# TITLE

Revealing beliefs: using ensemble ecosystem modelling to extrapolate expert beliefs to novel ecological scenarios.

# AUTHORS

Michael Bode [1*], Chris Baker [1], Joe Benshemesh [2], Tim Burnard [3], Libby Rumpff [1], Cindy Hauser [1], José Lahoz-Monfort [1], Brendan A. Wintle [1]

1. School of Biosciences, University of Melbourne, Melbourne, Australia.
2. School of Life Sciences, LaTrobe University, Melbourne, Australia.
3. Birdlife Australia, 60 Leicester St Carlton, Melbourne, Australia.

* Corresponding author

Email: mbode@unimelb.edu.au

Phone: +61 4 14 108 439

**Manuscript length:** 273 word abstract

7,063 words (includes tables, captions, references)

4 Figures; Supplementary materials.

**Running title:** Ensemble models of expert beliefs

**Keywords**: *Expert judgement; ensemble forecasting; qualitative modelling; uncertainty analysis; interaction networks.*

**SUMMARY**

1. Ecosystem-based management requires predictive models of ecosystem dynamics. There are typically insufficient empirical data available to parameterise these complex models, and so decision-makers commonly rely on beliefs elicited from experts. However, such expert beliefs are necessarily limited because (1) only a small proportion of ecosystem components and dynamics have been observed; (2) uncertainty about ecosystem dynamics can result in contradictory expert judgements; and; (3) elicitation time and resources are limited.
2. We use an ensemble of dynamic ecosystem models to extrapolate a limited set of stated expert beliefs into a wider range of revealed beliefs about how the ecosystem will respond to perturbations and management. Importantly, the method captures the expert uncertainty and propagates it through to predictions. We demonstrate this process and its potential value by applying it to the conservation of the threatened malleefowl (*Leipoa ocellata*) in the Murray mallee ecosystems of southern Australia.
3. In two workshops, we asked experts to construct a qualitative ecosystem interaction network and to describe their beliefs about how the ecosystem will respond to particular perturbations. We used this information to constrain an ensemble of $10^9$ community models, leaving a subset that could reproduce stated expert beliefs. We then interrogated this ensemble of models to reveal experts' implicit beliefs about management scenarios that were not a part of the initial elicitation exercises.
4. Our method uses straightforward questions to efficiently elicit expert beliefs, and then applies a flexible modelling approach to reveal those experts' beliefs about the dynamics of the entire ecosystem. It allows rapid planning of ecosystem based management informed by expert judgement, and provides a basis for value-of-information analyses and adaptive management.

## INTRODUCTION

Ecological management relies heavily on expert beliefs (Kuhnert *et al.* 2010; Burgman *et al.* 2011; Martin *et al.* 2012). Ecological systems are incredibly complex, with thousands of species interacting across space and time (Turchin 2003), and the time and resources available to study them are severely constrained. As a consequence, ecological communities and their dynamics are poorly understood (Lawton 1999; Kuhnert *et al.* 2010). Conservation managers nevertheless need to respond to multiple threats, often before experimental or observational evidence can be systematically collected. Expert beliefs allow managers to rapidly assess which management problems are most important, and which actions will best mitigate their effects (Kuhnert *et al.* 2010; Martin *et al.* 2012).

Unfortunately, while expert beliefs can offer decision-makers timely information, they have two key limitations. First, expert beliefs are incomplete, in the sense that they do not systematically describe all the components and dynamics of an ecosystem. Experts have generally only observed a small subset of possible dynamics and by definition cannot have observed novel circumstances (e.g., responses to untested management interventions). Second, expert beliefs are always uncertain. While structured elicitation methods can reduce the magnitude of uncertainty, uncertain beliefs about system dynamics are inevitable (Kuhnert *et al.* 2010; Martin *et al.* 2012; Wintle *et al.* 2013). We stress that these factors are not exclusive to elicited expert beliefs, but they do limit the utility of expert opinion for conservation decision-making. Furthermore, because elicitation is time-consuming and expert experience is limited, the solution is not simply to elicit more information (Kuhnert *et al.* 2010).

Expert beliefs are particularly limited when managing whole ecosystems. Conservation is increasingly moving from a single-species focus to the management of whole ecosystems (Garrett 1992; Grumbine 1994). This reflects a more expansive definition of conservation value that includes a greater range of biodiversity (Margules & Pressey 2000), the increasingly appreciated economic value of ecosystem processes and functions (Armsworth & Roughgarden 2001), and an awareness of how complex and indirect ecosystem interactions can determine the consequences of conservation actions (Raymond *et al.* 2010). As the components of an ecosystem being considered (its biotic and abiotic factors) increase in number, the number of ecosystem interactions and processes that need to be understood increase nonlinearly. We therefore need a method that can rapidly predict a wide range of ecosystem dynamics on the basis of uncertain and incomplete expert beliefs. This is the primary goal of this paper.

At the centre of this method will be an ecosystem model. Most ecosystems can be readily described by a network of interactions among ecosystem elements (Pimm *et al.* 1991). These qualitative models describe direct relationships between important ecosystem components (species, or environmental & anthropogenic drivers) using cause-and-effect connections, but without specifying the magnitude or functional form of the relationship (Levins 1974). A single *qualitative* network can therefore be represented by a very large set of *quantitative* community models. Rather than choose any particular model in this set (e.g., the best-fit to known data), we represent the interaction network by a very large ensemble of models. Most importantly, we ensure that each model in this ensemble can recreate any stated beliefs that we have been able to elicit from experts. The resulting model ensemble can be used to make predictions about any aspect of ecosystem dynamics, in response to any modelled perturbation or management action.

We describe and demonstrate this approach for the management of the malleefowl *Leipoa ocellata* (Gould 1840), a threatened bird species from Australia's semi-arid and arid zones that has experienced a substantial decline over the last two decades, but for uncertain reasons (Benshemesh 2007; Benshemesh *et al.* 2007). We undertook two workshops to elicit *stated* expert beliefs about the structure and dynamics of mallee ecosystems from a suite of relevant experts. Our method translates these limited and uncertain stated beliefs into a large ensemble of predictive, quantitative ecosystem models. This model ensemble can then be manipulated to answer new questions. The results of these simulations reveal expert beliefs about ecosystem dynamics that are not explicitly stated during the elicitation process. These *revealed* expert beliefs (embodied in the predictions of the model ensemble) can be used to inform management decisions, and guide future research.

## MATERIALS AND METHODS

### Expert workshops

Participants at two workshops were chosen to represent a cross-section of expertise on mallee ecosystems, including managers from nongovernmental conservation organisations; government; university researchers; ecological consultants; and conservation volunteers. Remnant mallee is broadly distributed across southern Australia, from New South Wales to Western Australia, but we focused our analyses on the Murray mallee ecosystem (MDD02 IBRA subregion) that contains high densities of malleefowl in well-studied populations (Benshemesh 2007). As with any ecosystem, there are a number of competing hypotheses about the drivers of malleefowl decline, which have been variously ascribed to the effects of invasive mammalian predation, herbivore competition, habitat degradation, altered fire regimes and climate change (Benshemesh *et al.* 2007; Bode & Brennan 2011; Garnett 2012; Walsh *et al.* 2012). We sought to include participants who represented a range of perspectives on the relative priority of these threats.

The first workshop constructed qualitative ecosystem interaction models. A set of important "ecosystem components" (species, or environmental drivers such as fire and rainfall), were joined by cause-and-effect connections. Connections were drawn if a change in one component was expected to directly cause a change in another component, with the sign of the relationship indicating whether the change in the recipient component will be negative or positive. Only direct interactions were included in the model. For example, if rabbit populations have a positive effect on dingoes through predation, but a potentially negative effect through damage to vegetated habitat, we only included the direct positive interaction, and allowed the model to incorporate the negative effect via links from rabbits to vegetation, and from vegetation to dingoes (Baker *et al.* 2016a). We incorporated structural uncertainty by allowing the experts to define relationships that they believed existed, but were of unknown sign (i.e., they could be either positive or negative), or that they were unsure existed but would be certain of the sign if they did (i.e., they could be zero or positive). A full description of the workshop and the results can be found in *Supplementary Information 1.*

The second workshop elicited uncertain information from participants that could be used to constrain the predictions of the qualitative interaction network. In 14 different scenarios, an abiotic or anthropogenic driver from the qualitative model changed by a particular magnitude (e.g., rainfall decreased by 75% for one year), following approximately 10 years of relatively constant ecosystem conditions. We explained that this period of unchanging conditions was to ensure that any large prior perturbations (e.g., a recent fire) were no longer playing a large role in the ecosystem dynamics. Experts were asked to quantitatively describe how a different ecosystem component would respond over the next 5 years, a length of time considered long enough to reveal dynamics over the short- to medium-term. The participants submitted their answers by drawing "envelopes" on a timeseries graph that described their belief and uncertainty about the response (Figure 1). While participants were encouraged to draw envelopes for all scenarios, they were free to not answer questions they felt were beyond their experience or intuition. A full description can be found in *Supplementary Information 2*.

When eliciting information from multiple experts, evidence shows that iterative rounds of anonymised feedback between experts (Kuhnert *et al.* 2010) improves the accuracy of estimates (Rowe & Wright 1999). We chose to elicit information from experts independently, in a single round. This allowed us to maximise the number of scenarios we could consider in one workshop, since we are primarily interested in the process of extrapolating from a range of stated beliefs, rather than eliciting the most accurate information.

## Ensemble ecosystem modelling

When predicting the future dynamics of a complex, nonlinear system, it is better to base decisions on the ensemble predictions of a large number of plausible models, rather than rely on a single model, even if that single model offers the best fit to the available validation data (Leith 1974). Ensemble prediction is an essential component of modern meteorology and the associated assessment of environmental risks, with the IPCC's multi-model predictions of global climate representing the best-known application of the technique (Stocker 2014). A recent variant of the approach is ensemble ecosystem modelling (EEM; Dexter *et al.* 2012; Gårdmark *et al.* 2013; Bode *et al.* 2015; Hunter *et al.* 2015), which applies the approach to models of ecosystems. Following our two workshops we generated an ensemble of models whose structure matched the beliefs of the first expert workshop, and whose parameterisations were consistent with expert beliefs (see *Supplementary Figure S1* for a schematic overview of the process).

For EEM, we define a large ensemble of models with a given qualitative structure (in our case, the interaction network identified in our first workshop). We use a system of Lotka-Volterra (LV) equations, where the amount of a component *i* in an ecosystem at time *t* (the abundance or density of a species, the volume of rainfall, etc) is defined as $N_i(t)$. This amount changes according to the component's internal dynamics, and its interactions with other components:

$$\frac{dN_i}{dt} = r_i N_i + \sum_{j=1}^{C} a_{ij} N_i N_j$$

(Equation 1)

where $C$ is the number of components in the ecosystem, $r_i$ is a component's intrinsic growth rate, and $a_{ij}$ are interaction coefficients describing the per-unit effect of component $j$ on each unit of component *i*. The interaction matrix **A** (Figure 2a) contains the elements $a_{ij}$ which match the sign structure of elicited interaction networks (Figure 2b). LV models are designed to describe the dynamics and stability of foodwebs (Pimm *et al.* 1991; Turchin 2003) but can be extended to describe abiotic components. An environmental driver such as rainfall is not affected by any other components of the ecosystem ($\alpha_{ij} = 0$ for $i \neq j$). Its intensity is therefore defined by its stable equilibrium value, determined by the ratio of $r_i$ and $\alpha_{ii}$.

We construct an ensemble of $10^9$ different models by choosing random values for growth rates and interaction terms. Growth rates are chosen at random from an inverse distribution (i.e., $1/(r_i + 1) \sim U(0,1)$), allowing them to take any positive value. The magnitudes of the interaction coefficients are chosen from unit uniform distributions ($\alpha_{ij} \sim U(0,1)$), with their signs assigned according to expert beliefs. Latin hypercube sampling can generate random numbers that efficiently sample this high-dimensional parameter space. Although the choice of bounded distributions for the interaction coefficients may seem limiting, any LV system can be rescaled to produce an ecosystem model with parameters within these bounds (*Supplementary Information 3*). In addition, simulations show that model predictions are robust to the distributions from which parameters are chosen (Baker *et al.* 2016b).

We then use expert beliefs about ecosystem dynamics to constrain the model ensemble. First, we remove any models from the ensemble that are not "viable"; that is, where not all the species that were listed can persist at equilibrium. To assess viability, we calculate the equilibrium state of the ecosystem, and determine whether all species have positive abundances (Baker *et al.* 2016b). In altered ecosystems, it may not be certain whether species that are currently extant will be able coexist over the medium to long term, and in these circumstances the equilibrium coexistence condition will be inappropriate. For example, malleefowl have coexisted with foxes for approximately 190 years, but malleefowl have a 10% probability of becoming extinct in the next 100 years (according to IUCN Red List Criterion E for Vulnerable), with foxes listed as a key threatening process. They may therefore be on a long trajectory towards extinction, and unable to persist alongside foxes. In these latter cases, we could simulate the models for the finite length of observed coexistence (e.g., 190 years), rather than calculate equilibrium abundances. We would then remove any models where at least one species declines below a threshold (e.g., malleefowl fall below 0.1% of their initial abundance). Given that all species eventually become extinct, a finite coexistence time is probably a more realistic constraint on the model ensemble, although it is more computationally demanding.

Next, we simulate the dynamics of each model in response to the perturbations (described in the second workshop), using the equilibrium as the initial condition. We compare the predicted changes in species' abundance to the uncertain envelopes drawn by the workshop participants, and penalise any models that disagree with the expert beliefs. We measure the "performance" of each model in the ensemble as shown graphically in Figure 1, by calculating the overlap between a model and the expert beliefs: for every time step the model falls within any expert envelope, its performance increases by a constant amount. Intersecting with an envelope for twice as long yields twice the benefit; intersecting two envelopes provides twice as much benefit as intersecting a single envelope. Under this measure of model performance, we include only the best 5% of models in the ensemble, an approach conceptually similar to Approximate Bayesian Computation (Beaumont 2010). Once the best performing models are identified, we calculate the proportional change in Shannon entropy associated with each model coefficient. This change measures the amount of information imparted to each uncertain coefficient by the set of envelope constraints (*Supplementary Figure S2*).

## Analyses

The remaining ensemble of models encapsulates the experts' beliefs – revealed as well as stated – about the dynamics of the mallee ecosystem. We undertake two sets of analyses to illustrate the potential of EEM, and the flexibility of revealed expert opinion.

Our first set of analyses illustrates how limited stated expert beliefs can be extrapolated to construct revealed expert beliefs. In our second workshop we asked questions about 14 different ecosystem perturbation scenarios (*Supplementary Information 2*). We used EEM to answer three additional perturbation questions about malleefowl populations that were purposely not explored in the workshop: (1) How will malleefowl abundance change in response to a 25% increase in dingo abundance over 5 years? (2) How will malleefowl abundance change in response to a 25% increase in the cat population over 5 years? (3) What will be the effect of additional annual migration of malleefowl into a population, equal to 10% of the equilibrium population, as a consequence of either natural dispersal, or a managed release from captive populations? EEM allows us to extrapolate experts' revealed beliefs regarding these three questions, by simulating the response of each model in the ensemble. We then extract and graph the range of malleefowl population responses, with the ecosystem simulated in weekly timesteps for a 5 year period. This time horizon matched the experts' stated beliefs, and is long enough to reflect the approximate timescale of malleefowl funding (e.g., the Malleefowl Management Committee funding lasted 7 years; the Australian Research Council Linkage grant that funded this work lasted 3 years). We repeat each simulation using both the unconstrained and constrained model ensemble (i.e., the set of models before and after we consider the envelopes), contrasting these simulations to illustrate the value of stated expert beliefs.

Our second set of analyses illustrates how EEM can offer management support that is formal and ecosystem-based, but is also rapid and efficient. The malleefowl National Recovery Plan

(NRP; Benshemesh 2007) lists six important threats to malleefowl, and we use the model ensemble to predict the impact of mitigating each in turn over 5 years. The key threats of the NRP and the specific details of our model simulations are: (1) *Habitat loss and fragmentation*. We assume that active restoration results in an exogenous 15% increase in suitable habitat (the seedling & vegetation components). (2) *Competition by grazing herbivores*. We model the effects of reducing feral goats by 30% through mustering, and rabbit populations by 30% through baiting. These are reasonable outcomes for ecosystems like the mallee (Parkes *et al.* 1996; Cooke 2010). (3) *Predation by introduced foxes*. We assume that effective baiting can reduce fox populations by 95%, in line with best practice in similar ecosystems (Saunders & McLeod 2007). (4) *Predation by introduced cats*. We model the effects of baiting that targets cat populations, reducing them by 85%. Reductions of this magnitude have been previously achieved in non-insular arid and semi-arid ecosystems (Algar & Burrows 2004). (5) *Fire intensity and severity*. Both dimensions of fire affect malleefowl negatively, and we model the effects of currently planned management changes to public land in the Murray mallee, which will increase the area burned by fire by at least 50% from current levels. (6) *Disease and inbreeding*. We assumed that these two factors act to reduce population growth rates, generally through increased mortality (Keller 2002). Although it is not clear how these threats would be addressed by managers, we assume that the benefit of managing disease and inbreeding will increase the population growth rate by 10%. In each case, we use EEM to simulate the range of consequences for malleefowl abundance. We note, however, that the results will reflect the above assumptions about management effectiveness, which are only based on a limited literature survey, and will vary with location and management actors.

## RESULTS

The first workshop generated three different interaction networks that connected similar ecosystem components in slightly different configurations (*Supplementary Information 1*). For the analyses that follow, we analyse the network produced by the first expert group (Figure 2). This network connects 14 ecosystem components with 80 direct interactions (in the matrix, we ignore intraspecific interactions and only consider off-diagonal elements). While this creates a complex interaction network (Figure 2b), it is fewer than half of the 182 possible direct connections, and the dynamics of many components are therefore only indirectly coupled. The majority (65%) of these direct interactions were qualitatively certain (either definitely positive or definitely negative), with the remainder being either of uncertain existence but known sign (28.8%), or of unknown existence and sign (6.2%).

In the second workshop, we were able to elicit 62 beliefs from 13 experts about 14 ecosystem perturbation scenarios (*Supplementary Figure S6*). Every scenario received between 2 and 7 different expert beliefs. The average expert was not able or willing to describe their beliefs about most scenarios, or did not have sufficient time (34% of 182 potential beliefs were elicited). While opinions about some scenarios were quite consistent (e.g., all experts believe that cat abundance will increase during fox baiting), others differed markedly (e.g., fox abundance could increase or decrease during overgrazing).

On the basis of the network structure, the viability constraint and the elicited envelopes, we were able to reduce the original set of $10^9$ LV models down to a ensemble of approximately $10^5$ models. The substantial constraints offered by the experts' beliefs (to <0.01%) were heterogeneously distributed across the unknown interaction parameters. The proportional change in Shannon entropy associated with each coefficient indicates that, while three-quarters of the coefficients only experienced a small reduction (<20%) in entropy, the coefficients associated with malleefowl, vegetation, predators and grazers experienced a large (>80%) reduction (*Supplementary Figure S2*). This concentration of information on a few parameters reflects the focus of the elicitation scenarios on the interactions between malleefowl and predators, and between malleefowl, vegetation and herbivores. Despite the reduction in the size of the model ensemble and the increase in information, the retained models are enormously variable. Although all are based on the same interaction network, and can replicate all stated expert beliefs, their parameterisations are vastly different (*Supplementary Figure S3*), and they therefore represent alternative hypotheses about what ecosystem dynamics could generate the stated expert beliefs. Models with as many as 16 structural differences were able to recreate the same dynamics. For example, about half of the models retained in the ensemble considered fire to have a positive effect on cat abundance; the other half considered it to have a negative effect. As a result of this variability, the correlation structure of the retained parameterisations is indistinguishable from purely random data, suggesting that the models remaining in the ensemble are dynamically very different from one another. The predictions of these retained models are similarly variable – species' responses to perturbations are generally of ambiguous sign. Thus some of the ensemble predicts increases in a given component, while others predict decreases. The magnitude of the changes also varies by more than an order of magnitude (e.g., some models predict a 10% increase in abundance, some predict a 100% increase).

Despite this variability, the first set of analyses shows that constraining the model ensemble with expert opinions reveals additional and informative expert beliefs. Figure 3 shows the change in malleefowl abundance that would result from an increase in dingo abundance; an increase in cat abundance; and increased malleefowl immigration. The grey envelopes show that the set of viable models is incredibly variable before they are constrained by the stated expert beliefs. This is even true when, as is the case for increased immigration (Figure 3c), the changes have a direct and positive impact on the malleefowl population. The blue envelopes show the revealed expert opinions, which are much narrower than the original set of possible trajectories. An increase in dingoes to the ecosystem will have an uncertain effect on malleefowl abundances, ranging from a decrease of 20% to an increase of 30% (Figure 3a). This range of revealed expert beliefs matches the uncertainty surrounding the effect of toppredators on prey species in the literature, particularly in Australia's semi-arid rangelands (Allen *et al.* 2013). Our other revealed expert beliefs show more confidence in the effects of ecosystem perturbations: if cat populations increase, malleefowl will most likely decline (a 0–15% decrease; Figure 3b); if malleefowl immigration increases, malleefowl populations will also experience a small increase (0–20%; Figure 3c).

Our second set of analyses uses EEM to calculate experts' revealed beliefs about the benefits of management actions that affect each NRP threat (Figure 4). The most striking result of these revealed beliefs is their uncertainty. While each management action could benefit malleefowl populations, the combined effect of direct and indirect ecosystem interactions could also result in a perverse negative outcome. Both habitat management and grazing management appear as likely to damage malleefowl populations as they are to benefit them. The revealed beliefs are less ambiguous about the effects of managing cat predation (likely positive) or an increase in fire intensity/frequency (likely negative), but the 95% EEM confidence intervals for both interventions still overlap zero. The sole exception to this qualitative uncertainty is the revealed belief about the positive effects of addressing inbreeding and disease in the populations. Interestingly, the management of foxes is arguably the most commonly undertaken action to benefit malleefowl, and could have the largest positive or the largest negative effects on malleefowl abundance.

## DISCUSSION

EEM allows limited stated expert beliefs to be extrapolated, revealing implicit beliefs about the broader dynamics of an ecosystem and its response to perturbations. Our application of these methods to malleefowl conservation produced a quantitative decision-support tool after two workshops and a relatively small amount of computational analysis. The method allows beliefs to be elicited at minimal cost, and therefore reduces burden on experts. It translated expert beliefs into a quantitative tool that we used to rapidly estimate the expected benefit and uncertainty of actions aimed at mitigating each threat. The total cost of the two workshops required to do this was approximately $10,000 (2015 Australian dollars).

The EEM process is computationally demanding but conceptually straightforward, and it offers decision-makers three primary benefits. First, EEM reveals a much broader range of expert beliefs about their ecosystems, without requiring them to answer an enormous number of questions. This process provides a logical and internally-consistent method of extending expertise to new and more complex problems. Once a few expert beliefs have been elicited, the decision-maker can ask an enormous number and range of questions at essentially no cost: expertise on tap. This provides substantial efficiencies: when eliciting information from 13 experts about a 14-component interaction network, there are 2,548 single-perturbation questions that can be asked. Our half-day workshop answered 63 of these questions; a desktop computer use EEM to answer the remainder in less than one minute (the model ensemble took approximately one day to create, but this can be pre-computed).

Second, we elicited stated beliefs about ecosystem dynamics by asking experts to construct interaction networks, and to draw uncertain envelopes describing the response of different components to perturbations. These forms of elicitation are simple and intuitive since both interaction networks and uncertain timeseries data are common elements of undergraduate biology degrees, ecological reports, and journal articles. In contrast, the alternative method of eliciting information about dynamic networks is to ask individual questions about the model parameters (Kuhnert *et al.* 2009), for example, about per-capita growth rates, interaction coefficients, or conditional probabilities for Bayesian networks (Martin *et al.* 2012). In contrast to our timeseries questions, these require difficult and numerically-precise statements about implicit and unobservable ecological quantities, and impose a high elicitation burden on experts. EEM allows these more difficult quantities to be computed from the envelopes.

Third, EEM can consider questions that are difficult or impossible to engage with using standard expert elicitation. Expert observations and beliefs concentrate on a subset of ecosystem components and dynamics: easily-observed species, recent perturbations, previous management actions, and contemporary environmental and climatic conditions. If we accept that ecosystems are in part driven by the deterministic interactions of a connected system with consistent dynamics, then observed phenomena can offer insights into unobserved events. Thus EEM allows us to extend stated expert beliefs to a much broader set of predictions about ecosystem dynamics and management actions. We expect that such predictions will be uncertain, and accept that they will often be ambiguous (e.g., Figure 4).

Our analyses demonstrate that an EEM approach, constrained by a reasonable number of stated expert beliefs, can provide useful predictions about the performance of different management actions (Figure 4). For example, the beliefs elicited from experts indicate that managing diseases or cat abundance are very likely to improve malleefowl populations, and that an increase in fire intensity and severity is likely to produce a negative impact. In contrast, the most commonly undertaken management action on behalf of malleefowl – fox baiting – has an entirely uncertain impact, although it could potentially offer the greatest benefits. These results concur with the deep uncertainty highlighted by previous modelling and empirical studies on fox predation (Walsh *et al.* 2012) and fire dynamics (Benshemesh 2007). They reveal that, despite more than a century of conservation research on malleefowl (Mellor 1911), we remain deeply uncertain about the impact of management actions on this threatened species.

Even when constrained by the expert-elicited timeseries, the forecasts made by our model ensemble are enormously variable, to the point of being qualitatively uncertain (Figure 3, 4). This variation is partly the result of over-fitting – we are estimating 108 free parameters using timeseries data on 14 perturbations – but this does not necessarily mean that our models are too complicated. The interaction network is complex (Figure 2), and so our models must also be complex if they aim to offer a fulsome mechanistic explanation of how ecosystem structure drives dynamics. Explicitly modelling the complexity of the interaction network is valuable for two reasons. First, management outcomes are often heavily affected by indirect interactions with the broader network (Raymond *et al.* 2010; Dexter *et al.* 2012; Buckley & Han 2014). Our ensemble offers a range of models that reproduce the stated expert beliefs, but offer competing hypotheses about which direct and indirect interactions produced them (*Supplementary Figure S3*). These competing hypotheses make different predictions about future dynamics, and this is partly responsible for the highly variable predictions. Second, in addition to forecasting future dynamics, a central goal of this method is to extrapolate from a limited set of stated beliefs, to create revealed beliefs about the broader ecosystem. A more parsimonious model might offer more accurate predictions about the future dynamics of observed ecosystem components, but it would be unable to extrapolate across the ecosystem.

Although we do not detail the required steps here, a model ensemble can answer a much wider range of questions. To give a few examples for malleefowl conservation:

- The mallee contains threatened species other than malleefowl. Will management actions that benefit malleefowl (e.g., particular fire regimes) detrimentally affect the viability of other species (Driscoll *et al.* 2016)? EEM models the future dynamics of multiple species simultaneously, identifying conservation trade-offs.
- The varying amount of uncertainty in different future predictions (Figure 3 & 4), and model parameters (*Supplementary Figure S2 & S3*) could be used to undertake a value-of-information analysis (Runge *et al.* 2011), focusing research on reducing uncertainties that most strongly hamper sound decision making and consequent improvements in outcomes.
- Monitoring data – particularly when gathered in response to perturbations or management interventions – can be used with EEM in the same manner as the stated

expert beliefs: to further constrain the model ensemble. Applied iteratively with VOI, managers can use the EEM method to undertake short-term active adaptive management (Benshemesh & Bode 2011), that explicitly considers ecosystem interactions.

Ecosystem dynamics are often modelled with sets of ordinary differential equations (Turchin 2003), and there are reasons to believe that our formulation may offer robust insights into future dynamics. However, two factors in particular should be kept in mind when interpreting the results: ecosystems are unlikely to be perfectly represented by the functional forms of the Lotka-Volterra equations, and ecosystem dynamics are stochastic and spatial. This is particularly true in arid environments, where stochastic and spatially-explicit models of environmental covariates are generally considered essential (Cadenhead *et al.* 2015). Despite these problematic assumptions, there are reasons to hope that EEM can offer useful predictions. First, we asked our experts to describe dynamics in the vicinity of the ecosystems' equilibrium point, and over short time periods (5 years). In this set of states, the precise functional form of the models (e.g., Lotka-Volterra) is much less important, since many different functional forms share the same dynamics (Raymond *et al.* 2010; Melbourne-Thomas *et al.* 2012). It may be appropriate to further account for this issue by eliciting and predicting dynamics over shorter time periods; by discounting the performance of models in the more distant future; by constructing model ensembles using more than one functional form; or by constructing models with alternative plausible interaction structures. Although we elicited three different interaction networks, their structure was very similar, as experts were allowed to move freely between the groups. This non-independence makes the three networks unsuitable for a structural sensitivity analysis. Second, although we model a stochastic ecosystem with an ensemble of deterministic models, the technique of ensemble modelling was adopted in meteorology precisely because it reduces inaccuracies caused by sub-grid-scale stochasticity and unmeasured variation in initial conditions (Leith 1974). We attempt to further reduce the influence of stochasticity by modelling ecosystem components with large abundances, as demographic stochasticity will have a smaller effect on the dynamics of large numbers of individuals (Gustafsson & Sternad 2013).

EEM is part of a field of ideas for making predictions about complex, uncertain, nonlinear systems. The approach is heavily based on ideas from qualitative modelling (QM) in both its loop-analysis (Levins 1974; Dambacher *et al.* 2003) and computational forms (Raymond *et al.* 2010; Dexter *et al.* 2012; Melbourne-Thomas *et al.* 2012). These QM approaches will offer complementary or superior perspectives to EEM for many problems, particularly when smaller interaction networks are sufficient (loop QM), or when elicited constraints and predictions concern short-term and small-magnitude perturbations (computational QM). "Sloppy modelling" analyses, which are increasingly influential in physics and systems biology, are another parallel set of ideas. This approach can offer deeper insights into the most important components and parameters in the system, rather than simply predicting the consequences of perturbations (Gutenkunst *et al.* 2007). Finally, model ensembles have proven invaluable in geophysical fluid dynamics (Leith 1974) and complex and nonlinear statistical modelling (Beaumont 2010). Both the motivation and justification for this approach to prediction can be found in reviews of these fields, as can a range of extensions that will add strength and robustness to our approach.

Resource and time constraints force conservation science to make important management decisions on the basis of limited information. Expert beliefs provide essential guidance in the face of such logistical constraints, but the elicited information is limited and uncertain. Conservation is also increasingly focused on making decisions that consider the highly interconnected nature of ecosystems, and the indirect and counter-intuitive dynamics that these connections create. Ecologists can construct interaction networks that outline such dynamics, but these cannot make the necessary quantitative predictions. Ensemble Ecosystem Modelling offers one solution to both these problems. By merging expert beliefs and qualitative modelling, EEM can systematically extrapolate a limited number of stated expert beliefs into a broader range of revealed implicit beliefs. Not only does this make expert-supported decisions more efficient and quantitative, it also provides a framework for extending them into unobserved future scenarios and untested management actions. The method therefore allows management options to be quickly and defensibly prioritised, and does so using a framework that explicitly takes ecosystem interactions and indirect effects into account. EEM therefore helps to address three key obstacles to effective conservation action: complex ecosystem interactions, limited information, and limited resources.

## ACKNOWLEDGEMENTS

We are grateful to the participants of the two expert workshops and their organisations. We also thank the partners on the ARC Linkage Project (LP120100490) that funded this work: the Victorian Malleefowl Recovery Group, Parks Victoria and Iluka Mining. MB and CB were supported by an ARC DECRA (DE130100572); BW was supported by an ARC Future Fellowship (FT100100819). TB, JLM and CH were supported by an ARC Linkage grant (LP120100490), the *National Environmental Research Program and the National Environmental Science Program - Threatened Species Recovery Hub*.

## DATA ACCESSIBILITY

This manuscript does not contain any data.

## AUTHOR CONTRIBUTIONS

All authors conceived the ideas, held the workshops, interpreted the results, and drafted the manuscript. MB and CB developed the methods. MB undertook the analysis. All authors gave final approval for publication.

## REFERENCES CITED

Algar, D. & Burrows, N.D. (2004). Feral cat control research: Western Shield review-February

2003. *Conservation Science Western Australia*, **5**, 131.

Allen, B.L., Fleming, P.J.S., Allen, L.R., Engeman, R.M., Ballard, G. & Leung, L.K.P. (2013). As clear as mud: A critical review of evidence for the ecological roles of Australian dingoes. *Biological Conservation*, **159**, 158–174.

Armsworth, P.R. & Roughgarden, J. (2001). An invitation to ecological economics. *Trends in Ecology & Evolution*, **16**, 229–234.

Baker, C.M., Bode, M. & McCarthy, M.A. (2016a). Models that predict ecosystem impacts of reintroductions should consider uncertainty and distinguish between direct and indirect effects. *Biological Conservation* **196**, 211–212.

Baker, C.M., Gordon, A. & Bode, M. (2016b). Ensemble ecosystem modeling for predicting ecosystem response to predator reintroduction. *Conservation Biology (Online Early, DOI: 10.1111/cobi.12798)*.

Beaumont, M.A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology*, **41**, 379–406.

Benshemesh, J. (2007). *National Recovery Plan for Malleefowl Leipoa ocellata*. Department of Environment and Heritage, South Australia, Canberra.

Benshemesh, J. & Bode, M. (2011). Adaptive management of Malleefowl. In: *Proceedings of the 4th National Malleefowl Conference, Renmark 2011,* pp. 126–135. SA Department of Environment and Natural Resources, Adelaide.

Benshemesh, J., Barker, R. & MacFarlane, R. (2007). *Trend analysis of malleefowl monitoring data*. National Malleefowl Monitoring, Population Assessment and Conservation Action Project, Bundoora.

Bode, M. & Brennan, K.E.C. (2011). Using population viability analysis to guide research and conservation actions for Australia's threatened malleefowl Leipoa ocellata. *Oryx*, **45**, 513–521.

Bode, M., Baker, C.M. & Plein, M. (2015). Eradicating down the food chain: optimal multispecies eradication schedules for a commonly encountered invaded island ecosystem. *Journal of Applied Ecology*, **52**, 571–579.

Buckley, Y.M. & Han, Y. (2014). Managing the side effects of invasion control. *Science*, **344**, 975–976.

Burgman, M.A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff, L. & Twardy, C. (2011). Expert Status and Performance. *PLoS ONE*, **6**, e22998.

Cadenhead, N., Kearney, M.R. & Moore, D. (2015). Climate and Fire Scenario Uncertainty Dominate the Evaluation of Options for Conserving the Great Desert Skink. *Conservation Biology*, **9**, 181–190.

Cooke, B. (2010). *Planning landscape-scale rabbit control*. Invasive Animals Cooperative Research Centre, Canberra.

Dambacher, J.M., Li, H.W. & Rossignol, P.A. (2003). Qualitative predictions in model ecosystems. *Ecological Modelling*, **161**, 79–93.

Dexter, N., Ramsey, D.S.L., MacGregor, C. & Lindenmayer, D. (2012). Predicting Ecosystem Wide Impacts of Wallaby Management Using a Fuzzy Cognitive Map. *Ecosystems*, **15**, 1363–1379.

Driscoll, D.A., Bode, M., Bradstock, R.A., Keith, D.A., Penman, T.D. & Price, O.F. (2016). Resolving future fire management conflicts using multicriteria decision making. *Conservation Biology*, **30**, 196–205.

Garnett, S.T. (2012). Unexpected outcomes of invasive predator control. *Animal Conservation*, **15**, 329–330.

Garrett, P. (1992). Correspondence. *Quarterly Essay*, **49**, 82–84.

Gårdmark, A., Lindegren, M., Neuenfeldt, S., Blenckner, T., Heikinheimo, O., Müller-Karulis, B., Niiranen, S., Tomczak, M.T., Aro, E., Wikström, A. & Möllmann, C. (2013). Biological ensemble modeling to evaluate potential futures of living marine resources. *Ecological Applications*, **23**, 742–754.

Grumbine, R.E. (1994). What is ecosystem management? *Conservation Biology*, **8**, 27–38.

Gustafsson, L. & Sternad, M. (2013). When can a deterministic model of a population system reveal what will happen on average? *Mathematical Biosciences,* **243**, 28-45.

Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R. & Sethna, J.P. (2007). Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLoS Computational Biology*, **3**, e189.

Hunter, D.O., Britz, T., Jones, M. & Letnic, M. (2015). Reintroduction of Tasmanian devils to mainland Australia can restore top-down control in ecosystems where dingoes have been extirpated. *Biological Conservation*, **191**, 428–435.

Keller, L. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, **17**, 230–241.

Kuhnert, P.M., Hayes, K., Martin, T.G. & McBride, M.F. (2009). Expert opinion in statistical models. In *Proceedings of the 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation* (eds R.S. Anderssen, R.D. Braddock & L.T.H. Newham), pp. 4262–4268.

Kuhnert, P.M., Martin, T.G. & Griffiths, S.P. (2010). A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters*, **13**, 900–914.

Lawton, J.H. (1999). Are There General Laws in Ecology? *Oikos*, **84**, 177.

Leith, C.E. (1974). Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, **102**, 409–418.

Levins, R. (1974). Discussion paper: the qualitative analysis of partially specified systems. *Annals of the New York Academy of Sciences*, **231**, 123–138.

Margules, C. & Pressey, R. (2000). Systematic conservation planning. *Nature*, **405**, 243–253.

Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low Choy, S., McBride, M. & Mengersen, K. (2012). Eliciting Expert Knowledge in Conservation Science. *Conservation Biology*, **26**, 29–38.

Melbourne-Thomas, J., Wotherspoon, S., Raymond, B. & Constable, A. (2012). Comprehensive evaluation of model uncertainty in qualitative network analyses. *Ecological Monographs*, **82**, 505–519.

Mellor, J.W. (1911). Mallee-fowl on Kangaroo Island. *Emu*, **11**, 35–37.

Parkes, J., Henzell, R. & Pickles, G.S. (1996). *Managing vertebrate pests: feral goats*. Canberra: Australian Government Publishing Service, 1996.

Pimm, S.L., Lawton, J.H. & Cohen, J.E. (1991). Food web patterns and their consequences. *Nature*, **350**, 669–674.

Raymond, B., McInnes, J., Dambacher, J.M., Way, S. & Bergstrom, D.M. (2010). Qualitative modelling of invasive species eradication on subantarctic Macquarie Island. *Journal of Applied Ecology*, **48**, 181–191.

Rowe, G. & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting,* **15**, 353-375.

Runge, M.C., Converse, S.J. & Lyons, J.E. (2011). Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. *Biological Conservation*, **144**, 1214–1223.

Saunders, G. & McLeod, L. (2007). *Improving fox management strategies in Australia*. Bureau of Rural Sciences, Adelaide.

Stocker, T.F. (2014). *Climate change 2013: the Physical Science Basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*.

Turchin, P. (2003). *Complex population dynamics: a theoretical/empirical synthesis*. Princeton University Press, Princeton, New Jersey.

Walsh, J.C., Wilson, K.A., Benshemesh, J. & Possingham, H.P. (2012). Unexpected outcomes of invasive predator control: the importance of evaluating conservation management actions. *Animal Conservation*, **15**, 319–328.

Wintle, B.C., Fidler, F., Vesk, P.A. & L Moore, J. (2013). Improving visual estimation through active feedback. *Methods in Ecology and Evolution*, **4**, 53–62.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the supporting information tab for this article:

**Supplementary Information 1:** Methods for eliciting interaction networks.

**Supplementary Information 2:** Methods for eliciting ecosystem dynamics.

**Supplementary Information 3:** Rescaling methods for Lotka-Volterra systems.

**Supplementary Table S1:** Participants and affiliations at the first workshop.

**Supplementary Table S2:** Participants and affiliations at the second workshop.

**Supplementary Figure S1:** Schematic overview of belief modelling process.

**Supplementary Figure S2:** Information contained in stated expert beliefs.

**Supplementary Figure S3:** Similarity of satisfactory ecosystem models.

**Supplementary Figure S4:** Example ecosystem scenario shown to workshop participants.

**Supplementary Figure S5:** Example ecosystem scenario with expert opinion superimposed.

**Supplementary Figure S6:** All ecosystem scenarios with all stated expert beliefs superimposed.

## FIGURE LEGENDS

**Figure 1:** Envelope method for eliciting expert beliefs and constraining the model ensemble. Grey shaded area indicates a 3 year period of overgrazing of mallee habitat, where native and introduced herbivore abundance was 200% of its long-term average. Experts were asked to draw envelopes that described their belief in the dynamics of the fox populations (y-axis) during this window. Two experts chose to answer this question; the coloured envelopes indicates their uncertain beliefs. The lines indicate the predictions of 4 viable models in the ensemble. One (green) is able to entirely replicate at least one expert belief; two (blue lines) are able to partly replicate the beliefs; one (black line) is unable to recreate them at all.

**Figure 2**: (a) Sign-structured interaction matrix elicited during the first workshop. Elements of the matrix indicate the qualitative direct impact that an increase in the component on the row would have on a component on the column. For example, an increase in rabbit abundance (row 11) will have a direct positive impact (+1) on fox abundance (column 5). +1 indicates a definite positive direct effect; -1 indicates a definite negative; 0 indicate a definite zero direct impact; +2 indicates either positive or zero; -2 indicates either negative or zero; 3 indicates either positive or negative. All diagonal values are negative to indicate density-dependence. (b) Graphical description of the interaction network shown above. This is the format in which

information on the structure of the interaction network was elicited from experts. Arrows indicate direct interactions with the sign indicated at the mid-point of each arrow.

**Figure 3.** Change in malleefowl abundance predicted by the unconstrained model ensemble ("unc": grey lines, with grey region enclosing 95% of the ensemble predictions) and expert belief-constrained ensemble ("con": blue lines and 95% region), following a perturbation made to another ecosystem component (black line). (a) Malleefowl abundance changes during a 25% increase in dingo abundance over 5 years. (b) Malleefowl abundance changes during a 25% increase in cat abundance over 5 years. (c) Malleefowl abundance changes during increased malleefowl immigration equal to 10% of the equilibrium population annually.

**Figure 4**: Predictions of the model ensemble when management interventions of reasonable intensity are applied to the 6 main threats in the malleefowl National Recovery Plan. Upper 6 plots show the relative change in malleefowl populations through time when each action is taken (shown in title), for a random sample of 200 models from the constrained ensemble. The lower plot synopsises the relative change of malleefowl population after 5 years (95% confidence intervals of final populations, with the mean shown by a circle).

**Supplementary Figure S1**: Schematic overview of how Ensemble Ecosystem Modelling (EEM) can be used to translate stated expert beliefs into revealed beliefs. The left-hand column describes the steps taken, the right hand column indicates where the information from each step is sourced from, the central column provides a diagrammatic flow-chart of the process.

Step 1 is to elicit a **qualitative ecosystem interaction network** from experts, which lists important ecosystem components and their direct cause-and-effect relationships. Step 2 is to transform this qualitative model into a large ensemble of **initial quantitative Lotka-Volterra (LV) models**, indicated by the black squares. Each model comprises a set of equations corresponding to the components of the interaction network, with the parameters chosen at random. Step 3 is to elicit a set of **stated belief envelopes** from experts, which describe their beliefs about how a particular ecosystem component will respond to a perturbation in a different ecosystem component. The degree of uncertainty surrounding this belief is indicated by the width of the envelopes. Step 4 is to use these stated belief envelopes to retain only those models in the ensemble that most closely recreate expert beliefs about system dynamics (green arrows), discarding any models which do not replicate these beliefs (red arrows). The envelopes therefore act as a filter or constraint on the model ensemble. The **remaining quantitative LV models** represent an ensemble of quantitative ecosystem models that concur with both sets of expert beliefs: the interaction structure of the qualitative networks, and the dynamic responses of the different envelopes. Step 5 is to interrogate the remaining ensemble of models, to produce implicit **revealed beliefs** about the ecosystem. These will include its response to perturbations, the relative priority of different management options, and the relative uncertainty surrounding different ecosystem processes and components.

**Supplementary Figure S2**: Information contained in the stated expert beliefs, expressed by the change in Shannon entropy $H$ of the interaction coefficients. To calculate $H$, we collect the values of each coefficient into 20 bins of width 0.05, where $n_i$ is the number of coefficients between $0.05(i-1)$ and $0.05i$. We then normalise these bins so that $p_i = n_i / \sum n_i$, and use Shannon's definition:

$$H = -\sum_{i=1}^{20} p_i \log_2 p_i.$$

The initial information associated with each coefficient (selected from a uniform distribution) is approximately $H = 4.32$. Once the expert beliefs are used to constrain the model ensemble, we are left with a subset of coefficient values with a lower or equal entropy. The decrease in entropy reflects the decrease in uncertainty caused by the constraints. Upper panel shows the proportional change in entropy for the distributions of each parameter in the interaction matrix (Figure 2a). A small proportion of coefficients have gained a large amount of information from the constraint process, while the majority have experienced a small improvement. The two lower panels detail two examples from the upper panel. Red bars indicate the original coefficient values in the unconstrained model ensemble (uniformly distributed); grey bars indicate the distribution in the constrained model. In the lower left hand panel (the direct effect of rainfall on fires), the constraints indicate that the coefficient is likely to be closer to one than zero. In the lower right hand panel (the direct effect of dingoes on kangaroo abundance), the model constraints offered little information, and the coefficient distribution remains relatively unchanged.

**Supplementary Figure S3**: (a–c) Scatter plots of interaction coefficient parameters for three randomly selected pairs of models that satisfy all stated expert beliefs. Blue markers represent the value of interaction parameter (e.g., $\alpha_{4,6}$ measures the per-capita / per-capita effect of dingoes on cats) in the two models. Red circles highlight structurally uncertain parameters. Note that the raw parameter correlation will appear artificially high because the signs of the different parameters are generally known. We therefore report the correlation of the absolute value of the parameters. (d) Scatter plot of the correlation results for 500 randomly selected model pairs. The black markers indicate Pearson's correlation coefficient (y-axis), and the significance of the correlation observed (x-axis). Green shaded area indicates the null expectation if these parameters were simply random numbers selected from a uniform distribution U(0,1). The two sets of correlation statistics are indistinguishable, indicating that the retained models in the ensemble are highly variable, and do not offer a single description of the ecosystem dynamics.

**Supplementary Figure S4:** An example ecosystem scenario, as used in the second workshop. Modelled dynamics correspond to a model based on Figure 2. Scenarios are defined by two factors: (1) the dynamics of a particular ecosystem species, shown by coloured lines. Each line offers an alternative response of the species to the perturbation. (2) The dynamics of a particular ecosystem driver, shown by coloured bars. In this case the driver is rainfall. In response to a perturbation in the driver, the abundance of the species may change. Experts were asked to choose between these alternatives, or to offer another option.

**Supplementary Figure S5:** As in *Supplementary Figure S4*, the bars indicate the change in an ecosystem driver. However in this figure the grey envelope indicates the opinions of an expert from the workshop.

**Supplementary Figure S6:** The aggregated results of the expert answers to each perturbation scenario (indicated in the title). Each expert has an assigned colour, and these are transparently overlaid on the timeseries to assess where experts agree and disagree on the likely consequences of an ecosystem perturbation. Note that since not all experts answered all questions, each scenario has a different number of responses.