Tutorial on Influence Diagrams


Timothy C. Haas
School of Business Administration
University of Wisconsin-Milwaukee
haas@uwm.edu
Tutorial Outline

1. Define Influence Diagrams (IDs).

2. Review causal modeling with an ID and give a simple example.

3. Review two ecosystem management IDs.

4. Summarize how to estimate ID parameters
   (ID Learning I).

5. Summarize how to learn ID structure
   (ID Learning II).

6. Discuss sensitivity analysis for IDs.

8. Discuss available software and other resources for developing IDs.

9. Describe how any stochastic model can be embedded in an ID.

10. Describe a prototype cheetah viability ID and give an example of its use.

11. Conclude.
ID Definitions

- An ID is a graphical representation of a multivariate probability distribution. The representation is in the form of a mathematical construction called a graph.

- A graph is a pair $G = (V, E)$ where $V$ is a finite set of vertices and $E$ is a subset of the $V \times V$ ordered pairs of distinct edges. For $\alpha, \beta \in V$, if the edge $(\alpha, \beta)$ is in $E$ but $(\beta, \alpha)$ is not in $E$, then the edge is directed; otherwise, the edge is undirected.

- If $\beta \rightarrow \alpha$ then $\beta$ is a parent of $\alpha$ and $\alpha$ is a child of $\beta$. In a graph consisting of only directed edges, vertices having no parents are roots, and vertices having no children are terminal.
• A path of length $n$ from $\alpha$ to $\beta$ is a sequence $\alpha = \alpha_0, \alpha_1, \ldots, \alpha_{n-1}, \alpha_n = \beta$ of distinct vertices such that $(\alpha_{i-1}, \alpha_i)$ is in $E$ for $i = 1, \ldots, n$.

• A cycle of length $n$ is a path wherein $\alpha_0 = \alpha_n$. The cycle is directed if it contains at least one directed edge. A graph whose edge set contains only directed edges is called a directed graph. A graph for which there are no directed cycles is called an acyclic graph.
• Kiiveri et al. (1984): the product of the unconditional and conditional distributions of a set of random variables indexed by the vertex set of a directed acyclic graph defines a joint probability distribution as long as this product is positive for all joint events.

• Random variables indexed by root vertices are assigned unconditional distributions, all other random variables are assigned conditional distributions for each possible event of the variables indexed by that variable's parent or parents. This joint distribution $P$ then, is said to recursively factorize according to the graph $G$.

• The ID model architecture helps to clarify variable dependencies, random components, and possible system control points. By conditioning on desired sets of control variables, the effect of different interventions can be explored.
• Previous work on including quantitative variables in an ID have focused on gaussian variables. The most developed of these models is the Conditional Gaussian model of Lauritzen and Wermuth (1989).

• **Example:** If the vertices in the above graph index the random variables $X_1$, $X_2$, and $X_3$, then

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = P(X_3 = x_3 | X_1 = x_1, X_2 = x_2)P(X_2 = x_2 | X_1 = x_1)P(X_1 = x_1).$$

• Two advantages of this graph theoretic representation of a joint distribution:

1. The above results allow the construction of a large, complicated joint distribution by specifying a set of much simpler, local unconditional and conditional distributions.

2. The directed graph provides in one picture, all probabilistic dependencies.
• An ID represents random variables as circles, decisions as squares, and value variables as diamonds. An ID having only random variables is often called a Bayesian Belief Network (BBN).

• Hereafter, any variable represented within an ID will be referred to as a node.
Environmental BBN Example

- SHADE: amount of shade present in an aspen stand – takes on the values shade, (0 to 25% full sunlight), and full sunlight (90 to 100% sun).

- MOISTURE: soil moisture of the stand – takes on the values dry, (5 to 25% soil moisture content by volume), and moist (25 to 50% moisture content).

- SITECNDs: qualitative measure of physiographic site conditions for aspen sucker growth response – takes on the values unfavorable, and favorable conditions.

- Causal relationships represented: amount of shade and amount of soil moisture independently influence the site conditions for sucker growth.
• Unconditional distributions for root nodes and conditional distributions for other nodes that need to be specified:

\[ P(\text{SHADE}), \]
\[ P(\text{MOISTURE}), \text{ and} \]
\[ P(\text{SITECNDS}|\text{SHADE, MOISTURE}). \]

• Substantive science represented by the distributions (see Haas (1991a)):

1. Aspen sucker growth is vigorous under the conditions of full sunlight and moist soil.

2. Suckers will tolerate dry conditions in full sunlight but heavy shade will severely restrict sucker growth no matter what the moisture content is.

3. Under heavy shade and dry soil, sucker growth is all but curtailed.
\begin{align*}
P(\text{favorable}|\text{shade, dry}) &= 0.05 \\
P(\text{unfavorable}|\text{shade, dry}) &= 1.0 - 0.05 \\
P(\text{favorable}|\text{shade, moist}) &= 0.1 \\
P(\text{unfavorable}|\text{shade, moist}) &= 1.0 - 0.1 \\
P(\text{favorable}|\text{full sun, dry}) &= 0.7 \\
P(\text{unfavorable}|\text{full sun, dry}) &= 1.0 - 0.7 \\
P(\text{favorable}|\text{full sun, moist}) &= 0.9 \\
P(\text{unfavorable}|\text{full sun, moist}) &= 1.0 - 0.9. \\
\end{align*}

- For example, the 0.05, 0.95 distribution is intended to represent very little conditional belief in a favorable site condition given shade and dry soil.

- These numbers are the IDs parameters and are stored in a Conditional Probability Table (CPT).
• Rationale for root node distributions.

SHADE: assume that this particular stand has been recently clearcut, but that due to a vigorous shrub and forb population, the belief that full sunlight reaching aspen suckers is only .8, implying a .2 belief that the suckers are in shade.

MOISTURE: based on soil moisture meter measurements and an examination of the types of forbs and shrubs growing on the stand, a belief of .9 can be assigned to the soil moisture value of 0-25% soil moisture; i.e., there is a .9 belief on the part of the individual examining the stand that the soil moisture is less than 25%. This implies that there is only .1 belief that this is a moist site.
• Combining the above conditional and unconditional distributions gives the joint
distribution for this three-node BBN:

\[
P(\text{SITECNDS, SHADE, MOISTURE}) \\
\quad = \quad P(\text{SITECNDS}|\text{SHADE, MOISTURE}) \\
\quad \times \quad P(\text{SHADE})P(\text{MOISTURE}).
\]

• For example, the overall or marginal belief in the site condition value of \textit{favorable}
implied by the above unconditional and conditional distributions is found by com-
puting:
\[ P(\text{SITECNDS} = \text{favorable}) = \sum_{x, y} P(\text{SITECNDS} = \text{favorable}, \ \text{SHADE} = x, \ \text{MOISTURE} = y) \]

\[ = P(\text{SITECNDS} = \text{favorable}|\text{full sun, moist soil}) \times P(\text{full sun})P(\text{moist soil}) \]
\[ + P(\text{SITECNDS} = \text{favorable}|\text{shade, moist soil}) \times P(\text{shade})P(\text{moist soil}) \]
\[ + P(\text{SITECNDS} = \text{favorable}|\text{full sun, dry soil}) \times P(\text{full sun})P(\text{dry soil}) \]
\[ + P(\text{SITECNDS} = \text{favorable}|\text{shade, dry soil}) \times P(\text{shade})P(\text{dry soil}) \]

and is

\[ = (0.9 \times 0.8 \times 0.1) + (0.1 \times 0.2 \times 0.1) \]
\[ + (0.7 \times 0.8 \times 0.9) + (0.05 \times 0.8 \times 0.9) \]
\[ = 0.614. \]
Since SITECNDS is dichotomous,

\[ P(\text{SITECNDS} = \text{unfavorable}) \]
\[ = 1 - P(\text{SITECNDS} = \text{favorable}) \]
\[ = 0.386. \]

- Overall, the model gives a .61 belief in this particular set of site conditions being favorable for aspen growth.
• This belief can be interpreted as a judgment of the site’s physiographic favorability for aspen sucker growth derived from expert opinion of aspen sucker growth sensitivities and the particulars of the stand under consideration.

• The value can be used as input to management decisions concerning what sites should be singled out as having high potential for aspen sucker growth.

• See Haas (1991a) for a more complete model of aspen stand sucker response.
• Finding all marginal probabilities of a BBN is called “solving the BBN.”

• Finding the optimal decision represented by an ID requires the expected value of the utility node be computed for each combination of decision nodes. The combination that maximizes this expected value is the optimal decision. Performing these computations is called “solving the ID.”

• Solving a BBN or ID is NP-hard.

• Exact methods based on graph manipulations are used in many software systems, e.g. Hugin and Netica.

• Simulation can also be used to approximately solve a BBN or ID. logic sampling (Henrion 1988) is the simplest of these methods. Analytica and my system (id) uses simulation to solve BBNs and IDs.
Modeling Causal Relationships
with an ID (Pearl 1995)

A mathematical language to express causal relationships can be derived from properties of an ID.

d-separation: Let \( X, Y, \) and \( Z \) be three disjoint subsets of nodes in a directed acyclic graph \( G \), and let \( p \) be any path between a node in \( X \) and a node in \( Y \). Then \( Z \) is said to block \( p \) if there is a node \( w \) on \( p \) satisfying one of the following two conditions: (i) \( w \) has converging arrows along \( p \), and neither \( w \) nor any of its descendants are in \( Z \), or (ii) \( w \) does not have converging arrows along \( p \), and \( w \) is in \( Z \). Further, \( Z \) is said to d-separate \( X \) from \( Y \), in \( G \), written \( (X \perp Y | Z)_G \), if and only if \( Z \) blocks every path from a node in \( X \) to a node in \( Y \).
Causal Effect: Given two disjoint sets of nodes, $X$ and $Y$, the causal effect of $X$ on $Y$, denoted $P(Y|\bar{x})$, is a function from $X$ to the space of distributions on $Y$. For each realization $\bar{x}$ of $X$, $P(Y|\bar{x})$ gives the probability of $Y = y$ induced by deleting from the graph all nodes that are parents of nodes in $X$ and substituting $\bar{x}$ for $X$ in the remainder.

the $\bar{\cdot}$ notation means that these values for $X$ are intentionally fixed.
**Identifiability:** The causal effect of $X$ on $Y$ is said to be identifiable if the quantity $P(Y|\tilde{x})$ can be computed uniquely from any positive distribution of the observed nodes that is compatible with $G$.

**Theorem:** Computing $P(Y = y|\tilde{x})$.

**Case 1:** If a set of nodes $Z$ satisfies a condition called the *back-door criterion* relative to $(X, Y)$, then the causal effect of $X$ on $Y$ is identifiable and is given by the formula

$$P(Y = y|\tilde{x}) = \sum_{\tilde{z}} P(Y = y|\tilde{x}, \tilde{z}) P(Z = \tilde{z}).$$
A set of nodes $Z$ satisfies the back-door criterion relative to an ordered pair of nodes $(X_i, X_j)$ in a directed acyclic graph $G$ if: (i) no node in $Z$ is a descendant of $X_i$ and (ii) $Z$ blocks every path between $X_i$ and $X_j$ which contains an arrow into $X_i$. If $X$ and $Y$ are two disjoint sets of nodes in $G$, $Z$ satisfies the back-door criterion relative to $(X, Y)$ if it satisfies it relative to any pair $(X_i, X_j)$ such that $X_i \in X$ and $X_j \in Y$. 
Back-Door Criterion Example

\{X_3, X_4\} and \{X_4, X_5\} satisfy the Back-Door criterion for the ordered pair \(X_i, X_j\).

\{X_4\} does NOT because \(X_4\) does not block the path \((X_i, X_3, X_1, X_4, X_2, X_5, X_j)\) because there are converging arrows onto \(X_4\) along this path.
Case 2: (Front-door) If a set of nodes $Z$ satisfies the following conditions relative to an ordered pair of nodes $(X, Y)$: (i) $Z$ intercepts all directed paths from $X$ to $Y$, (ii) there is no back-door path between $X$ and $Z$, and (iii) every back-door path between $Z$ and $Y$ is blocked by $X$. Then the causal effect of $X$ on $Y$ is identifiable and is given by the formula

$$P(Y = y | \bar{x}) = \sum_{\bar{z}} P(Z = \bar{z} | \bar{x}) \sum_{\bar{z}'} P(Y = y, \bar{z}') P(X = \bar{x}').$$
Front-Door Criterion Example

\[ \begin{align*}
U & \rightarrow X_1 \\
U & \rightarrow X_2 \\
X_2 & \rightarrow X_3
\end{align*} \]

$U$ is unobserved.

$X_2$ satisfies the Front-Door criterion for the ordered pair $X_1, X_3$.

- The Theorem allows many apparently useless nonexperimental samples to be used for discovering causal relationships.
- Ecosystem managers often have many such data sets.
Advantages of IDs in Natural Resource Management

- Domain experts can quickly learn how to build complex stochastic models of ecosystems.

- The graphical representation allows more intuitive model construction and more effective communication of the model.

- Direct assignment of the CPT’s allows stochastic models to be built of processes that lack other mathematical models such as systems of differential equations.
• Uncertainty about data, relationships, and outcomes can be made an integral part of the model from the beginning.

• Interventions and management activities can be explicitly represented with decision nodes.

• The value or utility of different future ecosystem states (e.g., extinction/non-extinction) can be modeled with utility nodes.
Two Applications of BBNs
to Ecosystem Management

Waterbody Eutrophication

- Reckhow (1999) describes an ID for managing eutrophication (and hence the possibility of a fishkill) of an estuary.

- Percent-Forested-Buffer is the management plan decision node and Onset-of-an-Anoxia-Condition (lack of oxygen in the water) is the utility node.
Wildlife Population Viability

- Marcot et al. (2000) describes a BBN approach to modeling fish and wildlife population viability as affected by land management and development to support management decisions by the USDA Forest Service as part of the Interior Columbia Basin Ecosystem Management Project (ICBEMP).

- Key Environmental Correlates (KEC’s) are defined by experts to be the characteristics most important to the success of the population.

- Proxy nodes computed from GIS maps are used to “cause” (indicate) the degree to which these KEC’s are present.

- Habitat suitability nodes are “caused” by these KEC’s and population response is “caused” by habitat suitability.
ID Parameter Estimation  
(ID Learning I)

- Estimating ID parameters (conditional probability tables) from data is the more-understood aspect of “learning” an ID.

- Currently, the most computationally efficient frequentist method for estimating the parameters of an influence diagram consisting solely of qualitative chance nodes from an incomplete sample is the EM(η) (for Expectation – Maximization with learning rate η) method of Bauer et al. (1997). Zhang (1996) is also able to speed up EM convergence.

- Also for discrete IDs, the most efficient bayesian method appears to be the Bound and Collapse (BC) method of Ramoni and Sebastiani (1997) although the Heckerman et al. (1994) method may be almost as efficient.
• Cheesman and Stutz (1997) give a bayesian method for fitting a mixed ID (some
discrete and some continuous-valued nodes) with missing data (equivalently, latent
nodes). A classification task is assumed in that each observation is assumed to have
come from a particular class. Neither these class assignments nor the number of
these classes are known and are to be estimated from the data.

• A software system (Autoclass) is available and can be viewed as a bayesian data
mining system.
All of these bayesian methods use a Dirichlet prior distribution. This is appropriate since ID parameters are probabilities and a Dirichlet distribution is a multivariate Beta distribution wherein each component has support \((0, 1)\).
ID Model Selection
(ID Learning II)

• Construction of a BBN shares the knowledge acquisition bottleneck common to any effort to build a decision support system.

• A solution would be to automate the construction of the BBN by applying a BBN structure learning algorithm to a set of observations.

• To do this, one must evaluate the “goodness” of a large number of candidate graphs (models) and decide on a structure that best represents dependencies in the data.

In other words, one must first define a score function that gives a score to each model and then use this score function to search the space of possible models and then select the model with the highest score value.
• Also called “structure learning.” Equivalent to model selection in statistics.

• Structure learning is much more difficult than estimating parameters due to the combinatoric explosion of the number of candidate models.

• Many score functions have been devised, both frequentist and bayesian, see Castillo et al. (1997, ch. 11).
• If prior probabilities on candidate models can be obtained, use the bayesian score function. Let $S$ be a sample, $G$ be a candidate network having $p$ nodes, and $\theta$ be the parameters defining the conditional and unconditional distributions in $G$. Then,

$$\text{Bayesian Score} \equiv \log(P(G)) + \log(P(S|G, \hat{\theta}))$$

$$- .5\text{Sizeof}(G)\log(\text{Sizeof}(S)).$$

The third term penalizes for complex models.
• If prior probabilities are difficult to justify, use the Minimum Description Length (MDL) criterion (see Castillo et al. (1997, p. 509). Let $v_i$ be the number of discrete values for node $i$ and $c_i$ be the number of unique combinations of values of node $i$’s parents. Then,

$$\text{MDL Score} \equiv \sum_{i=1}^{p} \sum_{j=1}^{v_i} \sum_{k=1}^{c_i} N_{ijk} \log(N_{ijk}/N_{ik})$$

where for each node $i$, $N_{ik}$ is the number of observations node $i$ with parent-value combination $k$ and $N_{ijk}$ is the number of observations on node $i$ with value $j$ under parent-value combination $k$. 
• **Recommendation:** use the K2-AS method of Provan and Singh (1997) because of its emphasis on finding a model that both has a high score and high predictive accuracy.

• Symbol Definitions:

  1. Consider $p$ discrete-valued nodes, $X_1, \ldots, X_p$, each measured on an experimental unit and $n$ experimental units observed to make up the sample, $S$.

  2. Partition $S$ into a training, testing, and evaluation subsamples, $S_{train}$, $S_{test}$, and $S_{eval}$, respectively.

  3. Let $Z$ be the set of the nodes. Let $\Delta$ be a subset of $Z$. 
• Definition of conditional independence test:

For nodes $a$ and $b$, find the set of nodes $C_{ab}$ for which $a$ is independent of $b$ given values on all nodes in $C_{ab}$. Do this by performing a $\chi^2$ test for conditional independence of $a$ and $b$ given a macro node $a_m$ where $a_m$ is a discrete node having values that are all the unique combinations of values of the nodes in $C_{ab}$.
• Definition of the structure learning algorithm, K2:

1. Given an ordering of the nodes, start with the first node. Temporarily make each remaining node a parent of $X_0$. Assign as a permanent parent to $X_0$, the node that makes the 2-node network have the largest score function. The score function used is the joint probability of the sample which is proportional to the probability of the model given the sample.

2. Repeat the above procedure on every node in the network as long as the score function of the new network is bigger than the previous network.

3. Do a greedy search for parents by only adding one candidate parent node at a time.
- Definition of CB: Use the K2 learning algorithm with the node ordering found with conditional independence tests.

- Definition of K2-AS: (K2 with Attribute (node) Selection):

  1. Variable Selection Phase:

     i. Divide the sample into $S_{train}$, $S_{test}$, and $S_{pred}$.

     ii. Add $X$ to $\Delta$ if this node creates a CB-built network using $X \cup \Delta$ nodes that has a uniquely higher predictive accuracy than any other node in $Z - \Delta$. Use $S_{train}$ to build each trial network and $S_{test}$ to compute the predictive accuracy of each of the networks.

     iii. Continue to add nodes in this manner to $\Delta$ until no one node creates a network that has a uniquely higher predictive accuracy.
2. Network Learning Phase:

Use only the nodes in $\Delta$ to create a network with CB using $S_{\text{train}}$.

3. Predictive Accuracy Phase:

Compute the predictive accuracy of the phase 2 network using $S_{\text{pred}}$.

ID Sensitivity Analysis

- Three main types of sensitivity analysis:

  1. Parameter sensitivity due to either (a) expert misspecification or (b) variability of estimated parameter values.

  2. Sensitivity of marginal probabilities to addition or deletion of dependency links in the ID.

  3. Variability of marginal probability distribution of outcome nodes.

- Parameter sensitivity has been studied in Pearl (1988), Haas (1991a), and Laskey (1993).
Pearl (1986) notes that the natural tendency of the effect of a node to be attenuated on nodes multiple links removed and cites this characteristic as an example of a BBNs stability:

The addition of any new node $x_i$ to the network requires only that the expert identify a set $S_i$ of nodes which "directly influence" $x_i$, locally assess the strength of this relation and make no commitment regarding the effect of $x_i$ on other nodes, outside $S_i$. Even though each judgment is performed locally, their sum total is guaranteed to be consistent. This model-building process permits people to express qualitative relationships perceived to be essential, and the network preserves these qualities, despite sloppy assignments of numerical estimates.
Haas (1991a) derives sensitivity ratios for particular cases. For example: $\beta$ parameterizes a parent of $X_2$ in a 2-node network. Let $P(X_1 = 0) = \beta_1$, $P(X_2 = 0|X_1 = 0) = \beta_2$, and $P(X_2 = 0|X_1 = 1) = \beta_3$. It is desired to study the effect of changes in $\beta_1$ on the marginal probability of $X_2$ ($P(X_2 = 0)$). Note,

$$
P(X_2 = 0) = P(X_2 = 0|X_1 = 0)P(X_1 = 0) + P(X_2 = 0|X_1 = 1)P(X_1 = 1)
$$

$$
= \beta_2\beta_1 + \beta_3(1 - \beta_1)
$$

$$
= \beta_1(\beta_2 - \beta_3) + \beta_3
$$

$$
= \beta_1a_2 + \beta_3
$$

where $-1 < a_2 < 1$. Hence, $\delta(\epsilon) = \epsilon a_2$. The sensitivity of the BBN to small errors in a parameter, now one dependency link removed, are less than the changes in the parameter itself.

Drawbacks to IDs

- Large number of parameters: can overparameterize, expensive to estimate.

- Difficult to represent time.

- Difficult to represent feedback.

- Currently, the only well-known extension to BBNs that allow continuous-valued nodes is the mixed, conditional Gaussian BBNs of Lauritzen. These networks suffer from being highly dependent on a very specific assumption: conditional multivariate normality.

- Although BBNs are ideal for structuring knowledge domains where there are few mathematical models, little work has been done on extending BBNs to model non-Gaussian and/or stochastic model-generated random nodes (but see below).
Software and Other Resources for IDs

• Software (see handouts)

• web tutorials: (see handouts)

• People.

  1. Marek J. Druzdzel (www.pitt.edu/~druzdzel): BBN causality, reasoning with BBNs.

  2. David Heckerman, Machine Learning and Applied Statistics Group (MLAS), Microsoft Research, Microsoft Corp. (heckerma@microsoft.com): learning BBNs.
3. Kenneth H. Reckhow, Director, The Water Resources Research Institute, North Carolina State University, and Professor, School of the Environment, Duke University (ken_reckhow@ncsu.edu, reckhow@duke.edu): BBNs applied to water quality.

4. Bruce G. Marcot, Research scientist, USDA Forest Service, Portland, Oregon (brucem@spiritone.com): applications of IDs to wildlife management.
IDs for Complex Stochastic Systems

- **statistical models**: stochastic models built solely for purposes of prediction and not derived from physical process relationships

- **mechanistic models**: models based at least in part on scientific theory of how the phenomenon’s state nodes are related to each other.

- Mechanistic models can be based on systems of Ordinary Differential Equations (ODE)’s or *Stochastic* Differential Equations (SDE)’s.

- The conditional distribution of a set of dependent nodes can be a parametric multivariate distribution such as a multivariate Beta, or implicit as outputs from a complex multivariate stochastic model.
• An ID provides a unified structure into which models of quantitative and/or qualitative system state nodes can be placed.

• Most mathematical mechanistic models have time explicitly present in the model description, e.g. the symbol \( t \) in a system of ODE’s or SDE’s.

• Because of this convention, time is explicitly represented in the proposed IDs as a decision node.

• Random model parameters or random independent nodes that are not time dependent are root nodes that generate a random value at \( t = 0 \).
ODE or SDE system output nodes are dependent ID nodes that produce a random value conditional on both the given end-time value ($t_{end}$) and all decision node values.

To obtain a realization of the model over the interval $(0, t_{end})$, the ID is sampled at times $t = 0, t = t_1, \ldots, t = t_{end}$. Other nodes may have time-dependent parameters and are therefore descendents of the time node.

- Logic sampling is used to solve the ID.
Species Viability ID: 
The Cheetah in Kenya

The cheetah management ID consists of three main sections

1. Decision nodes for representing management options, subregions or areas of Kenya, and a time value at which ID outputs are desired.

2. A vector SDE model of cheetah population size represented by a set of chance and deterministic nodes.

3. Presence/absence and associated utility represented by a chance and value node, respectively.

Probabilistic Representation of Geographic Characteristics

Areas of the study region that are large enough to have the potential of sustaining the species under favorable conditions are identified.

These areas should be as homogeneous in terms of vegetation and climate as possible.

Strict homogeneity is not required however, because the ID uses r.v.’s to represent within-area heterogeneity.

The Gross (1998) regions (Marsabit, Eastern, Samburu, Tsavo surroundings, Masailand, Laikipia, and Nakuru) are used here with the addition of regions for the densely settled western districts (Western), the central farmlands (Central), Turkana (Turkana), and coastal areas (Coastal).

Each area’s climate is modeled as a discrete, time-independent r.v. taking on values very-arid, arid, semi-arid, and non-arid (from Gros (1998)).

Land-use values of nomadic-camel, nomadic-cattle, ranching, and farming, are also modeled as time-independent discrete random nodes (also from Gros (1998)).
Probabilistic Representation of Carrying Capacity

A region’s cheetah carrying capacity is a deterministic function of herbivore biomass:

\[ K_t \equiv \text{nearest integer}(\beta_{K_t}^{(0)} + \beta_{K_t}^{(1)} B_t). \]

A single birth-death model for the meta-population size of cheetah-prey herbivores at time \( t \) is used to represent the population dynamics of the prey populations. Let \( k_0 \) be the herbivore carrying capacity at time \( t_0 \). The SDE for \( B_t \) with random carrying capacity is

\[ dB(t) = B(t)(k_0 - B(t))dt + B(t)\sigma dW_t^{(B)} \]

where \( dW_t^{(B)} \) is a zero mean, unit variance white noise process.
$B_t$ is influenced by the nodes $t$, $C$, and $m$. $U$ could also be an influence but was not used here due to the increased complexity.

This logistic growth SDE is added to the system of SDE's (see below) and this system of four equations is solved numerically.
Cheetah Population Dynamics Model

Wells et al. (1998) give a DE model of species population size (count) as a function of time. The authors define the following parameters:

\( f \in (0, 1) \): instantaneous birth rate,

\( r \in (0, 1) \): instantaneous mortality rate (death rate),

\( c \in (0, 1) \): proportion of the \( N \) animals that meet over a short period of time. This implies that \( cN \) is the number of meetings over a short period of time.

\( P \): the probability that any one meeting does not result in a litter.

\( k \): carrying capacity of the environment in terms of maximum number of animals that can be supported.
The model is
\[
\frac{dN}{dt} = f(1 - P^e)N - rN - (f - r)\frac{N^2}{k}.
\] (1)

\(N_0\), the initial count of the species within an area is modeled as an area-dependent parameter. The initial time at which this count exists is \(t_0\).
There are several sources of uncertainty when using (1) to predict population size through time. These are:

1. heterogeneity of land use, vegetation, and climate within an area,
2. the partial effect that herbivore density has on carrying capacity,
3. the partial effect that poaching and pest hunting has on an area’s death rate, and
4. the impact that unpredictable forces such as droughts have on birth and death rates.

Uncertainty effects on birth and death rates are modeled by making these rates stochastic processes defined by the solutions of their governing, independent SDEs as follows.
Let $W_t = (W_t^{(f)}, W_t^{(r)}, W_t^{(N)})$ be a vector of three independent Wiener processes.

Also, let
\[
a_f(X_t) \equiv -\left(\alpha_f + \beta_f X_t(1 - X_t^2)\right) \text{ \text{ (drift)},}
\]
\[
b_f(X_t) \equiv \beta_f (1 - X_t^2) \text{ \text{ (diffusion),}
}\]
and
\[
f_t = U(X_t) \equiv (1 + X_t)/2.
\]

The distribution of $f_t$ at $t$ is the solution to the SDE:
\[
df_t = \frac{1}{2}a_f(U^{-1}(f_t))dt + \frac{1}{2}b_f(U^{-1}(f_t))dW_t^{(f)}.
\]

This SDE was chosen because its solution is bounded between 0 and 1 making $f_t$ a well-defined birth rate.

A similar development leads to the death rate SDE:
\[
\]
\[
dr_t = \frac{1}{2}a_r(U^{-1}(r_t))dt + \frac{1}{2}b_r(U^{-1}(r_t))dW_t^{(r)}.
\]
The tendency of more females to have litters within protected areas (Gros 1998) is represented by having the parameter $\alpha_f$ be conditional on the area’s status.

Similarly, to represent the effect of poaching and pest hunting on $r_t$, $\alpha_r$ is conditional on the degree of poaching and pest hunting.

The variability of the sample paths of $f_t$ and $r_t$ are controlled by the parameters $\beta_f$ and $\beta_r$, respectively.
All other effects of uncertainty (e.g. model inadequacy, age-dependent parameters) on the within-area cheetah count differential \((dN_t)\) are represented by the derivative of a Wiener process — accomplished by converting (1) to an SDE:

\[
dN_t = \left[ f_t(1 - P^c N_t) N_t - r_t N_t - (f_t - r_t) \frac{N_t^2}{k} \right] dt + \beta_{\mu} dW^{(N)}_t.
\]

where \(P, c, N_0, \text{ and } \beta_{\mu} \) are fixed parameters, and \(k\) is a random parameter.

Conditional on \(k\), the random vector \((f_t, r_t, N_t)'\) is the solution to this nonlinear vector SDE.
If $f_t$, $r_t$, and $k$ are fixed, the distribution of $N_t$ is the solution of an SDE with constant coefficients. This is the causal model represented by the ID in terms of Pearl’s definition of causality.

The solution that is found when $k$ is a random node, and $f_t$ and $r_t$ are stochastic processes corresponds to finding the joint probability distribution of $(k, f_t, r_t, N_t)$ for given values of $t$, area, and management option.

Defining the viability model as an ID then, is critical to conveying causal structure to all involved parties. Such specific statements concerning causality cannot be deduced from examination of the vector SDE alone.
A species is defined to be viable within a particular area if the expected count within that area is nonzero at a distant future time point.

The ID gives a probability distribution of species count within each area at a given time and given management option.
**Fraction-of-Area-Detected Node**

The final output node, $D_t$, measures the fraction of a region’s area over which cheetah have been detected. Let $ra$ be a region’s surface area and $d = N_t/ra$, i.e., the density of cheetah in the region.

An observation on $D_t$ can be computed from maps of cheetah presence/absence by district: divide by $ra$, the sum of all areas of districts in the region on which cheetah have been detected.

$D_t$ is a deterministic function of $N_t$ and $ra$: Let $\xi$ be the minimum cheetah density that results in a cheetah detection report. Let $\rho$ be a cheetah density above which cheetah are certain to be reported. Then

$$D_t = \begin{cases} 
0, & d < \xi \\
(d - \xi)/(\rho - \xi), & d \in (\xi, \rho) \\
1, & d > \rho. 
\end{cases}$$

Note that it is possible for $N_t$ to be positive but $D_t$ to be zero, i.e., $\xi$ can be interpreted as the minimum density detection limit.

The loss node is a deterministic function of $D_t$. This loss function is arrived at through public debate and reflects a composite of the ecological and economic values that have been expressed by all stakeholders.

The loss function used here represents the values (a) high loss if cheetah go extinct, and (b) bounded economic loss from cheetah predation on livestock:

$$L = \begin{cases} 
10^4, & D_t < .15 \\
1, & .15 < D_t < .5 \\
5, & .5 < D_t < .8 \\
10, & .8 < D_t.
\end{cases}$$

The value of $m$ that minimizes the expected value of this loss node is chosen for implementation.
ID Parameter Estimation

- **Problem:** Huge environmental statistical models are difficult to estimate and validate because data sets are often not large enough or complete enough to allow the use of standard statistical methods.

- **Proposal:** fix some of the parameters at values based on the substantive literature and then use the available data to estimate the balance of the parameters. This approach was used by Speed (1993) to estimate a large state-space time series model of salmon population in the Pacific Northwest.
• Consistency Analysis (Haas 1997, 2001) extends this idea of fixing some of the parameters to substantive theory-justified values (called the **hypothesis values**) by producing parameter estimates such that the resultant fitted or *consistent* distribution deviates minimally from the hypothesis distribution while continuing to be consistent with the available data.
In situations in which it is difficult to justify a particular prior distribution on model parameters, a bayesian approach requires more from the analyst than the analyst can justify. In these situations, CA then, is one way to incorporate theoretical knowledge into parameter estimation without inheriting some of the criticisms of the application of bayesian methods to the assessment of ecological models (see Dennis (1996)).
Assessment of Fit

- Because this is a temporal model and is intended to be used to predict future cheetah viability, some assessment of its prediction skill is of interest.

- The statistic used here is the root mean squared error of predictions one time step ahead (RMSPE). For each prediction, the model is fitted with consistency analysis using all data that is strictly earlier than the current prediction time.

- When the sample contains unknown errors however, the RMSPE should not be viewed as the ultimate criterion for model selection. Rather, it should be examined in light of known weaknesses in the sample.
• Separate RMSPE values were computed for herbivore counts and detection fraction predictions over the observation times.

• For \( c_H = .5 \), these calculations yielded 3757.5, and .727 for \( B_t \) and \( D_t \), respectively.

• For \( c_H = 0. \), these values are 3716.9, and .720.
Use of the Consistent Influence Diagram for Cheetah Management

- Now that the influence diagram has been fitted to both the sample and prior knowledge, it can be used to aid cheetah management decisions.

- To illustrate the graphical outputs that are possible from the influence diagram, Figure 5 gives the mean $D_t$ values by region for the conditions $m = do$ nothing and $t = 2020$. Cheetah presence is low in regions of heavy farming and higher in less cultivated areas.
Selection of Optimal Management Option

- A hypothetical decision selection exercise is described to illustrate how the estimated influence diagram is used within the EMS to manage cheetah viability.

- The best management option for the Central region is to be found for a 20 year planning horizon. Hence, the decision nodes are set as follows: $t = 2020$, $q = Central$, and $m$ set in turn to each of its values.

- A Monte Carlo estimate of the expected loss under each option is computed from 100 simulated realizations drawn from the influence diagram.

- The option that minimizes the expected loss is increase anti-poaching enforcement.

- The standard deviations indicate this expected loss is significantly lower at the $\alpha = .1$ level than those of the other options.
Conclusions

- An ID allows more accurate representation of expert opinions in a decision support system than many other methods due to the modular nature of the IDs knowledge base.

- The marginal distribution of the IDs output nodes can be used to approximate the marginal expected value, e.g. $N_t$.

- The effect of interventions and management activities can be studied using either (a) best-available qualitative knowledge incorporated into a discrete or conditional gaussian ID, or (b) best-available population dynamics or other quantitative models incorporated into an extended ID as in Haas (2001).
Theoretically sound statistical estimation of an ID's parameters can be performed with an incomplete and/or small sample using either maximum likelihood, bayesian, or a mixed method as in CA.
References Not Included in Haas (2001)


