L. Thomas[1], J. Harwood[1,2]
Estimating Grey Seal Population Size using a Bayesian State-Space Model
[1] Centre for Research into Ecological and Environmental Modelling, The Observatory, University of St Andrews, St Andrews KY16 9LZ
[2] NERC Sea Mammal Research Unit, Gatty Marine Laboratory, University of St Andrews, St Andrews KY16 8LB

NOTE: THIS PAPER AND ITS CONTENTS SHOULD NOT BE REFERENCED WITHOUT PRIOIR PERMISSION OF THE AUTHORS

Summary
We develop a spatially-explicit, stochastic model of the British grey seal population. We fit the model using a computer-intensive Bayesian technique, using the pup production estimates for each region from 1984-2002 as input data. We give estimates of population parameters, historical population sizes and projections of future populations.

By fitting simpler models, various hypotheses about population processes can be tested. We find evidence for both density dependent juvenile survival and density dependent movement at the regional level.

Introduction
A principal aim of collecting annual pup production data is as a means to monitor total population size of seals. To do this, we must make inferences about the adult population based on the pup data, and this requires a model for the population dynamics. Here, we develop such a model, based on the state-space modelling framework proposed by Buckland *et al.* (in press). Recent advances in computer-intensive statistical inference allow us to fit this kind of stochastic, nonlinear model to the pup production data.

Our model allows for two density-dependent processes: survival of pups and recruitment of adult females to different breeding sites. Although we believe that these processes probably operate at the level of individual colonies, technical considerations have – so far – prevented us from fitting the model at that level. Instead, we have aggregated colony data into four regions: North Sea, Inner Hebrides, Outer Hebrides and Orkney.

We can use the modelling framework to test for evidence of the density dependent processes at the level of the region. To do this, we fit models with the density dependent processes taken out, and compare their fit with that of the full model.

Material and Methods
Here we briefly summarize the model and fitting methods. A more detailed description is given in Thomas *et al.* (in press), although both the movement model and the fitting algorithm have been further refined since that paper was written.

Full Model
A state space model has two components: (1) the state process, which models the true but unknown state of the population; and (2) the observation process, which models how the survey data are generated given the true states.

In constructing the state process, we divide the seal population in each region into 7 age classes: pups (age 0), age 1 – age 5 adult females (pre-breeding), and age 6 and older females. Note that our model does not include adult males.

The time step for the process model is 1 year, beginning just after the breeding season. The model is mode up of four sub-processes: survival, age incrementation, movement of recruiting females and breeding.

Survival is modelled as a binomial random process. Following Harwood (1981), we assume that pup survival is density dependent, and follows a Beverton-Holt function of the form:

$$ f_{p,r,t} = \frac{f_{p\max}}{1 + b_r n_{0,r,t-1}} $$

where $n_{0,r,t-1}$ is the number of pups born in region r in year t-1, $f_{p,r,t}$ is survival rate of these pups, $f_{p\max}$ is maximum pup survival rate, and $1/b_r$ reflects the carrying capacity of the region. Since half of the pups born will be male, the expected number of female pups surviving will be $0.5 f_{p,r,t} n_{0,r,t-1}$. We assume that adult female survival rate, f_a is constant across regions and time.

Age incrementation is deterministic – all seals age by one year (although those in the age 6+ category remain there).

To model movement, we assume that only females breeding for the first time may move from their natal region. Once a female has started breeding she remains faithful to that region. We assume that movement is fitness dependent (Ruxton and Rohani 1998), such that females will only move if the expected survival of their future offspring is higher elsewhere, and the probability of movement is proportional to the expected survival difference. In addition, we assume that females are more likely to move among regions that are close together, and that females show some degree of site fidelity – that is, they may not move even if conditions for their offspring will be better elsewhere. We model movement from each region as a multinomial random variable where probability of movement from region r to region i at time t is:

$$r_{r \rightarrow i,t} = \begin{cases} \dfrac{q_{r \rightarrow i,t}}{\sum_{j=1}^{4} q_{j \rightarrow i,t}} & : \sum_{j=1}^{4} q_{j \rightarrow i,t} > 0 \\[2ex] I_{i=r} & : \sum_{j=1}^{4} q_{j \rightarrow i,t} = 0 \end{cases}$$

where $I_{i=r}$ is an indicator that is 1 when $i=r$ and 0 otherwise, and

$$q_{r \rightarrow i,t} = \begin{cases} g_{sf} & : i = r \\[1ex] \dfrac{g_{dd} \max\left(\left[f_{p,r,t} - f_{p,i,t}\right],0\right)}{\exp\left(g_{dist} d_{r,i}\right)} & : i \neq r \end{cases}$$

where g_{sf}, g_{dd}, and g_{dist} are three movement parameters that index the strength of the site fidelity, density dependence and distance effects respectively, and $d_{r,i}$ is the 20% trimmed mean of the distances between colonies in regions r and those in region i (standardized so that the largest distance is 1.0).

We model breeding by assuming that the number of pups produced is a density independent function of the number of breeding females in the region, with binomial probability a.

For the observation process, we assume that pup production estimates follow a normal distribution with a constant coefficient of variation (CV). This CV is a model parameter, y, which is be estimated from the data.

In summary, the full model contains 11 parameters: adult survival f_a, maximum pup survival, $f_{p\max}$, one carrying capacity parameter

for each region b_1 - b_4, three movement parameters g_{sf}, g_{dd}, and g_{dist}, fecundity a and observation CV y.

Reduced Models

We fitted two reduced models to the data. In the first, we assumed no movement between regions. This removed the movement parameters, leaving 8 parameters in the model. In the second we assumed no movement and density independent juvenile survival. This removed both the movement and carrying capacity parameters, leaving 4 parameters in the model.

In the results, we denote the full model M_{full}, the first reduced model M_1 and the second M_2.

Fitting method

We used an implementation of a class of computer-intensive Bayesian techniques called sequential importance sampling (SIS) to fit the models. This technique is well suited to the analysis of time series data, as each data point is introduced one at a time into the algorithm, making it potentially much more efficient than other computer-intensive techniques such as Markov chain Monte Carlo (MCMC).

The basic idea is as follows. We start by defining prior distributions on the parameters and the states (i.e., the numbers of seals in each region and age class before the first time period). We simulate a large number of parameter and state vectors from these priors. Each pair of parameter and state vectors is called a 'particle'. We stochastically project each particle forward to the first time period using the state process (i.e., our model of the population dynamics), and calculate the likelihood of the simulated pup production generated for each particle, given the observed pup production in the first year and the observation model. These likelihoods form weights for each particle. We next resample according to the likelihood weights, so that particles with high likelihoods tend to be retained and those with low likelihood tend to be discarded. We then project forward to the second time period, calculate the likelihood of each particle given the second years' pup production estimates, resample, and so on to the end of the data. The distribution of particles at the end gives an estimate of the posterior of the parameters and states.

The basic method outlined above does not work well in practice because with a long enough dataset, we end up discarding almost all the

particles: a problem known as 'particle depletion'. A large number of strategies are available to combat this problem (see Lui 2001 and papers in Doucet *et al.* 2001). We used four strategies here: kernel smoothing of the parameter vectors, auxiliary particle filtering, residual sampling, and partial rejection control. The first three are described briefly in the context of the seal model by Thomas *et al.* (in press); the last is described in Lui (2001). For the record: for all runs shown here we used 400,000 particles, a kernel smoothing discount factor of 0.9 and a partial rejection criterion of the 99th percentile of the distribution of weights. The prior distributions for each parameter are given in Table 1, and are shown on Figure 4. We fit the models using pup production data for the four regions from 1984 to 2002. The first year of data was used to provide priors for the states (see Thomas *et al.* in press for details), and the SIS algorithm ran from 1985 to 2002.

Table 1. Prior parameter distributions

Param	Prior	Expected value
f_a	Beta(22.05,1.15)	0.95
$f_{p\max}$	Beta(14.53,6.23) Beta(5.45,3.38)[1]	0.7 0.62[1]
b_1	Gamma(4,2.07x10^{-4})	8.29x10^{-4}
b_2	Gamma(4, 2.96x10^{-4})	1.18x10^{-3}
b_3	Gamma(4,7.40x10^{-5})	2.96x10^{-4}
b_4	Gamma(4,5.76x10^{-5})	2.30x10^{-4}
g_{sf}	Gamma(2.25,1.33)	3
g_{dd}	Gamma(2.25,0.49)	ln(3)
g_{dist}	Gamma(2.25,0.22)	0.5
a	Beta(22.05,1.15)	0.95
y	Gamma(4,0.025)	0.10

[1]This prior on juvenile survival was used in model M_2, and comes from Hall *et al.* (2001).

Model comparison
For each model, we calculated the likelihood of the posterior pup production in all years, given the priors and all the data, averaged over all particles. We expect models with more parameters to fit better *a priori*, so we also computed Akaike's Information Criterion (AIC), a form of penalized likelihood, which adds a penalty proportional to the number of model parameters (Burnham and Anderson 1998). Models were compared using Akaike weights

(Burnham and Anderson 1998, p124), which can be thought of in the Bayesian context as the posterior probability of each model being the best approximating model (Akiake 1981).

Predicting future abundances
By projecting the models forward from the posterior distribution of states and parameters it is possible to generate predictions of future numbers of pups and adult females. The models do not include adult males, so it is not possible to say anything about their numbers without making additional assumptions. Here, we followed Hiby and Duck (unpublished) in assuming that the number of adult males is 73% of the number of adult females.

Results
Posterior estimates of true pup production for the three models are shown in Figures 1, 2 and 3 (these estimates are known technically as smoothed estimates; see Thomas *et al.* in press). Both the full model (M_{full}) and the no-movement model (M_1) provide reasonable representations of the observed changes in pup production. M_{full} is slightly better at capturing the very rapid levelling off of pup production in the Inner and Outer Hebrides in the mid-1990s, but neither M_{full} nor M_l completely capture this feature of the data. Without density dependence (model M_2) the fit to the Hebridean post-1995 data is very poor. This is supported by AIC statistics (Table 2), which show that model M_2 has very little support relative to M_{full} and M_1. In addition, M_{full} is favoured over M_l, providing evidence that density-dependent movement plays a role in regional population dynamics.

Table 2. Mean posterior log-likelihood, AIC and Akaike weights

Model	LnL	AIC	Akaike weight
M_{full}	-527.6	1077.4	0.915
M_1	-533.1	1082.1	0.085
M_2	-576.8	1161.6	0.000

Focussing on the full model alone, we give posterior parameter estimates in Figure 4 and an estimate of the 2002 pre-breeding population size in Table 3. Projecting the estimates from this model forward, we predict that abundance will continue to rise over the next 10 years (Figures 5 and 6, Table 3), but at a decreasing rate (2.4% in 2003-4, down to 1.8% in 2012-13). There is a moderate level of uncertainty about

these predictions (95% posterior confidence intervals for the 2003 growth rate are 1.0-3.7% and for 2012 are 0.8-2.5%) and the results are also conditional on the model being correct and on the prior parameter distributions. The increases are predicted to be less in the Inner and Outer Hebrides than in the North Sea and Orkney. Indeed, in the Inner and Outer Hebrides there is a small chance of the population declining slightly (posterior probabilities of a decline are 8.1% and 3.1% respectively).

Table 3. Estimated population size, in hundreds, of British grey seals from model M_{full} at the start of the 2002 breeding season and predicted population size before breeding in 2012. Numbers are posterior means with 95% confidence limits in brackets.

Region	2002	2012
North sea	10.2 (9.2-11.5)	11.3 (9.0-14.1)
Inner Hebrides	7.5 (6.9-8.5)	8.1 (6.5-10.1)
Outer Hebrides	30.4 (27.4-34.3)	33.2 (26.5-41.5)
Orkney	44.3 (40.1-50.4)	52.8 (42.7-64.8)
Total	92.5 (83.5-104.8)	105.3 (84.6-130.6)

Discussion

The full model is heavily favoured by the AIC statistic, but there is some Monte-Carlo variation in our results – repeat runs of the models produces slight differences in outputs. We therefore regard our conclusions as tentative at this stage, pending further methodological investigation and tuning of the SIS algorithm.

Our results are also conditional on the prior distributions used for the parameters. The posterior estimates for the movement parameters are similar to the priors (Figure 4), implying that there is relatively little information in the data to inform them. We anticipate that colony-level data will provide more information about movement, and plan on running an analysis at this level in the next year. A NERC/EPSRC funded PhD student is working on providing independent estimates of migration rates using genetic and mark-recapture information.

Although the current movement model encapsulates the main features we believe influence the dispersal behaviour of breeding females, it is over-simplistic in only allowing recruiting females to choose where to breed

based on the previous year's pup production. This produces over-compensatory behaviour, so we are working on an improved movement model that operates in continuous time (Matthiopoulos 2003).

An initial sensitivity analysis shows that the fit of M_{full} is not sensitive to reasonable variations in the assumed prior distribution of $\boldsymbol{f}_a$, $\boldsymbol{f}_{p\max}$ or $\boldsymbol{a}$. However the estimates and predictions of total population size (Table 3, Figures 5 and 6) are sensitive to these assumptions. One way to combat this is to expand the observation model to include independent estimates of adult survival and age structure (Harwood and Prime 1978) and pregnancy rates (Boyd 1985) from culled seals, and of adult survival (Pomeroy, unpublished) and juvenile survival (Hall et al. 2001) from tagged seals. An assumption must be made that these data are representative of the population they are applied to, and this is not necessarily valid as seal data is rarely collected from a random sample.

Hiby and Duck (unpublished) give an estimate of the total pre-breeding population size for 1999 of 109,000 with 95% CI 93,000-126,000. For comparison, our estimate for the same year (to the nearest thousand) is 85,000 with 95% CI 79,000-96,000 – about 20% lower. The main difference between Hiby and Duck's model and the one used there is that their model is density independent while ours includes density dependent pup survival (ours also includes movement, but that does not change the total population figures appreciably). Our estimates of pup survival are considerably lower (mean posterior estimates for 1999 in the 4 regions are 0.17, 0.15, 0.16 and 0.19, compared with Hiby and Duck's overall estimate of 0.39), which results in fewer non-breeding adult females. Hiby and Duck considered a model where pup survival showed a linear decline since 1989, but found insufficient evidence to support it in the data. They point out that such a model leads to significantly fewer females (their Figure 2). Hall et al. (2001) estimated first year female pup survival at the Isle of May (a North Sea colony) to be 0.617 (SE 0.155). This data could be incorporated into our observation model, but more information about juvenile survival in other regions will be needed to improve the overall reliability of the model predictions.

Other planned improvements include: addition of a random effect for fecundity, allowing it to vary each year around a pre-defined mean (c.f. unexplained dip in pup production in all regions

in 1999 in Figures 1-3); possible inclusion of covariates such as the North Atlantic Oscillation; and a more flexible form for density dependent survival that would allow survival to remain high until colonies become very crowded.

One strength of the Bayesian fitting method is that we can make projections of future population sizes that incorporate both uncertainty in parameter values and also uncertainty about which model is correct. If, for example, two models have posterior Akiake weights of 0.5 each, then we can project forward using equal numbers of particles from the two models. In the current case, with M_{full} getting a 91.5% posterior weight, this was not thought necessary – but this could change in the future as more models are considered.

The SIS algorithm is a relatively recent tool for fitting complex biological models to different sources of data, and there are a number of research avenues to be followed there. The model M_2 is linear and approximately normal, so could be fit using the well-established Kalman filter method, building on recent work by Besbeas et al (2002). The more complex models could be compared with the outputs from a MCMC analysis, although an efficient way to perform such an analysis would have to be devised. We also plan to investigate other strategies to avoid particle depletion in SIS, such as simulated tempering and multiple time-step look-ahead, and to further explore the optimum level of kernel smoothing.

References

Akiake, H. 1981. Likelihood of a model and information criteria. *Journal of Econometrics* 16:3-14.

Besbeas, P., J. Harwood, B.J.T. Morgan and P. Pomeroy. 2002. Integrated analysis of grey seal data. Report UKC/IMS/02/36, IMS, University of Kent, UK.

Boyd, I.L. 1985. Pregnancy and ovulation rates in grey seals *Halichoerus grypus* on the British coast. *Journal of Zoology* 205: 265-272.

Buckland, S.T., K.B. Newman, L. Thomas, and N. Kösters. In press. State-space models for the dynamics of wild animal populations. *Ecological Modelling.*

Burnham, K.P. and D.R Anderson. 1998. *Model Selection and Inference: A Practical Information-theoretic Approach.* Springer-Verlag, New York.

Doucet, A., N. de Freitas and N. Gordon [eds.]. 2001. *Sequential Monte Carlo Methods in Practice.* Springer-Verlag, New York.

Hall, A.J., B.J. McConnell and R.J. Barker. 2001. Factors affecting first-year survival in grey seals and their implications for life history strategy. *Journal of Animal Ecology* 70: 138-149.

Harwood, J. 1981. Managing grey seal populations for optimum stability. Pp 159-172 in: C.W. Fowler and T.D. Smith. (Eds.) *Dynamics of Large Mammal Populations.* Joun Wiley & Sons, New York.

Harwood, J. and J.H. Prime. 1978. Some factors affecting the size of British grey seal populations. *Journal of Applied Ecology* 15: 401-411.

Hiby, L. and C.D. Duck. Unpublished. Estimates of the size of the British grey seal *Halichoerus grypus* population and levels of uncertainty.

Lui, J.S. 2001. *Monte Carlo Strategies in Scientific Computing.* Springer-Verlag, New York.

Matthiopoulos, J. 2003. Modelling the between-year movement of grey seals and their population consequences. SCOS Briefing Paper.

Ruxton, G.D. and P. Rohani. 1998. Fitness-dependent dispersal in metapopulations and its consequences for persistence and synchrony. *Journal of Animal Ecology* 67: 530-539.

Thomas, L., S.T. Buckland, K.B. Newman, and J. Harwood. In press. A unified framework for modelling wildlife population dynamics. *Australian and New Zealand Journal of Statistics.*

Figure 1. Estimates of pup production from the full model, M_{full}. Input data are shown as circles, and the lines represent the mean of the particle values bracketed by 2.5th and 97.5th percentiles, which can be thought of as posterior 95% confidence intervals..
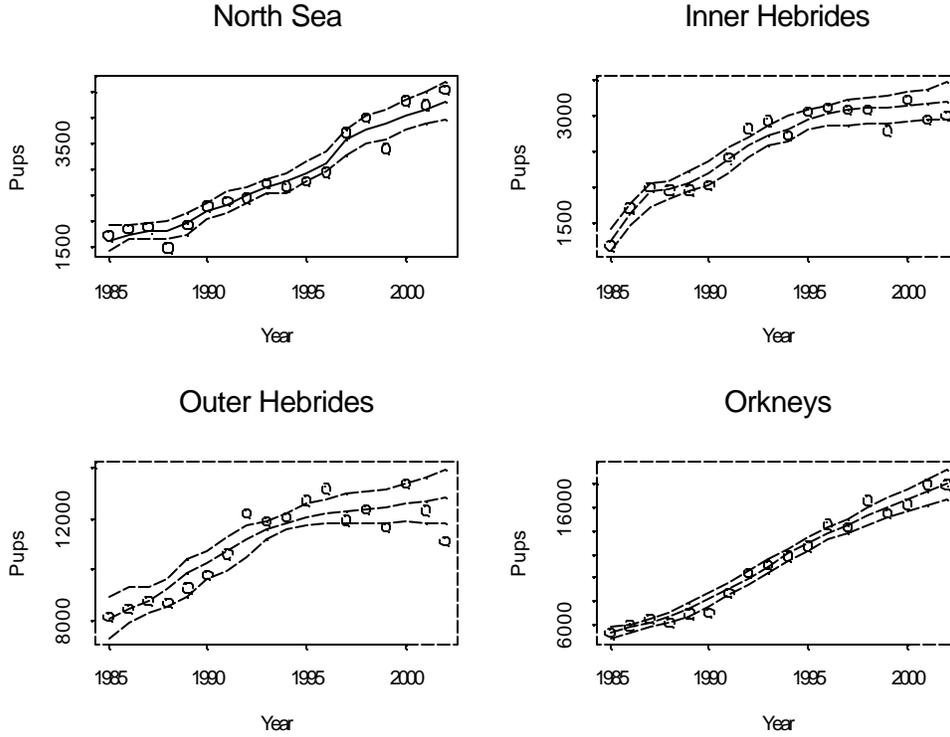

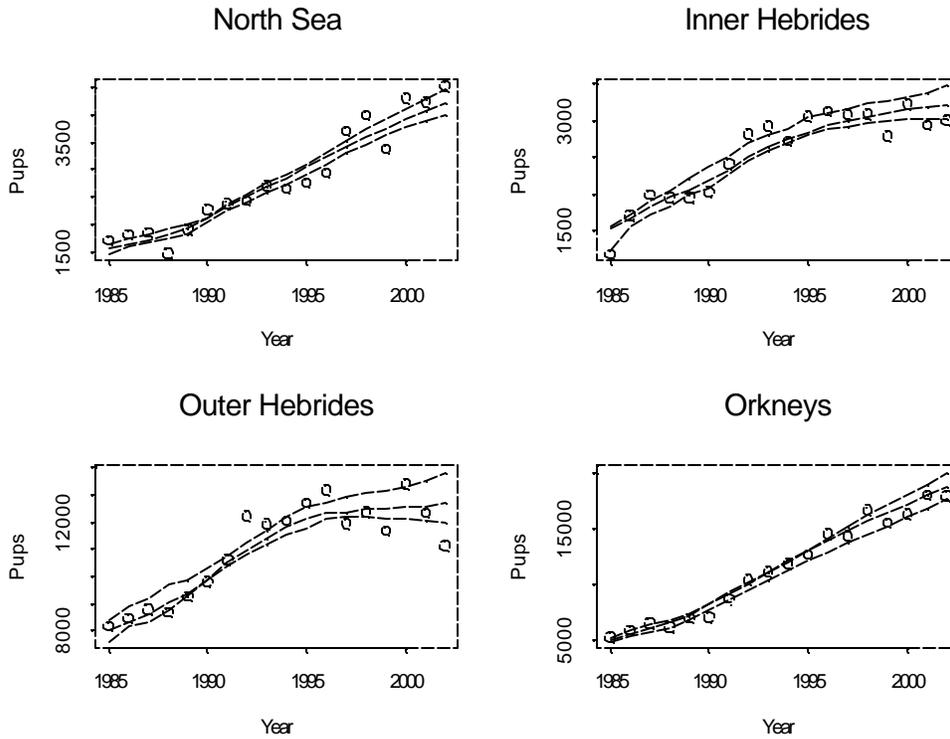*Figure 2. Estimates of pup production from the no-movement model, M_1.*

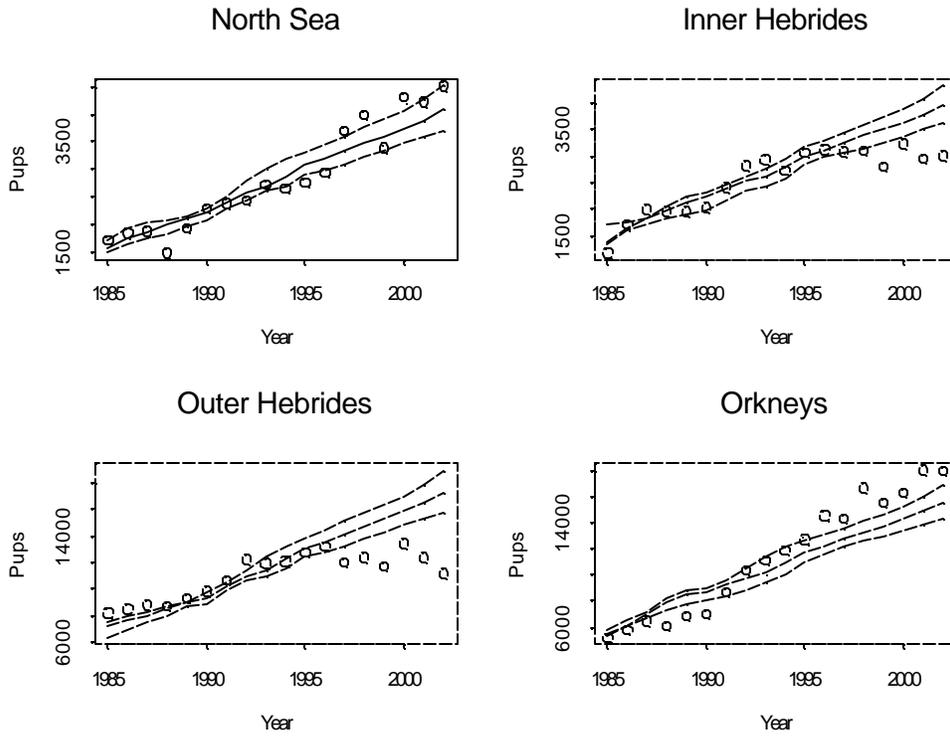*Figure 3. Estimates of pup production from the no-movement and no density dependence model, M_2.*

Figure 4. Posterior parameter estimates (histograms) and priors (solid lines) from M_{full}. Vertical line shows posterior mean, and this value is given in the title of each plot after the parameter name.
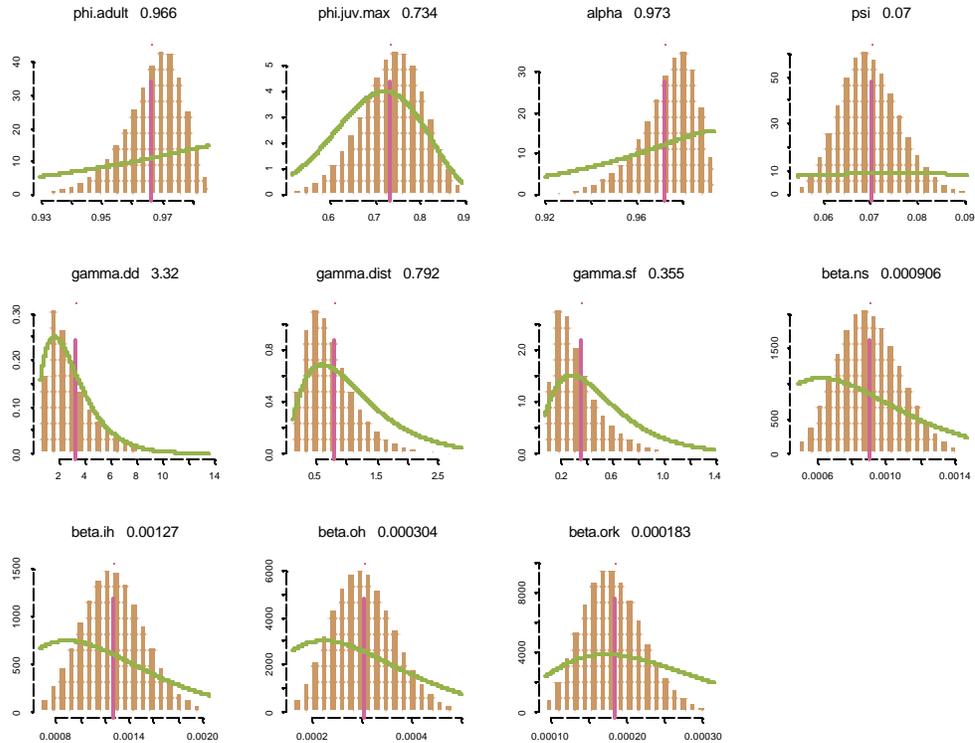
Figure 5. Predicted regional pre-breeding population size of British grey seals 2003-2012 from model M_{full}.
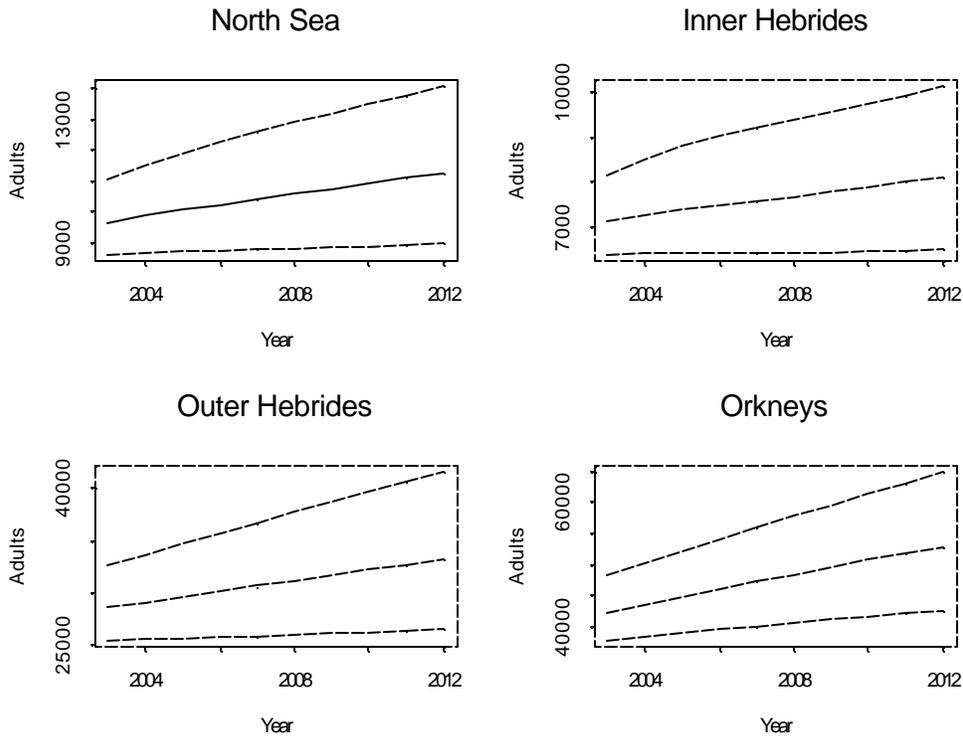

North Sea · Inner Hebrides · Outer Hebrides · Orkneys

Figure 6. Predicted total pre-breeding population size of British grey seals 2003-2012 from model M_{full}.