Structured Expert Judgement for innovative decision support for pollination and food security

Martine J. Barons ©
with
James Q Smith & Manuele Leonelli

University of Warwick

Martine.Barons@warwick.ac.uk

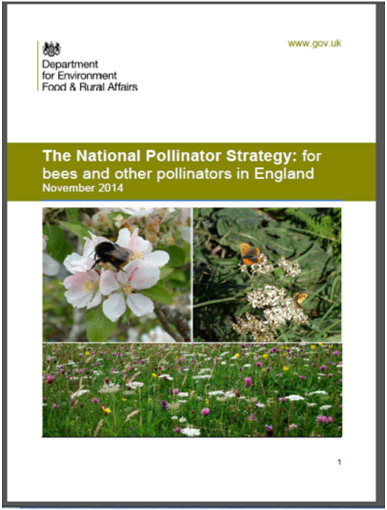
27th April 2017

Overview

What do these all have in common?

The pollinator problem

- There is concern over the decline of insect pollinators (Potts2010)
- National Pollinator strategy Nov 2014
- Identified gaps in knowledge
- Insect Pollinator initiative



The UK pollinator strategy

UK pollinator strategy published November 2014

Development of the strategy:

- Developed using research, NGOs retailers, professional bodies
- Public consultation
- Amendment

Contains an extensive list of what is known (little) and unknown (a lot).

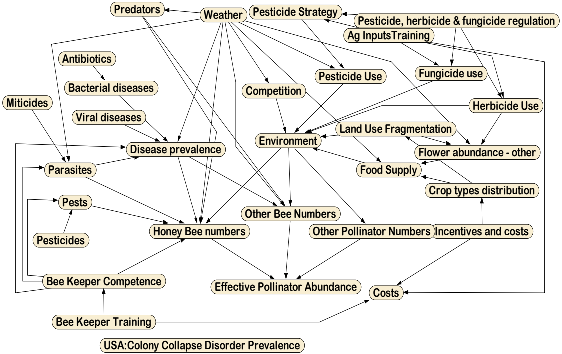
Not a very transparent policy decision-making process

The UK pollinator strategy

Recommendations

- Focus on environment
- Encourage field margins, hedges & similar farming practices
- Encourage growing wild flowers in urban and brownfield spaces
- Campaign for bee-friendly metric in BREEAM, environmental assessment method and rating system for buildings
- Research - starting with development of a monitoring framework and assessment of Neonicotinoid ban
- Improve taxonomic expertise to identify insect pollinators

Why do we need decision support?



Requires experts in many disparate domains and coherent combination of all variables.

Bees (and other pollinators) in human food security

Loss of pollination services provided by insects is one of the threats to human food security:

- 71% important crop species are bee-pollinated.
- Marked increase in bee populations disease levels (Varroa & other parasites, bacterial pathogens, parasitic insects and viruses such as Deformed wing Virus)
- Reduced pollination leading to lower crop yields.

Datta et al (2013) Modelling the spread of American foulbrood in honeybees J. R. Soc. Interface 2013 10,

Key question

Key question

How can we network together inputs form these disparate expert domains in a coherent way, taking account of inherent uncertainties, so that different policy options can be compared in order to support decision-making?

New methodology

We have contributed a methodology for doing this in a general system of this type and identified some frameworks which can be used to build such an integrating decision support system.

Integrating Decision Support systems

A formal & defensible statistical methodology to draw together inferences when:

- Users are decision *Centres*
- Expert judgements from disparate panels of experts
- Each component panel informed by complex models & huge data sets
- A single, comprehensive probabilistic model is inappropriate
  - infeasibly large
  - no shared structural assumptions so no centre can 'own' the full joint distribution
  - dynamic revisions lead to fast obsolescence

System requirements

- Dynamic for evolving environments.
- Distributed among disparate domain experts.
- Coherent: beliefs in different panels not contradictory.
- Networked probabilistic models.
- Can accommodate experimental and observational data.
- Balance strength of evidence.
- Compare risks of different policy decisions.
- Account for measures of uncertainty.
- Update in real time.
- Defensible decisions justifiable to external auditor or regulator.

Overview: Integrating Decision Support System (IDSS)

Many underlying technical requirements for IDSS

- **Common Knowledge assumptions** Policy, Utility & Structural consensus, Parametric union, Quantitative delegation
- **IDSS is** Adequate, Sound, Delegable, separately informed, Cutting, Commonly separated, Distributive, has separable likelihood, panel independence

Coherent methods for handling uncertainty are essential to robust decision analysis.
See *Coherent Frameworks for Statistical Inference serving Integrating Decision Support Systems Jim Q. Smith, Martine J. Barons and Manuele Leonelli arXiv 1507.07394* for full details.

Technical structure

- Processes are evaluated and overseen by m different panels of domain experts, $\{G_1, \ldots, G_m\}$
- Large vector of random variables measure various features of an unfolding future $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m)$ where $\boldsymbol{Y}_i$ takes values in $\mathbb{Y}_i(d)$, $i = 1, 2, \ldots, m$, $d \in \mathbb{D}$ the decision space, e.g. expected impact on the consumer given their socio-economic status.

Technical structure

- Panel G_i will be responsible for the output vector $\{\boldsymbol{Y}_i : \; i = i \ldots m, \}$
- The implicit (albeit virtual) owner of these beliefs, who needs to aggregate the individual panels' judgements, will henceforth referred to as the *supraBayesian*, **SB**

Technical structure continued:

- G_i, $i = 1, 2, \ldots, m$, will be required to deliver to the integrating decision support system (IDSS) belief summaries denoted by $\Pi_i^y \triangleq \{\Pi_i^y(d) : d \in \mathbb{D}, \}$. These summaries will typically be various expectations of certain functions of $\boldsymbol{Y}_i$ taken by some subvector of $\boldsymbol{Y}$ for each decision $d \in \mathbb{D}$
- all panellists make their inferences in a parametric or semi-parametric setting where $\boldsymbol{Y}$ is parametrised by $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_m) \in \Theta(d) : d \in \mathbb{D}$ and the parameter vector $\boldsymbol{\theta}_i$ parametrises the G_i's relevant sample distributions $i = 1, 2, \ldots, m$.
- the panels are *variationally independent* when the parameter space of the system can be written as the product space $\Theta(d) = \Theta_1(d) \times \Theta_2(d) \times \ldots \times \Theta_m(d), \; d \in \mathbb{D}$.

Technical structure continued:

In this parametric setting, for each decision $d \in \mathbb{D}$ that might be adopted, each panel G_i, $i = 1.2, \ldots, m$ has two quantities available to it.

- *sample densities* over the future measurements for which they have responsibility $\Pi_i^{y|\theta} \triangleq \left\{ \Pi_i^{y|\theta}(\boldsymbol{\theta}_i, d) : \boldsymbol{\theta}_i \in \Theta_i(d), d \in \mathbb{D} \right\}.$
- *beliefs about the parameters* $\Pi_i^{\theta} \triangleq \left\{ \Pi_i^{\theta}(d) : d \in \mathbb{D} \right\}$

Common Knowledge assumptions

- **Utility consensus:** *All agree on the class* $\mathbb{U}$ *of utility functions supported by the* **IDSS**.
- **Policy consensus:** *All agree the class of decision rules* $d \in \mathbb{D}$ *that might be examined by the* **IDSS**.
- **Structural consensus:** *All agree the variables* $\boldsymbol{Y}$ *defining the process of the developing crisis, where for each* $d \in \mathbb{D}$, *each* $U \in \mathbb{U}$ *is a function of* $\boldsymbol{Y}(d)$, *together with a set of qualitative statements about the dependence between various functions of* $\boldsymbol{Y}$ *and* $\boldsymbol{\theta}$. *Call this set of assumptions the* **structural consensus set** *and denote this by* $\mathbb{S}$.

Definition: CK class

Call the set of common knowledge assumptions shared by all panels and which contains the union of the utility, policy and structural consensus $(\mathbb{U}, \mathbb{D}, \mathbb{S})$ the *CK class*.

More common knowledge assumptions

- **Parametric union:** *All agree to adopt as their own G_i's beliefs about the sample families $\Pi_i^{y|\theta}$ $i = 1, 2, \ldots, m$.* This just assigns the specification of the future crisis variables to the appropriate panel.
- **Quantitative delegation:** *All agree to take on the sample summaries $\Pi_i^{y|\theta}$, the panel parameter distributions Π_i^{θ} and the panel marginal inputs Π_i^{y} provided by G_i as their own, $i = 1, 2, \ldots, m$.* i.e. it is appropriate to defer their judgements to the most well-informed panel about each domain vector.

Definition: Adequate IDSS

Call an IDSS *adequate* for a CK class $(\mathbb{U}, \mathbb{D}, \mathbb{S})$ when the SB can unambiguously calculate her expected utility score $\overline{U}(d)$ for any decisions $d \in \mathbb{D}$ she might take under any utility function $U \in \mathbb{U}$ she might be given by a user from the panel marginal inputs Π_i^y provided to her by the panels G_i, $i = 1, 2, \ldots, m$.

Definition: Sound IDSS

Call an IDSS *sound* for a CK class $(\mathbb{U}, \mathbb{D}, \mathbb{S})$ if it is adequate, and by adopting the structural consensus the SB would be able to admit coherently all the assessments $\Pi_i^{y|\theta}, \Pi_i^{\theta}$, (and hence Π_i^y), $i = 1, 2, \ldots, m$ as her own, the SB's underlying beliefs about a domain overseen by a panel G_i being $\left(\Pi_i^{y|\theta}, \Pi_i^{\theta}\right)$, $i = 1, 2, \ldots, m$.

Delegable IDSS

All useful information about parameters is the union of common knowledge and individual panels' specialist information.

Let I_0^t denote all the admissible evidence which is common knowledge to all panel members at time t. Let I_{ij}^t denote the subset of this admissible evidence panel G_i would use at time t if acting autonomously to assess their beliefs about $\boldsymbol{\theta}_j$, $i, j = 1, 2, \ldots, m$, were the SB to commit to policy $d \in \mathbb{D}$.

Define $I_+^t \triangleq \left\{ I_{ij}^t : 1 \leq i, j \leq m \right\}, I_*^t \triangleq \left\{ I_{jj}^t : 1 \leq j \leq m \right\}$

Definition: Delegable IDSS

Say that a CK class of an IDSS is *delegable* at time t if for any possible choice of policy $d \in \mathbb{D}$ and for $j = 1, 2, \ldots, m$ there is a consensus that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}(d) : d \in \mathbb{D}$ $I_+^t \perp\!\!\!\perp \boldsymbol{\theta} \mid I_0^t, I_*^t, d$

Distributivity

In order that panels may update their beliefs autonomously and the SB can use the existing IDSS with updated information $\Pi_i : 1 \leq i \leq m$ we require that the IDSS is panel separable.

Definition: panel separable IDSS

Call $l(\boldsymbol{\theta} \mid \boldsymbol{x}^t)$ *panel separable* over the panel subvectors $\boldsymbol{\theta}_i$, $i = 1, 2, \ldots, m$, when, given admissible evidence $\boldsymbol{x}^t$, it is in the CK class $\mathbb{S}$ that for all $d \in \mathbb{D}$

$$l(\boldsymbol{\theta} \mid \boldsymbol{x}^t) = \prod_{i=1}^{m} l_i(\boldsymbol{\theta}_i \mid \boldsymbol{t}_i(\boldsymbol{x}^t))$$

where $l_i(\boldsymbol{\theta}_i \mid \boldsymbol{t}_i(\boldsymbol{x}^t))$ is a function of $\boldsymbol{\theta}$ only through $\boldsymbol{\theta}_i$ and $\boldsymbol{t}_i(\boldsymbol{x}^t)$ is a statistic of $\boldsymbol{x}^t$, $i = 1, 2, \ldots, m$ known to G_i and perhaps others, collected under the admissibility protocol and accommodated formally by G_i into l_{ii}^t to form its own posterior assessment of $\boldsymbol{\theta}_i$.

i.e. separable likelihoods are key to distributivity

Examples of sound and distributive frameworks

- Staged trees
- Bayesian Networks
- Chain event graphs
- Decomposable graphs
- Multiregression dynamic models

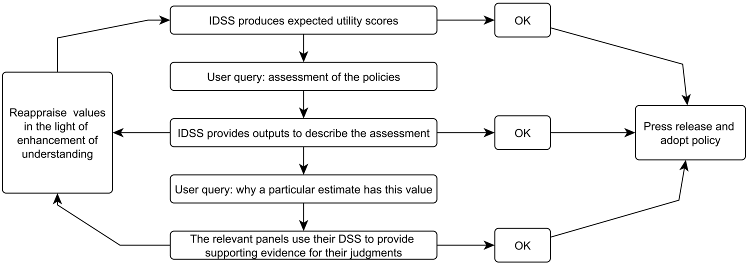


Figure: Manuele Leonelli & James Q. Smith(2015) Bayesian Decision Support for complex systems with many distributed experts Ann Op Res

An IDSS for UK Food security

arXiv 1507.07394

Martine J. Barons © with James Q Smith & Manuele Leonelli

From the literature, and in collaboration with academic experts from many domains we derive the qualitative structure:

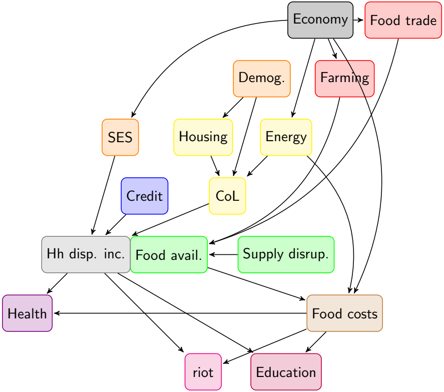


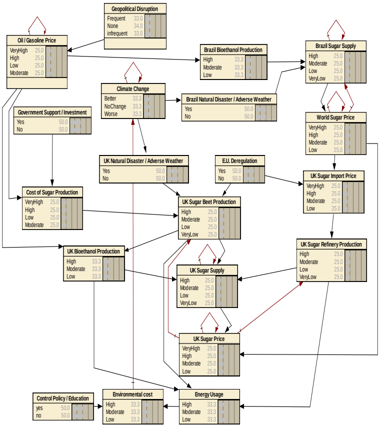
Figure: A plausible schematic of information flows for the modules of a UK food security IDSS. KEY: Economy: UK economic forecasts; Demog.:Demography; Farming: food production; SES: Socio-economic status; Credit: access to credit; CoL:cost of living; Food Avail: Food availability; Supply disrup: food supply disruption; disp. inc.: household disposable income

Calculation of consumer prices index

CPI divides food into 10 categories: Bread & cereals; Meat; Fish; milk, cheese & eggs; oils& fats; fruit; vegetables; sugar, jam, syrups, chocolate & confectionery; coffee, tea & cocoa; mineral waters, soft drinks & juices.

- Need a probabilistic model of supply for each food category
- Sugar category DBN well developed, submitted to AAI 2014
- Meat category by Dmitrijs Murins 2014/5, University of Warwick
- Fish category by Dominic Jones Jan - Feb 2015, Monash University
- Fruit category by Sophia Wright Sept 2015, Oxford & Warwick Universities

Supply sub-networks: Sugar



- Variables identified from research literature.
- List refined with domain experts.
- Conditional probabilities by domain experts
- Dynamic Bayesian Networks for decision support and sugar food security, Martine J. Barons, Xiaoyan Zhong and James Q. Smith, Applied Artificial Intelligence. CRiSM report 14-18

Supply sub-networks: Meat

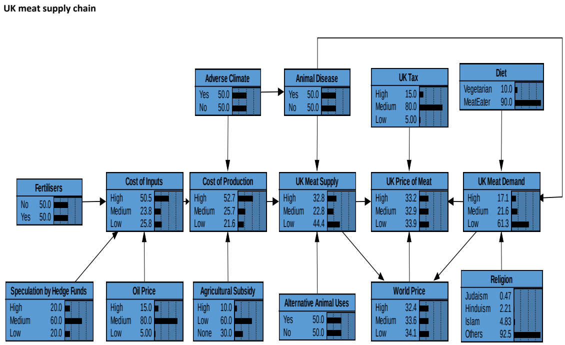


Figure: By Dmitrijs Murins, Department of Statistics, University of Warwick, supervised by Jim Q. Smith and Martine J. Barons

Supply sub-networks: Fish

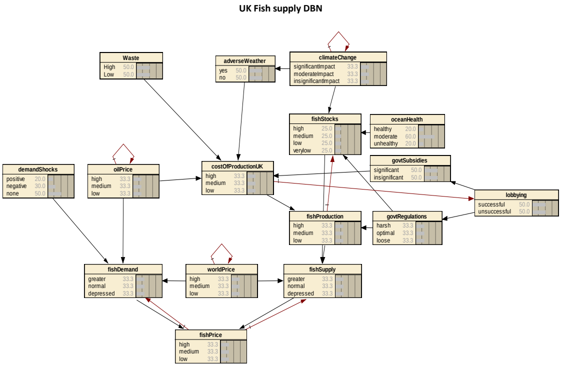


Figure: By Dominic Jones, Monash University, exchange student with Department of Statistics, University of Warwick, supervised by Jim Q. Smith and Martine J. Barons

Supply sub-networks: Fruit

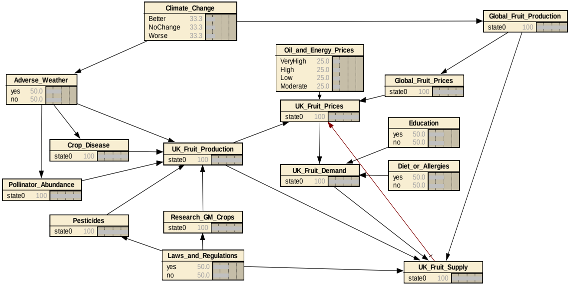


Figure: By Sophia Wright, OxWaSP student jointly with with Department of Statistics, University of Warwick and Oxford University supervised by Jim Q. Smith and Martine J. Barons

Farming sub-networks: Pollination

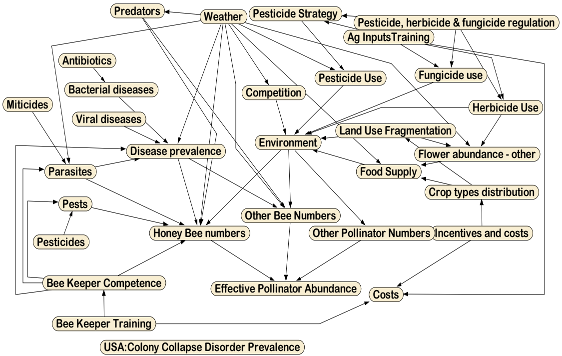


Figure: Expert-based approach to stucture representation by MJB

Utility & Policy consensus

Based on the National pollinator strategy & discussion with DEFRA

- Purpose of the IDSS: to evaluate policy affecting bees and other insect pollinators
- Attributes of the utility: pollinator abundance
- Attribute measurement: citizen science, bee-keeper surveys, species distribution modelling
- Range of decisions: includes interventions on bee-keeper education, pest control, planting, planning & land use

Data problems

- Difficult to measure
- Statistical challenges using this data
- Reliance on citizen science, biases (BWARS, HRS)
- Quality of data & experiments (Bees n beans)
- Experimental evidence mixed, sparse
- Marginal probabilities have some evidence, conditionals little or none

Qualitative structure elicitation

- Academic experts (UK)
- Domain literature (International)
- Domain experts (UK & Australia)
- Policymakers
  - Honey bees
  - Wild bees
  - Other pollinators
- Commercial beekeepers
- Queen breeders

Modelling as a Dynamic Bayesian Network

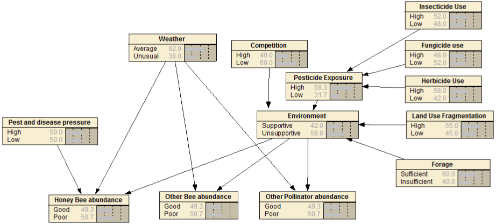
CK-class (Utility, Policy & Structural consensus) all complete; derived dynamic Bayesian network representation.

need to:

- Parameterise the model
- Elicit quantities not in literature
- Test & refine ready for use

Expert Judgement

To quantify key unknown elements Using IDEA protocol



IDEA protocol

IDEA protocol

- Experts first investigate and answer questions without engaging in discussion
- Individual judgements remain anonymous
- Experts see anonymised judgements of peers
- Discuss differences
- Second estimate strictly anonymous
- Expert estimates mathematically aggregated

A.M. Hanea, M.F. McBride, M.A. Burgman, B.C. Wintle, F. Fidler, L. Flander, C.R. Twardy, B. Manning, S. Mascaro, 2016, **I***nvestigate* **D***iscuss* **E***stimate* **A***ggregate* for structured expert judgement, International Journal of Forecasting, accepted for publication on 25.02.2016

Calibration exercise

Calibration exercise: IDEA protocol

- Same protocol as main elicitation: estimate, discuss, new estimate
- Correct answers known or will become known soon (Papers in press)
- Measure accuracy of individual experts
- Accuracy of domain knowledge
- Accuracy of probability estimation

Accuracy scoring for weighting

A scoring rule to measure accuracy of probabilistic predictions.

The Brier Score

Consider question i with two possible outcomes (j). The Brier score of participant k assessing question i is:

$$BrierScore_i^k = \sum_{j=1}^{2} (p_{ij}^k - x_{ij})^2$$

where p_{ij}^k is participant k's probability for question i, output j, and $x_{ij} = 1$ if output j occurs and 0 otherwise.
Usually, participants' accuracy is measured over N questions and averaged to estimate long term accuracy:

$$BrierScore^k = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{2} (p_{ij}^k - x_{ij})^2$$

The elicitation workshop

Question 1.3 – Round 1

What is the probability of observing **good** honey bee abundance, given that the environment is **supportive**, the weather is **unusual**, and the varroa control is **good**?

Question	Answer (in %)
Think about all of the data and reasons why the probability of observing good honey bee abundance, under the above conditions, is low. With these in mind, realistically, what do you think is the lowest plausible bound?	
Think about all of the data and reasons why the probability of observing good honey bee abundance, under the above conditions, is high. With these in mind, realistically, what do you think is the highest plausible bound?	
On balance, what is your best estimate for the probability of observing good honey bee abundance, under the above conditions?	
Comments and Reasons [Please enter any comments, additional knowledge or justification that you have about this question and /or your estimate]	

The elicitation workshop

Elicitation April 6th 2016

Warwick Statistics

cebra
Centre of Excellence for Biosecurity Risk Analysis

Parameters Elicited

Honey Bee abundance

- Good: overwinter losses are less than 30% (as defined by Coloss)
- Poor: overwinter losses are more than 30% (as defined by Coloss)

Other Bee abundance

- Good: if there are 500 or more observations of bees on BWARS in the spring season
- Poor: if there fewer than 500 observations of bees on BWARS in the spring season

Other Pollinator abundance

- Good: if there are 500 or more observations on the Hoverfly recording scheme in the spring season
- Poor: if there fewer than 500 observations on the Hoverfly recording scheme in the spring season

Parameters Elicited

Environment

- Supportive: If there is at least 1 patch of at least 1 hectare of open flowers within 1km spring forage range and pesticide usage is less than 0.3kg/hectare assuming the pesticide toxicity stays the same as at present.
- Unsupportive: : If there is no patch of 1 hectare of open flowers within 1km spring forage range and pesticide usage is more than 0.3kg/hectare assuming the pesticide toxicity stays the same as at present

Weather

- Average: if the number of days with more than 0.2mm of rain fall between 35-70, hours of sunshine fall between 240-480 and mean daily temperature falls between 3-10C
- Unusual: if rain, sunshine or temperature falls outside these ranges

Parameters Elicited

Varroa control

- Good: if there are fewer than 1500 mites per hive in spring, as calculated by BeeBase calculator
- Poor: if there are more than 1500 mites per hive in spring, as calculated by BeeBase calculator

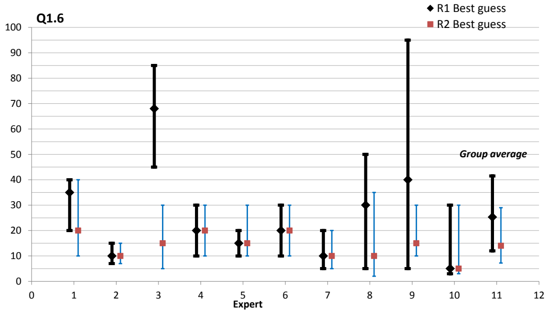
Modified definitions after the discussion

- Other Bee Abundance is all wild bees (Bumblebees and solitary bees) not simply Bombus (Bumble bees).
- Other Pollinator Abundance counts all hoverflies and flies but not butterflies.
- All numbers such as 500 observations in bee abundance definitions should be read as national average over the last five years.
- For a supportive environment rather than at least 1 hectare of open flowers it should be at least 15% proportion of semi-natural land.

Expert Judgement

Expert elicitation using IDEA protocol

Q1.6: What is the probability of observing good honey bee abundance, given that the environment is *unsupportive*, the weather is *average*, and varroa control is *poor*?

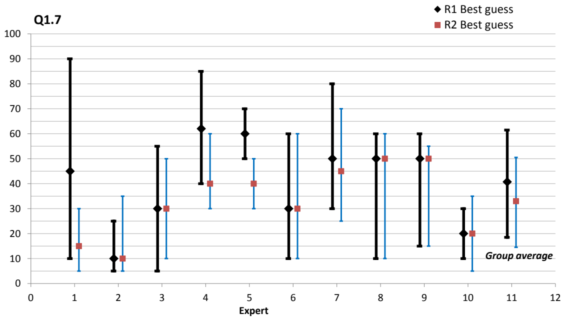


Expert Judgement

Martine J. Barons © with James Q Smith & Manuele Leonelli

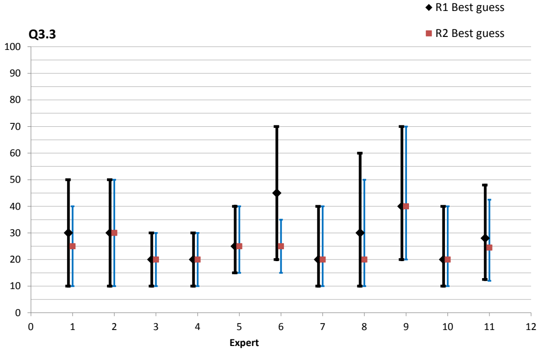
Expert elicitation using IDEA protocol

Q1.7: What is the probability of observing good honey bee abundance, given that the environment is *unsupportive*, the weather is *unusual*, and the varroa control is *good*?



Expert Judgement

Expert elicitation using IDEA protocol
Q3.3: What is the probability of observing good other pollinator abundance, given that the environment is unsupportive, the weather is average?



Calibration exercise

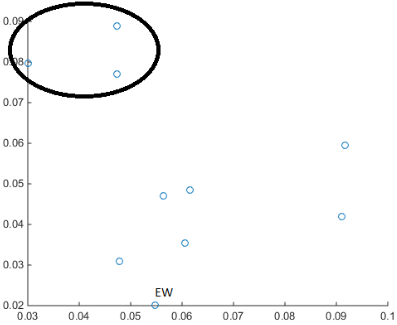
Questions from papers accepted but not published

The measures of performance we considered are:

- The Brier score (per question, per expert) scores close to 0 are good
- The average Brier score (per expert) scores close to 0 are good (a big score corresponds to poor performance; a 0.5 score can be achieved by setting all answers to 0.5)
- The length of the uncertainty interval (per question, per expert) small scores are better
- The calibration term of the Brier score (one number per expert calculated from all questions) smaller scores are better
- Relative informativeness (one score per expert calculated from all answers) departure from the [0.5 0.5] distribution larger scores are better

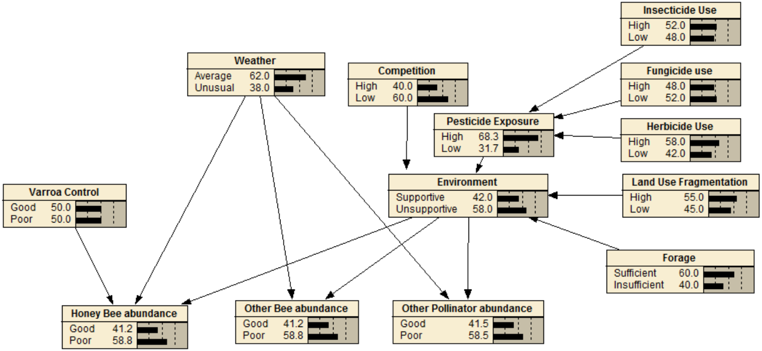
Calibration exercise

Brier score (x-axis) against relative informativeness (y-axis)



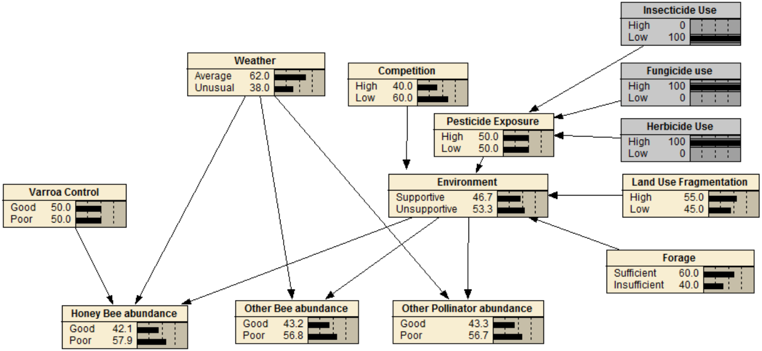
The optimal combinations are in the upper left corner. Three experts (2, 3, and 5) do quite well, based on this (& all) measures, but not significantly. This means that the original questions can be combined with equal weighting.

What-if analysis in Netica



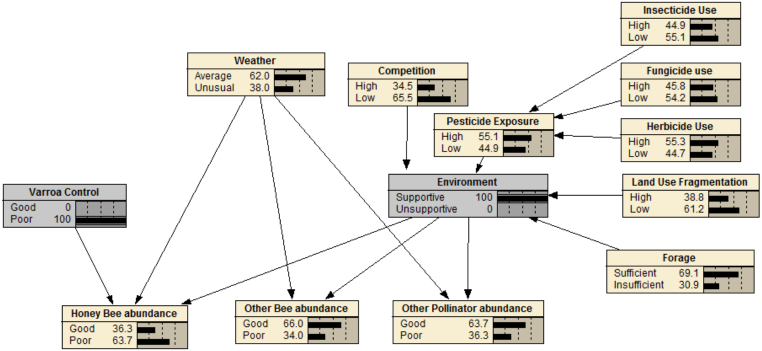
Simple group average; not accounting for uncertainties here

What-if analysis in Netica



Simple group average; not accounting for uncertainties here

What-if analysis in Netica



Simple group average; not accounting for uncertainties here

Next...

- Rest of pollination sub-network populated with probability distributions from data
- The pollinator sub-network able to be used for decision support for pollinator abundance.
- Domain paper, Statistics paper
- Integrate into expert panel for farming and food supply within the UK food security IDSS.
- WCC IDSS
- CCC survey linkage TT; modelling underlying food poverty

Martine.Barons@warwick.ac.uk
go.warwick.ac.uk/MJBarons

WARWICK Global Research Priorities FOOD

THE UNIVERSITY OF WARWICK

EPSRC Engineering and Physical Sciences Research Council

Warwick Statistics

Coherent Frameworks for Statistical Inference serving Integrating Decision Support Systems *Jim Q. Smith, Martine J. Barons & Manuele Leonelli*, on arXiv: 1507.07394 EPSRC EP/K039628/1