Introduction to Bayesian Statistical Modelling and Analysis

Kerrie Mengersen

Getting Started: Screening for cancer

- Medical diagnostic tests are very rigorous:
  - The chance of a positive test given cancer

    $$\Pr(+|CA) = 0.90$$

  - The chance of a negative test given no cancer

    $$\Pr(-|Not) = 0.95$$

But what is the question?

What is the question?

- From the perspective of the test:

Given a person has cancer, what is the probability that the test is positive?

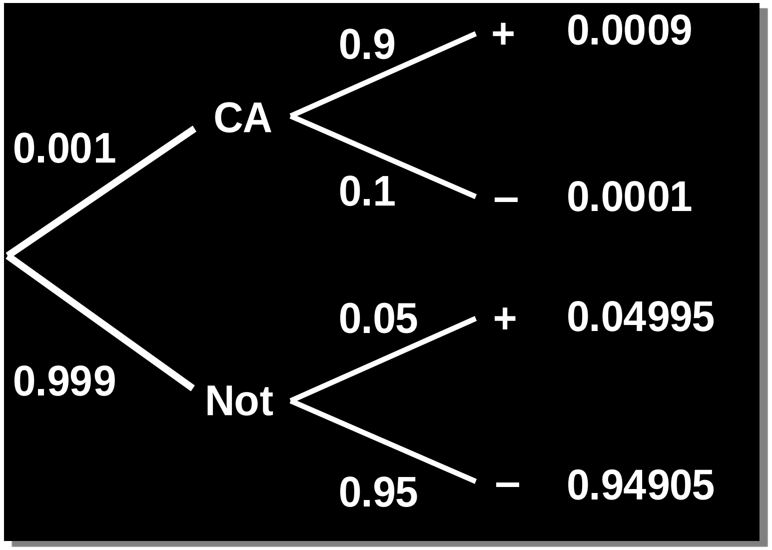
$$\Pr(+|CA) = 0.90$$

- From the perspective of the person:

Given that the test is positive, what is the probability that a person has cancer?

$$\Pr(CA|+) = ?$$

What is the answer? via Bayes' Theorem



Pr(CA|+) = Pr(CA and +) / Pr(+)

Pr(CA and +) = Pr(CA) Pr(+|CA)
= 0.0009

Pr(+) = 0.0009 + 0.04995 = 0.05085

Pr(CA|+) = 0.0009/0.05085 = 0.02

Pr(CA|+) = Pr(+|CA) Pr(CA) / Pr(+)

Your turn

Suppose the diagnostic test has the same sensitivity and specificity but the cancer is more common: 10% of the population has a chance of getting this cancer.

What is the probability that a person has cancer, given that the test is positive?

Solution

$Pr(+|CA) = 0.9$

$Pr(-|CA) = 0.95$

$Pr(CA) = 0.1$

$$Pr(+) = Pr(+|CA)\ Pr(CA) + Pr(+|Not)\ Pr(Not)$$
$$= 0.9*0.1 + 0.05*0.9 = 0.135$$

$$Pr(CA|+) = Pr(+|CA)\ Pr(CA)\ /\ Pr(+)$$
$$= 0.9*0.1\ /\ 0.135 = 0.67$$

Bayesian Modelling

Overall Aim: To learn about unknowns θ, given information *y*

$$p(\theta | y) = p(y | \theta)\, p(\theta) \,/\, p(y)$$

$$p(\theta | y) = p(\theta)\, p(y | \theta) \,/\, p(y)$$

Frequentist approach to modelling

We have some data Y, and want to know about θ

θ can be unknown parameters, missing data, latent variables, etc.

Frequentist: estimate θ through the likelihood: $p(Y|\theta)$

How likely is Y for given values of θ ?

Use moment estimators or maximum likelihood.

But we *really* want to know about $p(\theta|Y)$

⟹ **Bayes**

Why Bayes?

Bayesian methods allow us to:

- Think differently about estimating and interpreting unknown parameters

  "what are possible values of this parameter?"
- Combine prior information with the data

  "what else do I know about this parameter and model?"
- Describe many sources of uncertainty in the model

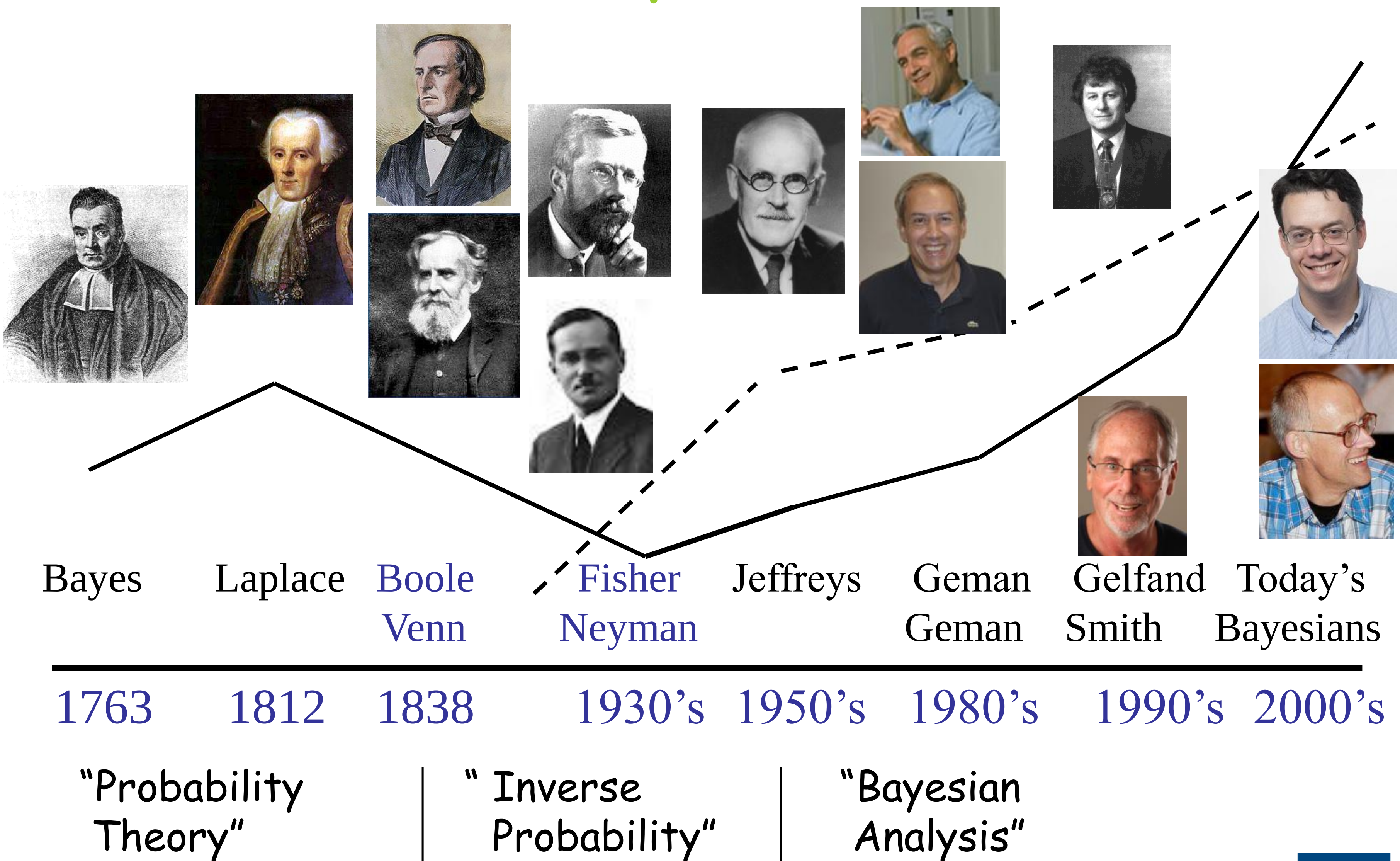
  "how sure am I about the inputs to my model?"
- Describe complex systems using hierarchical or multi-level models

Why Bayes?

Bayesian computational methods (such as MCMC) allow us to:

- Use non-standard distributions
- Fit very complex models
- Obtain a very wide variety of estimates
- Make a very wide range of inferences, based directly on posterior probabilities

Modern Bayesian Statistics



Binomial example

y = number of successes from n trials

Unknown: probability of success θ

$$y|\theta \sim \mathrm{Bin}(n, \theta)$$

$$p(y|\theta) \propto \theta^{y}(1-\theta)^{n-y}$$

Prior

We could use a point prior:
“the population probability can only be 0.1 or 0.3, with 70% and 30% chance respectively”.
That is: $p(\theta=0.1) = 0.7$ and $p(\theta=0.3) = 0.3$

We might not be this certain:
“θ could be any value [between 0 and 1] but based on previous studies it is more likely to be around 0.1, and very unlikely to be larger than 0.8.”

What is an appropriate prior distribution for θ?

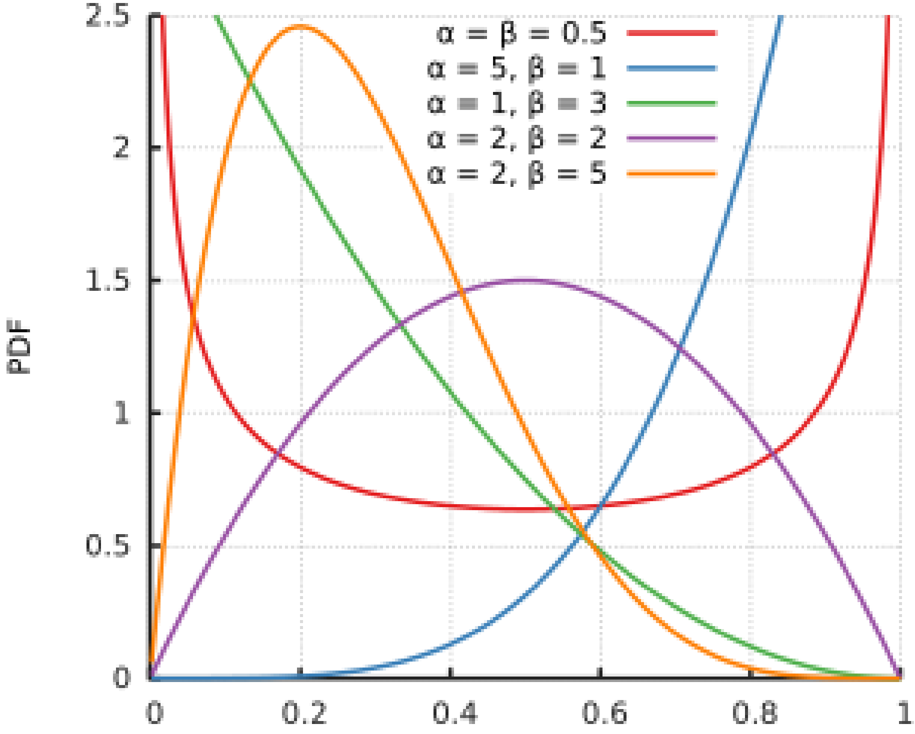
Possible Prior

$$\theta \sim \text{Beta}(\alpha, \beta)$$

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$E(\theta) = \frac{\alpha}{\alpha+\beta}$$

$$\text{Var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

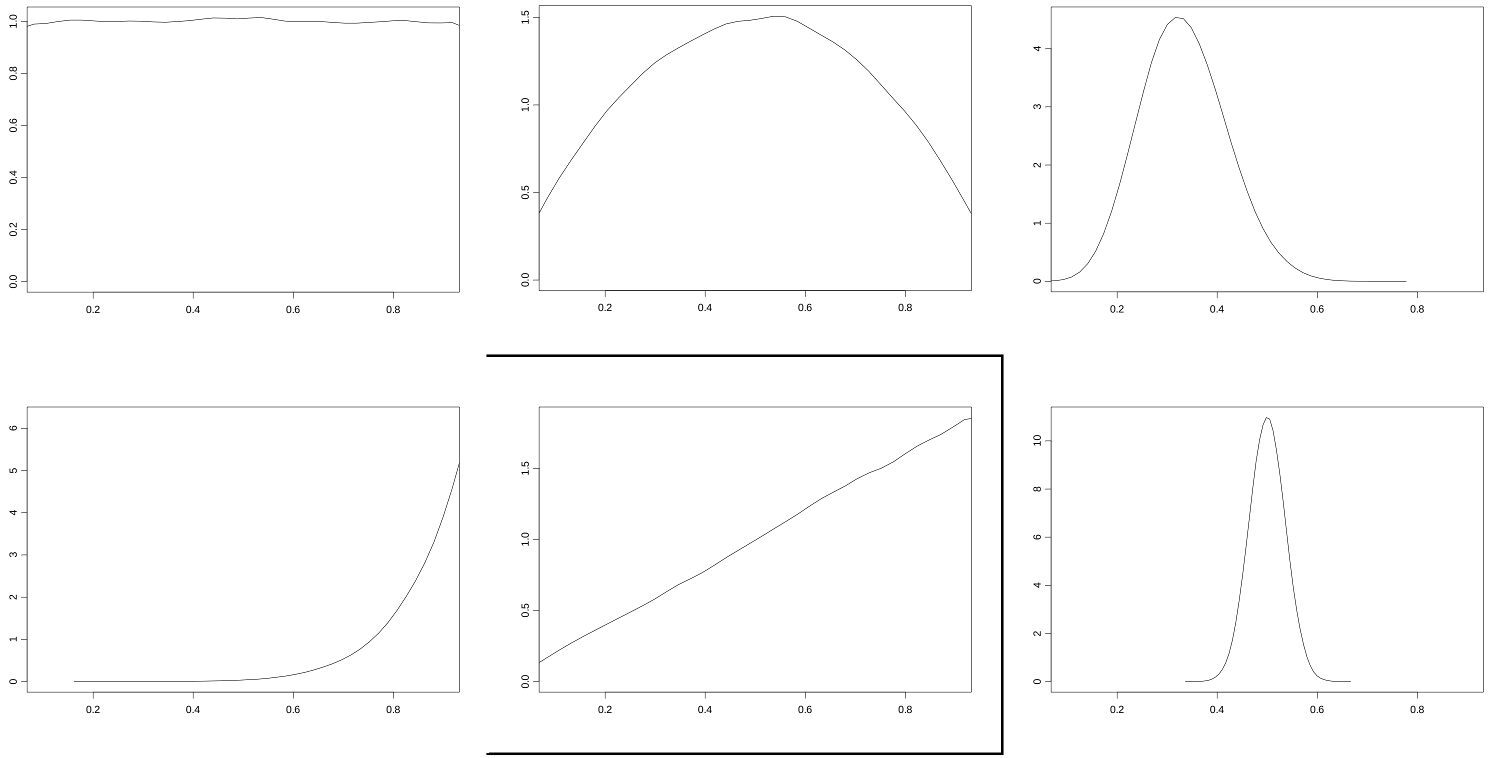


Wikipedia

Beta Distribution

Match the plots to the distributions.

Beta(1,1), Beta(2,2), Beta(100,100),
Beta(2,1), Beta(10,20), Beta(9,1)



Posterior

$$p(\theta \mid y) \propto likelihood \times prior$$

$$= \theta^{y}(1-\theta)^{y}\theta^{a-1}(1-\theta)^{b-1}$$

$$= \theta^{y+a-1}(1-\theta)^{n-y+b-1}$$

$$\theta|y \sim \mathrm{Beta}(y+a, n-y+b)$$

Your turn!

Binomial example with 22 successes, 7 failures:
Consider the following priors for θ:

Beta(1,1) Beta(9,1) Beta(100,100)

Choose one of these:

1. *What is the prior mean for θ?*
2. *What is the posterior distribution for θ?*
3. *What is the posterior mean for θ?*
4. *What general conclusions can you make about the influence of priors and sample size?*

Answers:

Sample proportion = 22/29 = 0.76

Beta(1,1):

Prior mean = (1)/(1+1) = 0.5

Posterior mean = (22+1)/(22+1+7+1) = 0.74

Beta(9,1):

Prior mean = (9)/(9+1) = 0.90

Posterior mean = (22+9)/(22+9+7+1) = 0.79

Beta(100,100):

Prior mean = (100)/(100+100) = 0.5

Posterior mean = (22+100)/(22+100+7+100) = 0.53

Dynamic Updating

If we obtain more data, we do not have to redo all of the analysis: our posterior from the first analysis simply becomes our prior for this next analysis.

Binomial example:

Stage 0. Prior $p(\theta) \sim Beta(1,1)$; ie $E(\theta)=0.5$.

Stage 1. Observe $y=22$ successes from $n=29$ trials.

Likelihood: $p(y|\theta) \sim Bin(n=29,\theta)$;
Posterior: $p(\theta|y) \sim Beta(23,8)$; ie $E(\theta|y) = 0.74$

Stage 2: Observe 5 successes from 10 new trials.

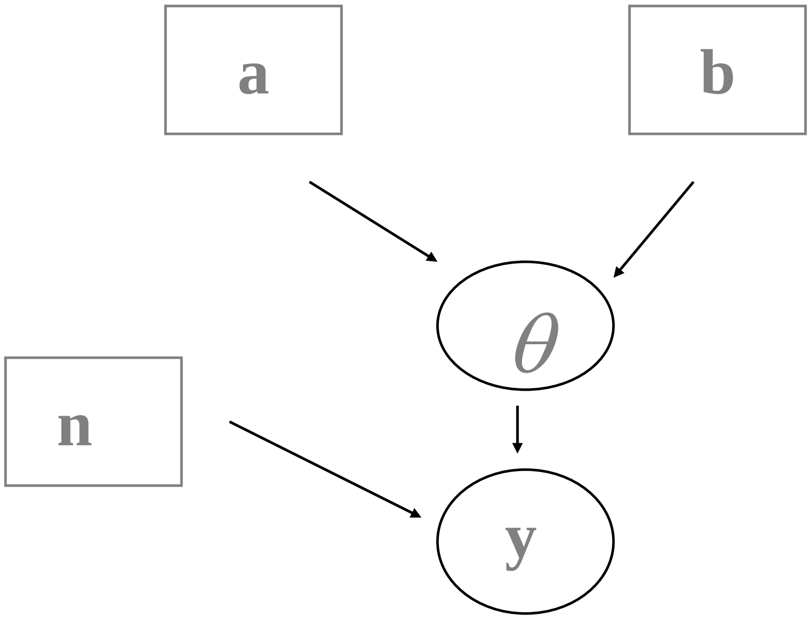
Likelihood: $p(y|\theta) \sim Bin(n=10,\theta)$;
Prior $p(\theta) \sim Beta(23,8)$;
Posterior $p(\theta|y) \sim Beta(28,13)$; ie $E(\theta|y) = 0.68$.

DAG: Binomial model

Model

$$y \sim Binomial\ (\theta, n)$$

$$\theta \sim Beta\ (a,b)$$



Normal Model

Data $D = (y_1, \dots, y_n)$; $\theta = \mu$ unknown; σ_y^2 known

- Likelihood: $D|\mu, \sigma_y^2 \sim N(\mu, \sigma_y^2)$
- Conjugate prior: $\mu \sim N(\mu_0, \sigma_0^2)$
- Posterior: $\mu|D \sim N(\mu_n, \sigma_n^2)$

$$\mu_n = \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{y}}{\sigma_y^2}\right) / \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma_y^2}\right)$$

$$\frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma_y^2}$$

Normal Model cont.

Data $D = (y_1, \ldots, y_n)$; $\theta = \mu$ unknown; σ_y^2 known

- Prior predictive:

$$p(y') = \int p\left(y' \middle| \mu, \sigma_y^2\right) p(\mu)\, d\mu$$
$$= \int N\left(y' | \mu, \sigma_y^2\right) N\left(\mu \middle| \mu_0, \sigma_0^2\right) d\mu$$
$$y' \sim N(\mu_0, \sigma_0^2 + \sigma_y^2)$$

- Posterior predictive:

$$p(y'|D) = \int p\left(y' \middle| \mu, \sigma_y^2, D\right) p(\mu|D) d\mu$$
$$= \int N\left(y' | \mu, \sigma_y^2\right) N(\mu | \mu_n, \sigma_n^2) d\mu$$
$$y' \sim N(\mu_n, \sigma_n^2 + \sigma_y^2)$$

Posterior predictive variance is the uncertainty due to the observation noise σ^2 plus the uncertainty due to the parameters, σ_n^2

Your turn!

1. Suppose that we observe $y = 2$ and wish to estimate the population mean μ.
2. Assume $p(y|\mu) \sim N(\mu, \sigma^2=3)$
   and our prior is $p(\mu) \sim N(\mu_0=0\ ,\ \sigma_0^2=1)$.
3. What is the posterior distribution for μ?
4. What if the prior is N(2,1)? N(0,10)?

5. What happens to the comparative weight of the likelihood and prior as the same size increases? As the prior variance increases?

Answers

- Observe $y = 2$; $p(y|\mu) \sim N(\mu,\sigma^2=3)$

- If prior $p(\mu) \sim N(\mu_0=0 , \sigma_0^2=1)$
  then the posterior is $p(\mu|y) \sim N(\mu_1, \sigma_1^2)$
  posterior mean: $\mu_1 = (0/1 + 2/3) / (1/1 + 1/3) = 0.50$
  posterior variance: $1/\sigma_1^2 = 1/1 + 1/3 + 1.333$ so $\sigma_1^2 = 0.75$

- If prior $p(\mu) \sim N(\mu_0=2 , \sigma_0^2=1)$
  then the posterior is $p(\mu|y) \sim N(\mu_1, \sigma_1^2)$
  $\mu_1 = (2/1 + 2/3) / (1/1 + 1/3) = 2$
  $1/\sigma_1^2 = 1/1 + 1/3 = 1.333$ so $\sigma_1^2 = 0.75$

- If prior $p(\mu) \sim N(\mu_0=0 , \sigma_0^2=10)$
  then the posterior is $p(\mu|y) \sim N(\mu_1, \sigma_1^2)$
  $\mu_1 = (0/10 + 2/3) / (1/10 + 1/3) = 1.54$
  $1/\sigma_1^2 = 1/10 + 1/3 = 0.4433$ so $\sigma_1^2 = 2.31$

Dynamic updating – normal model

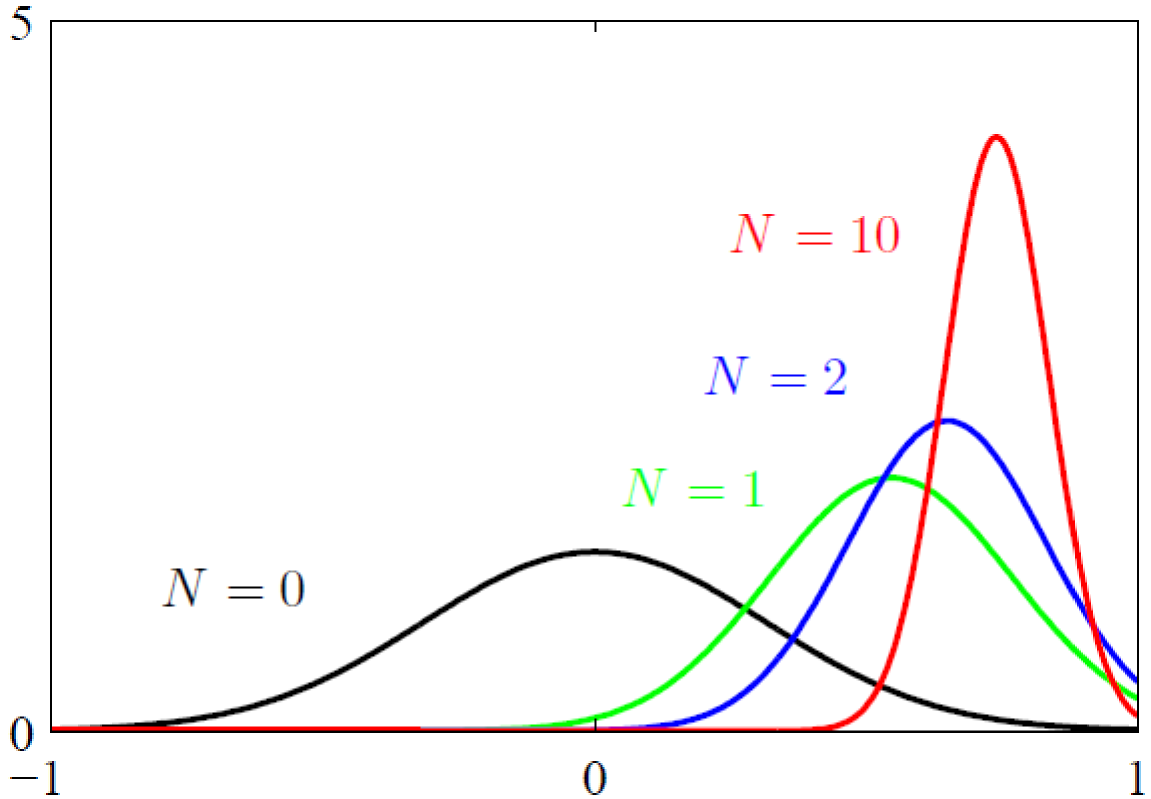


Figure 1: Sequentially updating a Gaussian mean starting with a prior centered on $\mu_0 = 0$. The true parameters are $\mu^* = 0.8$ (unknown), $(\sigma^2)^* = 0.1$ (known). Notice how the data quickly overwhelms the prior, and how the posterior becomes narrower. Source: Figure 2.12 [Bis06].

Kevin Murphy, 2007

Your turn!

Draw a DAG for the Normal model

Hierarchical models

- Useful when observations or parameters have a natural structure.
- Can be used to describe more complex priors.
- Can simplify computational strategies.

$$y_j|\theta_j,\phi \sim P(y_j|\theta_j,\phi)$$

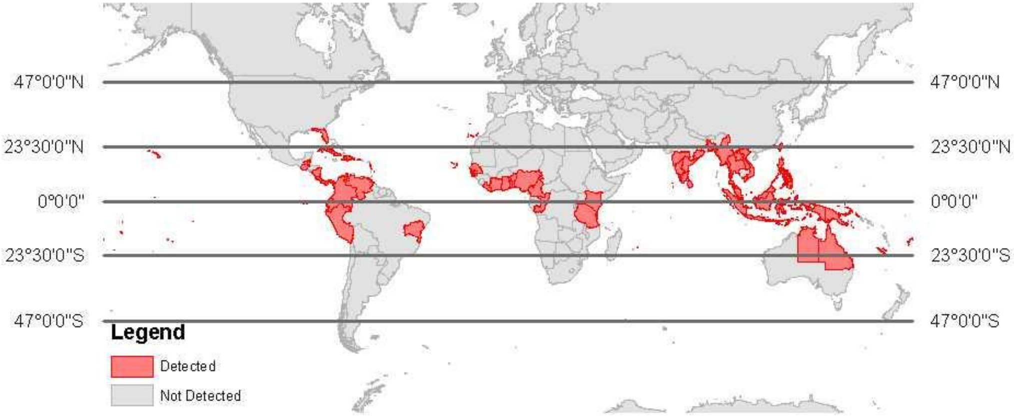
$$\theta_j|\phi \sim P(\theta_j|\phi)$$

$$\phi \sim P(\phi)$$

Hierarchical models in practice: Spiralling whitefly

Stanaway et al.

- The Problem
  - Major tropical plant pest
  - Lives on 100 hosts +
  - Restricts market access to other states
- Information
  - Literature: Characteristics, growth, spread
  - Detectability (inspectors)
  - Surveillance data (> 30 000 records)
- Scope of modelling
  - Local, district and statewide



Countries where spiralling whitefly has been detected. Administrative regions within some countries are shown when documented. Source (CABI 2004, Monteiro et al. 2005, CABI 2006). Personal communications (J.H. Martin, 2008, B.M. Waterhouse, 2008)

Hierarchical Bayesian model

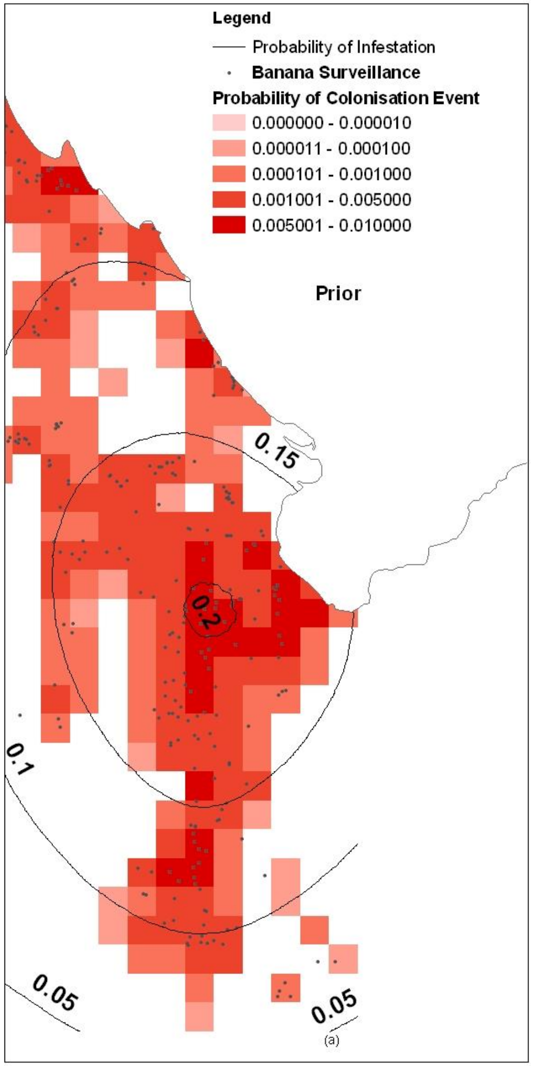
- **Data Model**: Pr(data | incursion process and data parameters)
  - How data is observed given underlying pest extent
- **Process Model**: Pr(incursion process | process parameters)
  - Potential extent given epidemiology / ecology
- **Parameter Model**: Pr(data and process parameters)
  - Prior distribution to describe uncertainty in detectability, exposure, growth …
- **The posterior distribution of the incursion process (and parameters) is related to the prior distribution and data by:**

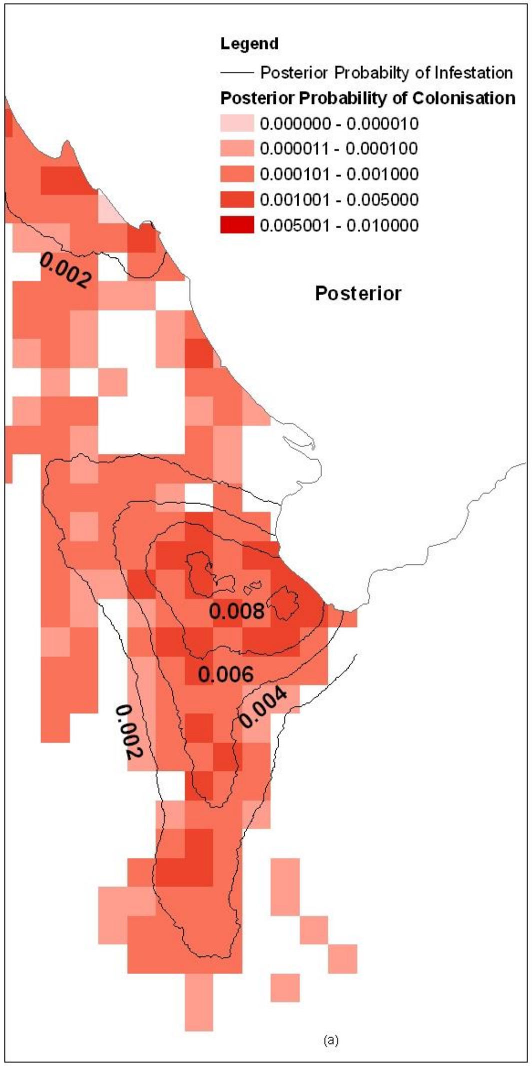
Pr(*process, parameters* | *data*) $\propto$

Pr(*data* | *process, parameters*) Pr(*process* | *parameters*) Pr(*parameters)*

Early Warning Surveillance

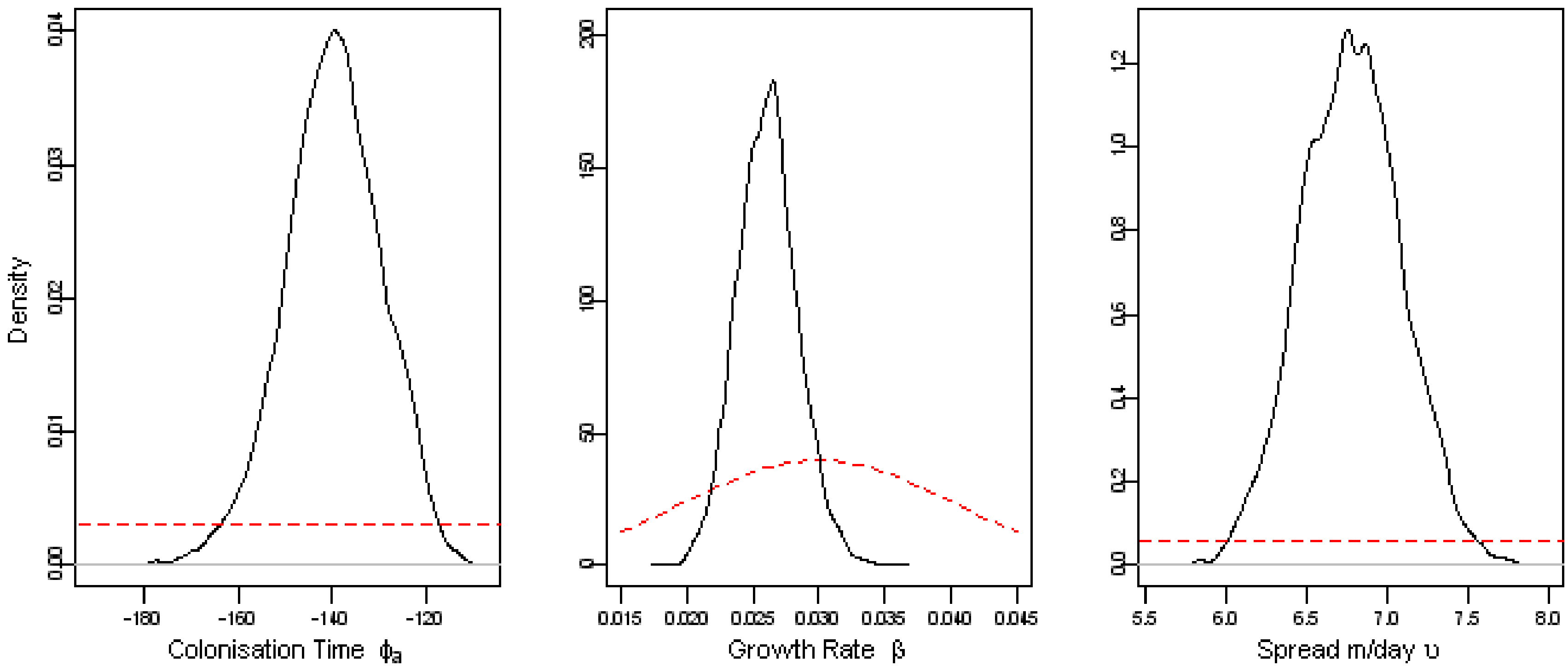
- Priors
- Surveillance data
- Posterior learning
  - modest reduction in area freedom
  - large reduction in estimated extent
  - residual “risk” maps to target surveillance





Invasion Parameter Estimates

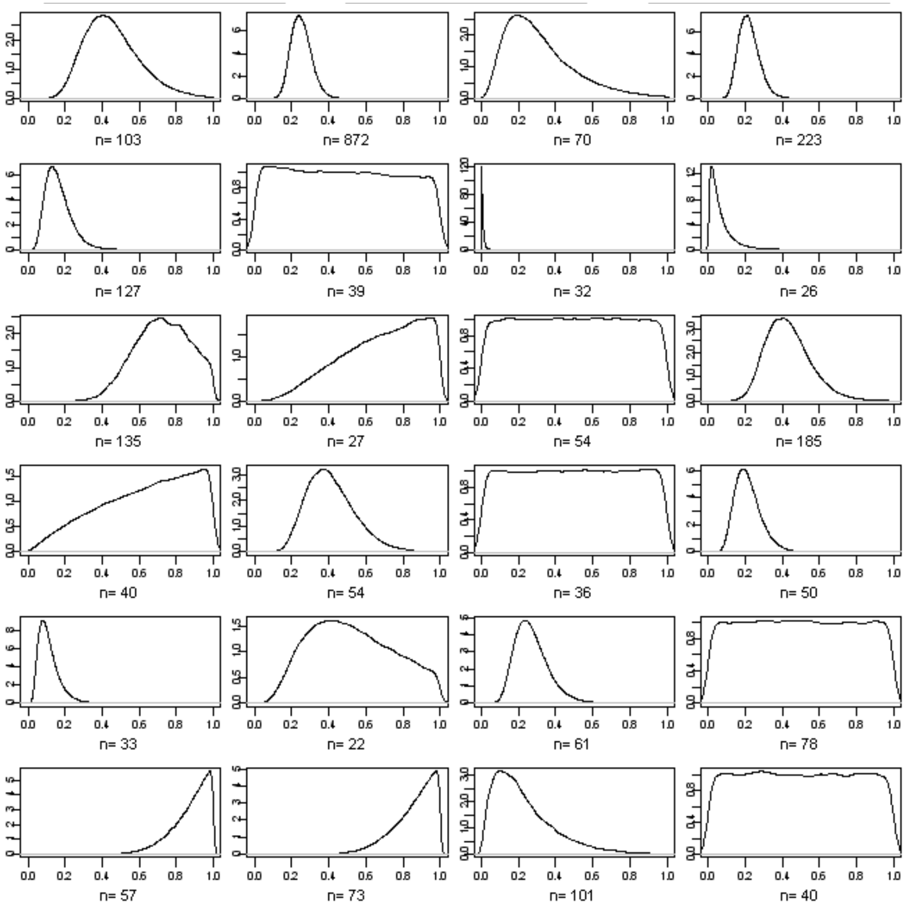
Useful for local management



Observation parameter estimates

Also learn about:

- Host suitability
- Inspector efficiency



Let's talk about Priors

- Conjugate priors
- Uninformative (objective, default)
- Weakly informative (vague)
- Informative

Conjugate priors

Family	Conjugate Prior
$\text{Binomial}(N, \theta)$	$\theta \sim \text{beta}(\alpha, \lambda)$
$\text{Poisson}(\theta)$	$\theta \sim \text{gamma}(\delta_0, \gamma_0)$
$N(\mu, \sigma^2),\ \sigma^2$ known	$\mu \sim N(\mu_0,\ \sigma_0^2)$
$N(\mu, \sigma^2)\ ,\quad \mu$ known, $\tau = 1/\sigma^2$	$\tau \sim \text{gamma}(\delta_0, \gamma_0)$
$\text{gamma}(\alpha, \lambda),\ \alpha$ known	$\lambda \sim \text{gamma}(\delta_0, \gamma_0)$
$\text{Beta}(\alpha, \lambda),\ \lambda$ known	$\alpha \sim \text{gamma}(\delta_0, \gamma_0)$

What did Bayes say?

Bayes: "... it is plain, that in the case of such an event as I there call M, from the number of times it happens and fails in a certain number of trials, without knowing any thing more concerning it, one may give a guess whereabouts it's probability is, and, by the usual methods computing the magnitudes of the areas there mentioned, see the chance that the guess is right. And that the same rule is the proper one to be used in the case of an event concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it, seems to appear from the following consideration; viz. that concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another."

Price, on Bayes' essay: "… the rule must be to suppose the chance the same that it should lie between any two equidifferent degrees"

Criticism: not invariant under transformation, eg if p is uniform in a binomial setup, p^2 is not uniform.

Plausible “objective” priors

Berger (1985); Ghosh (2011)

1. Bayes-Laplace $\theta \sim \text{Uniform}$ Beta(1,1)
2. Jeffreys/reference $\theta \propto \sqrt{\det I(\theta)}$ Beta(½ , ½)
3. Zellner $\theta \propto \theta^0 (1-\theta)^0$

Jeffreys' Prior

Example 1: $y|\mu \sim N(\mu, \sigma^2), \sigma^2$ known

What is the Jeffreys prior for μ?

$$p(y|\mu) = \sqrt{2\pi\sigma^2}\,\exp(-(x-\mu)^2)\,/2\sigma^2)$$

$$p(\mu) \propto \sqrt{I(\mu)} \;\; = \sqrt{E\left[\left(\frac{d}{d\mu}\log f(y|\mu)\right)^2\right]}$$

$$= \sqrt{E\left[\left(\frac{y-\mu}{\sigma}\right)^2\right]} = \sqrt{\int f(y|\mu)\left(\frac{y-\mu}{\sigma^2}\right)^2 dy} = 1$$

Improper, invariant

Jeffreys' Prior

Example 2:

$$y|\sigma \sim N(\mu, \sigma^2), \mu \text{ known}$$

Jeffreys' prior:

$$p(\sigma) \propto 1/\sigma$$

Improper, invariant

Example 3: Your turn!

$$y|\theta \sim \text{Bernoulli}(\theta),\ y\epsilon(0,1),\ 0 \leq \theta \leq 1$$

Jeffreys' prior?

Jeffreys' Prior: Binomial

$$y|\theta \sim \text{Bernoulli}(\theta) \propto \theta^{y}(1-\theta)^{y}$$

$$p(\theta) \propto \sqrt{I(\theta)} = \sqrt{E\left[\frac{y}{\theta} - \frac{(1-y)}{(1-\theta)}\right]^2}$$

$$= \sqrt{\theta\left[\frac{1}{\theta} - \frac{0}{(1-\theta)}\right]^2 + (1-\theta)[\frac{0}{\theta} - \frac{1}{(1-\theta)}]^2}$$

$$= \frac{1}{\sqrt{[\theta(1-\theta)]}}$$

arcsine distribution, Beta(½, ½)

Other priors

- Weakly Informative

$$\mu \sim \mathrm{N}(0, \text{large variance})$$

- Informative (subjective)

$$\mu \sim \mathrm{N}(M, V)$$

Normal model, unknown mean unknown variance

Conjugate:

$$\sigma^{-2} \sim \mathrm{Gamma}(\alpha, \beta)$$

Weakly/strongly informative:

$$\sigma \sim \mathrm{Uniform}(a,b)$$

Normal linear regression

$y = (y_1, .., y_n)^T$; $X = n \times k$ design matrix

$$y_i = {x_i}^T \beta + \varepsilon_i \;\; ; \;\; \varepsilon_i \sim N(0, \sigma^2)$$

or equivalently

$$y_i | \beta, \sigma^2 \sim N(\mu_i, \sigma^2)$$
$$\mu_i = {x_i}^T \beta$$

Conjugate prior

$$y_i|\beta,\sigma^2 \sim N(\mu_i,\sigma^2)$$
$$\mu_i = x_i{}^T\beta$$

Conjugate priors:

$$p(\beta,\sigma^2) = p(\beta|\sigma^2)p(\sigma^2)$$
$$p(\beta|\sigma^2) \sim \mathrm{N}(\beta_0,\sigma^2\Lambda_0^{-1})$$
$$p(\sigma^2) \sim IG(a_0 = \frac{\nu_0}{2}, b_0 = \frac{\nu_0 s_0^2}{2})$$

Posterior under conjugate prior

$$p(\beta, \sigma^2 | y, X) = p(y|\beta, \sigma^2)\, p(\beta|\sigma^2)\, p(\sigma^2)$$

$$p(\beta|\sigma^2, y, X) \sim \mathrm{N}(\mu_1, \sigma^2 \Lambda_1^{-2})$$

$$\mu_1 = (X^T X + \Lambda_0)^{-1} (\Lambda_0 \mu_0 + X^T y)$$

$$\Lambda_1 = (X^T X + \Lambda_0)$$

$$p(\sigma^2 | y, X) \sim IG(a_1, b_1)$$

$$a_1 = a_0 + n/2$$

$$b_1 = b_0 + 0.5(y^T y - \mu_1^T \Lambda_0 \mu_1 + \mu_0^T \Lambda_0 \mu_0)$$

Zellner's g-prior

- Objective prior for β: MVN with covariance matrix proportional to the inverse Fisher information matrix for β.
- The scalar constant g controls the weight assigned to the prior.

Prior: $$\beta|\varphi \sim MVN(\beta_0, g\varphi^{-1}(X^T X)^{-1})$$

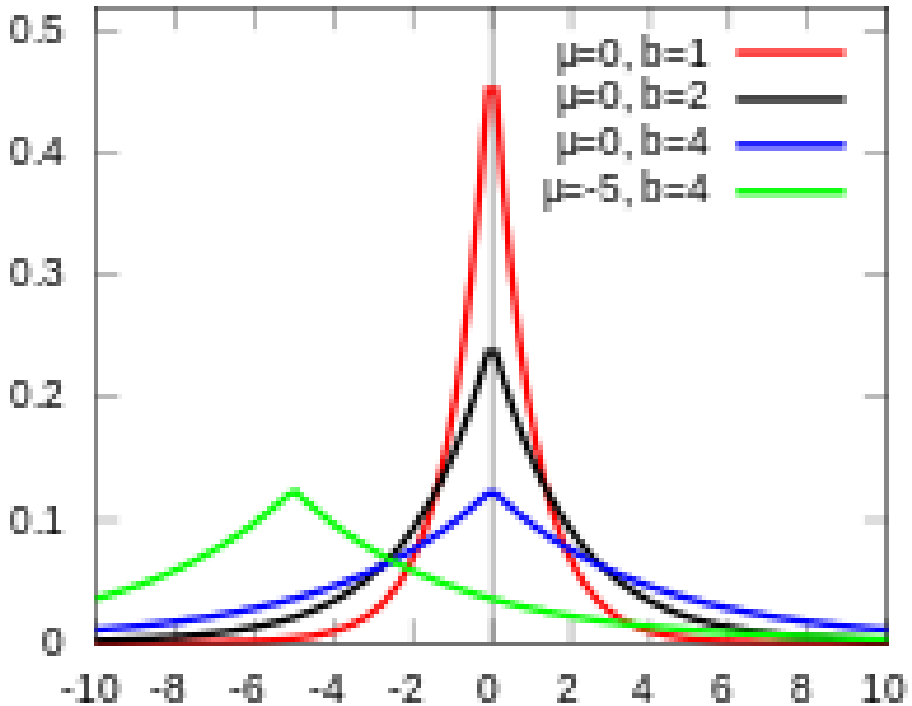
Posterior:

$$\beta|\varphi, x, y \sim MVN(q\hat{\beta} + (1-q)\beta_0, g/\varphi(X^T X)^{-1})$$

$$q = g/(g+1); \;\; \hat{\beta} = (X^T X)^{-1} X^T y$$

High dimensional regression

- Spike and slab priors
- Lasso regression (Laplace priors)
- Elastic net (combine lasso and ridge regression)



Wikipedia

Model Comparison

- Bayes factors, posterior odds

$$\frac{p(M_2|y)}{p(M_1|y)} \propto \frac{p(y|M_2)}{p(y|M_1)} \frac{p(M_2)}{p(M_1)}$$

Bayes Factor
(marginal likelihood ratio,
integrated over parameters)

- BIC, DIC

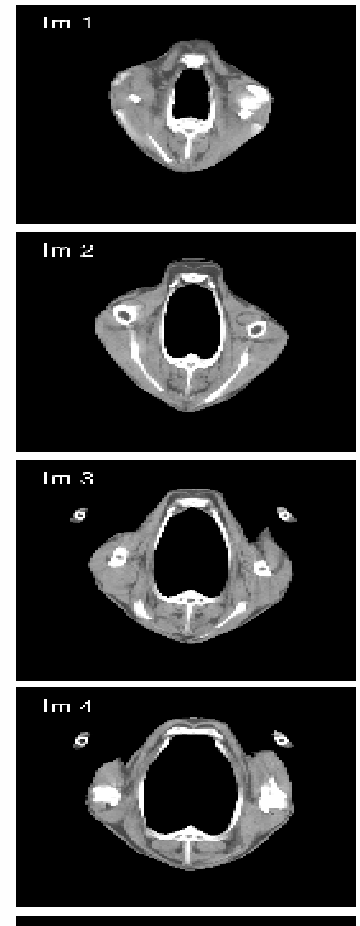
BIC = log P(y|θ*,M) – p/2 log n

- Posterior predictive fit
- Reversible jump MCMC, Birth and death MCMC

- Model averaging

Priors in practice: CAT-scanning sheep

with Alston, Robert, Besag et al.



$$p(y|\theta) = \sum^{K} w_k p(y|\,\theta_k)$$

Bayesian Mixture Models

Robert

$$Likelihood: \quad p(y|\mu,\sigma^2) = \sum_{k=1}^{K} w_k N(\mu_k, \sigma_k^2)$$

$$Priors: \quad \mu_k \sim Normal;\ w_k \sim Dirichlet$$

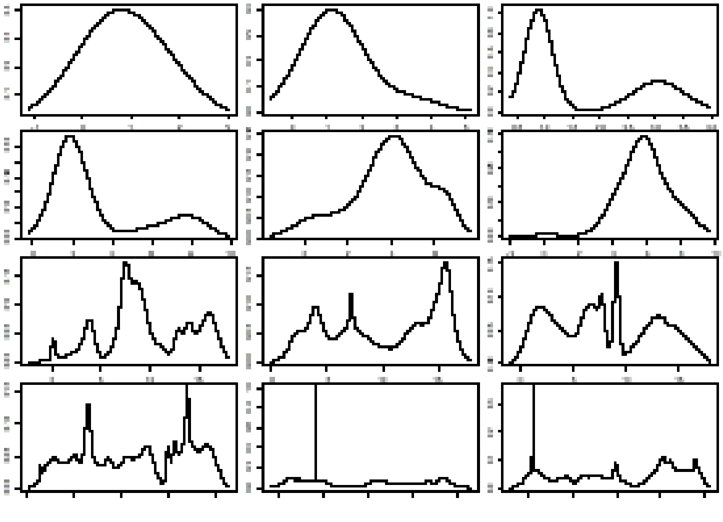
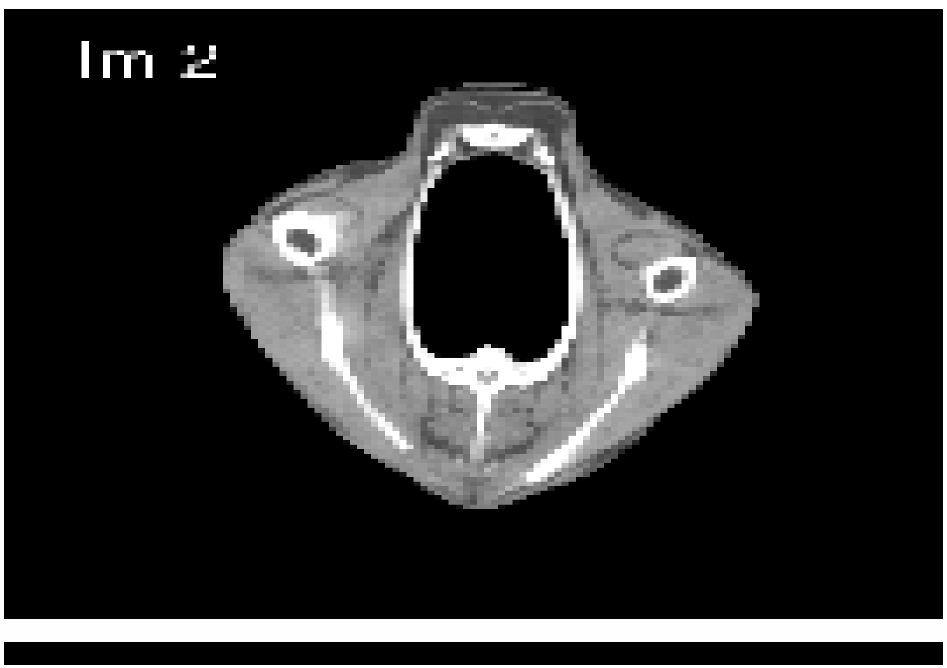


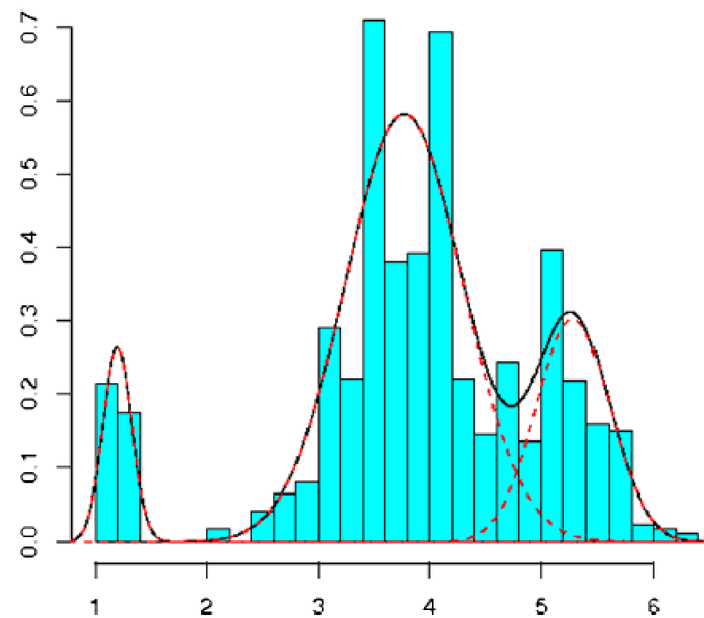
FIGURE 1. Some normal mixture densities for $K = 2$ *(first row)*, $K = 5$ *(second row)*, $K = 25$ *(third row)* and $K = 50$ *(last row)*.

Bayesian Mixture Models

Latent variable approach: ‘break down’ the likelihood

$$y_i|T_i = T \sim N(y|\mu_{T_i}, \sigma^2_{T_i})$$





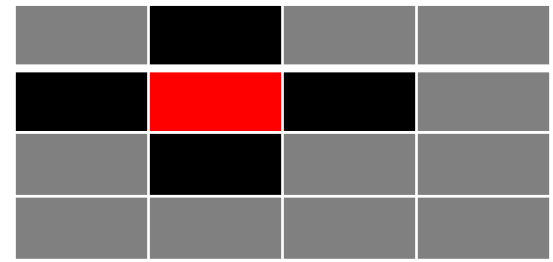
Bayesian Spatial Models

Conditional autoregressive prior:

$$y_i \sim \text{Poisson}(\mu_i)$$
$$\mu_i = E_i\,\theta_i$$

$$\theta_i = \exp(X\beta + \nu_i + \varepsilon_i)$$
$$\varepsilon_i \sim N(0,\ \sigma_\varepsilon^2)$$

$$\nu_i|\nu_j \sim N(\mu_i\ ,\ \sigma_i^2)$$

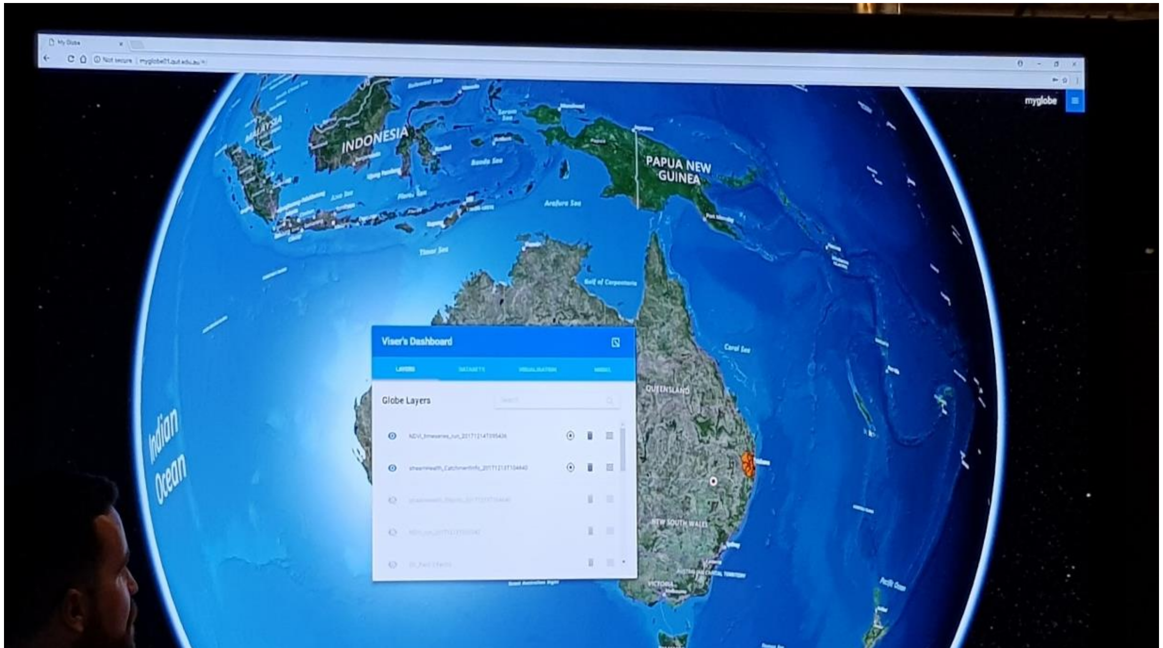


μ_i = weighted sum of neighbouring n's / no. neighbours

$\sigma_i^2 = \sigma^2$ / no. neighbours

From sheep to cancer

with Baade, Duncan Cramb et al.



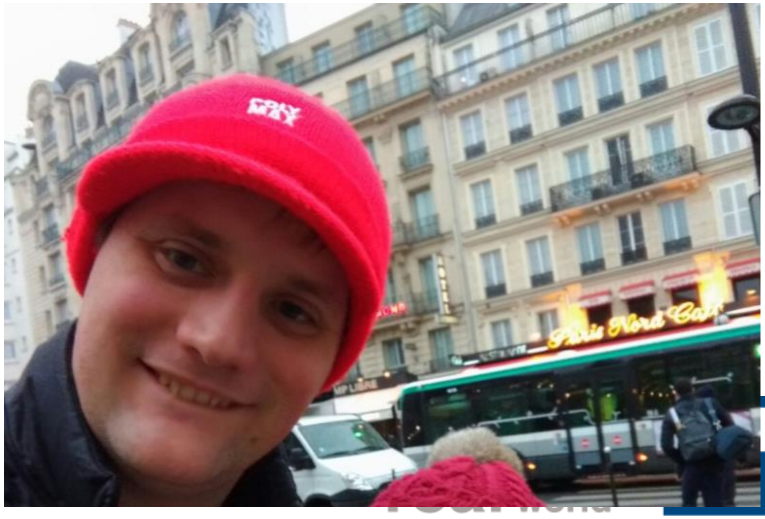
https://www.youtube.com/watch?v=sleXZ7EiSj8
https://atlas.cancer.org.au/

Does “place” impact on cancer?
What is the effect of agricultural/environmental exposures?

From cancer to ICU: predicting death after heart surgery

with Petra Graham

- Demographic (age, gender, hospital length of stay, etc)
- Comorbidity (other diseases)
- Physiological (heart rate, blood pressure, body temperature, etc)
- Biochemical (white cell count, potassium, sodium, etc)
- Treatment (use of mechanical ventilation, etc)
- Outcome (dead/alive after 30 days)



Adaptive logistic regression model

$$y_i \in (0,1); \quad i = 1, \ldots, n$$

$$y_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit}(p_i) = \log(p_i/(1 - p_i)) = X\beta + e_i$$

$$e_i \sim N(0, \sigma_e^2)$$

$$\beta \sim N(\beta_0, \sigma_0^2)$$

U.S.A. ICU risk score

“Doing” Bayesian Analysis: Markov chain Monte Carlo

- “Decompose” joint posterior distribution into a sequence of conditional distributions – these are often much simpler (eg, simple univariate normals, etc)

- Simulate from each conditional distribution in turn. We use a simulation method that resembles a Markov chain (so that the new simulated value relies only on the previous value), giving a set of simulated values

  $$\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(i)}, \ldots$$

  which converges to the required conditional,
  The resulting simulations will come from the required joint distribution.

- We can use Markov chain theory to make statements about behaviour and convergence of the chain

Markov chain Monte Carlo

- "Decompose" joint posterior distribution into a sequence of conditional distributions – these are often much simpler (eg, simple univariate normals, etc)

- Simulate from each conditional distribution in turn. We use a simulation method that resembles a Markov chain (so that the new simulated value relies only on the previous value), giving a set of simulated values

  $$\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(i)}, \ldots$$

  which converges to the required conditional,
  Under mild assumptions, the resulting simulations will come from the required joint distribution.

- Use Markov chain theory to make statements about behaviour and convergence of the chain

MCMC Algorithms

- **Gibbs sampling**: sample from the conditionals themselves
- **Metropolis-Hastings**: sample from an “easy” distribution and accept those values that conform to the conditional distribution
- Lots of variations: reversible jump, slice sampling, particle filters, perfect sampling, adaptive rejection sampling, etc
- Need to ensure conditions, eg *detailed balance*, *reversibility*

Gibbs sampling

Suppose you have a joint posterior $\boldsymbol{p(\theta_1, \theta_2|y, \ldots)}$

0. Choose starting values $\theta_1^{(0)}, \theta_1^{(0)}$
1. At *i*th iteration $^{(i)}$

   Sample $\theta_1^{(i)}$ from $p(\theta_1 \mid \theta_2, y, \ldots)$

   Sample $\theta_2^{(i)}$ from $p(\theta_2 \mid \theta_1, y, \ldots)$
2. Repeat step 1 many times
3. Make inferences based on simulated values

Example of Gibbs Sampler

Consider a single observation (y_1, y_2) from a bivariate normal population with unknown mean (θ_1, θ_2) and known variance-covariance matrix

$$\begin{bmatrix} \sigma_1^2 & \rho^2 \\ \rho^2 & \sigma_2^2 \end{bmatrix}$$

Let $\sigma_1^2 = \sigma_2^2 = 1$

With a uniform prior on θ_1, θ_2, the posterior distribution is

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \Bigg| \; y \sim N\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho^2 \\ \rho^2 & \sigma_2^2 \end{pmatrix} \right)$$

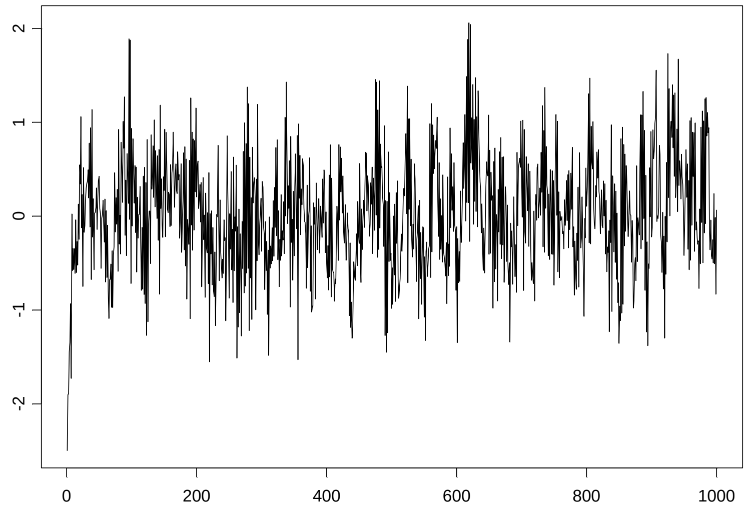
Gibbs sampler

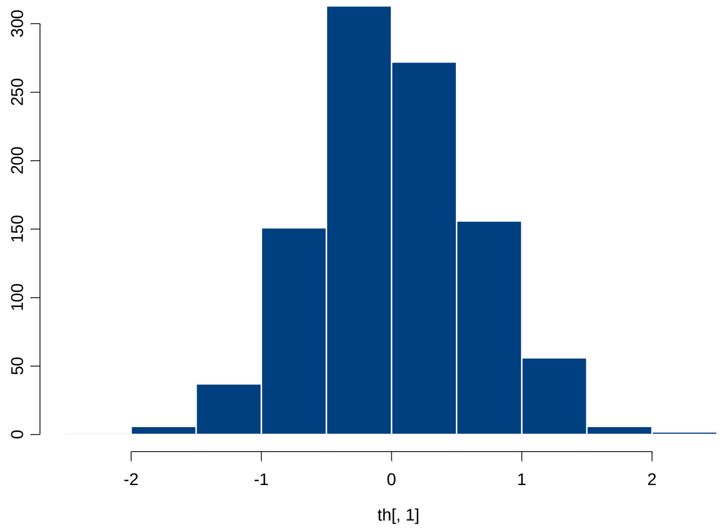
Sample from

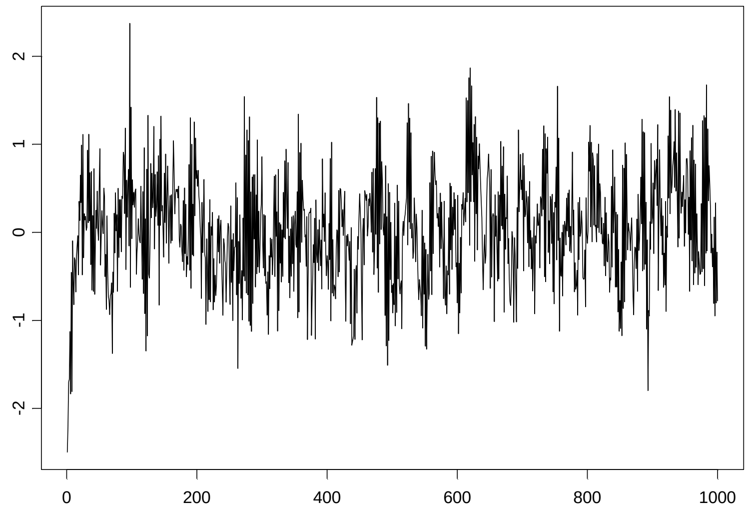
$$\theta_1 \mid \theta_2, y \sim N(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2)$$

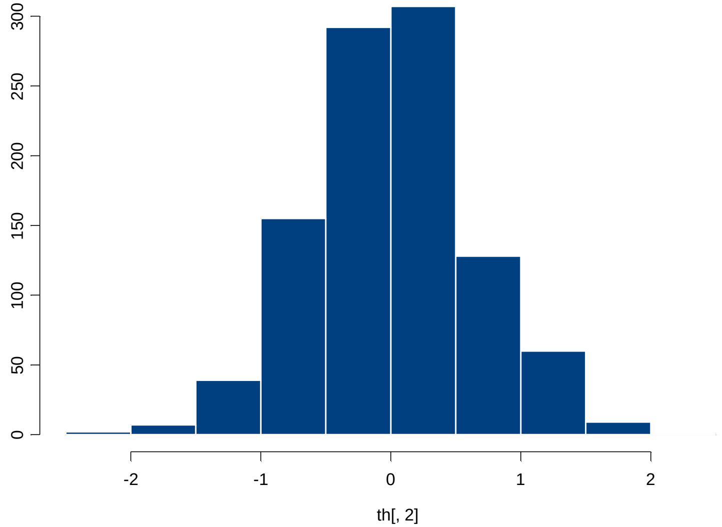
Then sample from

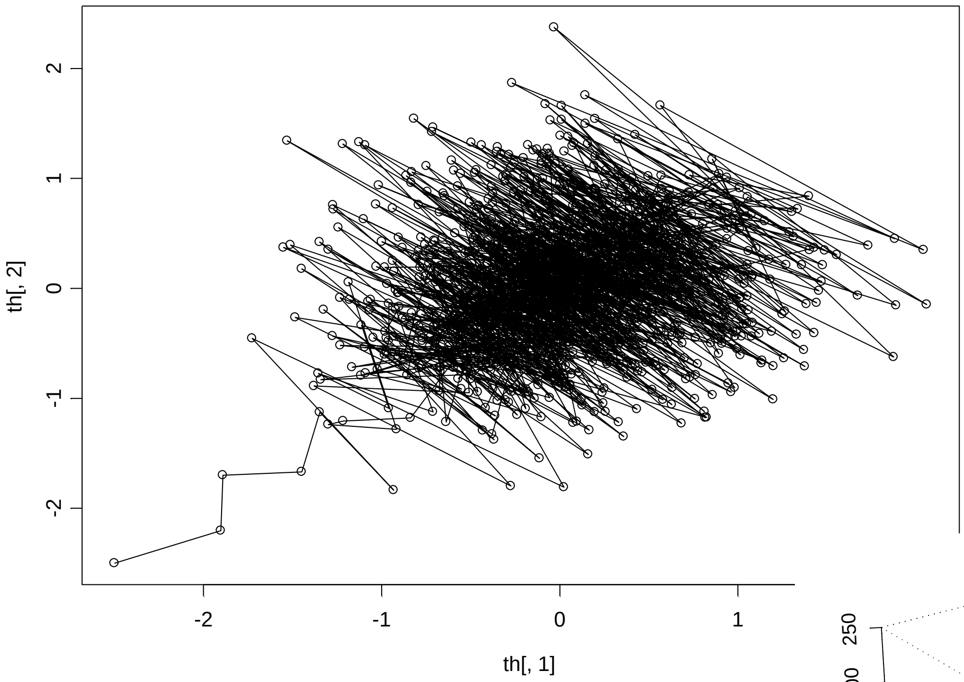
$$\theta_2 \mid \theta_1, y \sim N(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2)$$


2
1
0
-1
-2
0
200
400
600
800
1000

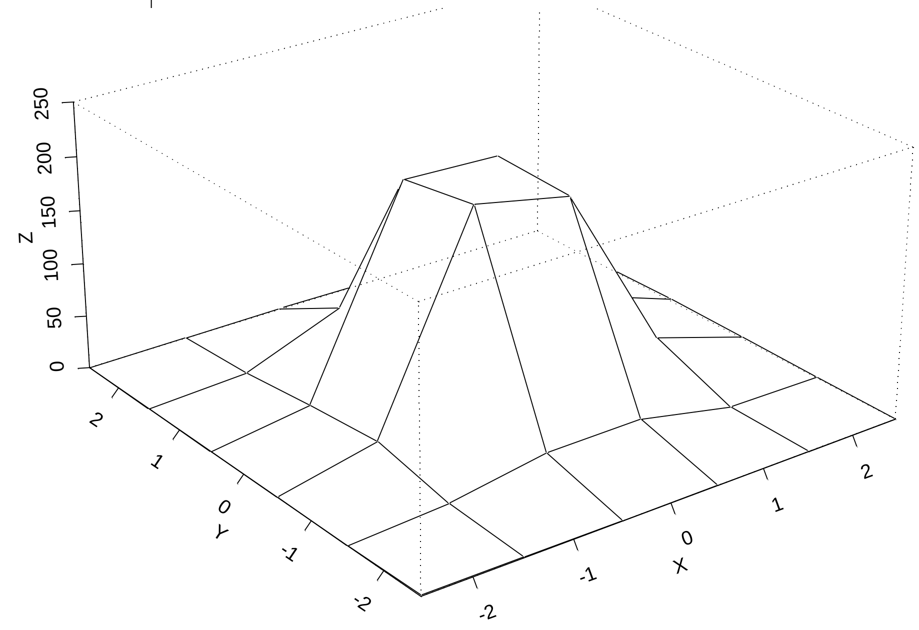

300
250
200
150
100
50
0
-2
-1
0
1
2
th[, 1]


2
1
0
-1
-2
0
200
400
600
800
1000


300
250
200
150
100
50
0
-2
-1
0
1
2
th[, 2]



th[, 2]
th[, 1]
-2
-1
0
1
2



Z
0
50
100
150
200
250
X
Y
-2
-1
0
1
2

Gibbs sampling for mixtures

0. *Initialisation:* Choose $\underline{p}^{(0)}$ and $\underline{\theta}^{(0)}$ arbitrarily

For t=1,…

1.1 Allocate observations to components:
Generate $T^{(t)}$ for each observation

1.2 Generate new weights for the components:
Generate $\underline{p}^{(t)}$

1.3 Generate new parameters for each component:
Generate $\underline{\theta}^{(t)}$

Other computational methods

- Other MCMC algorithms
  - Metropolis Hastings
  - Slice Sampling

- Approximations
  - Variational Bayes
  - INLA
  - ABC
  - SMC
  - Gaussian Processes

Bayesian Software

BUGS: Bayesian analysis using Gibbs sampling (OpenBUGS)
http://www.mrc-bsu.cam.ac.uk/software/bugs/
http://www.openbugs.net

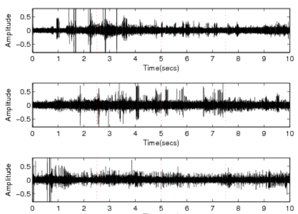
JAGS: Just Another Gibbs Sampler (RJAGS)
http://mcmc-jags.sourceforge.net/
https://www.r-bloggers.com/getting-started-with-jags-rjags-and-bayesian-modelling/

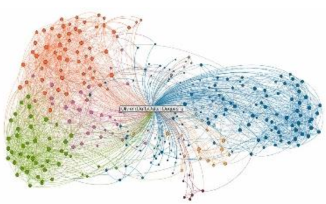
STAN (interface with R, Matlab, etc)
http://mc-stan.org/

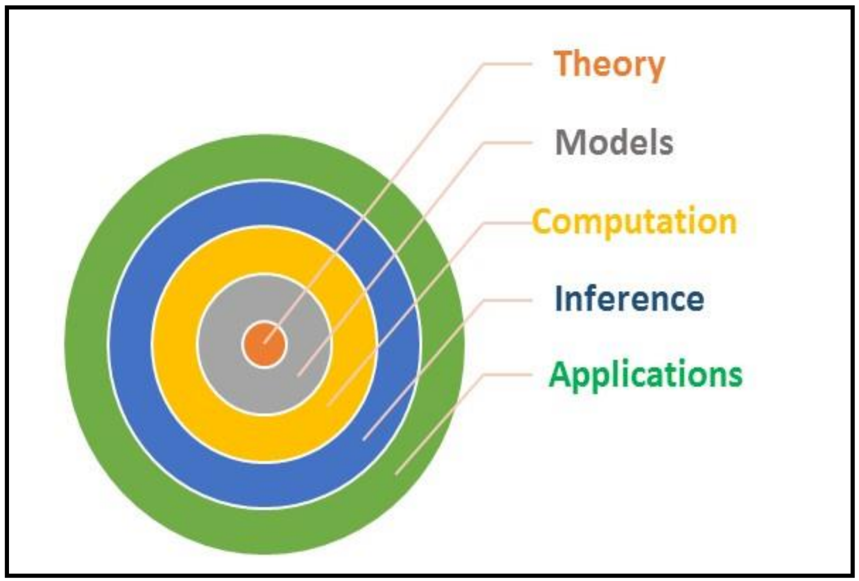
INLA: integrated nested Laplace approximation
www.r-inla.org

R packages
http://cran.r-project.org/web/views/Bayesian.html

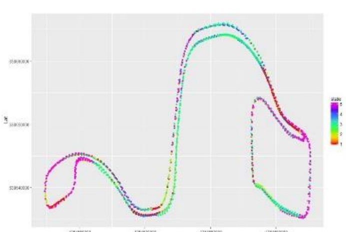
Bayes + Big Data

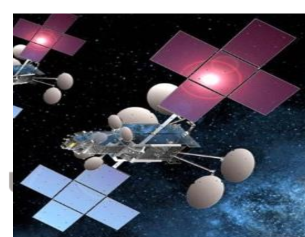






	age	sex
)	21.79603	female
)	21.71116	female
)	21.27584	male
)	27.37577	female
)	19.66598	male
)	24.37509	female





Meeting the challenge

Models:
- Probabilistic
- Regularised
- Flexible
- Robust
- Transferable
- Adaptive

Computation:
- Scalable (parallelisable)
- Subsampling
- Pre-computable
- Approximations (eg. ABC, SMC, VI, GP)

Inference:
- Estimation
- Optimisation
- Uncertainty quantification
- Testing
- Model averaging

"In the past ten years, it's hard to find anything that doesn't advocate a Bayesian approach." -Nate Silver

Monitoring the Great Barrier Reef

with Peterson et al.

Crown of thorns starfish pest 'the greatest impact on Great Barrier Reef'
By Kirsty Nancarrow
Posted Fri 14 Jun 2013

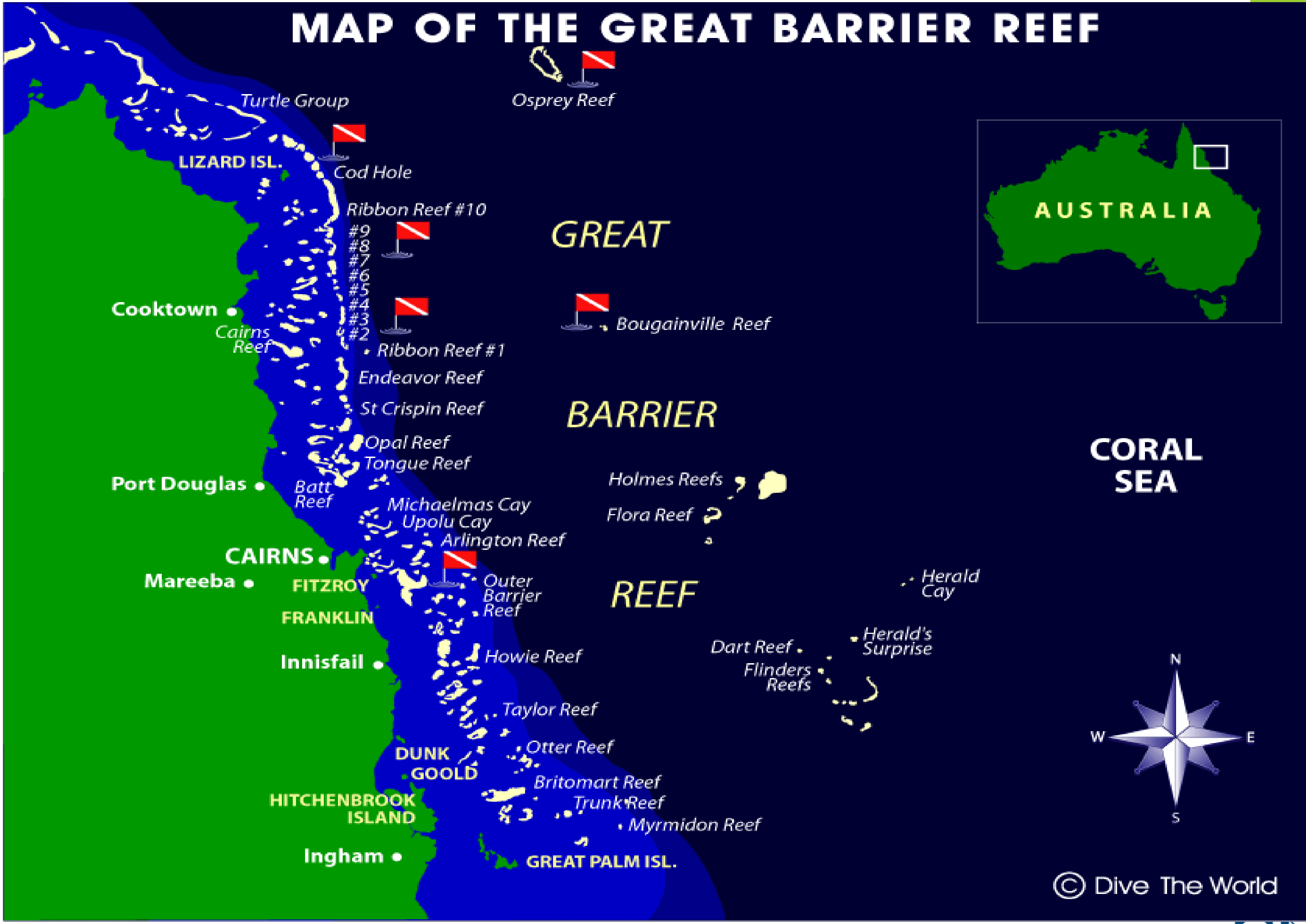
Jan 29 2014
Concerned scientists demand rejection of permit for dumping dredge spoil in Great Barrier Reef's waters

Cyclone damage devastating to the Great Barrier Reef

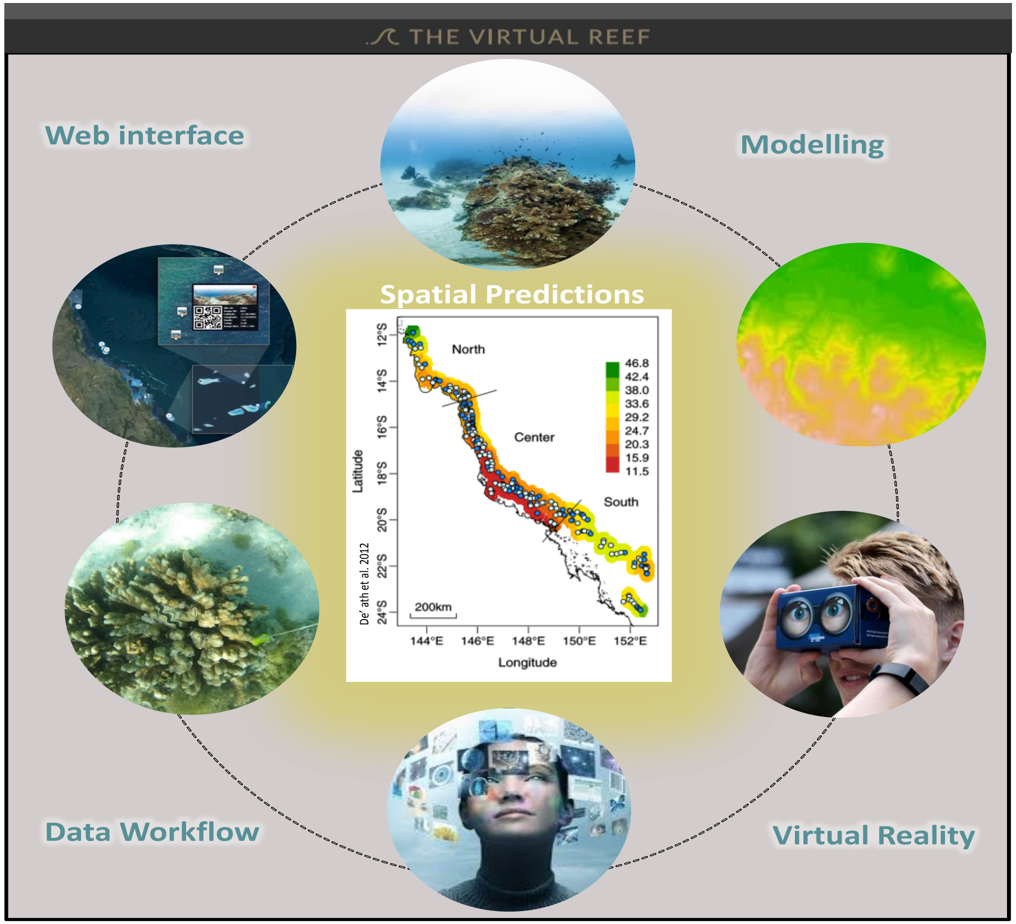
Sep 2 2013
Thousands gather for Rally for the Reef

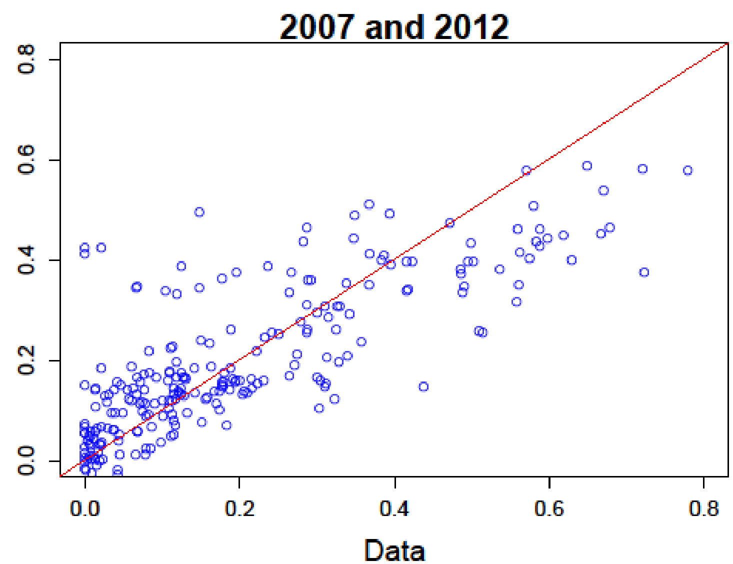
The world's largest living structure
Almost 3000 reefs
Over 2,600 km

Traditional data source: surveys


MAP OF THE GREAT BARRIER REEF
Osprey Reef
Turtle Group
LIZARD ISL.
Cod Hole
Ribbon Reef #10
#9
#8
#7
#6
#5
#4
#3
#2
GREAT
AUSTRALIA
Cooktown
Cairns Reef
Bougainville Reef
Ribbon Reef #1
Endeavor Reef
St Crispin Reef
BARRIER
CORAL SEA
Opal Reef
Tongue Reef
Port Douglas
Batt Reef
Holmes Reefs
Michaelmas Cay
Upolu Cay
Flora Reef
Arlington Reef
CAIRNS
Mareeba
FITZROY
Outer Barrier Reef
REEF
Herald Cay
FRANKLIN
Dart Reef
Herald's Surprise
Innisfail
Howie Reef
Flinders Reefs
N
W
E
S
Taylor Reef
Otter Reef
DUNK
GOOLD
Britomart Reef
HITCHENBROOK ISLAND
Trunk Reef
Myrmidon Reef
Ingham
GREAT PALM ISL.
© Dive The World

Virtual Reef Diver



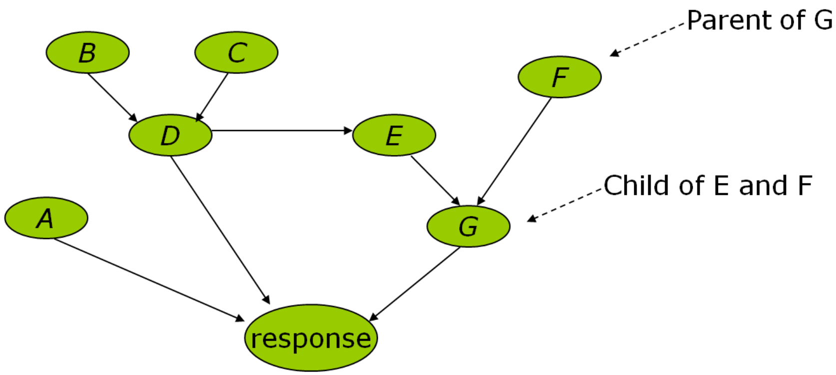


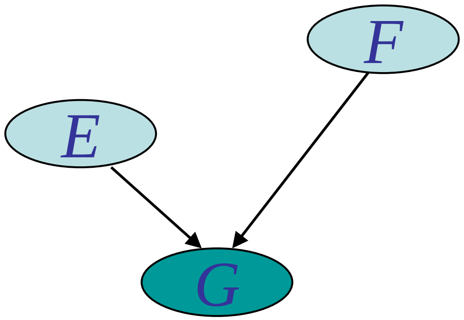
https://www.virtualreef.org.au/

Modelling complex systems

Environment
Biology
Social
Govt
Economic
Human impact
Unknowns
External factors

Bayesian Networks



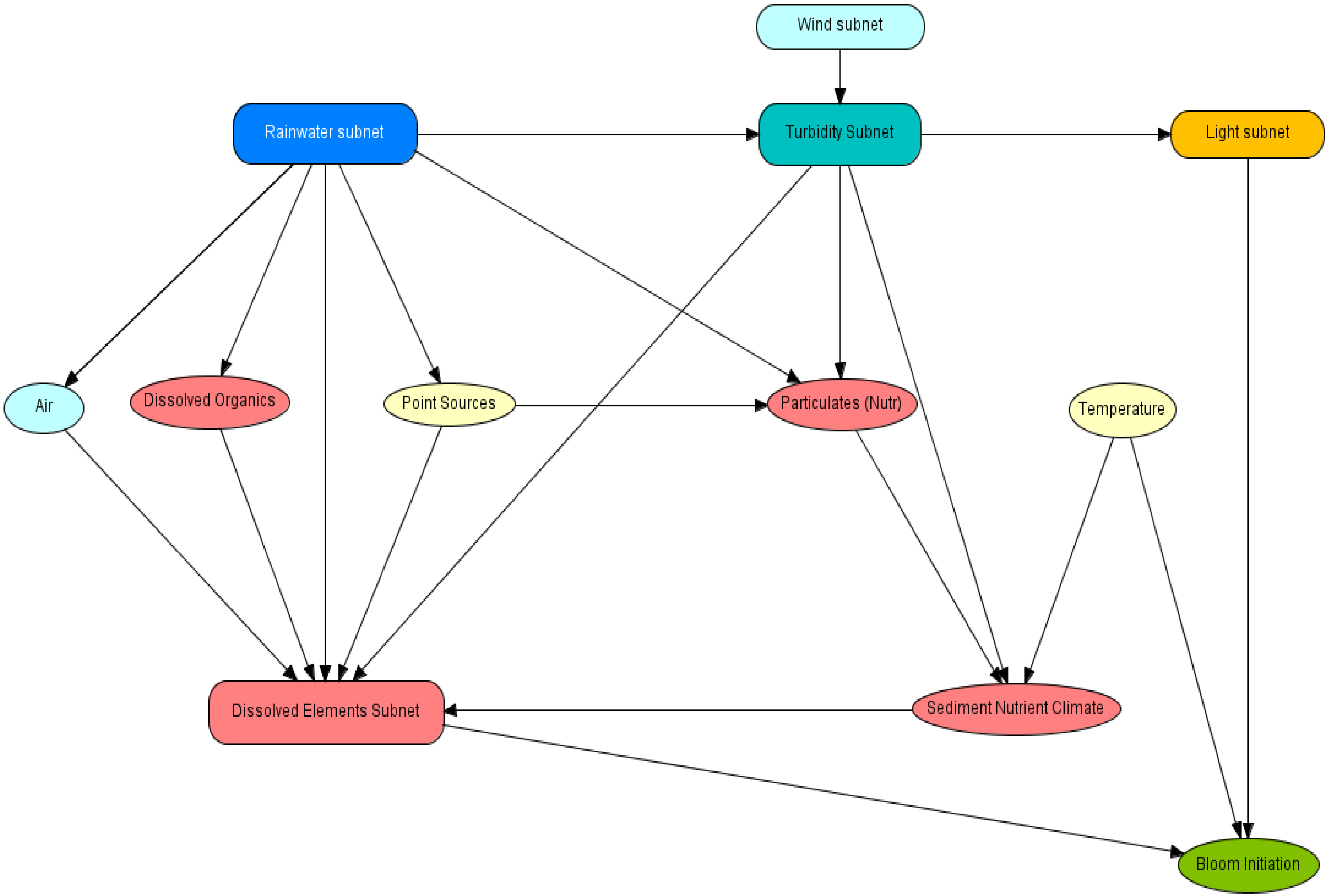


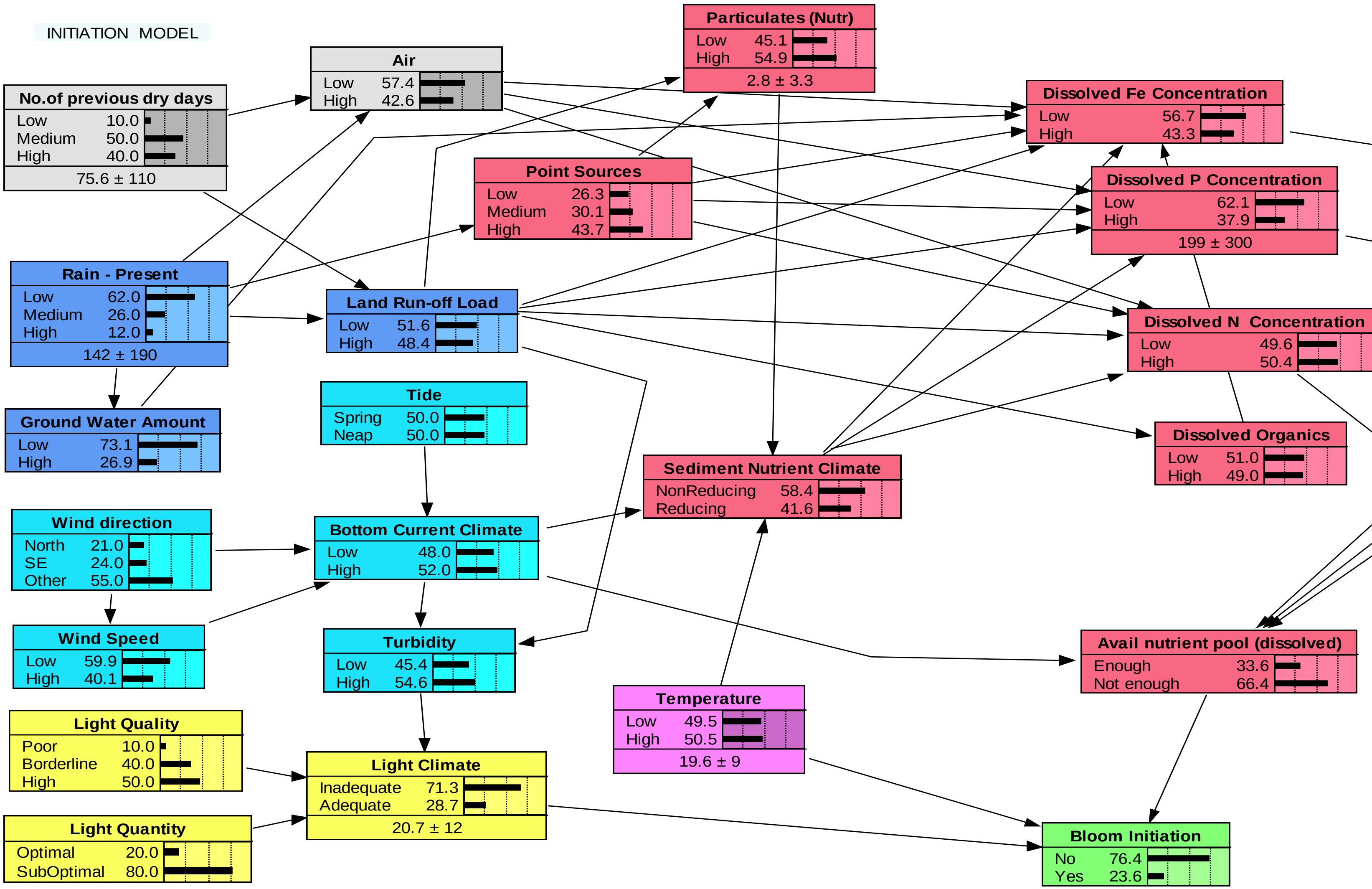
F	
low	0.7
medium	0.2
high	0.1

		G	
E	*F*	*normal*	*high*
yes	*low*	0.4	0.6
	medium	0.2	0.8
	high	0.1	0.9
no	*low*	0.5	0.5
	medium	0.6	0.4
	high	0.4	0.6

Managing lyngbya in Moreton Bay, Australia

- *What is the overall scientific consensus about the drivers of lyngbya?*
- *What management actions should be taken to reduce lyngbya in the Bay?*


Wind subnet
Rainwater subnet
Turbidity Subnet
Light subnet
Air
Dissolved Organics
Point Sources
Particulates (Nutr)
Temperature
Dissolved Elements Subnet
Sediment Nutrient Climate
Bloom Initiation



INITIATION MODEL
No.of previous dry days
Low 10.0
Medium 50.0
High 40.0
75.6 ± 110
Air
Low 57.4
High 42.6
Particulates (Nutr)
Low 45.1
High 54.9
2.8 ± 3.3
Dissolved Fe Concentration
Low 56.7
High 43.3
Point Sources
Low 26.3
Medium 30.1
High 43.7
Dissolved P Concentration
Low 62.1
High 37.9
199 ± 300
Rain - Present
Low 62.0
Medium 26.0
High 12.0
142 ± 190
Land Run-off Load
Low 51.6
High 48.4
Dissolved N Concentration
Low 49.6
High 50.4
Ground Water Amount
Low 73.1
High 26.9
Tide
Spring 50.0
Neap 50.0
Dissolved Organics
Low 51.0
High 49.0
Sediment Nutrient Climate
NonReducing 58.4
Reducing 41.6
Wind direction
North 21.0
SE 24.0
Other 55.0
Bottom Current Climate
Low 48.0
High 52.0
Wind Speed
Low 59.9
High 40.1
Turbidity
Low 45.4
High 54.6
Avail nutrient pool (dissolved)
Enough 33.6
Not enough 66.4
Temperature
Low 49.5
High 50.5
19.6 ± 9
Light Quality
Poor 10.0
Borderline 40.0
High 50.0
Light Climate
Inadequate 71.3
Adequate 28.7
20.7 ± 12
Bloom Initiation
No 76.4
Yes 23.6
Light Quantity
Optimal 20.0
SubOptimal 80.0

Most influential factors

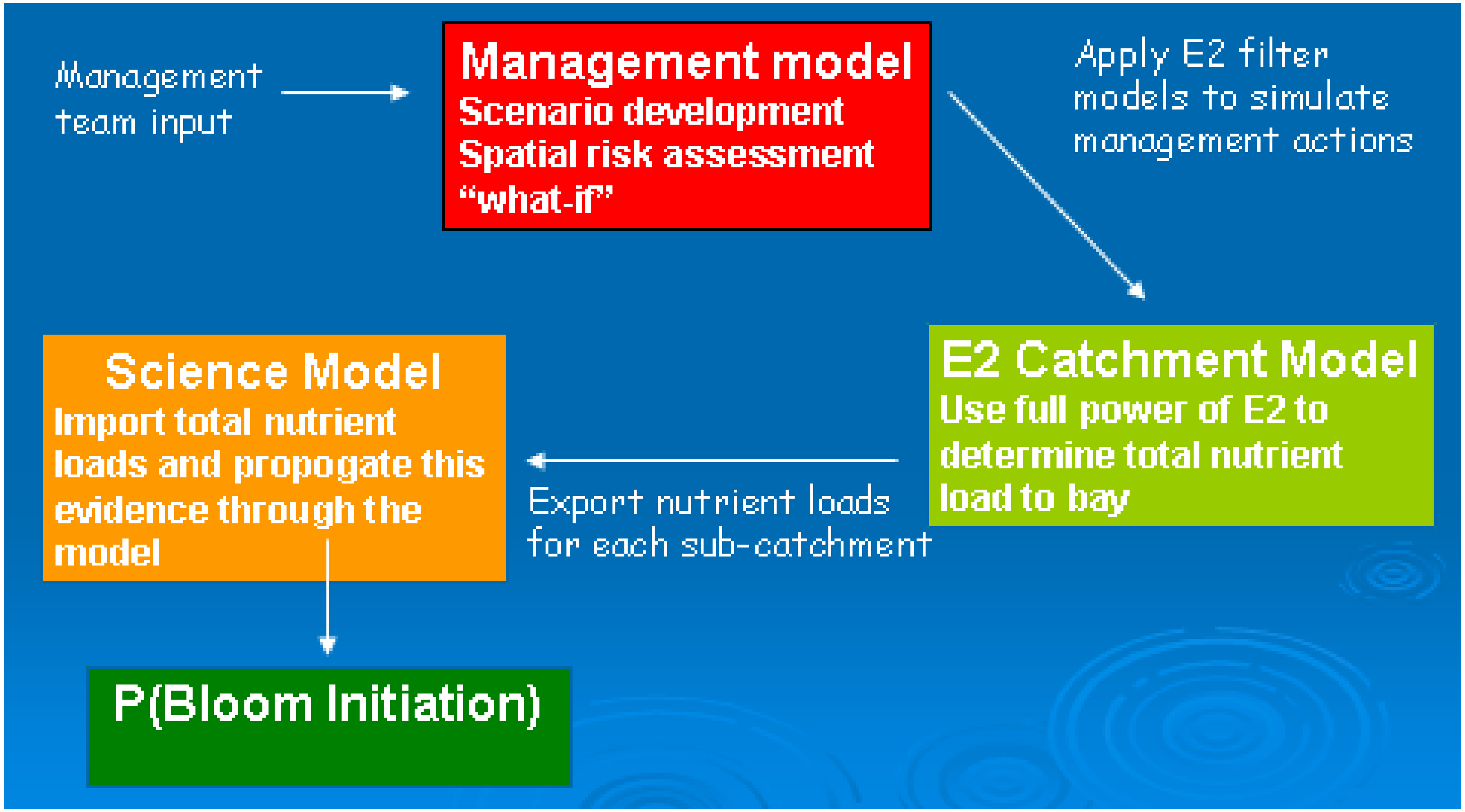
1. Available Nutrient Pool
2. Bottom Current Climate
3. Sediment Nutrients
4. Dissolved Iron
5. Dissolved Phosphorous
6. Light
7. Temperature

MANAGEMENT ACTIONS

“What-if” scenarios

Factor	Change in P(Bloom) (%)
Available Nutrient Pool	77 (3% - 80%)
Bottom Current Climate	28 (15% - 43%)
Sediment Nutrient Climate	17 (21% - 38%)
Dissolved Fe	16 (21% - 37%)
Dissolved P	15 (23% - 38%)
Light Climate	14 (18% - 32%)
Temperature	14 (21% - 35%)
Dissolved N	13 (22% - 35%)
Rain – present	10 (25% - 35%)
Light Quantity	9 (21% - 30%)

From Science to Management



Biosecurity: "Beyond Compliance"

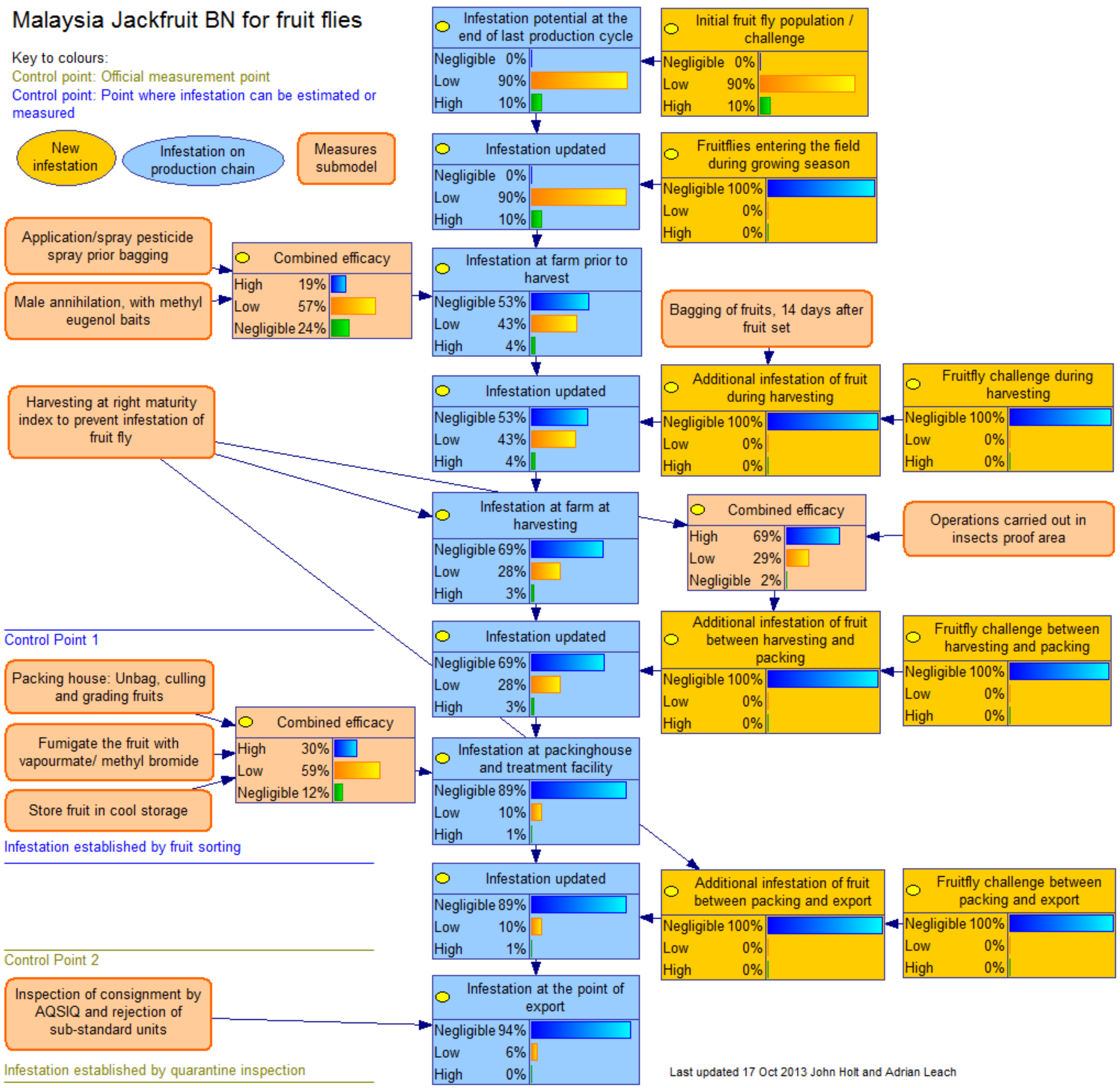
An integrated approach to pest risk management

STDF – WTO funded project

5 SEA partners + OC: + QUT

Mumford *et al.*

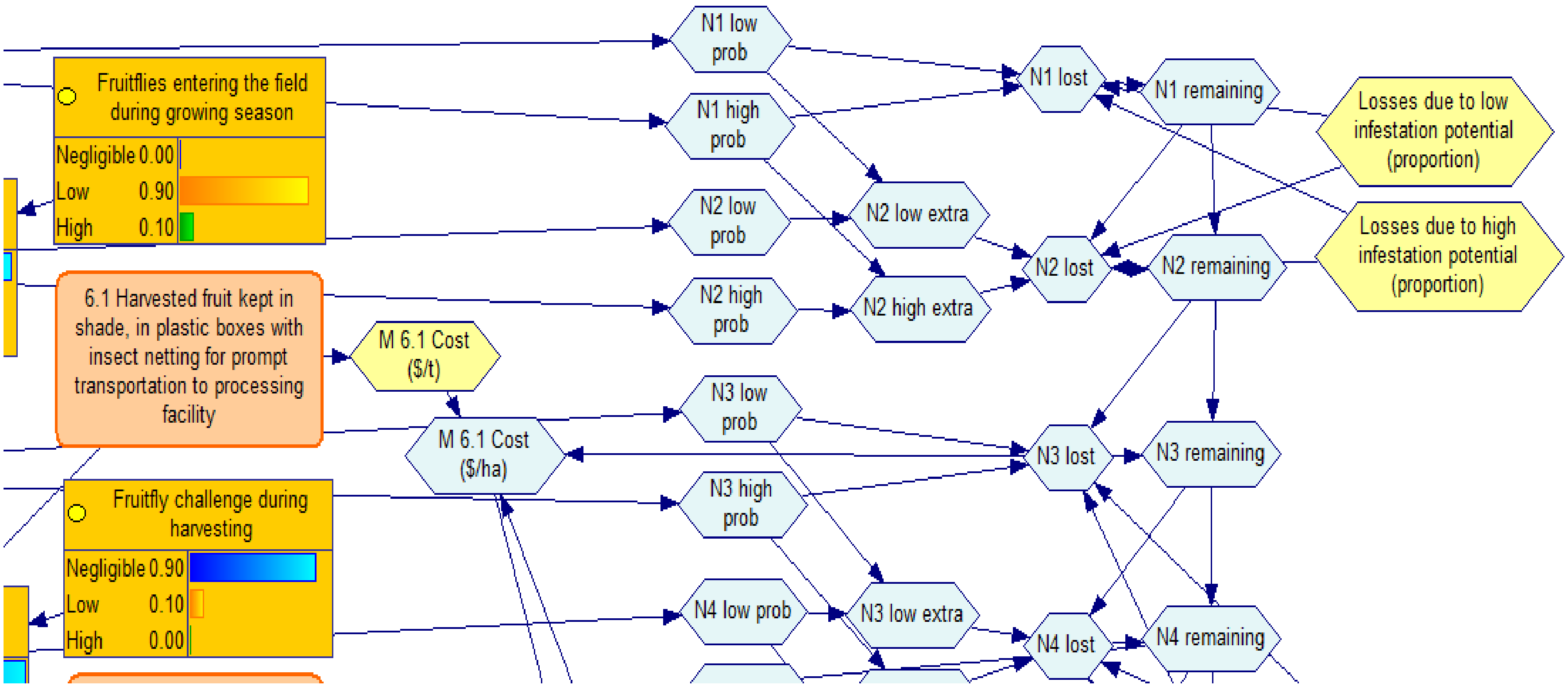
CP-BN



Malaysia Jackfruit BN for fruit flies
Key to colours:
Control point: Official measurement point
Control point: Point where infestation can be estimated or measured
New infestation
Infestation on production chain
Measures submodel
Infestation potential at the end of last production cycle
Negligible 0%
Low 90%
High 10%
Initial fruit fly population / challenge
Negligible 0%
Low 90%
High 10%
Infestation updated
Negligible 0%
Low 90%
High 10%
Fruitflies entering the field during growing season
Negligible 100%
Low 0%
High 0%
Application/spray pesticide spray prior bagging
Male annihilation, with methyl eugenol baits
Combined efficacy
High 19%
Low 57%
Negligible 24%
Infestation at farm prior to harvest
Negligible 53%
Low 43%
High 4%
Bagging of fruits, 14 days after fruit set
Harvesting at right maturity index to prevent infestation of fruit fly
Infestation updated
Negligible 53%
Low 43%
High 4%
Additional infestation of fruit during harvesting
Negligible 100%
Low 0%
High 0%
Fruitfly challenge during harvesting
Negligible 100%
Low 0%
High 0%
Infestation at farm at harvesting
Negligible 69%
Low 28%
High 3%
Combined efficacy
High 69%
Low 29%
Negligible 2%
Operations carried out in insects proof area
Control Point 1
Infestation updated
Negligible 69%
Low 28%
High 3%
Additional infestation of fruit between harvesting and packing
Negligible 100%
Low 0%
High 0%
Fruitfly challenge between harvesting and packing
Negligible 100%
Low 0%
High 0%
Packing house: Unbag, culling and grading fruits
Fumigate the fruit with vapourmate/ methyl bromide
Store fruit in cool storage
Combined efficacy
High 30%
Low 59%
Negligible 12%
Infestation at packinghouse and treatment facility
Negligible 89%
Low 10%
High 1%
Infestation established by fruit sorting
Infestation updated
Negligible 89%
Low 10%
High 1%
Additional infestation of fruit between packing and export
Negligible 100%
Low 0%
High 0%
Fruitfly challenge between packing and export
Negligible 100%
Low 0%
High 0%
Control Point 2
Inspection of consignment by AQSIQ and rejection of sub-standard units
Infestation at the point of export
Negligible 94%
Low 6%
High 0%
Infestation established by quarantine inspection
Last updated 17 Oct 2013 John Holt and Adrian Leach

Economics

adding costs and losses utility nodes



J. Holt, A. W. Leach, S. Johnson, D. M. Tu, D. T. Nhu, N. T. Anh, L. N. Quang, M. M. Quinlan, P. J. L.Whittle, K. Mengersen and J. D. Mumford (in prep.) Bayesian networks to compare pest control interventions on commodities along agricultural production chains.

Ongoing Questions

1. How to elicit information from experts?
2. How to combine information from multiple experts?

3. How to incorporate uncertainty into BNs?
4. How to combine BNs?

Reflection
and
Final Comments