



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

**A Bayesian State Space Approach to
Estimating Pest Management Models**

by

Russell Gorddard

Agricultural and Resource Economics
Faculty of Agriculture
University of Western Australia
Nedlands, WA 6907
gorddard@uniwa.uwa.edu.au

February 1995

This work was partially funded by the Grains research and Development Council. Presented at the 39th annual conference of the Australian Agricultural Economics Society, Perth 1995. Preliminary draft only. I would like to thank David Dole and Catherine Forbes for helpful comments.

Estimating a model for pest management

Abstract

Models for pest management need to incorporate information from many subject areas, including biology, agronomy and economics. The information sources are unrelated except by the structural assumptions of the bio-economic model that is used. Rigorous model estimation is often at the expense of ignoring much of the available information. This paper illustrates how a Bayesian approach using the Gibbs sampler can allow the incorporation of potentially valuable prior information. This approach provides posterior distributions for the important model relationships, giving an indication of confidence in the model and providing a natural starting point for stochastic decision making.

Estimating a model for pest management

1 Introduction

This paper uses a Monte Carlo Markov Chain method (MCMC) to estimate a nonlinear Bayesian state space model of weed populations in a cropping system. The system, like many natural resource systems, has the following features which make estimation difficult:

i. Weed population densities are not easily observed. There are several different sources of information on population available, all of them imperfect. Combining them requires updating of parameter estimates.

ii. Data is sparse, and what there is comes from many unrelated experiments of different design done under different conditions and measuring different variables. This poses a significant missing data problem.

iii. Model structure is complex. Many functions are nonlinear and stochastic, the plants life cycle needs to be considered, and there are many exogenous variables that should be taken into account. The problems is also explicitly dynamic.

While there are serious model specification problems, there are dependable basic model structures which provides valuable information; for example:

$$\text{population change} = \text{growth} - \text{harvest} - \text{natural mortality}.$$

Therefore abandoning structural modelling in favour of non-structural time series models would be wasting valuable information.

Part two below briefly reviews state space modelling and Gibbs sampling. In part three, a simplified state space model of ryegrass population dynamics and control is presented. A hypothetical data set generated from this model is used to examine the effectiveness of Gibbs sampling in estimating dynamic biological models.

2 Methods

State space models

State space models consist of an equation of motion(2), which describes how the states of the system (x_t) change over time, and an observation equation(1), which describes how the states are related to observations on the system..

$$1) y_t = f(x_t, \theta, u_t)$$

$$2) x_t = g(x_{t-1}, \theta, e_t)$$

The state space framework allows the relationship between the observations and underlying system to be explicitly defined, with distinct error terms associated with observation error (u_t) and with the randomness or uncertainty in the underlying process (e_t). Lags, exogenous influences, control variables and seasonality can also be built in. As such state space models provide a powerful framework for modelling pests and other natural resources

Linear state space models with normal additive errors can be estimated with the Kalman filter. (See Harvey 1989). However, for biological systems state equations must be nonlinear, and so are difficult to estimate. Carlin, Polson and Stoffer(1992) demonstrate how the Gibbs sampler can be used to estimate nonlinear state space models with nonnormal error terms in a Bayesian framework.

A brief overview of Gibbs sampling.

Gibbs sampling is a MCMC method for generating random draws from the full joint probability density function (jpdf) without having to calculate the density function. From the full jpdf the marginal densities of the variable of interest can be readily approximated. The technique is applicable to a wide range of problems and requires only that random draws from the full conditional probability distributions of all parameters are possible. Casella and George (1992) provide an introduction to the Gibbs sampler and Smith and Roberts (1993) supply a more detailed overview. Gelfand and Smith (1990) discuss how Gibbs sampling relates to other Monte Carlo Markov Chain procedures.

The method works as follows :

Consider a model with set of parameters U_i for $i = 1 \dots k$ and data matrix Y . Estimation of the model consists of finding the marginal density of the parameters conditional on the data, that is: $[U_i|Y]$, for $i = 1 \dots k$.

To do this with the Gibbs sampler:

1. Choose $\{U_1^{(0)}, U_2^{(0)}, U_3^{(0)}, \dots, U_k^{(0)}\}$, an arbitrary set of starting values.

2. Randomly draw a new set of values from the conditional probability functions, conditioning on these starting values and the data as follows:

$$U_1^{(1)} \sim [U_1^{(1)} | U_2^{(0)}, U_3^{(0)}, \dots, U_k^{(0)}, Y]$$

$$U_2^{(1)} \sim [U_2^{(1)} | U_1^{(1)}, U_3^{(0)}, \dots, U_k^{(0)}, Y]$$

$$U_k^{(1)} \sim [U_k^{(1)} | U_1^{(1)}, U_2^{(1)}, U_3^{(1)}, \dots, U_{k-1}^{(1)}, Y]$$

3 Repeat this process j times using the new set of values for U as a starting point.

Under certain conditions, and for j sufficiently large $\{U_1^{(j)}, U_2^{(j)}, U_3^{(j)}, \dots, U_k^{(j)}\}$ will effectively be a draw from the jpdf $[U_1, U_2, U_3, \dots, U_k | Y]$

4. Obtain multiple (K) draws from the posterior jpdf either by repeating this procedure or by increasing the number of iterations and sampling every j iterations.

The sample can then be used to calculate features of the posterior distribution of interest. For example the posterior mean of U_i is simply

$$\frac{1}{K} \sum_{k=1}^K U_i^k$$

Higher order moments and other functions of the posterior are equally straight forward to calculate. Multivariate features of the posterior can also be explored.

The posterior distribution can be estimated by averaging over the conditional density estimates based on all the sample sets. That is

$$f(U_i) \approx \frac{1}{K} \sum_{k=1}^K f(U_i | U_{-i}, Y)$$

The Gibbs sampler is extremely flexible, U_i can be scalar or vector valued, and missing data can be treated as simply another parameter to be estimated. The technique has successfully estimated models in many areas where the model specification or irregular error terms make other methods intractable. However the method does have its problems. It is difficult to test for convergence of the Gibbs sampler to the jpdf. (See Gelman and Rubin, 1992 and Geyer, 1992) Drawing random samples from nonstandard conditional distributions can prove difficult, however a range of new methods are available, see Smith and Gelfand (1992)

The technique is still relatively new, so many fundamental issues about the method are still being debated. The method has been applied to estimating Generalised linear models, hierarchical models in epidemiology and drug trials as well as state space modelling. See Casella and George(1992). Issues relevant to Gibbs sampling for state space modelling are discussed below.

Gibbs sampling for state space models

Carlin Polson and Stoffer (1992) demonstrate how the Gibbs sampler can be used to estimate nonlinear state space models with nonnormal error terms in a Bayesian framework.

Work by Liu, Wong and Kong (1994) suggests that where variables are highly correlated convergence of the Gibbs sampler to the true jpdf can be slow. Carter and Kohn (1994) noted that in state space models the state variables can be highly correlated. For linear state space models they use the Kalman filter to estimate the conditional distribution of the states. This estimate is then be used to generate the entire state vector simultaneously, within one step of the Gibbs sampler. This removes the correlation and avoided the problem slow convergence. Their results indicate that convergence can be dramatically speed up using this method. Scipone-Forbes (1994) extended this work to nonlinear models by using the extended Kalman filter to generate estimates of the states conditional distributions and used a Metropolis-Hastings algorithm, an alternative MCMC method, to generate the state variables. Other work on time series MCMC has been done by Shephard (1994), and McCulloch and Tsay(1994).

3 A stochastic weed management model.

This section presents a simplified model of ryegrass infestation of a wheat crop. The model is designed to capture the key features of the real problem in order to test the usefulness of Gibbs sampling in this situation. The model is loosely based on a model of ryegrass infestation of a wheat crop in the Western Australian wheat belt.

The model divides each year into four stages. In stage one, a proportion (M) of the seeds from the previous year die from natural causes over summer. In stage two, a proportion (G) of the surviving seed germinate. In stage three, the germinated weeds are sprayed with a selective herbicide, and a certain proportion survive and grow to maturity. Seed production occurs in stage four and the next year begins.

At the end of each stage there is an imperfect observation on the system (y_{it} , $i = 1..4$), and a state variable (x_{it} , $i = 1..4$) is defined as the true value of the weed or seed density at this time.

The model is defined as follows:

Observation Equations

Post summer seed count

$$3. \quad y_{1t} = x_{1t} * e_{1t} \quad e_{1t} \sim N(1, \sigma_1)$$

Seedling count

$$4. \quad y_{2t} = x_{2t} * e_{2t} \quad e_{2t} \sim N(1, \sigma_2)$$

Mature weed count

$$5. \quad y_{3t} = x_{3t} * e_{3t} \quad e_{3t} \sim N(1, \sigma_3)$$

Seed set count

$$6. \quad y_{4t} = x_{4t} * e_{4t} \quad e_{4t} \sim N(1, \sigma_4)$$

State Equations

Seed Mortality over summer

$$7. \quad x_{1t} = x_{4,t-1} * M \quad M \sim \text{beta}(a_m, b_m)$$

Seedling germination

$$8. \quad x_{2t} = x_{1,t-1} * G \quad G \sim \text{beta}(a_g, b_g)$$

Post emergent herbicide kill

$$9. \quad x_{3t} = x_{2,t-1} * \exp(A * H) \quad A \sim N(\mu_A, \sigma_A), H = \text{herbicide rate.}$$

Seed production

$$10. \quad x_{4t} = Nm (1 - \exp(-B * x_{3,t-1})) \quad Nm \sim N(\mu_n, \sigma_n), b \text{ fixed and known.}$$

Four further state variables are defined which indicate the stage of the model switch to the appropriate equations however these are not required in estimation of the model. The errors on the observations (y 's) are multiplicative and normally distributed with mean 1 and unknown variance. The possibility of negative seed counts is removed by keeping the variance small. Natural mortality over summer (M) and germination (G) are defined as proportions, they vary between 0 and 1. They are random variables drawn from beta distributions with unknown parameters a_m , b_m , and a_g , b_g respectively. In equation 9 a proportion $1 - \exp(-A * H)$ of the weeds are killed where H is the herbicide rate applied and A is the herbicide efficacy parameter. A is random variable from a normal distribution with unknown mean and variance.

Equation 10 is the seed production function. x_{4t} seeds are produced from $x_{3,t-1}$ mature weeds. For now b is considered fixed and known, while Nm , the maximum possible seed production is random and normally distributed with unknown parameters.

Note that a different random value for the parameters is drawn each year. Therefore the state equations are stochastic. They are also nonnormal and nonlinear making estimation by standard techniques intractable.

This model is designed to capture some important aspects of real weed models and to provide a challenging Monte Carlo test of the Gibbs state space estimation technique.

Model Estimation.

Ten years of simulated data was used to estimate the model. The model parameters and the priors used are listed in table one. Parameter priors were chosen to be relatively uninformative and slightly biased. The model was estimated using Excel 5 and Visual Basic, however lower level languages are required for reasonable computer efficiency. Fortran and C are popular choices. Three chains of 3000 iterations were generated, and after visual inspection for convergence, the first 500 iterations were disregarded. Of the remainder every third iteration was used to estimate

the posterior densities. The state variables were not simultaneously estimated as recommended by Carter and Cohn and only visual inspection was used to assess convergence. Conjugate priors were used for the priors on the parameters of normal and Beta distributions. Table 1 reports the parameter values used to generate the data and the prior distributions used in estimation.

4 Preliminary results from model estimation.

Figures one to four report the true values of the four state variables, the mean of the posterior distributions and the 90 percent confidence interval. Confidence intervals were calculated using the posterior densities. 67% of the state variables were within their 90% confidence intervals. In particular x_1 and x_4 are consistently overestimated. Apart from this the estimates appear to track the system very well.

Figures five and six show the prior, posterior and true distributions of germination (G) and Natural Mortality (Nm) variables respectively. The prior and posterior distributions were calculated using the mean of the distributions of a_m , b_m , and a_g , a_m . Note that for each parameters there were only ten observations with which to update the prior density. The results indicate a move towards the true distribution however not enough data was available to over ride the biased priors. In future applications I hope to use data from multiple sites in a hieratical model to overcome this data deficiency.

Table two reports the complete list of the means of the priors and posteriors and their relationship to the true values. Some variables are well estimated, however the posterior estimates of the variance parameters tend to be overestimated, and the posterior mean of Nm is biased upwards. This result appears to be due to an overly strict prior on one of the parameters. This may also explain the consistent overestimation of x_4 and x_1 .

Figure 7 and 8 show histograms of the posterior densities of the mean of Nm and A. The posterior of μ_N is biased upwards as mentioned. The true value of μ_A is within the 90% confidence interval of the posterior distribution.

While the overall estimates were poor, this appears to be due to an overly strict and biased prior. Despite this the method seems capable of estimating a model of this type with very little data.

5 Discussion and Conclusions.

The purpose of this paper was to point out the potential of Gibbs sampling and related MCMC's for the estimation of natural resource models. The results reported here suggest that the technique is capable of providing accurate and informative posterior distributions for a model with many of the features of a real natural resource problem. The model used here has relatively few observations (40), nonlinear state equations, some nonnormal error terms and a moderately complicated model structure. The true value of the state variables fell within the 90% confidence interval 67% of the time, despite problems with incorrect priors.

The important difference between this hypothetical model and a real application is knowledge of the model structure. However there is the scope within this framework for testing of restrictions and assumptions and the possibility of using flexible functional forms to capture uncertainty about nonlinear functions. Other real world problems missing from this example include the problem of missing observations and the use of nonconjugate and informative priors.

Missing observations are easily handled by Gibbs sampling. By treating the missing values as another parameter to be estimated the true posterior density for the missing observations is obtained and inference is based on this distribution. (See Gelfand *et al* 1990).

The use of nonconjugate prior increases the computational burden but otherwise cause no problems. Obviously, as the problems here illustrate, it is important that the prior density provides support over the range of all feasible values.

In summary the Gibbs sampler, although not without its problems, has the potential to allow estimation of complete biological and bioeconomic models making full use of the information available. This represents a significant improvement over the alternatives currently available.

References

- Carlin B.P., Polson N.G., and Stoffer D.S (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*. 87 493-500
- Carter C.K and Kohn R (1994) On Gibbs sampling for state space models *Biometrika* 81 541-553
- Casella G. and George E.I. (1992) Explaining the Gibbs sampler. *The American Statistician*. 46 167-174
- Gelfand A.E, Hills S.E, Racine-Poon A, and Smith A.F.M. (1990) Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*. 85 973-985
- Gelfand A.E and Smith A.F.M. (1990) Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*. 85 398-409
- Gelman A and Rubin D.B (1992) Inference from Iterative simulation using multiple sequences. *Statistical Science* 7 457-511
- Geyer C.J (1992). Practical Markov chain Monte Carlo *Statistical Science* 7 473-511
- Harvey A.C. (1989) Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge.
- Liu J S, Wong W.H and Kong A. (1994) Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* 81 27-40
- McCulloch R.E and Tsay R.S (1994) Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis* 15 235-250
- Scipione Forbes C, (1994) On Markov Chain Monte Carlo methods for nonlinear state space models. Paper presented at the 1994 Australasian meeting of the Econometrics Society, Amidale, NSW.
- Shephard N. (1994) Partial non-Gaussian state space. *Biometrika* 81 115-31
- Smith A.F.M and Gelfand A.E (1992) Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician* 46 84-88
- Smith A.F.M and Roberts G.O (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series. B*. 55 3-23

Table 1. Parameter Values and Priors.

Parameter	True Value	Mean of Prior	Std Dev of Prior	Form of Prior
σ_1	0.3	0.1	10	Inverse gamma
σ_2	0.3	0.1	10	Inverse gamma
σ_3	0.3	0.1	10	Inverse gamma
σ_4	0.2	0.1	10	Inverse gamma
x_{10}	3032	2500	800	Normal
a_m	4	0.83	0.36	Beta Likelihood
b_m	6	1.2	0.73	Beta Likelihood
a_e	9	2.6	4.4	Beta Likelihood
b_e	1	1.0	0.53	Beta Likelihood
μ_A	-7.5	-9.3	10	Normal
σ_A	0.25	4	13	Inverse gamma
μ_{Nm}	5000	4379	1000	Normal
σ_{Nm}	500	11000	10500	Inverse gamma

Table 2
Comparison of the True parameter values with the mean of the prior and posterior estimates.

Parameter	Prior Mean	Posterior Mean	True Value	% Improvement in Estimate
σ_1	0.32	0.59	0.3	-1681
σ_2	0.32	0.38	0.3	-387
σ_3	0.32	0.83	0.3	-3156
σ_4	0.32	0.39	0.2	-63
x_{10}	2500	3134.49	3032	81
a_m	0.83	2.53	4	54
b_m	1.18	1.72	6	11
a_e	2.63	6.14	9	55
b_e	1.04	1.70	1	-1650
μ_A	-9.3	-8.41	-7.5	50
σ_A	2	1.03	0.25	55
μ_n	4379	9480.76	5000	-622
σ_n	11000	1460.61	500	91

Figure 1 Seed Bank (X1)
 Prediced, True and 90% Confidence Interval

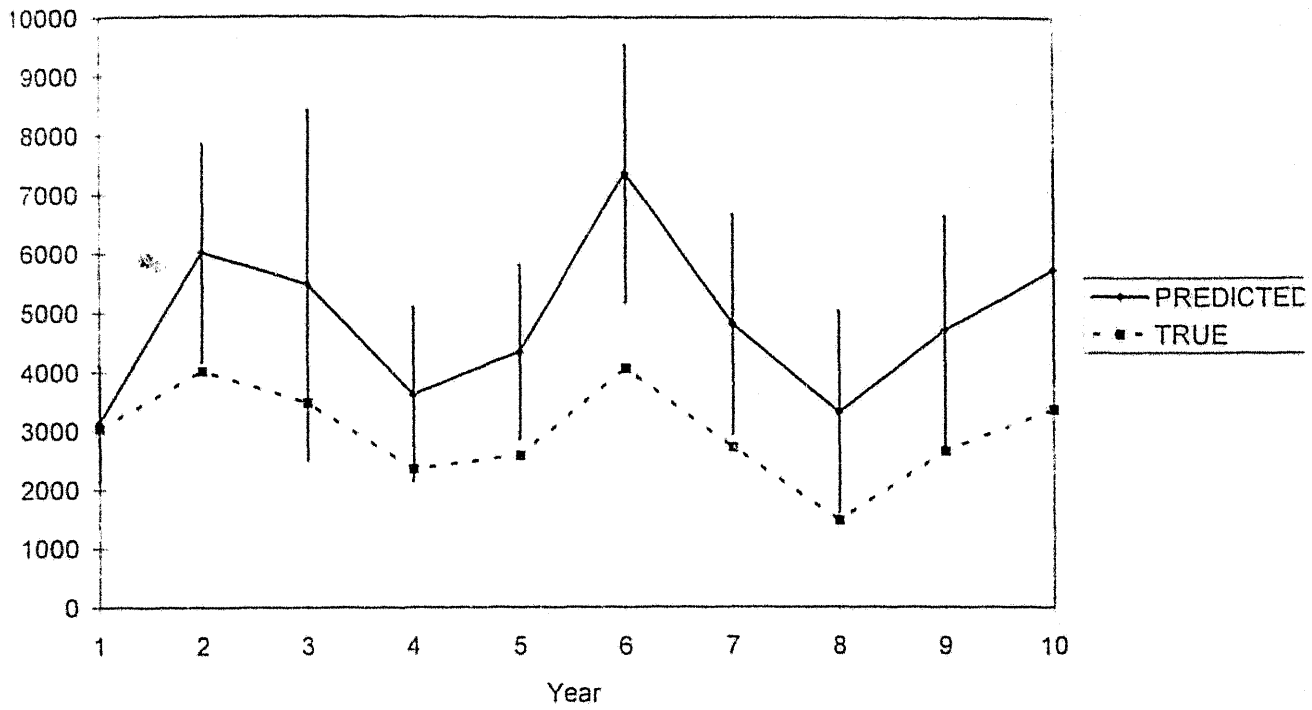


Figure 2 Seedling Density (X2)
 Prediced, True and 90% Confidence Interval

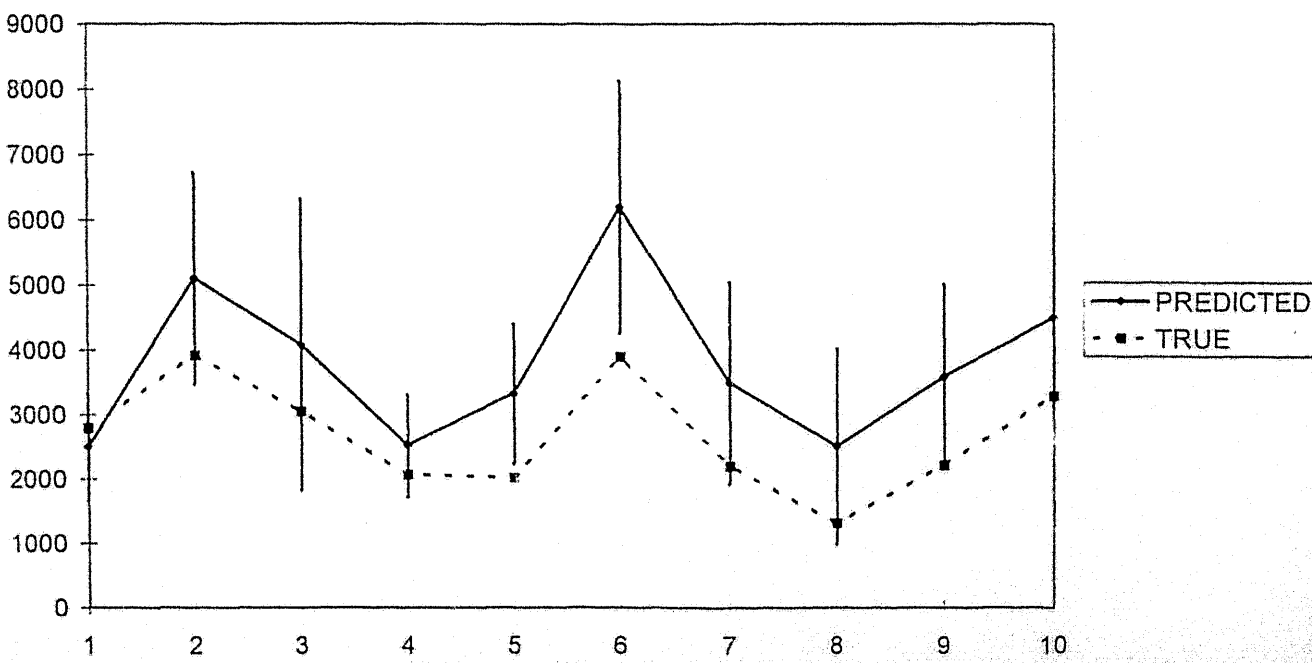


Figure 3 Mature Weed Density (X3)
 Predicted, True and 90% Confidence Interval

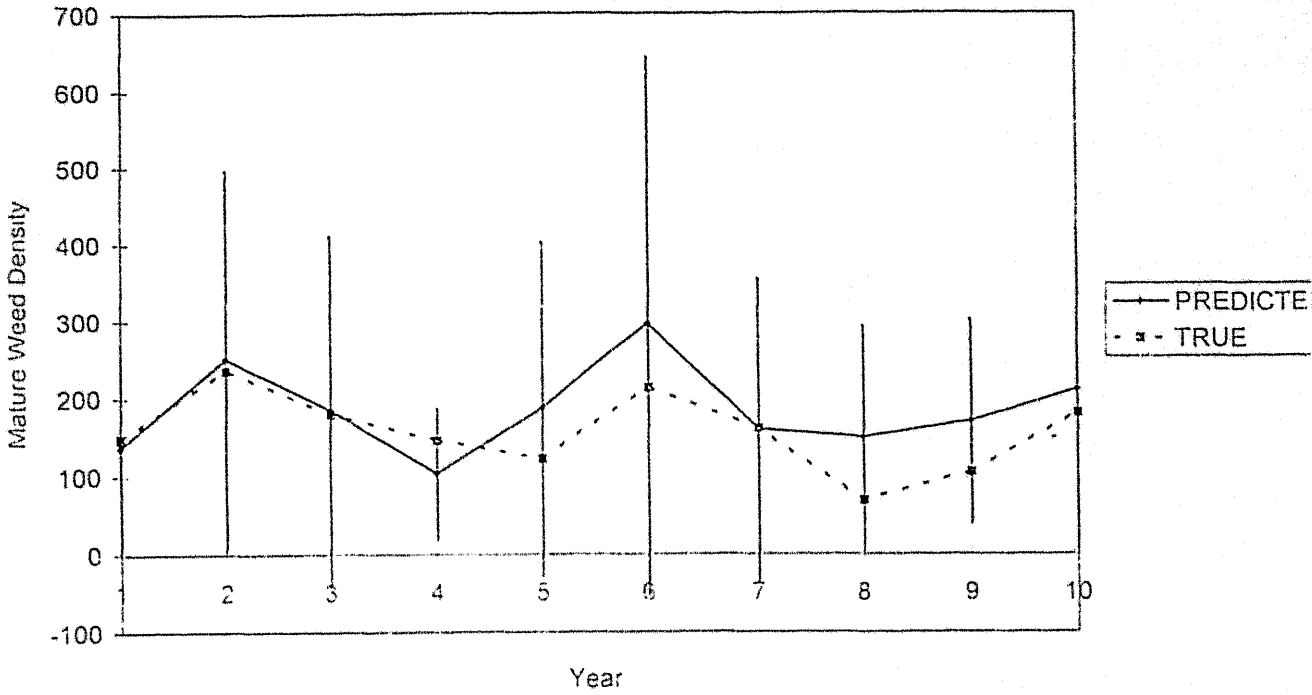


Figure 4 Seed Production (X4)
 Predicted, True and 90% Confidence Interval

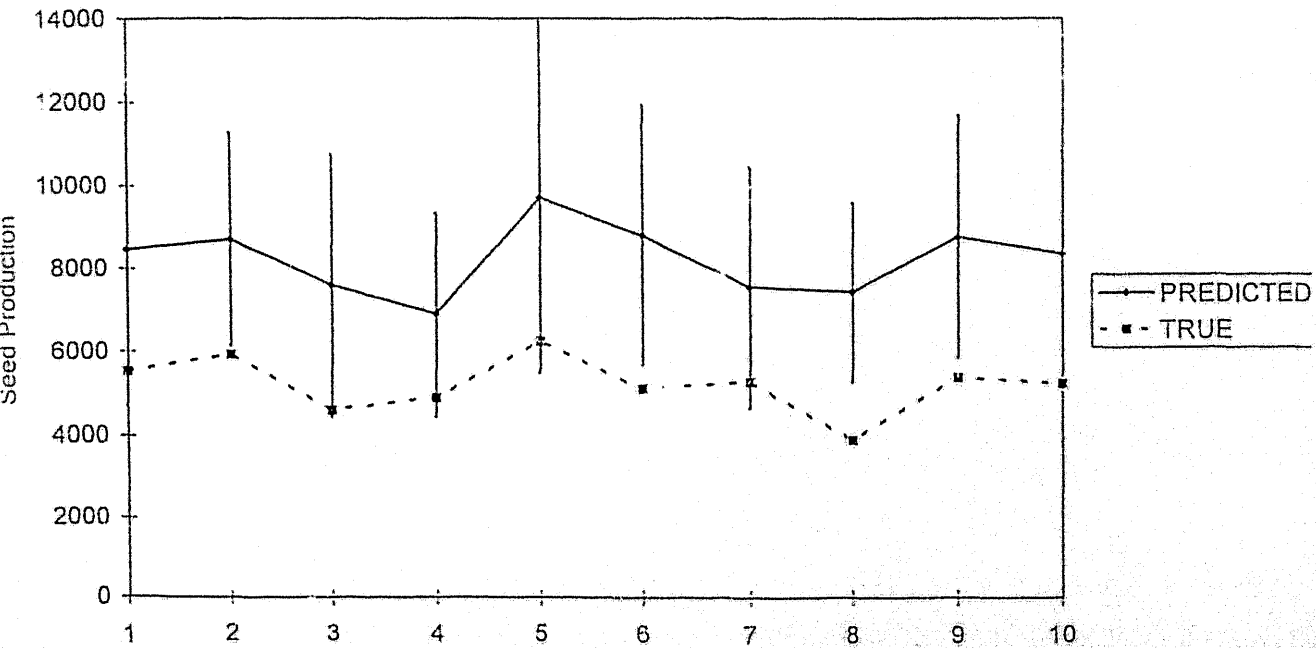


Figure 5
Prior, True and Estimated Distribution of Seedling Germination (G)

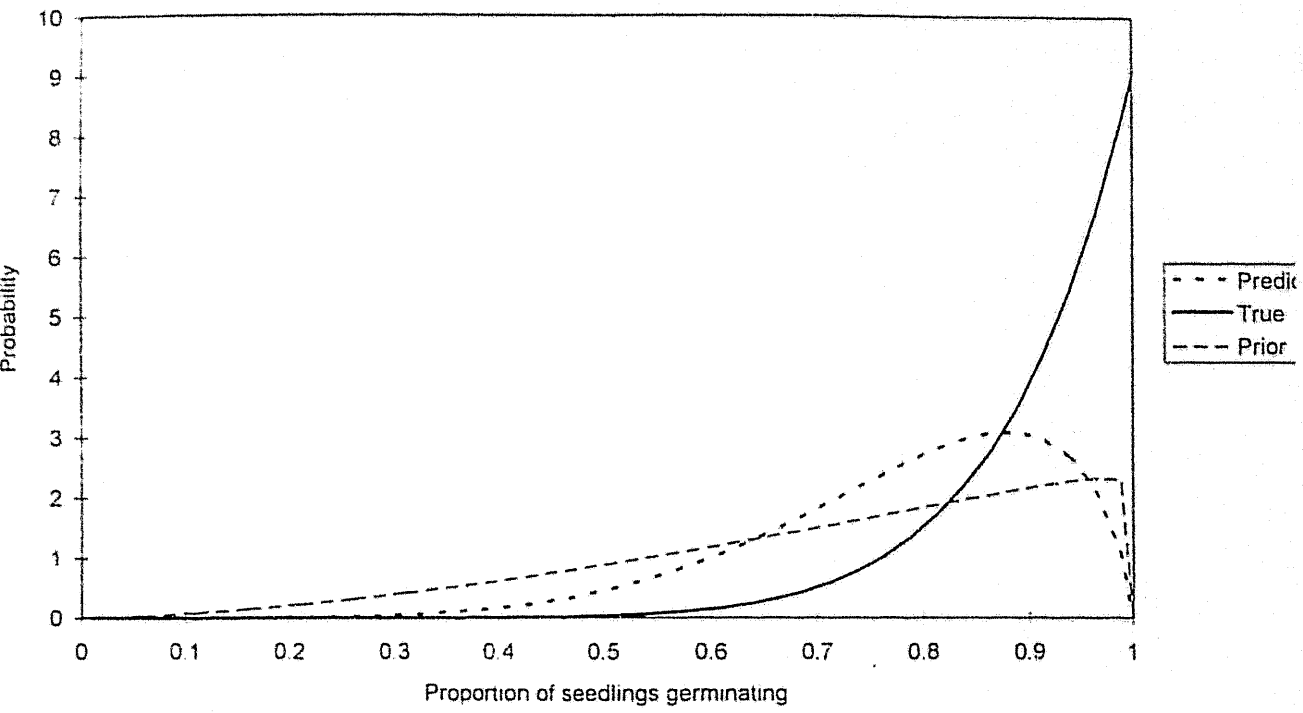


Figure 6
Prior, True and Estimated Distribution of Seed Mortality over Summer (M)

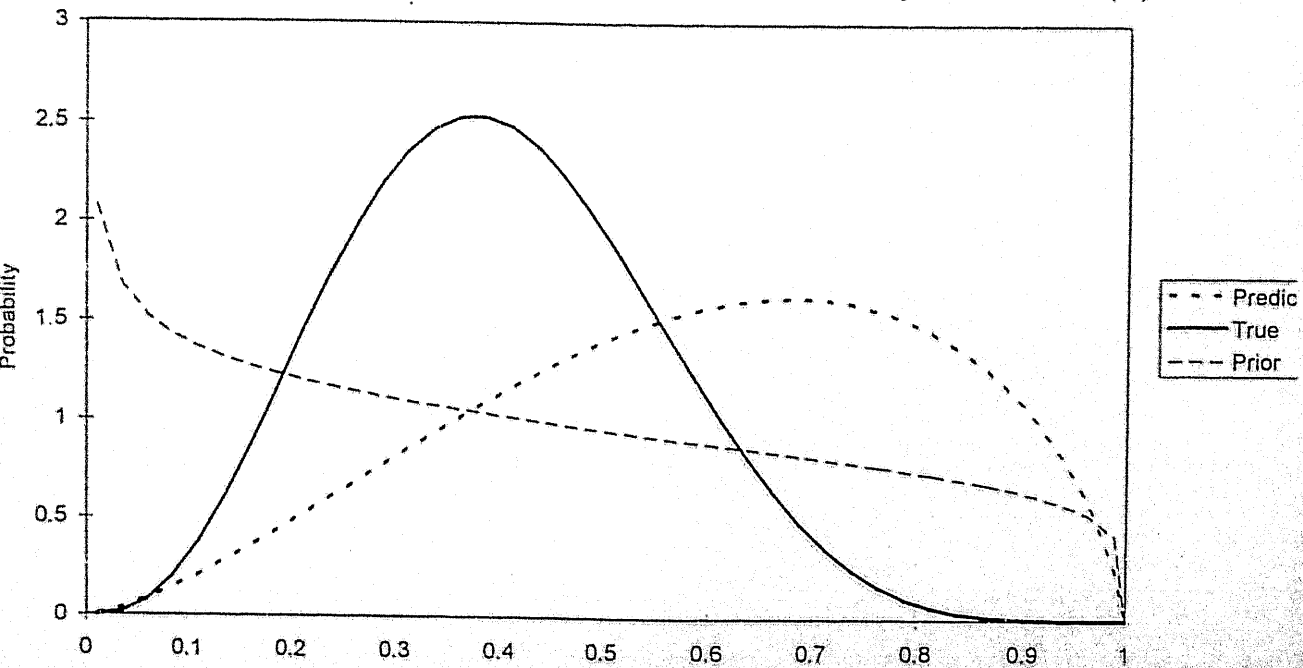


Figure 7
Histogram of the mean of Nm

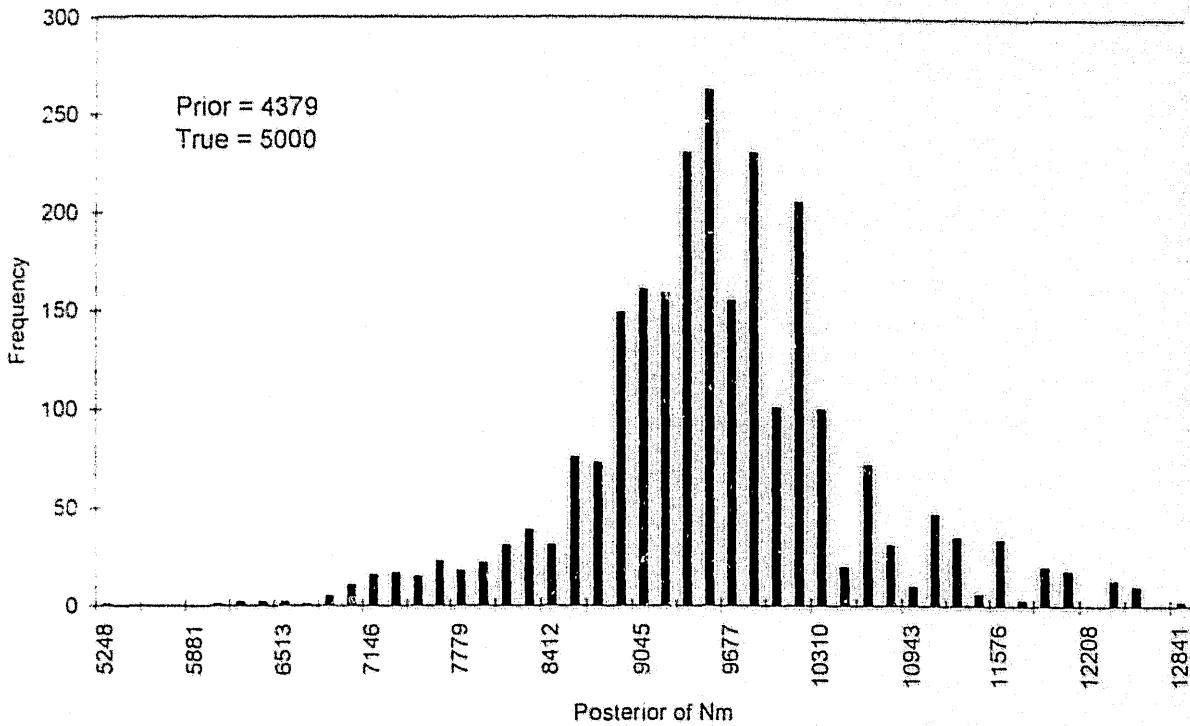


Figure 8
Histogram of the posterior density of A, the Herbicide kill coefficient

