REPRESENTATION OF UNCERTAINTY IN A MONTE CARLO SIMULATION MODEL OF A SCAVENGING CHICKEN RODUCTION SYSTEM

Anders Ringgaard Kristensen, ark@dina.kvl.dk, and Charlotte Vesterlund Pedersen
Department of Animal Science and Animal Health
The Royal Veterinary and Agricultural University
Grønnegårdsvej 8, DK-1870 Frederiksberg C
Copenhagen, Denmark

A framework for representation of uncertainty concerning true parameter variables in Monte Carlo simulation models is presented. The framework is valid for a situation where the purpose of the model is to use it for inference on a population of flocks of a particular type in order to evaluate the *general* consequences of changes in a management and/or production system. Results from the model show that the variation in output results between states of nature is far bigger than between replications within the same state of nature. It is therefore concluded that it is very important to represent the state of nature as a simultaneous distribution instead of just a set of estimates. The presented framework is, however, still not suitable if the purpose of the model is to be used for inference on a specific real life flock.

1. Introduction

Chicken production in most African countries is traditionally based on scavenging systems. This low input/output practise has been a component of small farms for centuries and is expected to continue as such in the future. In spite of the importance of the scavenging system, few activities have been initiated in this field in order to improve the output. A variety of factors act over time to influence productivity and thus a dynamic management tool could be useful. A dynamic stochastic model "SimFlock" of the traditional African chicken production system has been developed as a supplement to on-farm and on-station trials.

Numerous Monte Carlo simulation models of livestock herds have been developed (Blackie & Dent, 1976; Cacho et al., 1995; Chang et al., 1994; Faust et al., 1993; Harris et al., 1989; Henken et al., 1994; Herrero et al., 1999; Pomar et al., 1991; Singh, 1986; Østergaard et al., 2000), and a common trait is that they represent knowledge about the production system. In some cases this knowledge is based on the experience of human experts and in some cases it is based on empirical data analyses. Very often the same model relies on knowledge from both sources. In most cases the knowledge is entered into the model in the form of biological input parameters (growth rates, mortality levels etc.). In stochastic models, the variation between animals is also represented through appropriate distributions. The growth rate of the animals could for instance be specified through a mean value and a standard deviation, so that the actual growth rate of a particular animal is drawn from a normal distribution having specified parameters.

The basic assumption behind the described approach is that the true parameter values of the population being simulated are known with certainty. In the growth rate example the assumption is that the specified mean and standard deviations are the true values. In the real world the specified parameters are at best estimated with a certain precision. At worst they are just based on a vague experience of an expert in the field.

As long as the model is only used for system comprehension, the assumption of complete knowledge is not a problem, because all results of the model are produced under the assumption that the input parameters are known with certainty. If, on the other hand, the model is used for inference on real world systems (livestock herds), ignoring the uncertainty concerning the true parameter values may be fatal.

As opposed to many other livestock herd simulation models, SimFlock takes the uncertainty concerning the true parameter values into account. The basic idea is that, in this context, uncertainty is not the opposite of knowledge - it is a property of knowledge. In principle, each time a parameter is defined in the model, a hyper distribution representing the uncertainty concerning the true value is specified as well. A complete set of parameter values drawn from these hyper distributions is referred to as a "state of nature". For each state of nature, multiple simulation runs will reveal the output range under these specific conditions (concerning growth, mortality, egg laying probabilities etc.). The variation in output between the simulation runs of different states of nature expresses the uncertainty concerning the true parameter values.

2. The scavenging chicken production system

Chicken production in most African countries is traditionally based on scavenging systems. This low input/output practice has been a component of small farms for centuries and is thought to continue as such in the future. The main product is chicken for household consumption or marketing whereas egg production is negligible. In spite of the importance of the scavenging system, few activities have been initiated in this field in order to improve the output. There are many advantages in this production system but also constraints. The advantages consist of a production based on free feed resources available in the surrounding environment and kitchen leftovers, using local chicken breeds, which have adapted to the existing conditions having preserved their ability to incubate and brood naturally. All these factors maintain an inexpensive level of production and minimal management requirements. The disadvantages are high mortalities, low egg production and slow growth. Thus in recent years more focus has been on how the existing constraints can be minimised and thereby increase the overall production.

A simplified flow diagram of the traditional African production system is shown in figure 1.

3. Monte Carlo simulation

As described by Jørgensen (2000), Monte Carlo simulation is a method for evaluating an integral

$$\Psi = E_p\{U(X)\} = \int U(x)p(x)\mathrm{d}x \tag{1}$$

where $E_p\{\}$ is the expectation with respect to the probability density p and $X = \{\Theta, \Phi\}$ is a vector of decision parameters, Θ, and a combination, Φ, of system parameters (mortality, daily gain etc.) and state variables calculated by the model. The symbol $U()$ denotes some kind of response function like economic net returns, number of animals produced etc. In general, we may consider $U()$ to be a utility function. The Monte Carlo method is a numeric method for evaluating the integral in Eq. (1) by generating random draws $X = x^{(j)}$ from the target distribution p and then estimating Ψ as the average of $U(x^{(1)}),\ldots,U(x^{(k)})$, where k is the number of replications.

As mentioned, Φ is a combination of system parameters (often referred to as the *state of nature*), Φ_0, and state variables $\Phi_{\bullet s} = \{\Phi_{1s},\ldots\Phi_{Ts}\}$ calculated by the model. This splitting leads to the following reformulation of Eq. (1):

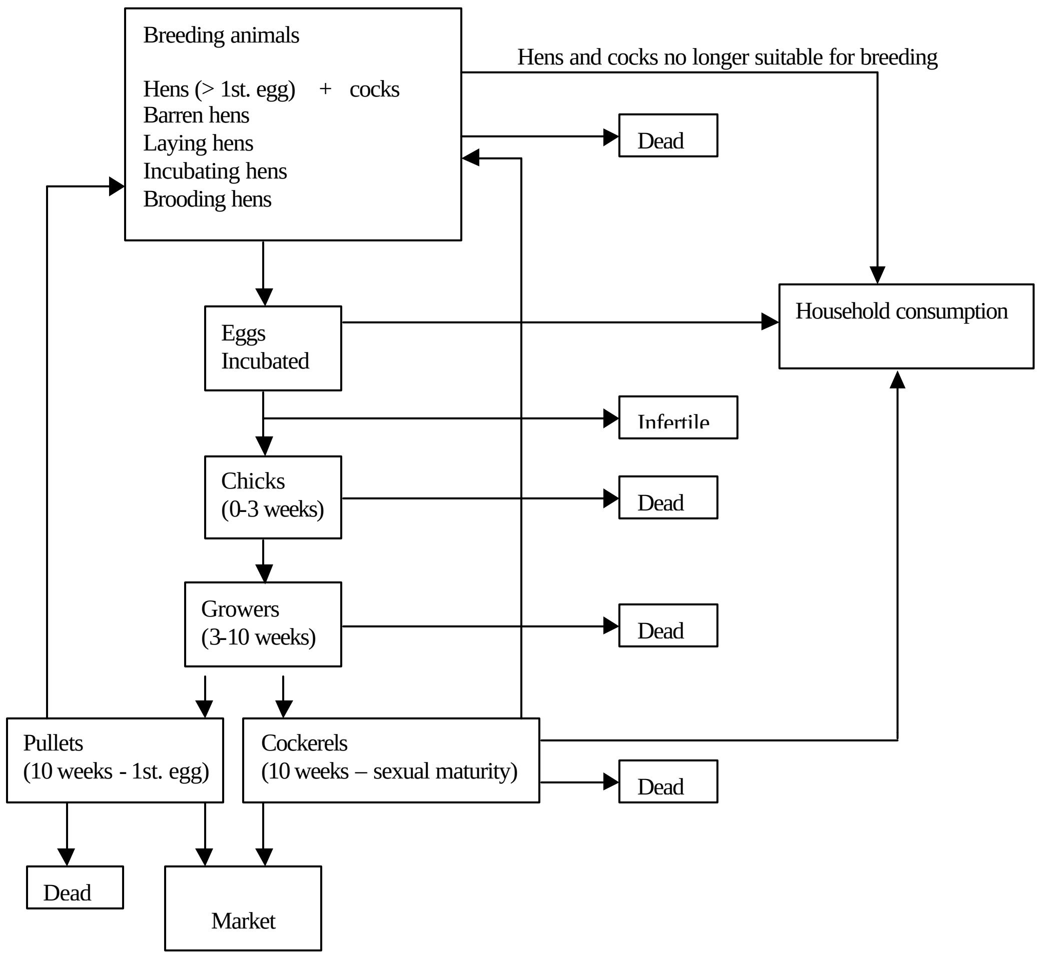

Figure 1. Flow diagram of a traditional African chicken production system

$$\Psi = E_{p_0}\{E_{p_{s|0}}\{U(X)\}\} = \int\left\{\int U(x)\frac{p(x)}{p_0(\Phi_0)}\mathrm{d}\{\Theta,\Phi_{\bullet s}\}\right\}p_0(\Phi_0)\,\mathrm{d}\Phi_0 \qquad (2)$$

where $E_{p_0}\{\}$ is the expectation with respect to the probability density p_0 of the state of nature, and $E_{p_{s|0}}\{\}$ is the expectation with respect to the conditional probability density given a state of nature.

The double integral of Eq. (2) may be evaluated numerically by first drawing j states of nature, $\Phi_0^{(1)},\ldots,\Phi_0^{(j)}$, at random from the distribution p_0. For each state of nature a random sample consisting of k observations from the distribution $p_{s|0}$ is created by running the simulation model k times. Since we have j states of nature the total number of simulation runs equals $n = j \times k$.

Most simulation models (with Jørgensen, 2000, as a remarkable exception) assume that the state of nature is known with certainty. In relation to Eq. (2) this means that a certain state of nature (set of parameters), Φ_0', has probability 1 (i.e., $p_0(\Phi_0')=1$). Under that assumption the outermost integral of Eq. (2) vanishes and the evaluation reduces to the innermost integral.

Since, however, the state of nature represents biological parameters like mortality or daily gain, the true values are seldom known with certainty. Assessing the simultaneous distribution of the parameters of the state of nature therefore becomes a major issue if the model is intended to be used for inference on real life systems (livestock herds, flocks) as discussed in the introduction.

4. Assessing the distribution of the state of nature for the simulation model

Parameters describing a scavenging chicken flock

The parameters used for characterization of a chicken flock are summarized in Table 1. Most of the rows of the table represent several parameters of the state of nature, Φ_0. There is for instance a daily gain parameter for each bird category (male chicks, female chicks, male growers, female growers, cockerels and pullets). Several parameters are defined by a mean and a standard deviation (for instance age at puberty). The total number of parameters is 42. A complete set of values for the parameters defines a state of nature.

Table 1. Parameters of the state of nature characterizing a chicken flock .

Parameter	Distribution/Model
Daily gain of chicks, growers, cockerels and pullets	General linear model, Eq. (6)
Survival rate of chicks, growers, cockerels and pullets	Logistic model, Eq. (7)
Full grown weight, hens and cocks, mean and std. dev.	Normal distribution, Eq. (3)
Age at puberty for cockerels and pullets, mean and std. dev.	Normal distribution, Eq. (3)
Cock fertilization probability	Beta distribution
Egg hatching probability	Logistic model, Eq. (4)
Number of eggs before incubation, mean and std. dev.	Normal distribution, Eq. (3)

The theoretical basis of the concept of state of nature is that a particular flock has a set of fixed, but unknown, parameter values characterising that particular flock. Another flock will have another set of values. If we have sufficient data from a flock we may estimate the parameters, but only with a certain degree of precision. The true values will remain unknown and the degree of uncertainty will depend on the amount of data. As mentioned, the traditional approach is to use the estimated parameters directly and ignore the uncertainty, but if we want to draw conclusions concerning a particular flock, it is important that this uncertainty is reflected by representing the simultaneous distribution of the state of nature. Otherwise there is a severe risk of overestimating the effect of changes in the production system and/or management of the flock. Use of the simulation model for inference on particular flocks is, however, outside the scope of this study, even though the model presented has the potential to be further developed for that purpose as well.

In this study, the purpose of the model is to use it for inference on a population of flocks of a particular type. In other words, the purpose is to evaluate the *general* consequences of changes in a management and/or production system. Since each flock has its own state of nature, the consequences will also differ. In order to be able to estimate these differences, the distribution of the state of nature must reflect the variation in parameters between flocks.

The population used for estimation of the distribution were flocks participating in on-farm studies in Zimbabwe over a 2-year period (Pedersen et al., 2002).

Estimating the distribution of the state of nature: independent parameters

If the parameters of the state of nature are mutually independent, the estimation procedure is very simple. We may for instance for continuous values use the general model

$$Y_{ij} = m + A_i + e_{ij} \tag{3}$$

where Y_{ij} is the observation (e.g. daily gain) for bird j in flock i, m is the overall mean, $A_i \sim N(0, s_A^2)$ is the random effect of flock i, and $e_{ij} \sim N(0, s^2)$ is a random term representing the variation between birds within flock. In relation to the state of nature, $\Phi_0 = (j_1, \ldots, j_K)$, this model defines two flock specific parameters, j_k and j_{k+1}. One of them is the flock specific average value $j_k = (m + A_i) \sim N(m, s_A^2)$ implying that the probability density function p_{0k} is the probability density function of the normal distribution $N(m, s_A^2)$. We shall refer to m and s_A as *hyper parameters* of the simulation model. The other parameter of the state of nature is the standard deviation within flock, $j_{k+1} = s$, where $p_{0,k+1}(j_{k+1} = s) = 1$ (ignoring the uncertainty concerning the true value of the standard deviation within flock).

For binary variables (e.g. hatching of eggs) we might accordingly use the logistic model

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = m + A_i + e_{ij} \tag{4}$$

where p_{ij} is the observed ratio between success (e.g. hatching) and failure (e.g. foetus death) in flock i, $A_i \sim N(0, s_A^2)$ is the random effect of flock and e_{ij} is a random term representing the variation within flock. Also this model defines a parameter, j_k, of the state of nature. It is the probability parameter

$$j_k = 1/\left(\exp\left(- m - A_i\right) + 1\right) \tag{5}$$

of a binomial distribution defining the probability of success. Again, m and s_A are hyper parameters, but the distribution p_{0k} is in this case more complicated due to the transformation in Eq. (5).

A few parameters describing the distribution of binary variables are difficult to estimate because of lacking registrations. An example is the probability that an egg has been fertilized by the cock of the flock or (if no cock is available) by the cock of a neighbour flock. In such cases the beta distribution is well suited for qualified expert guesses. It takes two hyper parameters, a and b, and the *average* probability of success (between flocks) is simply $a/(a+b)$. The level certainty is also very easy to assess, since the sum $a + b$ corresponds to the number of assumed observations behind the guess. If, for instance, the expert expects that 90% of all eggs are fertilized, and his level of certainty corresponds to a situation where he had observed 100 cases, the parameters should be $a = 90$ and $b = 10$. The distribution is also very easy to update, if observations later become available. If $n = n_a + n_b$ new cases are observed, the new parameters a' and b' become $a' = a + n_a$ and $b' = b + n_b$.

The advantage of mutually independent parameters of the state of nature is that the individual parameters may be sampled one by one. In the models (3) and (4), A_i is drawn from the relevant normal distribution, and the corresponding flock specific parameter is calculated as $j_k = (m + A_i)$ or $j_k = 1/\left(\exp\left(- m - A_i\right) + 1\right)$, respectively. In case of qualified guesses, the probability parameter is drawn from the relevant beta distribution. This one by one sampling is valid because in case of mutually independent parameters, the simultaneous probability density function p_0 is just the product of the individual probability density functions, i.e. $p_0 = \prod_k p_k$.

Estimating the distribution of the state of nature: dependent parameters

In many cases an assumption of mutually independent parameters is obviously wrong. Consider for instance the daily gain of different bird categories of a flock. It is reasonable to assume, that some flocks in general have high daily gain (i.e. for *all* categories) whereas other have a lower daily gain. Furthermore it must be expected, that some individual birds have a higher growth capacity than others. In other words, if a chick has a relatively high growth rate it is expected also to grow faster when it becomes a grower and later a cockerel/pullet). In such cases we need more sophisticated models in order to represent the distribution correctly.

The following model was used for daily weight gain

$$Y_{ijk} = m + a_i + F_j + (aF)_{ij} + B_k + e_{ijk} \tag{6}$$

where Y_{ijkl} is the observed logarithm to the daily gain of bird k in age group (category) i in flock j, a_i, (i = 1, …, 6) is the systematic effect of age group/category (male and female chicks, male and female growers, cockerels and pullets), $F_j \sim N(0, s_F^2)$ is the random effect of flock, $(aF)_{ij} \sim N(0, s_{aF}^2)$ is the random interaction between age group and flock, $B_k \sim N(0, s_B^2)$ is the random (permanent) effect of bird (the growth potential), and $e_{ijk} \sim N(0, s^2)$ is a random term representing the variation between age groups within flock and bird. This model defines 8 parameters $j_{n+1}, \ldots, j_{n+8}$ of the state of nature:

- 6 flock and age group specific growth parameters, $j_{n+i} = m + a_i + F_j + (aF)_{ij}$, where for i = 1, …, 6, $j_{n+i} \sim N(m + a_i, s_F^2 + s_{aF}^2)$. As the distribution of the parameters is normal, the hyper parameters needed are simply $m, a_1, \ldots, a_6, s_F$ and s_{aF}.
- The standard deviation between birds within flock, $j_{n+7} = s_B$. In other words, we again assume the degenerate distribution $p_{0,n+7}(j_{n+7} = s_B) = 1$, which means that we must consider s_B as a hyper parameter describing $p_{0,n+7}$.
- The standard deviation between age groups within flock and bird, $j_{n+8} = s$. Also here we assume $p_{0,n+8}(j_{n+8} = s) = 1$ making s a hyper parameter.

The procedure for sampling these flock specific state of nature parameters from the hyper distribution is as follows:

- A random value F_j is drawn from $N(0, s_F^2)$.
- 6 random values $(aF)_{1j}, \ldots, (aF)_{6j}$ are drawn from $N(0, s_{aF}^2)$.
- The 6 state of nature variables are calculated as $j_{n+i} = m + a_i + F_j + (aF)_{ij}$, where i = 1, …, 6. Since the same value of F_j is used for all 6 parameters, they become mutually correllated.
- The two standard deviations s and s_B are (trivially) sampled.

For binary variables a logistic model is used instead of Eq. (6). This applies for instance for the survival rates, where a similar interdependence is assumed. The following model is applied:

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = m + a_i + F_j + (aF)_{ij} + e_{ij} \tag{7}$$

where p_{ij} is the observed ratio between survival and death in flock j, age group i, a_i, (i = 1, …, 4) is the systematic effect of age group/category (chicks, growers, cockerels and pullets), $F_j \sim N(0, s_F^2)$ is the random effect of flock, $(aF)_{ij} \sim N(0, s_{aF}^2)$ is the random interaction between age group and flock and e_{ij} is a random term representing the variation within flock and age group.

This model defines 4 parameters $j_{n+1}, \ldots, j_{n+4}$ of the state of nature. They are 4 flock and age group specific parameters, $j_k = 1/\left(\exp\left(-m - a_i - F_j - (aF)_{ij}\right) + 1\right)$, i = 1, …, 4. The hyper parameters needed are simply $m, a_1, \ldots, a_4, s_F$ and s_{aF}. The procedure for sampling these flock specific state of nature parameters from the hyper distribution is as follows:

- A random value F_j is drawn from $N(0, s_F^2)$.
- 4 random values $(aF)_{1j}, \ldots, (aF)_{4j}$ are drawn from $N(0, s_{aF}^2)$.
- The 4 state of nature variables are calculated as $j_k = 1/\left(\exp\left(-m - a_i - F_j - (aF)_{ij}\right) + 1\right)$ where i = 1, …, 4. Since the same value of F_j is used for all 4 parameters, they become mutually correllated.

5. Simulation procedure

The simulation model has been constructed by use of an object oriented approach. It has a graphical user interface and it has been programmed entirely in Java. For a more detailed presentation of the simulation model reference is made to Pedersen & Kristensen (2002).

The idea of the model is that the user specifies the hyper parameters estimated as described in the previous section. By drawing random numbers from the relevant hyper distributions, the model is able to generate an arbitrary number of states of nature each representing realistic flock conditions under the circumstances in question (i.e. the circumstances of the flocks used for estimation of the hyper parameters).

The user must also define a production strategy to be used (i.e. the decision parameters, Θ). The production strategy is evaluated through a simulation job defined by the user. The simulation job definition includes the specification of the number of states of nature to generate, the number of replications per state of nature (i.e. evaluation of the innermost integral of Eq. (2)), the number of days to simulate per replication and, finally, the burn-in period (initial period where the results are discarded).

The model calculates the desired results (technical and economic key figures) for each combination of state of nature and replication. Calculation of the average over replications under the same state of nature corresponds to a numerical evaluation of the innermost integral of Eq. (2). When, furthermore, the average over states of nature are calculated, the outermost integral is numerically evaluated as well.

For an illustration of the potential of the model, reference is made to Pedersen & Kristensen (2002). In this paper we shall only illustrate the importance of a careful representation of the uncertainty concerning true parameter values by the result of a single simulation job, where 100 states of nature were generated, the number of replications per state of nature were 50, the simulation period was 5 years and the burn-in period 2 years. The results are expressed as number of chickens produced, and they are shown in Figure 2 as average and standard deviation for each of the 100 states of nature (interpreted as different flocks). Simulation results clearly illustrate that at least under the present conditions, the uncertainty concerning true parameter values has a marked influence on the output range.

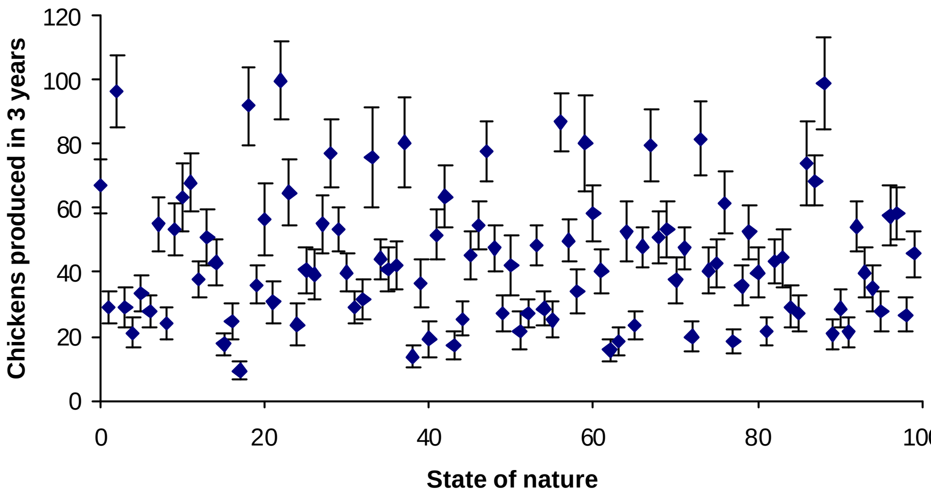

Figure 2. Variation in number of chickens produced within and between farms

Given the fact that almost all existing livestock simulation models ignore the uncertainty concerning the state of nature, it is remarkable to see the effect of the uncertainty concerning the parameters of the state of nature in Figure 2.

6. Conclusion

Results from the model show that the variation in output results between states of nature is far bigger than between replications within the same state of nature. It is therefore concluded that it is very important to represent the state of nature as a simultaneous distribution instead of just a set of estimates.

The presented framework is, however, still not suitable if the purpose of the model had been to use it for inference on a specific real life flock.

REFERENCES

Blackie, M.J. & J.B. Dent (1976). Analyzing Hog Production Strategies with a simulation Model. *American Journal of Agricultural Economics*, 58, 39–46.

Cacho, O.J., J.D. Finlayson, & A.C. Bywater (1995). A simulation model of grazing sheep: II. Whole Farm Model. *Agricultural Systems*, 48, 27–50.

Chang,W., D.H. Streeter, & L.R. Jones (1994). An Object-Oriented Model for Simulating Milking Parlor Operations. *Journal of Dairy Science.*, 77, 84–93.

Faust, M.A., O.W. Robison, & M.W. Tess (1993). Integrated Systems Analysis of Sow Replacement Rates in a Hierarchical Swine Breeding Structure. *Journal of Animal Science*, 71 pp. 2885–2890.

Harris, D.L., C. Pomar-Goma, & F. Minvielle (1989). Computer Simulation of Biological Aspects of Swine Production. *Swine Research, Progress Report No 3. Roman L. Hruska U. S: Meat Animal Research Center.*, pp. 14–16.

Henken, A.M., E.A.M. Graat, H.W. Ploeger, & T.E. Carpenter (1994). Description of a model to simulate effects of Eimeria acervulina infection on broiler production. *Parasitology*, 108, 513–518.

Herrero, M., R.H. Fawcett, & J.B. Dent (1999). Bio-economic evaluation og dairy farm management scenarios using integrated simulation and multiple-criteria models. *Agricultural Systems*, 62, 169–188.

Jørgensen, E. 2000. Calibration of a Monte Carlo simulation model of disease spread in slaughter pig units. *Computers and Electronics in Agriculture 25*, 245-259.

Pedersen C.V. & A. R. Kristensen. 2002. Dynamic modeling of a traditional African chicken production system. In: Pedersen, C.V. *Production of Semi-scavenging Chickens in Zimbabwe.* PhD Thesis, Department of Animal Science and Animal Health, Royal Veterinary and Agricultural University, Chapter 7, p 91-116.

Pedersen C.V., J. Kusina & J. Madsen. 2002. Egg production, mortality, growth, and parasite burden of village chickens under improved management and feeding conditions using locally available feed sources. In: Pedersen, C.V. *Production of Semi-scavenging Chickens in Zimbabwe.* PhD Thesis, Department of Animal Science and Animal Health, Royal Veterinary and Agricultural University, Chapter 5, p 49-77.

Pomar, C., D.L. Harris, P. Savoie, & F. Minvielle (1991). Computer simulation model of swine production. III. A dynamic herd simulation model including reproduction. *Journal of Animal Science*, 69, 2822–2836.

Singh, D. (1986). Simulation of swine herd population dynamics. *Agricultural Systems*, 22 pp. 157–183.

Østergaard, S., J.T. Sørensen & A.R. Kristensen. 2000. A stochastic model simulating the feeding-health-production complex in a dairy herd. *Journal of Dairy Science 83*, 721-733.