Bayesian modelling of surveillance and proof of freedom

The mathematical, logical & psychological challenges

Samantha Low-Choy
Queensland University of Technology / Plant Biosecurity CRC project funded by GRDC
slowchoy@quteduau

Work with Jo Slattery, Sharyn Taylor on CRCNPB and PBCRC projects
And on SRG/SNPHS with Nichole Hammond, Lindsay Penrose, Mark Stanaway

Looking for a needle in a haystack?

THE CHALLENGE:

SHOULD WE RELY ON SURVEILLANCE?
IF SO: WHEN, WHERE, HOW MUCH?

Zeros can be:

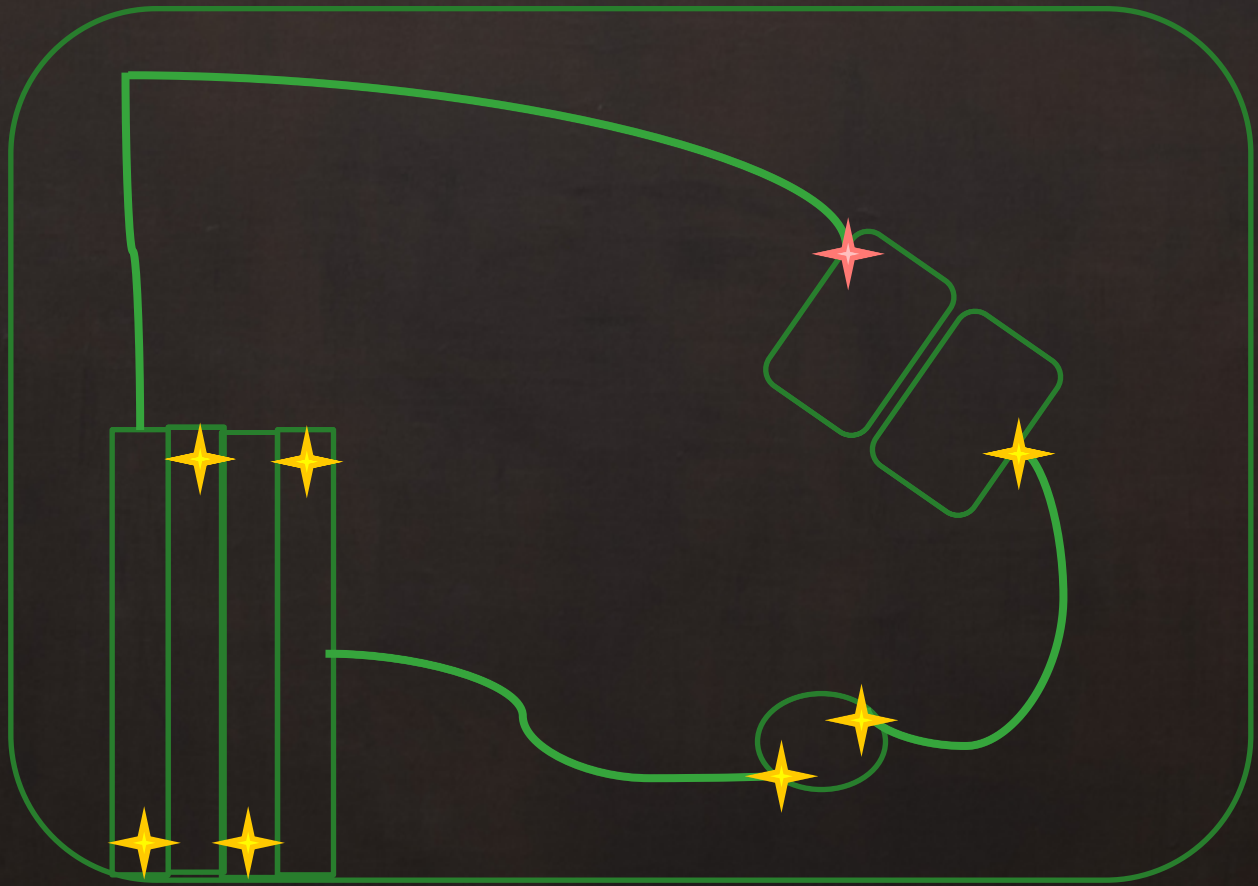
Ambiguous

Excess

Naughty

or

Everywhere



Sampling + Biological
process process

→ Observed Data

Bayesian hierarchical models provide a natural framework

Exchangeability *cf* Independence

Royle & Dozario (2008, Hierarchical modelling & inference in Ecology)

The logical vs perceptual nuances of claims about pest status

AREA FREEDOM
LEGAL LOGICAL CHALLENGE

Maintaining trade agreements

The pest is not known to occur

The pest is known to occur

The pest is known not to occur

Which statement(s) are weak? (In terms of evidence?)

Which statement(s) are strong (in terms of evidence?)

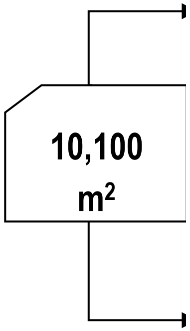
Which statement corresponds best to “Area Freedom”?

Defining what you (really) need …
not necessarily what is easiest to compute

LOGICAL CHALLENGE
GETTING THE QUESTION RIGHT

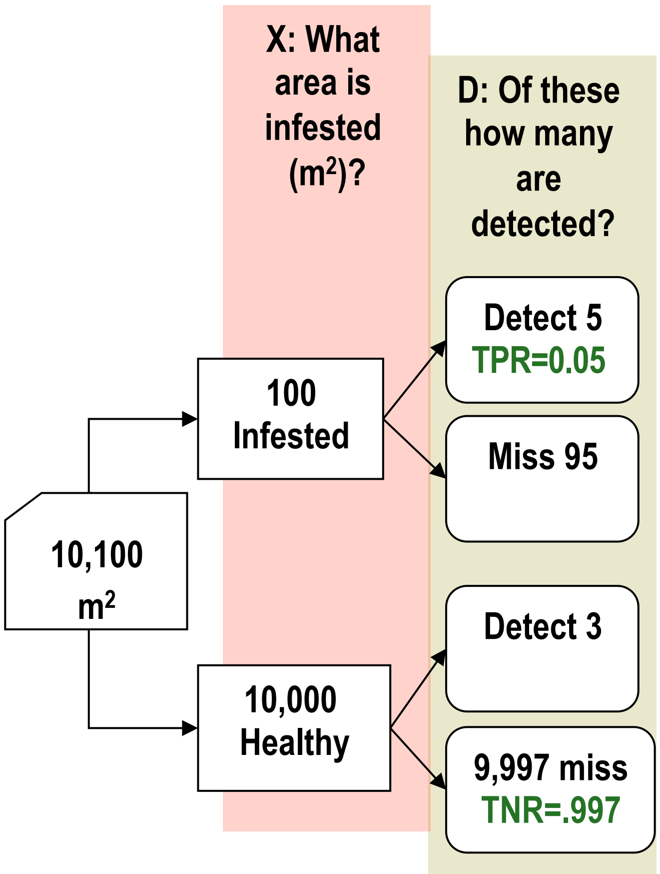
A logical perspective

Assume you know pest status & deduce the evidence you would get OR For a given piece of evidence, infer the plausible pest status



A logical perspective

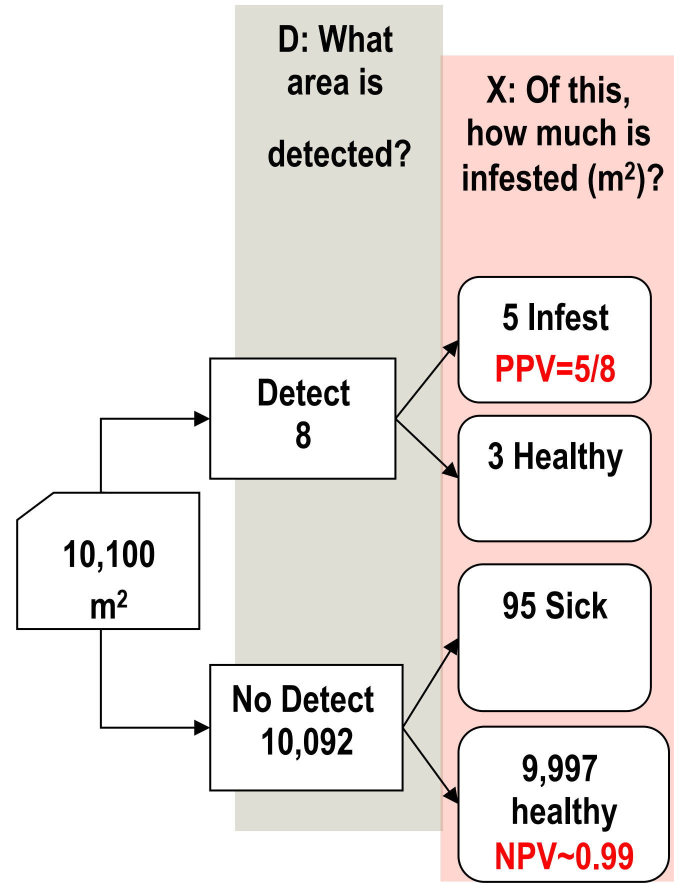
Assume you know pest status & deduce the evidence you would get OR For a given piece of evidence, infer the plausible pest status



Getting it right:

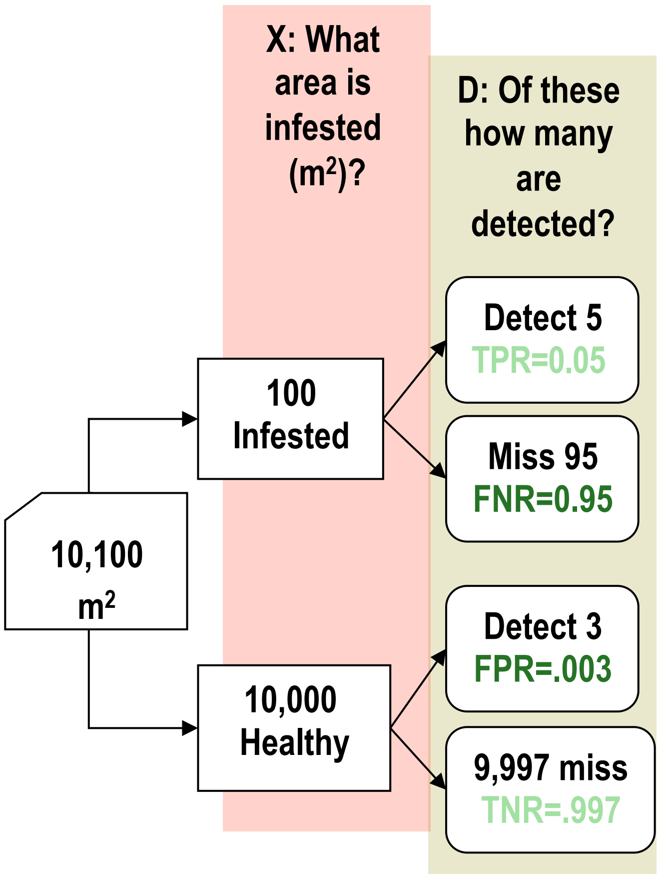
TPR, TNR

PPV, NPV



A logical perspective

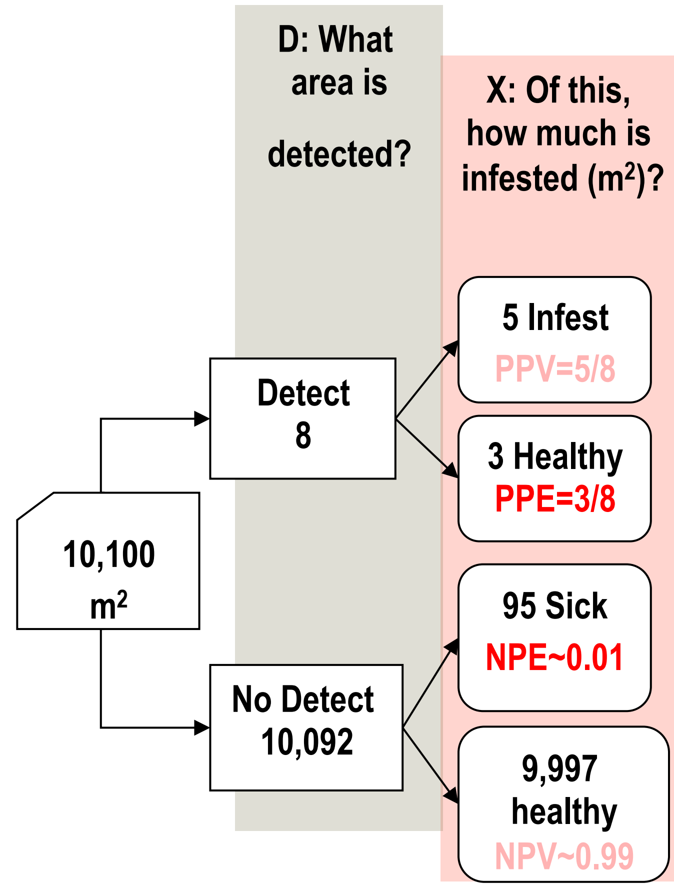
Assume you know pest status & deduce the evidence you would get OR For a given piece of evidence, infer the plausible pest status



Getting it wrong:

FNR, FPR

PPE, NPE



Bayes Theorem:
A Bridge between logical perspectives

Bayes Theorem tells us: $\Pr(X \mid Y) = \dfrac{\Pr(Y \mid X)\Pr(X)}{\sum_k \Pr(Y \mid X_k)\Pr(X_k)}$

Thus: $PPV = \dfrac{TPR\pi}{FPR(1-\pi) + TPR\pi}$ $\quad NPV = \dfrac{TNR(1-\pi)}{FNR\pi + TNR(1-\pi)}$

Equations: just another way of *seeing* the *rules* for the decision tree

Expressing Bayes Theorem for Inference

The logical challenge here: TNR or NPV

What does Area freedom mean?

TNR: When the pest is absent, How often is it not reported?	**99.7%**
NPV: When the pest is not reported, How often does that mean it's absent?	**99%**

What errors can we make about Area freedom?

FNR: When the pest is present, How often is it not reported?	**95%**
NPE: When the pest is not reported, How often does that mean it's really present?	**1%**

Significance is the FNR of hypotheses

The chance of rejecting the null hypothesis when it is true

Invited Paper:

THE INSIGNIFICANCE OF STATISTICAL SIGNIFICANCE TESTING

DOUGLAS H. JOHNSON,[1] U.S. Geological Survey, Biological Resources Division, Northern Prairie Wildlife Research Center, Jamestown, ND 58401, USA

Abstract: Despite their wide use in scientific journals such as *The Journal of Wildlife Management,* statistical hypothesis tests add very little value to the products of research. Indeed, they frequently confuse the interpretation of data. This paper describes how statistical hypothesis tests are often viewed, and then contrasts that interpretation with the correct one. I discuss the arbitrariness of *P*-values, conclusions that the null hypothesis is true, power analysis, and distinctions between statistical and biological significance. Statistical hypothesis testing, in which the null hypothesis about the properties of a population is almost always known a priori to be false, is contrasted with scientific hypothesis testing, which examines a credible null hypothesis about phenomena in nature. More meaningful alternatives are briefly outlined, including estimation and confidence intervals for determining the importance of factors, decision theory for guiding actions in the face of uncertainty, and Bayesian approaches to hypothesis testing and other statistical practices.

JOURNAL OF WILDLIFE MANAGEMENT 63(3):763–772

The trouble with significance

Sifting the evidence—what's wrong with significance tests?

Jonathan A C Sterne, George Davey Smith

BMJ VOLUME 322 27 JANUARY 2001 bmj.com

The findings of medical research are often met with considerable scepticism, even when they have apparently come from studies with sound methodologies that have been subjected to appropriate statistical analysis. This is perhaps particularly the case with respect to epidemiological findings that suggest that some aspect of everyday life is bad for people. Indeed, one recent popular history, the medical journalist James Le Fanu's *The Rise and Fall of Modern Medicine*, went so far as to suggest that the solution to medicine's ills would be the closure of all departments of epidemiology.[1]

One contributory factor is that the medical literature shows a strong tendency to accentuate the positive; positive outcomes are more likely to be reported than null results.[2–4] By this means alone a host of purely chance findings will be published, as by conventional reasoning examining 20 associations will produce one result that is "significant at $P = 0.05$" by chance alone. If only positive findings are published then they may be mistakenly considered to be of importance rather than being the necessary chance results produced by the application of criteria for meaningfulness based on statistical significance. As many studies contain long questionnaires collecting information on hundreds of variables, and measure a wide range of potential outcomes, several false positive findings are virtually guaranteed. The high volume and often contradictory nature[5] of medical research findings, however, is not only because of publication bias. A more fundamental problem is

Summary points

P values, or significance levels, measure the strength of the evidence against the null hypothesis; the smaller the P value, the stronger the evidence against the null hypothesis

An arbitrary division of results, into "significant" or "non-significant" according to the P value, was not the intention of the founders of statistical inference

A P value of 0.05 need not provide strong evidence against the null hypothesis, but it is reasonable to say that $P < 0.001$ does. In the results sections of papers the precise P value should be presented, without reference to arbitrary thresholds

Results of medical research should not be reported as "significant" or "non-significant" but should be interpreted in the context of the type of study and other available evidence. Bias or confounding should always be considered for findings with low P values

To stop the discrediting of medical research by chance findings we need more powerful studies

When the data don’t tell you about some/all of the parameters in the model…

Ask the experts!

MATHEMATICAL CHALLENGE
NOT ENOUGH DATA

Bayesian Learning

A focus on data … the prior is a silent partner

$$\pi(\theta \mid x) \propto f(x \mid \theta)\boxed{\pi_0(\theta)}$$

Parameter estimates and plausible range of values

Data (uncertainty due to sampling)

Expressing Bayes Theorem for Bayesian statistical modelling

CRC PLANT biosecurity

Bayes & Price (1763, Philosophical Transactions of the Royal Society of London)

Bayesian Learning

A focus on data … the prior is a silent partner

$$\pi(\theta \mid x) \propto f(x \mid \theta) \boxed{\pi_0(\theta)}$$

WARNING

"Ignoring" the prior *presumes* it is (locally) uniform (on the scale of the parameter in the likelihood).

Omitting this presumption, leads to the widespread "Inversion Fallacy" where Pr(A|B) is mistaken for Pr(B|A)

Uniform priors in Box & Tiao (1982, B Inf Stat Analysis);
Inversion fallacy in Low-Choy & Wilson (2009, IASE)

Bayesian Learning

A focus on updating … the prior is an active partner
Implies investment in >1 study!

$$\pi(\theta \mid x) \propto f(x \mid \theta)\pi_0(\theta)$$

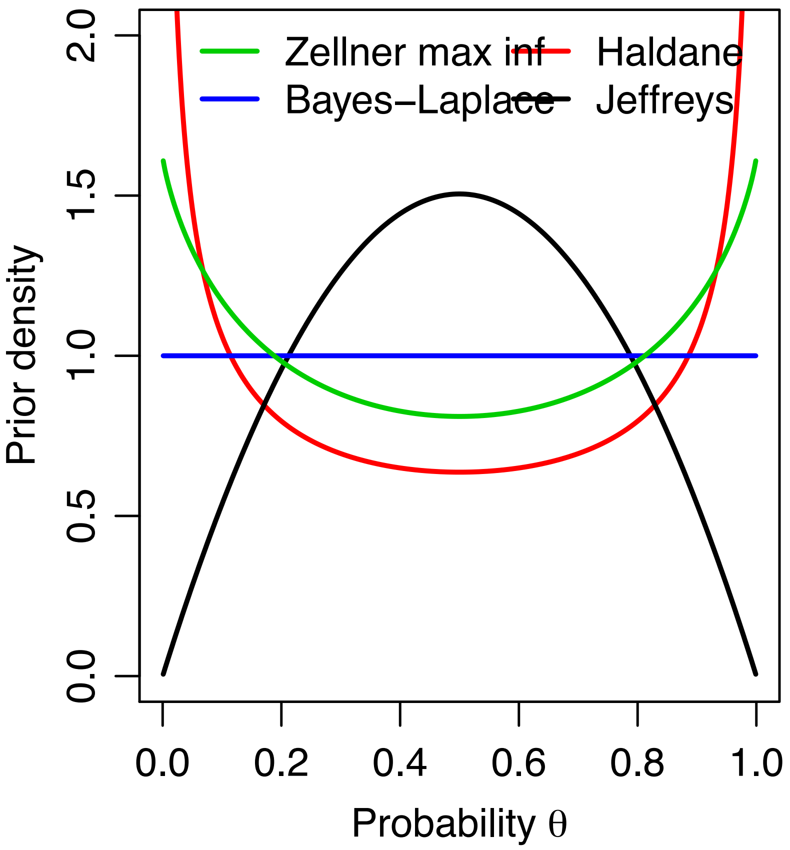
Parameter estimates adjusted to related data

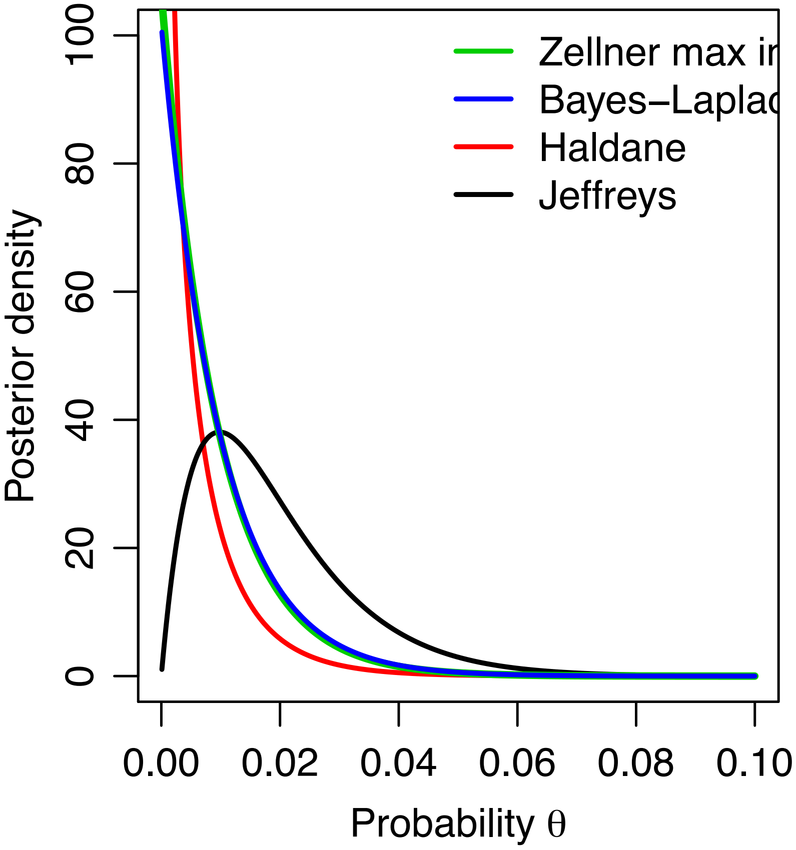
Related information

What the experts think re: parameters

Low Choy et al (2012, CS-BSMA)

The prior has impact with small data

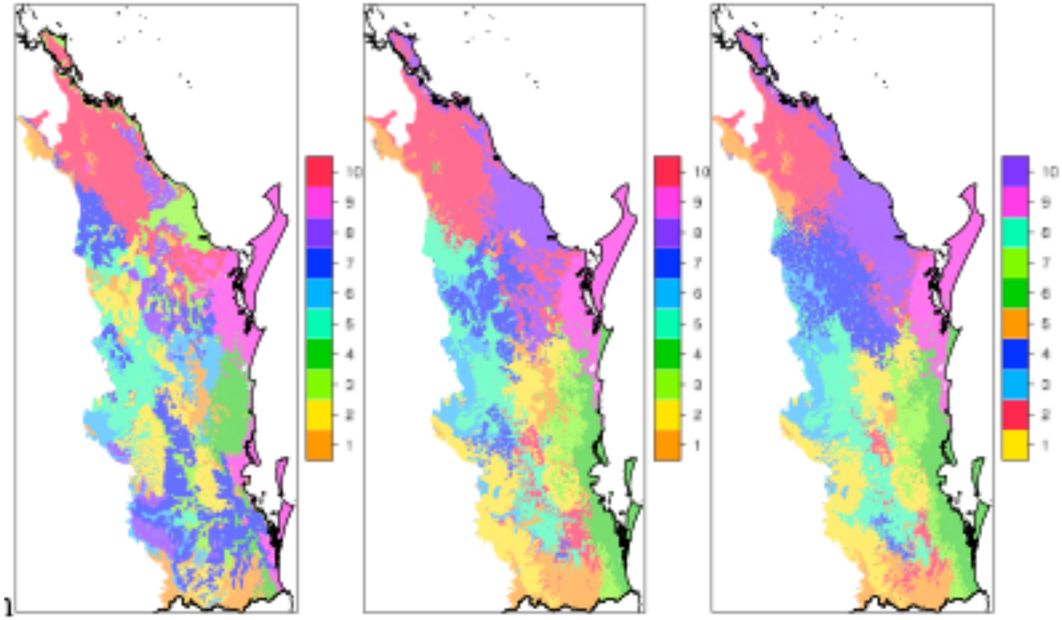




CRC PLANTbiosecurity

For more details: Low-Choy et al (2012, CS-BSMA) and Tuyl et al (2008, Amer. Statistician)

The prior even has impact with big data



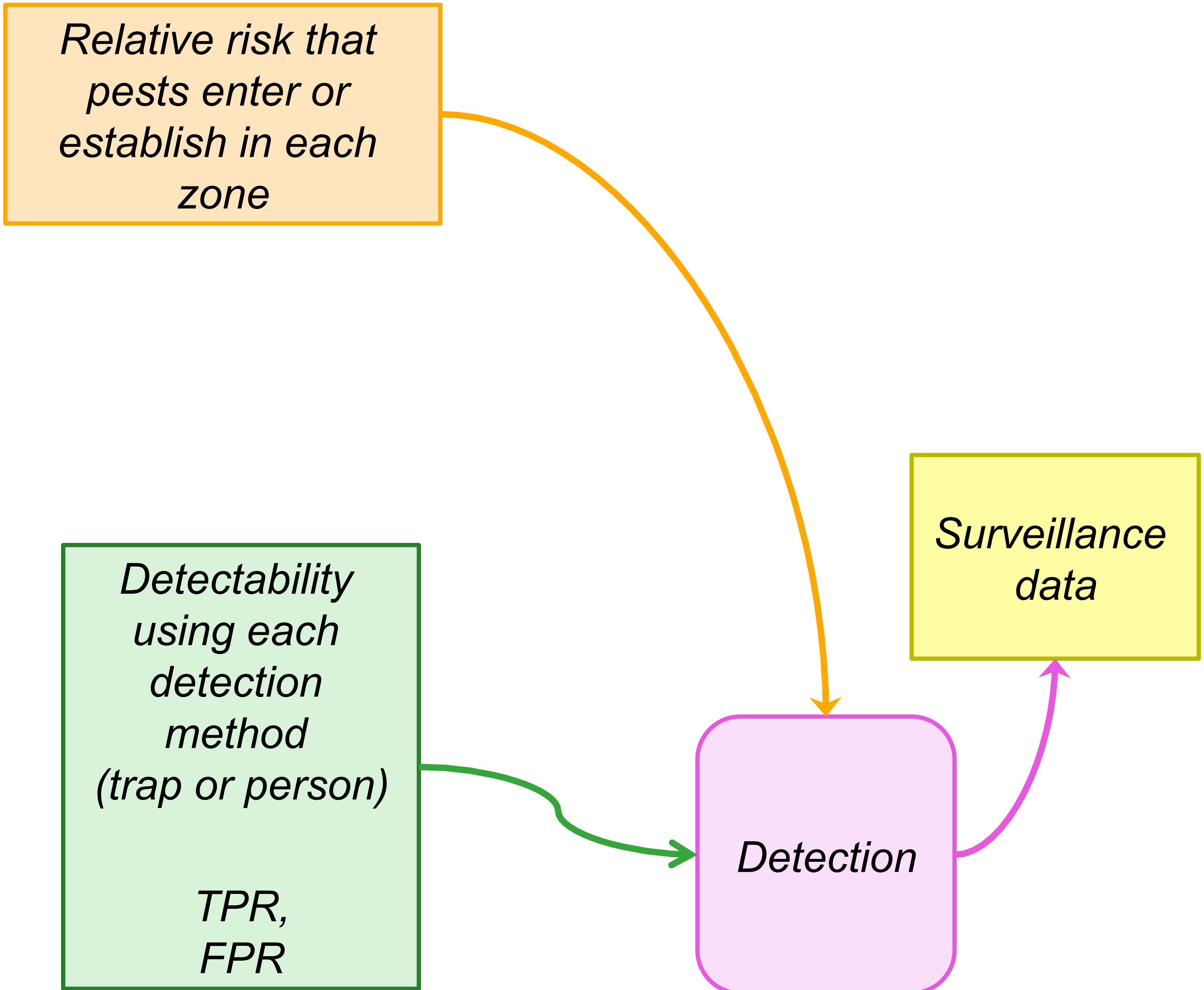
MVN mixture model with 10 components (regions) and 8 GIS attributes (variables), with varying weight on prior knowledge: "vaguely" informative (left), informative (middle), no data (right)

Low-Choy et al (2012, CS-BSMA) and Tuyl et al (2008, The Amer. Statistician)

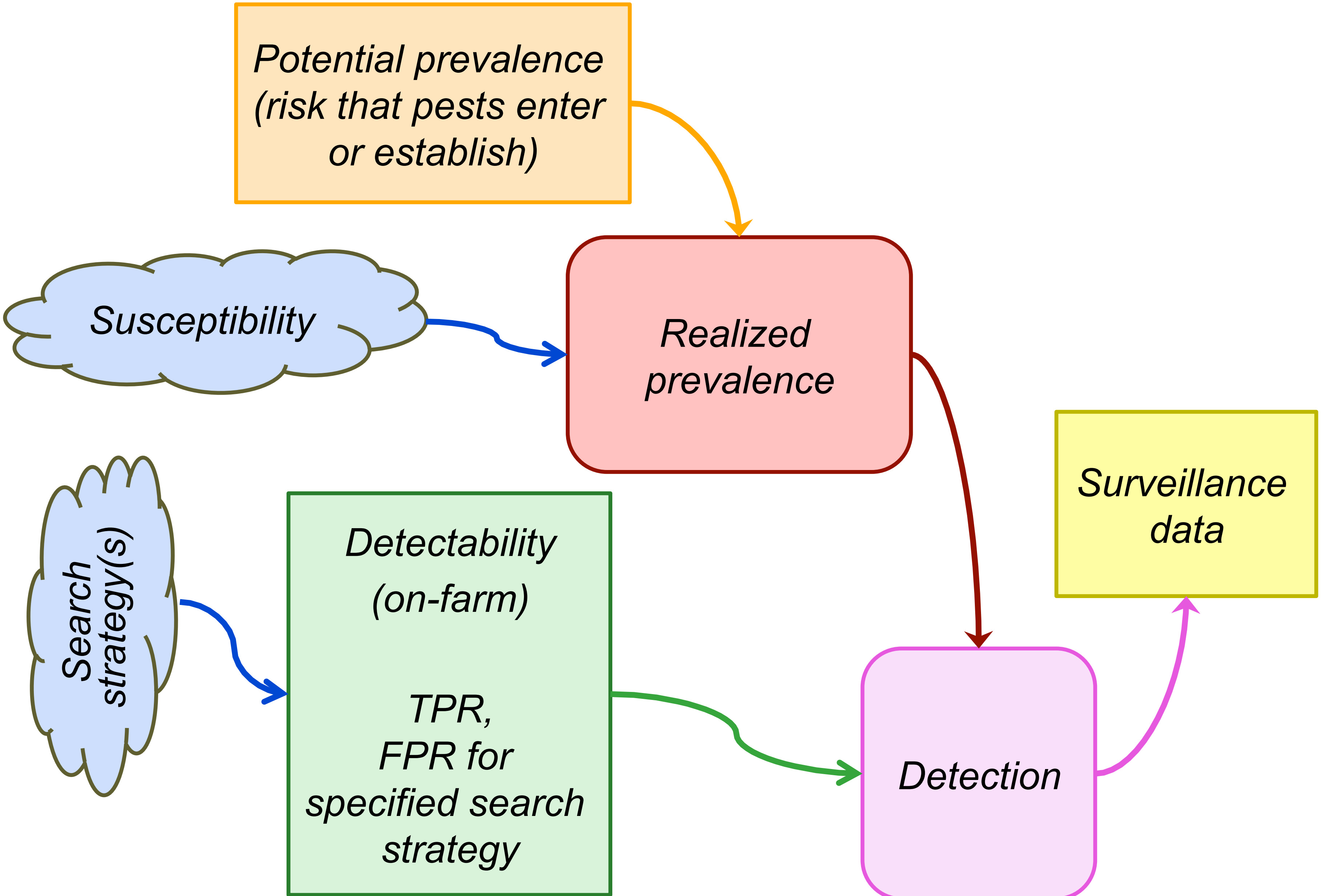
Experts can integrate what is relevant from the literature and their own field experience, in similar situations.

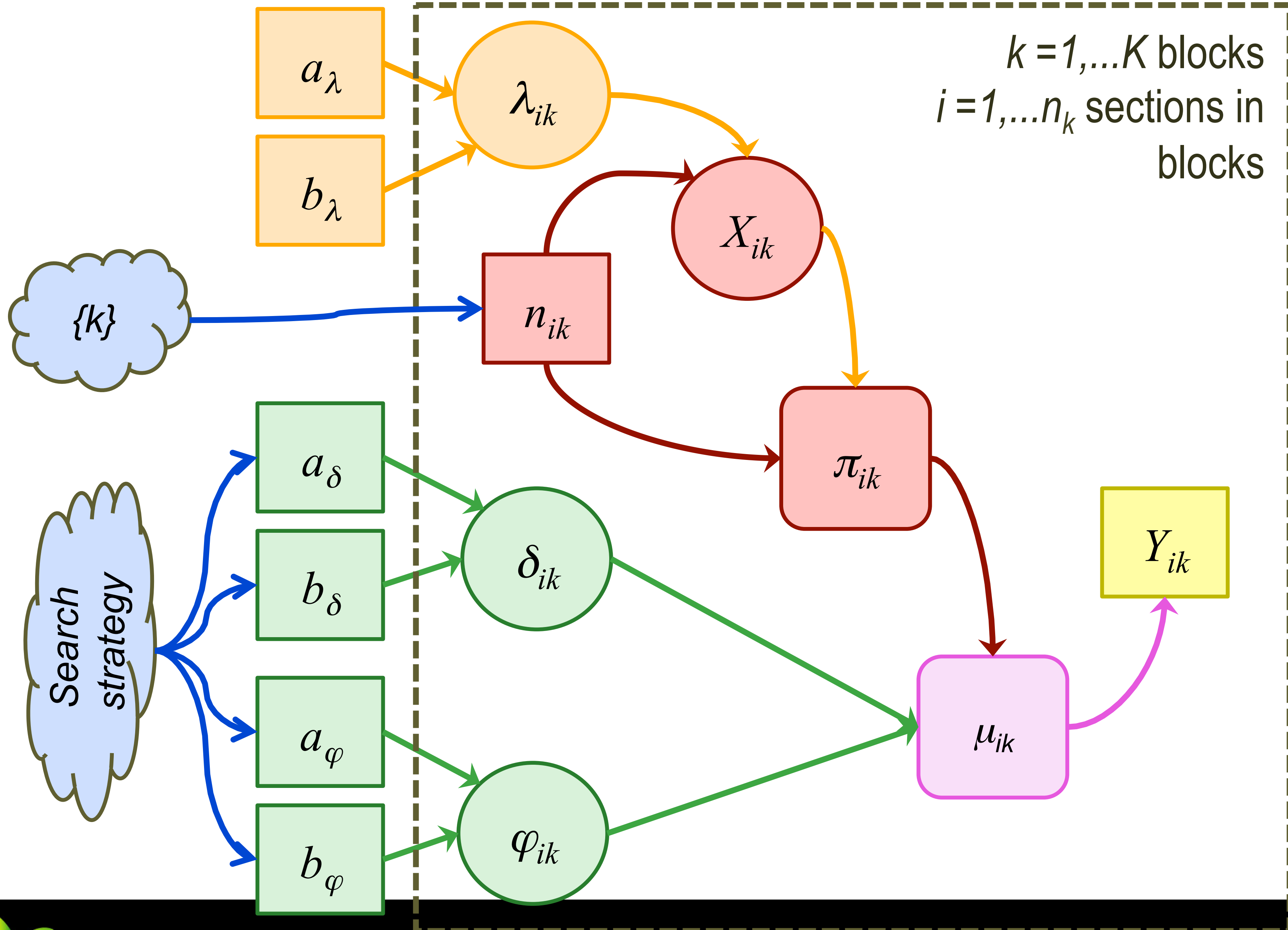
Make explicit what the current state of knowledge is…

MATHEMATICAL CHALLENGE
STRUCTURING THE MODEL



Barrett+2010


Potential prevalence (risk that pests enter or establish)
Susceptibility
Realized prevalence
Search strategy(s)
Detectability (on-farm)
TPR, FPR for specified search strategy
Detection
Surveillance data



CRC PLANT biosecurity

```
for (k in 1:nblocks) {
        for (j in 1:nsections[k]) {
                # number of infested plants in jth section, kth block
                x[j,k] ~ dbin(lambda[j,k], Nplants.per.section[k])
                lambda[j,k] ~ dbeta(a.lambda, b.lambda)

                # prob that any plant in section is the infested one
                pinfest[j,k] <- x[j,k] / Nplants.per.section[k]

                # probability of detecting each infested plant
                pdetect[j,k] <- pinfest[j,k]*delta[k] + (1-pinfest[j,k])*phi[k]

                # number of detections depends on the number inspected
                y[j,k] ~ dbin( pdetect[j,k], ninspect[k])
        }
        # true infestation in each block
        xsum[k] <- sum(x[1:nsections[k],k])

        # detection depends on TPR and FPR
        delta[k] ~ dbeta(a.delta, b.delta)
        phi[k] ~ dbeta(a.phi, b.phi)
}
# missed infestations across blocks
xtot <- sum(xsum[1:nblocks])
```

How do you capture expert knowledge on δ, φ into a statistical distribution?

PSYCHOLOGICAL CHALLENGE

Defining what is being elicited

Factors considered to affect the Chance of Reporting

Detection

Evidence

Skill and resourcing

Defining the Chance of Reporting (CoR)

Pr(Report | Detection, Compelling evidence, Skill of observer)

Probability that suspicious evidence, consistent with RWA damage, is reported; given whether it is detected, how compelling it is (to the observer & their networks), and the skill of the observer.

If suspicious evidence is detected in the field, ***whether it is reported*** to the next level depends on***:***

Detection: *whether the evidence was detected – yes or no*

Compelling Evidence: depends on

- the level of evidence detected (mild symptoms or devastation)
- the level of awareness and networking to evaluate the evidence – little or substantial

Skill: of the observer – inexperienced (low) or trained (moderate).

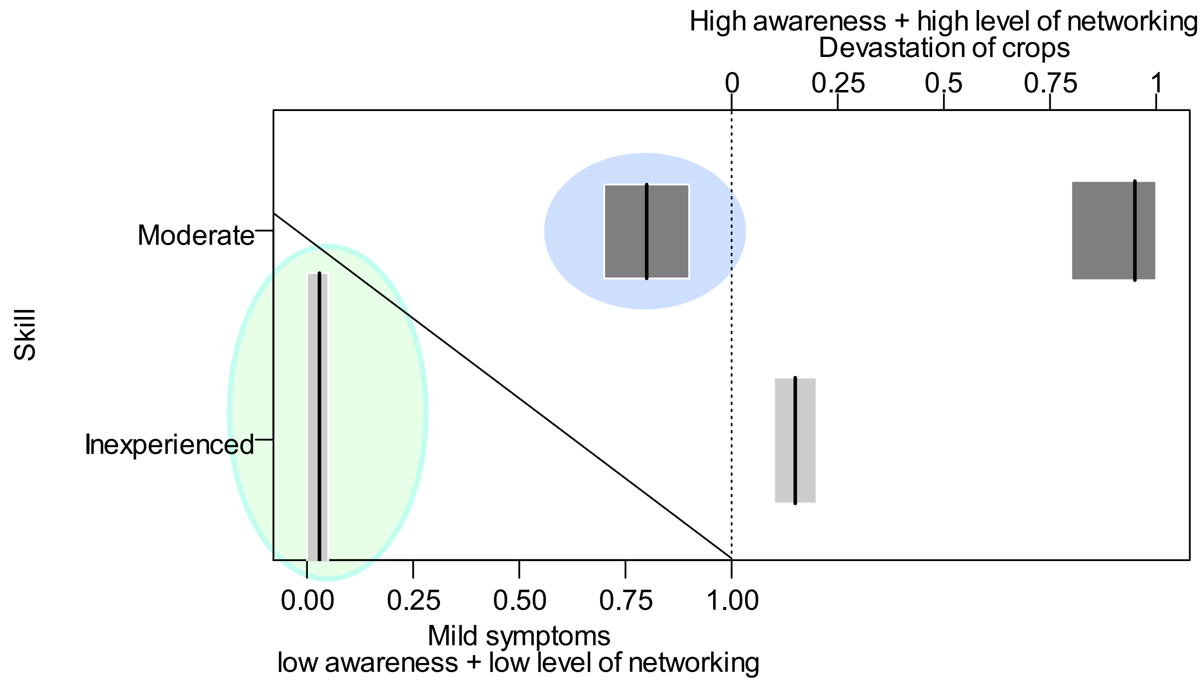
- NB It was considered unlikely to have highly skilled observers undertaking general surveillance.

Reporting

Level of infestation	Skill of inspector	Likelihood of reporting
Mild symptoms, low awareness, and low level of networking	Inexperienced inspector	0-5% (80% plausible), with best estimate 3%
	Moderately experienced inspector	70-90% (90% plausible), with best estimate 80%
Intermediate symptoms		Depends on threshold for spraying a few paddocks affected, and whether visitors with relevant knowledge.
Devastation of crops, high level of awareness and high level of networking	Moderately experienced inspector	80-100% (95% plausible), with best estimate 95%
	Inexperienced inspector	10-20% (60% plausible), with best estimate 15%

Blue scenario

- **spreads plausibility (shorter) over wider range of values (fatter)**
- **very distinct from green scenario**



Reporting

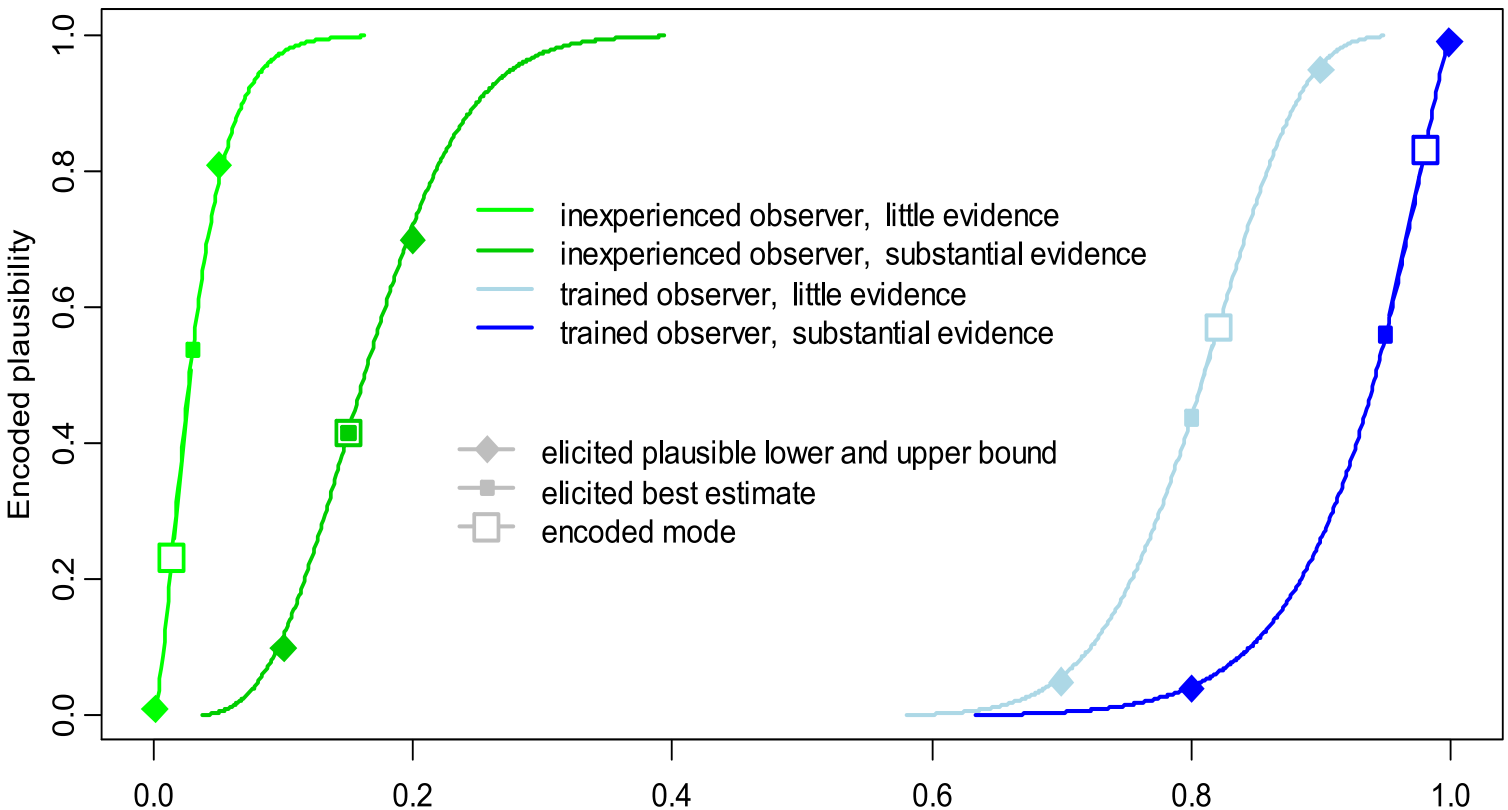
Level of infestation	Skill of inspector	Likelihood of reporting
Mild symptoms, low awareness, and low level of networking	Inexperienced inspector	0-5% (80% plausible), with best estimate 3%
	Moderately experienced inspector	70-90% (90% plausible), with best estimate 80%
Intermediate symptoms		Depends on threshold for spraying a few paddocks affected, and whether visitors with relevant knowledge.
Devastation of crops, high level of awareness and high level of networking	Moderately experienced inspector	80-100% (95% plausible), with best estimate 95%
	Inexperienced inspector	10-20% (60% plausible), with best estimate 15%

Reporting factors		Elicited information			Translation into statistical information		Encoded Beta(a,b)	
Evidence	Skill	Best estimate	Range	Plausibility of range	Target quantiles	Target cprob*	a	b
Mild symptoms, little aware and networked	Low	3%	0-5%	80%	0.1%, 5%	.01-.81	1.66	47.30
	Moderate	80%	70-90%	90%	70%, 90%	.05-.95	32.20	7.84
Devastation, highly aware and networked	Low	15%	10-20%	60%	10%, 20%	.10-.70	6.40	31.60
	Moderate	95%	80-100%	95%	80%, 99.9%	.04-.99	16.70	1.32

CRC PLANT biosecurity

Reporting

Level of infestation	Skill of inspector	Likelihood of reporting
Mild symptoms, low awareness, and low level of networking	Inexperienced inspector	0-5% (80% sure), with best estimate 3%
	Moderately experienced inspector	70-90% (90% sure), with best estimate 80%
Devastation of crops, high level of awareness and high level of networking	Moderately experienced inspector	80-100% (95% sure), with best estimate 95%
	Inexperienced inspector	10-20% (60% sure), with best estimate 15%



The mathematics

Use a Bayesian hierarchical model for surveillance given the pest process

The logic

A Bayesian posterior probability gives NPV for Area Freedom

The psychology

Encoding expert knowledge & uncertainty to inform subjective priors in the Bayesian framework

Combining expert knowledge

Albert I, Donnet S, Guihenneuc-Joyaux C, Low-Choy S, Mengersen K, Rousseau J (2012). Combining expert opinions in prior elicitation, with discussion, *Bayesian Analysis*, *7(3):502–532*, http://wwwquteduau/e-prints

Encoding expert knowledge, methods & software

Fisher R, O'Leary R, Low-Choy S, Mengersen K, Caley J (2012). A software tool for elicitation of expert knowledge about species richness or similar counts, *Environmental Modelling & Software*, 3:1-14

Johnson S, Low-Choy S, Mengersen K (2012) "Integrating Bayesian networks and Geographic information systems", *Integ Environ Assess Mgmt*, 8(3): 473-9.

Low Choy S, Murray J, James A, Mengersen K (2010) Indirect elicitation from ecological experts: from methods and software to habitat modelling and rock-wallabies *in* O'Hagan A, West M (eds) *Oxford Handbook Appl. Bayesian Analysis*, OUP:UK, pp 511-544.

Low-Choy S, James A, Murray J, Mengersen K (2012) Elicitator: a user-friendly, interactive tool to support the elicitation of expert knowledge. In Perera AH, Drew CA, Johnson CJ (eds) *Expert Knowledge & Its Applications in Landscape Ecology*. Springer, NY.

Low-Choy S (2013b). Priors: Silent or active partners in Bayesian inference? In C. Alston, Mengersen, K, and Pettitt, A. N, editors, *Case Studies in Bayesian Statistical Modelling & Analysis*, pp30–65. John Wiley & Sons, Inc: London.

Martin TG, Burgman MA, Fidler F, Kuhnert PM, Low-Choy S, McBride M, Mengersen K. (2012) Eliciting Expert Knowledge in Conservation Science, *Conservation Biology*, 26(1): 29-38.

O'Leary R, Fisher R, Low-Choy S, Mengersen K, Caley MJ (2011) What is an expert? In Chan, F. et al (eds) *Proceedings MODSIM2011*, wwwmssanzorgau/modsim2011/e9/oleary.pdf

Search effort and detectability

Falk M, O'Leary RA, Nayak MK, Collins PJ, Low-Choy S (submitted) A Bayesian Hurdle Model for Analysis of an Insect Resistance Monitoring Database.

Low-Choy S, Daglish G, Ridley A, Burrill P. (submitted) "Bayesian adjustment of sampling biases for small intensive surveys on farm management practices relevant to biosecurity"

Low-Choy S, Hammond N, Penrose L, Anderson C, Taylor S (2011). In Chan et al (eds) *Proceedings MODSIM 2011*, www.mssanz.org.au/modsim2011/E16/low_choy.pdf

Low-Choy S, Slattery J, Falk M, Taylor S. (2012b). Eliciting expert knowledge on general surveillance: parameterizing design and evaluation of general surveillance for early detection of exemplar pests. Part 1: Methodology. Technical report, CRNNPB

Low-Choy S (submitted). Looking for plant pests: when is 600 samples enough? *Quantitative methods for Designing Surveillance in Plant Biosecurity*

What are the benefits of a Bayesian approach?

SOME RESULTS

Surveillance is like Battleships

You need more effort (for field-detection) of ships in a bigger area

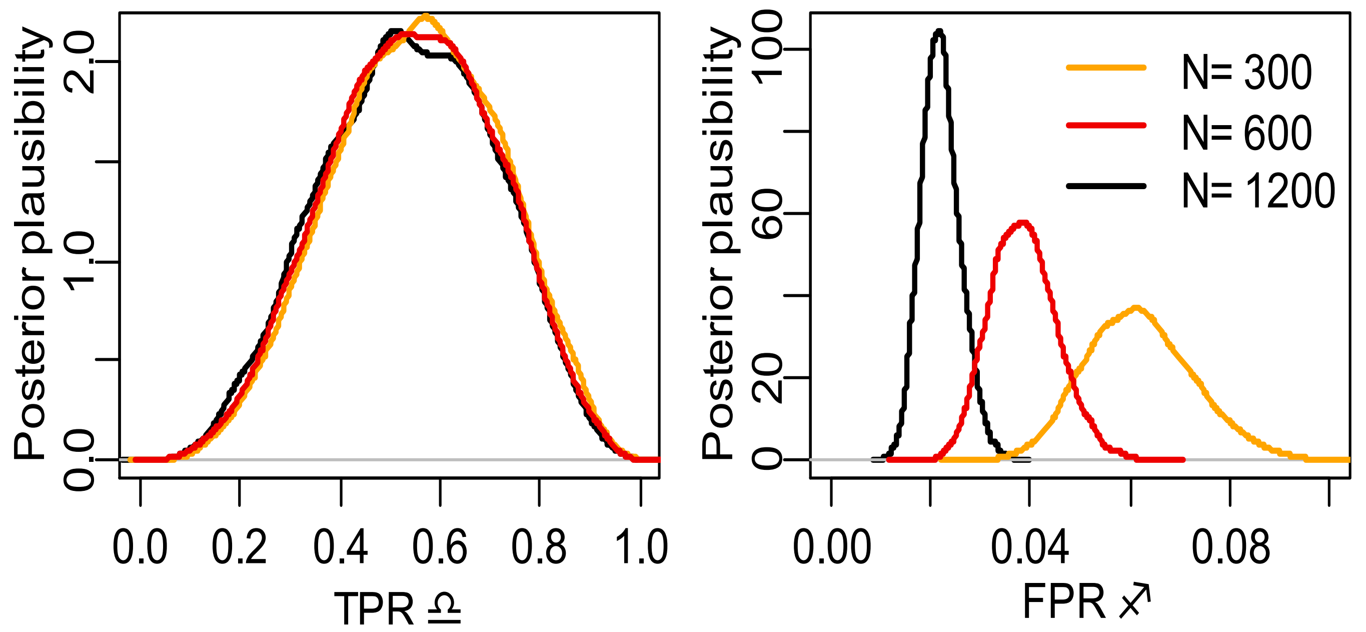
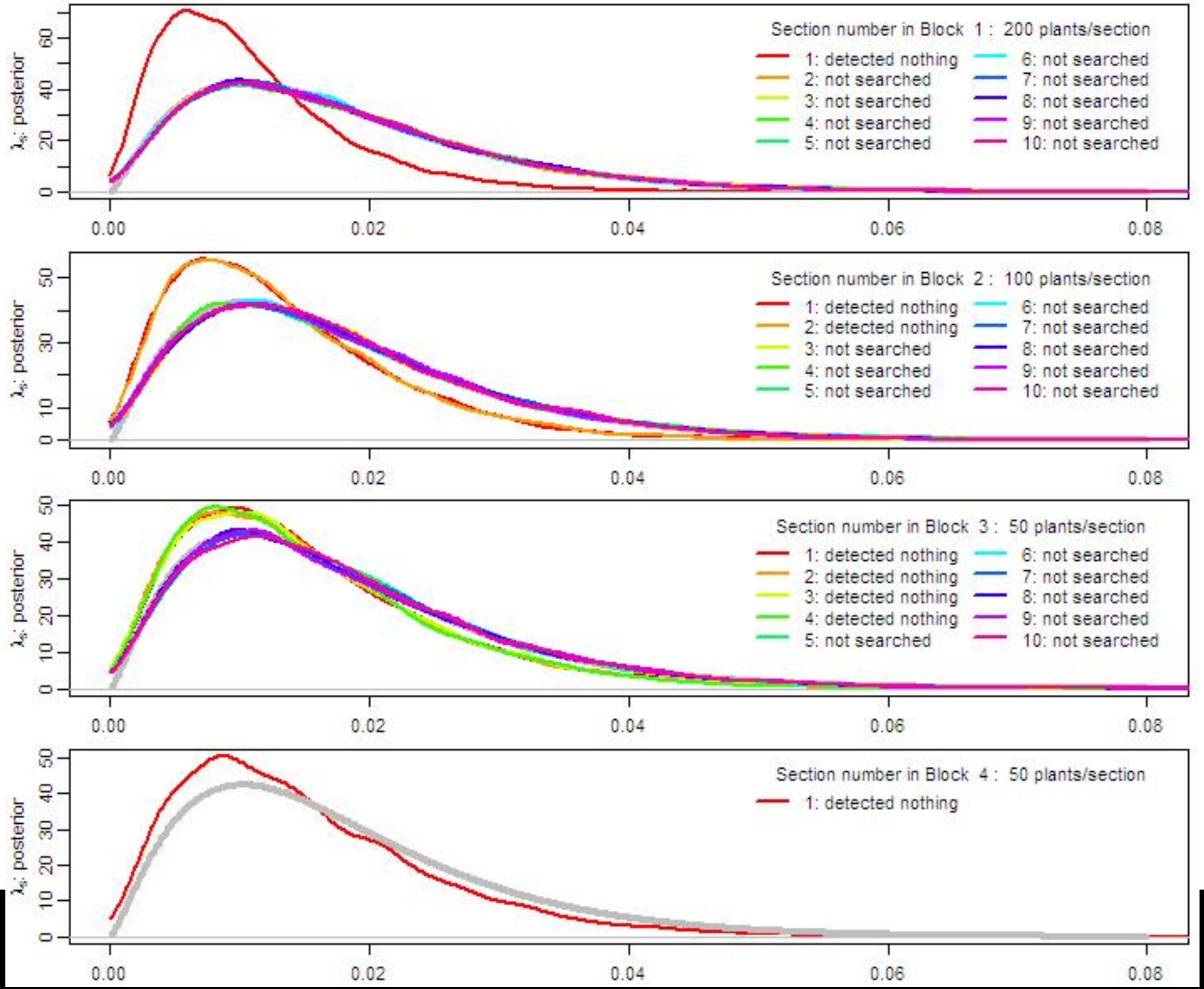


Figure : First sampling occasion. Effect of changing the number of blocks searched, with no detections, on detectability parameters.

CRC PLANTbiosecurity

Surveillance is like Battleships

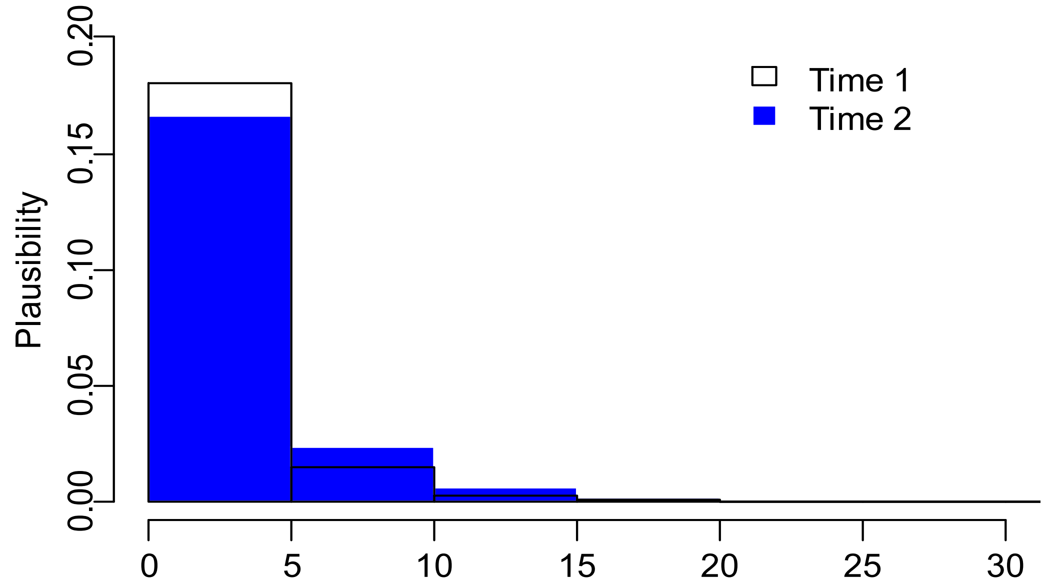
We learn by looking, and we don't learn by not looking



CRC PLANTbiosecurity

Surveillance is like Battleships

but ships grow, and our knowledge grows



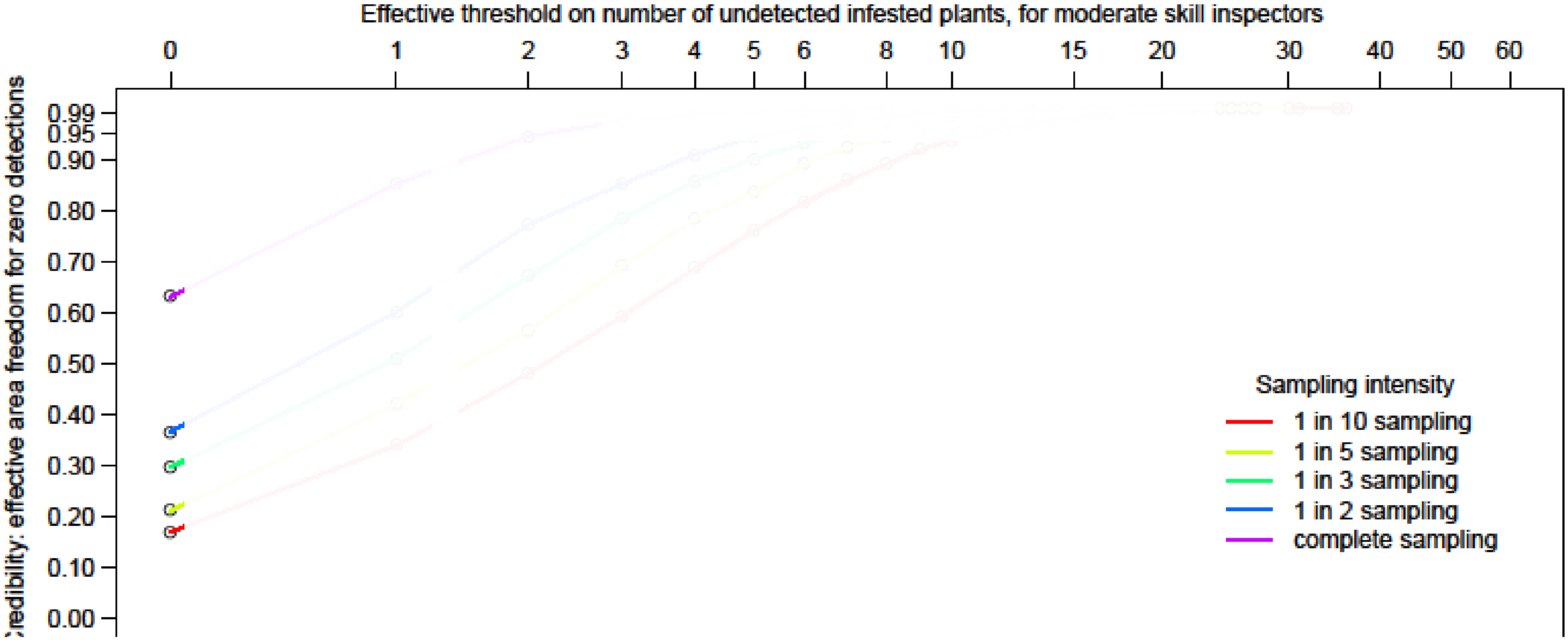
After 4 weeks, typical scenario (40 blocks searched

- the mean infested #plants doubles (5.97→12.08)
- 95% sure infested #plants >doubles (17→46)

Can harness Bayesian cycle of learning to adapt as information gained & knowledge refined.

Surveillance is like Battleships

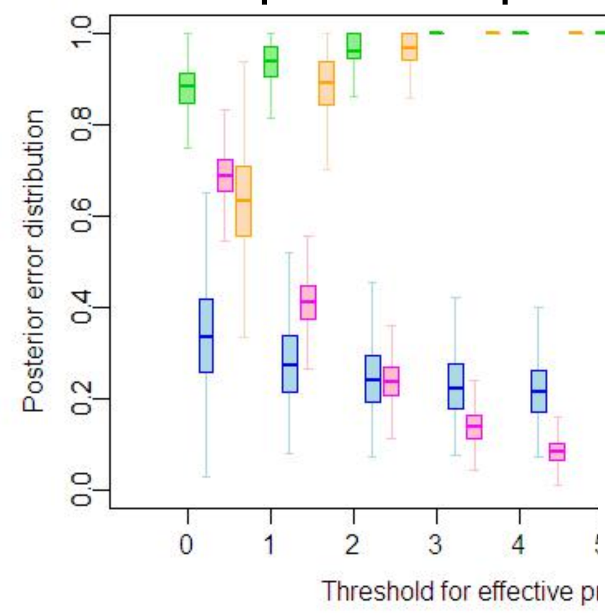
Looking harder is more effective



CRC PLANT biosecurity

Sampling performance

6f. Sample 600 of 9K plants



7a. Sample 600 of 3K plants