NOTE TO USERS

The original manuscript received by UMI contains pages with slanted print. Pages were microfilmed as received.

This reproduction is the best copy available

UMI

MODELING A SWINE BREEDING HERD: DATA NEEDS AND THE VALUE OF INFORMATION

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

ALVARO SOLER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR IN PHILOSOPHY

JULY, 1998

UMI Number: 9838668

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

UNIVERSITY OF MINNESOTA

This is to certify that I have examined this bound copy of a doctoral thesis by

Alvaro Juan Soler

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made

William Ernest Marsh

July 1998

GRADUATE SCHOOL

ACKNOWLEDGEMENTS

To my family; I owe them everything. To the Marsh family; their relentless encouragement, support and understanding made this endeavor possible. To Doris Mold and Andrew Whyte; they put warmth to the coldest times.

I cannot forget my fellow Uruguayan friends and colleagues, as well as the many International and American fellow students with whom I shared a lot of hours of both frustration and joy. They always had a supporting word to keep me going, and allow me to bring my dreams to fruition. A special recognition to the support staff at the CAPS Department; they always lent a helping hand.

Last, but not least, I want to thank the members of my Graduate Committee: Dr. Will Marsh, Dr. Robert Morrison, Dr. R. Ashley Robinson, Dr. Robert King, Dr. Thomas Blaha and Dr. Gary Dial, for their support, advice and helpful comments to this work.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

LIST OF FIGURES

ABSTRACT

In a business world where decreasing profit margins have turned efficiency into the cornerstone of profitability, making the best possible management decisions with a better understanding of their most probable outcome, has become the difference between staying in, or going out of business.

Unfortunately, perfect knowledge cannot be achieved. Technology, however, has provided ways to mitigate the shortcomings of imperfect knowledge, and to minimize the uncertainty surrounding business decisions. On the other hand, technological progress has also meant increased dependency on data gathering and processing, which may sometimes be counterproductive to the ends sought.

A simulation model for a swine breeding herd was developed and validated, in order to make available a tool to attain better management decisions in swine breeding. The model was used to assess the influence of data type and quality on simulation performance.

Results show that the model performs very well, and its output is reliable and accurate, with satisfactory agreement over all the range of farm sizes studied. Model performance was less volatile than actual system performance, and appeared more accurate for larger than smaller farms.

Results show that the density of data supplied to the model did not affect its simulation performance, and that for our test conditions, input data quality did not affect simulation performance.

CHAPTER 1

Simulation modeling of livestock production systems

1.1 Introduction

During the past 30 years, manifestations of important animal diseases have become more and more subtle. As it is virtually impossible to permanently eliminate infectious agents from intensive livestock production systems, management has become increasingly recognized as having a major role in determining the production efficiency of livestock farming.

Economic decision-making is a procedure that assigns scarce resources among alternative productive uses which are consistent with the goals and objectives of the business (Kay, 1986; Jalvingh et al., 1992). Having a priori knowledge of the expected impact of alternative courses of action on the enterprise bottom line (i.e., net profit), is an integral part of making the best decisions (Jalvingh, 1992). The combination of the amount of work involved in calculating the probable outcomes of different courses of action and the increased availability and computational power of personal computers has led to the use of computerized decision support systems (DSS) as a tool for better production management (Marsh, 1986; Lloyd, 1989).

Farm managers employ four types of information: descriptive, diagnostic, predictive, and prescriptive (Harsh et al., 1981). Computerized decision support systems typically comprise a management information system (MIS) and a simulation model. Descriptive and sometimes diagnostic information is provided by the MIS through collection and processing of production data into useful summary information. This information, describing the critical aspects of the production process, can be input in the simulation model portion of the DSS to provide automated diagnosis and predictive functions (Lloyd, 1989). Appropriate use of simulation models can contribute to the strategic and tactical planning functions of management, particularly in improving understanding of technical and economic relationships among factors of production (de Hoop, 1993).

During the past decade computerized production record systems have become widely adopted by pork producers, and have become the core component of on-farm management information systems. Stored data are used to prepare reports of past performance, to determine current system status and, in some cases, provide information to support putative diagnosis of underlying causes of suboptimal productivity. A database of detailed animal life histories constitutes a wealth of information that can be exploited to provide insight into the likely outcomes of possible alternative courses of action. It is the purpose of this thesis to draw upon those resources to develop a practical decision support management tool, and to determine the minimum requirements for record keeping practices that maximize the usefulness of such tool.

1.2 Systems approach and modeling

The systems approach and its relation to modeling is described by Spedding (Spedding, 1993). He defines system as "a group of inter-related components operating together for a common purpose". Systems are capable of reacting as a whole to external stimuli, remain unaffected directly by their own outputs, and have specified boundaries based on the inclusion of all significant feedbacks (Sørensen, 1987).

Within this context, models are useful tools in furthering our understanding of livestock production systems. A computer model is the best mechanism for defining the boundaries of the system, explaining existing knowledge of the system, in terms of the components and their interactions, specifying quantifiable inputs and outputs, and perhaps most importantly, exposing the gaps in knowledge necessary to construct a functional model to help us fully understand the working system.

No change in any part of a system can be regarded as an improvement if it does not result in an improvement of the system as a whole (Spedding, 1993).

Improvement must be sought for the system as a whole, and any attempt to improve has first to define, delineate and describe the system to be improved. In the real world, there has to be an agreement on what constitutes an improvement, and it is clear that it cannot be achieved by changes in one component, and certainly not without regard to the rest of the system.

A model is an abstraction and simplification of the real world. It is specified so as to capture the main interactions and behavior of the system it portrays, and it must be capable of experimental manipulation in order to project the consequences of changes in the determinants of system behavior. Five areas in which modeling can be applied, include:

a - scientific research,
b - teaching,
c - advisory work,
d - management activities on the farm, and
e - political decisions.

While understanding a process within a system increases our capacity to improve the system, our understanding must be related to a purpose. Just because a process occurs within a system does not necessarily mean that the process is important, primarily in the sense that the operator can influence it. Also, solving a problem is not enough; whatever is done, must be done within an economic context. This decision-taking process is the essence of modeling actual systems and it depends upon knowledge of what is or is not essential. By essential we can only mean whether it significantly affects the operation of the system in fulfilling its purpose. Clarity of purpose is thus vital, since reference to purpose supplies the only criterion for making the choice between essential and non-essential constituents (Spedding, 1993).

1.3 Approaches to modeling

To define a model as "an equation or set of equations which represents the behavior of a system" is both too specific and too general. It is too specific in referring only to mathematical expressions, and too general in relating the definition to a system. Models may be of many forms; they are not confined to mathematical versions, and may relate to only parts of systems.

For the study of quantified and complex relationships a mathematical expression is necessary. A practical application requires that the relationships are quantified, and that they adequately reflect the essential complexity of real life relationships. Spedding (Spedding, 1993) defines an adequate model as the simplest that will serve its purpose (which must be clearly specified in advance), and one that has been tested against the real-world behavior of the system modeled (to an accuracy determined in advance).

The difference between simulation and other traditional approaches to understanding a system, arises only because the concepts and data are transformed into mathematical equations which can be solved rapidly by computer to provide a quantitative and dynamic appraisal of the system. Such a rapid and thorough appraisal of complex biological systems is almost impossible to obtain from intellectual effort alone. There are two main parts to mathematical representation. The first is designing the form of the mathematical equation used to describe each component of the system. The second is parameterisation, that is, to establish the quantitative values for the constants within the equation (Black, 1993).

A model contains three essential components; variables, parameters and constants, and differential equations. Variables are the quantities that tend to change over time. There are four types of variables (France and Thornley,1984) : state variables, rate variables, auxiliary variables, and driving variables.

- A *state* variable is a quantity that helps define the state of the system at a given point in time; they are independent from the values of other variables at that point in time. For example, sow inventory at the beginning of the planning period.

•A *rate* variable is a quantity that defines a process within the system at a given time. Rate variables always have dimensions of quantity per unit time. For example, number of sows farrowed per month.

• *Auxiliary* variables are variables calculated from state variables that are used to assist in understanding the system and for comparison with measurements. Some auxiliary variables can be rates such as growth rate. For example, average female inventory per farrowing crate.

•*Driving* variables are data inputs to a model that vary with time; typical are those that describe the environment (temperature, wind). For example, average number of farrowing crates available per week.

Variables may be input to the model as data tables or they may be calculated by algorithms as a function of time (Black, 1993).

Parameters and constants are quantities appearing in the equations of a model that do not vary with time. They are arbitrarily distinguished on the basis of the reliability of their numerical value. Often parameters are adjusted to improve the goodness-of-fit between predictions and experimental results. This process is called model calibration or parameterisation of the model (Black, 1993).

Differential equations describe how the state variables of a model change with time. The number of differential equations in the model must equal the number of state variables (Black, 1993).

1.4 Computer models in livestock production

Computer models of livestock systems have been developed and used to help improve understanding of how production systems components interrelate, as tools to predict future performance, and to support decision making. By condensing several years' worth of production into just a few minutes, computer simulation techniques allow users to analyze interactions between systems components that otherwise would be impossible to carry out (Marsh, 1986; Damrongwatanapokin, 1993).

As management becomes predominant among the production factors, and production efficiency is paramount, management information systems (MIS) take on an ever increasingly important role among management tools. In spite of this, MIS provide support mostly to problem identification and partly to alternative delineation, which are but the first two steps in the decision process (Bohelje and Eidman, 1984). It is the combination of available MIS and computer simulation models that allows managers to complete the decision process, by getting insight on the potential consequences of different possible future alternatives courses of action (Jalvingh et al., 1992).

There are different types of models . They can be defined as either static or dynamic, deterministic or stochastic, and either empirical or mechanistic. There are also different levels of models, which correspond to the degree of detail with which they represent the system they portray. Models may range in their level from the cell level, to the herd level, to the farm level. At higher levels of representation (i.e.: farm or enterprise level) models may also differ in the scope of their work, in that they may either simulate, or they may optimize the functioning of the system they represent (Black, 1993). The higher the level of the simulation, and the more holistic the approach taken by the model builder, the higher the chances of finding several lower level models used as building blocks, then integrated into one seamless model (Harris, 1995).

The best animal models are likely to be an integration of mechanistic modules each developed by subject specialists. Research should be directed towards providing the concepts necessary for the replacement of any empirical equations used, with either mechanistic or conceptual equations (Black, 1993).

A herd model based on events occurring to individual animals makes it possible to follow the life histories of single animals. This approach provides a high level of realism to the model user, while providing the model builder with large volumes of detailed data for the purposes of testing, verification, and refinement of the model. Where the individual animal is chosen as the simulation unit, stochastic variables must be used in the prediction of discrete events. It is therefore necessary to make replicated runs in order to estimate expected mean values of output measures. This approach makes the use of the model more complicated and more expensive, but information regarding the variances around the means will contribute to the knowledge of the system behavior (Sørensen, 1993).

1.4.1 Static vs dynamic

A static model represents the state of a system at one instant in time. A dynamic model explicitly describes the behavior of a system over time. The time interval that elapses between consecutive updates of the system (the time step) varies from model to model, and greatly depends on the type (level) of application (Black, 1993). A static model is a simpler, more efficient way to predict the outcome of a simulated event under a narrow range, or prescribed set of values for the driving variables. But a static model cannot substitute for a dynamic model when the state of the system portrayed must be continually predicted over time.

1.4.2 Deterministic vs stochastic

In deterministic models no element of chance is considered, so results are determined solely by the value of a prescribed set of input variables (Marsh, 1986). Deterministic models have only one possible outcome from a calculation based on the particular values of a matrix of input values, Stochastic models provide a range of values for each outcome measure, which are typically summarized as means (representing expected values), and standard deviations (representing variance) (Black, 1993). (Where the underlying distributions of outcome variables are non-normal, or where the statistic of interest is non-parametric, then it is convenient to describe the range of possible outcomes using percentiles.)

Stochastic models have the advantage over deterministic models that variation between replicates can be quantified. This allows to asses the effect of varying levels of the driving variables on system output. Choice of type of model is subjective, and highly dependent on the objective of the modeling effort.

1.4.3 Empirical vs mechanistic

In empirical models, the interrelation between variables is depicted by equations that have no real association to the real mechanisms regulating the system portrayed. These models are often based on correlations and associations between variables, valid only under the study conditions they were derived from; these associations may have no implications about the mechanisms that control operation of the system. On the other hand, mechanistic models simulate the basic mechanisms that control the functioning of the system portrayed, such as flight simulators, which rely on the laws of physics to predict the outcome of events. Unfortunately there are some important predictions that cannot be based easily on concepts describing mechanisms. That happens when knowledge of factors determining the events is inadequate to develop a mechanistic concept (Black, 1993). For example, prediction of backfat in pigs relies on regression equations (empirical models) due to incomplete knowledge of all factors influencing fat deposition across all breeds and environmental conditions. Thus, an empirical model may be regarded as a valid first step in modeling a system, until research provides enough concepts to develop a mechanistic or conceptual representation.

The distinction between empirical and mechanistic models is clear at the individual level but less obvious at the farm and population level.

1.4.4 Level

The level of the model follows from the extent (degree) of detail with which the system is depicted. Some categorization has been put forward. Lloyd (Lloyd, 1989) points out that agricultural production processes were categorized into four levels by Dent (Dent, 1975). Dent's classification included biochemical and physical systems, plant and animal systems, farm business systems, and national and international systems. More recently France and Thornley (1984) set forth a hierarchy of animal systems that covers from biochemical reactions to the enterprise level, including metabolic functions, physiological functions, the individual animal and the flock.

Applied basic scientists will focus on the first two levels proposed by Dent (Dent, 1975), whereas management consultants will tend to use models constructed by social scientists at upper levels. There is no "better" level; there is always an adequate level depending on the perceived task the model must carry out, and what it is designed to accomplish. Nevertheless, it is paramount that the model be based one level below that where it is supposed to accurately predict events. This also implies that each level has its own language and concepts (Black, 1993).

1.5 Types of models related to swine

Given the different types of models available, as already described, and the wide range of aspects related to swine production, a great deal of modeling effort has taken place in the last decades; a trend that has gained impulse with the advent of more powerful and less expensive computer systems.

In the area of simulation, efforts have covered a wide array of aspects. Harris (Harris, 1995) mentions growth nutrition (Baldwin et al., 1979; Whittemore et al., 1976; Whittemore et al., 1981; Whittemore et al., 1983; Whittemore et al., 1986; Moughan et al., 1984; Moughan et al., 1987), life cycle genetics (Tess et al., 1983), environmental effects (Bruce et al., 1979; DeShazer et al., 1988; Christianson et al., 1982), reproduction (Allen et al., 1983; Pettigrew et al., 1986), herd dynamics (Singh, 1986), life-cycle nutritional and environmental concerns (Black et al., 1986), heat production and interaction with the environment in the individual pig, individual animal growth process (Pomar et al., 1991), individual animal reproduction process (Pomar et al., 1991b), and herd level simulation (Marsh, 1987; Pordesimo et al., 1993). Damrongwatanapokin (Damrongwatanapokin, 1993) mentions a few simulation models related to the epidemiology of diseases in swine (Smith and Grenfell, 1990; Grenfell and Smith, 1990; Rodriguez et al., 1990).

In the specific area of dynamic optimization, efforts in swine have been less numerous than for other species such as bovines (Zeddies, 1972; Renkema and Stelwagen, 1979; Gartner, 1981; Dijkhuizen et al., 1985; Van Arendonk, 1985; Kristensen, 1987; Sørensen, 1987; Spath et al., 1984), and concentrated to the optimization of replacement policies (Huirne et al., 1990; Dijkhuizen, 1985; Dijkhuizen et al. 1986).

In spite of the above mentioned body of literature related to modeling efforts, it persists a lack of information on considerations of quality and quantity of input data for the published models. Therefore the need for this particular study.

1.6 References

Allen MA and Stewart TS (1983). A simulation model for a swine breeding unit producing feeder pigs. Agric Syst, 10(4): 193.

Baldwin RL and Black JL (1979). Simulation of the effects of nutritional and physiological status on the growth of mammalian tissues: description and evaluation of a computer program. CSIRO Animal Research Laboratories Technical Paper N° 6. CSIRO, Melbourne.

Black JL (1993). In: Modelling growth in the pig. International Postgraduate

Training Course, Wageningen Agricultural University, The Netherlands.

Black JL, Campbell RG, Williams IH, James KJ and Davies GT (1986). Simulation of energy and amino acid utilization in the pig. Res Dev Agric 3:121.

Boehlje MD and Eidman VR (1984). Farm Management. John Wiley & Sons, New York.

Bruce JM and Clark JJ (1979). Models of heat production and critical temperature for pigs. Anim Prod, 28:353.

Christianson I, Hahnn GL and Meador N (1982). Swine performance model for summer conditions. Int J Biometeor, 26:137.

Damrongwatanapokin T (1993). Modeling the spread of infectious agents among and within swine herds: Minnesota Pseudorabies epidemics. Ph. D., University of Minnesota.

de Hoop DW (1989). Management processes in dairy and pig farming and the construction of systems. In: Korver S and van Arendonk JAM (editors): Modelling of Livestock Production Systems. Kluwer Academic Publishers, 1989.

dee-Roo G (1987). Stochastic model to study breeding schemes in a small pig population. Agric Syst, 25:1.

Dent JB (1975). The application of system theory in agriculture. In: Dalton GE (editor): Study of agricultural systems. Applied Science Publishers, London.

DeShazer JA, Milanuk MJ, Walt DL, Xin H, Vansteelant B and Ewan RC (1988).

NCCISWINE: The environmental and housing component. In: Livestock Environment III, Proc 3rd Int Livest Environ Symp, Apr. 25-28, 1988, Toronto. Amer Soc Agric Engineers, St. Joseph, MI.

Dijkhuizen AA, Renkema JA and Stelwagen J (1985). Economic aspects of reproductive failure in dairy cattle. II: the decision to replace animals. Prev Vet Med 3:265.

Dijkhuizen AA, Stelwagen J and Renkema JA (1986). A stochastic model for the simulation of management decisions in dairy herds, with special reference to production, reproduction, culling and income. Prev Vet Med, 4:273.

Dijkhuizen AA, Morris RS and Morrow M (1986). Economic optimization of culling strategies in swine breeding herds, using the "PorkCHOP" computer program. Prev Vet Med, 4:341.

Faust MA, Tess MW and Robison OW (1992). A bioeconomic simulation model for a hierarchical swine breeding structure. J Anim Sci, 70:1760.

France J and Thornley JHM (1984). Mathematical models in agriculture: a quantitative approach to problems in agriculture and related sciences. Butterworths, London.

Gartner J (1981). Replacement policy in dairy herds on farms where heifers compete with cows for grassland. Part I: model construction and validation. Agric Syst, 7:289.

Glen JJ (1983). A dynamic programming model for pig production. J Opl Res Soc, 34 (6):511.

Grenfell BT and Smith G (1990). Mathematical model for the impact of a pseudorabies epizootic on the productivity of a farrow-to-finish operation. Am J Vet Res, 51(1): 156.

Harris DL (1995). S.P.E.C.I.A.L. - Simulation model to support decision making in pork production. Proc FACTs '95 (Farm Animal Computer Technologies) Conference, Orlando, Florida: 11.

Harsh SB, Connor LJ and Schwab GD (1981). Managing the farm business, Prentice Hall Inc., New Jersey.

Huirne RBM, Hendricks THB, Dijkhuizen AA and Giesen GW (1990). The economic optimization of sow replacement decisions by stochastic dynamic programming. J Agric Econ, 39:426.

Huirne RBM, Hendricks THB, Dijkhuizen AA and Giesen GWJ (1988). The economic optimisation of sow replacement decisions by stochastic dynamic programming. J Agric Econ, 39:426.

Huirne RBM, Dijkhuizen AA and Renkema JA (1991). Economic optimization of sow replacement decisions on the personal computer by method of stochastic dynamic programming. Livestock Production Science, 16:131.

Jalvingh AW (1992). The possible role of existing models in on-farm decision-support in dairy cattle and swine production. Livest Prod Sci, 31:351.

Jalvingh AW, Dijkhuizen AA and van Arendonk JAM (1992). Dynamic probabilistic modelling of reproduction and replacement management in sow herds. General aspects and model description. Agric Syst, 39:133.

Kay RD (1986). Farm Management. Planning, control and implementation . McGraw Hill, New York.

Kristensen AR (1987). Optimal replacement and ranking of dairy cows determined by a hierarchic Markov process. Livestock Production Science, 16:131.

Lloyd JW (1989). Decision support for livestock production: integration of information management systems, modeling, and computer simulation techniques. Ph. D., University of Michigan.

Marsh WE (1986) Economic decision making on health and management in livestock herds: examining complex problems through computer simulation. Ph. D., University of Minnesota.

Morris RS and Marsh WE (1985). Quantitative methods for studying the epidemiology of virus diseases. In: Della-Porta Aj (Editor): Veterinary viral diseases: Their significance in South-East Asia and the Western Pacific. Academic Press, New York.

Moughan PJ and Smith WC (1984). Prediction of dietary protein quality based on a model of the digestion and metabolism of nitrogen in the growing pig. N Z J Agric Res, 27:501.

Moughan PJ, Smith WC and Pearson G (1987). Decription and validation of a model simulating growth in the pig (20-90 kg liveweight). N Z J Agric Res, 30:481.

Pettigrew JE, Cornelius SG, Eidman VR and Moser RL (1987). Integration of factors affecting sow efficiency: a modeling approach. J Anim Sci, 63:1314.

Pomar C, Harris DL, and Minvielle F (1991a). Computer simulation model of swine production system: I. Modeling the growth of young pigs. J Anim Sci, 69:1468.

Pomar C, Harris DL, and Minvielle F (1991b). Computer simulation model of swine production systems: II. Modeling body composition and weight of female pigs, fetal development, milk production, and growth of suckling pigs. J Anim Sci, 69:1489.

Pomar C, Harris DL, Savoie P, and Minvielle F (1991c). Computer simulation model of swine production systems: III. A dynamic herd simulation model including reproduction. J Anim Sci, 69:2822.

Pordesimo LO, Harris DL, Nienaber JA, Hahn GL an Korthals RL (1993). Proc 4th Int Livest Environ Symp, Coventry, England. IV: 945.

Renkema JA and Stelwagen J (1979). Economic evaluation of replacement rates in dairy herds. part I: Reduction of replacement rates through improved health. Livestock Production Science, 6:15.

Rodrigues Ca, Gardner IA and Carpenter TE (1990). Financial analysis of pseudorabies control and eradication in swine. J Am Vet Med Assoc, 197(10):1316.

Smith G and Grenfell BT (1990). Population biology of pseudorabies in swine. Am J Vet Res, 51(1):148.

Singh D (1986). Simulation of swine herd population dynamics. Agric Sys, 22:157.

Sørensen JT (1993). In: Korver S and van Arendonk JAM (editors):Modelling of Livestock Production Systems. Kluwer Academic Publishers,1993.

Sørensen JT (1987). Udvikling af en matematisk model af systemet, der producerer kælvekvier. Ph. D., The Royal Veterinary and Agricultural University, Copenhagen.

Spath EJA, Morley FMW and White DH (1984). A computer model of a self-replacing beef herd. Proc Aust Soc Anim Prod, 15:597.

Spedding CRW (1993). In: Korver S and van Arendonk JAM (editors):Modelling of Livestock Production Systems. Kluwer Academic Publishers,1993.

Tess MW, Bennett GL and Dickerson GE (1983). Simulation of genetic changes in life cycle efficiency of prok production. 1. A bioeconomic model. J Anim Sci, 56:336.

Usry JL, Turner LW, Stahly TS, Bridges TC and Gates RS (1989). GI tract simulation model of the growing pig. Paper American Society of Agricultural Engineers 89-7621:1.

Van Arendonk JAM (1985). Studies on replacement policies in dairy cattle. Ph. D., Wageningen Agricultural University.

Whittemore CT (1981). Animal production response prediction. In : GM Hillyer, Whittemore CT and Gunn RG (editors): Computers in animal production. Br Soc Anim Prod Occ Publ Nº 5, 47.

Whittemore CT and Fawcett RH (1976). Theoretical aspects of a flexible model to simulate protein and lipid growth in pigs. Anim Prod, 22:87.

Whittemore CT (1983). Development of recommended energy and protein allowances for growing pigs. Agric Sys, 11:159.

Whittemore CT (1986). An approach to pig growth modeling. J Anim Sci, 63:615.

Zeddies J (1972). Ökonomische entscheidungshilfen für die selektion in milchviehherden, Züchtungskunde, 44:149.

CHAPTER 2

Simulation model of a swine breeding herd unit

2.1 Introduction

PigORACLE© is one of two simulation models developed from a common underlying skeleton model (Marsh, 1986). The aims of that work were to (a) investigate the feasibility of the skeleton model approach to simulation model implementation; and (b) produce a series of species specific simulation tools to help study the effects of management decisions on the reproductive efficiency of livestock herds, and financial performance of farms.

In its original form, PigORACLE© simulation runs are limited to using only annual average values for the main driving variables. This supports study of the effects of reproductive performance on population dynamics and financial performance only when the simulated scenario does not involve seasonal fluctuations in productivity. This and other limitations have been overcome by changes introduced in the model's second development cycle, as detailed in chapter 3. The most significant change was the replacement of a single set of annual values for the main variables, by a matrix of monthly values. Our work was designed to find out whether this amount of "fine tunning" was really warranted. Modifications of functions used to predict the timing and occurrence of various events, are based on the analysis of PigCHAMP® records from 18 farms (hereafter referred to as "database"), which include 64851 parity records and 9928 removal events. Farm data files were selected from a larger set comprising respondents to a survey of PigCHAMP® users (Polson et al., 1992). Based on the integrity of their production records, data files were considered eligible if less than 5 % of their breeding or farrowing records were missing, incomplete or wrong, and if they had a stable herd size indicated by less than a 10% change in breeding female inventory.

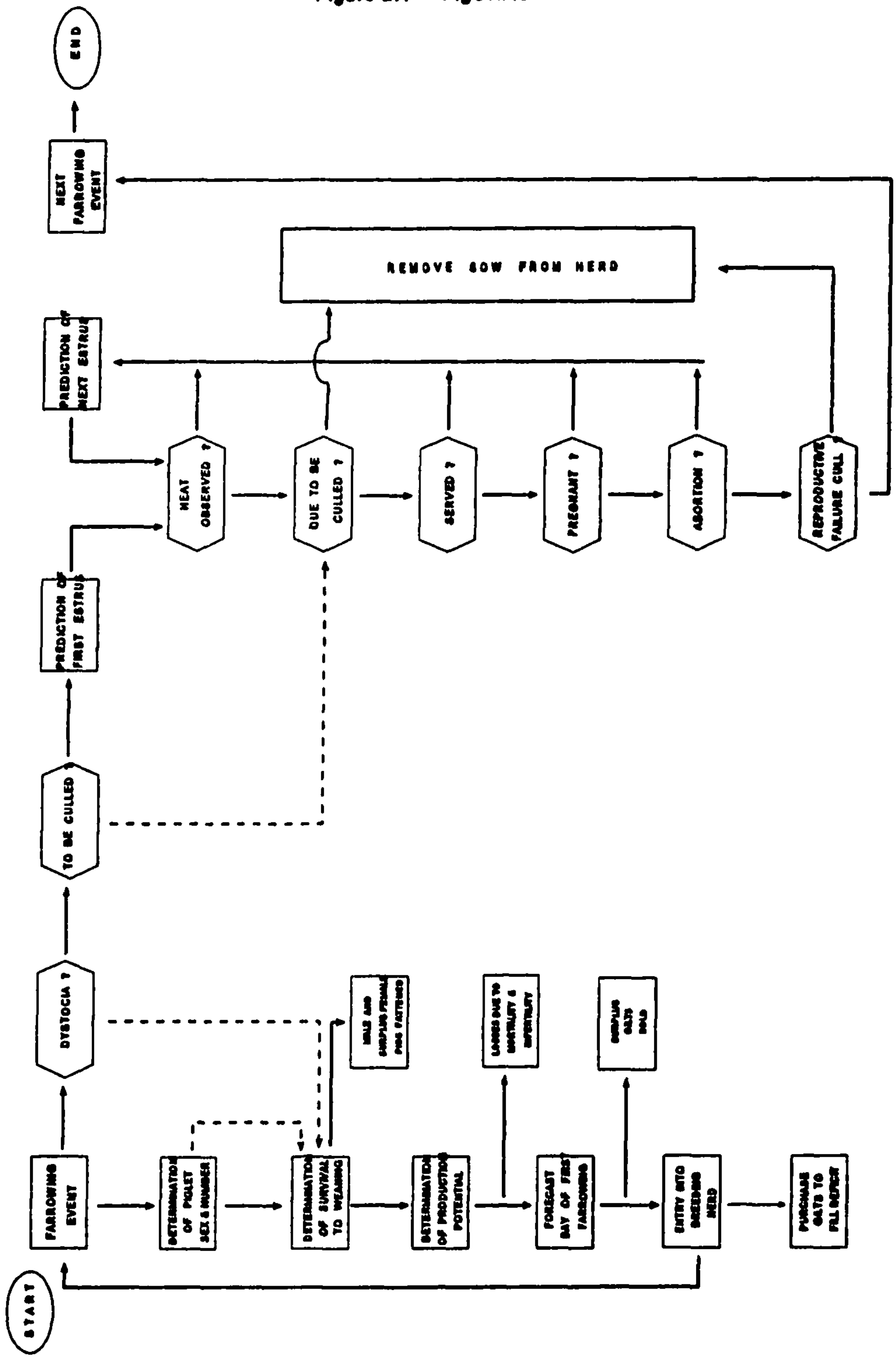
What follows is a brief summary of the main model characteristics as explained in Marsh (Marsh, 1986). This will serve as an introduction to the model and to the modifications introduced in this new development cycle, which are further detailed in the next chapter.

The model simulates the lives of individual female breeding animals; it maintains animal records which include 11 fields: (1) animal ID; (2) current reproductive status; (3) productivity index; (4) date of birth; (5) date of first service; (6) date of last conception; (7) current parity number; (8) date of last farrowing; (9) date of last weaning; (10) date due to farrow; and (11) date due to be culled. The animal record mimics those found in production record systems and is designed as such to interface directly with herd record systems such as PigCHAMP®. Unlike a true production record system, PigORACLE© keeps only current data; each field in each animal's history is updated throughout the simulation, as events are predicted.

The "status set" for breeding females is exhaustive and mutually exclusive; it includes the following possibilities: (1) gilt, selected for and entered to the breeding herd; (2) gilt, served; (3) gilt, diagnosed pregnant; (4) sow, lactating; (5) sow weaned; (6) sow served; (7) sow, served before weaning; (8) sow diagnosed pregnant; and (9) sow diagnosed not pregnant.

The flowchart in figure 2.1 summarizes information flow during simulation. The time step of the model is one calendar day. The algorithm checks the herd database each day and identifies any breeding females due to farrow that day. It then simulates a complete reproductive cycle for each qualifying animal. Each reproductive cycle is completed with either a future due-to-farrow date or predicted removal date. The occurrence and timing of reproductive, disease, and removal events are determined by random sampling from probabilistic distributions. The simulation model uses a Monte Carlo simulation technique, by taking random observations on the relevant distributions to determine the outcome of the particular event simulated.

Figure 2.1 - PigORACLE flowchart



2.2 Predictive functions for the breeding herd

2.2.1 Litter size

Initial approach - Mean total litter size has been reported to increase up to parity four, peak from parity four through parity seven, and decline afterwards (Rasbech, 1969; Penney, 1971; Peterson et al., 1980; Skjervold, 1975; Rybalko, 1975; Kroes and van Male, 1982; Hillyer, 1979; Joo and Kang, 1981). Thus, PigORACLE© allowed setting of different expected mean total litter sizes by parity. Total born litter size was determined through a random observation on a normal distribution, with a mean equal to the parity specified mean, and a standard deviation equal to one fourth the value of the mean. Litter size was restricted to positive values up to a maximum bound by a constant, set at 25, so that the random observation may not yield unusual values.

Modified approach - An analysis of the database is carried out to validate the approach taken to litter size determination. Results ranked according to the value of their chi square statistic (CS), show that the best fitting distribution for total born litter size is the Logistic distribution. However, if totalborn litter size is considered as a continuous variable and results are ranked according to the value of their Kolmogorov-Smirnov statistic (KS) , the Normal distribution shows a fit as good as that of the Logistic; its KS value and visual inspection of the fitted distribution differ very little from those yielded by fitting the Logistic distribution. In view of this, it is felt that the normal distribution provides a good approximation to reality for simulation purposes, with the added benefit of compactness and speed of program source code. The modified approach keeps the same general mechanism to determine total born litter size, but it allows users to specify parity specific average litter sizes by month, so seasonal variation may be better simulated as it may affect different parities differently.

2.2.2 Lactation length

Initial approach - Minimum lactation length was set by the user, to which a variable amount of days was added, such that all weanings occurred at a pre-set, specific day of the week. The model treated cross-fostering by means of a "fostering pool", where litters from sows that die before weaning were sent to complete the lactation period before being moved to the next phase. Sows with liveborn litter size equal to or less than three were immediately culled and the litter sent to the "fostering pool".

Modified Approach - In the modified approach, model users are prompted to specify a preferred weaning "day" of the week; options available include a specific day of the week, any day of the week, and any weekday (Monday - Friday). With this modification, the user specified minimum weaning age still acts as a "floor", but the number of days added to the lactation period after minimum weaning age is reached, depends on the weaning day preference set by the user.

2.2.3 Sow productivity index

Initial approach - The Ohio sow productivity index (Irvin et al., 1981) was calculated for all weaned sows, and used to rank them according to productivity for culling purposes. The index measures productivity as a function of number of pigs born alive, and the adjusted weaning weight of the litter as a proxy of the milking ability of the sow.

Modified Approach - No changes.

2.2.4 Culling and removal

Initial approach - To generate probability distributions for the timing of removal, removal reasons were divided into several categories as follows:

- lameness, injuries and degenerative problems;
- specific systemic diseases;
- miscellaneous problems;
- reproductive problems related to farrowing and litter;
- reproductive problems related to fertility.

Timing for removal for reproductive failure was unnecessary, since the model automatically removed sows that failed to conceive or complete a gestation successfully. Time of removal was determined through random observations on a family of Poisson distributions; this approach was derived from the assessment of very limited data available before the commercial release of PigCHAMP®.

Modified Approach - Underlying distributions for removal events have been reassessed and changed to be consistent with the analysis of the database. Event occurrences were tallied, probabilities calculated, and distributions fitted. In this process (explained in further detail in chapter 3), best fitting distributions were ranked. It was determined that the LogNormal distribution provides an adequate fit for simulation purposes, so timing of removal is determined through random observations on a LogNormal distribution with appropriate means and standard deviations for the various removal reasons.

2.2.5 Post weaning estrus and inter-estral intervals

Initial approach - Determination of time of estrus after weaning was carried out through random observations on a LogNormal distribution, with a default mean of seven days, and a standard deviation of one day. These parameters demarcated a distribution such that 90% of sows were predicted to show estrus within 8 days post weaning. The inter-estral interval was set by default at 21 days, with a 1.5 day standard deviation sampled on a Normal distribution.

Modified Approach - No changes.

2.2.6 Estrus detection and pregnancy rates.

Initial approach - Since different management styles determine different mating strategies, PigORACLE© allowed the user to input values for the probability of a sow being served at first estrus post-farrowing (default = 90%), and for the expected pregnancy rate (default = 80%).

Modified Approach - The mechanism is unchanged with the same default values. However, users are allowed to input monthly values for the percentages of sows served at first estrus, pregnant at first service, and pregnant at other services. This allows more flexibility in modeling seasonal fluctuations in the reproductive performance of the herd.

2.2.7 Re-breeding policy

Initial approach - Re-breeding policy was described by two variables. The first variable was the minimum sow productivity index (SPI) value a sow had to attain in order to qualify for re-breeding; the second variable was the maximum number of times a sow could be bred in a single parity before being culled. Thus, breeding decisions were carried out at two levels; at the first level, the decision to breed or cull after weaning was carried out based on the sow's current productivity level, and at the second level, the decision to re-breed was made depending on the number of prior failed services, and the maximum number of services allowed by management according to the sow's current productivity level. The program allowed the user to set different maximum number of services for low to average SPI sows, and for high SPI sows. This two tier system determined that low SPI sows be culled immediately postweaning, and other sows be bred as many times as allowed by their current relative productivity before being culled.

Modified Approach - Breeding criteria are unchanged, but maximum number of times a sow may be re-bred may be adjusted monthly. This allows a better simulation of management adjustments in breeding patterns to compensate for seasonal infertility problems.

2.2.8 Abortion

Initial approach - An abortion was the loss of all fetuses before the completion of the gestation period. Sows which were predicted to abort were culled immediately after the event. Prediction of abortion was determined by a random observation on a uniform distribution, which was compared to a probability of abortion set by the user in accordance with the production environment. Timing of abortion was determined by a random observation on an exponential distribution.

Modified Approach - The basic mechanism is kept unchanged, but probability of abortion is now set on a monthly basis. This allows the model to follow better any seasonal patterns of abortions.

2.2.9 Sow feeding

Initial approach - The model allowed for two types of sow feed (lactation and gestation), which were specified on a per-head per-day basis. When different feeds were utilized through each phase, their prices and levels of usage had to be compounded into a single figure to portray average pounds of feed per sow, per day, over a complete phase (lactation or gestation) of the production cycle.

Modified Approach - Allows users to specify monthly average intake values, as opposed to annual averages, which results in better estimates of monthly expenses.

2.3 Predictive functions for the replacement herd.

2.3.1 Piglet viability

Initial approach - The model predicted stillbirths and mummies on an individual litter basis, so as to mimic variability of mortality among litters. The users set the expected proportion of piglets born alive. An initial random observation on a Normal distribution, the mean of which was a function of parity of the dam, determined litter size. Once total born litter size was determined, a random observation on a uniform distribution was drawn for each pig in the litter, and the value compared to the proportion set by the user. Although the observation for each individual piglet was carried out on a uniform distribution, the end result for a whole simulation run (i.e. over a large number of litters simulated), was a distribution of mummies and stillborn per litter similar to that of a Poisson, with the majority of litters showing 0, 1 or 2 stillbirths, and very few showing four or more deaths. Pigs born dead included both stillbirths and mummies. Survival until weaning was determined through a similar procedure.

Modified Approach - The mechanism to determine viability is unchanged, but the user is allowed to set monthly values for the proportion of piglets born alive.

2.3.2 Pre-weaning mortality

Initial approach - Pre-weaning mortality levels were set with no adjustments for litter size, parity and fostering. Pre-weaning mortality was assessed individually for each born alive piglet, through a random observation on an uniform distribution. The process was similar to that of viability determination, where observations for each individual piglet were carried out on a uniform distribution, but the resulting distribution of pig deaths per litter across litters resembled a Poisson distribution. After weaning, the model assumed that piglets were pooled, and that both growth and mortality rates distributed uniformly across groups.

Modified Approach - Determination of piglet per-weaning mortality is similar, but the user is allowed to input monthly values for per-weaning mortality rate.

2.3.3 Nursery mortality

Initial approach - Nursery mortality level was set apart from pre-weaning mortality. The model assumed that piglets were pooled, and that both growth and mortality rates distributed uniformly across groups.

Modified Approach - No change.

2.3.4 Piglet growth rate and consumption (nursery phase)

Initial approach - Feed consumption was based on body weight, and it was estimated weekly. Feed conversion ratio could be modified to mirror general management conditions (housing, health status, etc.). With these parameters the model determined the growth curve, and pigs' development was simulated from birth until time of sale set by the specified average target weight.

Modified Approach - The mechanism to simulate pig growth rate and feed consumption is the same.

2.3.5 Replacement gilts

Initial approach - The flow of replacement gilts into the breeding herd is governed by the target number of sows and gilts to be mated each week; which in turn, is governed by weekly farrowing-crate availability, and anticipated farrowing rates. Sows not removed immediately post-weaning and predicted to shows estrus were determined a post-weaning interval (see 2.2.5). In any week, the difference between the number of sows mated and the target week matings was made up by introducing and mating replacement gilts. Thus the model assumed that gilts were readily available, and were drawn from an inexhaustible gilt pool. Gilts that failed to conceive were sold in a similar manner to cull sows. The day of first farrowing by a gilt was predicted through a random observation on a gamma distribution to determine time of conception, then adding an average 114 days gestation length and a random deviate.

Modified Approach - Treatment of replacement gilts in the modified model is unchanged.

2.4 Output reports

Initial approach - There were six different types of reports: demographic, reproduction, time series, performance indices, cash flow statement, income statement and livestock valuation reports.

The demographic report yielded an account of all animals in the herd at the time of the simulation period when the report is generated. The six reproductive performance reports covered the distribution of lactation lengths, weaning first service intervals, weaning to conception intervals, total born litter size, born alive litter size and pigs weaned per litter by parity. Time series reports provided monthly or weekly charts of sows farrowed, sows and gilts weaned, sows weaned, sows and gilts culled, gilts entered, total pigs born, pigs born alive, pigs weaned and feeder pigs sold. Performance Indices reports provided yearly values for selected breeding, farrowing and weaning performance indices, plus population related data and financial measures.

The three financial reports covered yearly income statements (total cash income, variable expenses, fixed expenses, net income and cash flows before and after taxes), quarterly cash flows, and a year end livestock valuation (breeding sows, replacement gilts and pigs).

Modified Approach - Output reports have been kept the same as in the first version of PigORACLE©. There is however a fundamental modification that changes both the way simulation results should be analyzed, and the information available from simulation data. The model now allows running multiple simulations (i.e. years) consecutively, and it also allows the user to specify the number of repetitions (i.e. farms) to perform; the model then creates an output file for each farm simulated, which contains herd data that can be read directly into PigCHAMP® using the data entry feature. With this change the stochasticity of the results is greatly enhanced; users can now run several repetitions or farms, read the output herds from those runs into PigCHAMP®, and run a multiple farms report (farm comparison) to summarize performance figures between the different runs. In this approach, with an adequate number of repetitions, the summary column of the farm comparison report provides a much better and effortless portrayal of the simular herd, smoothing the extremes that any one single simulation run may yield. This addition also opens the possibility of running all PigCHAMP® breeding herd reports on the simular herd, allowing access to a much wider array of analytical reports, in a more familiar format for PigCHAMP® users.

2.5 References

Hillyer GM (1979). Body condition and reproduction in pigs. In: Haresign W and Lewis D (Editors): Recent advances in animal nutrition. Butterworths, London.

Irvin K, Drewry K and Isler G (1981). Sow productivity indexing. National Swine Improvement Federation. Publication #NSIF4.

Joo HS and Kang BJ (1981). Reproductive performance on intensive swine farms in Korea. J Korean Vet Med Assoc, 17:40.

Kroes Y and van Male JP (1982). Reproductive lifetime of sows in relation to economy of production. Lives Prod Sci, 6:179.

Marsh WE (1986). Economic decision making on health and management in livestock herds: examining complex problems through computer simulation. Ph. D. dissertation, University of Minnesota, St. Paul.

Paterson AM, Barker I and Lindsay DR (1980). Analysis of the reproductive performance records of an intensive piggery in Australia. Proc Aust Soc Anim Prod, 13:389.

Penney RHC, Edwards MJ and Mulley R (1971). The reproductive efficiency of pigs in Australia with particular reference to litter size. Aust Vet J, 47:194.

Polson DD, Marsh WE and Dial GD (1992). A survey of North American swine breeding herds using the PigCHAMP® information system: I. Profiles of swine production system alternatives. Proc 12th IPVS Congress, August 17-20, The

Hague. pp 501.

Rasbech NO (1969). A review of the causes of reproductive failure in swine. Brit Vet J, 125:599.

Rybalko VP (1975). Sroki ispolzovaniya i vozrastnye izmeneniya produktivnosti suinomatok. Zhivotnovodstvo, 9:56.

Skjervold H (1975). Comparisons of litter size by rise of natural and artificial mating pigs. Z Tierz Zuchtgsbiol, 92:252.

CHAPTER 3

Use and fitting of probability distributions in stochastic modeling

3.1 Summary

A database of production records from 18 breeding swine herds was analyzed. The database comprised 64,851 parity records, 9,928 of which terminated with removal events. The purpose of the analysis was to develop a set of probability distributions which represent the timing and occurrence of important events in the reproductive lives of breeding females. Mathematical expressions were fitted to frequency distributions of empirical data describing total born litter size, weaning to first service interval, inter-estral interval and farrow to removal interval for nine main non-reproductive culling reasons. In most instances, up to five families of standard mathematical distributions were judged to adequately fit the empirical data. The most commonly fitted distributions were Weibull, LogNormal, Pearson, Erlang, and Beta.

3.2 Introduction

PigORACLE© is a dynamic stochastic simulation model of a swine breeding herd (Marsh, 1986). Complete reproductive cycles for individual breeding females are simulated. Each complete reproductive cycle begins with a farrowing event and ends with either a predicted farrowing or removal date (Fig. 3.1). As the simulation proceeds through the cycle, the occurrence and timing of each event is determined by drawing random observations from appropriate probability distributions.

The model time step is one calendar day. At each simulated day, the record of every breeding female is examined and checked for a due-to-farrow or removal event which has been previously predicted to occur on that date. If the predicted event is a removal, the animal is removed from the herd (effective of that date), its record deleted, and herd demographics adjusted accordingly. If the predicted event is a farrowing, the model proceeds to simulate the next production cycle for that particular animal.

Determination of the occurrence, timing, and magnitude of each simulated event predicted to occur during the production cycle are determined stochastically by taking random observations on mathematically defined probability distributions. The exact shape of each distribution is determined by its mathematical form, modified by the values (e.g. mean, standard deviation) of several management variables which are under the control of the person running the simulation.

At each simulated farrowing event, the number, sex, and the extent of peri-natal mortality of total born pigs are determined. The number of total born pigs is determined by taking the integer portion of a random observation from a truncated Normal distribution having a mean and standard deviation chosen to represent litter sizes of females of particular parity numbers. Determination of whether each pig is born alive is made by taking random observations on a uniform distribution ranging between zero and one. For example, if the model is set to simulate 90% of total born pigs being born alive, the random observation for each pig is compared with the value of 0.9000. Pigs drawing numbers in the range 0.0000 to 0.9000 are considered "born alive"; those drawing numbers above 0.9000 are considered "born dead". Sex of born alive pigs is determined in a similar fashion. Thus, in extreme cases in individual litters, the number of "born dead" pigs may be zero or all. However, across many simulated litters, the proportion of live born pigs will be 90%, and the spread of perinatal mortality rates among litters will mimic the pattern of variability among litters in a real herd. Pre-weaning mortality is also determined on an individual pig basis. Thus, simulated weaned litter sizes are calculated as live born litter size minus the number of Pre-weaning deaths. No attempt is made to model cross-fostering. Doing so would add a considerable amount of unnecessary complexity which would not significantly improve simulation of weaned pig production.

Once weaned pig production has been simulated, the model initiates simulation of estrus events. As estrous behavior in breeding pigs is initiated by weaning, the timing of the first estrus (weaning to estrus interval) is predicted by drawing a random observation on a LogNormal distribution. The (integer) number of days is added to the weaning date to yield the predicted estrus date. Mating however, is not automatically assumed. Whether a service will take place or not will depend on the probability of estrus being detected, and perhaps, the time since farrowing. Following a mating event, determination of pregnancy depends on the value of a random observation from a uniform distribution based on the pregnancy date set for the herd for that time period. If conception is deemed to have taken place, then the predicted farrowing date is forecast by adding the mean gestation length (plus or minus a random deviate) to the date of conception. If a breeding female fails to conceive at the first estrus post-weaning, then its next estrus is simulated by sampling from a normal distribution (e.g., N (21, 1.5)) which would typically predict the next estrus event to occur in range of 17-24 days later. Animals returning to estrus may be mated or culled, depending on the setting of the repeat service management variable. Animals presumed pregnant may also abort.

3.2.1 Predictive functions

At the time of PigORACLE's first development cycle, frequency distributions used in the prediction of reproductive, health and removal events were described on the basis of analysis of a very limited data set. Ten years later, the PigCHAMP® research database comprises millions of breeding female parity records from a wide array of production systems across North America. Animal records in this database were analyzed to construct more realistic predictive distributions for incorporation in the simulation model.

The choice of the most suitable predictive distributions was determined on a trial-and-error basis. The process involved fitting likely mathematical expressions to frequency distributions of empirical data derived from events recorded in individual breeding female records in PigCHAMP® files. Once programmed into PigORACLE©, the simulation model uses a process of sampling random observations from these probability distributions to determine the occurrence, timing, and if appropriate, the magnitude of events occurring in the life cycles of breeding females. The better the fit of the mathematical equations to the empirical distributions, then the more realistically the simulated data will mimic the real world situation. In Fig. 3.1, the numbered circles indicate those steps where random numbers are drawn from specific probability distributions in order to simulate real-world variation in the occurrence of events.

The main stochastically simulated events and the characteristics of the predictive functions used in their determination are listed in Table 3.1.

Figure 3.1 - PigORACLE® flowchart

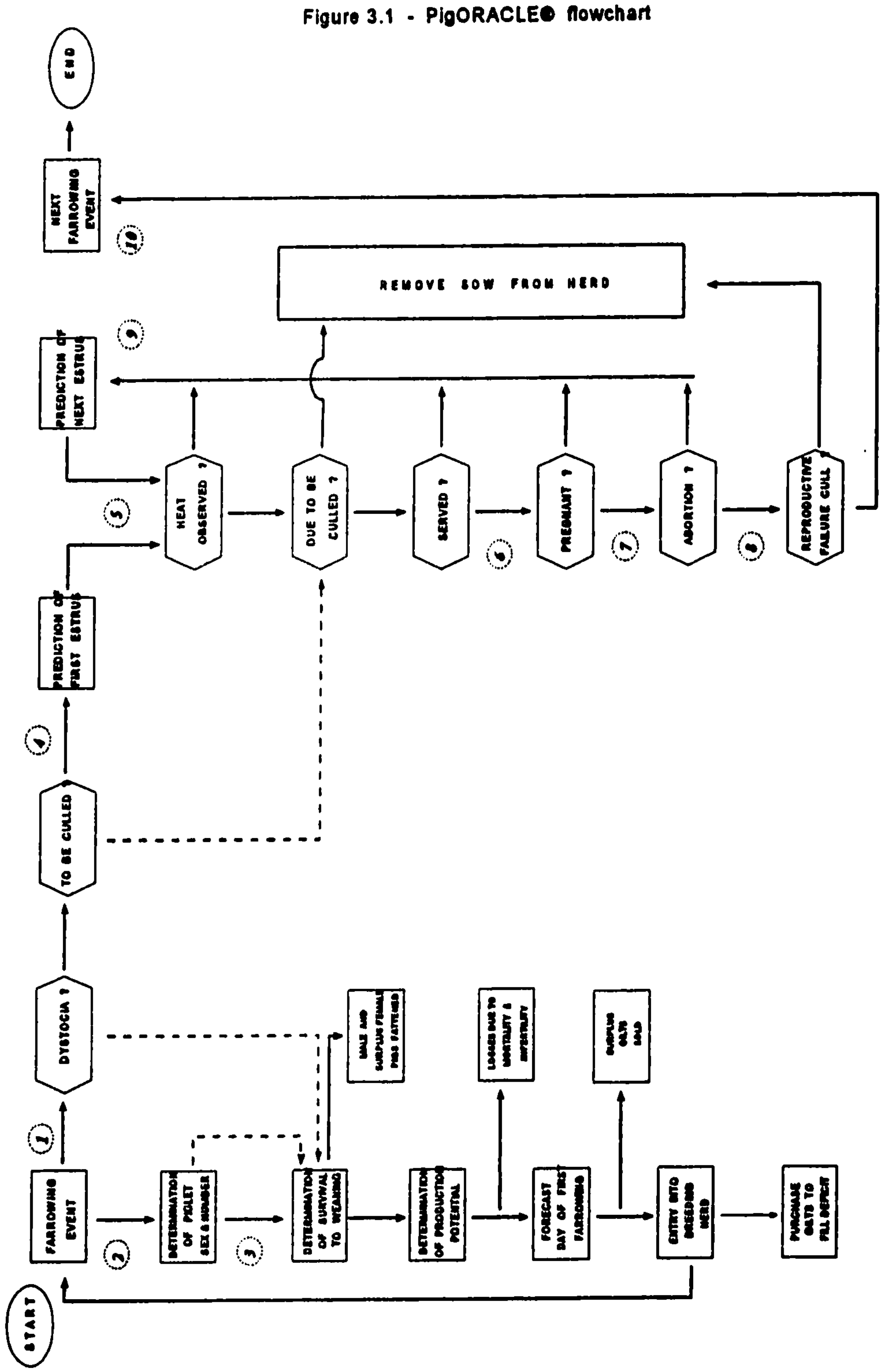


Table 3.1 - Main stochastic events simulated and their associated distributions

Event or Interval		Unit or Outcome	Type	Distribution	Typical Setting
1	Dystokia	Y / N	Binary	Uniform	Prob ≤ .055
2	Total born litter size	pigs	Discrete	Normal (μ, σ)	(10.8, 2.95)
3	Pig born alive	Y / N	Binary	Uniform	Prob ≥ .90
4	Farrow ->Removal	days	Continuous	Lognormal (μ, σ)	Variable
5	Weaning -> 1st Service	days	Continuous	Lognormal (μ, σ)	(5.5,1.5)
6	Detected in estrus and served	Y / N	Binary	Uniform	Prob ≥ .90
7	Conception	Y / N	Binary	Uniform	Prob ≥ .90
8	Abortion	Y / N	Binary	Uniform	Prob ≤ .025
9	Inter-estrus interval	days	Continuous	Lognormal (μ, σ)	(21.0, 3.0)
10	Gestation length	days	Continuous	Normal (μ, σ)	(114.0, 3.0)

3.3 Materials and Methods

PigCHAMP® production records from 18 herds were assembled as an aggregate database comprising 64,851 parity records with 9,928 removal events. Herd files originated from a larger set made up of the respondents to a PigCHAMP® users survey (Polson et al., 1992), and were selected for this analysis based on data integrity and completeness. Selection criteria included less than 5% missing mating or farrowing events, with special emphasis on completeness of removal records, and stable breeding female inventory.

Breeding females in the PigORACLE© model that experience reproductive failure are automatically removed following a prescribed number of failed matings, an abortion event, or a failure to farrow. Therefore it was not necessary to fit mathematical expressions to predict those events. Thus, only parity records containing removal events classified as non reproductive culls, deaths or destroyed animals were considered. This reduced the total number of events analyzed to

4,983.

Qualifying records were identified and exported to ASCII data files using the *Database Applications* feature in PigCHAMP® ver. 3.05 for DOS[1]. Data abstracted from each of 18 herd files were merged into a single data file which was imported into Statistix® ver. 4.0 for DOS [2]. Frequency distributions generated from Statistix® were read into BestFit ® ver. 1.02 for Windows [3] to analyze and determine the best fitting distributions for each event under study.

3.4 Results and Discussion

3.4.1 Best fitting distributions

Results of the analyses for the three best fitting mathematical distributions for the empirical data corresponding to each event are shown in Table 3.2 . Results are ranked according to goodness of fit, as indicated by the appropriate coefficient (Kolgomorov - Smirnov for continuous variables, or Chi - Square for discrete critical expressions superimposed upon them. Values in the column headed "Classes" indicate the number of discrete frequency values or "bins" that were observed for each variable considered.

---

[1] University of Minnesota, CAPS, 1365 Gortner Ave., St. Paul, MN 55108, USA.

[2] Analytical Software, PO Box 13204, St. Paul, MN 55113, USA.

[3] Palisade Corporation, 31 Decker Road, Newfield, NY 14867, USA.

Table 3.2 - Ranking of best fitting distributions for all variables analyzed

Variable	Classes	Distribution Rank									
		# 1	Coeff (*)	# 2	Coeff (*)	# 3	Coeff (*)	# 4	Coeff (*)	# 5	Coeff (*)
Continuous											
Inter Estral Interval	23	Ext. Value	0.057	Lognorm	0.077	Normal	0.064	Pearson V	0.064	Beta	0.112
Weaning -> 1st Service	16	Inv. Gaussian	0.052	Lognorm	0.054	Lognorm2	0.054	Pearson V	0.063	Ext. Value	0.065
Farrow - Removal Interval due to:											
- Abortion	83	Pearson VI	0.031	Ext. value	0.043	Inv. Gaussian	0.047	Weibull	0.057	Beta	0.060
- Downer sow	23	Lognorm 2	0.054	Gamma	0.067	Lognorm	0.071	Weibull	0.072	Ext. Value	0.069
- Gastro-intestinal	14	Beta	0.025	Weibull	0.155	Exponential	0.173	Erlang	0.209	Logistic	0.209
- Injury	28	Pearson V	0.071	Pearson VI	0.078	Inv. Gaussian	0.081	Lognorm	0.085	Beta	0.098
- Lame/Unsoundness	38	Lognorm	0.047	Inv. Gaussian	0.071	Erlang	0.081	Weibull	0.083	Gamma	0.087
- Mastitis	11	Gamma	0.079	Lognorm 2	0.082	Exponential	0.098	Erlang	0.101	Weibull	0.125
- Old Age	38	Pearson V	0.080	Log Logistic	0.088	Lognorm	0.090	Lognorm	0.090	Pearson VI	0.091
- Rectal/Uterine Prolapse	22	Pearson V	0.074	Weibull	0.089	Erlang	0.089	Lognorm	0.093	Log Logistic	0.101
- Unthrifty	25	Beta	0.094	Pearson VI	0.105	Pearson V	0.105	Exponential	0.110	Weibull	0.112
All culls	69	Pearson V	0.044	Lognorm	0.080	Inv. Gaussian	0.081	Pearson V	0.083	Ext. Value	0.085
Discrete											
Total Born	28	Logistic	0.014	Neg Binomial	0.131	Poisson	0.140	Exponential	1.928	Geometric	2.011

(*) all coefficients indicate non significant differences between real and fitted distributions

Reading across the table, the values of the coefficients for each variable share the same order of magnitude in most cases. This may be interpreted as the fit of all three distributions shown should be regarded as being equivalent. Comparison of these with the distributions used in the original version of PigORACLE©, shows that LogNormal was a good choice to use for prediction of interestral intervals; as it is ranked second for that variable, with no significant difference in goodness of fit with respect to the first ranked. Conversely , the use of the Poisson distribution to predict farrow to removal intervals was inappropriate, as it is not ranked in the top three distributions for farrow-to-removal intervals for any of the nine removal categories.

On balance, results of the analyses using the Bestfit software indicate that the LogNormal distribution is the predominantly appropriate distribution for the continuous variables. It is second best fit for the Inter Estral and Weaning - First Service Intervals, and it also appears within the three best fitting distributions in four out of nine types of Farrow - Removal Intervals analyzed, and three times in fourth place among the five other types; in all cases the fit is significant. These results, and the ease of conceptual understanding of the underlying mathematical expression, lead to the adoption of the LogNormal distribution for the simulation of the timing (days post-farrowing) of all removal events in the model, and for the inter estral interval. Section 3.1 in the appendix to this chapter includes a graphical comparison of the best fitting distribution for the data of each event considered, and the LogNormal distribution. This graphical comparison allows a visual reassurance of the appropriateness of the LogNormal as the distribution of choice for simulation modeling purposes.

Total born litter size was the only discrete variable analyzed using Bestfit. Preliminary analysis of 64,851 farrowing records with total born litter sizes ranging between 0 and 35, showed the Logistic distribution to be the best fitting expression, followed distantly by the Negative Binomial and the Poisson distributions. However, close inpection of the frequency distribution of total born litter sizes revealed that only 6 cases (0.00925% of total observations) were in the range of 26-35.

Since the original code of the PigORACLE© model restricts the values of predicted total born litter sizes to integers in the range 0-25, data from 64,845 farrowing records consistent with those values was re-analyzed. For this data set, the Logistic and Normal distributions were best fitting, with Kolmogorov-Smirnov coefficients of 0.011 and 0.012, respectively. In this case, the third-best fitting distribution (Poisson) was by far a less suitable fit to the data with a coefficient of 0.055.

This exercise illustrates the important distinction between identifying and ranking the best fitting distributions, and the selection of the distribution to be used to simulate real-world events. These may necessarily be the same for practical reasons such as the availability of a computer code to reliably generate random observations on known mathematical distributions, as well as the rapid execution of that code. In the case of total born litter sizes there was no way to validate the very few - less than 0.01% - extremely large recorded litter sizes in excess of 25. Some or all of them may very well have resulted from PigCHAMP® data entry errors, as total pigs born are calculated as the sum of liveborn, stillbirths, and mummies. As fitting the Normal distribution to 99.99% of the observations in the empirical data set resulted in a virtual tie with the Logistic distribution for best fit, choice of the Normal distribution has a clear practical and conceptual advantage over the Logistic distribution, and thus is preferred.

The issue of goodness of fit as opposed to practical or conceptual advantages is prevalent throughout the analysis. The Bestfit software permitted trial-and-error fitting of numerous distributions for each event, which resulted in several distributions having similar characteristics being reasonable approximations to the empirical data. This could be expected since there are many distributions whose mathematical expressions really cover a family of distributions, with various, very different shapes, depending on the value of their shape parameters (e.g.: Beta, Gamma, Weibull). These more "flexible" distributions were expected to be ubiquitous among the best fitting for the different events, but their better fit must be analyzed under the light of computational ease and conceptual suitability. From this point of view, the choice of the LogNormal distribution for the continuous variables is quite adequate; it provides an overall fit which is good enough for the type of data involved, it simplifies and speeds computations, and it is conceptually more suitable than the rest of the top ranking distributions.

3.4.2 Probabilities of events

In conjunction with the work on distributions, the main removal events of interest for simulation purposes were tallied from the database. Removals were categorized as deaths or destructions, and culls. Culls were subsequently subdivided into eight main subcategories, including old age, lameness or unsoundness, injuries, unthriftiness, downer, prolapse, mastitis, and other causes. Other causes encompass a wide array of culling events, including behavior problems, size, urogenital problems, ulcer, vulvar discharge, multiple systems, gastro intestinal, respiratory, metritis, central nervous, etc.

The results are presented in Table 3.3 as a quantitative contribution to the understanding of the main non-reproductive removal causes in swine breeding herds. The table summarizes the probabilities found associated to each type of removal event by parity, derived from the analysis of 4,983 removal events not related to reproductive reasons from the study database. Results are in agreement with those found by Lucia (Lucia, 1993), and confirm that, as expected, culls are more prevalent in first parity sows and old sows.

Table 3.3 - Probability of removal by parity and removal type for 4,983 removal events

	Parity							
Removal Type	All (%)	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	7+ (%)
Death / Destroy	7.2	7.7	9.7	9.6	12.6	8.2	8.4	3.8
Cull :	92.8	92.3	90.3	90.4	87.4	91.8	91.6	96.2
- *Old age*	30.7	—	—	—	5.4	16.4	39.0	73.6
- *Lame/Unsound*	19.9	23.4	30.0	34.2	34.6	27.1	17.5	6.8
- *Injury*	5.8	6.0	7.8	10.5	5.9	10.2	4.8	3.2
- *Unthrifty*	3.6	5.1	4.8	6.8	5.1	4.6	3.5	0.9
- *Downer*	2.0	2.2	3.8	4.2	2.6	2.8	2.0	0.6
- *Prolapse*	1.3	0.7	1.5	2.2	1.5	1.5	3.8	0.7
- *Mastitis*	1.1	1.0	2.1	1.1	1.8	1.5	1.5	0.5
- *Other*	28.3	53.8	40.3	31.4	30.5	27.6	19.5	10.0

3.4.3 Comparing non-reproductive culling reasons throughout parities

Data in Table 3.3 were contrasted with the use of the CATMOD procedure in base SAS® software [4]. The Statistical analysis shows that, as expected, there is an effect of parity on age since culling due to age does not take place until parity 4. After parity 4 there are no statistical differences between age and other non-reproductive culling reasons.

In spite of less evident number differences, there are also parity effects for most of the other non-reproductive culling reasons. The only non-reproductive culling reason where parity effects are less evident is mastitis, for which parity effect is significant at the 5% probability level, but not at the 1% level.

Lameness appears more frequent during the first parities, and declines after the fifth. Injuries are more or less ubiquitous throughout all parities, with statistically

[4] SAS Institute Inc., Box 8000, Cary, NC 27511-8000.

significant peaks in parities 3, 5 and older sows. Unthriftiness and downer sows appear equally distributed throughout parities 1 to 6. Prolapses are also equally common throughout most parities but for a small low peak in parity 6 which is statistically different from culling levels in parities 1, 2 and older sows. Finally, mastitis is also quite evenly distributed throughout all parities, with the exception of older sows, which appear to be statistically less affected than 2, 4, 5 and 6 parity sows.

3.4.4 Comparing non-reproductive culling reasons within each parity

The analysis of how reason affects non-reproductive culling within each parity shows, as expected, that it has an effect in all parities. The main effects are introduced by age, which is significantly different than all other causes in parities 1 through 4, and in older sows, but only significantly different from downer, lame/unsoundness, injuries and unthriftiness in parity 5, and from lame/unsoundness or prolapses in parity 6.

Other than age there are few cases of culling reasons significantly different from any other within any one parity. Among them: prolapses are significantly different from unthriftiness in parity 1, prolapses are also significantly different from all other causes (but mastitis) in parity 6, and injuries are significantly different from lameness, unthriftiness and prolapses in parity seven.

So overall, age appears as the main differential non-reproductive culling cause throughout most parities, while most other causes tend to show similar weight within any one parity.

3.5 Conclusion

Based on analysis of real-world data, we are confident that the improved model provides a more realistic simulation of the timing and occurrence of reproductive and culling events. Some of the original guesses were adequate, while others, like use of the Poisson distribution, were inadequate.

It is recommended that in the future, real world data be checked to ensure that simulation techniques are appropriate for the existing production systems. This is especially important in the case of early weaning and multisite production.

3.6 References

Analytical Software. Statistix® User's Manual: Statistix®, Version 4.0, 1992.

Lucia T (1997). Lifetime productivity of female swine. PhD dissertation, College of Veterinary Medicine, University of Minnesota, St. Paul.

Marsh WE (1986). Economic decision making on health and management in livestock herds: examining complex problems through computer simulation. Ph. D., University of Minnesota, St. Paul.

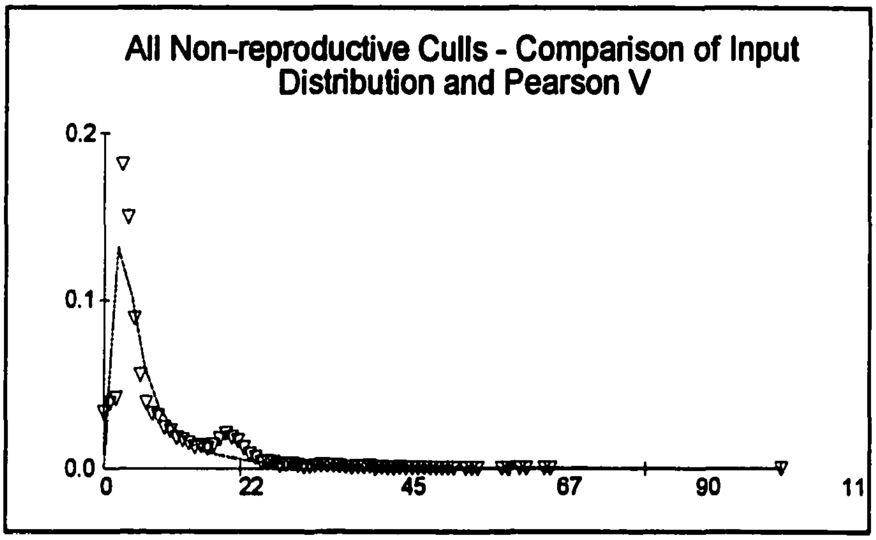
Palisade Corp. BestFit© User's Manual: BestFit© Distribution Fitting Software for Windows, Release 1.02, 1994.

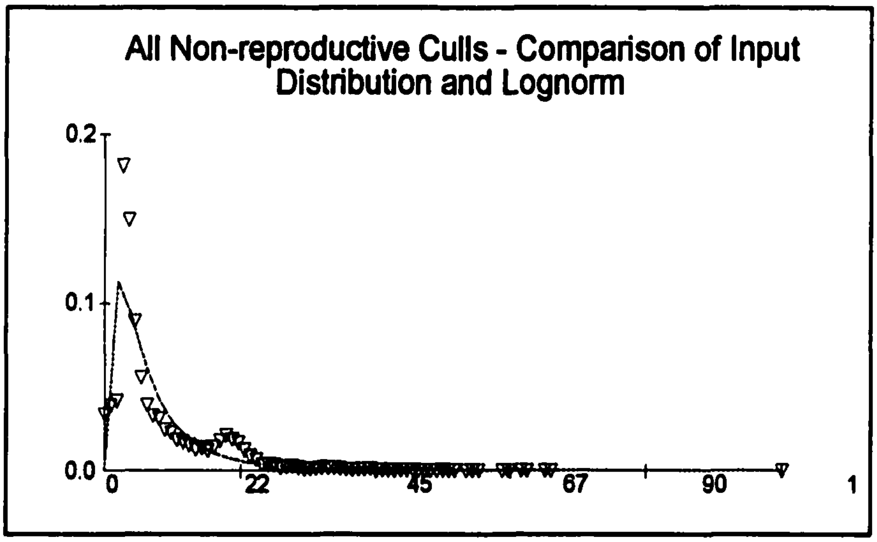
Polson DD, Marsh WE and Dial GD (1992). A survey of north american swine breeding herds using the PigCHAMP® information system: I. Profiles of swine production system alternatives. Proc 12th IPVS Congress, August 17-20, The Hague, pp 501.

SAS Institute Inc. SAS® User's Guide: Statistics, Version 5 Edition. Cary, NC: SAS Institute Inc, 1985. 956 pp.

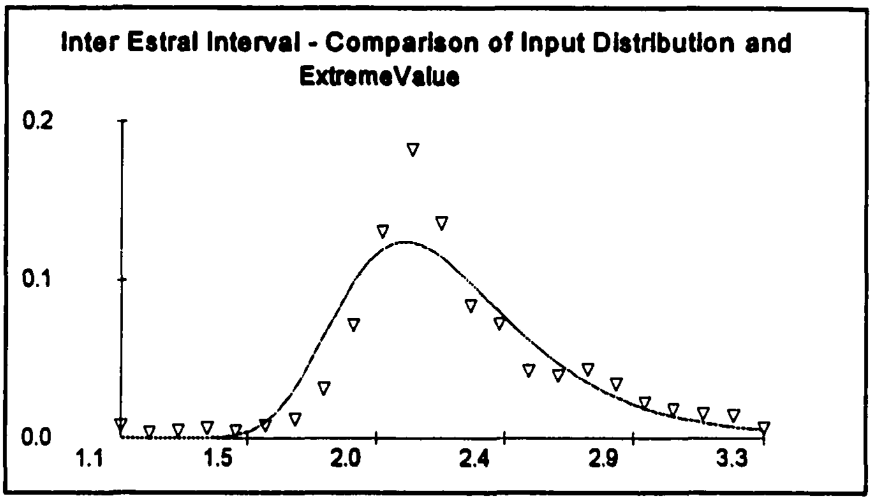
Annex Chapter 3

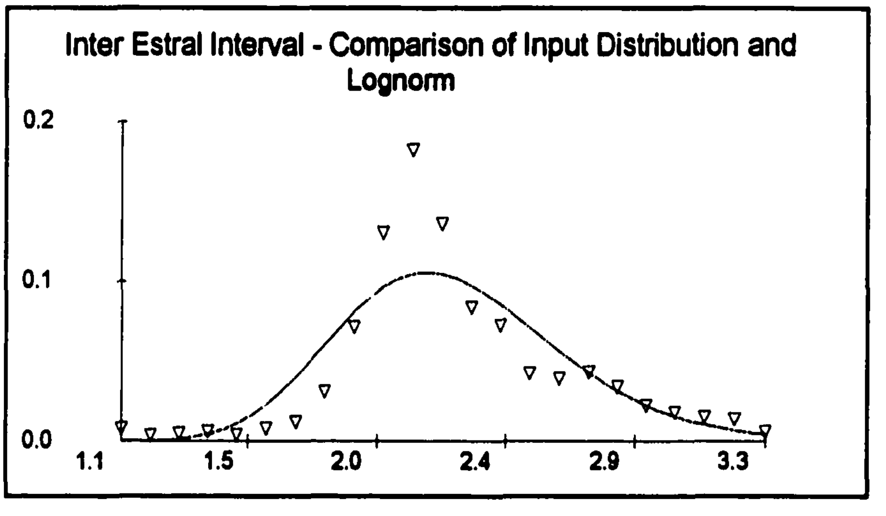
3.1 Graphical[1] comparison of best fitting and chosen distributions for selected events

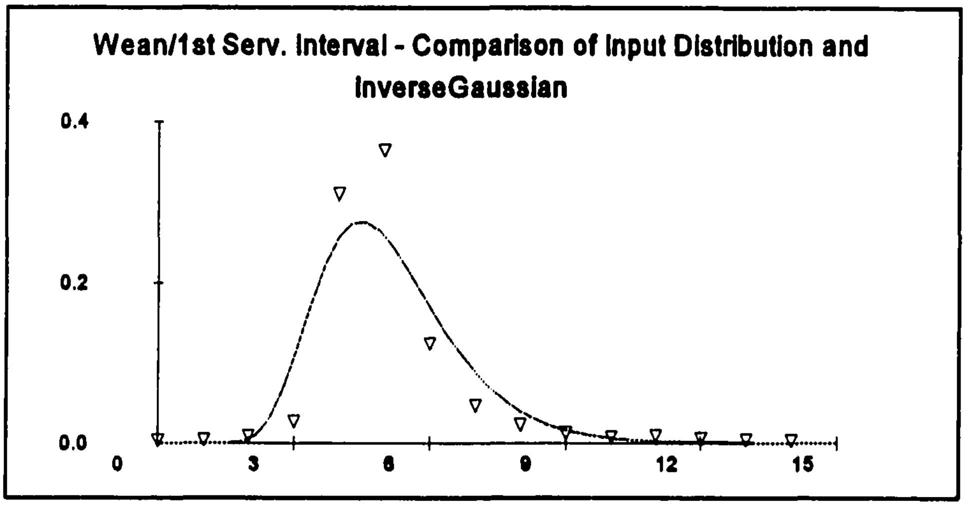


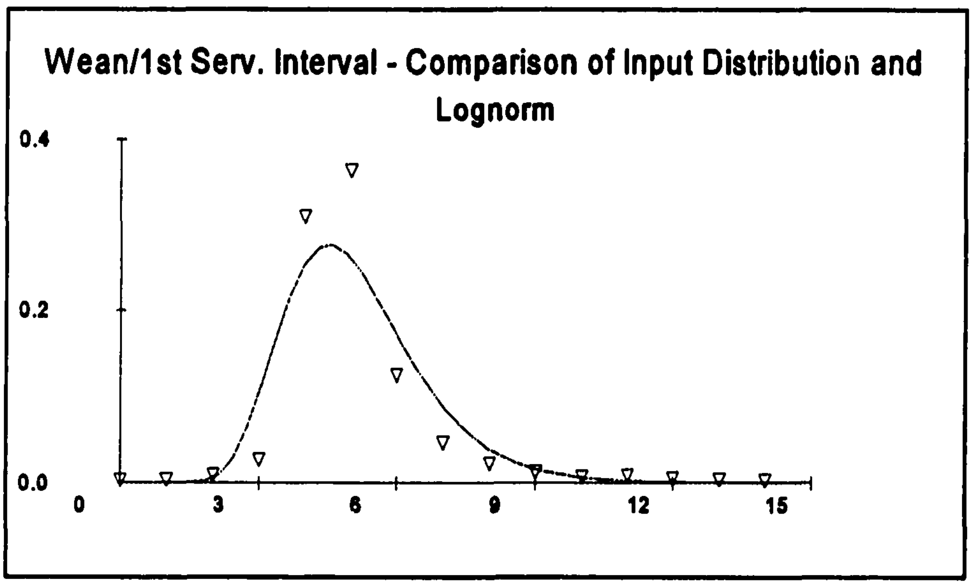


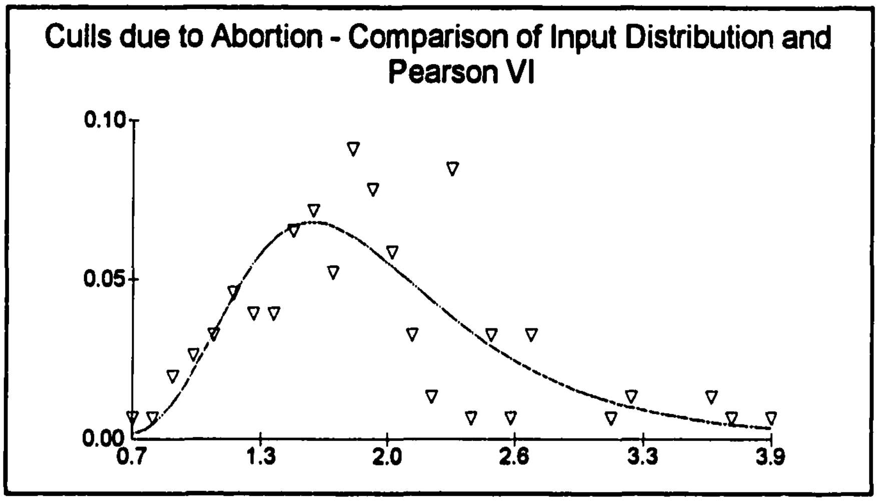
[1] All Y axes represent proportions. All X axes express days after farrowing. Triangles mark data points (observations) and show the input distribution. Continuous lines show the particular fitted distribution.

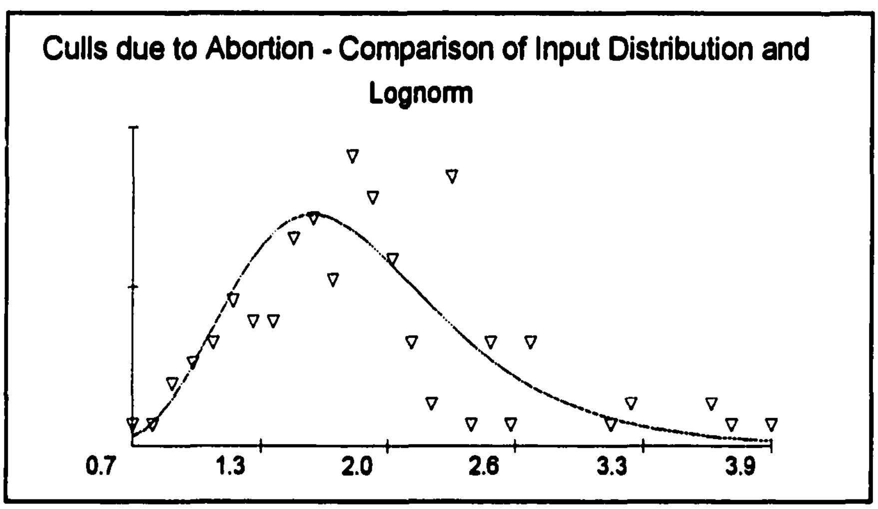

Inter Estral Interval - Comparison of Input Distribution and ExtremeValue
0.2
0.1
0.0
1.1
1.5
2.0
2.4
2.9
3.3

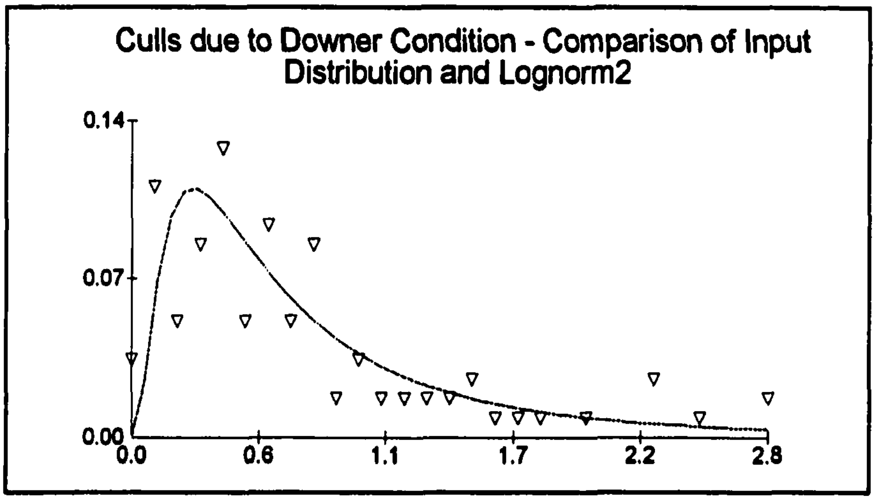

Inter Estral Interval - Comparison of Input Distribution and Lognorm
0.2
0.1
0.0
1.1
1.5
2.0
2.4
2.9
3.3

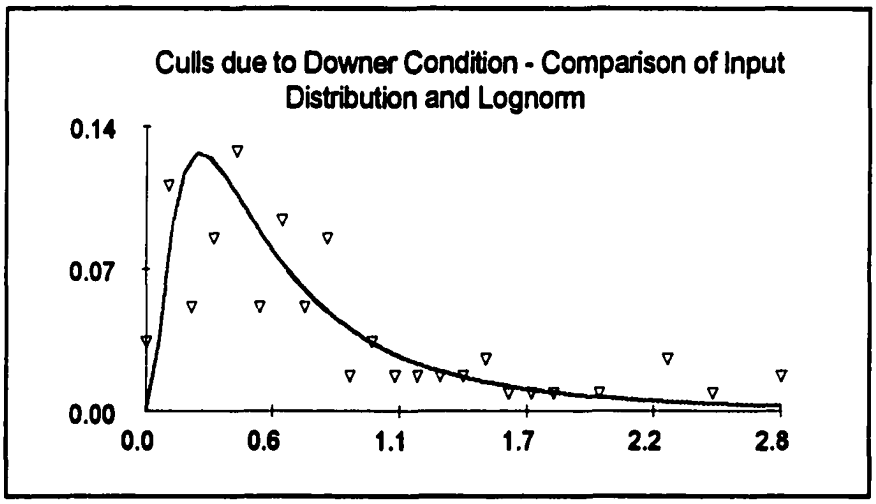

Wean/1st Serv. Interval - Comparison of Input Distribution and InverseGaussian
0.4
0.2
0.0
0
3
6
9
12
15

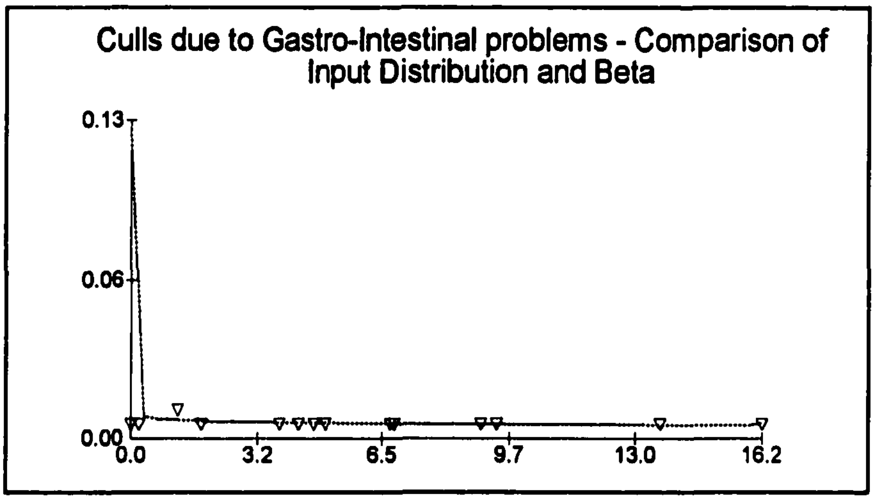

Wean/1st Serv. Interval - Comparison of Input Distribution and Lognorm
0.4
0.2
0.0
0
3
6
9
12
15

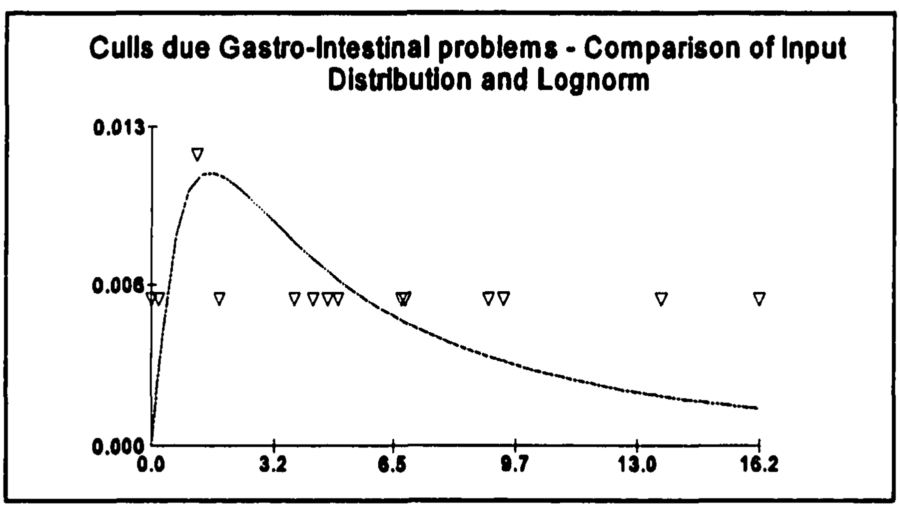

Culls due to Abortion - Comparison of Input Distribution and Pearson VI
0.10
0.05
0.00
0.7
1.3
2.0
2.6
3.3
3.9

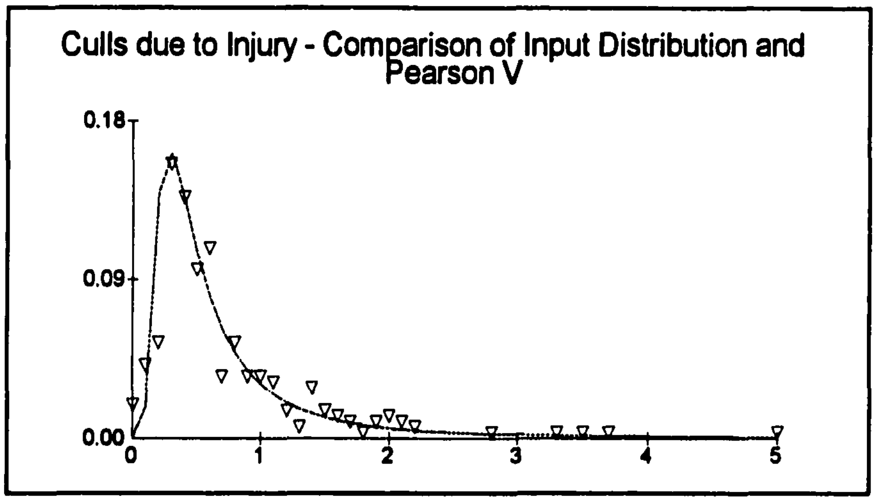

Culls due to Abortion - Comparison of Input Distribution and Lognorm
0.7
1.3
2.0
2.6
3.3
3.9


Culls due to Downer Condition - Comparison of Input Distribution and Lognorm2
0.14
0.07
0.00
0.0
0.6
1.1
1.7
2.2
2.8

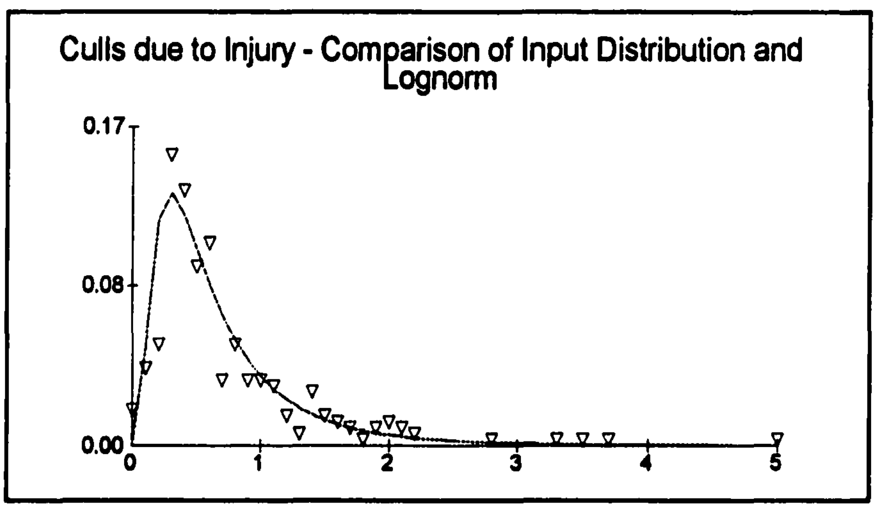

Culls due to Downer Condition - Comparison of Input Distribution and Lognorm
0.14
0.07
0.00
0.0
0.6
1.1
1.7
2.2
2.8


Culls due to Gastro-Intestinal problems - Comparison of Input Distribution and Beta
0.13
0.06
0.00
0.0
3.2
6.5
9.7
13.0
16.2


Culls due Gastro-Intestinal problems - Comparison of Input Distribution and Lognorm
0.013
0.006
0.000
0.0
3.2
6.5
9.7
13.0
16.2

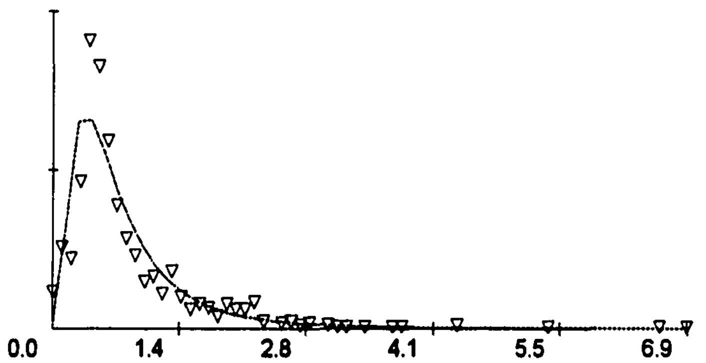


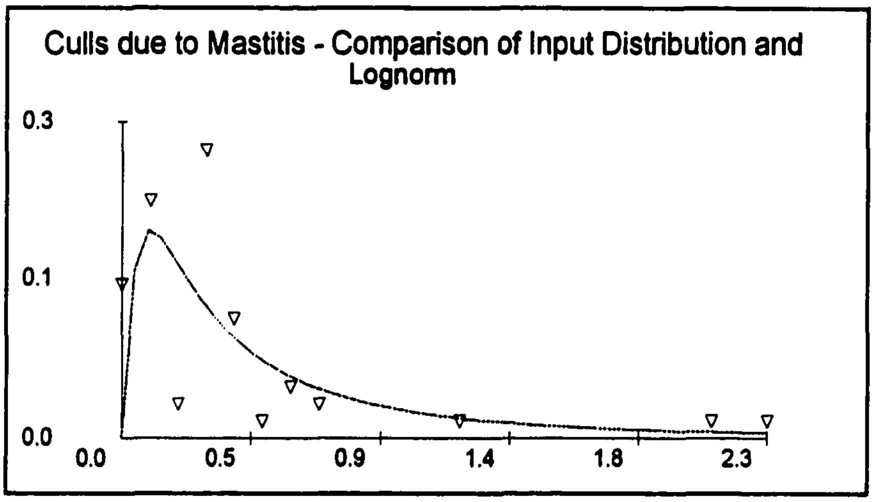
Culls due to Injury - Comparison of Input Distribution and Pearson V
0.18
0.09
0.00
0
1
2
3
4
5

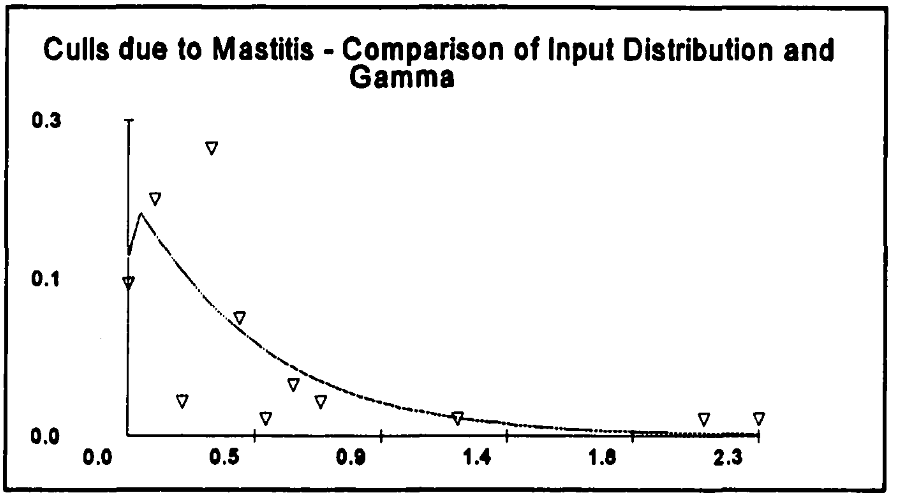


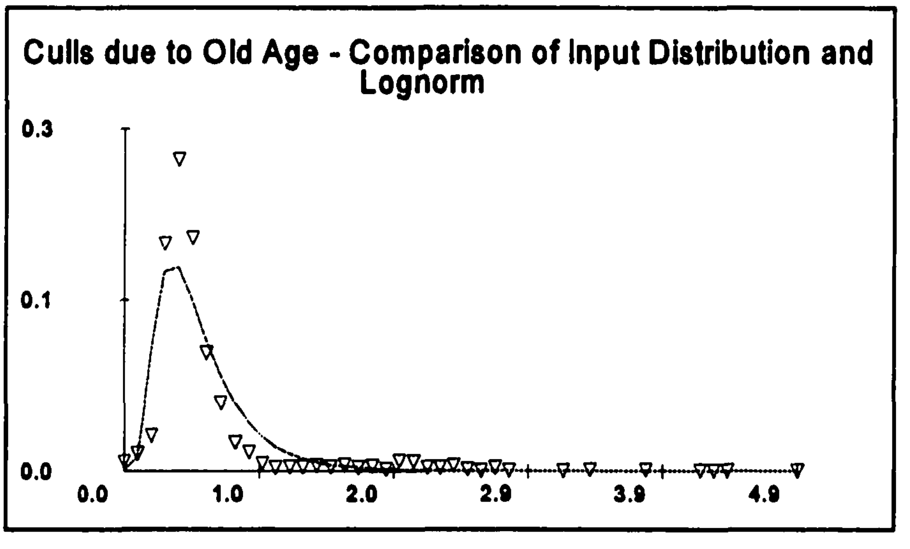
Culls due to Injury - Comparison of Input Distribution and Lognorm
0.17
0.08
0.00
0
1
2
3
4
5

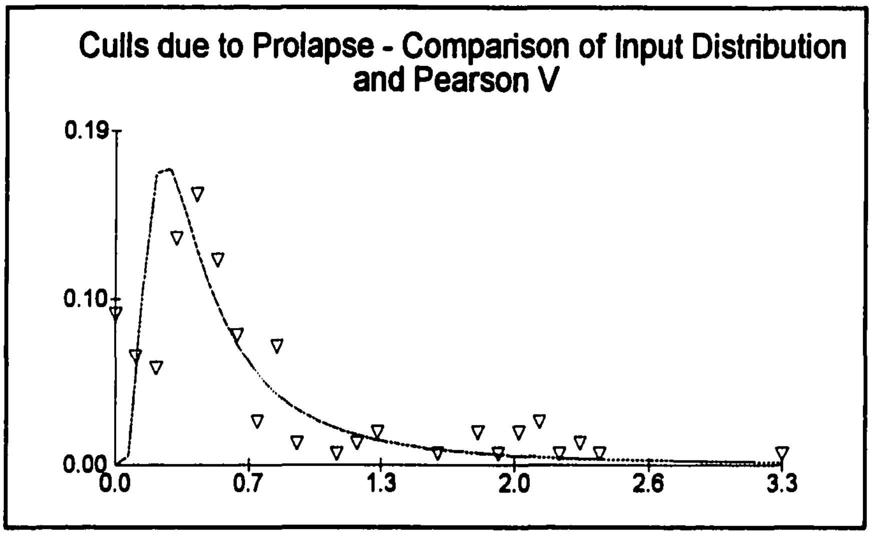
Culls due to Lame/Unsoundness - Comparison of Input Distribution and Lognormal

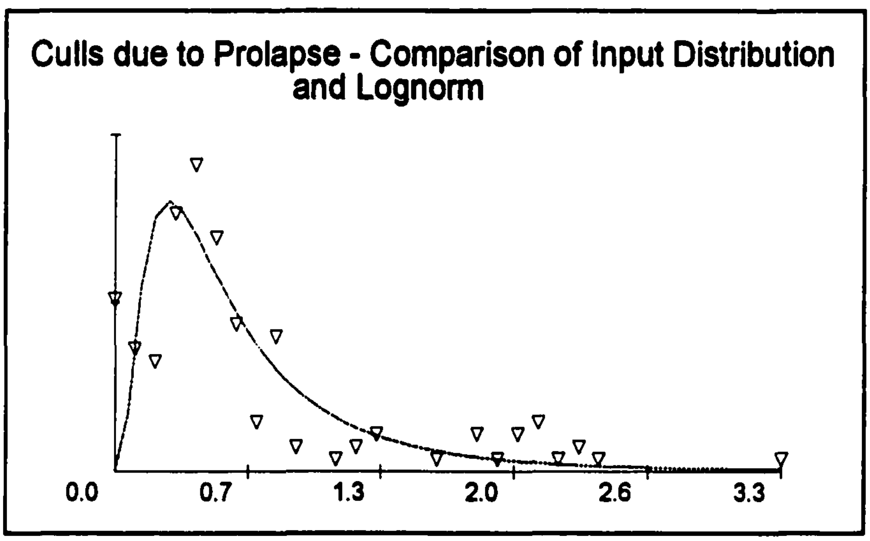



Culls due to Mastitis - Comparison of Input Distribution and Lognorm
0.3
0.1
0.0
0.0
0.5
0.9
1.4
1.8
2.3

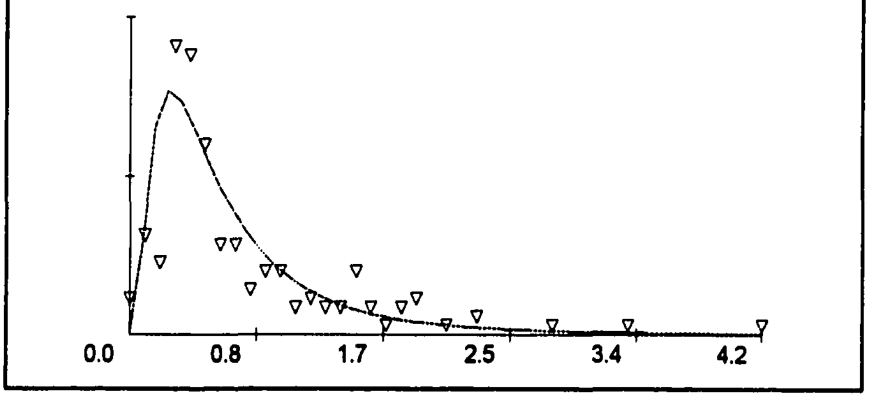

Culls due to Mastitis - Comparison of Input Distribution and Gamma
0.3
0.1
0.0
0.0
0.5
0.9
1.4
1.8
2.3

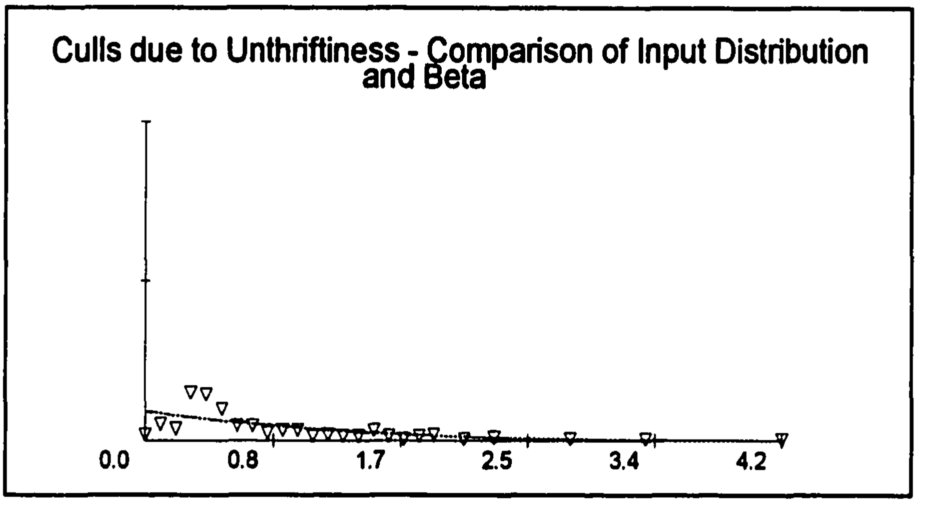

Culls due to Old Age - Comparison of Input Distribution and Lognorm
0.3
0.1
0.0
0.0
1.0
2.0
2.9
3.9
4.9


Culls due to Prolapse - Comparison of Input Distribution and Pearson V
0.19
0.10
0.00
0.0
0.7
1.3
2.0
2.6
3.3

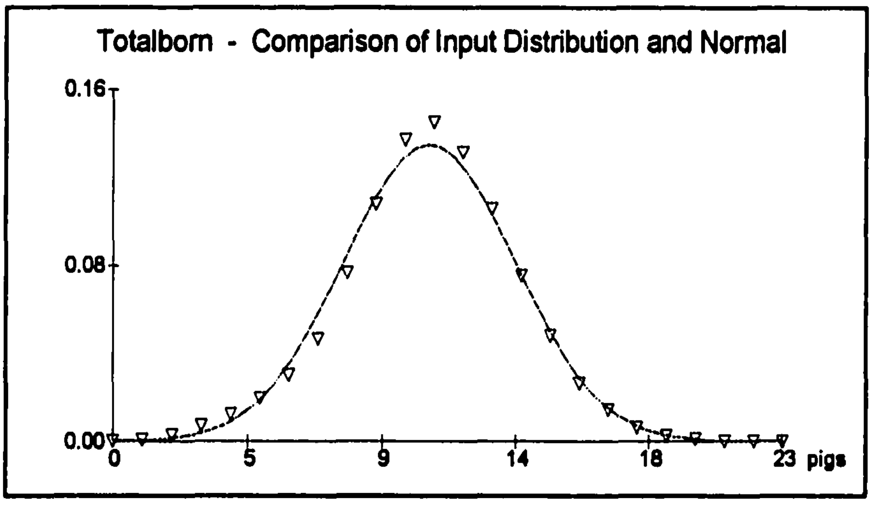

Culls due to Prolapse - Comparison of Input Distribution and Lognorm
0.0
0.7
1.3
2.0
2.6
3.3

Culls due to Unthriftiness - Comparison of Input Distribution and Lognorm





Culls due to Unthriftiness - Comparison of Input Distribution and Beta
0.0
0.8
1.7
2.5
3.4
4.2



Totalborn - Comparison of Input Distribution and Normal
0.16
0.08
0.00
0
5
9
14
18
23
pigs

3.2 Comparison of non-reproductive culling reasons throughout and within parities

Table A.3.1 Significance (p values) of comparisons of culling parities within culling reasons

Reason	Contrast	Parity 7+	Parity 6	Parity 5	Parity 4	Parity 3	Parity 2
Age	Parity 1	< 0.0001	< 0.0001	< 0.0001	0.0007	0.6616	0.7124
	Parity 2	< 0.0001	< 0.0001	0.0002	0.0040	0.9450	
	Parity 3	< 0.0001	< 0.0001	0.0003	0.0054		
	Parity 4	< 0.0001	< 0.0001	< 0.0001			
	Parity 5	< 0.0001	< 0.0001				
	Parity 6	0.0001					
Lame	Parity 1	< 0.0001	0.0159	0.1275	< 0.0001	< 0.0001	0.0019
	Parity 2	< 0.0001	< 0.0001	0.2602	0.0607	0.1473	
	Parity 3	< 0.0001	< 0.0001	0.0150	0.6284		
	Parity 4	< 0.0001	< 0.0001	0.0052			
	Parity 5	< 0.0001	0.0011				
	Parity 6	< 0.0001					
Injury	Parity 1	0.0002	0.4081	0.0050	0.8601	0.0013	0.1293
	Parity 2	< 0.0001	0.0605	0.2343	0.3157	0.1425	
	Parity 3	< 0.0001	0.0020	0.8247	0.0234		
	Parity 4	0.0034	0.4040	0.0442			
	Parity 5	< 0.0001	0.0047				
	Parity 6	0.0780					
Unthrifty	Parity 1	< 0.0001	0.2129	0.6949	0.8338	0.1646	08283
	Parity 2	< 0.0001	0.3378	0.8609	0.7212	0.1745	
	Parity 3	< 0.0001	0.0330	0.1568	0.3683		
	Parity 4	< 0.0001	0.2204	0.6197			
	Parity 5	< 0.0001	0.4580				
	Parity 6	0.0001					

continued ...

Reason	Contrast	Parity 7+	Parity 6	Parity 5	Parity 4	Parity 3	Parity 2
Downer	Parity 1	0.0002	0.8166	0.5163	0.6103	0.0336	0.0625
	Parity 2	< 0.0001	0.1145	0.3818	0.3356	0.7817	
	Parity 3	< 0.0001	0.0754	0.2709	0.2372		
	Parity 4	0.0003	0.5457	0.9188			
	Parity 5	0.0002	0.4760				
	Parity 6	0.0053					
Prolapse	Parity 1	0.9936	0.0001	0.1206	0.0999	0.0118	0.0906
	Parity 2	0.0562	0.0378	0.9823	0.9411	0.4416	
	Parity 3	0.0039	0.1828	0.4652	0.5289		
	Parity 4	0.0685	0.0716	0.9284			
	Parity 5	0.0857	0.0562				
	Parity 6	< 0.0001					
Mastitis	Parity 1	0.1003	0.4289	0.4193	0.2018	0.8708	0.0812
	Parity 2	0.0010	0.5001	0.5107	0.7941	0.2235	
	Parity 3	0.1387	0.6041	0.5945	0.3654		
	Parity 4	0.0072	0.6979	0.7091			
	Parity 5	0.0295	0.9884				
	Parity 6	0.0307					

Table A.3.2 **Significance (p values) of comparisons of culling reasons within parities**

Parity	Contrast	Unthrifty	Prolapse	Mastitis	Lame	Injury	Downer
1	Age	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	Downer	0.1855	0.0475	0.6126	0.7025	0.7488	
	Injury	0.0322	0.0504	0.7423	0.2562		
	Lame	0.1150	0.0129	0.3828			
	Mastitis	0.1201	0.1745				
	Prolapse	0.0028					
2	Age	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	Downer	0.2244	0.2375	0.9530	0.3201	0.1873	
	Injury	0.9893	0.7387	0.2633	0.4787		
	Lame	0.5530	0.4826	0.4093			
	Mastitis	0.2880	0.2692				
	Prolapse	0.7575					
3	Age	< 0.0001	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001
	Downer	0.7989	0.6204	0.1206	0.4419	0.6223	
	Injury	0.7960	0.8647	0.1690	0.7424		
	Lame	0.5637	0.9882	0.1927			
	Mastitis	0.1427	0.2896				
	Prolapse	0.7434					
4	Age	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	Downer	0.7095	0.9274	0.5749	0.2860	0.5523	
	Injury	0.2271	0.6976	0.2483	0.0104		
	Lame	0.3957	0.3405	0.8570			
	Mastitis	0.7626	0.5603				
	Prolapse	0.6805					

continued ...

Parity	Contrast	Unthrifty	Prolapse	Mastitis	Lame	Injury	Downer
5	Age	0.0007	0.0538	0.0222	< 0.0001	< 0.0001	0.0031
	Downer	0.8591	0.7708	0.9808	0.9701	0.4449	
	Injury	0.2478	0.3490	0.5708	0.1462		
	Lame	0.8256	0.7455	0.9543			
	Mastitis	0.8651	0.7819				
	Prolapse	0.8656					
6	Age	0.3522	0.0011	0.8384	0.0072	0.0614	0.4687
	Downer	0.9947	0.0067	0.5219	0.7346	0.6521	
	Injury	0.5835	0.0001	0.2552	0.8065		
	Lame	0.6584	0.0000	0.2709			
	Mastitis	0.4831	0.0801				
	Prolapse	0.0016					
7+	Age	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
	Downer	0.6976	0.1032	0.2338	0.5007	0.0314	
	Injury	0.0020	0.9276	0.6164	0.0019		
	Lame	0.1630	0.1208	0.3478			
	Mastitis	0.0963	0.7337				
	Prolapse	0.0281					

CHAPTER 4

Assessing the accuracy of a swine breeding herd simulation model

4.1 Summary

Performance of a swine breeding herd simulation model was assessed using a suite of statistical measures. Monthly numbers of total pigs weaned was chosen as the key output variable. Actual and simulated values were compared. Comparison criteria included mean differences, amount of variability explained by the line of perfect agreement, value of the regression coefficient when simulated values are regressed on measured values and the calculation of a zero intercept is forced, and value of the correlation coefficient. Model accuracy was found to be good, with satisfactory agreement over all the range of farm sizes studied, when judged by the mean difference, the correlation coefficient, and the simple linear regression coefficient between measured and simulated values. Amount of variability explained by the model, as measured by the line of perfect agreement between simulated and measured values, averaged 46%, while average standard deviation of simular data represented 79% of that of actual data. Thus, model performance was more even than actual system performance, and appeared more accurate for larger than smaller farms.

4.2 Introduction

Simulation models developed for predictive purposes need to be verified to assess the reliability and accuracy of their predictions and to build confidence that they are useful tools. Unfortunately, there is no single, proven method for evaluating simulation model behavior and output. Rather, where reported in the literature, the verification process has consisted of performing a series of statistical tests to compare model output with real-world data.

The simplest approach is to plot simulated values against measured (actual) values. The degree of agreement is indicated by how closely the resulting graph approximates a straight line. The appropriate statistical test is to regress model output on measured values and to perform F test for zero intercept and unit slope. Although this approach is intuitively appealing, Harrison (1990) has warned about the validity of this test, since bias in parameter estimates can lead to rejection of valid models.

Some authors (e.g. Addiscott & Whitmore, 1987; Whitmore, 1991) favor the use of multiple methods when assessing validity. Suggested methods include:

- The product moment correlation coefficient (r) as a measurement of the strength of the linear relationship between simulated and measured values. Values close to either +1 or -1 indicate strong linear relationship;
- The mean difference (M) between simulated and measured values. Mean differences close to 0 are preferred;
- The dispersion of the difference ($y_i - x_i$) between measured and simulated values (Richter, 1985), where some preset percentage of simulated values must be within some arbitrary, but important, range of measurements;
- comparison of the size of the sum of the squares of the residuals with the total sum of squares in the data about their mean (Greenwood, 1985); and
- Student's t-test to verify whether a simulation is within the experimental error of replicate measurements.

The objective of the verification process is to generate and assemble various statistics which, as a set, provide an objective basis for comparing and summarizing the degree of deviation between simulated data and real-world measurements. Thus, the more closely selected simulation model output mirrors actual measurements, the higher the degree of confidence in the overall performance of the simulation model.

4.3 Materials and Methods

4.3.1 Software

The model tested, PigORACLE©, is the second development stage of one of two available modules from a single skeleton simulation model. This module simulates the breeding phase of a swine production system up to the end of the nursery stage. Since it interfaces with the main swine farm record system available in North America, its validation is especially important because, if performance is proved adequate, it will enable a direct link between farm records and grow/finish simulation models.

4.3.2 Data Sources

Breeding herd performance was simulated for each of twelve different herds. Agreement between model output and actual performance was assessed using each of the methods listed above. In selecting herds, production records were subjected to a strict quality control procedure to ensure that, as far as possible, data represented a complete and accurate record of events that occurred on the farm. In order to qualify for this study, each herd's production record data files could contain no more than five percent missing events when cross-checked against all recorded and related events. For example, where a farrowing event was recorded, the breeding female's record was checked for the existence of a corresponding mating event; each weaning event was checked for an associated farrowing event. In addition, breeding herd female inventory must not have varied more than 10% during each 12-month period for which actual and simulated output were to be compared. This was necessary to avoid depopulations or herd expansions. These events typically result in increased volatility among statistics reported by records systems, and so would require more complex management input data patterns for proper simulation than were considered in these "steady state" verification exercises.

PigCHAMP®[1] Performance Monitor, Parity Distribution, and other reports were generated for each herd-year to be simulated. Each year's output was divided into 12 calendar months. . Monthly values for selected statistics were manually transcribed to PigORACLE© management and production variable input matrix files. Actual management and production variables used to drive the simulation model runs included: average total litter sizes by parity (1..6, 7+), percent pigs born live, pre-weaning mortality (%), minimum lactation length (d), percent abortions, nursery pigs mortality (%), annualized parity specific culling rates (%), and preferred day of week for weaning. Other input matrix variables not reported or recorded by PigCHAMP®, such as percent difficult farrowings, were set at constant values across all farms.

Herd demographic data was transferred from PigCHAMP® to PigORACLE© using the *PigCHAMP PigORACLE Interface Report*. In each instance, the report date was set at the date corresponding to the beginning of the comparison period. To establish the simular population, the following data were transferred from PigCHAMP® to PigORACLE© for each breeding female in the herd: ID, parity number, date last mated, date last farrowed, total born litter size, date last weaned.

Each farm-year was simulated ten times, with each simulation run spanning two years. Output from the first year was discarded as trial runs indicated that an initialization period of one year was necessary in most cases in order for the system to attain a steady state, and attain initial simulation state closer to real conditions. Thus output from the second simulated year was considered to represent the period of interest. Output from sets of 10 simulation runs per farm were aggregated, and monthly averages compared with corresponding values from PigCHAMP® reports.

4.3.3 Statistical Analysis

A suite of statistical measures was used to assess the effectiveness of the

[1] University of Minnesota, CAPS, 1365 Gortner Ave., St. Paul, MN 55108, USA.

swine breeding herd simulation model to predict the monthly totals of pigs weaned, pigs born, and farrowings for each herd.

The set of statistics calculated included:

a) the difference between actual monthly and simulated values, expressed both in absolute terms, as a percent deviation, and measured in actual data standard deviation units;

b) value and statistical significance (p-value) of the correlation coefficient between actual and simulated data;

c) value of the regression coefficient when actual values are regressed on simulated values, and the model is forced through the origin;

d) the variability explained by the line of perfect agreement (LPA) between simulated and measured values, which is the straight line drawn at a 45° degree angle, that results when simulated values exactly match real values. The amount of variability explained is derived from the comparison between the sum of squares about the line of perfect agreement, and the sum of squares about the mean;

e) the comparison between the standard deviations of actual and simular data;

f) the maximum percent deviation between simulated and actual values;

4.4 Results and Discussion

4.4.1 Overall accuracy assessment

Values of selected statistics depicting accuracy of simulated total pigs weaned per month for each of the 12 simular farms, are shown in Table 4.1. Tables A.4.1 and A.4.5 in the Annex summarize results for other two important variables: average monthly total pigs born, and number of farrowings.

These results indicate an excellent level of agreement between simulated and real data. In almost all herds, the percent mean difference for total pigs weaned is well within the pre-set 15% tolerance level, and overall average difference is 7% (5% median difference) for the average 516 pigs weaned per month. Percent mean differences for the other two relevant variables (sows farrowed and total pigs born) are similar or slightly better, as shown in the tables of the Annex (tables A.4.2 and A.4.6). This is because simulated values for these two variables are generated earlier in the simulation cycle for each sow, thus accumulate less stochastic deviation. These results indicate consistency of model performance through the mating, gestation, farrowing, and weaning phases of the simulated production system.

Correlation coefficients, in the range of 0.854 to 0.997, are all approaching one, with very small P values ($p < 0.001$), indicating significant linear relationships between simulated and measured values. The value of the weakest correlation

TABLE 4.1 - Correlation and observed differences between real and simulated average monthly pigs weaned

Farm	Recorded monthly total pigs weaned	Mean (*) Difference (real - simulated)			Correlation(**) Coefficient	Regression(**) Coefficient
		Numeric (a1)	Percent (a2)	In SD (a3)	(b)	(c)
A	100	20	20	0.53	0.933	1.232
B	91	-4	4	0.07	0.854	0.961
F	154	-8	5	0.27	0.970	0.926
I	189	14	7	0.20	0.957	1.075
G	110	13	12	0.51	0.951	1.092
M	552	34	6	0.38	0.980	1.048
J	523	-18	3	0.30	0.981	0.943
E	621	18	3	0.27	0.981	1.010
H	849	100	12	1.11	0.992	1.123
C	869	8	1	0.08	0.987	0.996
K	897	-6	1	0.12	0.997	0.990
L	1238	-72	6	0.72	0.996	0.942

(*) Mean actual monthly total pigs weaned - Mean simulated monthly total pigs weaned
(**) Model forced through the origin

coefficient (0.854) is influenced by the fact that records for the corresponding farm show zero pigs weaned for one of the months considered. Analysis of the original farm records show that this was a recording error; which highlights the difficulties involved in simulation of biological systems. Overall, the values for the correlation coefficients for pigs weaned are very high, as are those for recorded and simulated farrowings, and recorded and simulated pigs born. Regression coefficients were not tested, since it has been reported that the F-test for zero intercept and unit slope may lead to rejection of valid models due to bias in parameter estimates (Harrison, 1990). Most regression coefficients are nevertheless very close to unity, and given the high correlation coefficients, r^2 values are mostly well above 0.95, and show a median value of 0.96.

Despite the highly satisfactory overall performance of the model as indicated by level of agreement between simulated and real data, two issues are worthy of comment. First, some of the simulated number of farrowings per week (mainly for smaller herds with less than 10 farrowings per week) should, ideally, have been closer to observed values. However, the model controls the introduction of replacement females into the herd by limiting weekly matings as a function of weekly farrowing capacity and anticipated farrowing rate. Thus, model performance drops when attempting to accurately simulate situations where different-sized farrowing rooms, or failure to observe strict all-out farrowing room management causes the number of available farrowing crates to vary from week to week, or not to be an integer number. Second, partly as a consequence of the above, and possibly due to managerial aspects involved in running a larger operation, agreement between simulated and observed values seem to improve with simular farm size. In spite of this, there are no statistically significant differences in simulation model performance for different herd sizes.

In Table 4.2 simulation results are broken down by breeding herd size. The 12 farms are divided into three groups according to actual monthly total pigs weaned.

TABLE 4.2 - Correlation between real and simulated data by farm size

Farm size	Actual monthly weaned	Difference: real-simulated (*)			Correlation(**) Coefficient	Regression(**) Coefficient	r^2 (**)
		Numeric	Percent	in SD			
0 - 499 monthly total born (n=5)							
Average	129	7	10	0.31	0.933	1.057	0.8725
Median	110	13	7	0.27	0.951	1.075	0.9044
500 - 999 monthly total born (n=4)							
Average	636	34	6	0.52	0.983	1.031	0.9671
Median	587	26	5	0.34	0.981	1.029	0.9617
1000 + monthly total born (n=3)							
Average	1001	-23	2	0.31	0.993	0.976	0.9862
Median	897	-6	1	0.12	0.996	0.99	0.9911
All farms							
Average	516	8	7	0.38	0.965	1.028	0.9320
Median	538	10	5	0.29	0.98	1.003	0.9612

(*) Mean actual monthly total weaned - Mean simulated monthly total weaned
(**) Model forced through the origin

Despite the good overall accuracy, and the similar level of performance for all herd size groups, the average percent difference between real and simulated data is as much as five times larger for small than for large farms, and the average correlation and regression coefficients are better as well.

4.4.2 Month to month (variability) accuracy assessment

Having determined the overall accuracy of model output, this section addresses the issue of how well the model mimics month-to-month variability. Table 4.3 summarizes results for the twelve farms simulated. Numbers show a relatively large variation in maximum deviation of individual monthly simulation results with respect to actual data, but mostly for smaller farms. This is better

TABLE 4.3 - Analysis of monthly simular data deviation and amount of data variability explained

Farm	Recorded monthly total weaned	CV (*)	SD (**)	% Variability explained by LPA (***) (d)	Max (%) deviation of simular data (f)	Proportion SDs/SDa(****) (e)
A	100	0.38	38	28	121	0.31
B	91	0.66	60	7	105	0.22
G	110	0.23	26	51	64	0.62
F	154	0.19	29	79	58	0.97
I	189	0.37	70	20	57	0.49
J	523	0.11	58	76	35	1.47
M	552	0.16	89	28	41	0.81
E	621	0.11	67	39	31	1.06
H	849	0.11	91	46	30	0.99
C	869	0.11	94	28	29	0.98
K	897	0.06	51	105	19	0.90
L	1238	0.08	100	45	20	0.68

(*) Coefficient of Variation of actual data
(**) Standard Deviation of actual data
(***) Line of Perfect Agreement
(****) Standard Deviation simular data / Standard Deviation actual data

characterized in table 4.4 where farms are divided in three groups by size. Numbers in this table show an approximate sixty percentage point drop in maximum deviation between the smaller and larger farm groups.

Nevertheless, this is a "snapshot" analysis since it depicts only the worst deviation encountered in all the monthly values simulated. On the other hand, the last column in table 4.3 attempts to analyze model performance from a different

TABLE 4.4- Analysis of monthly simular data deviation by farm size

Farm size	Actual monthly total weaned	CV (*)	% Variability explained by LPA (**) (d)	SDs/ SDa (***)	Deviation of simular data		
					Max (+)	Max (-)	Max (%) (f)
0 - 499 total born (n=5)							
Average	129	0.37	37	0.52	87	-72	81
Median	110	0.37	28	0.49	78	-72	64
500 - 999 total born (n=4)							
Average	636	0.12	47	1.08	206	-156	34
Median	587	0.11	43	1.02	211	-171	33
1000 + total born (n=3)							
Average	1001	0.08	59	0.86	170	-220	23
Median	897	0.08	45	0.90	159	-244	20
All farms							
Average	516	0.21	48	.79	147	-137	51
Median	538	0.14	42	.86	135	-129	38

(*) Coefficient of Variation of actual data
(**) Line of Perfect Agreement
(***) Standard Deviation simular data / Standard Deviation actual data

scope. It portrays how model output dispersion mimics (follows) actual data dispersion, by showing the standard deviation (SD) in simular data as a proportion of the standard deviation of actual data. Observing these values it is clear that the simulation algorithm follows natural variations in output to a large degree, and quite consistently across the larger farm size groups. These results, in association with the small mean percent differences found between real and simulated values for the output variable, indicate a very good overall performance of the simulation model.

Even though the model takes into account variability, it performs more (efficiently and more) evenly than real systems, mainly small ones. Model performance is steadier than real world performance, so it tends to be more consistent and produce less variability than real world production units. It follows that when measured both by the Line of Perfect Agreement and by the proportion of SD of simular data to that of actual data, the model accounts for somewhat less than real variation, a fact that seems to be more true for smaller than medium and large producing units.

In summary, these results, indicate a high level of agreement between simulation model output and recorded farm data. Results of statistical tests show that agreement in monthly numbers of farrowings, pigs born and pigs weaned was significant for all farms simulated. As the ultimate measure of output of the breeding herd, we are satisfied that the PigORACLE© model can reliably predict the expected average and variability in the number of pigs weaned per month in steady-state continuous farrowing production systems of at least 250 breeding females.

4.5 References

Addiscott TM and Whitmore AP (1987). Computer simulation of changes in soil mineral nitrogen and crop nitrogen during autumn, winter and spring. J Agric Sci, 109:141.

Greenwood DJ, Neeteson JJ and Draycott A (1985). Response of potatoes to N fertilizer: Dynamic model. Plant and Soil, 85:185.

Harrison, SR (1990). Regression of a model on real-system output: an invalid test of model validity. Agric Syst, 34:183

Richter J, Nordmeyer H and Kersebaum CK (1985). Simulation of nitrogen regime in loess soils in the winter half-year: comparison between field measurements and simulations. Plant and Soil, 83:419.

Whitmore AP (1991). A method for assessing the goodness of computer simulation of soil processes. J of Soil Sci, 42:289.

Annex Chapter 4

Tables in this Annex summarize data for the same types of analyses performed on monthly total pigs weaned, but performed on monthly total sows farrowed and total pigs born, which are the main events (variables) that must be simulated prior to obtaining the chosen outcome variable.

TABLE A.4.1 - Correlation and observed differences between real and simulated average monthly total pigs born

Farm	Recorded monthly total pigs born	Mean (*) Difference (real - simulated)			Correlation(**) Coefficient	Regression(**) Coefficient
		Numeric (a1)	Percent (a2)	in SD (a3)	(b)	(c)
A	123	-23	18	0.67	0.973	1.220
B	126	-10	8	0.23	0.950	0.925
F	204	-7	4	0.17	0.955	0.926
I	243	27	11	0.46	0.958	1.096
G	142	15	10	0.37	0.963	1.104
M	663	63	10	1.00	0.993	1.097
J	663	10	1	0.13	0.994	1.009
E	793	14	2	0.30	0.998	1.017
H	976	78	8	0.81	0.995	1.081
C	1101	43	4	0.43	0.997	1.038
K	1082	-48	4	0.86	0.997	0.954
L	1481	-68	5	0.68	0.995	0.952

(*) Mean actual monthly total pigs born - Mean simulated monthly total pigs born
(**) Model forced through the origin

TABLE A.4.2 - . Correlation between real and simulated monthly total pigs born data, by farm size

Farm size	Actual monthly total born	Difference: real-simulated (*)			Correlation(**) Coefficient	Regression(**) Coefficient	r^2 (**)
		Numeric	Percent	in SD			
0 - 499 monthly total born (n=5)							
Average	167	9	10	0.38	0.960	1.054	0.9209
Median	142	15	10	0.37	0.958	1.096	0.9170
500 - 999 monthly total born (n=4)							
Average	774	47	5	0.56	0.995	1.051	0.9895
Median	728	39	5	0.55	0.958	1.049	0.9887
1000 + monthly total born (n=3)							
Average	1221	-24	4	0.65	0.996	0.962	0.9927
Median	1101	-48	4	0.68	0.997	0.954	0.9932
All farms							
Average	633	12	7	0.51	0.981	1.035	0.9817
Median	663	15	6	0.44	0.993	1.028	0.9864

(*) Mean actual monthly total born - Mean simulated monthly total born
(**) Model forced through the origin

TABLE A.4.3 - Analysis of monthly simular data deviation and amount of data variability explained

Farm	Actual average monthly total born	CV (*)	SD (**)	% Variability explained by LPA (***) (d)	Max (%) deviation of simular data (f)	Proportion Sds/SDa (****) (e)
A	123	0.25	34	48	81	0.46
B	126	0.28	43	26	72	0.28
G	142	0.30	39	37	63	0.36
F	204	0.20	43	53	62	0.91
I	243	0.25	58	30	52	0.46
J	663	0.10	76	44	20	0.98
M	663	0.08	63	57	27	0.91
E	793	0.06	49	120	12	0.60
H	976	0.08	96	54	26	0.93
C	1101	0.06	99	36	20	0.76
K	1062	0.06	56	104	17	1.33
L	1481	0.07	100	148	17	0.96

(*) Coefficient of Variation of actual data
(**) Standard Deviation of actual data
(***) Line of Perfect Agreement
(****) Standard Deviation simular data / Standard Deviation actual data

TABLE A.4.4- Analysis of monthly simular data deviation by farm size

Farm size	Actual monthly total born	CV (*)	% Variability explained by LPA (**) (d)	SDs/SDa (***)	Deviation of simular data		
					Max (+)	Max (-)	Max (%) (f)
0 - 499 total born (n=5)							
Average	167	0.27	49	0.49	103	-66	66
Median	142	0.27	37	0.46	99	-53	63
500 - 999 total born (n=4)							
Average	774	0.09	69	0.86	166	-108	21
Median	728	0.10	56	0.92	158	-113	23
1000 + total born (n=3)							
Average	1221	0.07	63	1.02	144	-186	18
Median	1101	0.07	48	0.98	155	-185	17
All farms							
Average	633	0.16	55	0.75	134	-110	39
Median	663	0.11	48	0.83	117	-108	27

(*) Coefficient of Variation of actual data

(**) Line of Perfect Agreement

(***) Standard Deviation simular data / Standard Deviation actual data

TABLE A.4.5 - Correlation and observed differences between real and simulated average monthly sows farrowed

Farm	Recorded average monthly sows farrowed	Mean (*) Difference (real - simulated)			Correlation(**) Coefficient	Regression(**) Coefficient
		Numeric (a1)	Percent (a2)	in SD (a3)	(b)	(c)
A	12	2	16	0.63	0.974	1.184
B	12	-1	7	0.28	0.961	0.924
F	18	-1	3	0.14	0.979	0.968
I	20	1	7	0.29	0.973	1.074
G	14	0	1	0.04	0.960	1.008
M	60	3	6	0.73	0.995	1.055
J	56	0	1	0.08	0.995	0.991
E	70	0	0	0.02	0.997	0.999
H	89	1	1	0.14	0.995	1.006
C	97	0	0	0.02	0.997	1.002
K	101	-3	3	0.59	0.998	0.966
L	127	-6	5	0.60	0.997	0.955

(*) Mean actual monthly sows farrowed - Mean simulated monthly sows farrowed
(**) Model forced through the origin

TABLE A.4.6 - Correlation between real and simulated monthly sows farrowed data, by farm size

Farm size	Actual monthly sows farrowed	Difference: real-simulated (*)			Correlation(**) Coefficient	Regression(**) Coefficient	r^2 (**)
		Numeric	Percent	In SD			
0 - 499 monthly total born (n=5)							
Average	15	0	7	0.27	0.969	1.032	0.9394
Median	14	0	7	0.28	0.973	1.008	0.9457
500 - 999 monthly total born (n=4)							
Average	69	1	2	0.24	0.995	1.013	0.9909
Median	65	1	1	0.11	0.995	1.002	0.9901
1000 + monthly total born (n=3)							
Average	108	-3	3	0.40	0.997	0.9743	0.9945
Median	101	-3	3	0.59	0.997	0.9660	0.9947
All farms							
Average	56	0	4	0.30	0.985	1.011	0.9704
Median	58	0	3	0.20	0.995	1.022	0.9897

(*) Mean actual monthly sows farrowed - Mean simulated monthly sows farrowed
(**) Model forced through the origin

TABLE A.4.7 - Analysis of monthly simular data deviation and amount of data variability explained

Farm	Actual average monthly sows farrowed	CV (*)	SD (**)	% Variability explained by LPA (***) (d)	Max (%) deviation of simular data (f)	Proportion SDs/SDa (****) (e)
A	12	0.25	3	48	69	0.33
B	12	0.28	3	27	65	0.30
G	14	0.30	4	31	52	0.22
F	18	0.20	4	51	50	0.30
I	20	0.25	5	22	41	0.30
J	56	0.10	5	53	22	0.31
M	60	0.08	5	82	23	0.58
E	70	0.06	4	102	16	0.82
H	89	0.08	7	83	20	1.10
C	97	0.08	8	40	18	0.45
K	101	0.06	6	67	18	0.55
L	127	0.08	10	34	17	0.29

(*) Coefficient of Variation of actual data
(**) Standard Deviation of actual data
(***) Line of Perfect Agreement
(****) Standard Deviation simular data / Standard Deviation actual data

TABLE A.4.8- Analysis of monthly simular data deviation by farm size

Farm size	Actual monthly sows farrowed	CV (*)	% Variability explained by LPA (**) (d)	SDs/ SDa (***)	Deviation of simular data		
					Max (+)	Max (-)	Max (%) (f)
0 - 499 total born (n=5)							
Average	15	0.267	3	0.29	9	-9	67
Median	14	0.25	2	0.30	8	-9	65
500 - 999 total born (n=4)							
Average	69	0.08	136	0.70	7	-9	13
Median	65	0.08	110	0.70	7	-8	11
1000 + total born (n=3)							
Average	108	0.07	71	0.43	12	-14	15
Median	101	0.08	87	0.45	11	-12	14
All farms							
Average	56	0.15	64	0.46	9	-10	36
Median	58	0.09	28	0.46	8	-10	25

(*) Coefficient of Variation of actual data
(**) Line of Perfect Agreement
(***) Standard Deviation simular data / Standard Deviation actual data

CHAPTER 5

Assessing the impact of data type on a swine breeding herd simulation model performance

5.1 Summary

Performance of a swine breeding herd simulation model was assessed under different periodicity of key driving variables. Monthly numbers of total pigs weaned was chosen as the principal output variable. Actual and simulated values originated from data sets containing 1, 2, 3, 4 and 6 month averages for the main driving variables were compared. Comparison criteria included mean differences between actual and simulated values, amount of variability explained by the line of perfect agreement, value of the regression coefficient when simulated values are regressed on measured values and the calculation of a zero intercept is forced, and value of the correlation coefficient. Model performance was found to be consistent throughout the different levels of input data periodicity studied.

5.2 Introduction

Performance of simulation models developed for predictive purposes depends not only on the appropriateness of the underlying algorithm, but also to a great extent on the type and quality of input data they are supplied. Long-term use of farm production record systems has allowed producers to accumulate considerable amounts of individual animal and group data. Besides their obvious use for production monitoring and historical reporting and analysis purposes, these data represent unique sources of information which enable simulation models to be closely tailored to individual herds. Simulation allows producers to develop reliable forecasts which illustrate the spread of probable outcomes suggested interventions designed to correct or improve either the magnitude or variability in current or past productivity. The reporting capabilities of most swine herd information systems provide users with numerous reporting options. Most offer considerable flexibility when choosing the incremental time slice when generating reports which provide a series of statistics over time. For example, the PigCHAMP® program allows users to choose a reporting period as short as one day or as long as 10 years. Time slices may be declared as any number of days, weeks, months, or years. However, in the case of swine production modeling, we are unaware of any published information regarding how model performance is affected by the periodicity, or "density" of the driving variables.

This purpose of this study is to measure the effects of changing periodicity of input data on the performance of the PigORACLE© simulation model. Our hypothesis is that there is a baseline periodicity, beyond which, the cost and extra effort required to assemble and input driving variable data is not rewarded by a change or improvement in simulation model performance.

5.3 Materials and Methods

5.3.1 Software

The simulation model used is PigORACLE©, a swine breeding herd simulation model. The model simulates the breeding phase of a swine production system up to the end of the nursery stage. Results from the simulations are recorded into text files, and analyzed with the help Statistix® ver. 4.0 for DOS [1], and Microsoft® Excel ver. 5.0 for Windows [2].

The reproductive performance of 12 swine breeding herds was simulated repeatedly using five different periodicity levels of the model's main driving variables. Monthly, bimonthly, quarterly, four monthly and semestral driving data sets were derived from production records and used to drive the model over a series of replications. Results from these trials were compared against actual herd performance, and against each other. Assessment of model performance under each scenario was achieved by calculating and comparing the values of several statistics associated with key output variables simulated

---

[1] Analytical Software, PO Box 13204, St. Paul, MN 55113, USA.

[2] Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-6399, USA.

under the differing input data periodicity situations. The statistics monitored include:

- the difference between actual monthly and simulated values of pigs weaned (the outcome variable), expressed both in absolute terms and as percent deviation;

- the product moment correlation coefficient between measured and simulated values of the outcome variable;

- the amount of variability explained by the line of perfect agreement between actual and simulated values;

This set of statistics summarize, and allow to analyze, the random and systematic deviations of simulated data from actual measurements, and to compare how they behave as the periodicity of the underlying driving data changes.

5.3.2 Data Sources

Production records of potential study herds were examined. Our objective was to select, as far as possible, data files which represented complete and accurate account of breeding female events that occurred on each farm during the period of interest. Use was made of the PigCHAMP® Data Integrity Report for this process. The proportion of missing events could not exceed five percent of all recorded linked events in qualifying herds. Additionally, the breeding female inventory in qualifying herds was not permitted to fluctuate more than 10% during the simular year. This criterion was necessary in order to avoid depopulations or rapid herd expansions which are beyond the scope of this experiment.

Monthly PigCHAMP®[3] Performance Monitor Reports were produced for each herd for the year to be simulated. Monthly values for the driving variables were transferred to the PigORACLE© management and production variable input matrix files. These variables

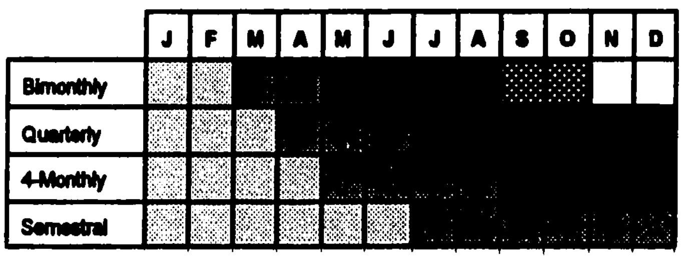
---

[3]University of Minnesota, CAPS, 1365 Gortner Ave., St. Paul, MN 55108, USA.

include: average total litter sizes by parity (1..6, 7+), percent pigs born live, pre-weaning mortality (%), minimum lactation length (d), percent abortions, nursery pigs mortality (%), annualized parity specific culling rates (%), and preferred day of week for weaning. Other input matrix variables not reported or recorded by PigCHAMP®, such as percent difficult farrowing, were set at typical values across all farms.

Herd demographic data was transferred from PigCHAMP® to PigORACLE© using the *PigCHAMP PigORACLE Interface Report*.

FIGURE 5.1 - Data patterns for the different periodicity levels



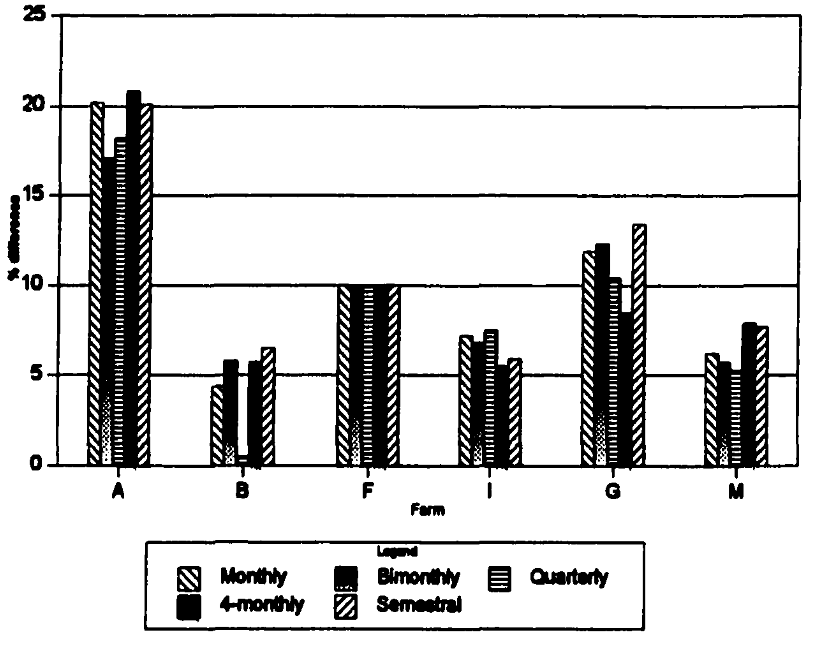
Each farm-year was simulated 10 times. Simulation model output data were aggregated by farm. Mean values of simulated data were compared with corresponding values from various PigCHAMP® reports. This process was repeated using five different input data periodicity levels: 1, 2, 3, 4 and 6 months averages for the main driving variables, following the patterns shown in Table 5.1 (e.g.: four quarterly values were used, the result of averaging monthly values for the periods January-March, April-June, July-September and October-December respectively). The main driving variables included: conception rate (%), pre- weaning mortality (%), nursery pig mortality (%), percent abortions, and parity-specific litter sizes.

5.4 Results and Discussion

Graphical and numerical summaries of the values for selected statistics, obtained from simulations performed at different levels of data periodicity, are shown in Figure 5.1

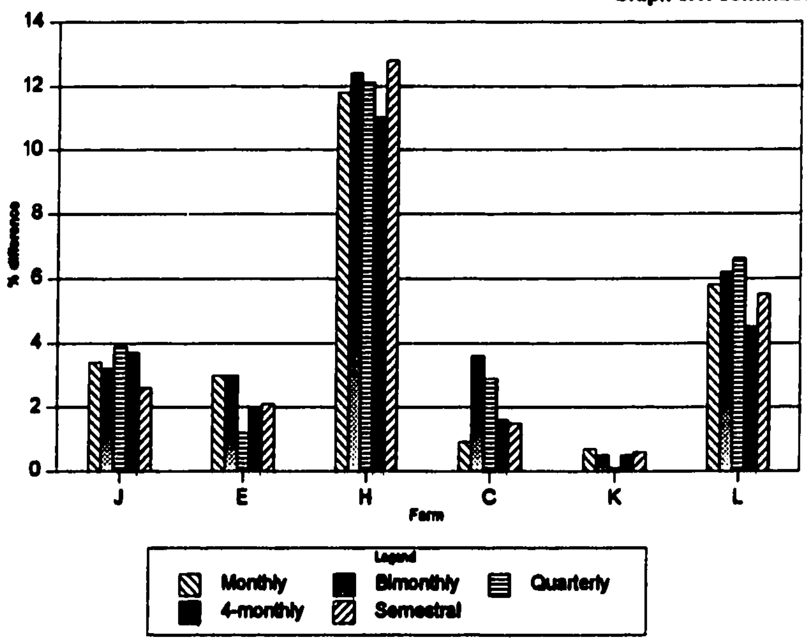
and Table 5.1. Visual inspection of the graphs, and the analysis of the table reveal very small differences between percent difference in actual vs. simulated values for number of pigs weaned; for correlation coefficients between real and simulated values; or in the coefficients of Simple Linear Regressions (SLR) between real and simulated values. Similarities between

Graph 5.1 - Percent difference between actual and simulated monthly pigs weaned at different input data periodicity levels



Graph 5.1 - Percent difference between actual and simulated monthly pigs weaned at different input data periodicity levels

Graph 5.1: continued



simulated and actual data are further supported by the results of a more rigorous statistical analysis. Table 5.2 summarizes the results of statistical significance tests performed on the variables in table 5.1.

These results indicate no statistically significant differences in simulation performance at differing periodicity levels for the main driving variables values: percent difference between real and simulated values, correlation coefficient between real and simulated values, regression coefficient for a SLR between real and simulated values at the different levels of data periodicity tried.

TABLE 5.1 - Values of selected statistics correlating actual and simulated average monthly pigs weaned at different periodicity levels for the main driving variables

Farm	Pigs Weaned per Month	Periodicity	Mean (*) Difference		Correlation Coefficient (**)	Regression Coefficient (**)
			Numeric	Percent		
A	100	Monthly	20	20.2	0.9330	1.2322
		Bimonthly	17	17.1	0.9257	1.1794
		Quarterly	18	18.2	0.9303	1.1976
		4-monthly	21	20.8	0.9409	1.2599
		Semestral	20	20.1	0.9356	1.2338
B	91	Monthly	-4	4.4	0.8544	0.9607
		Bimonthly	-5	5.8	0.8574	0.9473
		Quarterly	0	0.5	0.8302	0.9607
		4-monthly	-5	5.7	0.8439	0.9353
		Semestral	-6	6.5	0.8594	0.9438
F	154	Monthly	-2	1.0	0.9702	0.9257
		Bimonthly	-7	4.5	0.9701	0.6416
		Quarterly	-10	6.4	0.9642	0.6322
		4-monthly	-3	1.7	0.9671	0.6503
		Semestral	-7	4.5	0.9722	0.6499
I	189	Monthly	14	7.2	0.9572	1.0754
		Bimonthly	13	6.8	0.9434	1.0565
		Quarterly	14	7.5	0.9586	1.0859
		4-monthly	10	5.5	0.9551	1.0596
		Semestral	11	5.9	0.9562	1.0679
G	110	Monthly	13	11.9	0.9510	1.0922
		Bimonthly	14	12.3	0.9488	1.0945
		Quarterly	12	10.4	0.9554	1.0755
		4-monthly	9	8.5	0.9658	1.0749
		Semestral	15	13.4	0.9668	1.1392
M	552	Monthly	34	6.2	0.9803	1.0479
		Bimonthly	32	5.7	0.9842	1.0478
		Quarterly	29	5.3	0.9806	1.0377
		4-monthly	44	7.9	0.9878	1.0787
		Semestral	43	7.7	0.9849	1.0728

Continued ...

(*) Mean actual monthly total pigs weaned - Mean simulated monthly total pigs weaned

(**) Model forced through the origin

TABLE 5.1 - Values of selected statistics correlating actual and simulated average monthly pigs weaned at different periodicity levels for the main driving variables

Continued...

Farm	Pigs Weaned per Month	Periodicity	Mean (*) Difference		Correlation Coefficient (**)	Regression Coefficient (**)
			Numeric	Percent		
J	523	Monthly	-18	3.4	0.9805	0.9432
		Bimonthly	-17	3.2	0.9829	0.9481
		Quarterly	-20	3.9	0.9855	0.9475
		4-monthly	-19	3.7	0.9878	0.9541
		Semestral	-13	2.6	0.9852	0.9612
E	621	Monthly	18	3.0	0.9808	1.0097
		Bimonthly	19	3.0	0.9810	1.0106
		Quarterly	7	1.2	0.9820	0.9936
		4-monthly	12	2.0	0.9813	0.9999
		Semestral	13	2.1	0.9828	1.0040
H	849	Monthly	100	11.8	0.9919	1.1230
		Bimonthly	105	12.4	0.9902	1.1290
		Quarterly	102	12.1	0.9891	1.1220
		4-monthly	94	11.0	0.9915	1.1135
		Semestral	108	12.8	0.9890	1.1327
C	869	Monthly	8	0.9	0.9870	0.9961
		Bimonthly	31	3.6	0.9881	1.0243
		Quarterly	25	2.9	0.9852	1.0125
		4-monthly	14	1.6	0.9862	1.0013
		Semestral	13	1.5	0.9878	1.0030
K	897	Monthly	-6	0.7	0.9968	0.9902
		Bimonthly	4	0.5	0.9976	1.0032
		Quarterly	-1	0.1	0.9968	0.9959
		4-monthly	5	0.5	0.9970	1.0033
		Semestral	-5	0.6	0.9979	0.9933
L	1238	Monthly	-72	5.8	0.9956	0.9424
		Bimonthly	-77	6.2	0.9960	0.9391
		Quarterly	-81	6.6	0.9958	0.9367
		4-monthly	-55	4.5	0.9965	0.9564
		Semestral	-68	5.5	0.9955	0.9459

(*) Mean actual monthly total pigs weaned - Mean simulated monthly total pigs weaned

(**) Model forced through the origin

TABLE 5.2 - Summary of One-Way Analysis of Variance Results

Variable	Kruskal -Wallis Statistic	P-value
Precision		
- Percent difference (*)	0.5404	0.9694
- Correlation Coefficient	0.3518	0.9862
- Regression Coefficient (**)	0.0710	0.9994
Variability		
- Variability explained by LPA	0.4963	0.9739

(*) Mean actual monthly total weaned - Mean simulated monthly total weaned
(**) Model forced through the origin

These results show that the accuracy of the model is not compromised by the level of data density of the main driving variables. With regards to the amount of variability of the original data that the simulation model output can account for, this also seems to be quite leveled for different periodicities. The bottom section of Table 5.2 shows that there are no statistically significant differences in the amount of variability accounted for by the model at the different levels of data density tested.

5.5 Discussion

In summary, results show that the density of data supplied to the model did not affect its simulation performance, both in its accuracy and in the amount of variability it can account for. This may be explained by the fact that even smaller producing units have, on average, "industrial type" production systems, where the monthly amount of pigs weaned, and other biological performance indicators are quite stable.

5.6 References

Addiscott TM and Whitmore AP (1987). Computer simulation of changes in soil mineral nitrogen and crop nitrogen during autumn, winter and spring. J Agric Sci, 109:141.

Analytical Software. Statistix® User's Manual: Statistix®, Version 4.0, 1992.

Greenwood DJ, Neeteson JJ and Draycott A (1985). Response of potatoes to N fertilizer: Dynamic model. Plant and Soil, 85:185.

Marsh WE (1986). Economic decision making on health and management in livestock herds: examining complex problems through computer simulation. Ph. D., University of Minnesota, St. Paul.

Microsoft Corporation. Excel® User's Manual: Excel®, Version 5.0, 1993.

Polson DD, Marsh WE and Dial GD (1992). A survey of north american swine breeding herds using the PigCHAMP® information system: I. Profiles of swine production system alternatives. Proc 12th IPVS Congress, August 17-20, The Hague. pp 501.

Richter J, Nordmeyer H and Kersebaum CK (1985). Simulation of nitrogen regime in loess soils in the winter half-year: comparison between field measurements and simulations. Plant and Soil, 83:419.

Whitmore AP (1991). A method for assessing the goodness of computer simulation of soil processes. J of Soil Sci, 42:289.

CHAPTER 6

Assessing the impact of missing data on a swine breeding herd simulation model performance

6.1 Summary

Performance of a swine breeding herd simulation model was assessed under different quality levels of data for a key herd data input variable. Monthly numbers of total pigs weaned was chosen as the main output variable. Simulated values originated from data sets containing five different levels (0%, 5%, 15%, 25%, and 35% missing data) of data quality for one input variable. The comparison criterion was the value of the main output variable for the different levels of missing data. It was found that under the test conditions, quality of data, measured as percentage of missing values for one input variable, had no influence on model performance between simulations at the different levels of data quality.

The amount of simulation months needed for the model to attain steady state was also determined. It was found that, on average, the model can achieve final performance level within a three month simulation period.

6.2 Introduction

Performance of simulation models developed for predictive purposes depends not only on the appropriateness of the underlying algorithm, but also to a large degree on the type and quality of input data they are supplied. It was shown for swine production simulation models, that type of input data (expressed by data periodicity) may not affect simulation performance when stable, good quality data herds are simulated (Soler, 1997).

However, in the case of swine production modeling, we are unaware of any published information regarding how model performance may be affected by the quality of input data variables. This study will analyze the effect data quality may have on simulation output performance. Its objective is to measure the effects of different levels of missing input data on the performance of the PigORACLE© simulation model.

The model has a built in feature that allows it to compensate for missing input data, by generating it according to data patterns contained in the management data section. The purpose of the exercise is to determine how poor input data quality must be before the model cannot fully compensate for missing data. Our hypothesis is that there is a minimum baseline quality, beyond which, the model cannot fully compensate for the inadequacy of the input data it is provided, and simulation performance is affected.

6.3 Materials and Methods

6.3.1 Software

The simulation model used is PigORACLE©, a swine breeding herd simulation model. PigCHAMP®[1] production records of study herds are examined for completeness and accuracy of breeding female events, and transferred to PigORACLE© for simulation. The model simulates the breeding phase of a swine production system up to the end of the nursery stage. Results from the simulations are recorded into text files, and analyzed with the help of Statistix software[2], and Microsoft® Excel ver. 5.0 for Windows[3] .

6.3.2 Data Sources

Production records of potential study herds were examined to ensure best possible quality, and to allow optimum baseline simulation conditions. Our objective was to select, as far as possible, data files which represented complete and

---

[1] University of Minnesota, CAPS, 1365 Gortner Ave., St. Paul, MN 55108, USA.

[2] Analytical Software, PO Box 13204, St. Paul, MN 55113, USA.

[3] Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-6399, USA .

accurate account of breeding female events that occurred on each farm during the period of interest. Use was made of the PigCHAMP® Data Integrity Report for this process. The proportion of missing events could not exceed five percent of all recorded linked events in qualifying herds. Additionally, the breeding female inventory in qualifying herds was not permitted to fluctuate more than 10% during the simular year. This criterion was necessary in order to avoid depopulations or rapid herd expansions which are beyond the scope of this experiment.

Herd demographic data was transferred from PigCHAMP® to PigORACLE® using the *PigCHAMP - PigORACLE Interface Report*. Each farm-year was simulated 10 times. Simulation model output data were aggregated by farm, and those aggregates obtained from full herd data were used as gold standards for comparison purposes.

6.3.3. Method

The reproductive performance of 12 swine breeding herds was simulated repeatedly using five different quality levels for one of the model's main input variables. Data sets containing from 5% to 35% missing data for one input variable were derived from production records and used to drive the model over a series of replications. Results from these trials were compared against results of simulations with full actual herd data, and against each other. Assessment of model performance under each scenario was done by comparing the values of average monthly number of simulated pigs weaned (the outcome variable) with full, and missing data.

A program was developed to stochastically cull a user defined level of data for one of the variables in the input set. The variable chosen was "last farrowing date". The purpose specific program was designed to read each female record, and stochastically determine whether the data in the field for the chosen variable would be missing (depending on the user defined probability for missing data). This process was repeated using five different input data quality levels: 0% through 35%

missing data in 5% steps.

The model has a built in feature that allows it to compensate for missing input data, by generating it according to data patterns contained in the management data section. The purpose of the exercise was to determine how poor input data quality must be before the model cannot fully compensate for missing data.

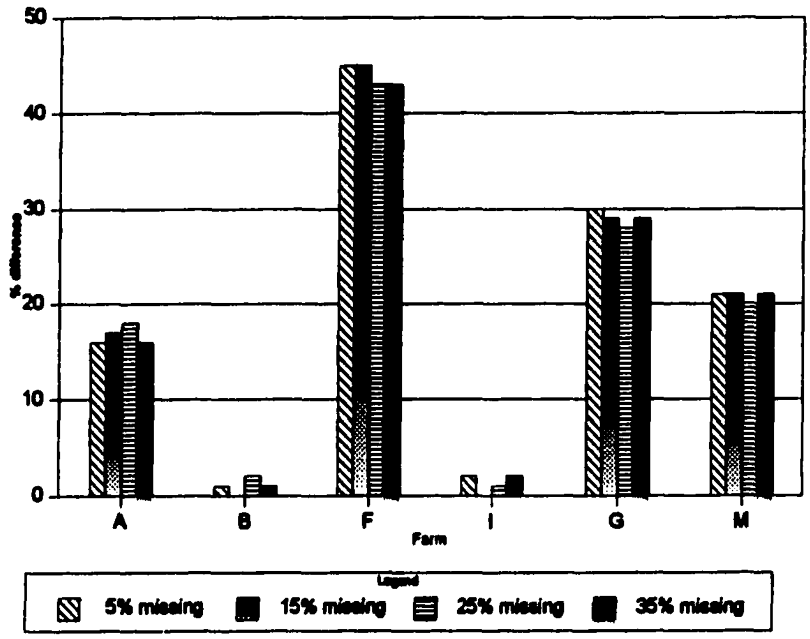
6.4 Results and Discussion

A summary of the values for selected statistics, obtained from simulations performed at the different levels of data quality, is shown in Table 6.1. Inspection of the table reveals for some cases, apparent differences between values for average number of pigs weaned generated with full data, as opposed to those generated with missing data. There are, however, no noticeable differences in pigs

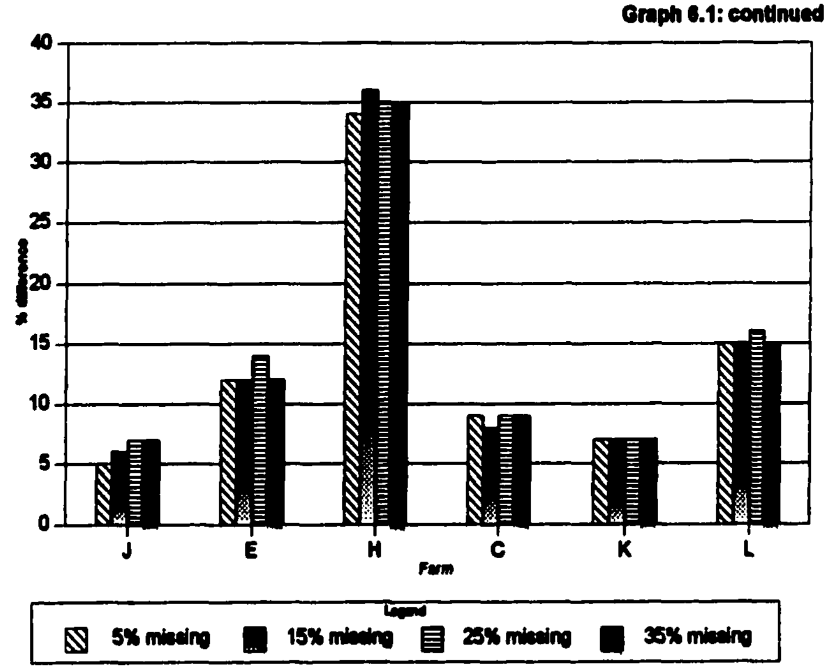
TABLE 6.1 - Values of simulated average monthly pigs weaned at different levels of data quality

Farm	Statistic	Full Data	Percent missing data			
			5	15	25	35
A	Value	80	92	93	94	92
	% Difference		16	17	18	16
B	Value	95	96	95	96	93
	% Difference		1	0	2	-1
F	Value	162	234	234	232	232
	% Difference		45	45	43	43
I	Value	176	172	176	177	172
	% Difference		-2	0	1	-2
G	Value	97	126	126	124	125
	% Difference		30	29	28	29
M	Value	518	625	625	623	626
	% Difference		21	21	20	21
J	Value	540	566	575	580	585
	% Difference		5	6	7	7
E	Value	603	674	678	686	674
	% Difference		12	12	14	12
H	Value	749	1007	1019	1010	1011
	% Difference		34	36	35	35
C	Value	861	936	934	939	936
	% Difference		9	8	9	9
K	Value	903	966	966	966	968
	% Difference		7	7	7	7
L	Value	1310	1513	1509	1515	1503
	% Difference		15	15	16	15

Graph 6.1 - Percent difference between actual and simulated monthly pigs weaned at different levels of data quality (missing data)



Graph 6.1: continued



weaned simulated at various levels of missing data. These observations are, however, contradicted by the results of a One Way Analysis of Variance performed on simulation data results. Table 6.2 summarizes this result, which indicates no statistically significant differences in simulation output for all five data quality levels included in the trial.

Table 6.2 shows that as much as 35% missing data for last farrowing date, does not significantly affect simulation performance. Up to this level, the model's internal routine can, on average, fully compensate for the lack of original data, so that there are no significant differences between increasing percentages of missing data.

TABLE 6.2 - Summary of One-Way Analysis of Variance Results for Simulated Average Monthly Pigs Weaned by Input Data Quality

Variable	Kruskal -Wallis Statistic	P-value
Average monthly pigs weaned (*)	0.6824	0.9535

(*) full input data set, and 5%, 15%, 25% and 35% missing input data sets considered

Another issue related to simulation performance concerns the simulation time period it takes the model to attain the final level of accuracy for each specific simulation exercise.

In the specific case of PigORACLE©, at this point in its development, the model reads in herd data provided, and simulates each female's future reproductive cycle based on the data it is provided. The model, however, does not account for pigs already born, and in the production process at the starting simulation date. This means that there is always a four month period between the start simulation date, and the time the first simulation pigs are weaned. In view of this, the validation protocol for PigORACLE© included an initial 12 month period to allow for the model to reach a equilibrium ("steady state"), before simulation output results were taken into consideration.

After taking these facts into account, one question that remained unanswered was: how long does it take for the model to reach "steady state", whereby output results for average monthly pigs weaned, show accuracy levels similar to those achieved under the validation protocol.

In order to answer this question, a special trial was set to determine how many simulation months would it take to attain an accuracy level within 10% of a pre-specified target. For the purpose of this exercise, target accuracy was defined as the percent difference between real and simulated monthly number of pigs weaned, achieved when following the model's validation protocol. Additionally, this level of accuracy had to be maintained for at least three consecutive months.

It was found that for the conditions tested, the model reaches final simulation performance levels within three monthly simulation periods, after the initial four month simulation period needed for the first simulation pigs to start being weaned.

6.5 Conclusion

In summary, results show that:

a) input data quality does not affect simulation performance under the conditions tested.

b) on average, for the conditions tested, the model attains final simulation performance levels within three monthly simulation periods, after the initial four month simulation period needed for the first simulation pigs to start being weaned.

6.6 References

Analytical Software. Statistix® User's Manual: Statistix®, Version 4.0, 1992.

Microsoft Corporation. Excel® User's Manual: Excel®, Version 5.0, 1993.

Soler, A. (1997). Assessing the impact of data type on a swine breeding herd simulation model performance. PhD dissertation, College of Veterinary Medicine, University of Minnesota, St. Paul.

CHAPTER 7

Summary and conclusions

7.1 Summary

This study had three objectives:

Objective 1: to complete the development and testing of a microcomputer-based swine breeding herd simulation model (PigORACLE©) designed to be interfaced with general purpose swine herd management information system software (PigCHAMP®).

Objective 2: To assess how the performance of the swine breeding herd simulation model is affected by the periodicity of values of driving variables derived from management information system reports.

Objective 3: To assess how the performance of the swine breeding herd simulation model is affected by incomplete or erratic recording of animal events comprising breeding female life histories that are used to establish the simular population for model runs.

Objective 1

A high-quality database of lifetime production records from 18 breeding swine herds collected over a 5-year period was analyzed. The database comprised 64,851 individual parity records, 9,928 of which were terminated with removal events. Patterns of, and intervals between, events recorded in individual animal lifetime histories provided the basis for the derivation of a set of probability distributions. These probability distributions were used in the simulation model to predict the timing and occurrence of important reproductive and removal events in the simulated lives of breeding females. The BestFit software was used to fit standard mathematical expressions to frequency distributions of empirical data describing total born litter size, weaning to first service interval, inter-estral interval, and farrow to removal interval for nine main non-reproductive culling reasons.

In most instances, up to five families of standard mathematical distributions were judged to adequately fit the empirical data. The basis for evaluation of goodness of fit were Chi-square and Kolmogorov-Schmirnov statistics produced in comparing empirical distributions with mathematical expressions fitted using the BestFit software. The most commonly fitted distributions for continuous variables were Weibull, Lognormal, Pearson, Erlang, and Beta. For ease of simulation, since there were no appreciable differences with the quality of fit among the different distributions, the Lognormal distribution was chosen for simulating intervals between events.

Next, performance of the simulation model was assessed using a suite of statistical measures. Monthly numbers of total pigs weaned was chosen as the key output variable. Actual and simulated values were compared. Comparison criteria included mean differences, amount of variability explained by the line of perfect agreement, value of the regression coefficient when simulated values are regressed on measured values and the calculation of a zero intercept is forced, and value of the correlation coefficient. Model accuracy was found to be good, with satisfactory agreement over all the range of farm sizes studied, when judged by the mean difference, the correlation coefficient, and the simple linear regression coefficient between measured and simulated values. Amount of variability explained by the model, as measured by the line of perfect agreement between simulated and measured values, averaged 46%, while average standard deviation of simular data represented 79% of that of actual data. Not surprisingly, model performance was less volatile than actual system performance, and appeared more accurate for larger than smaller farms.

Objective 2

Model performance was assessed under different periodicity of key driving variables. Monthly totals of pigs weaned was chosen as the principal output variable of interest. Actual and simulated values originated from data sets containing 1, 2, 3, 4 and 6 months averages for the main driving variables were compared. Comparison criteria included mean differences between actual and simulated values, amount of variability explained by the line of perfect agreement, value of the regression coefficient when simulated values are regressed on measured values and the calculation of a zero intercept is forced, and value of the correlation coefficient. Model performance was found to be consistent throughout the different levels of input data periodicity studied. Therefore, it was concluded that the extra work involved in compiling and entering a matrix of monthly values for the suite of key driving variables was not warranted for the Midwest U.S. pork production systems simulated. Under these circumstances, it appears that the derivation and use of 6-monthly averages for mean litter size and reproductive driving variables are sufficient to provide satisfactory simulation of actual performance.

Objective 3

Model performance considering varying quality of beginning simular herd data was assessed. The simulation model attempts to compensate for missing events in animal records copied from the management information system. Given this feature, and because it is general practice to delay entry of farrowing information into management information systems until after weaning, the number of simulated months needed for the model to attain steady state were assessed. Monthly numbers of total pigs weaned was chosen as the main output variable. Simulated values originated from data sets containing five different levels (0%, 5%, 15%, 25%, and 35% missing data) of data quality for one input variable. Comparison criterion was the value of the main output variable for the different levels of missing data. It was found that under the test conditions, quality of data, measured as percentage of missing values for one input variable, had no significant influence on model performance between simulations at the different levels of data quality tested.

It was also determined that, on average, approximately three simulated months are needed for the model to attain steady state.

In summary:

- model accuracy was found to be good, with satisfactory agreement over all the range of farm sizes studied;
- model performance was less volatile than actual system performance, and appeared more accurate for larger than smaller farms;
- results show that the density of data supplied to the model did not affect its simulation performance;
- for our test conditions, input data quality did not affect simulation performance; and,
- on average, the model attains final simulation performance levels within three monthly simulation periods.

7.2 Conclusions

Based on analysis of real-world data, we are confident that the stochastic sampling from a family of Lognormal distributions programmed into the model provides a realistic simulation f the timing and occurrence of reproductive and culling events in the life histories of breeding female swine.

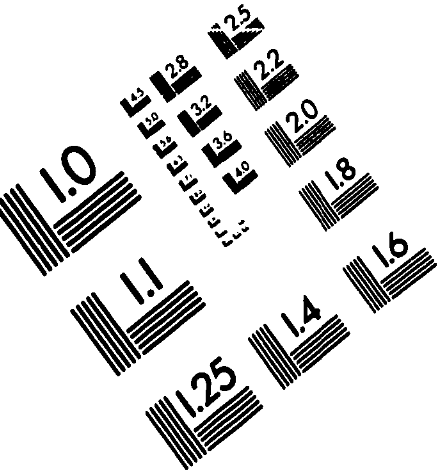
Model performance was found to be robust through varying levels of periodicity of key driving variables and completeness of events in records of animals which formed the populations for beginning simulations. One reason for this may be that the consistency of breeding female management and performance in large, modern pork production systems in the Midwest United States causes simulations to rapidly reach and maintain a steady state. This begs the question of the necessity for establishing a simular population from the management information system in order to achieve a reasonable set of simulation runs. This is in direct contrast to previous experience with a dairy herd simulation model derived from the same skeleton model as this swine breeding herd model. Because of longer gestation lengths, seasonal calving patterns, less systematic culling practices, and lower culling rates, the establishment of the initial simular herd from management information system records was considered to be fundamental to reliable simulation of patterns of reproductive performance and milk production in 5 Minnesota dairy herds (Marsh, 1986), and later in New Zealand production systems (Marsh, personal communication).

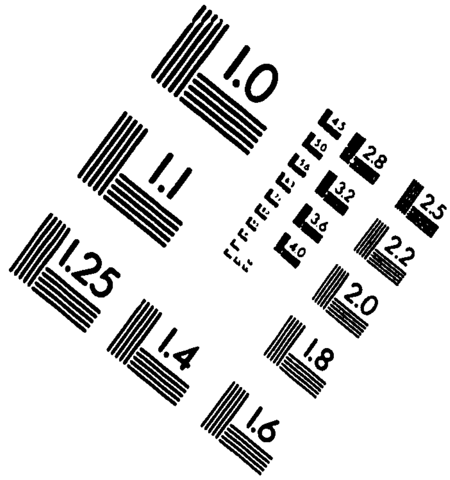
However, while data density may be less of a factor in determining good simulation performance, given the restricted testing conditions, input data completeness should always be a goal.

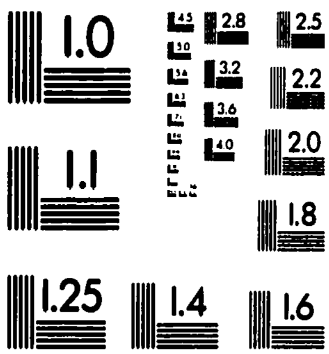
As developed, validated and tested, the model is a useful tool to support managerial decision making. It not only provides realistic production forecasts under stable conditions, but given its flexibility, it is most useful for ex-ante impact assessment of management changes.

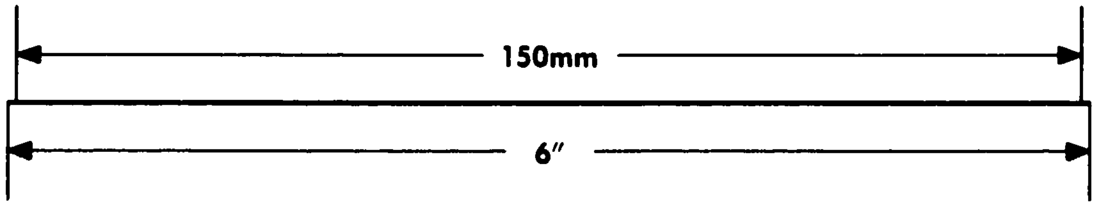
It is recommended that in the future, real world data be checked periodically to ensure that simulation techniques remain appropriate for the evolving production systems. This is especially important in the case of early weaning and multisite production. Model performance could also be enhanced by improving the realism of the simulation procedure for replacement gilts, and by estimating the population of piglets already in the production process at the starting simulation date.

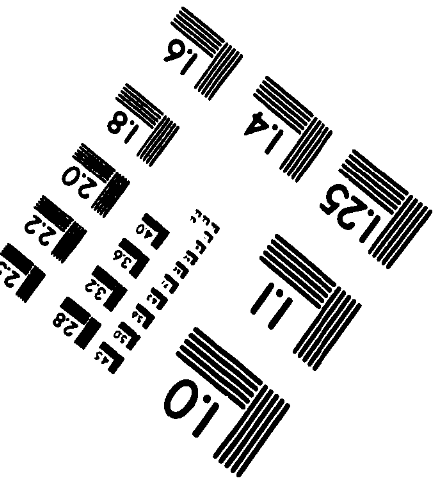
IMAGE EVALUATION TEST TARGET (QA-3)

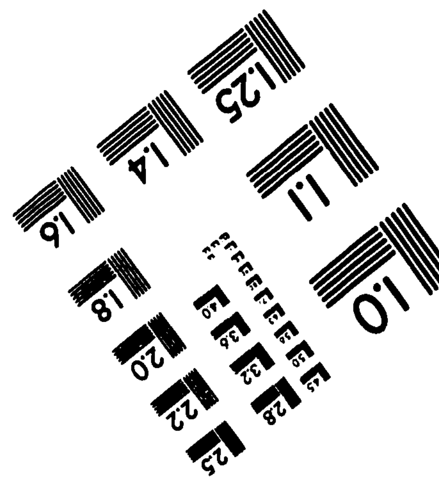












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5999