# Bayesian network modeling

# Probabilistic vs. deterministic modeling approaches

**Probabilistic**

↑ Explanatory power (e.g., $r^2$)
↓ Explanation why
Based on *inductive reasoning*

Good when you:
• Want to explore for patterns
• Want to see if real-world patterns conform to theory
• Have incomplete datasets or high uncertainty

Pitfalls (among others):
• Putting too much faith into patterns found in the data that lack a reasonable theoretical foundation

**Deterministic/mechanistic**

↑ Explanation why
↓ Explanatory power (e.g., $r^2$)
Based on *deductive reasoning*

Good when you:
• Want to test/understand why something works the way it does
• Have strong understanding of how something works

Pitfalls (among others):
• Sloppy model construction

# Use of Bayesian modeling in ARIES

- We used Bayesian Networks (BNs) in most of our early ARIES case studies
   (see http://aries.integratedmodelling.org/?page_id=546**)**
- Our current core global models are *not* BN models
- Early motivation to use BNs:
   - Great for expert elicitation, data-driven/inductive modeling
   - Account for uncertainty
   - Work well when data are incomplete or processes poorly known
- Current recommendation on using BNs:
   - Use physical/process models where those models are well known & trusted (Tier 1 & 2 models)
   - Use BNs for cases where you can take advantage of their strengths (Willcock et al. 2018, biodiversity modeling in Sicily; recreation & streambank erosion in Hawai'i)
   - Like any model, there's a time & place for BNs; know and use them then! (intelligent modeling)

# Recent examples of Bayesian models in ARIES



4
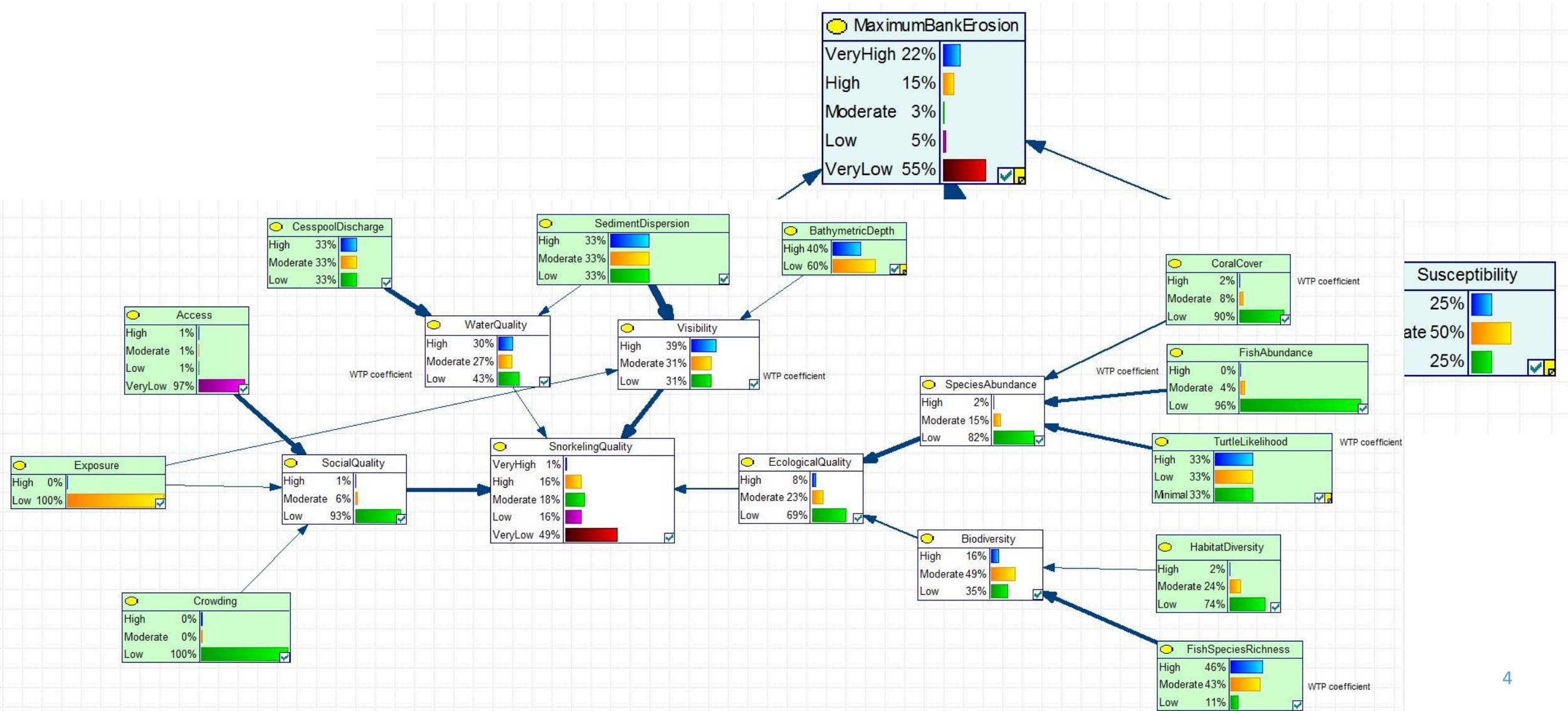
# Steps in a typical modeling process

1. Define system boundaries
2. Define model elements/variables
3. Build conceptual model
4. Identify potential feedback loops, thresholds, equilibria
5. Collect & prepare data to parameterize model
6. Formalize mathematical relationships
7. Testing, validation, calibration, sensitivity analysis

# Bayes' theorem

$$p(A|X) = \frac{p(X|A)*p(A)}{p(X|A)*p(A) + p(X|\sim A)*p(\sim A)}$$

How do we update the probability of A when we get new evidence, X?

# Bayesian Inference

**Experiment** Judy picks a jar at random, and then a cookie at random. The cookie is plain. What's the probability that Judy picked from jar #1?

**Prior probabilities** $P(J_1) = P(J_2) = 0.5$

**Event** $E$ = observation of plain cookie

**Conditional** $P(E|J_1) = 30/40 = 0.75$
**Probabilities** $P(E|J_2) = 20/40 = 0.50$

# Bayesian Inference

**Experiment** Judy picks a jar at random, and then a cookie at random. The cookie is plain. What's the probability that Judy picked from jar #1?

**Bayes Theorem**

$$P(J_1|E) = \frac{P(E|J_1)\,P(J_1)}{P(E|J_1)\,P(J_1)\ +\ P(E|J_2)\,P(J_2)}$$

**Posterior**

$$P(J_1|E) = \frac{0.75 \times 0.5}{0.75 \times 0.5\ +\ 0.5 \times 0.5} =$$

**Probability** = **0.6**

# Bayesian/probabilistic modeling

- Elements are assigned probabilities of occurrence (in the absence of data) – *conditional* and *prior* probabilities

- Data replace prior and conditional probabilities when available

- Provides results as a distribution of values without requiring stochiastic variables

# Uncertainty in deterministic models

- All else being equal (i.e., same input data & equations), you'll get the same results every time

- Change input parameters, use stochastic inputs & run repeatedly to generate a *distribution of results* (Monte Carlo simulation)

# Uncertainty in probabilistic models

Uncertainty estimates "built in" with prior probabilities & conditional probability tables
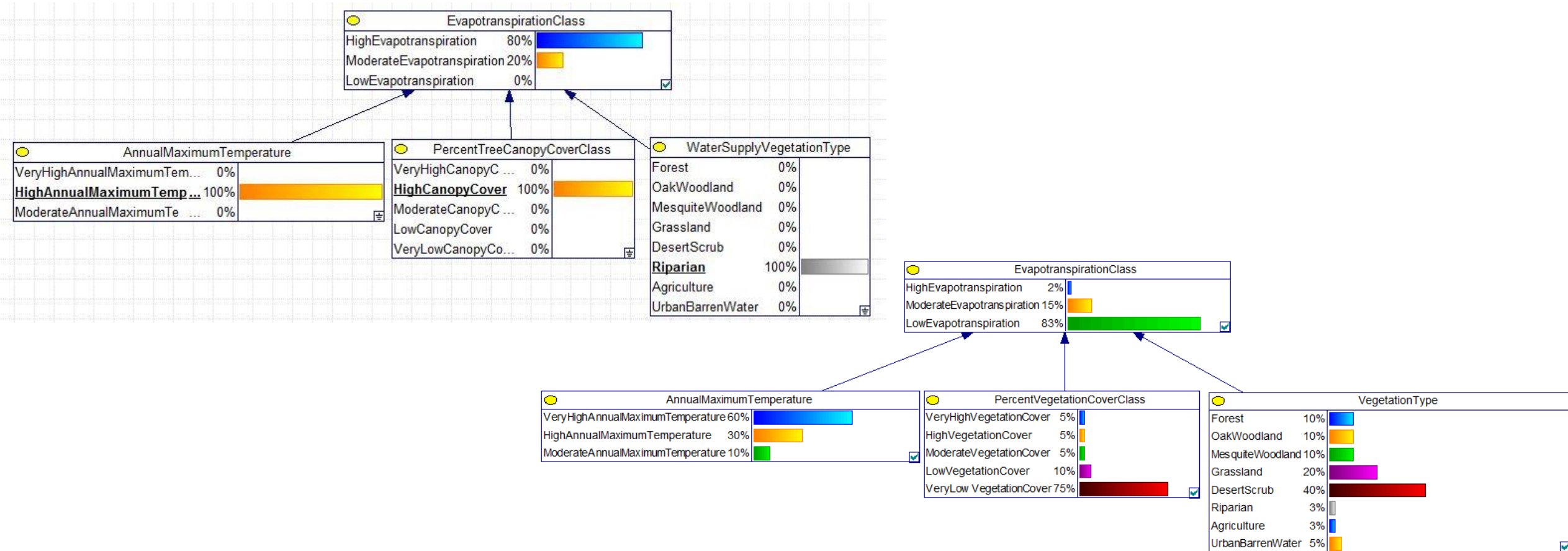
# Guidelines for Bayesian modeling
# (Marcot et al. 2006)

1. Develop causal model (i.e., influence diagram/directed acyclic graph)

2. Discretize each node

3. Assign prior probabilities

4. Assign conditional probabilities ("alpha-level model")

5. Peer review ("beta-level model")

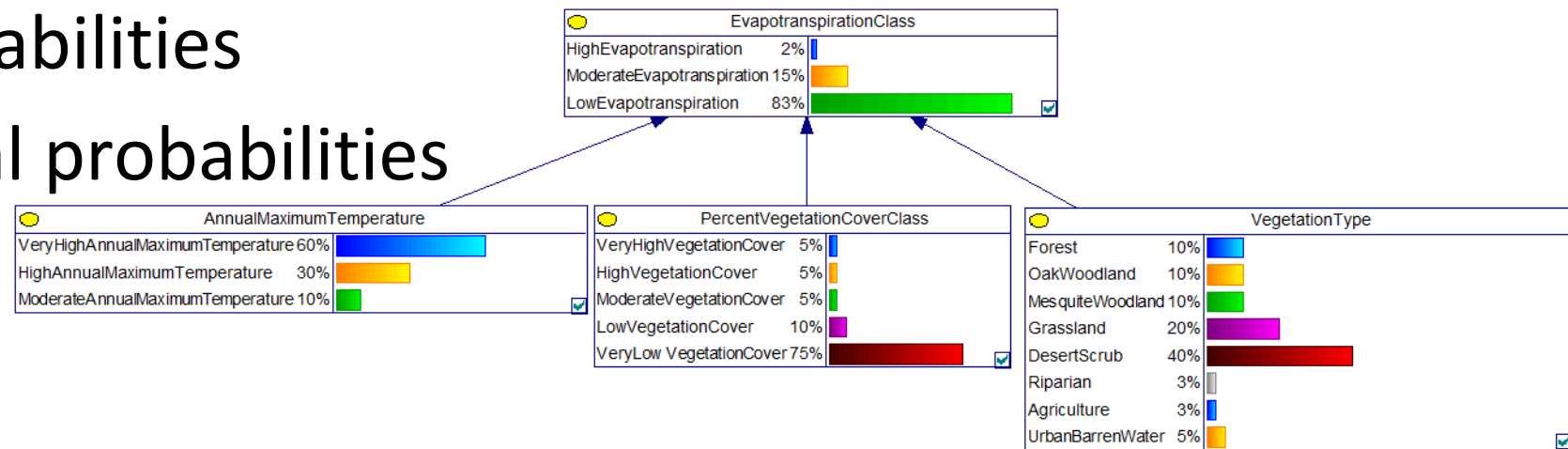6. Test with data and train the BN ("gamma-level model")

# General tips (Marcot et al. 2006)

- Keep # of input (parent) nodes & their # of discrete states tractable relative to each child node
- Role of intermediate variables
- Avoid unnecessarily "deep" models (problems with uncertainty propagation)
- Using training data
- CPTs: can use equations or "peg the corners;" potential role when thresholds are known

# Building the mathematical model:
# Probabilistic models

- Discretize variables

- Assign prior probabilities
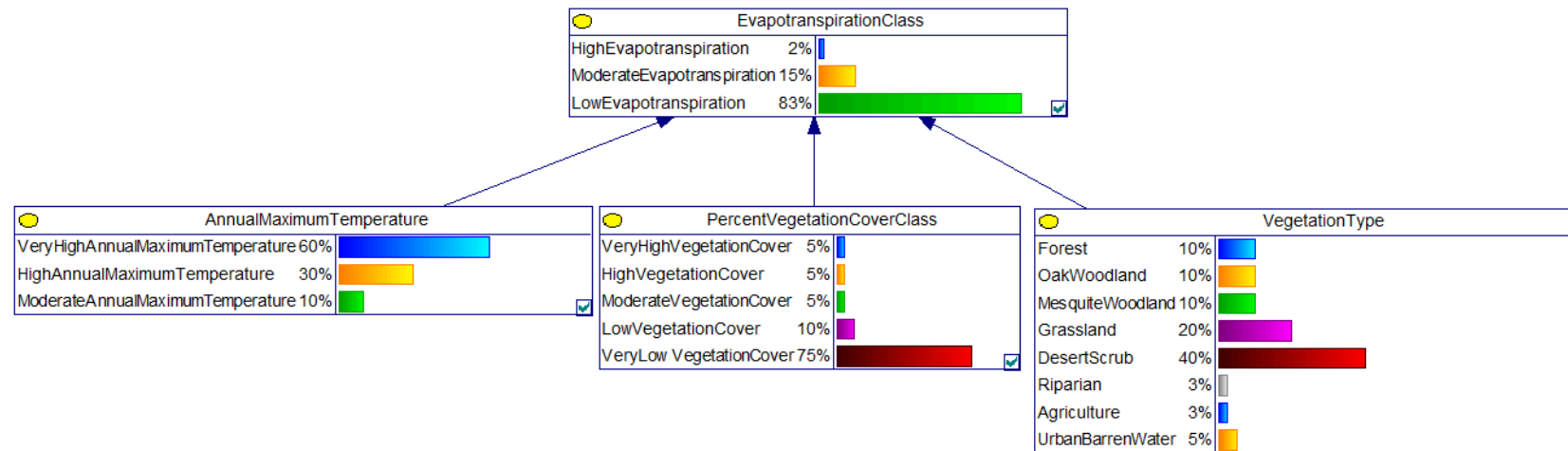
- Assign conditional probabilities

| | EvapotranspirationClass |
|---|---|
| HighEvapotranspiration | 2% |
| ModerateEvapotranspiration | 15% |
| LowEvapotranspiration | 83% |

| | AnnualMaximumTemperature |
|---|---|
| VeryHighAnnualMaximumTemperature | 60% |
| HighAnnualMaximumTemperature | 30% |
| ModerateAnnualMaximumTemperature | 10% |

| | PercentVegetationCoverClass |
|---|---|
| VeryHighVegetationCover | 5% |
| HighVegetationCover | 5% |
| ModerateVegetationCover | 5% |
| LowVegetationCover | 10% |
| VeryLow VegetationCover | 75% |

| | VegetationType |
|---|---|
| Forest | 10% |
| OakWoodland | 10% |
| MesquiteWoodland | 10% |
| Grassland | 20% |
| DesertScrub | 40% |
| Riparian | 3% |
| Agriculture | 3% |
| UrbanBarrenWater | 5% |

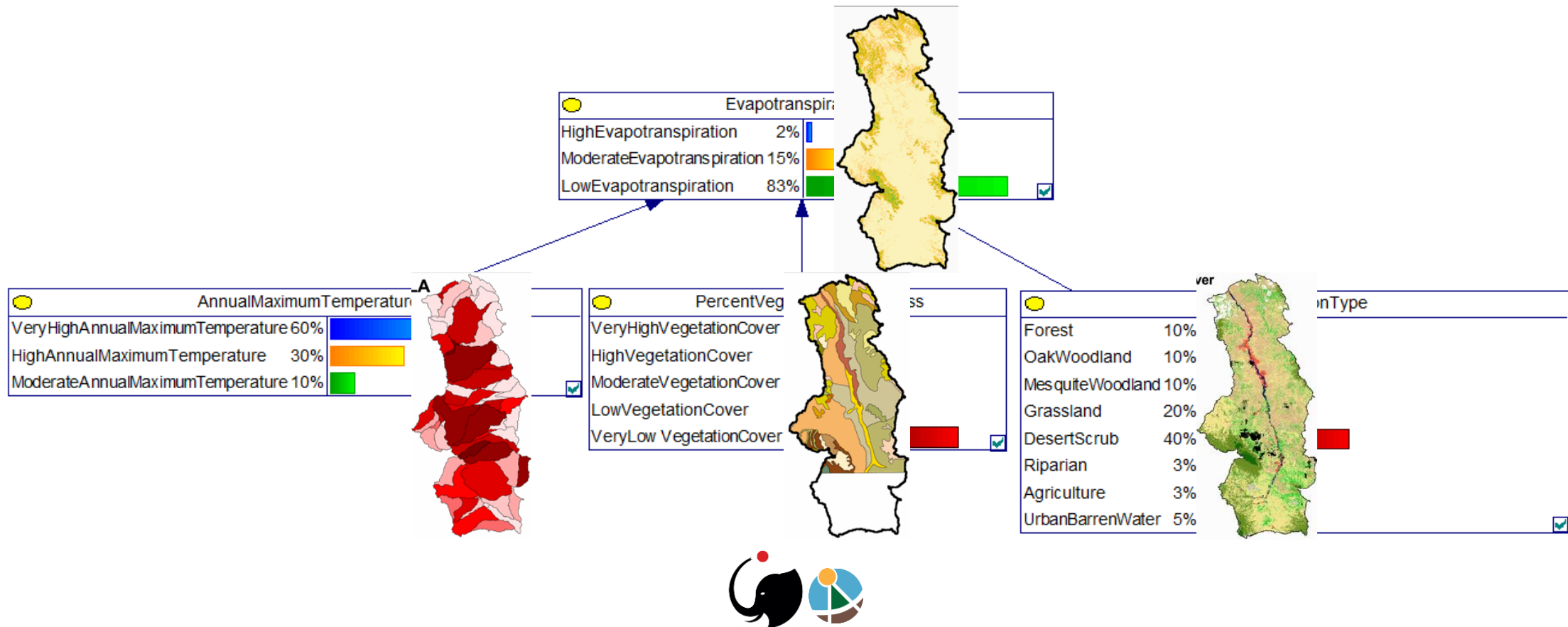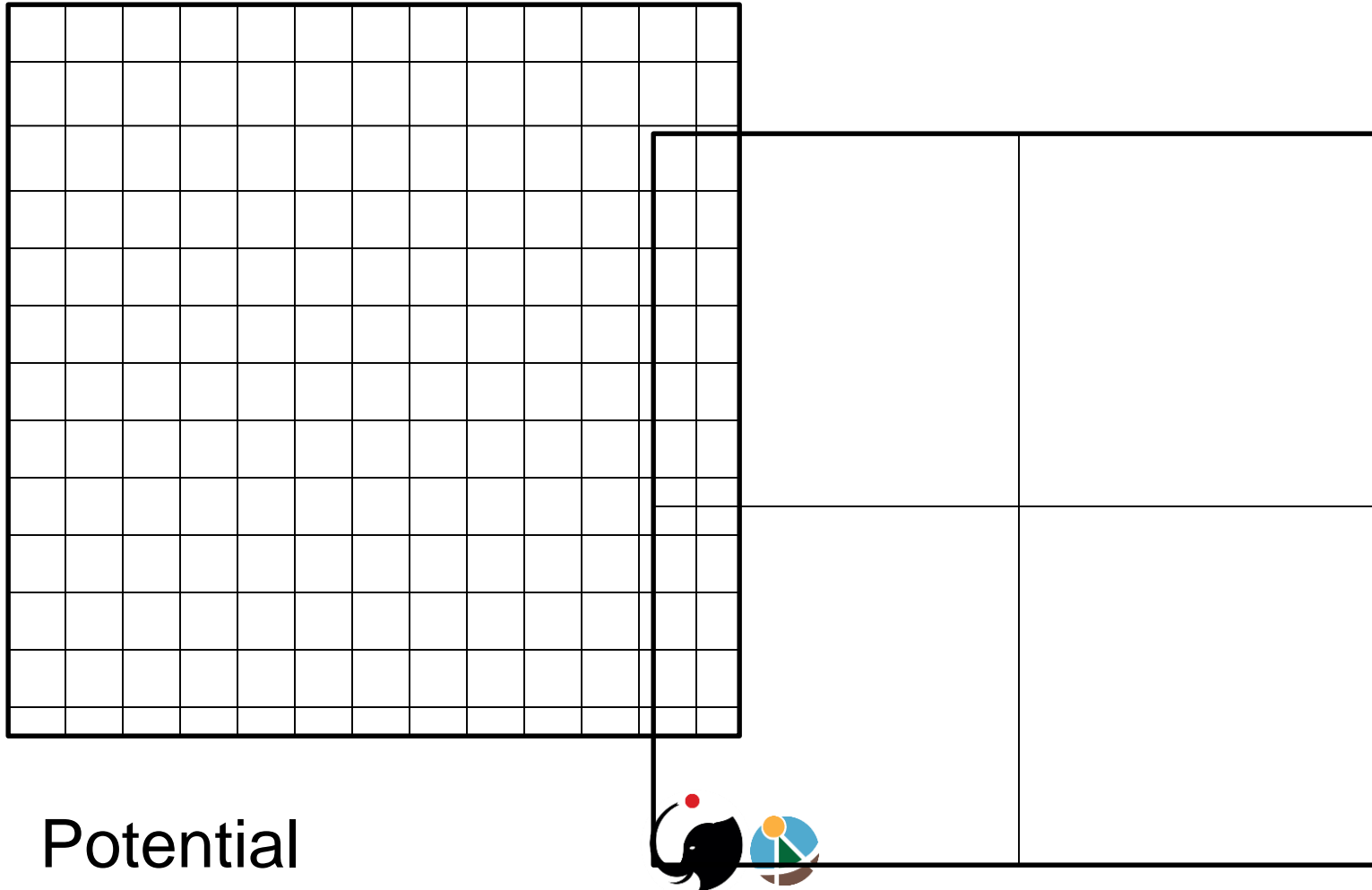| VegetationType | OakWoodland | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PercentVegetationCoverClass | VeryHighVegetationCover | | | HighVegetationCover | | | ModerateVegetationCover | | | LowVegetationCover | | | VeryLowVegetationCover | | |
| AnnualMaximumTemperature | VeryHighA... | HighAnnual... | ModerateA... | VeryHighA... | HighAnnual... | ModerateA... | VeryHighA... | HighAnnual... | ModerateA... | VeryHighA... | HighAnnual... | ModerateA... | VeryHighA... | HighAnnual... | ModerateA... |
| HighEvapotranspiration | 0.2 | 0.1 | 0.1 | 0.15 | 0.05 | 0.05 | 0.15 | 0.05 | 0.05 | 0.1 | 0.05 | 0.05 | 0 | 0 | 0 |
| ModerateEvapotranspiration | 0.7 | 0.7 | 0.6 | 0.7 | 0.75 | 0.7 | 0.6 | 0.6 | 0.5 | 0.45 | 0.4 | 0.3 | 0.4 | 0.3 | 0.2 |
| LowEvapotranspiration | 0.1 | 0.2 | 0.3 | 0.15 | 0.2 | 0.25 | 0.25 | 0.35 | 0.45 | 0.45 | 0.55 | 0.65 | 0.6 | 0.7 | 0.8 |

# Bayesian network training

- Bayesian training: Process where the system quantifies the relative contribution of parent nodes to child node in a BN
- User-specified CPT becomes much less relevant

| EvapotranspirationClass | |
|---|---|
| HighEvapotranspiration | 2% |
| ModerateEvapotranspiration | 15% |
| LowEvapotranspiration | 83% |

| AnnualMaximumTemperature | |
|---|---|
| VeryHighAnnualMaximumTemperature | 60% |
| HighAnnualMaximumTemperature | 30% |
| ModerateAnnualMaximumTemperature | 10% |

| PercentVegetationCoverClass | |
|---|---|
| VeryHighVegetationCover | 5% |
| HighVegetationCover | 5% |
| ModerateVegetationCover | 5% |
| LowVegetationCover | 10% |
| VeryLow VegetationCover | 75% |

| VegetationType | |
|---|---|
| Forest | 10% |
| OakWoodland | 10% |
| MesquiteWoodland | 10% |
| Grassland | 20% |
| DesertScrub | 40% |
| Riparian | 3% |
| Agriculture | 3% |
| UrbanBarrenWater | 5% |

# Bayesian network training

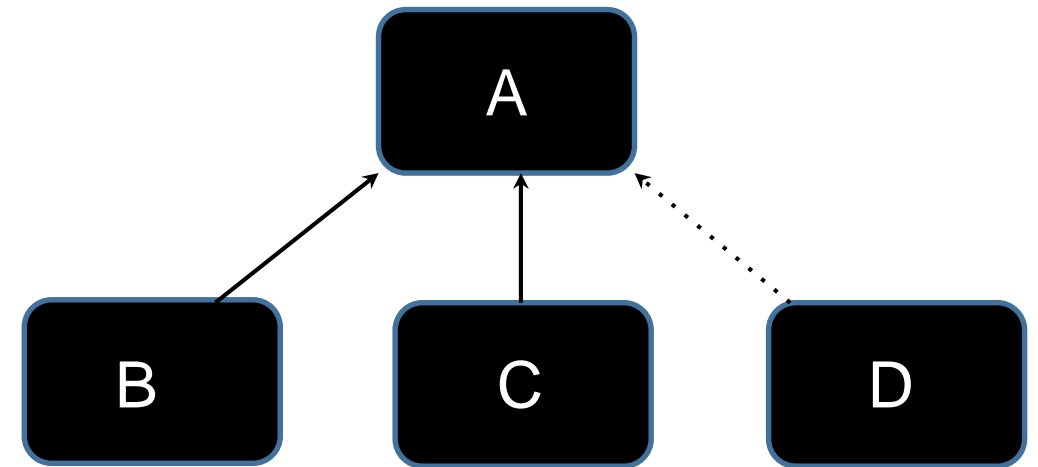# Spatial resolution & Bayesian network training

Potential problems?

# What if the system could determine the optimal model structure?

- Structural learning (see Willcock et al. 2018, "Machine-learning for ecosystem services," Ecosystem Services)
- Built into ARIES using Weka (more on this Friday)
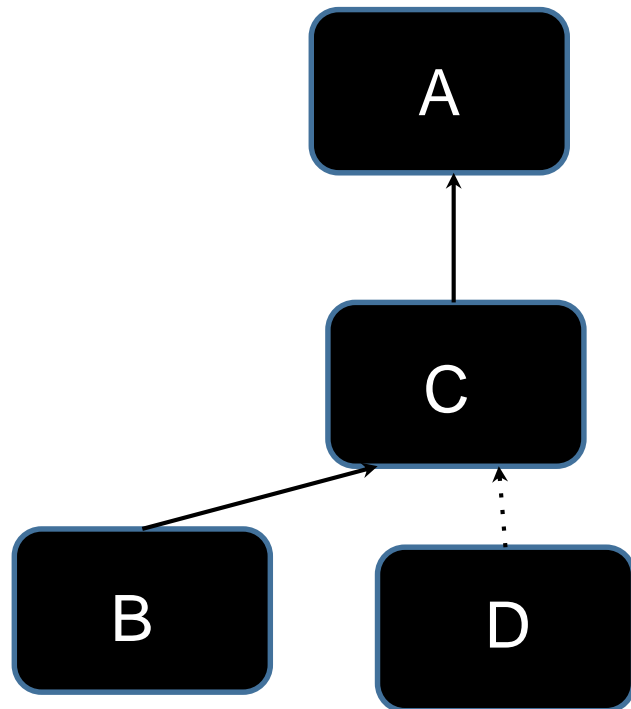
# Parting words

- Ockham's Razor/parsimony principle
- Start simple, continuously test the model, and add features/complexity slowly and carefully
- Keep your eye on the ball (original goals)
- Use best available data & assumptions
- Peer review is always valuable
- Document everything!

# For more information

- http://yudkowsky.net/rational/bayes
- Pearl, J. 1988. Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan-Kaufmann: San Francisco, CA.
- Marcot et al. 2006 & McCann et al. 2006 articles (distributed with course materials)

# On Bayesian modeling

"Some would argue that incorporating beliefs about models other than those implied by empirical measurement is a subjective, or unscientific, approach. In response, it could be stated that, certainly, Bayesianism has the potential for this problem to arise, and so one must have a strict 'code of conduct' for prior distribution specification. For example, making use of the outcomes of previous studies to provide prior beliefs is a reasonable scientific standpoint. Indeed, it could be argued that it is unscientific to ignore these prior results! Another way of avoiding subjectivity is to use non-informative priors in cases where prior information is unavailable or unobserved. Of course, one could argue that even a non-informative prior gives us some form of information about the distribution of an unknown parameter: after all, a specific distribution is being supplied rather than the information that any distribution might apply. However, in many cases non-informative priors do make reasonable models for a state of no subjective knowledge. In several 'text-book' examples of Bayesian analysis, for example multiple linear regression analysis assuming normal error terms, the adoption of non-informative priors results in tests algebraically identical to classical inferential procedures. In most cases, analysts are reasonably satisfied with regarding such classical approaches as 'objective'."

- Brundson & Willis 2002

# Bayes' theorem: cancer screening example

Convert the plain English to mathematical notation:

1% of women over 40 that are routinely screened have breast cancer

      $p(c) = 0.01$

80% of women with breast cancer test positive for cancer with a mammography

      $p(m+|c) = 0.8$

9.6% of women without breast cancer also test positive for cancer with a mammography (false positive)

      $p(m+|\sim c) = 0.096$

We want to know the likelihood of cancer, given a positive test

      $p(c|m+) = ?$

# Bayes' theorem: cancer screening example

$$p(c|m+) = \frac{p(m+|c)*p(c)}{p(m+|c)*p(c) + p(m+|\sim c)*p(\sim c)}$$

P (c|m+) = (0.8*0.01)/[(0.8*0.01) + (0.096*0.99)] = 0.07764 = 7.8%

How do we update the probability of c when we get new evidence, m+?