

Where were they from? Modelling the source stock of dolphins stranded after the *Deepwater Horizon* oil spill using genetic and stable isotope data

L. Thomas*, C. G. Booth, P. E. Rosel, A. Hohn⁴, J. Litz, L. H. Schwacke

*Corresponding author: len.thomas@st-andrews.ac.uk

Endangered Species Research 33: 253–264 (2017)

Supplement

Table S1. Estimated probability of being from the coastal stock (p_{ij}) for stranded animals where more than one data source was available, and where the estimates differed by >0.2 . Animals are listed in order from largest absolute difference to smallest. Blank cells indicate cases where a data source was not available for an animal. (No animals with genetic-ID data are in the list; hence all rows are blank for that column.)

Animal ID	p_{i1} Genetic	p_{i2} SIR	p_{i3} Photo-ID	p_{i4} Genetic-ID
JSH-20130417-LA001	0.1232	0.9574		
RIB-20101217-LA001	0.0175	0.7483		
LFH-20120214-LA001	0.0000	0.7287		
JSH-20130429-LA001	0.0005	0.6889		
JSH-20130318-LA002	0.0412	0.6922		
RIB-20101130-LA001	0.0000	0.5598		
MCT-20110608-LA002	0.0045	0.4633		
JSH-20130517-LA003	0.0000	0.4461	0.0001	
BCF-20110822-LA001	0.0000	0.4457	0.0001	
MRB-20110517-LA001	0.0023	0.3619		
LFH-20120314-LA001	0.3107	0.1832	0.0001	
RAH-20130214-LA001	0.2713	0.0264		

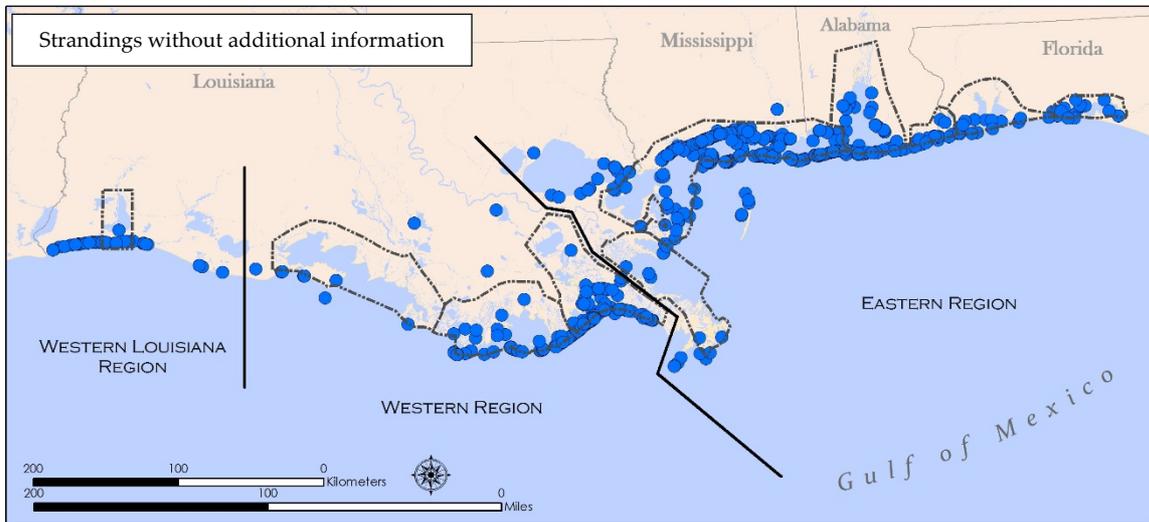
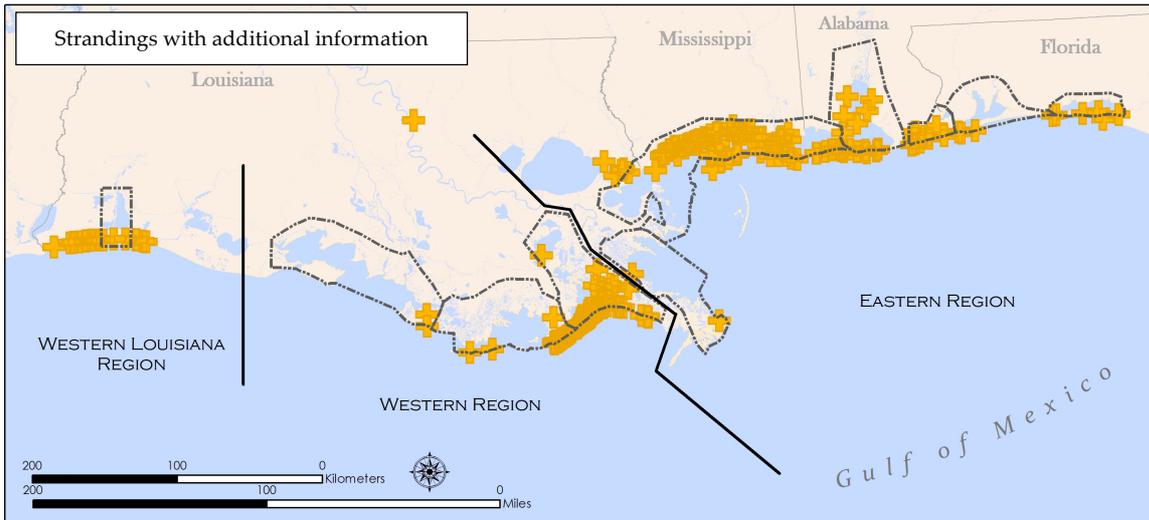


Fig. S1 - Maps showing reported locations of strandings for animals with (yellow crosses, top map) and without (blue dots, bottom map) additional information about stock of origin (i.e., either genetic assignment, stable isotopes, photo-ID or genetic ID). Analysis regions are shown, together with total strandings per region and approximate region boundaries (thick black lines). Approximate bay sound and estuary (BSE) stock boundaries are shown as dot-dashed lines, with the location of the Barataria Bay and Mississippi Sound stocks indicated.

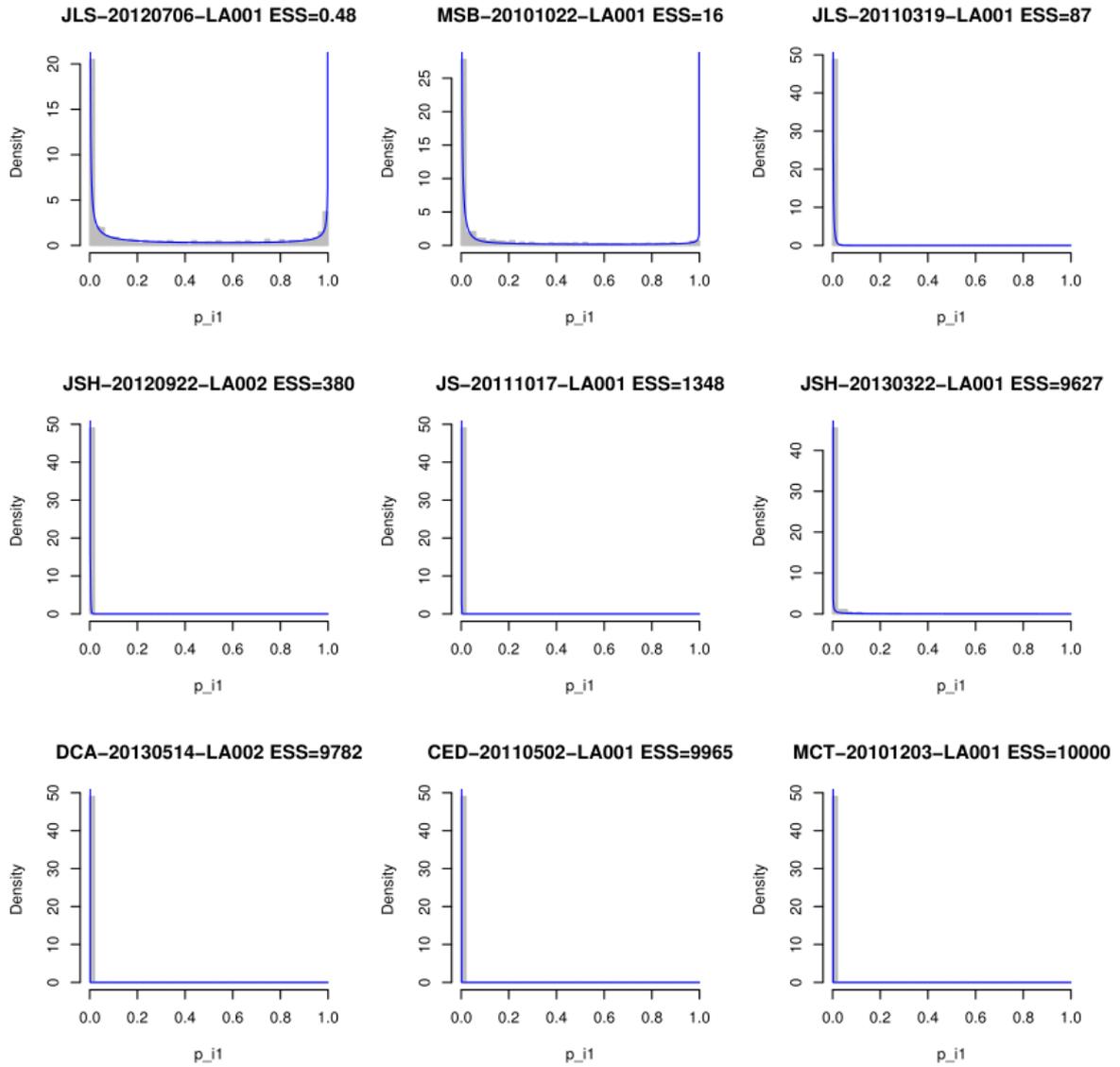


Fig. S2 - Histograms showing bootstrap estimates of probability of originating from the coastal stock for stranded animals estimated from genetic assignment data, p_{i1} , with fitted mixture of beta distributions superimposed, for nine sample animals selected systematically to represent lowest to highest effective sample size (ESS). The title in each plot gives the animal ID and ESS.

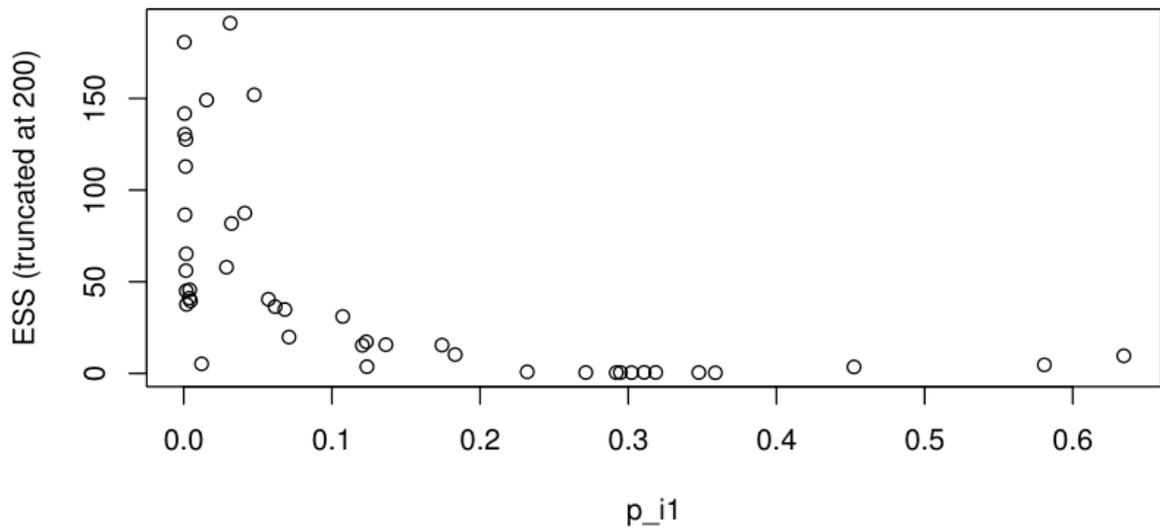
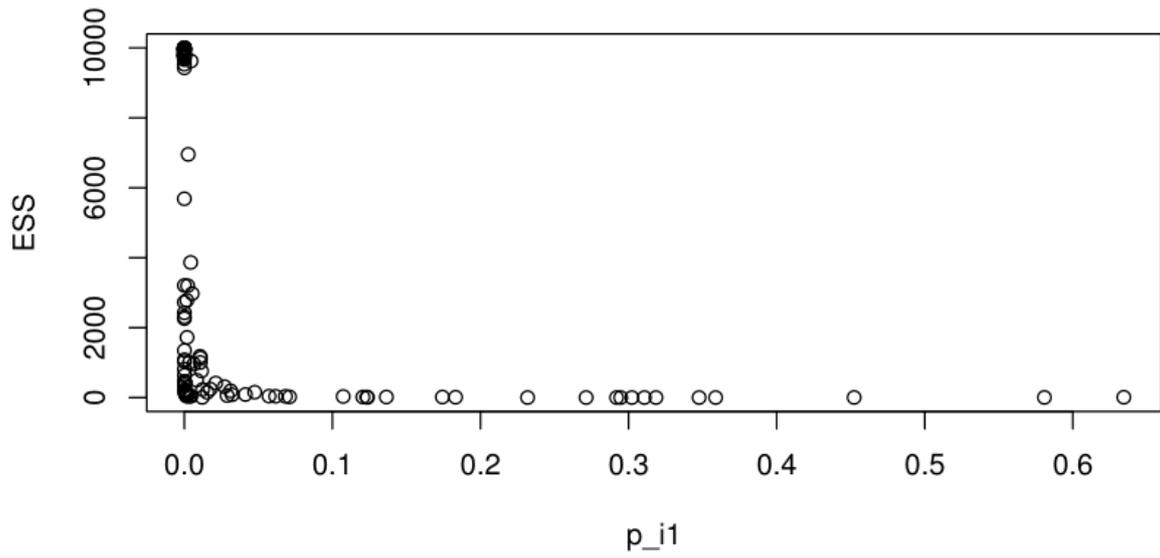


Fig. S3 - Estimated probability of being from the coastal stock for stranded animals from genetic assignment data, p_{i1} , vs ESS. Top plot is all animals; bottom plot is truncated on the y-axis at 200.

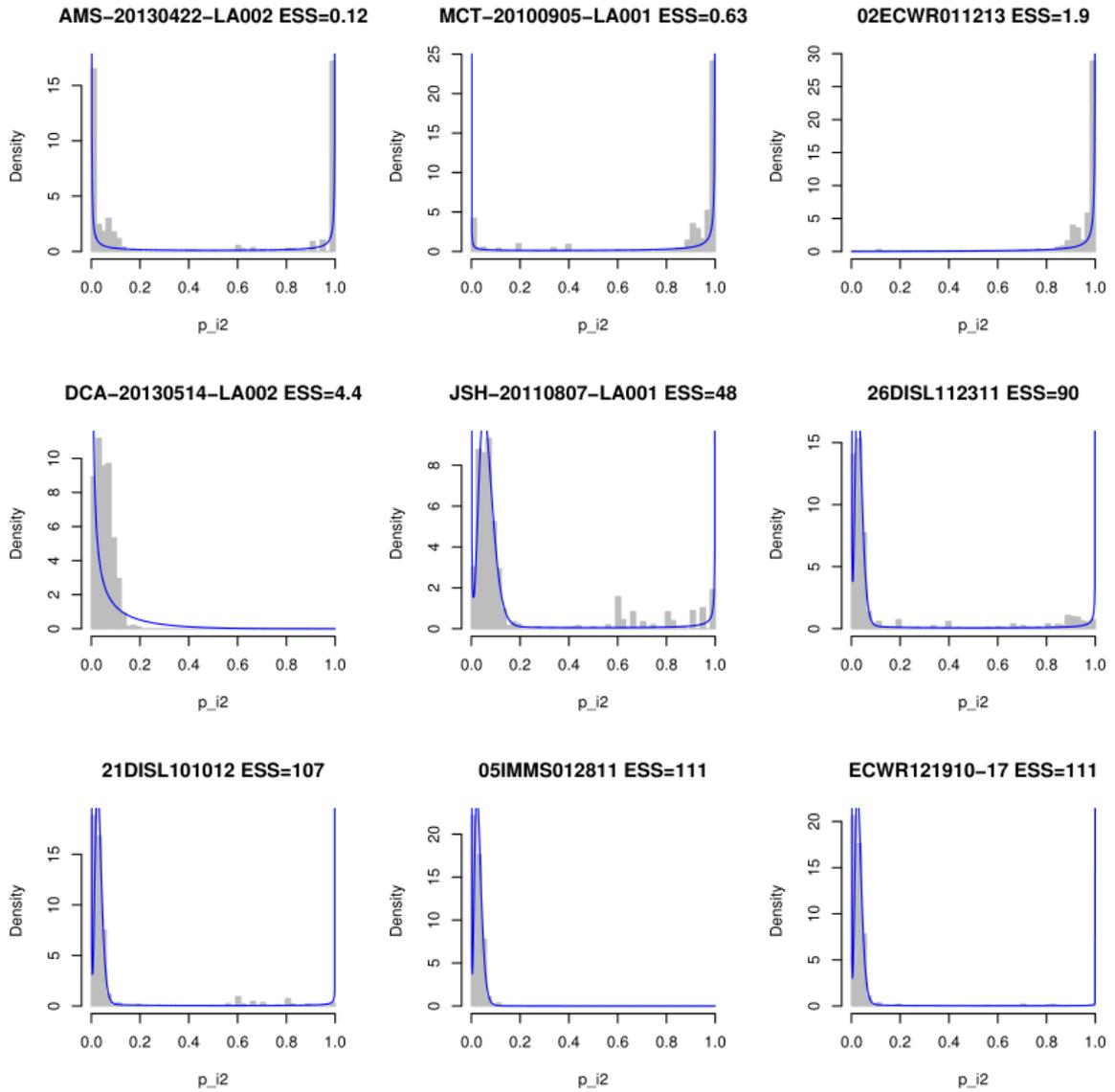


Fig. S4 - Histograms showing bootstrap estimates of probability of originating from the coastal stock for stranded animals estimated from stable isotope ratio (SIR) data, p_{i2} , with fitted mixture of beta distributions superimposed, for nine sample animals selected systematically to represent lowest to highest effective sample size (ESS). The title in each plot gives the animal ID and ESS.

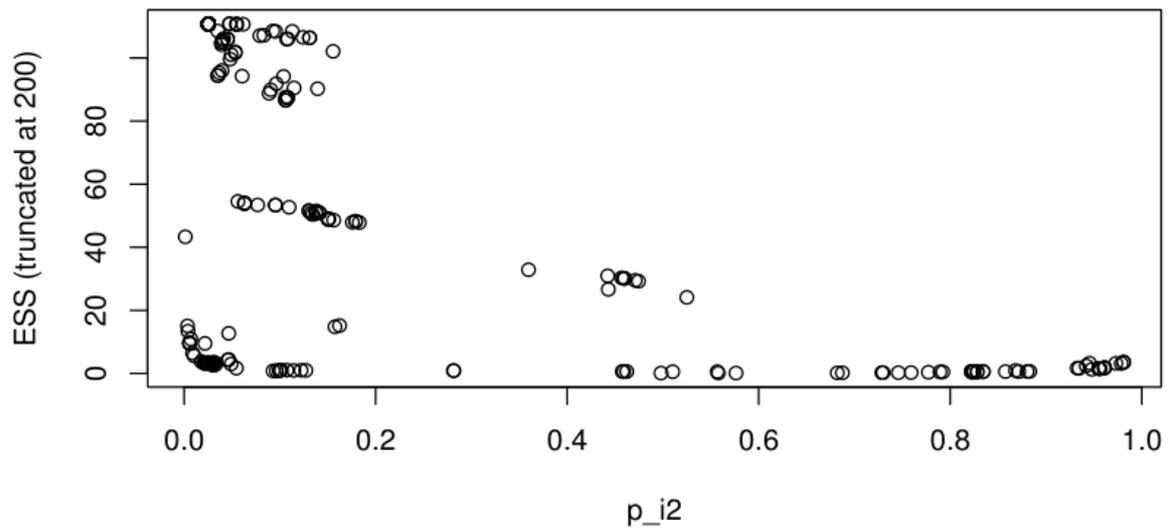
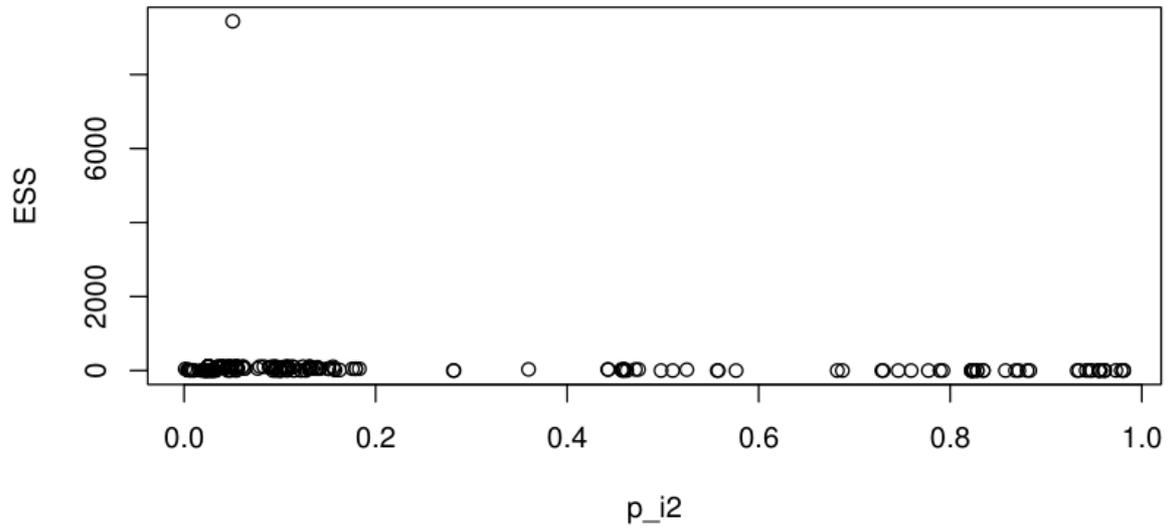


Fig. S5 - Estimated probability of being from the coastal stock for stranded animals from SIR data, p_{i2} , vs ESS. Top plot is all animals; bottom plot is truncated on the y-axis at 200.

Supplemental methods: alternative treatments of strandings with no additional information

There are several possibilities for estimating n_b , the number of coastal stranded animals without additional information, all of which use the basic assumption that the animals with no additional information and those with additional information are exchangeable; in other words that the probability distribution of being from coastal stock is the same for both groups.

1. The simplest assumption along these lines is to assume the number of coastal animals out of those with no additional information follows a binomial distribution, i.e.,

$$n_b \sim \text{binomial}(N_b, P)$$

where

$$P = E_p(p)$$

There are (at least) two ways to estimate P :

- a. As the mean of the \hat{p}_i s:

$$\hat{P}^{*1} = \frac{1}{N_a} \sum_{i=1}^{N_a} \hat{p}_i$$

- b. As the mean of the \hat{x}_i s:

$$\hat{P}^{*2} = \frac{1}{N_a} \sum_{i=1}^{N_a} \hat{x}_i = \frac{\hat{n}_a}{N_a}$$

A disadvantage of these approaches is that they may lead to an under-estimate of uncertainty, because they assume the same P applies to all animals in N_b . The second approach will have slightly more uncertainty, because it includes uncertainty due to the Bernoulli trials that create x_i from p_i .

2. A more general assumption would be that n_b follows a beta-binomial distribution, with the mean and variance of the beta component coming from the mean and variance of the p_i s. However, there is no guarantee that the beta-binomial accurately models the distribution of the p_i s. This would need checking on the dataset being used.

3. An even more general assumption is that of a mixture distribution on the stock probability for animals in N_b , where the mixing is over the animals for which there is additional information, and each of these animals is included with equal probability. That is

$$f(p_k; \alpha, \beta) = \frac{1}{N_a} \sum_{i=1}^{N_a} f(p_i; \alpha_i, \beta_i) \quad (1)$$

where k indexes the animals with no additional information, hence $k = 1, \dots, N_b$.

In practice, we implement this mixture by augmenting the model with a latent allocation vector d such that each d_i takes a value from $1, \dots, N_a$, indicating which of the N_a animals' p to use in the mixture. To create the mixture given in Eqn 1 we set

$$d_j \sim \text{categorical}(v_1, \dots, v_{N_a})$$

where all v take the value $1/N_a$, and then

$$p_j | (d_j = j) = p_{d_j}$$

Finally, let y_i be the true stock of the j th animal. We assume

$$y_j \sim \text{Bernoulli}(p_j)$$

and

$$n_b = \sum_{j=1}^{N_b} y_j$$

This option will take longer to run in JAGs as one has to track each of the N_b animals with no additional info; in the previous options, there is a single distribution to generate n_b . However, it allows each of the N_b animals to have a different p , and also does not make any assumption about the distribution of these p s except that the distribution is the same as that for the N_a animals. This was, therefore, the option that was implemented in the work reported here.

JAGS code for hierarchical model

The following code implements the “Extended model” described in the methods and shown as a DAG in Figure 3.

```

model{
  N<-N.a+N.b

  #model for n.a
  for(i in 1:N.a){
    for(j in 1:M){
      #Choose which component of the mixture is used
      e[i,j]~dcat(rho[i,j,])
      #Use this component to work out the pi
      pi.unnorm[i,j]<-(alpha[i,j,e[i,j]]+beta[i,j,e[i,j]])
      pi[i,j]<-pi.unnorm[i,j]/pi.norm.const[i]
    }
    #Normalizing constant to ensure pi[i,j] sum to 1 over js
    pi.norm.const[i]<-sum(pi.unnorm[i,])
    c[i]~dcat(pi[i,])
    #Use the correct beta distribution, depending on the
    # mixture component chosen
    p.i[i]~dbeta(alpha[i,c[i],e[i,c[i]]],beta[i,c[i],e[i,c[i]]])
    x[i]~dbern(p.i[i])
  }
  n.a<-sum(x)

  #Model 3 for n.b
  for(k in 1:N.b){
    d[k]~dcat(nu[])
    p.k[k]<-p.i[d[k]]
    y[k]~dbern(p.k[k])
  }
  n.b<-sum(y)

  #Other derived quantities of interest
  n<-n.a+n.b
  p<-n/N
}

```