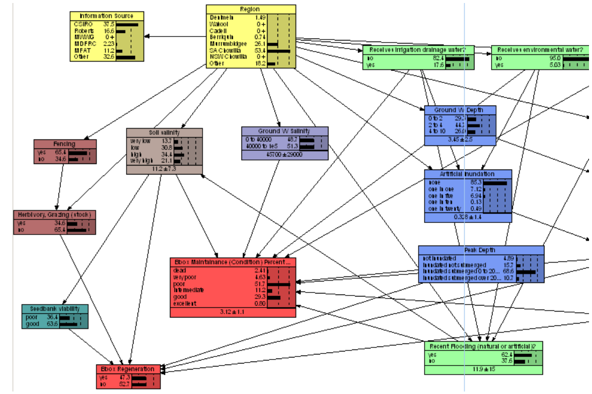
Graphical and Numerical Analysis of BNs

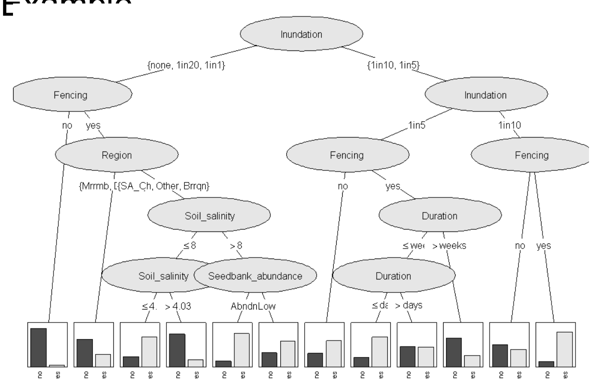
Felix Andrews
iCAM, Fenner School
The Australian National University

2009-11-27

BNs: transparent, or a black box?



Example

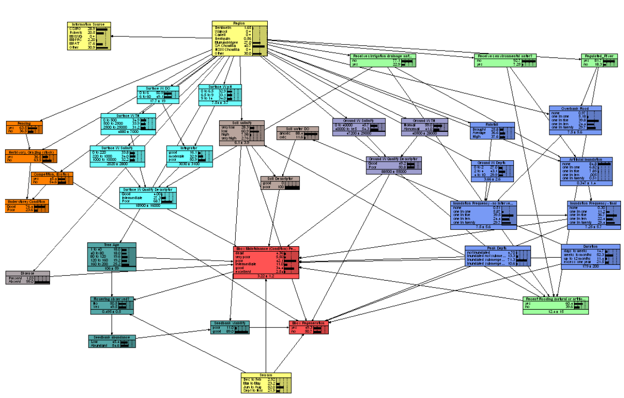


Sensitivity Analysis

For:
- revealing model behaviour
- assessing variable importance
  - to focus model development
  - uncertainty analysis

Sensitivity Analysis for modellers?
Would you go to an orthopaedist who didn't use X-ray?
(Furbinger, 1996)

Ecological example (Pollino et. al.)



Sensitivity of what?

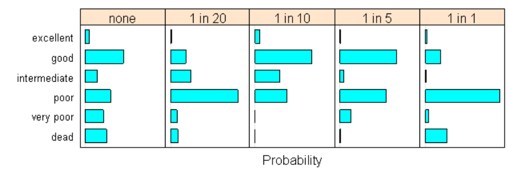
Formulate the relevant model result, depending on the model's purpose, e.g.
- a quantity
- 95% upper bound on a quantity
- a probability
- a difference in two quantities
- a ratio of two probabilities
- ranking (order) of a set of scenarios

Sensitivity to what?

1. Sensitivity to inputs, giving a summary of the model behaviour
2. Sensitivity to uncertain assumptions (or parameters), revealing aspects of uncertainty in the results

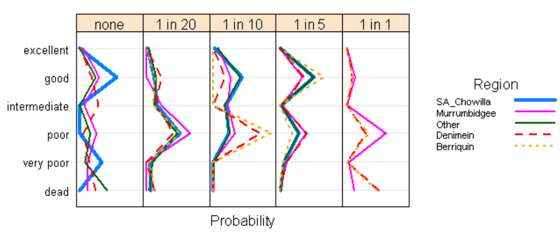
Simple example

Endpoint distribution for each state of "Inundation
(all other variables fixed).



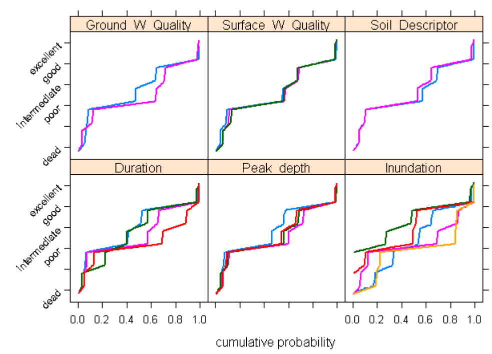
Simple example of interactions

Showing interactions (superposed and counterposed).
Effect of "Region" within each state of "Inundation".



"Sensitivity to findings" - graphical

Comparing the cumulative distribution for each state.

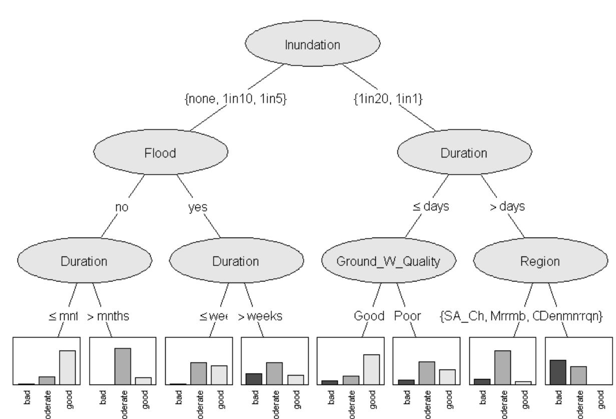


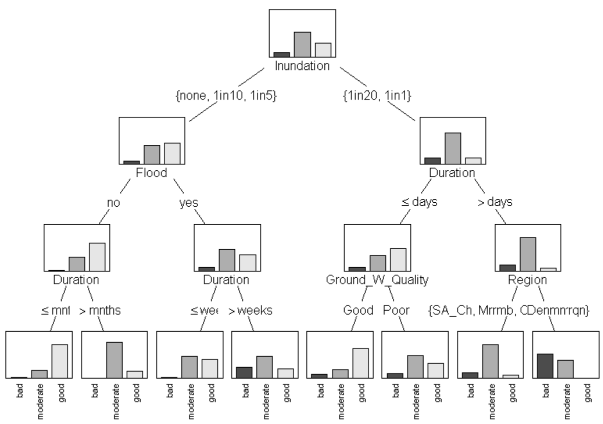
Model behaviour

This approach can begin to reveal some interactions.

But it is still difficult to get a sense of the overall model behaviour. Would like a global summary.

One idea: simulate from BN and fit a decision tree as a meta-model.





Problems with this approach

Not a complete representation of the original BN.

Variables are correlated: splits in Decision Tree are ill-defined (best to choose a representative set).

Effects modelled by the Decision Tree are statistical, not necessarily causal (requires careful interpretation).

Quite sensitive to number of states of variables...

Random or deterministic simulation?

States of endpoint node are not generally determined even if all its parents' states are known. Unmodelled randomness. Just take most probable state (deterministic) for some uses.

Most Probable State	Simulated State						Total
	dead	very poor	poor	interm.	good	excellent	
dead	73	1	8	16	2	1	100%
very poor	1	72	17	1	7	2	100%
poor	2	5	79	4	9	1	100%
intermediate	6	1	13	71	8	0	100%
good	1	1	10	3	84	1	100%

Random or deterministic simulation?

With Full simulation from BN:
terminal nodes of Decision Tree represent the <u>marginal posterior</u> distribution under each set of conditions.

With Deterministic simulation from BN:
terminal nodes of Decision Tree show the <u>approximation error</u> (behaviour of the BN that was not captured).

Example of Importance Assessment

"Sensitivity to findings" as in Netica
(average entropy reduction / Mutual Information).
Note that rankings depend on the current BN state.

No findings
<u>(priors distributions)</u>

1. Inundation_Final
2. Overbank_Flood
3. Inundation_no_intervention
4. Region
5. Information_Source
6. Duration
7. ...

With finding
<u>Region = Murrumbidgee</u>

1. Duration
2. Inundation_Final
3. Bbox_Regeneration
4. Artificial_Inundation
5. Inundation_no_intervention
6. Overbank_Flood
7. ...

Sensitivity Analysis for Importance

Proposed method for "global" sensitivity analysis (i.e. integrating all interactions between variables):

Look at direct parents of the endpoint node. In the example this has 10 parents for a total of 76800 combinations.

Calculate the <u>effect size</u> (by imposing an ordinal scale) induced by changing from one state to another; do this for all possible states of the other parent nodes.

Summarise the distribution of effects of each variable.

Ranking by effect sizes

	Quantile of effect size from changing var. state							mean effect size	st.dev (inter-action)	
	0%	10%	25%	50%	75%	90%	100%			both
Inundation_Final	0	0.01	0.07	0.38	1.04	1.73	4	0.36	0.86	0.93
Disease	0	0.04	0.15	0.46	1.12	1.46	3.94	0.29	0.84	0.89
Duration	0	0	0.01	0.14	0.73	1.62	3.92	0.27	0.76	0.8
Peak_depth	0	0	0.01	0.12	0.49	1.27	3.96	0.06	0.65	0.66
Season	0	0	0.01	0.1	0.41	0.95	3.95	0.13	0.56	0.57
Flood	0	0	0.01	0.06	0.27	0.82	3	0.11	0.51	0.53
Ground_W_Quality	0	0	0	0.08	0.38	0.9	3.82	0.06	0.5	0.51
Tree_Age	0	0	0	0.01	0.07	0.32	3.89	0.12	0.4	0.42
Soil_Descriptor	0	0	0.01	0.06	0.22	0.5	3.54	0.02	0.35	0.35
Surface_W_Quality	0	0	0	0.04	0.16	0.38	2.17	0.02	0.25	0.25

Sensitivity Analysis for Importance

Note massive, unrealistic effects. Useful for debugging but not so good for importance assessment.

Problem: naive approach does not account for the (joint) prior probability of observing such effects.

So - calculate probabilities of each combination of states in the CPT. Use these to calculate expected effect sizes.

Nodes with many states are more "uncertain" and therefore state-changes "probable". Best to take a single state-change.

Ranking by expected effects

	expected consistent effect	std dev weighted	raw score	likeli-hood	score (expected)	score (most important)	most important change
Inundation_Final	1.06	0.92	1.40	0.0045	1.00	1.00	one.in.twenty / one.in.ten
Flood	0.33	0.57	0.66	0.0019	0.20	0.68	no / yes
Duration	0.43	0.97	1.06	0.0037	0.63	0.60	years / weeks.to.months
Tree_Age	0.10	0.39	0.40	0.0045	0.29	0.54	160.to.200 / 120.to.160
Ground_W_Quality	0.25	0.41	0.48	0.002	0.16	0.54	Poor / Good
Season	0.20	0.53	0.57	0.0035	0.32	0.45	Sept.to.Nov / Jun.to.Aug
Peak_depth	0.14	0.64	0.66	0.0008	0.08	0.18	submerged / not.submerged
Soil_Descriptor	0.05	0.25	0.25	0.0012	0.05	0.16	poor / good
Surface_W_Quality	0.03	0.10	0.10	0.0016	0.03	0.07	Poor / Intermediate
Disease	0.45	0.70	0.83	0.0001	0.01	0.03	Absent / Present

Summary & questions

BNs: transparent, or a black box?

Decision trees as meta-models?

"Global" sensitivity analysis and importance assessment

the end.