An Introduction to Bayesian Networks

Martin Neil

Agena Ltd &
Risk Assessment and Decision Analysis Research Group,
Department of Computer Science, Queen Mary, University of London
London, UK

Web: www.agenarisk.com
Email: martin@agena.co.uk

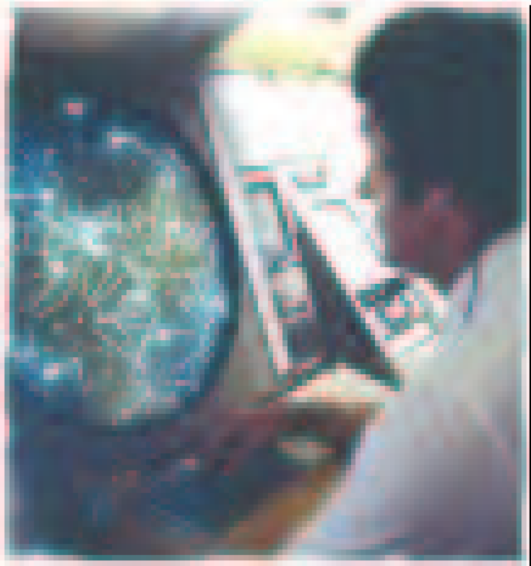
Contents

Bayesian and Bayesian Network Applications

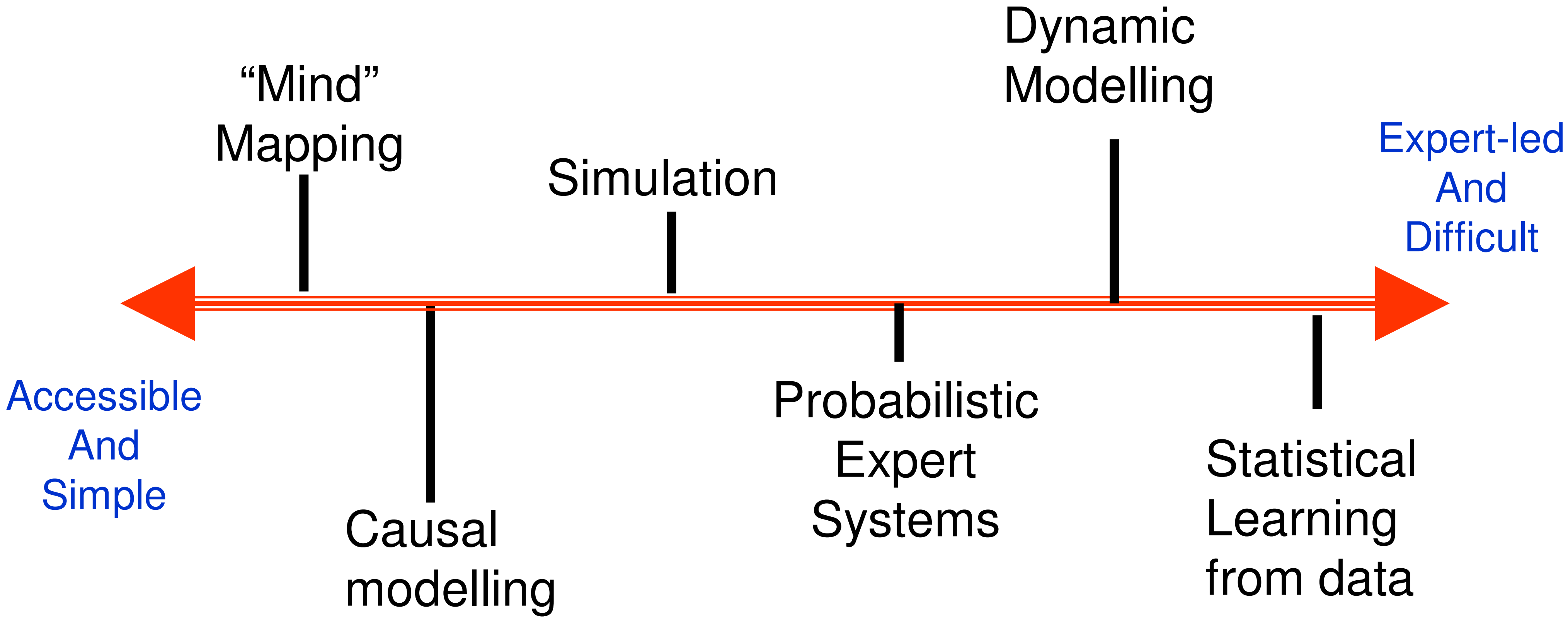
- Google for intelligent search
- Autonomy corporation's information retrieval Agent technology
- Collaborative filtering and recommendation technology for Internet and DigitalTV
- Expert systems for medical diagnosis
- Data mining
- Risk assessment and quality prediction is systems and software engineering
- Air traffic risk prediction
- Computer Vision

“Risky” Applications

- Aircraft Mid-air collision
- Software defects
- Systems reliability and availability
- Warranty return rates of electronic parts
- Operational risk in financial institutions
- Portfolio of IT project risks (ITIL)



AgenaRisk
Modelling Spectrum



Introduction to Bayes Theorem and Bayesian Networks

Features of rational decision making

- **Philosophical Requirements:**
  - Scientific
  - Coherent
  - Prescriptive
  - Optimising
- **Technical requirements:**
  - Simulation model of "system"
  - Decision support for human or as an AI
  - Identification of variability and risks (Epistemic and otherwise)
  - Quantification for learning, estimation and prediction

Rev Thomas Bayes

Derivation of Bayes Theorem

$$p(A,B) = p(A \mid B)\,p(B)$$

$$p(B,A) = p(B \mid A)\,p(A)$$

$$\Rightarrow\ p(A \mid B) = \frac{p(B \mid A)\,p(A)}{p(B)}$$

Bayes' Theorem

A: 'Person has cancer' p(*A*) = 0.1 (*priors*)

B: 'Person is smoker' p(*B*) = 0.5

What is p(*A* | *B*)? (*posterior*)

p(*B* | *A*) = 0.8 (*likelihood*)

Posterior probability **Likelihood** **Prior probability**

$$p(A \mid B) = \frac{p(B \mid A)\, p(A)}{p(B)}$$

So p(*A*|*B*)=0.16

The Frequentist Viewpoint

- **A frequentist believes that probability:**
  - can be legitimately applied only to repeatable problems
  - is an objective property in the real world
  - applies only to events generated by a random process
  - is associated only with collectives not individual events
- **Frequentist Inference**
  - Data are drawn from a distribution of known form but with an unknown parameter
  - Often this distribution arises from explicit randomization
  - Inferences regard the data as random and the parameter as fixed (even though the data are known and the parameter is unknown)

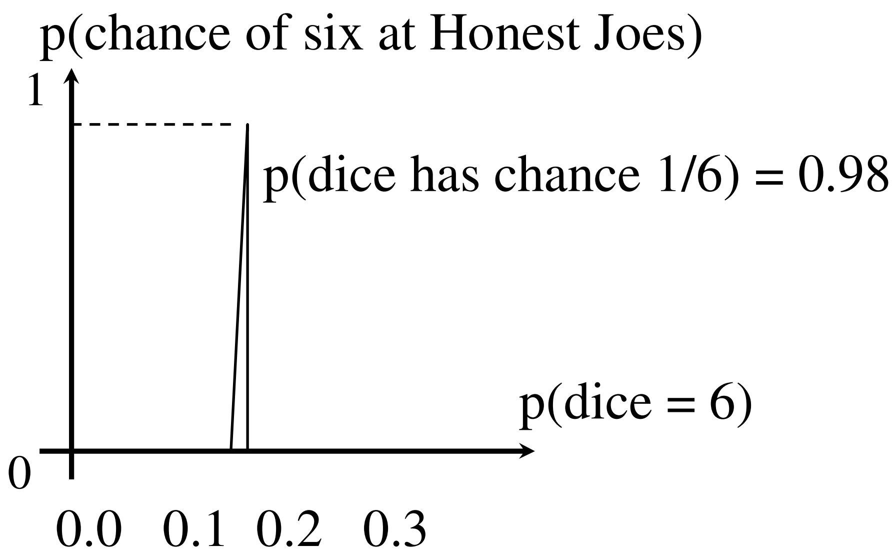
The Subjectivist Viewpoint

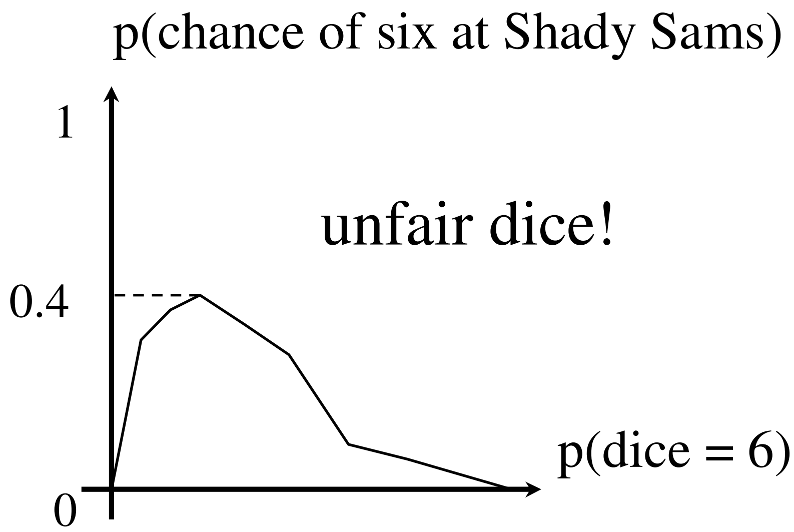
- **A subjectivist believes:**
  - Probability as an expression of a rational agent's degrees of belief about uncertain propositions.
  - Rational agents may disagree. There is no "one correct probability."
  - If she receives feedback her assessed probabilities will in the limit converge to observed frequencies
- **Subjectivist Inference:**
  - Probability distributions are assigned to the unknown parameters.
  - Inferences are conditional on the prior distribution and the observed data

Combining Subjective and Objective information

- **Casino 1- Honest Joe's.**
  - You visit a reputable casino at midnight in a good neighbourhood in a city you know well. When there you see various civic dignitaries (judges etc.). You decide to play a dice game where you win if the die comes up six.
  - What is the probability of a six?
- **Casino 2 - Shady Sams.**
  - More than a few drinks later the Casino closes forcing you to gamble elsewhere. You know the only place open is Shady Sam's but you have never been. The doormen give you a hard time, there are prostitutes at the bar and hustlers all around. Yet you decide to play the same dice game.
  - What is the probability of a six?

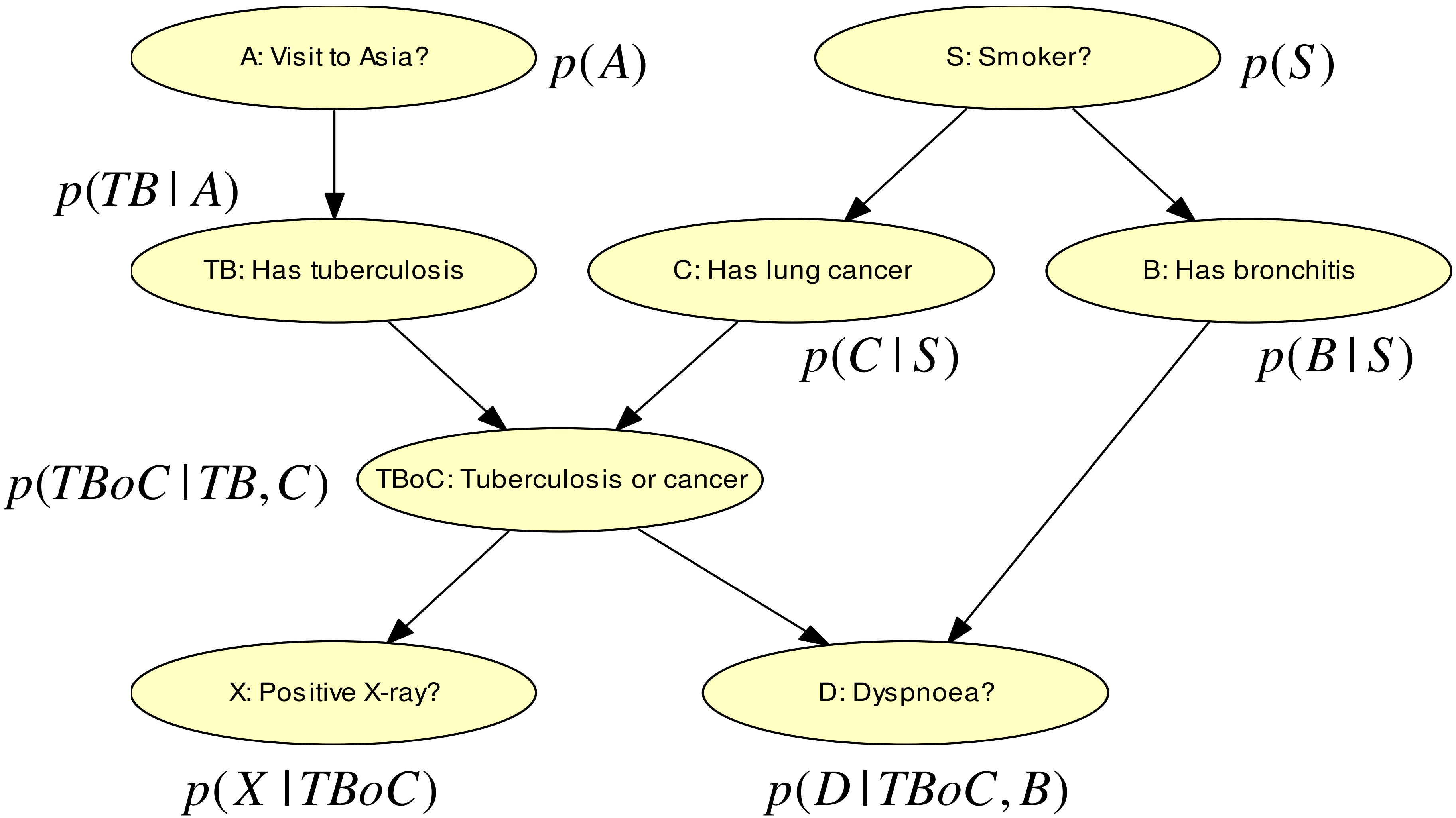
Honest Joe's Vs Shady Sams



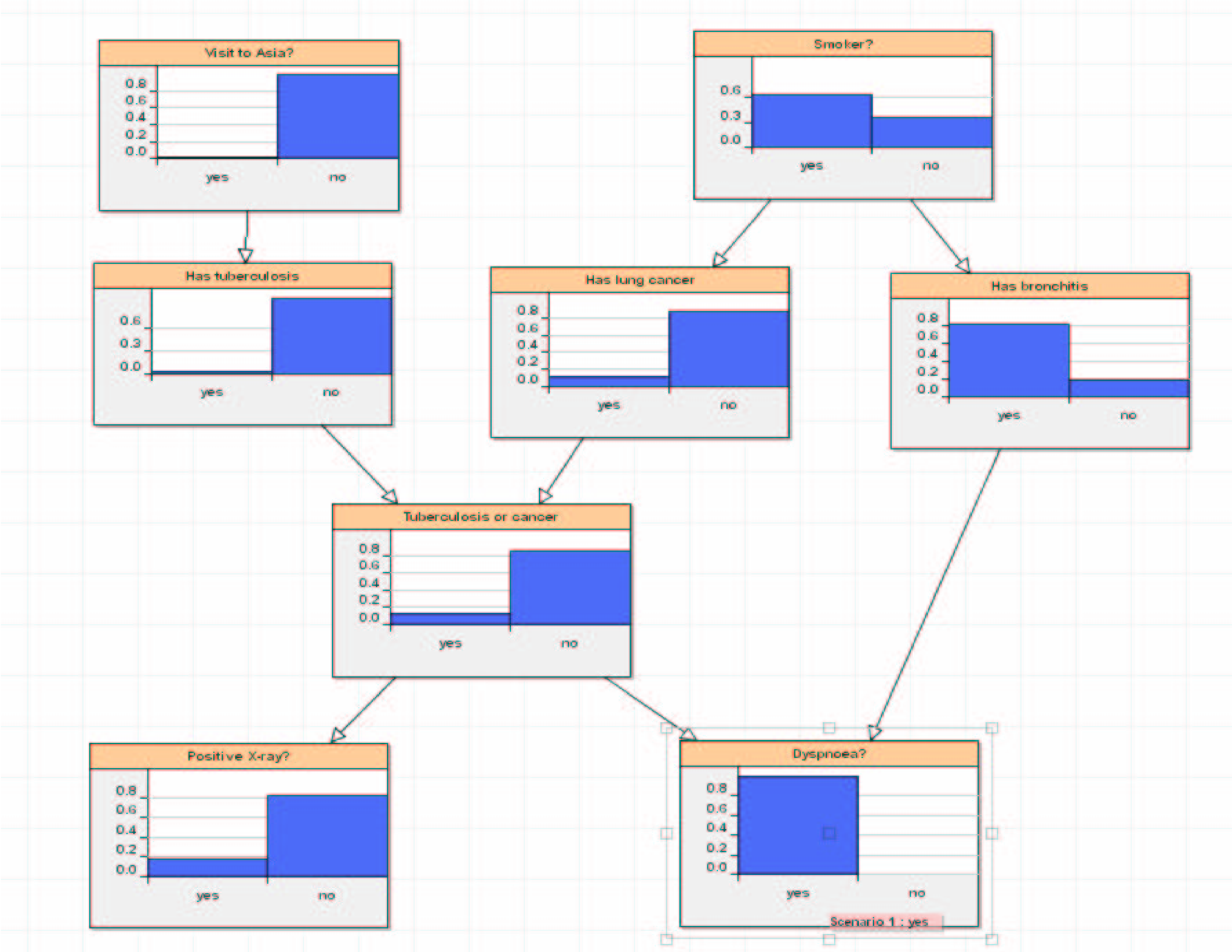


Both of these graphs may be produced by subjective guesses or by long-run observation of dice or indeed by combination of frequencies, as data, and guesses, as prior dispositions.

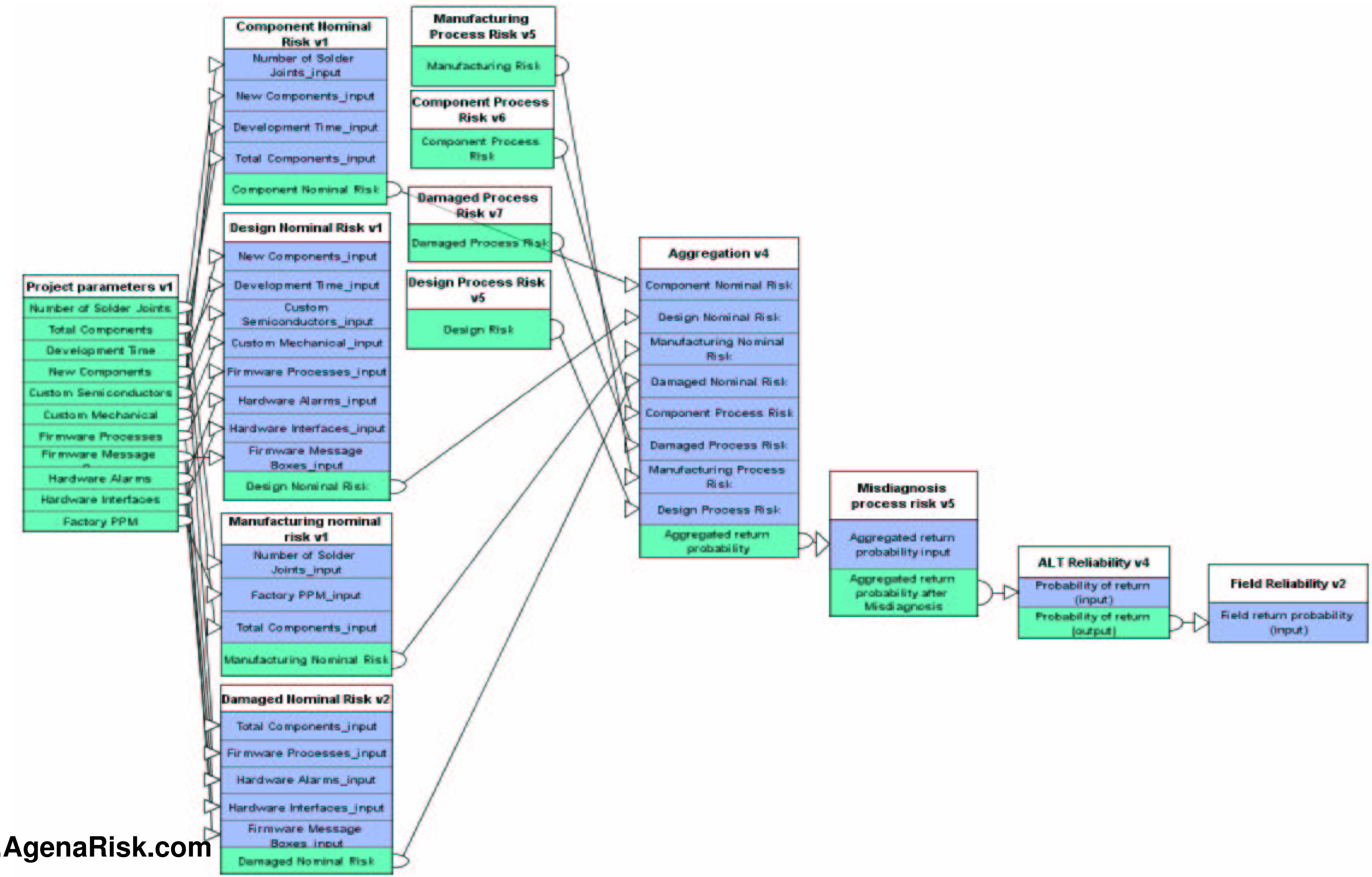
Bayesian Network Example



Executing a BN in AgenaRisk

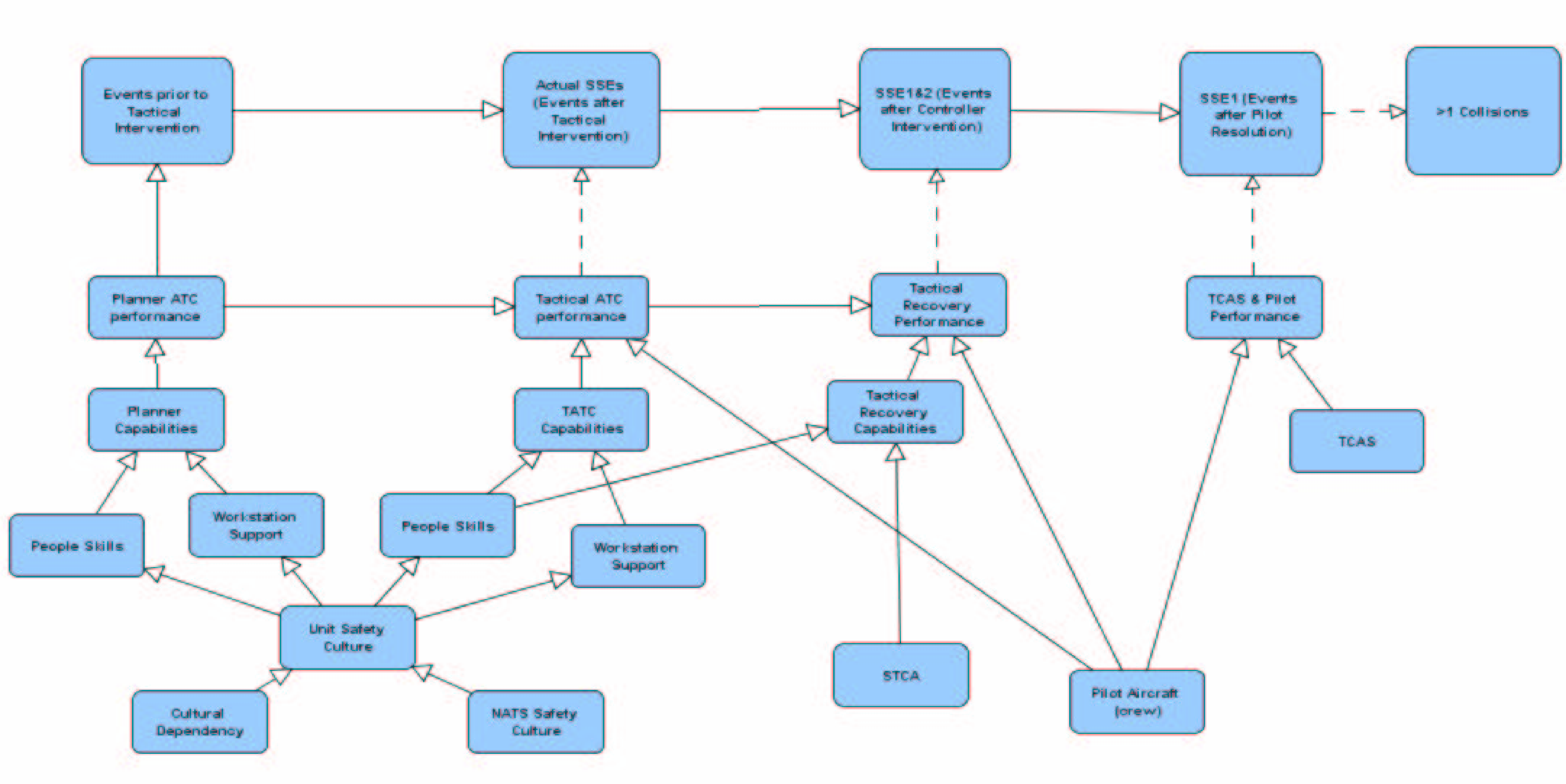


Six Sigma Quality Control



www.AgenaRisk.com

Mid Air Collision Prediction

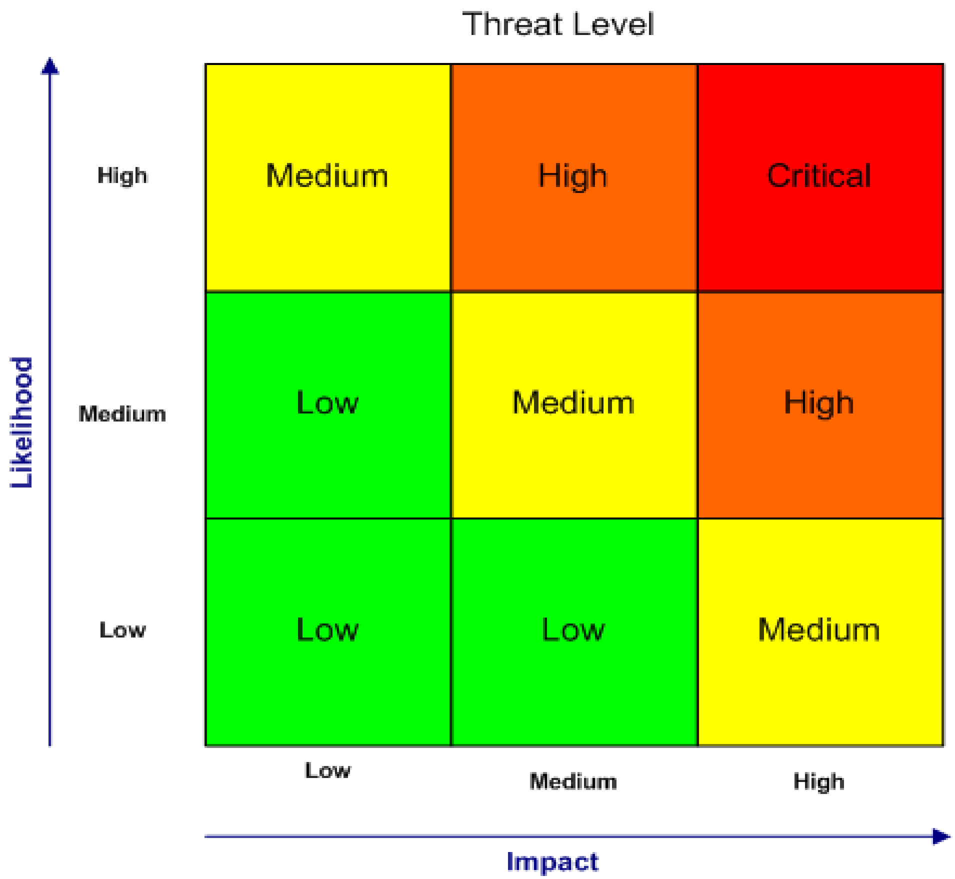


Using Bayesian Networks for "Risk Maps"

Risk Register

- “There are tight budget constraints”
- “The project overruns its schedule”
- “The company’s reputation is damaged externally by publicity about poor final system”
- “The customer refuses to pay”
- “The delivered system has many faults”
- “The requirements are especially complex”
- “The development staff are incompetent”
- “Key staff leave the project”
- “The staff are poorly motivated”
- “Generally cannot recruit good staff because of location”
- “There is a major terrorist attack”

Risk Heat Maps and Profiles



Risk = Likelihood x Impact

Spreadsheets

BINOMDIST =C16/ABS(A16-B16)

	A	B	C	D	E	F	G
4							
5							
6	**Xlower**	**Xupper**	**p(x)**	**midpoint**	**m*p(x)**	**p(x)*x^2**	**density**
7	-100	-1	1.39E-05	-50.5	-7.02E-04	3.55E-02	1.40E-07
8	-1	0	5.57E-05	-0.5	-2.79E-05	1.39E-05	5.57E-05
9	0	1	0.001217	0.5	6.09E-04	3.04E-04	1.22E-03
10	1	1.5625	0.003352	1.28125	4.29E-03	5.50E-03	5.96E-03
11	1.5625	1.84375	0.006036	1.703125	1.03E-02	1.75E-02	2.15E-02
12	1.84375	1.984375	0.007557	1.914063	1.45E-02	2.77E-02	5.37E-02
13	1.984375	2.125	0.014305	2.054688	2.94E-02	6.04E-02	1.02E-01
14	2.125	2.265625	0.025505	2.195313	5.60E-02	1.23E-01	1.81E-01
15	2.265625	2.40625	0.041875	2.335938	9.78E-02	2.28E-01	2.98E-01
16	2.40625	2.546875	0.062771	2.476563	1.55E-01	3.85E-01	=C16/ABS(A16
17	2.546875	2.6875	0.085438	2.617188	2.24E-01	5.85E-01	ABS(num
18	2.6875	2.828125	0.10558	2.757813	2.91E-01	8.03E-01	7.51E-01
19	2.828125	2.96875	0.11851	2.898438	3.43E-01	9.96E-01	8.43E-01
20	2.96875	3.109375	0.12108	3.039063	3.68E-01	1.12E+00	8.61E-01

Expert Judgement - "I Assume"

- On the one hand….
  - Obvious risk of being wrong
  - Dangerous if unverified, checked or agreed
  - Political
- On the other hand….
  - Absolutely necessary
  - Unavoidable
  - We employ people for a reason!
- Model Risk: If you want to analyse risk you are going to have to take them….

How good are people at estimating risk?

- **Evidence from psychology is worrying!**
  - **Availability of more recent cases**
  - **Emphasis on easier to remember dramatic events**
  - **Large single consequence often outweighs multiple small consequences**
- **Framing Problem: Answer you get depends how you ask the question!**

"What is the chance of disease?"

Vs

"Given positive test result what is the chance of disease?"

Vs

"Chance of disease given test positive?"

If you cannot trust people then trust the data

- Statistical validity restricted to controlled experiments
- Data sets must represent homogeneous samples and correlations clear
  - High correlation between shoe size and IQ!
- Do you even have the data?
  - New business ventures?
  - Rare events?

……

Age	Accidents
23	5
26	3
45	2
67	3
56	2
12	0
65	1
34	4

The lure of objective irrationality

Decomposing (Exposing) Risk Measure

- **Standard Definition:**

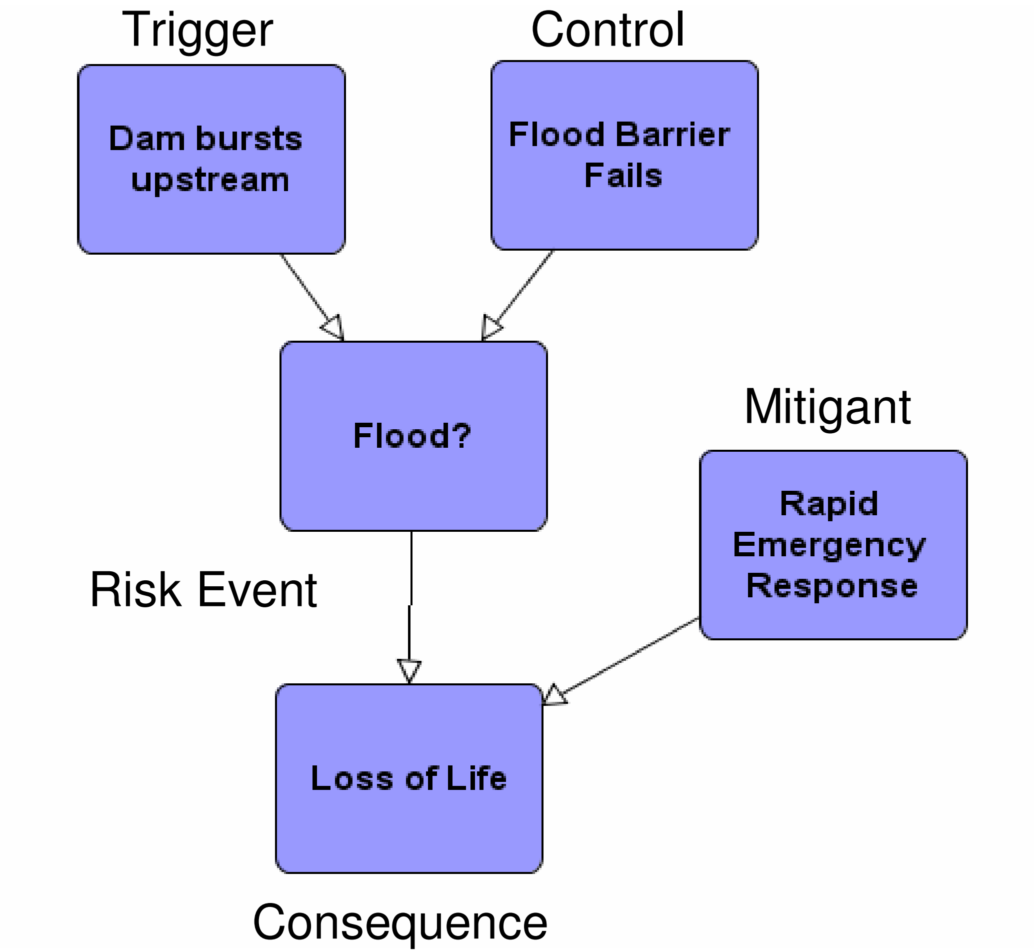
Risk = Impact x Probability

- **Is this decomposition enough?**
- **Expose the assumptions!**
  - **What is the context driving the numbers?**
  - **Who's risk is it?**
  - **Is it a risk to me?**
  - **Is it really a risk?**
  - **An indicator of a risk?**
  - **A mitigant…..?**

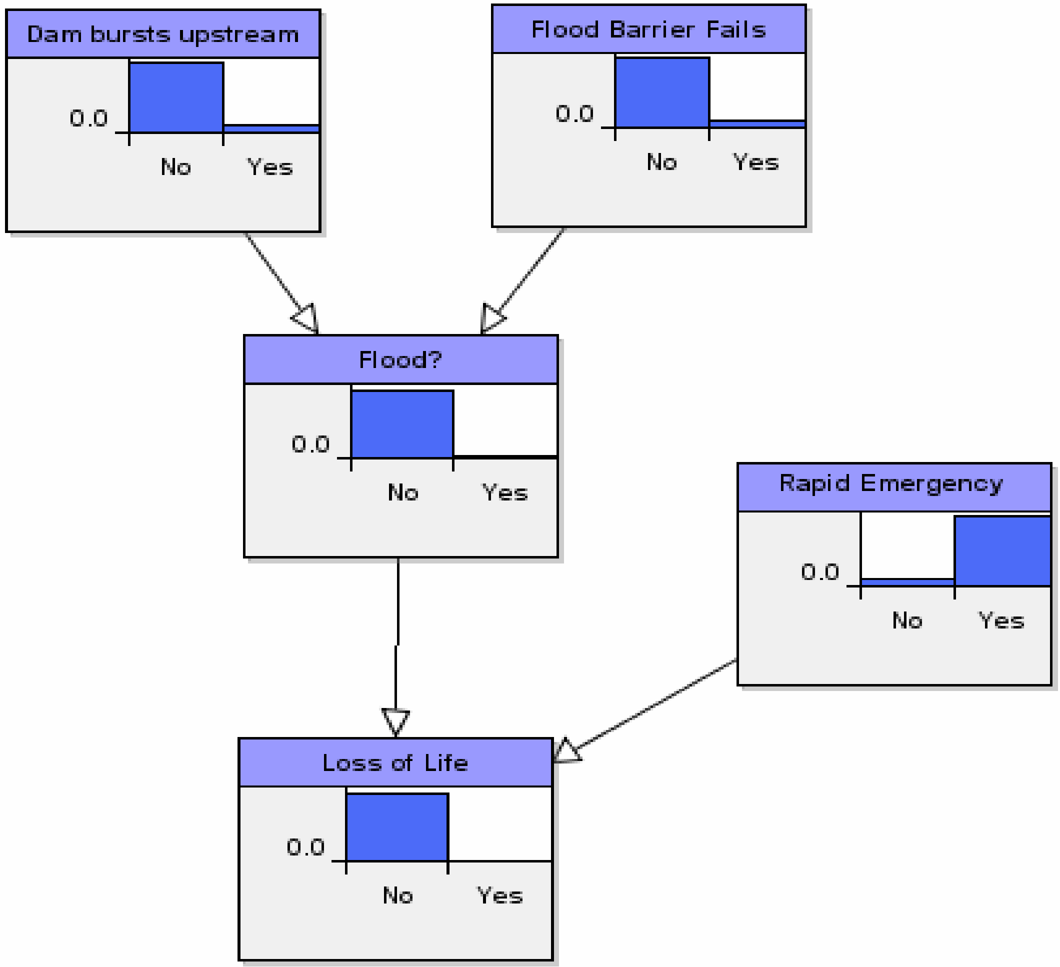
Causal Framework for Risk

- Replace oversimplistic measure of risk with a causal approach
- Characterise risk by event chain involving:
  - The risk itself (at least)
  - One consequence event
  - One or more trigger events
  - One or more mitigant events
- Context "tells a story" and depends on perspective

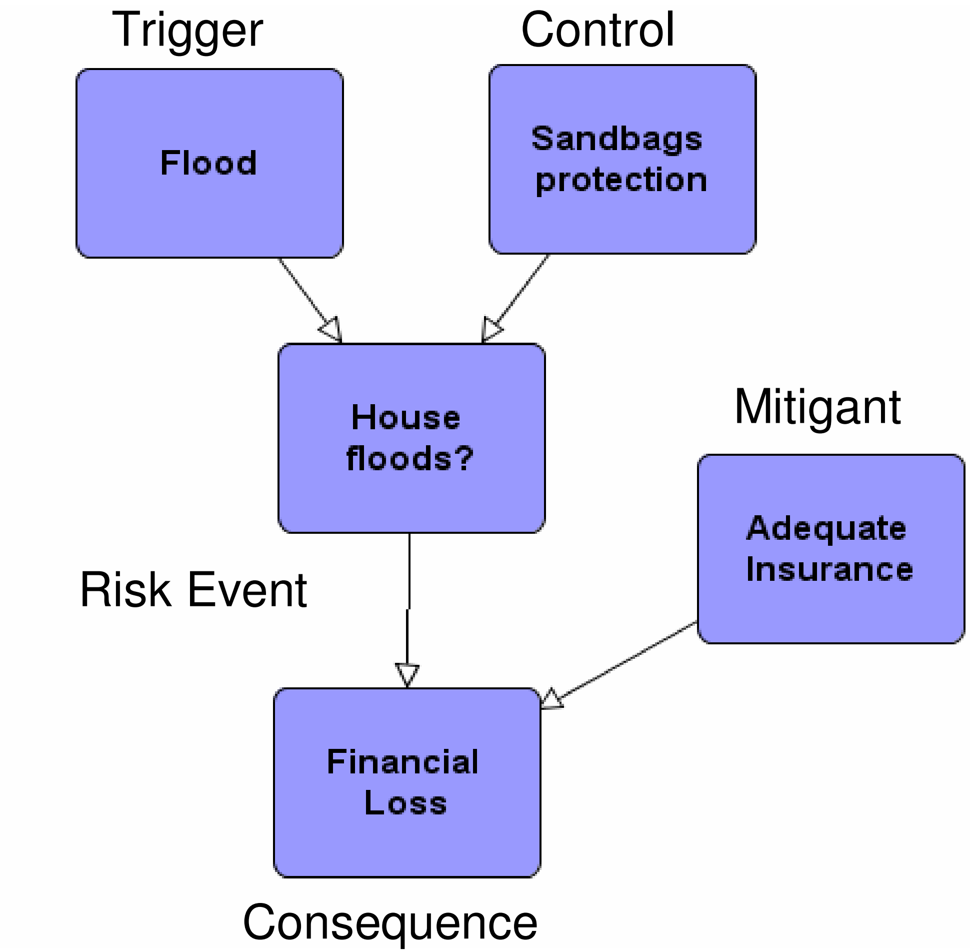
Town Flood Example



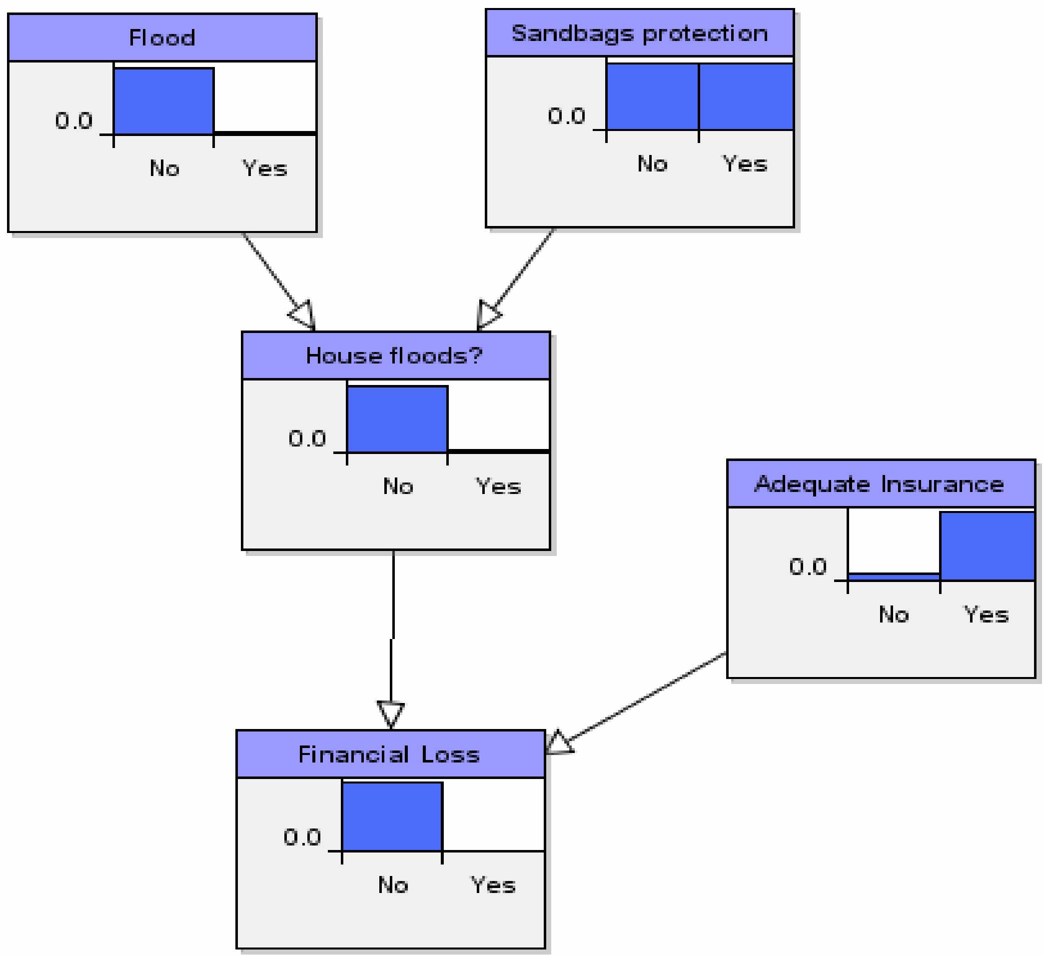
Calculation of Town Flood Risk



Flood Example – Homeowners Perspective



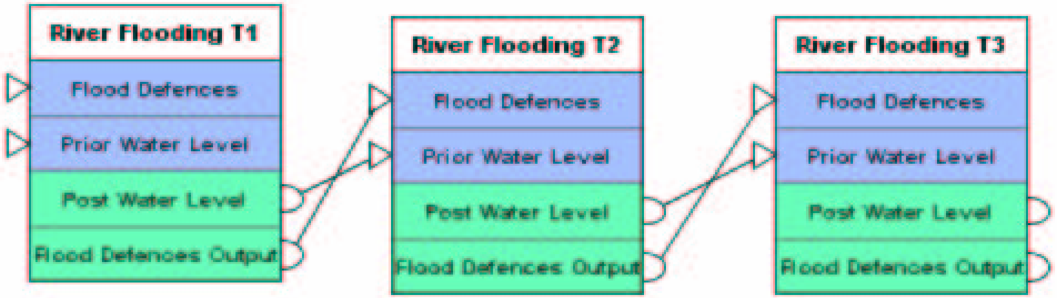
Calculation of Home Flood Risk

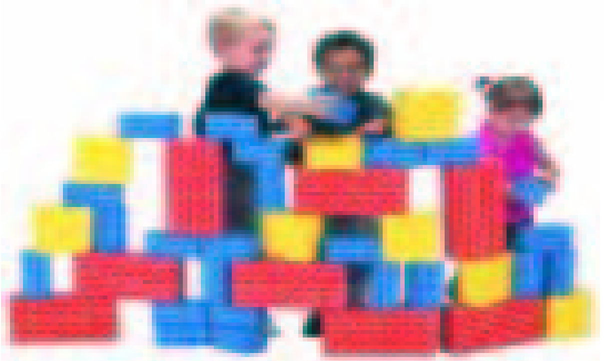


Connecting Risk Maps using Building Blocks

- Connect risk maps via input/output risk nodes

- Create complex time based or complex structural models





Benefits

- "A picture tells a thousand words"
- Explicitly quantifies uncertainty
- Connecting models "connects perspectives"
- Dynamic calculation of risk values
- Great for "what if" analysis

Information fusion & AI applications

Object Tracking
Learning from data

Motivation

- **Aim to model complex systems:**
  - Develop a system model that accounts for direct and indirect uncertainties in a system behaviour.
  - Optimally estimate the quantities of interest in the presence of uncertainty
  - Optimise the control of a system in the face of incomplete and noise corrupted data
- **Deterministic control theories are not enough!**
  - No mathematical system is perfect
  - Mathematical laws can be built in but various system parameters will be imprecisely understood, so we need to embrace uncertainty
  - Our measurements and sensors provide imperfect knowledge of the world

State Space Models

- **A state space model consists of:**
  - **Prior state** $p(X_{t=0})$
  - **State transition function** $p(X_t \mid X_{t-1})$
  - **Observation function** $p(Y_t \mid X_t)$
- **Usually interested in inference over time, *t*, but can apply to any non-stationary system**
- **Popular methods:**
  - **Hidden Markov Models (HMMs)**
  - **Kalman Filter Models (KFMs)**
  - **Dynamic Bayesian Networks (DBNs)**

Components of KFM

Key assumption: All distributions unimodal Gaussian

Unreliability of each sensor: $X_i \sim N(Y, {\sigma_{X_i}}^2)$

State of latent node: $Y \sim N(\mu_Y, {\sigma_Y}^2)$

Information from two sensors is weighted:

$$\mu_Y = \left[\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right] x_1 + \left[\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right] x_2$$

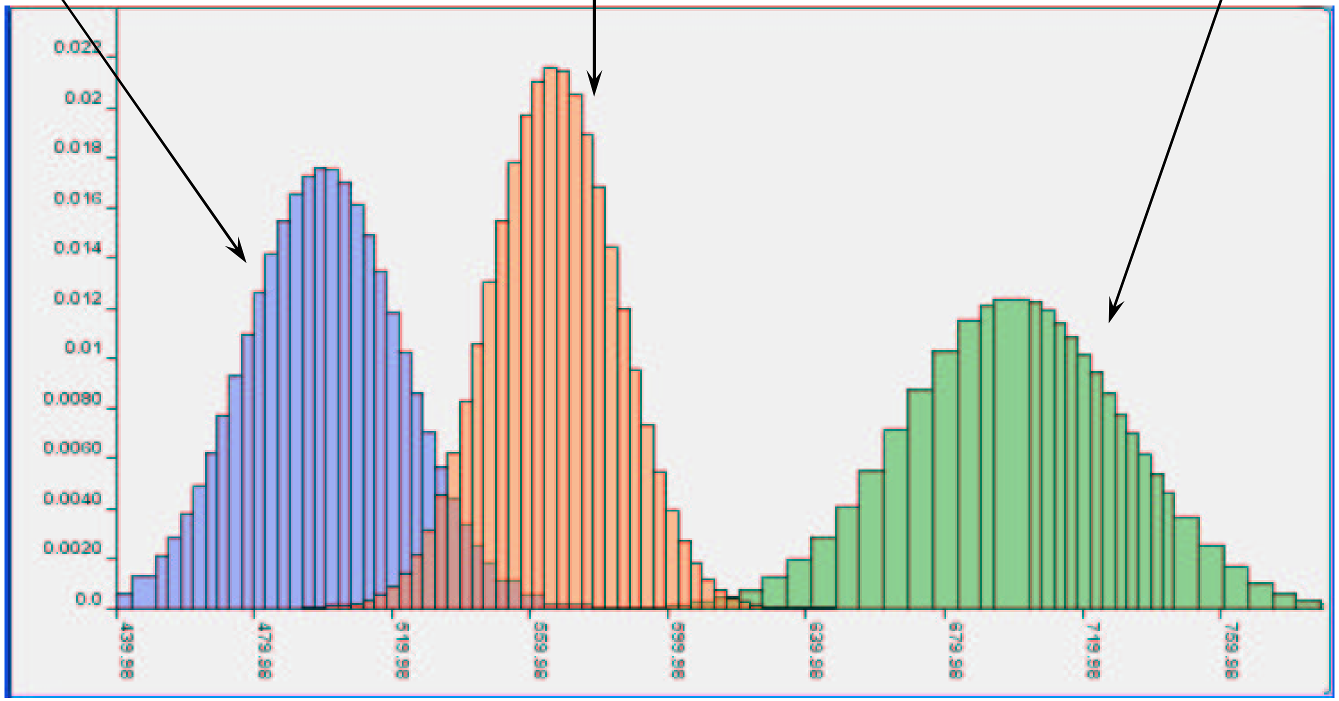
$$\frac{1}{{\sigma_Y}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

Example of fusion from two sensors

Sensor 2 worse than sensor 1: $\sigma_1^2 = 500 \quad \sigma_2^2 = 1000$

$$Var(Y \mid X_1 = 500, X_2 = 700) < Var(Y \mid X_1 = 500) < Var(Y \mid X_2 = 700)$$

$p(Y \mid X_1 = 500)$ $\quad$ $p(Y \mid X_1 = 500, X_2 = 700)$ $\quad$ $p(Y \mid X_2 = 700)$



Linear Dynamical System

- **Linearity a requirement in KFM but not a problem in DBNs**
- **Example difference equations for position and velocity:**

$$P_t = P_{t-1} + V_{t-1}$$

$$V_t = V_{t-1}$$

- **Difference equations ignore noise/transition terms**

KFM Example Specification

Observations $p(O_t \mid P_t) = Normal(P_t, \sigma^2)$

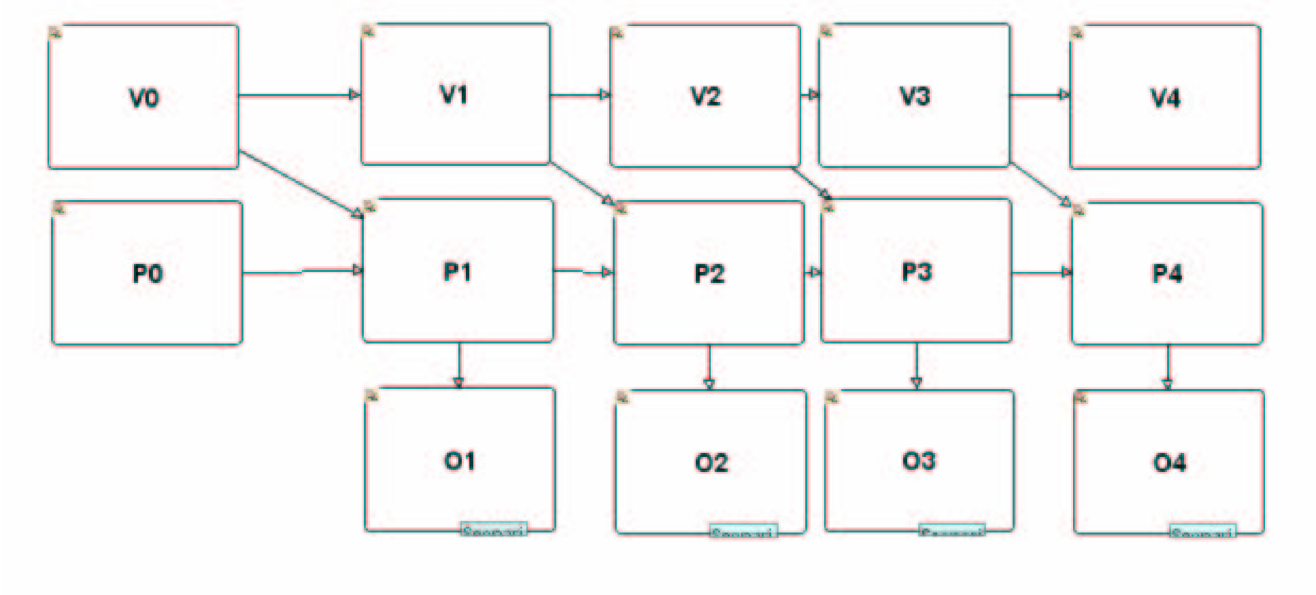
Position Transition $P_t = P_{t-1} + V_{t-1}$

Velocity Transition $V_t = V_{t-1}$

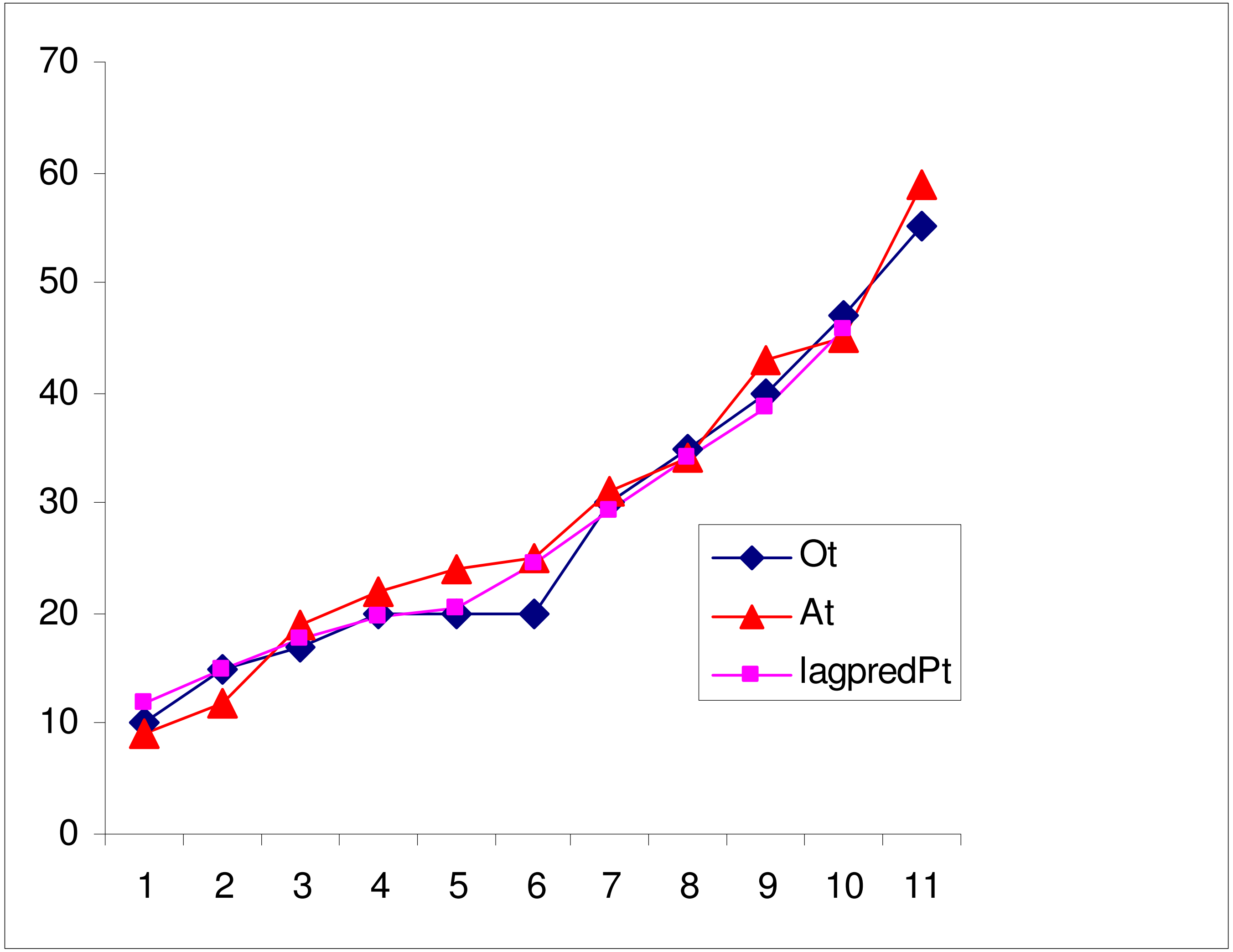
Initial Conditions

$$V_0 = N(0, \theta_1)$$

$$P_0 = N(0, \theta_2)$$



Tracking accuracy – Lag Prediction Vs Actual and Observed



Detecting unreliable sensors

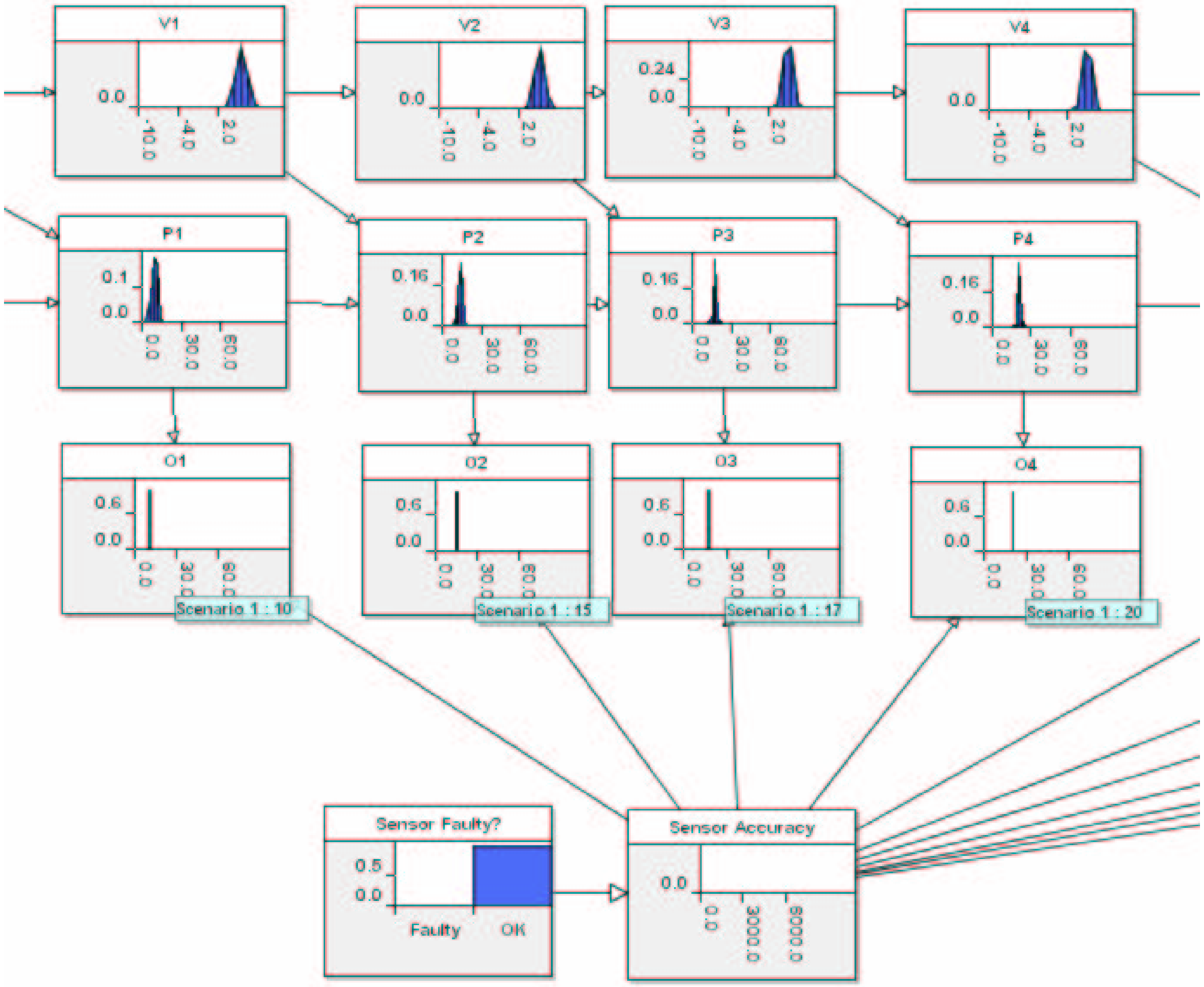
- From observations we can learn, online, whether a sensor is unreliable or not
- Consider a sensor with two states {OK, Faulty}
- When the sensor is OK the variance is 10
- When the sensor is faulty the variance is 1000
- Normal data:

10, 15, 17, 20, 20, 20, 30, 35, 40, 47, 55

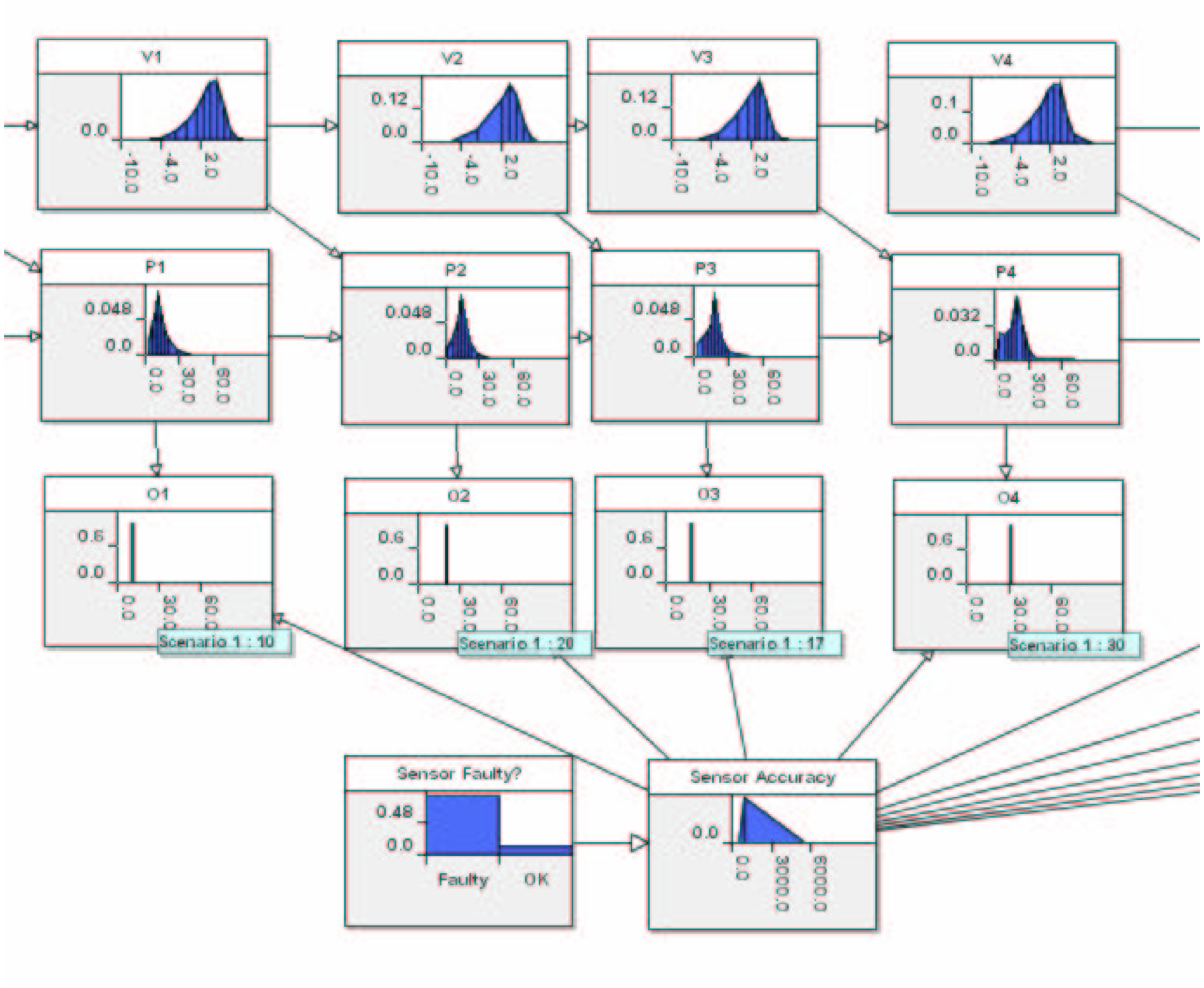
- Abnormal data:

10, 20, 17, 30, 20, 20, 10, 25, 40, 47, 55

Unreliable sensors? – Normal data



Unreliable sensors? – Abnormal data



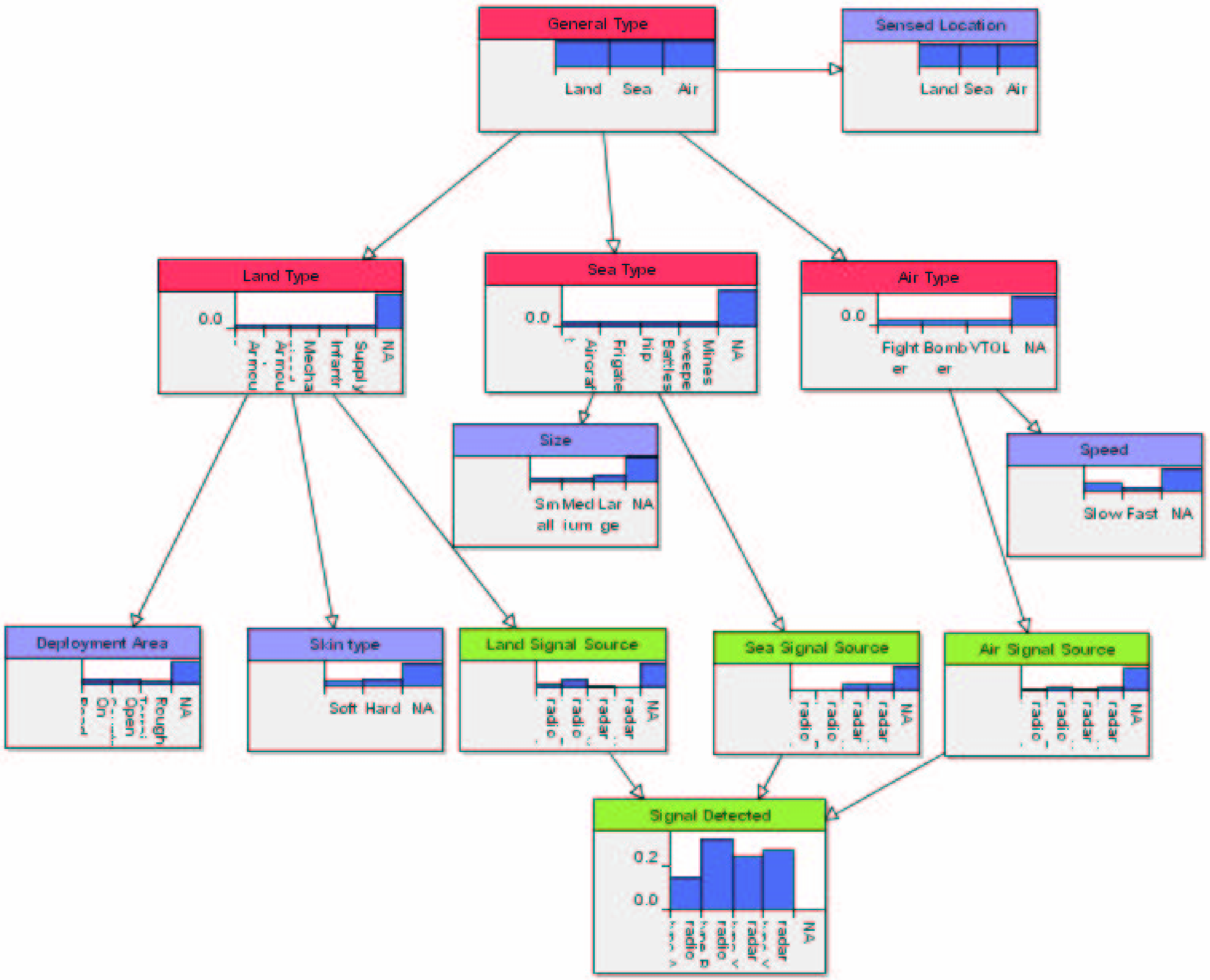
Information Fusion

Classification

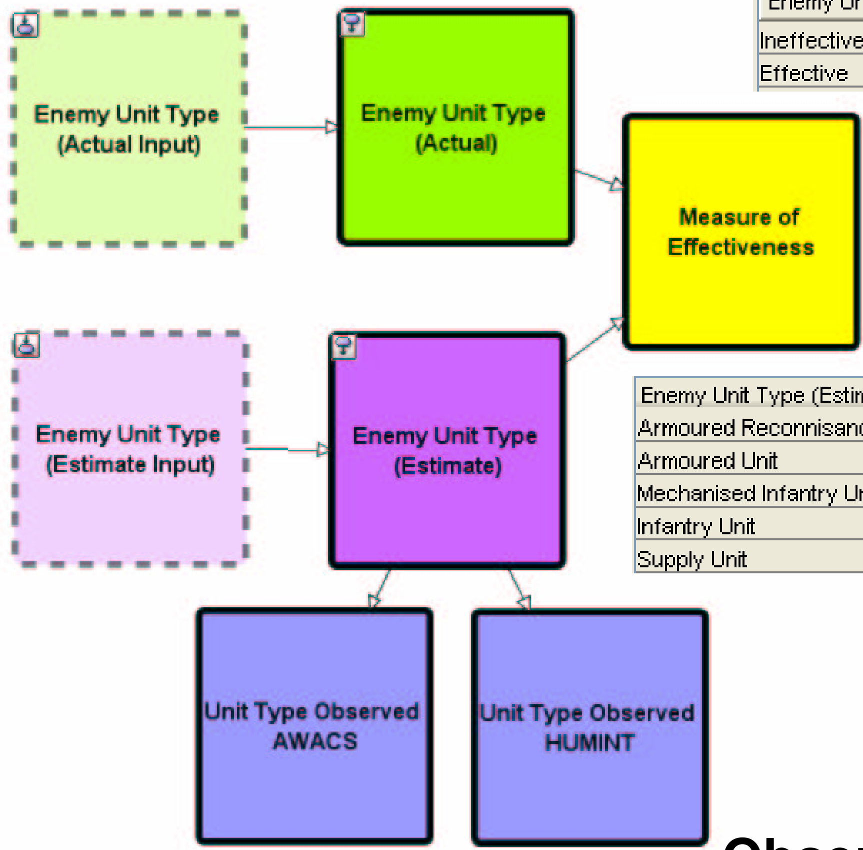
Classification

- Aim to classify hidden attributes of an object using direct or inferred knowledge
- Prior knowledge about possible attribute values (probabilistic)
  – Existence {0, 1} or probability [0,1]
- Classification hierarchy (logical constraints)
  – {Mammal, Dog, Alsatian}
- Effects on other objects (causal)
  – Signals received by sensors (infrared, radar, etc.)
  – Indirect measures from tracking and filtering (max speed, location)
  – Relationship with other objects (proximity to valued asset)

Classification Model Example



Temporal Fusion Model



Enemy Unit ...	Armoured Reconnisance Unit					Armoured Unit				
Enemy Unit ...	Armou...	Armou...	Mecha...	Infantr...	Supply...	Armou...	Armou...	Mecha...	Infantr...	Supply...
Ineffective	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0
Effective	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

MoE Model

Transition Model for Enemy Unit type

Enemy Unit Type (Estimat...	Armoured Recon...	Armoured Unit	Mechanised Infan...	Infantry Unit	Supply Unit
Armoured Reconnisance ...	1.0	0.0	0.0	0.0	0.0
Armoured Unit	0.0	1.0	0.0	0.0	0.0
Mechanised Infantry Unit	0.0	0.0	1.0	0.0	0.0
Infantry Unit	0.0	0.0	0.0	1.0	0.0
Supply Unit	0.0	0.0	0.0	0.0	1.0

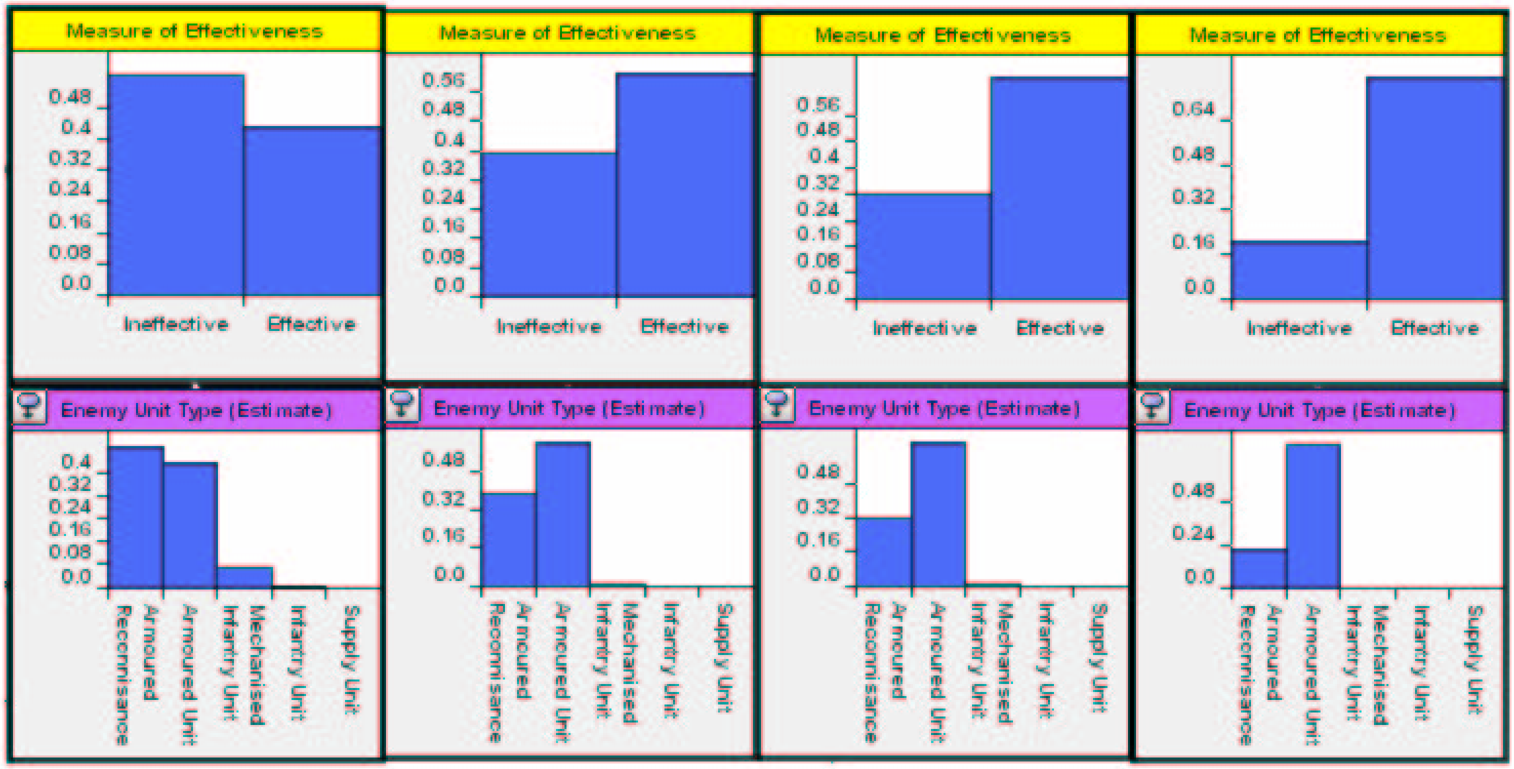
Observation Model for AWACS

Enemy Unit Type (Estimate)	Armoured Reconni...	Armoured Unit	Mechanised Infan...	Infantry Unit	Supply Unit
Armoured Reconnisance Unit	0.4716981	0.30973452	0.1923077	0.13333334	0.009852217
Armoured Unit	0.3301887	0.44247788	0.115384616	0.06666667	0.0049261083
Mechanised Infantry Unit	0.14150943	0.22123894	0.3846154	0.23333333	0.14778325
Infantry Unit	0.047169812	0.022123894	0.26923078	0.33333334	0.3448276
Supply Unit	0.009433962	0.0044247787	0.03846154	0.23333333	0.49261084

Running the model over four time periods

Data:					
	AWACS	AR	A	A	A
	HUMINT	A	A	-	A

p(MoE | Data):



p(Armoured | Data):	43%	61%	68%	79%

Information Fusion & AI: Benefits

- Dynamic Bayesian Network is infinitely more flexible and more general than competing (older) approaches
- Graph model is easy to understand and debug
- Can cope with non-Guassian assumptions
- Supports mix of subjective probabilities, derived from judgement, with observed data
- Copes with mixture of continuous and discrete random variables

Final Remarks

- **Structured Method**
  - Based on 300 year old proven Bayes' theorem
  - Enabled by modern computer power & technology
  - Beyond current statistical & Monte Carlo techniques
  - Combines subjective judgements with data
  - Flexible and general purpose
- **AgenaRisk**
  - Enables scalable, reusable & auditable risk models
  - Integrates easily with DBMS & Excel
  - Enables professional developers to build end-user applications
  - Free 30 Trial Evaluation available from:

    www.AgenaRisk.com