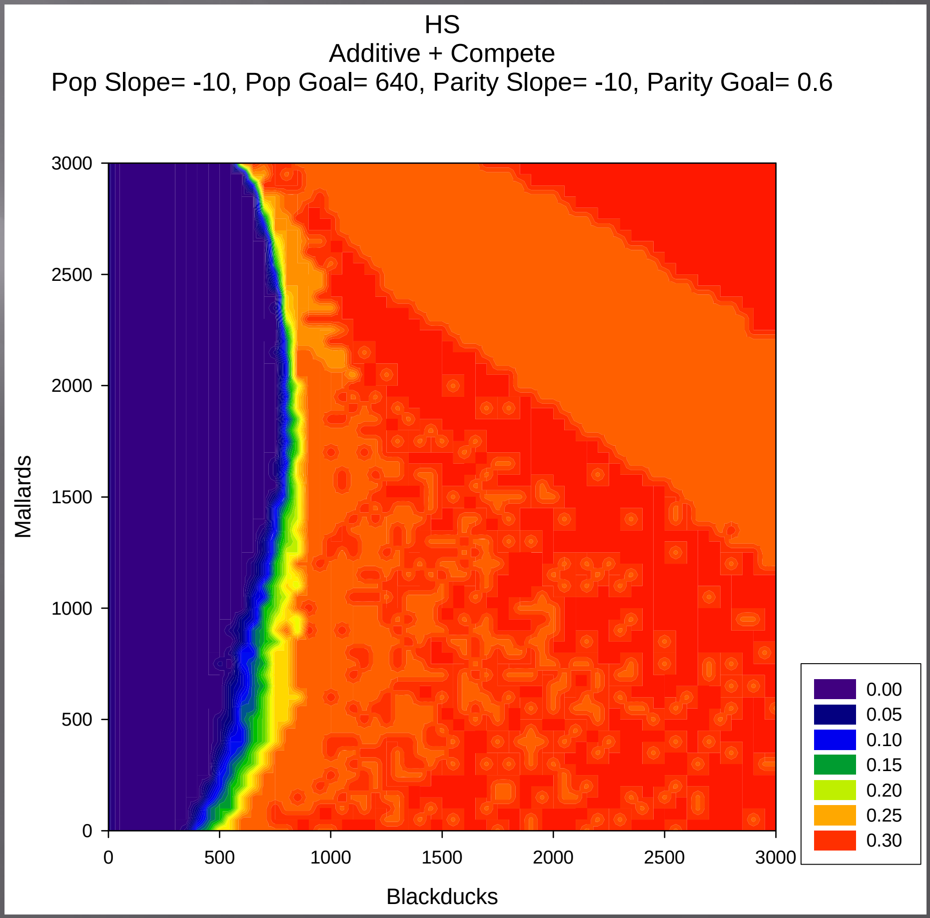
Additive, competition

Additive, no competition

Optimal strategy (Strong population and parity constraints, Additive/compet.)

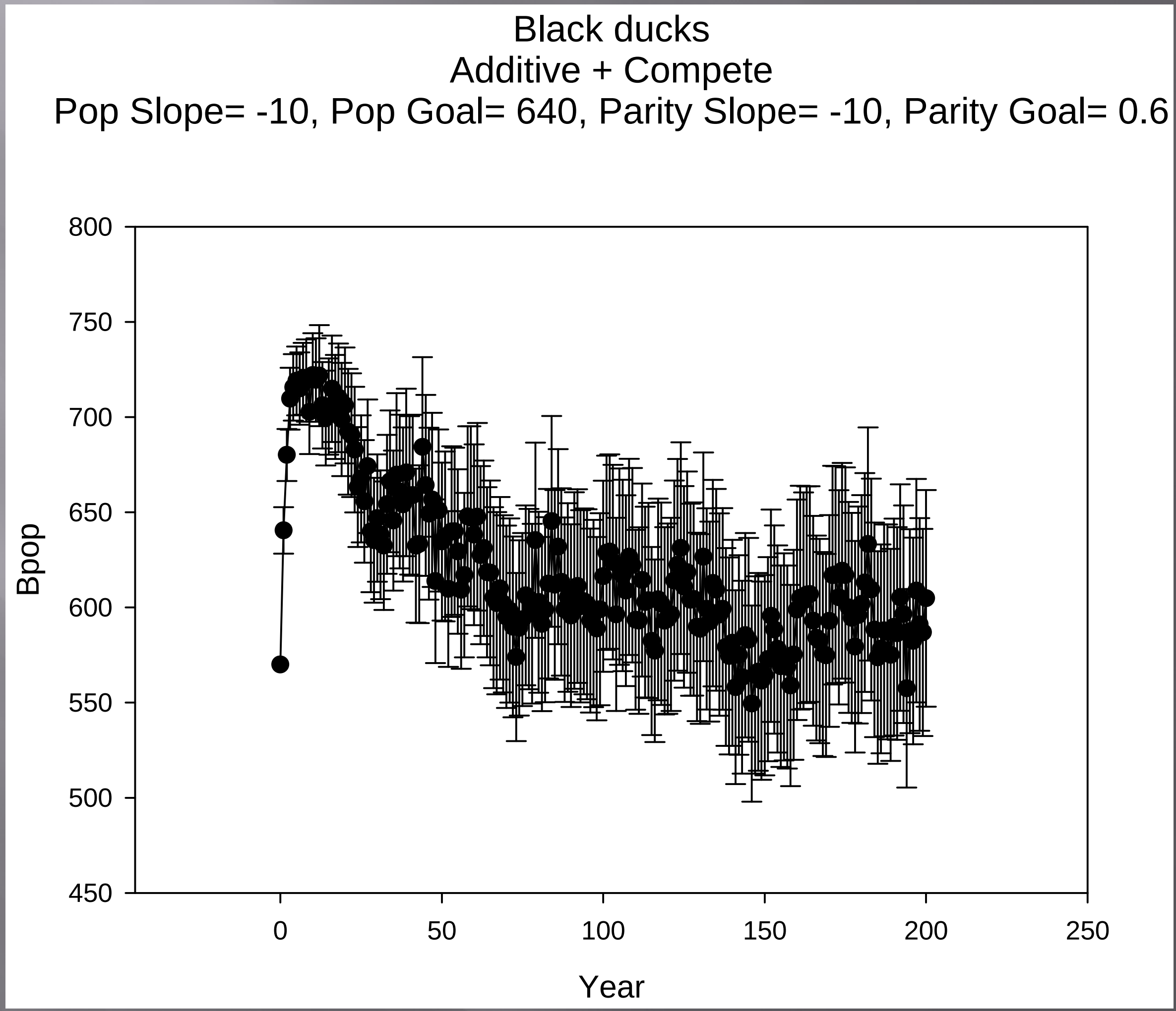


Canada

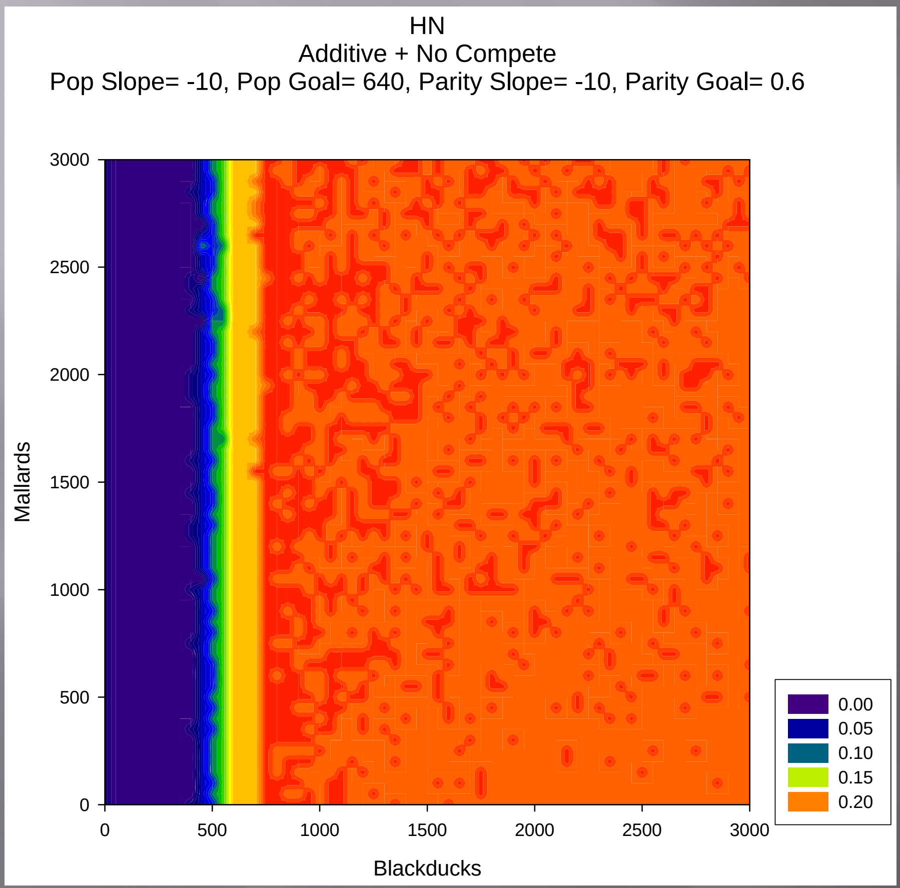


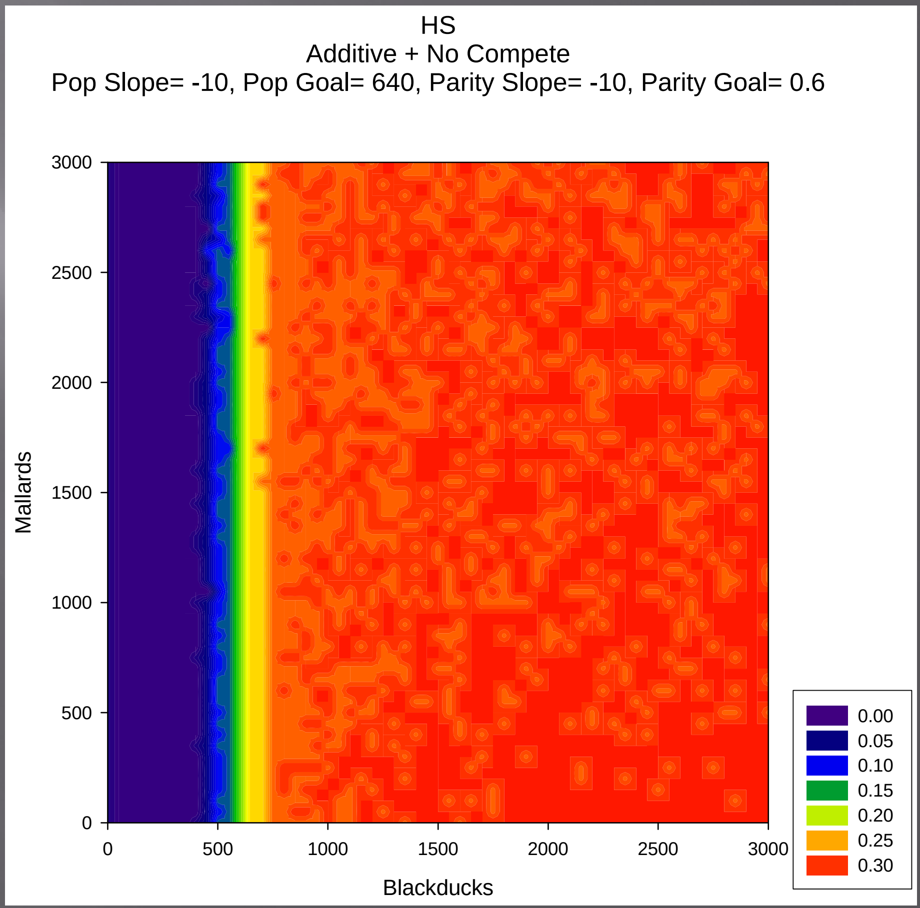
US

Simulation of Optimal strategy (Strong population and parity constraints, Additive/compet.)



Optimal strategy (Strong population and parity constraints, Additive/no compet.)

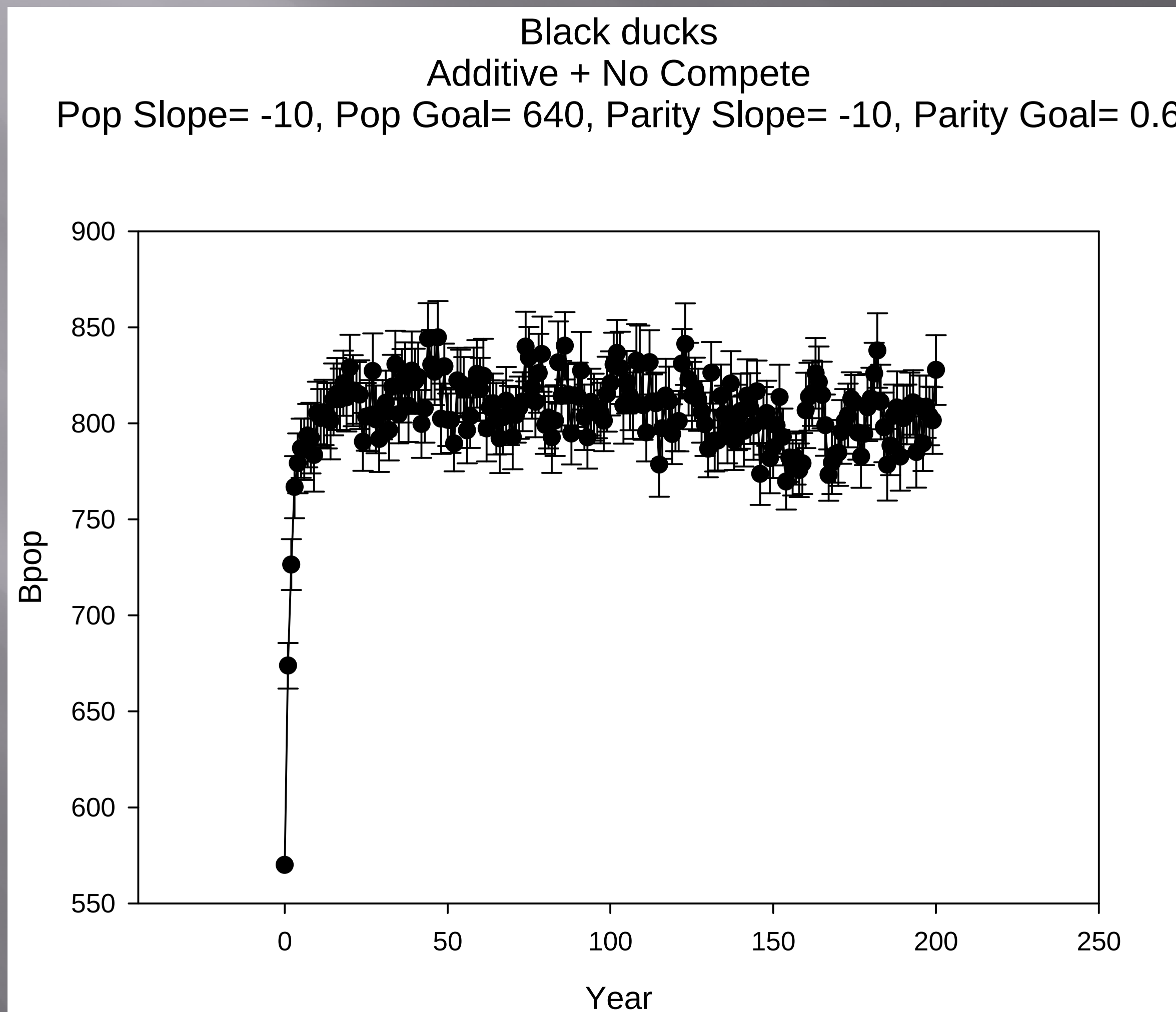




Canada

US

Simulation of Optimal strategy (Strong population and parity constraints, Additive/no compet.)



Problem extensions

- Incorporation of partial controllability
  - 14 random harvest rate outcomes per harvest decision (4-5 levels)
- Spatial stratification
  - 3 breeding populations
  - 6 harvest zones
- State – decision- RV dimensions (independent populations and harvest zones)
  - 60^6 X 5^6 X14^9 =1.5 X 10^{25}
  - Haven't done this!
  - Still trying to get buy-in on single population, 2 – harvest international strategy

ADAPTIVE HARVEST MANAGEMENT FOR WESTERN MALLARDS

Motivation

- Mallard AHM based (c. 2005) on single stock ("Midcontinent Population")
- Pacific Flyway mallards
  - Derive much of harvestable population from coastal and trans-Rockies west
  - However substantial intermixing with midcontinent population
- Work explored feasibility of western AHM
  - 2-stock "virtual model"
    - Independent stochastic effects and dynamics
    - Independent harvest regulations

Properties of a candidate model

- Equal or less complexity than MCP
  - Take state space = D^2
- Harvest decisions and population states independently determined, of similar dimension to MCP
  - Could reduce dimension by linkage
- No current model of population interaction
  - Assume independent for now
  - Interaction structure potentially reduces dimension
- Stochastic variation
  - Assumed independent for now
  - Covariance structure would reduce dimension

Initial model

- "Cloned" MCP model
- Joint model
  - States, decisions, random variables completely independent
  - Dimensionality = D^2 where D= dimensionality of MCP model

Evaluations of performance

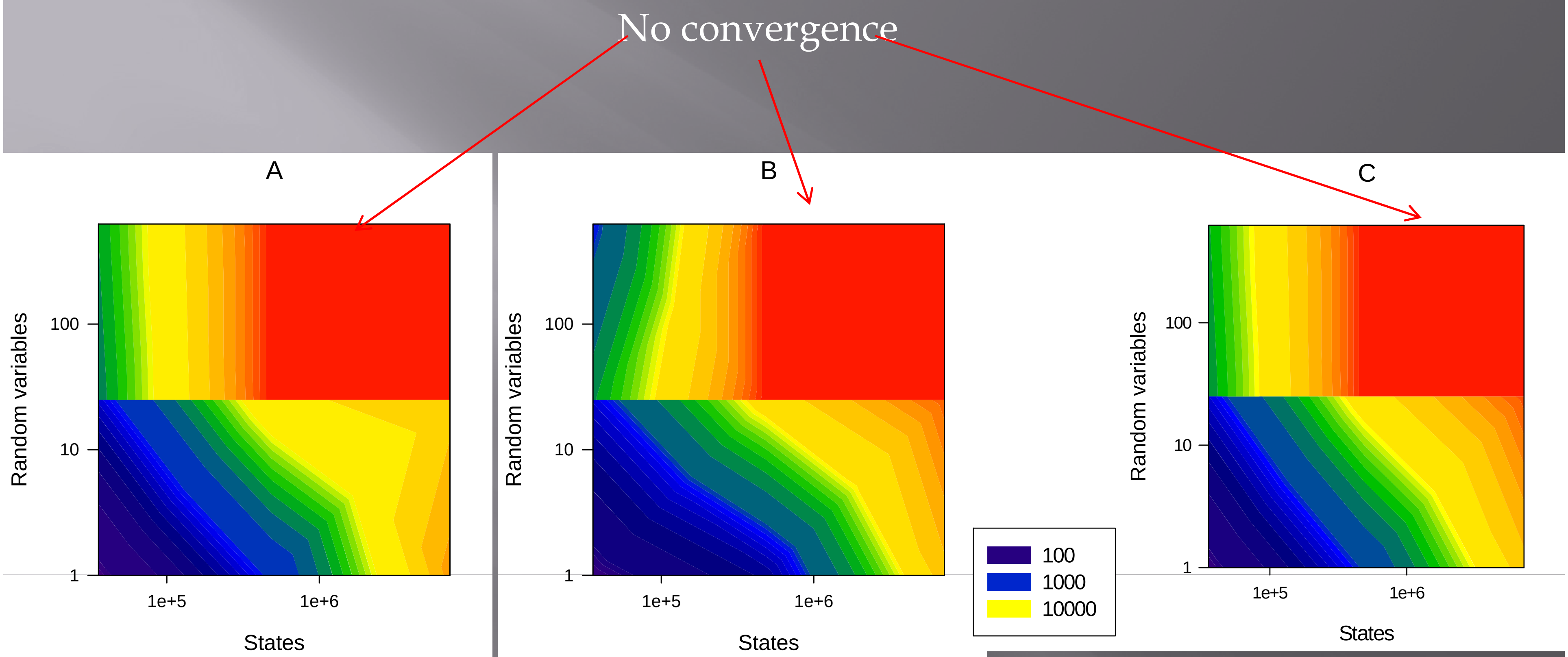
- Scenarios
  - 4 independent harvest alternatives per population
  - Population states 0-20 M , ponds 1-9 M per population
    - Discretization from 0.25 to 1 M
    - RV dimensions from 1 (deterministic) to 400K
- Platforms IBM & Dell desktops
  - IBM 2.40 GHZ 640MB
  - DELL 2.8 GHZ 512 MB
  - DELL 2.8GHZ 1GB

Table 1. Dimensions of optimization/ simulation problems investigated.

Scenario file	Number of state combinations	Number of random variables	Number of decision combinations	Total dimension
D1	7,144,929	1	16	114,318,864
D2	35,721	1	16	571,536
D4	485,809	1	16	7,772,944
R1	7,144,929	25	16	2,857,971,600
R2	4,85,809	25	16	194,323,600
R3	35,721	25	16	14,288,400
R4	7,144,929	625	16	7,1449,290,000
R5	485,809	625	16	4,858,090,000
R6	35,721	625	16	357,210,000
R7	7144929	25	16	2,857,971,600
R8	485809	25	16	194,323,600
R9	35721	25	16	14,288,400

Results

- Attempted to obtain stationary ASDP solutions for 36 scenario-platform combinations
  - 12 failed to converge in <24 h, several still running after 1 wk
    - Scenarios R4,5,7,8
    - All 3 platforms
  - Remaining 24 convergence time from <100 s (D2) to > 50,000 s (R1)
  - Convergence time function of both state dimension and RV dimension
    - As RV → 100 even low-dimension problems were slow to converge
- If convergence occurred, simulations took only a modest amount of additional time



IBM 2.40GHZ 640MB

DELL2.8GHZ 512MB

DELL 2.8GHZ 1GB

Conclusions/ recommendations

- Currently not practicable to obtain full DP solution to joint AHM problem involving
  - Relatively fine discretization of states and decisions
  - Full incorporation of stochastic effects
- Alternatives
  - Brute force computing power (suck it up)
  - Simplify
    - Simpler model structure and random variable distributions
    - Coarser discretization
    - Non-independent decisions (e.g., proportional)
  - Deterministic DP followed by stochastic simulation
  - Heuristics

Use of heuristics for optimal harvest management

Rationale

- Fully optimal closed loop (DP) solutions not always practicable
  - The Curse happens quickly
  - Resource managers do not have supercomputers
- Heuristic methods may get us "close enough" to the optimal solution
- Some heuristic methods
  - Simulation-optimization
  - Genetic algorithms
  - Reinforcement learning
  - Simulated annealing
- I'll discuss the first 3 and mainly the 2nd and 3rd

Simulation-optimization

- Forward stochastic simulation through time
- Exponentially increasing complexity of decisions
  - In practice draw candidate decisions at each time and simulate these
- For each simulation evaluate harvest utility
- Advantages
  - Arbitrary complexity possible
  - Can represent states, RVs, and transitions continuously
- Downsides
  - No process for culling suboptimal decisions as in DP
  - Requires very large number of replications even for short time horizons
  - No assurance of global optimality

Genetic Algorithms

- Evolutionary model for optimization
- Alternative decisions represented by combinations of “alleles”
- Decision space explored via mathematical analogs to recombination and mutation
- Achievement of objective measured by a “fitness function” (e.g., harvest utility)

Genetic algorithms

- Advantages
  - Do not require state discretization, dynamics can follow continuous functions
  - Can be arbitrarily complex with little if any computational penalty
  - Can apparently be efficient
- Disadvantage
  - No general conclusions about optimality possible

Application to Harvest Optimization

- Moore (2002) Appendix E

- Johnson et al (1997) formulation of Anderson (1975) mallard harvest model
  - Duck abundance and pond states
  - Dynamics under 4 alternative models
  - Stochastic rainfall and harvest outcomes
  - Harvest utility simple total cumulative harvest

GA trials

- Fixed (15-y) time frame
- 81 levels of harvest rate from 0 - 0.5
- GA
  - Each annual decision =1 “gene” on a 15-gene “chromosome”
  - “Chromosome” encoded a particular 15-y harvest decision schedule
  - Fixed population followed over fixed number of generations
  - “Organisms” pair, exchange genetic material, and are replaced by offspring
    - Bernoulli trials to determine mutation

Steps

1. Input initial system state and model
2. Initialize population of C organism with 15C random alleles
3. g=0
4. Do until g=G
    1. Evaluate expected fitness of all organisms
    2. Construct mating pool
    3. Crossover genetic material between parents
    4. Mutate alleles of offspring (or not)
    5. Create replacement population from offspring plus elite-selected parents
    6. g=g+1
5. Retrieve organism with greatest fitness, interpret allele A1= optimal state-specific harvest rate

Comparison of GA to DP

- Solutions mostly consistent for 2 models of compensatory harvest mortality
  - However GA underestimated optimal harvest rate for high-abundance initial state
- Solutions diverged for 2 models of additive harvest mortality
  - For high initial abundance GA underestimated optimal harvest rate
  - For low initial abundance GA overestimated optimal harvest rate
- GA generally outperformed random search algorithm
- GA tended to be risk aversive compared to DP
  - Maintained a higher than optimal stock, lower harvest

Conclusions

- GA may perform reasonably well in searching for optimal harvest strategies in complex systems
- Still many issues regarding implementation
  - Subjectivity of decisions regarding population size, mutation rate, etc.
  - No general statements possible from this example
  - Problem: how do we judge relative performance when DP is infeasible?

Reinforcement learning

- Broad definition (Sutton and Barto 1998)
  - "Any goal-directed learning problem based upon interaction with a system or a model thereof"
- RL "learns" an optimal policy by receiving reinforcement from a dynamic environment
- Feedback guides exploration of the space of feasible policies by evaluating actions taken
- RL is *unsupervised* (e.g., in contrast to neural networks)
- RL combines trial-and-error search with delayed reward from the environment to achieve its goals

Fonnesbeck (2003)

- Imbedded a MDP in RL
- Constructed an "action-value" function in terms of a state-action pair $Q^{\pi}(s,a)$
  - Calculates a value for each available action at state s assuming that future actions are chosen according to stated decision policy π
  - When value function is maximized for each state $s \in S$ then policy is optimal $\pi = \pi^*$ and Q is equivalent to the H-J-B equation
- Average accumulated rewards from n sample visits to each state

Fonnesbeck (2003)

- Formulation in terms of Bellman's equation

$$Q^*(s,a) = \max_{a \in A_s} \sum_{s' \in S} p(s' \mid s,a)\{r_{t+1}(s,a) + \gamma Q^*(s',a')\}$$

- Estimated optimal policy should converge on π^*
- Optimal policies evaluated and improved by temporal difference learning (TDL)
  - Blends elements of DP and Monte Carlo learning to produce effective and efficient learning algorithm
  - Rather than evaluating every action at each step, TDL chooses 1 action for current state
  - Evaluates return by 1-step ahead search (like DP)

RL

- Based on difference between estimate value of *(s,a)* before and after the execution of *a*

$$Q'(s_t, a_t) = Q(s_t, a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$

RL Basic Steps ("SARSA" method)

- Initialize Q(s,a) arbitrarily for all *s* and *a*
- Initialize *s* arbitrarily
- Choose initial action *a* from policy π
- Repeat until convergence
  - Execute *a* , observer *r* , *s′*
  - Choose action *a′* at *s′* from policy π
  - Update

$$Q'(s, a_t) = Q(s\ \ , a\ \ ) + \alpha[r + \gamma Q(s', a') - Q(s_t, a_t)]$$

- Produces a Markov chain of state-action pairs and associated rewards
  - Parallel chain of policies that converge on optimal policy

Application: Anderson (1975) Mallard model

- RL using tabular Q-learning algorithm
  - Mallard populations 0-17M by 1 M
  - Ponds 0-4M by 0.5 M
  - Harvest rates 0-0.6 by 0.05
- Compared to DP results with like discretization

Comparison results

- Under compensatory model
  - Estimated policy from RL close to DP only when mallard abundance low to moderate
  - Diverge >8 M
  - Similarity related to amount of state-specific experience by the RL algorithm
- Under addition model
  - RL algorithm generally failed to converge to the optimal policy
- Comparison of cumulative harvest and abundance (200 y)
  - Similar between DP and RL (overlap of 95% CI)
  - Suggest that even though policies differ, resulting objective outcomes are similar

General conclusions

- Global optimality lacking in RL
- RL Strategies likely perform poorly in extreme regions of state space (little experience)
- Other criteria (Anderson 1975 desirable properties) all fulfilled
  - Adequate consideration of environmental - uncertainty √Allows for error in observed state
  - State-specific decision making
  - Ergodicity
  - Allows for objective constraints

What now?

Some random thoughts

Brute force vs. clever and close approximations

- How will we know when we're close if the "true globally optimal" strategy cannot be found
  - And if we can find it, why would we settle for "close"
- When is "close" close enough?
- Do we really need optimal strategies?
  - Are we trying to get the best possible resource outcome?
  - Or are we trying to avoid really bad outcomes?

This…

Not this…

The use of information

- Dealing with parametric uncertainty
  - Not handled well in current DP paradigm

- Dealing with structural uncertainty
  - ASDP can explicitly deal with this via “information states”
  - Adds dimensionality and brings down The Curse

- Dealing with partial observability
  - Not handled properly in current ASDP approach
  - POMDP ?
  - Why the distinction between these 3 types of uncertainty?

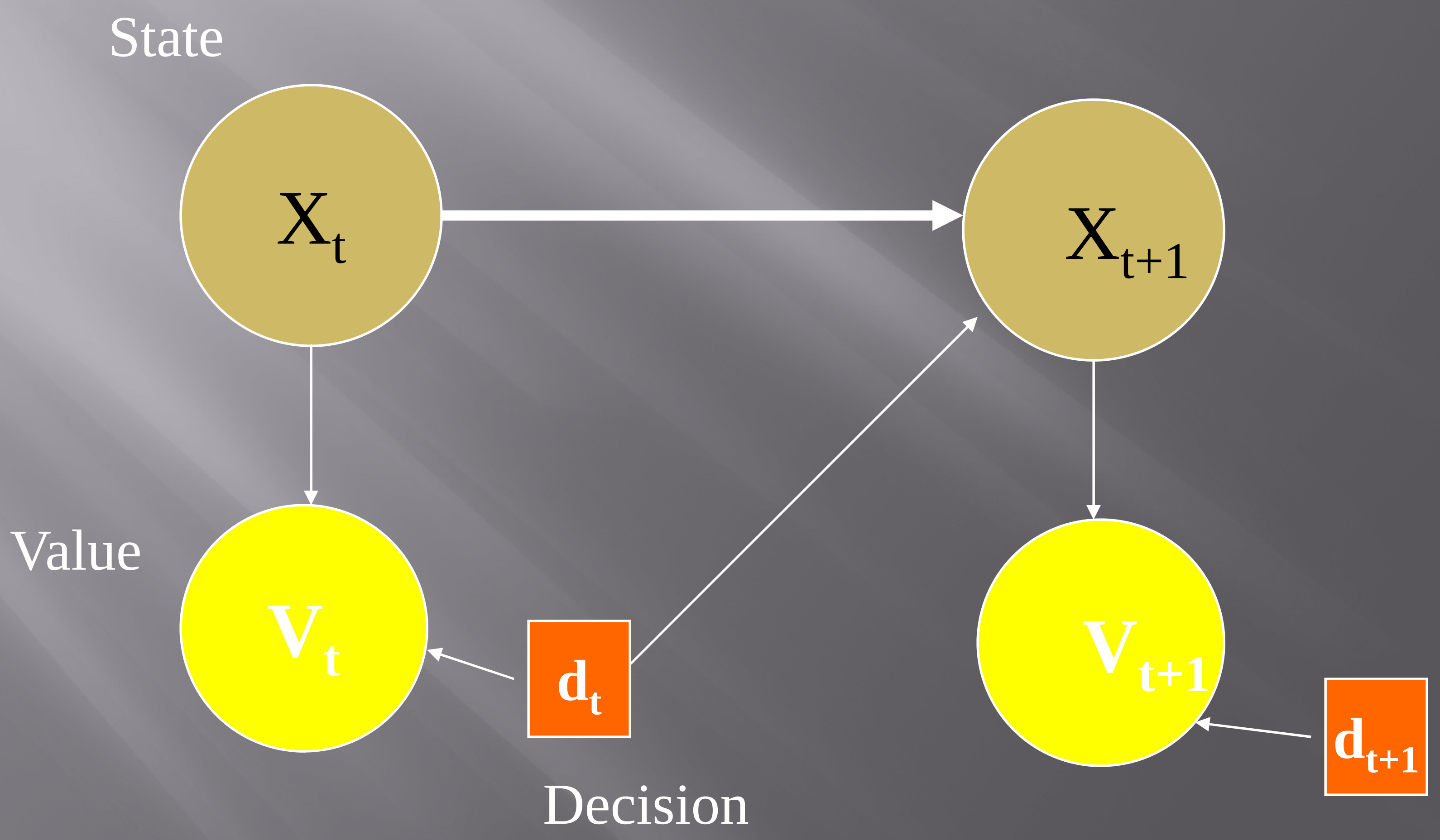
A Bayesian information / decision making paradigm

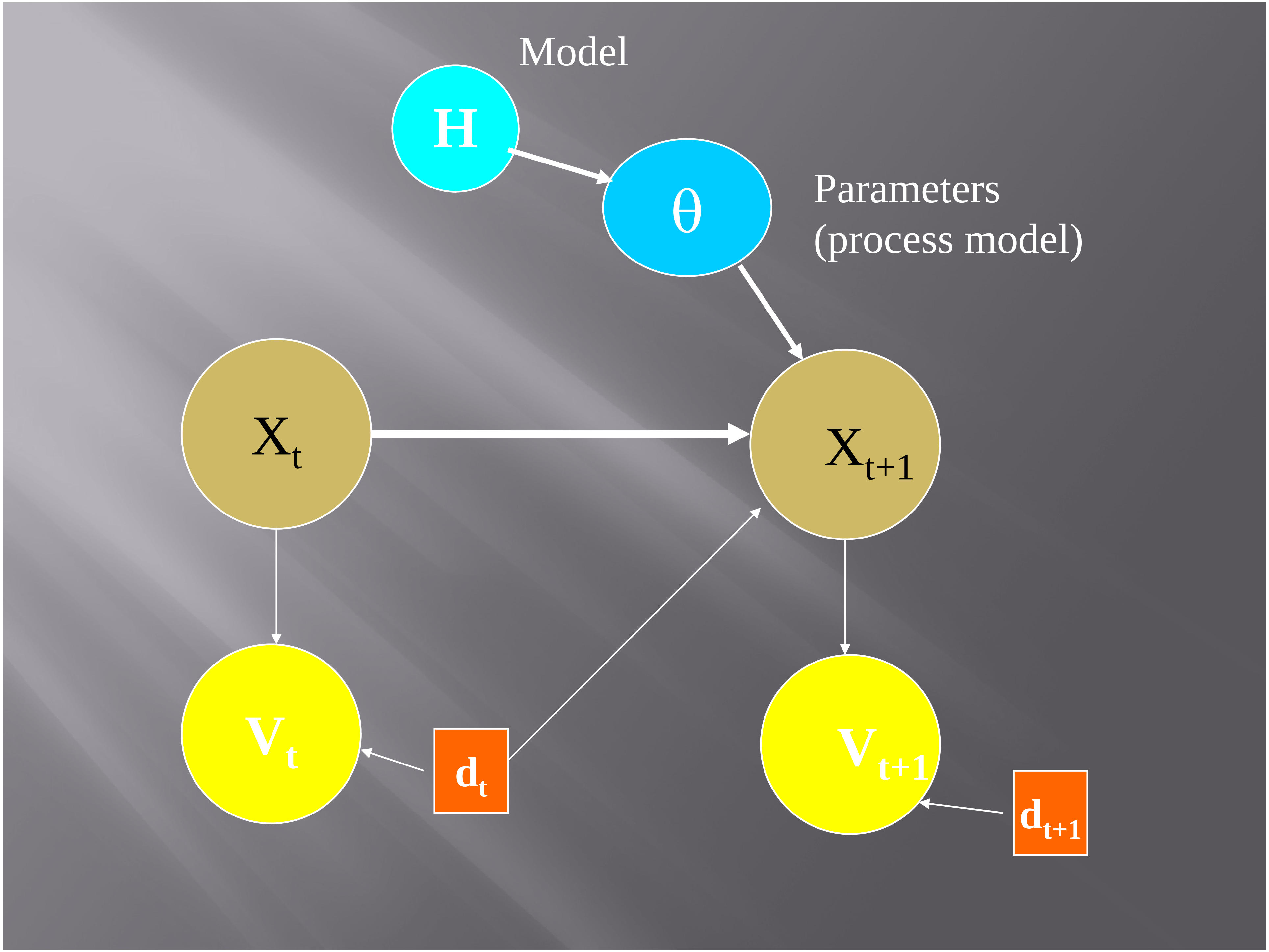
Utilization of information

- Current approach: Optimization and estimation/ adaptation are modeled separately

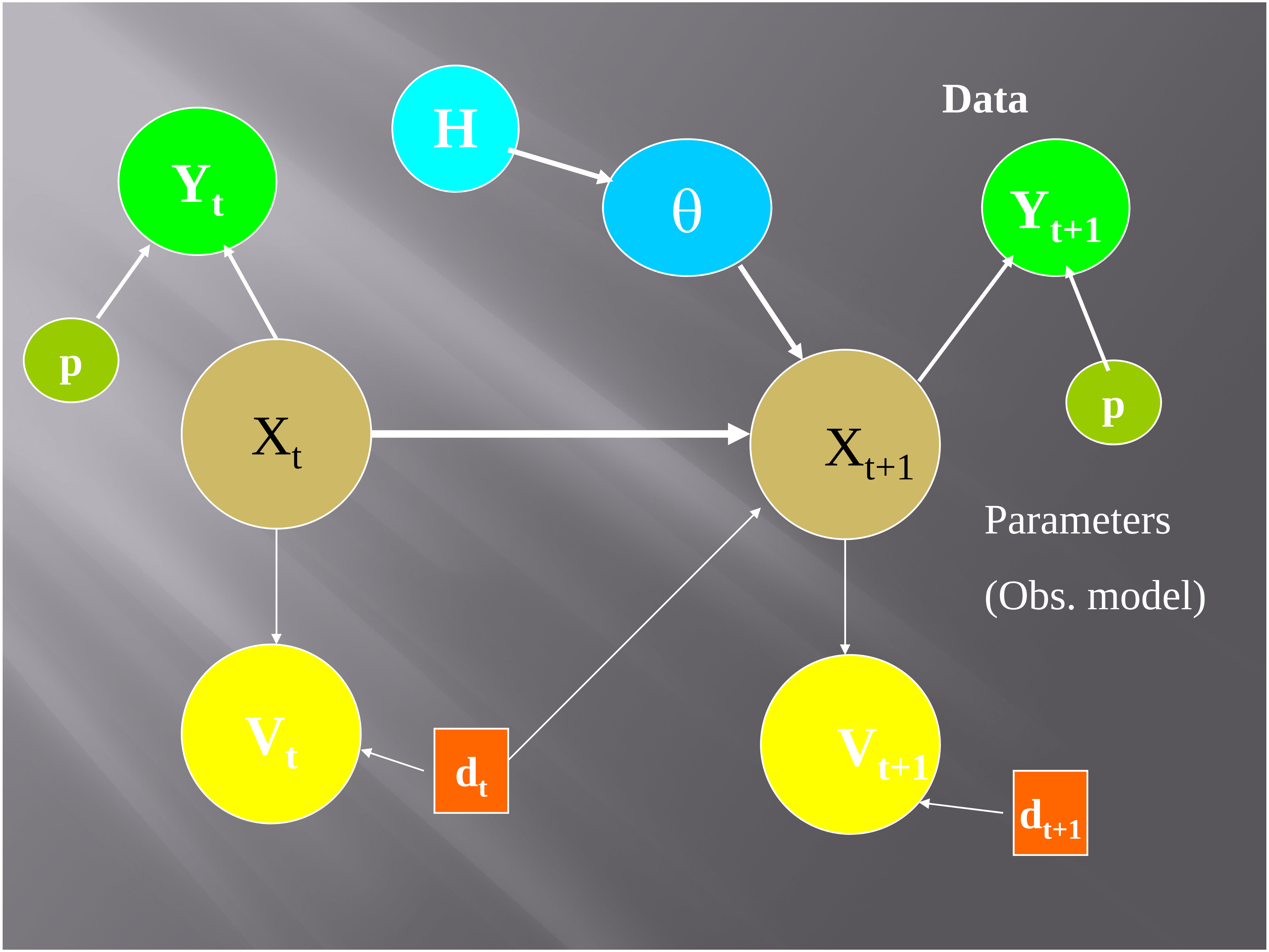
- Possible solution: Full Bayesian treatment of the Markov decision problem

Markov decision problem





Model
H
θ
Parameters
(process model)
X_t
X_{t+1}
V_t
V_{t+1}
d_t
d_{t+1}



Data
H
θ
Yt
Yt+1
p
p
Xt
Xt+1
Parameters
(Obs. model)
Vt
Vt+1
dt
dt+1

Joint distribution

Decision value

Parameters

State

$$[V(d), X, \Theta, p, H, Y]$$

Model

Data

Implementation?

- Combine Bayesian updating of parameters and information weights with RL updating
- Produce a joint trace of state-action pairs, rewards, parameter values, and model (information weights)

Thanks for listening