COMPARING ARTIFICIAL NEURAL NETWORK AND STATISTICAL METHODS FOR ESTIMATING THE SURVIVAL OF *ESCHERICHIA COLI* O157:H7 IN DRY FERMENTED SAUSAGE

BY

ANANDAKUMAR PALANICHAMY

A Thesis Submitted to
the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Department of Biosystems Engineering
University of Manitoba
Winnipeg, Manitoba

The Faculty of Graduate Studies
500 University Centre, University of Manitoba,
Winnipeg, Manitoba R3T 2N2

Phone: (204) 474-9377
Fax: (204) 474-7553
graduate_studies@umanitoba.ca

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

COMPARING ARTIFICAL NEURAL NETWORK AND STATISTICAL METHODS FOR ESTIMATING THE SURVIVAL OF *ESCHERICHIA COLI* 0157:H7 IN DRY FERMENTED SAUSAGE

BY

ANANDAKUMAR PALANICHAMY

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

OF

MASTER OF SCIENCE

ABSTRACT

The Canadian Food Inspection Agency (CFIA) requires the meat industry to ensure *Escherichia coli* (*E.coli*) O157:H7 does not survive in dry fermented sausage (salami) because a series of food-borne illness outbreaks have resulted from the presence of this pathogenic bacterium. The industry is in need of an alternative technique like 'predictive modeling' for estimating bacterial viability because traditional microbiological enumeration is a time-consuming and laborious method. Testing the accuracy and speed of artificial neural networks (ANNs) for this purpose is a current trend in predictive microbiological research, especially for online processing in industries. Two experimental data sets, one on interactive effects of different levels of pH, water activity (A_w), the concentration of allyl isothyocyanate (AIT) at various time intervals and the second on interactive effects of Lactoferrin, ethylene-diamine-tetraacetic-acid and sodium bi-carbonate during sausage manufacture in reducing *Escherichia coli* O157:H7 were used to develop predictive models using General Regression Neural Network (GRNN) (a form of ANN) and a statistical linear polynomial regression technique. Both models were compared for their prediction error using various statistical indices. GRNN predictions for training and test data sets had fewer and less serious errors when compared with the statistical model predictions. GRNN models were far superior and considerably superior respectively, for training and test sets than the statistical model. Because it is simple, fast and quite accurate, the ANN model can be used for online processing by research and development departments or quality control sections of the meat

processing industry to ensure product safety, and specifically for processing to eliminate *Escherichia coli* O157:H7 from dry fermented sausage.

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude and worships to my family deities for making my stay possible in Winnipeg and completing this Master program at University of Manitoba (U of M).

My sincere thanks are due to my advisor Dr. Digivir S. Jayas from Department of Biosystems Engineering, for his guidance, review of my research work, support and his time for discussions even during his very busy hours through out the program.

I am grateful to my Coadvisor Dr. Richard A. Holley from Department of Food Science, for funding this project and for his cherished guidance; expertise; time for discussions; help in accessing all type of the necessary literature; review of my research work and support throughout the project. The financial support for this study from Natural Science and Engineering Research Council of Canada is gratefully acknowledged.

I also would like to thank Dr. N.D.G. White (Agriculture and Agri-Food Canada) for serving on my program advisory committee and reviewing the thesis.

I was proud and thankful to God made me to have the above three eminent researchers as sources of inspiration for my research.

Many thanks are to Mr. Pedro Chacon and Dr. Anas Al-Nabulsi, the fellow graduate students from Department of Food Science who gave me their experimental data for developing the models. I thank Pedro and Dr. Parthiban Muthukumarasamy for giving me an exposure to the salami manufacturing process

and experimental data generation in Pilot Plant and Microbiological Laboratory in Department of Food Science.

Many thanks are to all the faculty members who taught me courses during this program at U of M. I owe my sincere thanks to Dr.Gary Crow from Department of Animal Science for his expertise and time for discussion on statistical modeling.

I also thank my friends Dr. Jeyamkondan Subbiah (University of Nebraska) and Dr. Chithra Karunakaran (Canadian Light Source) for their encouragement to take graduate study at U of M; their support to establish my stay in Winnipeg and to get started on continuation of my academic learning for my mid career development.

I owe thanks to Dr. J. Paliwal from Department of Biosystems Engineering for his support and thought provoking questions that facilitated in doing a good research.

I also thank Mr. M.P. Vasimalai, Executive Director, DHAN Foundation, India and other DHANites who permitted me to leave my responsibilities and sanctioned the sabbatical leave to undergo this graduate program at U of M.

I also thank my Mother, Wife, Siblings and other relatives who tolerated my absence around them during my stay in Winnipeg.

My thanks are to other fellow research scholars working in Grain Storage Research Centre, Wenyue Wu, Samuel Ima, and Wenbo Yang in the department for their company during this program.

Finally, I thank all who directly and indirectly supported me to complete this program.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

1.0 INTRODUCTION

A series of food borne illness outbreaks caused by *Escherichia coli* O157:H7 in fermented sausages (Anon, 1995; WHO, 1997) led the Canadian government adopting new U.S. food safety regulations which require sausage manufacturers to protect these products from this pathogen (Reed, 1995) by ensuring at least a 5 log reduction. *E.coli* O157:H7 is a pathogenic microorganism which rarely causes host animals (ruminants) to become ill but causes illness in humans who consume contaminated meat. About 4.1 cases per 100,000 population are reported annually. Ground beef is a common vehicle for spread of contamination by *E.coli* O157:H7 and, raw or undercooked meat products prepared using ground beef have been implicated (Health Canada, 2000).

It is generally reported that the salami industry does not use any modeling techniques to predict the presence of *E.coli* O157:H7 in the end product because the existing models have not been easily understood or are not easy to use. If such systems were available, industry would be in a better position to understand the most important factors and how these could be changed to reduce pathogen viability during processing. The growth of micro organisms has high biological variability in relation to the intrinsic and extrinsic factors of food. Their physiological responses to these factors are very complex and poorly understood (Geeraerd et al., 1998a).

The primary objective of food microbiologists is to identify and quantify the micro organisms that have both beneficial and deleterious effects on the safety and quality of raw or processed foods. An extensive amount of work is needed to generate and accumulate data on behavior of micro organisms in food. It is expensive too. Unfortunately, these data provide little insight in explaining the relationship between manufacturing processes and the growth or survival of micro organisms. The models developed based on these data are helpful in solving this limitation and in predicting unknown values through interpolation. Predictive modeling is a rapidly emerging food safety engineering technique with microbiological applications.

Predictive microbiology has rapidly emerged in the last two decades. Now, quality control sections of the meat industry and the food safety regulators are able to make reasonable assessment of the relative risk posed by a food or food process. Simple primary statistical models were successfully developed to model the bacterial growth rate using one or two environmental factors (Ratkowsky et al., 1991; McMeekin et al., 1993).

Polynomial models are the most common and attractive secondary models applied within predictive microbiology for describing the interplay of many factors affecting bacterial growth. Pond et al. (2001) developed polynomial models to

describe the survival of *E.coli* O157:H7 by using published data on inactivation of *E.coli* O157:H7 in uncooked fermented salami. Polynomial models are considered simple and relatively easy to fit to experimental data by multiple linear regression, which is available in most statistical packages (Ross et al., 2004). The application software packages identified as the "Pathogen Modeling Program" and "Food Micro Model" rely primarily on the use of polynomial models (Buchanan, 1993; McClure et al. 1994a).

Artificial Neural Networks (ANNs) have been used to generate complex models using predictor and response variables. The ANNs learn and remember the underlying implicitly non-linear relationship between the input and output variables following 'back propagation' techniques.

Secondary models have also been developed using ANNs for estimating microbial growth rates (Garcia-Gimeno et al., 2003; Geeraerd et al., 1998b; Jeyamkondan et al., 2001; Lou and Nakai, 2001; Najjar et al., 1997), under fluctuating environmental conditions (Cheroutre-Vialette and Lebert, 2000; Geeraerd et al., 1998a), and have also been used to predict microbial inactivation (Geeraerd et al., 1998b). ANNs have been suggested as an alternative to logistic regression polynomial modeling techniques (Tu, 1996). Generalized Regression Neural Network (GRNN) predictions were found to be superior to statistical modeling for the training data set, whereas they were

similar or slightly worse than statistical models for test data (Jeyamkondan et al., 2001).

A data set from Chacon (2006) (Appendix I) for the reduction of *E.coli O157:H7* (36 data sets of mean values of 3 replications) in response to changes in the level of the natural antimicrobial AIT (0, 500, 750, and 1000 ppm), days of processing (0 to 45 days), water activity (0.864-0.948), and pH (4.70-5.57) were used to develop a GRNN model for survival of *E.coli O157:H7.* The fermentation temperature during salami production was 26°C (< 3days) and drying was done at 13°C (25 days). These are common temperatures used by the industry for salami production.

Another data set from Al-Nabulsi (2006) (Appendix II) for the reduction of *E.coli* O157:H7(324 data sets) in response to changes in the level of the natural antimicrobial LF alone (0, 1.8, 3.0, and 6.0 mg / g sausage batter) or with different chelating agents namely sodium bicarbonate (SB) (0, 2.5, 5.0 mM) which is used linearly proportionate with ethylene diamine tetraacetic acid (EDTA) (0, 250, 500) for 0, 1.8, 6.0 levels of LF, days of processing (0 to 28 days) were used to develop a GRNN model for predicting the survival of *E.coli* O157:H7. The data on injured cells recovered on All Purpose Tween (APT) broth overlaid with selective medium cefixime-tellurite sorbitol McConkey agar (ct-SMAC).

For both data sets, the salami was manufactured in the pilot plant of the Department of Food Science at the University of Manitoba. The same method which is followed in Pillers' sausages plant, Waterloo, ON was adopted for manufacturing the salami.

The hypothesis of this research is that ANNs could be used for developing models for predicting the survival of *E.coli* O157:H7 for ensuring food safety in meat products. The developed models may be used by the meat industry as an online processing tool, because of its accuracy, prediction performance and speed.

The objectives formulated in this study to validate the hypothesis are:

to develop suitable statistical and ANN models using commercial neural network software for predicting the survival of *E.coli* O157:H7 in dry fermented sausage processing of two experiments where allylisothiocyanate (AIT) and lactoferrin (LF) were used as antimicrobial agents, and

to compare both the models using some statistical indices for determining their suitability so that the best modeling technique may be used as a tool for predicting process performance in the meat industry from a safety perspective.

2.0 REVIEW OF LITERATURE

2.1 General Emphasis on Food Quality and Safety

Most countries and the Food and Agriculture Organization reaffirmed at the World Food Summit that it is *the right of everyone to have access to* ***safe and nutritious food****, consistent with the right to adequate food and the fundamental right of everyone to be free from hunger.* This is the Rome declaration of world food security (FAO, 1996). This declaration emphasizes the need for every country to concentrate on ensuring the availability of a safe and nutritious food to its citizens.

About 30 – 40 percent of world food production is wasted because of improper storage. Micro organisms play a major role in reducing quality; spoiling and poisoning of stored foods and rendering them unsuitable for consumption. Food Technology has developed methods for controlling the growth of micro organisms physically, chemically and biologically. To assist in activities to prevent losses due to spoilage there also exists a need for predicting the growth and/or inactivation of micro organisms in foods during processing, storing and preservation. In order to ensure the stability and safety of food products until they reach consumer's hands, the growth or activity of micro organisms should be below levels found in the Food Safety Objectives (Jay et al. 2005) for specific foods to ensure quality and safety. Predictive microbiology is a rapidly emerging science which has attained separate sub-discipline status in the field of food

microbiology. It has the potential to be a powerful tool for development and evaluation of methods to improve the stability and safety of food.

2.2. Why Predictive Models?

Why do we need a model when the accumulated knowledge of the response pattern itself can provide significant information on the behavior of micro organisms with respect to the combined effect various intrinsic and extrinsic factors of food?

To answer this question in a practical sense, it is clear that the additional efforts made to develop and validate a model will, if properly done, lead to formulation of a general rule for describing the effect of environmental factors on microbial activity. Whereas the response pattern itself, if no model is developed, is more likely to describe only the output of limited experimental trials. Its applicability will again need challenge testing often under slightly changed conditions. Challenge testing is also expensive and time consuming.

The answer to the above question in a philosophical sense lies in the nature of science itself. "When you can measure what you are speaking about and express it in numbers, you have something about it; but when you can not measure it, when you can not express it in numbers your knowledge is of a meager and unsatisfactory kind"(sic)-Lord Kelvin (McMeekin 2004). So modeling is nothing but a quantitative science which is inherently more useful than the qualitative

description of a phenomenon. Modeling will be helpful in interpreting, analyzing data and inferring decisions.

In a general sense, models which are a combination of descriptions, mathematical functions or equations, and specific starting conditions can be used to simplify and study a variety of system.

2.3 Origin and History of Predictive Microbial Modeling

The canning industry in 1920 started and developed the methods for calculating the thermal death time of bacteria. The modeling of behavior of micro organisms with respect to controlling parameters like temperature began from then. Interest in predictive modeling resurged again from 1980 with a growing number of refrigerated foods being developed, the development of hurdle technology and of computer technology. Predictive microbiology is centered on the assumption that the growth or inactivation of micro organisms is due to varying levels of controlling parameters that can be predicted within biological variability.

2.3.1. Inactivation models for thermal processing of foods

Among the earliest work on predictive modeling of microbial activity, Scott (1937) did a study on the effect of temperature on microbial growth on meat. He emphasized the importance of the knowledge of microbial growth rates at different temperatures while studying meat spoilage. The relative influence of spoilage caused by various microbes at each storage temperature and their

changes in population could be predicted using these data. He also did a study on the effect of water content on microbial growth and spoilage (Scott 1936). The accumulated knowledge from both the studies, at a time when there was no explicit model, allowed the shipment of non-frozen meat from Australia to Europe.

2.3.2. Origin of mathematical expressions

A mathematical equation can describe the effect of integrated functions of controlling parameters on microbial growth in food. Models should be validated with a number of selected tests. Use of the Arrhenius law equation is a basic theoretical approach for finding the relationship between the reaction rate and temperature in theories, but it is more applicable for chemistry reactions. Instead of using this equation, the following model was developed to describe the relationship between the growth rate of micro organisms and the temperature minimum and optimum (Ratkowsky, 1982)

$$\sqrt{r} = b\,(T - T_o) \qquad \text{(Eq.1)}$$

Where r = growth rate, CFU per hour;

b = slope of the regression line drawn between growth and time, constant, $hr^{-0.5}$ / °C;

T_o = theoretical minimum temperature for growth, the intercept between the model and the temperature axis, °C;

and T = temperature of food, °C.

Using this formula, (Ratkowsky, 1982) showed a linear relationship between temperature and the growth of spoilage bacteria in foods or in broth media utilizing amino acids.

The spoilage of highly perishable foods depends on the temperature of storage. According to Olley and Ratkowsky (1973) spoilage can be predicted by a spoilage rate curve; the general spoilage curve was incorporated into the circuitry of a temperature function integrator that read out the equivalent days of storage at 0°C. So, it could predict the remaining shelf life at 0°C.

The model has been further developed (Wijtzes et al., 2001) to include the effects of other factors like pH and Aw on growth rate.

$$\sqrt{r} = b\,(T - T\,\min)\sqrt{(Aw - Aw\min)(pH - pH\,\min)(pH - pH\,\max)} \qquad \text{(Eq.2)}$$

where,

Aw, pH and T - water activity, pH and temperature of given food;

Aw min, pH min and T min - lower or minimum limits of above factors; and

pH max – Maximum limits of pH of food.

This model is popularly known as the square root model.

2.3.3 Development of predictive model studies

The lab-generated data from studies on the effect of cultural parameters on microbial growth rate at different pH, Aw, temperature and preservative concentrations was used to develop several types of mathematical models. Computers are used for rapid analysis of this multi-factorial data. The traditional viable count method is still the dominant method used to monitor bacterial response in modeling studies. This method is standard where other methods need validation. Though it is laborious and time consuming it can be done with good lab practice, and has adequate sensitivity, accuracy, precision, reproducibility and repeatability.

2.3.3.1 Sigmoidal model

A kinetic-based statistical model which is capable of indicating the growth rate for both lag and exponential phases is discussed below: The growth curves were statistically fitted by using non-linear regression analysis in conjunction with Gompertz functions. The results were then analyzed to develop this sigmoidal model which was introduced by Gibson et al. (1987). This model has been recognized and also developed by the United States Department of Agriculture for predicting microbial growth in food environments containing many control parameters (Ray, 2004). The expression of this asymmetric sigmoidal model is:

$$N = A + C * e^{-e^{[-B(t-M)]}}$$

Eq. (3)

Where A = $\log_{10}$ CFU per ml at initial time (asymptotic count)

N = $\log_{10}$ CFU per ml at time 't'

C = the difference in value of the upper and lower asymptote

M = time at which maximum growth rate occurs

B = relative growth rate at time M

This model gives us the lag time, the maximum growth rate constant, and the maximum microbial load directly from non-linear regression between the growth numbers and time data.

2.3.3.2 Polynomial and multiple linear regression models

Polynomial models are the most common models used in predictive microbiology, especially for modeling the survival of microorganism against the environmental conditions. Pond et al. (2001) developed four polynomial models to describe the survival of *E.coli* O157:H7 by using published data on inactivation of *E.coli* O157:H7 in uncooked fermented salami. The variables included in the models were significant at the $P<0.0001$ level. The correlation between predicted and observed values was at R^2 value of 0.888, 0.828, 0.836, and 0.818 for those models. Over-prediction of reductions in *E.coli* O157:H7 was found by these models which is 'fail-dangerous'. They concluded that modeling can be a useful tool in assessing manufacturing practices for uncooked fermented sausage

processes. However, the authors did not include all relevant variables such as drying temperature and relative humidity.

Polynomial models are considered simple and relatively easy to fit to experimental data by multiple linear regression, which is available in most statistical packages (Ross et al., 2004). The application software packages identified as the "Pathogen Modeling Program" and "Food Micro Model" rely primarily on the use of polynomial models (Buchanan, 1993; McClure et al. 1994a). Logistic regression is comparatively good against linear polynomial regression for modeling the percentage of data that are "bounded" and may be considered as rescaled probability values (Zhao et al. 2001). Lebert et al. (2000) developed a polynomial model that satisfactorily predicted the growth of *Pseudomonas* spp. in meat.

Polynomial models allow virtually any input variable and its interaction to be taken into account. It is easy to use the estimated coefficients when the model is included in application software. However, polynomial models also have some limitations. For a few parameters, there are many coefficients that have no biological interpretation. This makes it difficult to compare the polynomial model with other secondary predictive models. Baranyi et al. (1996) recommended that the interpolation region of a polynomial model be within the minimum convex polyhedron (MCP) defined by the ranges of environmental conditions used for developing the model.

2.3.3.3 Need for *E.coli* O157:H7 survival models in the salami industry

Tomicka et al. (1997) explored the survival of *E.coli* O157:H7 in a model representing fermented salami production. They determined the combined effect of starter culture (10^7 CFU of lactic acid bacteria per ml), dextrose (0.8%), sodium chloride (2%), nitrite (200 ppm), plus temperature (37 and 22° C) and concluded that a lower temperature and longer fermentation time ("European style") were better for elimination of *E.coli O157:H7* from a model system than fermentation at high temperature and short time ("American style"). Thus, there is a need for predictive models to facilitate choice of the optimal levels of ingredients and environment factors to use in processed meat manufacturing.

2.3.3.4 Effect of full and partial experimental data on the models

McClure et al. (1994b) demonstrated the stages in developing a predictive mathematical model for estimating the growth of *Aeromonas hydrophila* when the effects of variables, temperature (3-20°C), NaCl concentration (0.5-4.5% w/v) and pH (4.6-7.0) were considered. The growth curves were generated from viable counts and fitted using the Gompertz equation. Quadratic response surface equations were fitted to the log of lag and generation times, in response to the above variables. This model was compared with other models for growth of *A. hydrophila* developed with viable count data and optical density measurements which were used to obtain predicted growth rates and lag times. Also, the same was done with data from the literature on the growth of this bacterium in food. This study concluded that there is a potential for combining

data sets from different studies and to extend the useful range of the resulting models and allow their application for all kinds of variables and all kinds of foods.

Bratchell et al. (1989) attempted to determine the result of systematic removal of data from a model and its consequences. Their study illustrated the consequences of using insufficient data and demonstrated the risk of using an erroneous model. A mathematical model of the growth responses of salmonellae in a laboratory medium was developed by including the factors pH, NaCl concentration, and storage temperature. A part of the data was systematically removed from the whole data set and then the model was examined. Results highlighted the difference between the 3-D plots of fitted response surfaces for the whole data set and the reduced data sets. This study was useful in understanding the risks arising from inaccurate or incorrect models which are more acute when predicting bacterial growth than in any other system. However, a model with insufficient sampling will become less important with increasing use of automated sampling. Also, from this study, it was understood that the robustness of a model is characterized by the ease of identification of outliers, and other unusual observations. Data reduction imposes limitations on the ability of the model to predict with accuracy.

2.4 Model Categories

Primary vs. Secondary: Whiting and Buchanan (1993) have classified models into primary and secondary types. The models which describe the response of micro organisms with change in a small set of parameters, for example, temperature over time are called primary models. Whereas the models which describe the effect of environmental conditions, say, physical, chemical, and biotic features, like storage atmosphere, water activity, pH, food preservatives and additives on the values of the parameters of primary models are called secondary models. The various primary and secondary models are detailed by McKellar and Lu (2004).

Growth vs. Survival: Growth models and survival models are respectively, describing the growth and inactivation of microorganism during the process.

Descriptive vs. Explanatory: Descriptive models are data driven. They are observational, empirical, "block box" or inductive approaches such as polynomial functions, artificial neural nets, and principal component analysis. As these models can not be extrapolated beyond the data used to build them, true predictions are difficult to make. In spite of that, they have been widely used with considerable success in predictive microbiology. Explanatory models are mechanistic; 'white box'; or deductive models which will relate the given data with fundamental scientific principles, or at least to measurable physiological processes. They are composed of analytical and numerical models.

Stochastic or Probabilistic models are the ones that can recognize and account for the uncertainty or variability in an experimental systems.

Tertiary models are the models incorporated with environmental values of interest continuously entered into the secondary models in order to obtain the values of the predicted variable. Application software, risk assessment simulations and expert or decision support systems are the main examples of tertiary model systems. These make the modeling technology and data bases readily available for application in industry. This paves the way for choosing the type of new data collection needed in such a way that it can be merged with existing data bases.

The rules (parsimony, parameter estimation properties, and range of variables, stochastic assumption, and interpretability of parameters) for model selection were formulated by Ratkowsky (1993). Selection of an appropriate experimental design consistent with the purpose of the study should be an important pre-requisite. No model is mechanistic in predictive microbiology. Predictions can be done by interpolation only. The interpolation region only defines the applicability of a model, and the interpolation region is affected not only by the range of individual variables but also by the experimental design (Ross et al., 2000)

Moving from an empirical or phenomenological description to a mechanistic or deterministic description of a process indicates advancement in 'good science' hierarchy. The former stochastic models give a just mathematical relationship, whereas the latter have a theoretical basis i.e., interpreting the observed biological response on the basis of underlying theory. None of the secondary models is truly mechanistic. So Arrhenius type dependence models have a greater mechanistic basis than the Square root or Ratkowsky type models. But Belehradek (1930) did not support use of these because the Arrhenius models are based on chemical kinetics which can not be used for biological reactions. The secondary models like square root type or Ratkowsky type (1982) were developed based on the Belehradek type.

2.5 Use of Artificial Neural Networks for Predictive Modeling

The use of ANN in predictive modeling remains limited. ANNs are a data driven, black box approach to predictive modeling in contrast to other secondary models that can be written as an equation with coefficients and parameters. Neural network models are empirical ones and many methodological issues remain to be resolved. ANNs are robust and are able to handle high biological variability and non-linear data as long as enough data from well planned studies are used during their development.

Lou and Nakai (2001) developed an ANN model for predicting the thermal inactivation of *E.coli* due to combined effects of temperature, pH and water

activity, and compared this model with two others using root-mean-square-error and R^2. They showed that the ANN prediction performance - 0.144, 0.949 was better to that of response surface methodology model - 0.232, 0.868 and Cerf's model - 0.234, 0.815 (Cerf et al., 1996).

Cheroutre-Vialette and Lebert (2002) developed a dynamic model based on recurrent neural networks (RNN) and concluded that the complex effects of environmental variable conditions on microorganism behavior can be represented by this kind of model.

Two neural networks were developed by Mittal and Zhang (2002) to predict thermal process evaluation parameters like g (retort temperature – temperature of food at slow heating location) and found that ANN models closely followed the observed values.

Hajmeer and Basheer (2003) developed a hybrid model by integrating ANN and Bayes' statistical theorem for computing the probabilistic modeling of the bacterial growth and no growth interface. It outperformed the other approaches in its accuracy as well as flexibility to extract the implicit interrelationships between the various parameters.

Yu et al. (2006) developed and compared an ANN model of a three-layer back-propagation neural network trained using the survival and growth interfacing data-set and the model of McKellar et al. (2002). ANN's accuracy was more than

99 % for training data; 90% on classification accuracy for additional literature data-sets used for validation; and 100 % on all observed growth. The ANN model has been recommended as an alternative tool for evaluation of survival and growth conditions in predictive microbiology.

The effect of temperature, pH and NaCl on the heat resistance of *Bacillus stearothermophilus* spores was described using low-complexity, black box models based on ANN by Esnoz et al. (2006). Published data were used to build and train the neural network. The ANN models gave better predictions than the classical quadratic response surface model in all the experiments tried. Good predictions were also obtained when the neural networks were evaluated using new experimental data, providing fail-safe predictions of *D* values in all cases.

2.5.1 Generalized Regression Neural Networks

GRNN perform regression rather than classification tasks (Specht, 1991). There are three hidden layers in this network. The first layer containing the radial units is the input layer. The second hidden layer units help to estimate the weighted average as the transformation is applied to them at the hidden nodes. It is a specialized procedure. Here, the weighting is related to the distance of the point from the point being estimated (so that points nearby contribute most heavily to the estimate). A single special unit in this layer calculates the weighted sum. Hence, the second layer always has one more unit than the output layer. The

weighted sum should be divided by the sum of weighting factors to get a weighted average. Third layer is the output layer, and performs actual division (using special division units). Gaussian kernel functions are located at each training case. Each case can be regarded as evidence. The GRNN copies the training cases into the network for estimating the response to new points (Statsoft, 2006). GRNN applications are able to produce continuous valued output values. This is useful for continuous function approximation. GRNN can have multidimensional input, and it will fit multidimensional surfaces through the data.

2.5.2. Comparing ANN and Statistical models

By using published data, Jeyamkondan et al. (2001) used Generalized Regression Neural Networks (GRNN) to develop a model for predicting the lag phase period and generation time of *Brochothrix thermosphacta*, *Aeromonas hydrophila,* and *Shigella flexneri.* They compared the predictions of GRNN and of the published statistical models with the observed data. Six statistical indices, namely graphical plot, mean relative percentage residual (MRPR), bias factor, mean absolute relative residual (MARR), accuracy factor, and root mean square residual (RMSR) were used for comparison. GRNN predictions were found to be far superior to statistical modeling for the training data set, whereas they were similar or slightly worse than statistical models for test data (Jeyamkondan et al., 2001).

Several secondary ANN models have been developed so far for *Aeromonos hydrophila, Brochothrix thermosphacta, Shigella flexneri, E.coli, Listeria monocytogenes*, and lactic acid bacteria (Jeyamkondan et al. 2001; Garcia-Gimeno et al., 2003; Cheroutre-Vialette and Lebert 2000; and Lou and Nakai 2001). Predictions from the ANN models were compared with polynomial, square root type models and in general, ANN models provided slightly improved predictions. Currently, the development of ANN models has become relatively easy with the use of available commercial neural network software. But it can not provide classical secondary models (equations with coefficients and parameters), which are essential in order to allow incorporation with user-friendly application software in industry, teaching and research. This issue has to be sorted out by future researchers of predictive microbiology (Ross et al., 2004).

2.5.2.1 Advantages and disadvantages of ANN models

Tu (1996) compared the advantages and disadvantages of the ANN approach with those of statistical regression modeling.

Advantages:

i. Require less formal statistical training to develop.

ii. Detect implicitly the non-linear relationship between predictor and response variables.

iii. Detect the possible interactions between predictor variables.

iv. There would not be many methodological issues during model development and usage because it is empirical.

Disadvantages:

i. It is a black box approach and has limited ability to specifically identify casual relationships.

ii. Requires greater computational resources.

iii. Prone to over-prediction.

2.6 Limitations of Predictive Models

Some researchers contend that because predictive models are often developed based on the data generated from well controlled laboratory conditions, there are possibilities for the model to fail in predicting the behavior of target organisms in real food and environments which represent actual situations that occur during production, processing, and storage. Brocklehurst (2004) directs our attention towards the effects of food structure, including emulsions and surfaces which may significantly affect microbial behavior.

The predictive power of a model will always be constrained by the complexities of interactions of food microbes. Most of the models are developed based on only one targeted microorganism or at most a few strains in a homogeneous broth. But a variety of heterogeneous strains of the same organism may be present in the food and behave differently (Barbosa et al., 1994).

A model can be satisfactory only if it overestimates the observed growth. A model that over predicts the generation time or under estimates the growth rate is always 'fail-dangerous' (Ratkowsky, 2004).

A chosen model should be used to predict within the applicable boundary conditions and prediction should not be made beyond these (Ross, 1999).

2.7 Industrial Applications of Predictive Microbial Modeling

Presently, the predictive modeling technique is not being used at all by the meat industry. Even though they may be willing to use the one or two predictive models so far developed, there is no simple and easy way for the meat industry to incorporate these models into their online processing systems. An ANN model with the data informatics of salami processing variables would be very useful in 'online processing' for choosing the best combination of environmental variables for salami processing or predicting the residual population of *E.coli O157:H7* under any environmental conditions within a minute by a single person. Little time, labour or laboratory equipment is needed to complete the ANN analysis. In addition, the plate count requirement is eliminated.

Devices that can monitor environmental conditions, a tertiary model (spreadsheet program) that can convert the temperature history into estimates of microbial growth, a decision support expert system including software packages

(bioinformatics data bases and biomathematics models), set up together in the food industry can support on-line processing systems (McKellar and Lu 2004).

The 'front end' of the modeling process including data collection, model development, and model fitting are reviewed and discussed above. As a result of a substantial amount of work, this area of research now has a firm scientific foundation. While looking at the industrial application part i.e., the 'middle bit' like tertiary models, applications software, expert systems should be given attention in order to make this technology readily available to provide solutions to industrial problems. Potential users of this concept should adopt predictive modeling as a food safety management tool in the short term (McMeekin, 2004). A suitable model should be developed using numerical techniques for modeling the lag phase of bacteria. In spite of this limitation, predictive modeling can be a potentially valuable food safety management tool.

3.0 MATERIALS AND METHODS

3.1 Neural network software

Commercial neural network software, Neuroshell®2 (1993) (Release 4.0, Ward Systems Group, Inc.,), was used in this study. The purpose of using commercial software was that the industries could easily get and use this software. GRNN structure, which is a most suitable network for predictive microbiology purposes especially when regression is involved, is available in this software (Specht, 1991; Neuroshell®2, 1993; Bishop, 1995; Patterson, 1996; Jeyamkondan et al., 2001; and Statsoft, 2006). The other probabilistic neural networks were meant for classification purposes. As this was a user-friendly program, the user needed to know only the basics of neural networks and did not need to be an expert in programming neural network structures.

GRNN consisted of one input, one hidden and one output layer, with the number of neurons respectively, in each layer, equal to the number of input variables, the number of training sets and the number of output variables (Specht 1991). Different scaling functions (linear, logistic, or hyperbolic tangent) could be used to transform the variables at the input nodes. Each neuron of the hidden layer received input data from the input layer and computed the output by using transfer, estimator, or kernel functions (a Gaussian function defined by the parameter and standard deviation). This function was also known as the smoothing factor. A lower smoothing factor tightly fits (over-fit) the data and

higher smoothing factor loosely fits (under fit) the data. The appropriate selection of this smoothing factor determines the success of future predictions. If the smoothing factor is lower than needed for prediction, the GRNN would not output the predicted value for production sets. The smoothing factor used ranged from 0.01 to 1.0. The default terminal condition for the learning process was no improvement in mean squared error by at least 1 % for 20 successive reproductions of the whole training data set. This method prevented overtraining of networks and minimized memorizing problems. For feeding the unseen data, the production set was chosen as the last pattern input and the test set was randomly chosen for the rest of the seen data.

GRNN is randomly extracts the test set for every run of a software application, in order to validate the model. So, the model predictions will vary for each run. An analysis (Appendix II) was done on this issue and proved that there would not be any significant difference in the predictions for different runs.

3.2 SAS Program

The windows version of SAS (Statistical Analysis Systems, SAS Institute Inc., Cary, North Carolina) was used to develop a statistical model for this study. SAS version *9.1.3* was used. The advantage of selecting a SAS program was that it was easy to write and it took less than half a minute to run and generate the results.

3.3 Data

Data (Appendix 1) from Chacon (2006) for the reduction of *E.coli O157:H7* (36 data sets that are mean of three replicated values) in response to changes in the level of the natural antimicrobial allyl isothiocyanate (AIT) (0, 500, 750, and 1000 ppm), days of processing (0 to 45 days), water activity (0.864-0.948), and pH (4.70-5.57) were used to develop a GRNN model for survival of *E.coli* O157:H7. The product was inoculated with 6.45 $\log_{10}$CFU/g of *E.coli* O157:H7 on the day 0 of manufacturing with all the treatments. The fermentation temperature during salami production was 26°C (< 3 days) and drying was done at 13°C (25 days). These are common temperatures used by industry for salami production. Three replications of the parameters, population of *E.coli* O157:H7, water activity and pH were observed and the mean values were used as data for model development.

Another data set from Al-Nabulsi (2006) (Appendix VI) for the reduction of *E.coli* O157:H7(324 data sets) in response to changes in the level of the natural antimicrobial lactoferrin (LF) alone (0, 1.8, 3.0, and 6.0 mg / g sausage batter) or with different chelating agents namely sodium bicarbonate (SB) (0, 2.5, 5.0 mM) which was used in linear proportion with ethylene-diamine-tetraacetic-acid (EDTA) (0, 250, 500) for 0,1.8,6.0 levels of LF, for 0 to 28 days of processing were used to develop a GRNN model for predicting the survival of *E.coli* O157:H7. Six replications for each measurement were done using cefixime-

tellurite sorbitol McConkey agar (ct-SMAC) alone in the experiments. All those measurements were used for developing the models. The mean inoculation levels of *E.coli* O157:H7 on day 0 of manufacturing for the six treatments were 5.76, 5.79, 5.83, 5.65, 5.75, and 5.90, respectively. Normal LF was used for treatments 2 and 3. Paste-like micro capsules of LF were used for treatment 4. Dried powder microcapsules of LF were used for treatment 5 and 6. It was also observed in the experiment that the reduction of *E.coli* O157:H7 was due to cell injury (and not lethality) since significantly greater numbers of cells were recovered (Fig. 2 and Fig. 3) on All Purpose Tween (APT) agar overlaid with the selective medium ct-SMAC on the 28^{th} day of observation for all the treatments. The salami was manufactured in the pilot plant of the Department of Food Science at the University of Manitoba. The same method which is followed in Pillers' sausages plant, Waterloo, ON was adopted for manufacturing the salami.

3.4 GRNN modeling

An excel format of the data set (Chacon, 2006) was imported as input. The logarithmic transformation of *E.coli* O157:H7 numbers was modeled with GRNN. No function was used to scale the input values at the input nodes because the transformed value was given as input. The data set was split randomly into two groups (4:1), the best ratio (Jeyamkondan et al., 2001) as training and test sets for training and validating the GRNN model. An analysis (Appendix III) was also done and proved that 80:20 would be the best split ratio to be adopted. The

scaling function 'none' was used for equalization. The production set (unseen data for modeling) was created in such a way that the predicted value would be near to 1.45 (a 5 log reduction from the initial count 6.45) at 28 and 35 days. The production set (Table 1) was fed as last patterns along with training and test set. The maximum smoothing factor 1.0 was chosen as the other values less than 1.0 could not produce the predicted values for all patterns of the production set. The configuration was saved. The output file contained the GRNN predictions for the training, test and production sets.

3.4.1 Creation of Production set for Chacon's data

Table 1: Production set (man made unseen data) for future model predictions of survival of *E.coli* O157:H7 based on Chacon's Data

Aw	AIT in ppm	pH	No. of days
0.89	500	4.93	28
0.89	505	4.93	28
0.89	510	4.93	28
0.89	515	4.93	28
0.89	520	4.93	28
0.89	530	4.93	28
0.89	540	4.93	28
0.89	550	4.93	28
0.89	560	4.93	28
0.89	570	4.93	28
0.89	580	4.93	28
0.877	500	4.93	35
0.877	490	4.93	35
0.877	480	4.93	35
0.877	470	4.93	35
0.877	460	4.93	35
0.877	450	4.93	35
0.877	440	4.93	35
0.877	430	4.93	35
0.877	420	4.93	35
0.877	410	4.93	35

Of 19 data sets (Table 1), 10 data sets are targeting to know the *E.coli* population at the 35th day and 9 data target the 28th day. The AIT levels have been chosen from 500 to 410 ppm at 10 ppm intervals for 35 days and from 500 to 580 ppm at 10 ppm intervals for 28 days to know the combination of factors which can reduce *E.coli* by 5 $\log_{10}$ CFU/g *E.coli* cells. So, the value of the *E.coli* target population of interest is 1.45 (initial population 6.45 minus 5.00). The water activity levels 0.89 and 0.877 represent, respectively, levels achieved at 28 and 35 days during processing of salami.

The same procedure was followed for the other data set (Al-Nabulsi, 2006) used for developing the validated GRNN model. The default initial smoothing factor was 0.3. The production set was not created to find a 5 log reduction because the highest reduction achieved in this experiment was 4.2 $\log_{10}$CFU/g and the GRNN could not perform an extrapolation. So, the modeling was done only for the experimentally observed values.

3.5 Statistical Modeling

Chacon data: A polynomial regression modeling technique was followed to develop the model and describe the effect of the parameters AIT, days of processing, water activity and pH. The interaction between these parameters

was also taken into account in developing the model. All 36 data sets were used to develop this model.

Al-Nabulsi data: The same polynomial multiple linear regression procedure was followed here too. All 324 data sets were used to develop the statistical model. The contributions of each individual factor LF, SB, EDTA and Day; the interactions of LF, Day with the interaction of SB and EDTA; LF with Day; and interaction of all these four factors in the model was estimated using the SAS Reg procedure, and then the model was formulated. As there was no production set, future prediction was not done using this model.

3.6 Criteria for comparison

The same criteria used by Jeyamkondan et al. (2001) for comparing the GRNN model and statistical models were also used in this study. They are briefly explained (see Table 2 for mathematical expression) and here for clarity.

Graphical plots: A bias plot is the graph of predicted versus observed values to show the region of over-prediction and under-prediction. A plot between residuals and the observed values is the best way of comparing competitive models as it gives a clear picture of the magnitude and distribution of residuals (Ratkowsky, 1990). Mean relative percentage residual (MRPR) is equal to zero when there is no bias in the prediction; a positive value occurs when there is under-prediction and *vice-versa*. Bias factor: When the model has no bias, it will be equal to 1. If it

is greater than 1, it indicates the model overestimates survival of the pathogen. If it is less than 1, it indicates the model underestimates the bacterial numbers. Mean absolute relative residual (MARR) is the percentage by which the predicted values deviate from observed values (either above or below). The accuracy factor indicates the average deviation of the predicted values from the observed values. Root Mean Square Residual (RMSR): If it is 0, it indicates there is no bias between predicted and observed values. While considering the reduction of *E.coli,* an under-prediction is considered as 'fail-dangerous' and over-prediction as 'fail-safe'.

Table 2: Mathematical expressions of statistical indices for comparing the model prediction performance

Statistical indices	Mathematical Expressions
Mean Relative Percentage Residual (MRPR)	$1/N * \sum(O-P)*100 / O$
Bias Factor	$A.\log10 \sum\log10(P/O)/N$
Mean Absolute Percentage Residual (MARR)	$1/N * \sum \lvert O-P \rvert *100 / O$
Accuracy Factor	$A.\log10 \sum \lvert \log10(P/O) \rvert /N$
Root Mean Square Residual	$\sqrt{\sum}(O-P)2/N$

O – Observed value; P – Predicted value

4.0 RESULTS AND DISCUSSIONS

4.1 GRNN Models

Chacon's data: GRNN converged to the local minima (smoothing factor) as 0.9650588 and gave the predictions for all training, test (Table 3) and production patterns (Table 4). The value of the correlation factor R^2 was 0.869. The correlation coefficient was 0.9588. The GRNN model curves for the combined training and test sets and the trend of observed values are shown in Fig. 1.

Since the data set having the mean values was used for the GRNN model development, there needs to be justification that the model developed with mean data will be able to predict the values for the whole size range (for all three replications) of each mean value. This was analyzed and reported in Appendix IV.

Al-Nabulsi's data: GRNN finally converged to local minima of 0.96117646 and gave the predictions for all training and test set (Table 5). The R^2 value was 0.858. The GRNN model curves for the combined training and test sets and the trend of observed values of the treatments 2, 3, and 6 (Fig. 2) and of the treatments 1, 4, and 5 (Fig. 2) are separately shown in the figures.

Table 3: Chacon's data and the predictions of the GRNN model for training and test sets and of statistical model for its full set

Aw	AIT in ppm	pH	No. of days	*E.coli* population log CFU/g		
				Observed	GRNN	Statistical
					Training set	**Full set**
0.948	0	5.57	0	6.453	6.453	6.591
0.936	0	4.70	6	5.140	5.140	5.000
0.908	0	4.79	16	4.670	4.670	4.049
0.88	0	4.93	28	4.583	4.583	4.495
0.878	0	4.93	35	4.117	4.117	3.923
0.874	0	4.93	40	4.079	4.079	3.843
0.87	0	4.93	45	4.023	4.023	3.849
0.943	500	5.57	0	6.453	6.453	6.702
0.938	500	4.81	3	4.102	4.102	4.345
0.939	500	4.81	6	4.112	4.103	3.671
0.91	500	4.80	16	2.280	2.268	1.994
0.89	500	4.93	28	1.702	1.686	1.444
0.877	500	4.93	35	1.068	1.058	1.229
0.873	500	4.93	40	0.000	0.000	0.961
0.87	500	4.93	45	0.000	0.000	0.690
0.946	750	5.57	0	6.453	6.453	6.819
0.925	750	4.71	6	3.165	3.170	2.946
0.9	750	4.81	9	2.162	2.161	2.183
0.902	750	4.80	16	1.068	1.069	1.017
0.886	750	5.01	28	0.000	0.016	1.208
0.873	750	5.01	35	0.000	0.010	0.887
0.94	1000	4.87	3	3.455	3.455	3.420
0.923	1000	4.80	6	2.704	2.708	2.180
0.911	1000	4.86	9	1.962	1.964	1.485
0.905	1000	4.87	16	0.000	0.010	0.379
0.874	1000	5.01	28	0.000	0.000	0.026
0.87	1000	5.01	35	0.000	0.000	-0.355
0.868	1000	5.01	40	0.000	0.000	-0.591
0.864	1000	5.01	45	0.000	0.000	-0.835
					Test set	
0.941	0	4.87	3	4.916	4.916	5.196
0.92	0	4.78	9	4.735	5.140	4.842
0.918	500	4.77	9	2.580	2.162	3.183
0.936	750	4.78	3	3.550	3.779	3.768
0.87	750	5.01	40	0.000	0.000	0.663
0.866	750	5.01	45	0.000	0.000	0.494
0.941	1000	5.57	0	6.453	6.453	6.618

Data source, Chacon (2006)

Table 4: Results of GRNN and statistical model predictions of *E.coli* O157:H7 population in log_{10}CFU/g for production sets (man made unseen data) for industrial use

Aw	AIT in ppm	pH	No. of days	*E.coli* population log_{10} CFU/g	
				GRNN	Statistical
				Production set	
0.89	500	4.93	28	1.686	1.444
0.89	505	4.93	28	*	1.423
0.89	510	4.93	28	1.679	1.401
0.89	515	4.93	28	*	1.380
0.89	520	4.93	28	1.669	1.358
0.89	530	4.93	28	1.654	**
0.89	540	4.93	28	1.633	**
0.89	550	4.93	28	1.604	**
0.89	560	4.93	28	1.563	**
0.89	570	4.93	28	1.508	**
0.89	**580**	**4.93**	**28**	**1.434**	**
0.877	500	4.93	35	1.058	1.229
0.877	490	4.93	35	1.061	1.285
0.877	480	4.93	35	1.063	1.341
0.877	470	4.93	35	1.065	1.397
0.877	**460**	**4.93**	**35**	1.066	**1.453**
0.877	450	4.93	35	1.066	1.509
0.877	440	4.93	35	1.067	**
0.877	430	4.93	35	1.067	**
0.877	420	4.93	35	1.068	**
0.877	410	4.93	35	1.068	**

* - corresponding input set was not applied to GRNN model.

**- corresponding input set was not applied to statistical model.

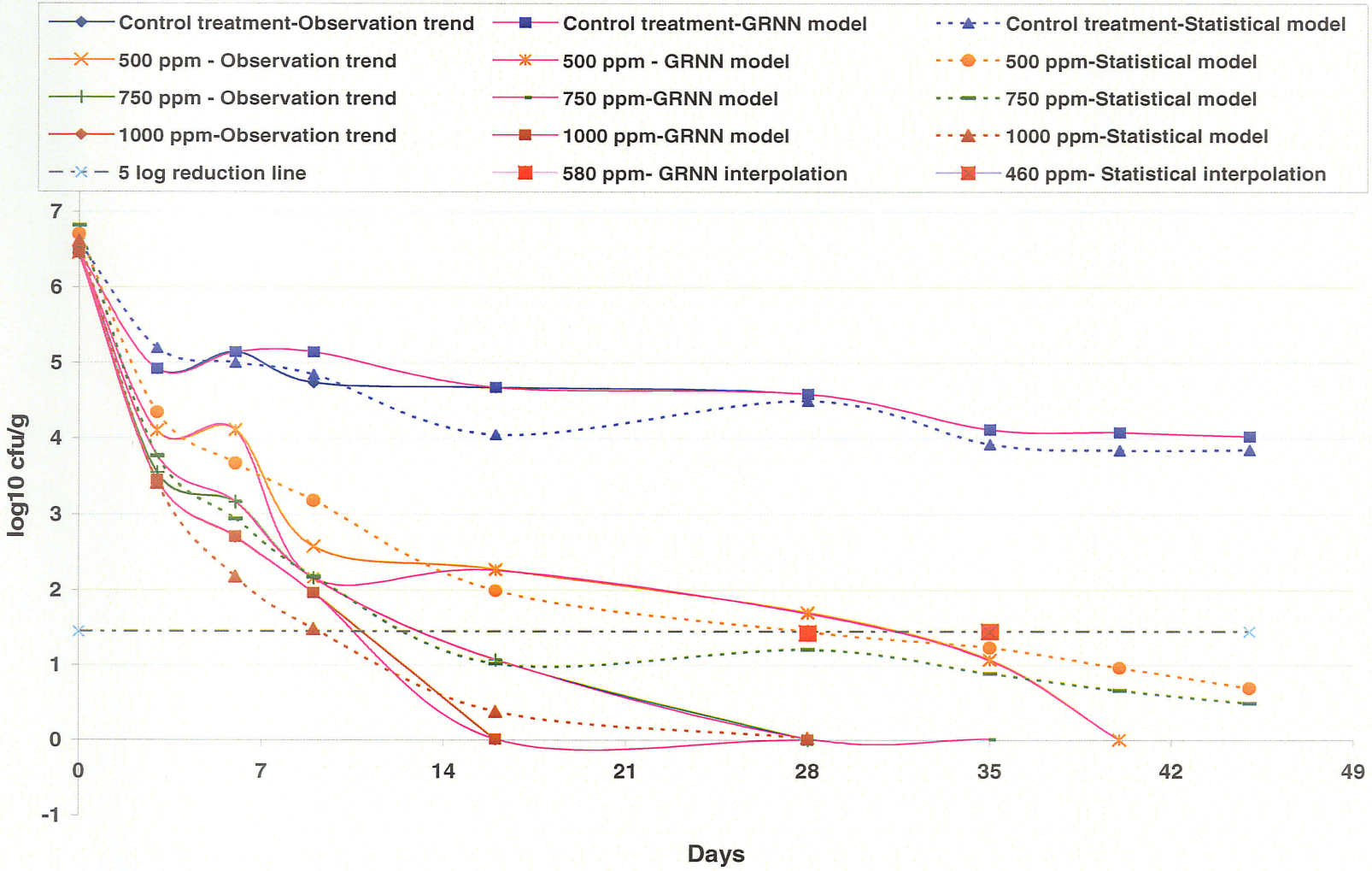

Fig. 1: Observation trend, GRNN and Statistical models of survival of E.coli O157:H7 for Control (0 ppm), 500 ppm, 750 ppm, and 1000 ppm of allyl isothiocyanate.

Table 5: The Al-Nabulsi's data and the predictions of the GRNN model for training and test set and of statistical model for its full set

Treatment #	mg LF / g sausage batter	EDTA ppm	mM SB	Days	Replications	*E.coli* population log10 CFU/g		
						Observed	GRNN Training set	Statistical Full set
Trt1	0	0	0	0	2	5.78	5.65	5.17
Trt1	0	0	0	0	3	5.30	5.65	5.17
Trt1	0	0	0	0	4	6.09	5.65	5.17
Trt1	0	0	0	0	5	5.82	5.65	5.17
Trt1	0	0	0	0	6	5.64	5.65	5.17
Trt1	0	0	0	1	2	5.03	4.95	5.09
Trt1	0	0	0	1	3	4.94	4.95	5.09
Trt1	0	0	0	1	6	4.30	4.95	5.09
Trt1	0	0	0	2	2	4.96	4.85	5.01
Trt1	0	0	0	2	3	4.88	4.85	5.01
Trt1	0	0	0	3	1	4.86	4.72	4.92
Trt1	0	0	0	3	2	4.96	4.72	4.92
Trt1	0	0	0	3	3	4.58	4.72	4.92
Trt1	0	0	0	3	4	4.48	4.72	4.92
Trt1	0	0	0	3	5	4.65	4.72	4.92
Trt1	0	0	0	6	1	4.40	4.39	4.67
Trt1	0	0	0	6	3	4.57	4.39	4.67
Trt1	0	0	0	6	4	4.58	4.39	4.67
Trt1	0	0	0	6	5	4.20	4.39	4.67
Trt1	0	0	0	6	6	4.20	4.39	4.67
Trt1	0	0	0	9	1	4.72	4.00	4.43
Trt1	0	0	0	9	2	3.86	4.00	4.43
Trt1	0	0	0	9	3	3.92	4.00	4.43
Trt1	0	0	0	9	4	3.48	4.00	4.43
Trt1	0	0	0	15	1	4.47	3.91	3.93
Trt1	0	0	0	15	2	3.68	3.91	3.93
Trt1	0	0	0	15	4	3.62	3.91	3.93
Trt1	0	0	0	15	5	3.97	3.91	3.93
Trt1	0	0	0	15	6	3.81	3.91	3.93
Trt1	0	0	0	21	1	3.20	3.35	3.44
Trt1	0	0	0	21	3	3.31	3.35	3.44
Trt1	0	0	0	21	4	3.53	3.35	3.44
Trt1	0	0	0	28	1	3.00	3.22	2.86
Trt1	0	0	0	28	2	3.15	3.22	2.86
Trt1	0	0	0	28	3	3.29	3.22	2.86
Trt1	0	0	0	28	4	3.34	3.22	2.86

Table 5 Continued

Treatment #	mg LF / g sausage batter	EDTA ppm	mM SB	Days	Replications	*E.coli* population log10 CFU/g		
						Observed	GRNN	Statistical
							Training set	Full set
Trt1	0	0	0	28	5	3.50	3.22	2.86
Trt1	0	0	0	28	6	3.04	3.22	2.86
Trt-2	6	0	0	0	1	5.91	5.72	5.16
Trt2	6	0	0	0	2	5.60	5.72	5.16
Trt2	6	0	0	0	3	5.53	5.72	5.16
Trt2	6	0	0	0	4	5.75	5.72	5.16
Trt2	6	0	0	0	5	5.94	5.72	5.16
Trt2	6	0	0	0	6	5.99	5.72	5.16
Trt2	6	0	0	1	1	5.19	4.93	5.01
Trt2	6	0	0	1	3	5.06	4.93	5.01
Trt2	6	0	0	1	6	4.10	4.93	5.01
Trt2	6	0	0	2	1	4.84	4.50	4.87
Trt2	6	0	0	2	2	4.69	4.50	4.87
Trt2	6	0	0	2	3	4.82	4.50	4.87
Trt2	6	0	0	2	4	4.15	4.50	4.87
Trt2	6	0	0	2	5	3.90	4.50	4.87
Trt2	6	0	0	3	1	4.84	4.47	4.72
Trt2	6	0	0	3	3	4.82	4.47	4.72
Trt2	6	0	0	3	4	4.30	4.47	4.72
Trt2	6	0	0	3	6	3.90	4.47	4.72
Trt2	6	0	0	6	2	4.28	4.40	4.29
Trt2	6	0	0	6	4	4.30	4.40	4.29
Trt2	6	0	0	6	5	4.57	4.40	4.29
Trt2	6	0	0	6	6	4.44	4.40	4.29
Trt2	6	0	0	9	1	3.60	3.63	3.85
Trt2	6	0	0	9	2	3.62	3.63	3.85
Trt2	6	0	0	9	3	3.45	3.63	3.85
Trt2	6	0	0	9	4	3.81	3.63	3.85
Trt2	6	0	0	9	5	3.66	3.63	3.85
Trt2	6	0	0	15	2	3.45	3.48	2.98
Trt2	6	0	0	15	3	3.45	3.48	2.98
Trt2	6	0	0	15	4	3.20	3.48	2.98
Trt2	6	0	0	15	5	3.83	3.48	2.98
Trt2	6	0	0	15	6	3.48	3.48	2.98
Trt2	6	0	0	21	6	1.00	1.00	2.11
Trt2	6	0	0	28	2	2.08	1.66	1.09
Trt2	6	0	0	28	3	1.58	1.66	1.09
Trt2	6	0	0	28	4	1.00	1.66	1.09
Trt2	6	0	0	28	5	1.45	1.66	1.09

Table 5 Continued

Treatment #	mg LF / g sausage batter	EDTA ppm	mM SB	Days	Replications	*E.coli* population log10 CFU/g		
						Observed	GRNN Training set	Statistical Full set
Trt2	6	0	0	28	6	2.20	1.66	1.09
Trt3	6	500	5	0	1	5.84	5.76	5.04
Trt3	6	500	5	0	2	6.09	5.76	5.04
Trt3	6	500	5	0	3	5.48	5.76	5.04
Trt3	6	500	5	0	4	5.91	5.76	5.04
Trt3	6	500	5	0	5	5.92	5.76	5.04
Trt3	6	500	5	0	6	5.76	5.76	5.04
Trt3	6	500	5	1	1	5.52	5.21	5.00
Trt3	6	500	5	1	3	4.87	5.21	5.00
Trt3	6	500	5	1	4	5.23	5.21	5.00
Trt3	6	500	5	1	5	4.97	5.21	5.00
Trt3	6	500	5	1	6	5.17	5.21	5.00
Trt3	6	500	5	2	1	4.95	4.96	4.96
Trt3	6	500	5	2	2	4.76	4.96	4.96
Trt3	6	500	5	2	3	4.91	4.96	4.96
Trt3	6	500	5	2	4	5.12	4.96	4.96
Trt3	6	500	5	2	5	4.88	4.96	4.96
Trt3	6	500	5	2	6	4.92	4.96	4.96
Trt3	6	500	5	3	1	4.94	5.01	4.91
Trt3	6	500	5	3	2	5.11	5.01	4.91
Trt3	6	500	5	3	3	5.05	5.01	4.91
Trt3	6	500	5	3	4	5.24	5.01	4.91
Trt3	6	500	5	3	5	5.09	5.01	4.91
Trt3	6	500	5	3	6	4.70	5.01	4.91
Trt3	6	500	5	6	2	4.89	4.13	4.78
Trt3	6	500	5	6	3	3.34	4.13	4.78
Trt3	6	500	5	6	4	3.50	4.13	4.78
Trt3	6	500	5	6	5	4.48	4.13	4.78
Trt3	6	500	5	6	6	4.46	4.13	4.78
Trt3	6	500	5	9	2	3.60	3.69	4.66
Trt3	6	500	5	9	3	4.36	3.69	4.66
Trt3	6	500	5	9	4	4.35	3.69	4.66
Trt3	6	500	5	9	5	3.08	3.69	4.66
Trt3	6	500	5	9	6	3.06	3.69	4.66
Trt3	6	500	5	15	1	4.59	4.50	4.40
Trt3	6	500	5	15	2	4.72	4.50	4.40
Trt3	6	500	5	15	3	4.48	4.50	4.40
Trt3	6	500	5	15	4	4.40	4.50	4.40
Trt3	6	500	5	15	6	4.32	4.50	4.40

Table 5 Continued

Treatment #	mg LF / g sausage batter	EDTA ppm	mM SB	Days	Replications	*E.coli* population log10 cFU/g		
						Observed	GRNN Training set	Statistical Full set
Trt3	6	500	5	21	2	4.20	4.34	4.14
Trt3	6	500	5	21	3	4.40	4.34	4.14
Trt3	6	500	5	21	4	4.50	4.34	4.14
Trt3	6	500	5	21	5	4.10	4.34	4.14
Trt3	6	500	5	21	6	4.50	4.34	4.14
Trt3	6	500	5	28	1	3.50	4.07	3.84
Trt3	6	500	5	28	2	3.73	4.07	3.84
Trt3	6	500	5	28	3	4.60	4.07	3.84
Trt3	6	500	5	28	4	4.44	4.07	3.84
Trt3	6	500	5	28	5	3.97	4.07	3.84
Trt3	6	500	5	28	6	4.17	4.07	3.84
Trt4	3	0	0	0	1	5.83	5.63	5.16
Trt4	3	0	0	0	2	5.69	5.63	5.16
Trt4	3	0	0	0	4	5.72	5.63	5.16
Trt4	3	0	0	0	5	5.66	5.63	5.16
Trt4	3	0	0	1	1	5.08	5.04	5.05
Trt4	3	0	0	1	2	4.99	5.04	5.05
Trt4	3	0	0	1	3	5.06	5.04	5.05
Trt4	3	0	0	1	4	5.01	5.04	5.05
Trt4	3	0	0	2	1	4.66	4.62	4.94
Trt4	3	0	0	2	2	4.47	4.62	4.94
Trt4	3	0	0	2	3	4.64	4.62	4.94
Trt4	3	0	0	2	4	4.61	4.62	4.94
Trt4	3	0	0	2	5	4.72	4.62	4.94
Trt4	3	0	0	2	6	4.51	4.62	4.94
Trt4	3	0	0	3	1	4.68	4.45	4.82
Trt4	3	0	0	3	2	4.60	4.45	4.82
Trt4	3	0	0	3	3	4.56	4.45	4.82
Trt4	3	0	0	3	4	4.71	4.45	4.82
Trt4	3	0	0	3	5	4.72	4.45	4.82
Trt4	3	0	0	3	6	3.30	4.45	4.82
Trt4	3	0	0	6	1	4.52	4.53	4.48
Trt4	3	0	0	6	2	4.56	4.53	4.48
Trt4	3	0	0	6	3	4.52	4.53	4.48
Trt4	3	0	0	6	4	4.68	4.53	4.48
Trt4	3	0	0	6	5	4.58	4.53	4.48
Trt4	3	0	0	6	6	4.31	4.53	4.48
Trt4	3	0	0	9	2	4.00	4.04	4.14
Trt4	3	0	0	9	3	4.15	4.04	4.14

Table 5 Continued

Treatment #	mg LF / g sausage batter	EDTA ppm	mM SB	Days	Replications	*E.coli* population log10 CFU/g		
						Observed	GRNN	Statistical
							Training set	Full set
Trt4	3	0	0	9	4	3.83	4.04	4.14
Trt4	3	0	0	9	5	3.45	4.04	4.14
Trt4	3	0	0	9	6	4.73	4.04	4.14
Trt4	3	0	0	15	1	3.58	3.68	3.45
Trt4	3	0	0	15	3	3.53	3.68	3.45
Trt4	3	0	0	15	4	3.66	3.68	3.45
Trt4	3	0	0	15	5	3.84	3.68	3.45
Trt4	3	0	0	15	6	3.75	3.68	3.45
Trt4	3	0	0	21	2	1.80	1.92	2.77
Trt4	3	0	0	21	3	2.00	1.92	2.77
Trt4	3	0	0	21	4	1.70	1.92	2.77
Trt4	3	0	0	21	5	1.75	1.92	2.77
Trt4	3	0	0	21	6	2.30	1.92	2.77
Trt4	3	0	0	28	1	2.45	2.13	1.97
Trt4	3	0	0	28	3	2.38	2.13	1.97
Trt4	3	0	0	28	4	2.20	2.13	1.97
Trt4	3	0	0	28	5	1.48	2.13	1.97
Trt5	1.8	0	0	0	2	5.73	5.65	5.17
Trt5	1.8	0	0	0	3	5.56	5.65	5.17
Trt5	1.8	0	0	0	4	5.83	5.65	5.17
Trt5	1.8	0	0	0	6	5.79	5.65	5.17
Trt5	1.8	0	0	1	1	5.13	5.15	5.06
Trt5	1.8	0	0	1	4	4.98	5.15	5.06
Trt5	1.8	0	0	1	5	5.26	5.15	5.06
Trt5	1.8	0	0	1	6	5.18	5.15	5.06
Trt5	1.8	0	0	2	1	4.93	4.75	4.96
Trt5	1.8	0	0	2	3	4.75	4.75	4.96
Trt5	1.8	0	0	2	5	4.95	4.75	4.96
Trt5	1.8	0	0	2	6	4.20	4.75	4.96
Trt5	1.8	0	0	3	1	4.68	4.66	4.86
Trt5	1.8	0	0	3	2	4.82	4.66	4.86
Trt5	1.8	0	0	3	4	4.88	4.66	4.86
Trt5	1.8	0	0	3	5	5.13	4.66	4.86
Trt5	1.8	0	0	3	6	3.78	4.66	4.86
Trt5	1.8	0	0	6	1	4.50	4.64	4.56
Trt5	1.8	0	0	6	2	4.60	4.64	4.56
Trt5	1.8	0	0	6	4	4.64	4.64	4.56
Trt5	1.8	0	0	6	5	4.76	4.64	4.56
Trt5	1.8	0	0	6	6	4.71	4.64	4.56

Table 5 Continued

Treatment #	mg LF / g sausage batter	EDTA ppm	mM SB	Days	Replications	*E.coli* population log10 CFU/g		
						Observed	GRNN	Statistical
							Training set	Full set
Trt5	1.8	0	0	9	1	4.47	5.04	4.25
Trt5	1.8	0	0	9	3	5.13	5.04	4.25
Trt5	1.8	0	0	9	4	5.38	5.04	4.25
Trt5	1.8	0	0	9	5	5.22	5.04	4.25
Trt5	1.8	0	0	9	6	5.05	5.04	4.25
Trt5	1.8	0	0	15	1	4.68	4.52	3.65
Trt5	1.8	0	0	15	3	4.58	4.52	3.65
Trt5	1.8	0	0	15	4	4.64	4.52	3.65
Trt5	1.8	0	0	15	5	4.54	4.52	3.65
Trt5	1.8	0	0	15	6	4.20	4.52	3.65
Trt5	1.8	0	0	21	1	2.60	2.62	3.04
Trt5	1.8	0	0	21	2	2.78	2.62	3.04
Trt5	1.8	0	0	21	3	3.58	2.62	3.04
Trt5	1.8	0	0	21	4	4.15	2.62	3.04
Trt5	1.8	0	0	21	6	0.00	2.62	3.04
Trt5	1.8	0	0	28	1	2.94	2.38	2.33
Trt5	1.8	0	0	28	2	2.83	2.38	2.33
Trt5	1.8	0	0	28	3	2.08	2.38	2.33
Trt5	1.8	0	0	28	5	2.48	2.38	2.33
Trt5	1.8	0	0	28	6	1.60	2.38	2.33
Trt6	1.8	250	2.5	0	1	5.78	5.79	5.45
Trt6	1.8	250	2.5	0	2	6.53	5.79	5.45
Trt6	1.8	250	2.5	0	3	5.92	5.79	5.45
Trt6	1.8	250	2.5	0	4	6.15	5.79	5.45
Trt6	1.8	250	2.5	0	5	6.44	5.79	5.45
Trt6	1.8	250	2.5	0	6	4.60	5.79	5.45
Trt6	1.8	250	2.5	1	1	5.25	5.11	5.33
Trt6	1.8	250	2.5	1	2	4.95	5.11	5.33
Trt6	1.8	250	2.5	1	3	4.91	5.11	5.33
Trt6	1.8	250	2.5	1	4	5.19	5.11	5.33
Trt6	1.8	250	2.5	1	5	4.96	5.11	5.33
Trt6	1.8	250	2.5	1	6	4.80	5.11	5.33
Trt6	1.8	250	2.5	2	1	5.07	4.92	5.22
Trt6	1.8	250	2.5	2	2	5.06	4.92	5.22
Trt6	1.8	250	2.5	2	3	5.18	4.92	5.22
Trt6	1.8	250	2.5	2	4	4.96	4.92	5.22
Trt6	1.8	250	2.5	2	5	4.91	4.92	5.22
Trt6	1.8	250	2.5	2	6	4.50	4.92	5.22
Trt6	1.8	250	2.5	3	2	4.88	4.65	5.10

Table 5 Continued

Treatment #	mg LF / g sausage batter	EDTA ppm	mM SB	Days	Replications	*E.coli* population log10 cFU/g		
						Observed	GRNN	Statistical
							Training set	Full set
Trt6	1.8	250	2.5	3	3	5.04	4.65	5.10
Trt6	1.8	250	2.5	3	4	4.69	4.65	5.10
Trt6	1.8	250	2.5	3	5	4.61	4.65	5.10
Trt6	1.8	250	2.5	3	6	3.78	4.65	5.10
Trt6	1.8	250	2.5	6	1	4.68	4.75	4.75
Trt6	1.8	250	2.5	6	2	4.72	4.75	4.75
Trt6	1.8	250	2.5	6	3	4.66	4.75	4.75
Trt6	1.8	250	2.5	6	4	4.53	4.75	4.75
Trt6	1.8	250	2.5	6	5	5.10	4.75	4.75
Trt6	1.8	250	2.5	6	6	4.79	4.75	4.75
Trt6	1.8	250	2.5	9	1	4.67	4.71	4.41
Trt6	1.8	250	2.5	9	2	4.88	4.71	4.41
Trt6	1.8	250	2.5	9	3	4.83	4.71	4.41
Trt6	1.8	250	2.5	9	4	4.99	4.71	4.41
Trt6	1.8	250	2.5	9	6	4.20	4.71	4.41
Trt6	1.8	250	2.5	15	3	5.05	4.98	3.71
Trt6	1.8	250	2.5	15	4	4.89	4.98	3.71
Trt6	1.8	250	2.5	15	5	4.99	4.98	3.71
Trt6	1.8	250	2.5	21	1	3.15	2.33	3.02
Trt6	1.8	250	2.5	21	2	3.53	2.33	3.02
Trt6	1.8	250	2.5	21	3	3.82	2.33	3.02
Trt6	1.8	250	2.5	21	4	3.50	2.33	3.02
Trt6	1.8	250	2.5	21	5	0.00	2.33	3.02
Trt6	1.8	250	2.5	21	6	0.00	2.33	3.02
Trt6	1.8	250	2.5	28	2	1.90	2.24	2.21
Trt6	1.8	250	2.5	28	3	1.70	2.24	2.21
Trt6	1.8	250	2.5	28	4	1.60	2.24	2.21
Trt6	1.8	250	2.5	28	5	2.78	2.24	2.21
Trt6	1.8	250	2.5	28	6	3.23	2.24	2.21
							Test Set	
Trt1	0	0	0	0	1	5.93	5.65	5.17
Trt1	0	0	0	1	1	5.50	4.95	5.09
Trt1	0	0	0	1	4	4.60	4.95	5.09
Trt1	0	0	0	1	5	4.78	4.95	5.09
Trt1	0	0	0	2	1	4.86	4.85	5.01
Trt1	0	0	0	2	4	4.45	4.85	5.01
Trt1	0	0	0	2	5	4.76	4.85	5.01
Trt1	0	0	0	2	6	4.81	4.85	5.01
Trt1	0	0	0	3	6	4.77	4.72	4.92

Table 5 Continued

Treatment #	mg LF / g sausage batter	EDTA ppm	mM SB	Days	Replications	*E.coli* population log10 cFU/g		
						Observed	GRNN	Statistical
							Test set	Full set
Trt1	0	0	0	6	2	4.20	4.39	4.67
Trt1	0	0	0	9	5	4.00	4.00	4.43
Trt1	0	0	0	9	6	4.00	4.00	4.43
Trt1	0	0	0	15	3	3.83	3.91	3.93
Trt1	0	0	0	21	2	3.51	3.35	3.44
Trt1	0	0	0	21	5	3.68	3.35	3.44
Trt1	0	0	0	21	6	3.28	3.35	3.44
Trt2	6	0	0	1	2	5.23	4.93	5.01
Trt2	6	0	0	1	4	5.08	4.93	5.01
Trt2	6	0	0	1	5	4.40	4.93	5.01
Trt2	6	0	0	2	6	4.76	4.50	4.87
Trt2	6	0	0	3	2	4.69	4.47	4.72
Trt2	6	0	0	3	5	4.15	4.47	4.72
Trt2	6	0	0	6	1	4.16	4.40	4.29
Trt2	6	0	0	6	3	4.29	4.40	4.29
Trt2	6	0	0	9	6	3.59	3.63	3.85
Trt2	6	0	0	15	1	3.48	3.48	2.98
Trt2	6	0	0	21	1	2.00	1.00	2.11
Trt2	6	0	0	21	2	2.30	1.00	2.11
Trt2	6	0	0	21	3	1.00	1.00	2.11
Trt2	6	0	0	21	4	1.00	1.00	2.11
Trt2	6	0	0	21	5	1.48	1.00	2.11
Trt2	6	0	0	28	1	1.60	1.66	1.09
Trt3	6	500	5	1	2	4.99	5.21	5.00
Trt3	6	500	5	6	1	4.67	4.13	4.78
Trt3	6	500	5	9	1	3.90	3.69	4.66
Trt3	6	500	5	15	5	4.44	4.50	4.40
Trt3	6	500	5	21	1	4.15	4.34	4.14
Trt4	3	0	0	0	3	5.63	5.63	5.16
Trt4	3	0	0	0	6	5.39	5.63	5.16
Trt4	3	0	0	1	5	4.79	5.04	5.05
Trt4	3	0	0	1	6	4.66	5.04	5.05
Trt4	3	0	0	9	1	4.00	4.04	4.14
Trt4	3	0	0	15	2	3.68	3.68	3.45
Trt4	3	0	0	21	1	2.45	1.92	2.77
Trt4	3	0	0	28	2	1.30	2.13	1.97
Trt4	3	0	0	28	6	1.30	2.13	1.97
Trt5	1.8	0	0	0	1	5.75	5.65	5.17
Trt5	1.8	0	0	0	5	5.82	5.65	5.17

Table 5 Continued

Treatment #	mg LF / g sausage batter	EDTA ppm	mM SB	Days	Replications	*E.coli* population log10 CFU/g		
						Observed	GRNN	Statistical
							Test set	Full set
Trt5	1.8	0	0	1	2	5.34	5.15	5.06
Trt5	1.8	0	0	1	3	5.06	5.15	5.06
Trt5	1.8	0	0	2	2	4.80	4.75	4.96
Trt5	1.8	0	0	2	4	4.53	4.75	4.96
Trt5	1.8	0	0	3	3	4.67	4.66	4.86
Trt5	1.8	0	0	6	3	4.67	4.64	4.56
Trt5	1.8	0	0	9	2	4.67	5.04	4.25
Trt5	1.8	0	0	15	2	4.29	4.52	3.65
Trt5	1.8	0	0	21	5	3.30	2.62	3.04
Trt5	1.8	0	0	28	4	1.60	2.38	2.33
Trt6	1.8	250	2.5	3	1	4.85	4.65	5.10
Trt6	1.8	250	2.5	9	5	3.98	4.71	4.41
Trt6	1.8	250	2.5	15	1	5.02	4.98	3.71
Trt6	1.8	250	2.5	15	2	4.77	4.98	3.71
Trt6	1.8	250	2.5	15	6	4.87	4.98	3.71
Trt6	1.8	250	2.5	28	1	1.30	2.24	2.21

Source: Al-Nabulsi (2006).

4.2 Statistical Models

Chacon's Data: As the days of processing were increased, the *E.coli* O157:H7 numbers were decreased due to anti-microbial effects of processing and use of AIT. This effect was higher at high concentrations of AIT. The pH was reduced from 5.57 to around 4.70 and then slightly increased at week 5. Thus, it was assumed that a quadratic relationship existed between the *E.coli* O157:H7 numbers and pH. The water activity was reduced from 0.948 to 0.864. So, a negative linear relationship existed between water activity and *E.coli* O157:H7

numbers. The interaction between factors played a major role in determining the reduction of *E.coli* O157:H7.

The Regression Procedure of SAS version 9.1.3 was used to develop the statistical model. For the production set the near interpolation region was created in such a way that the reduction of the Y value was 5 $\log_{10}$CFU at 28 and 35 days from the initial value, and the *E.coli* O157:H7 numbers were predicted using this model. One pattern among the created unseen data set (Table 1) which gave a Y value as 1.45 (5 log reduction from the initial count of 6.45) was chosen for use with this model. The statistical polynomial model developed was as follows:

$$Y = 256.85 - 0.082X1 + 0.778X2 - 127.44X3 - 69.951X4 - 0.0012X1X2 + 0.071X1X3 + 0.003X1X4 - 2.144X2X3 + 0.202X2X4 + 19.745X3X4 + 5.026X4X4 + 0.003X1X2X3X4 \quad \text{(Eq.4)}$$

Where Y = 10 based logarithm of the *E.coli* O157:H7 number;

X1 – AIT in ppm;

X2 – number of days of processing;

X3 – water activity; and

X4 – pH.

The correlation factor (R^2) for this model was 0.964.

Model predictions are shown in Table 3 and Fig.1.

It was understood from the SAS program result (Appendix V) that none of the individual factors or interaction of factors significantly contributed to the reduction of the pathogen except the interaction of AIT and pH (Probability of t value is less than 0.05). It was true because the AIT use was more effective at pH values below 5.0 to reduce the numbers of the pathogen. Low pH by itself contributed to the reduction of pathogen.

When we consider other factors which have a probability of t value less than 0.30, the factor AIT alone (<0.09), pH alone (<0.12), the interaction of AIT and water activity (<0.16), the quadratic pH with time (<0.29) contributed for the reduction of the pathogen.

Because the interaction of water activity and AIT has contributed, to pathogen reduction, it is understood that the drying process was also important in reducing pathogen viability.

Al-Nabulsi's data: The factors EDTA and SB level used in the experimental design were proportionately linear. So, both the factors were confounding in the model. Therefore, 0 degrees of freedom was reported for EDTA and the parameter estimate for SB was found to be biased. The statistical model for Al-Nabulsi's data was

$$Y = 5.1703 - 0.0022X1 + 0.1369X2 - 0.08261X4 - 0.0000534X1X2X3 \\ -0.01044X1X4 - 0.000051X2X3X4 + 0.0000153X1X2X3X4$$

(Eq.5)

Where Y = 10 based logarithm of the *E.coli* O157:H7 numbers

X1 – LF in mg / g sausage batter;

X2 – SB in mM;

X3 – EDTA in ppm; and

X4 – Day of processing.

The correlation factor (R^2) for this model was 0.736.

The statistical model curves for the combined training and test sets and the trend of observed values for treatments 2, 3, and 6 (Fig 3) and for treatments 1, 4, and 5 (Fig 4) are separately shown in figures.

The SAS program result (Appendix VII) of this model indicates that the individual factor 'Day of processing'(X4) and the interactions of factors LF and Day of processing (X1*X4); LF, SB and EDTA (X1*X2*X3); SB, EDTA and Day (X2*X3*X4); LF, SB, EDTA and Day (X1*X2*X3*X4) significantly contributed to the reduction of *E.coli* O157:H7 with the probability of t-value less than 0.05. But, the LF(X1) alone did not have significance in the reduction. For this, the probability of t value was only less than 0.936. The common reason was the presence of the factor X4 (Day) in all the interactions. So, the drying process that is taking place from 3 to 28 days alone and with LF has made the significant reduction in all the treatments.

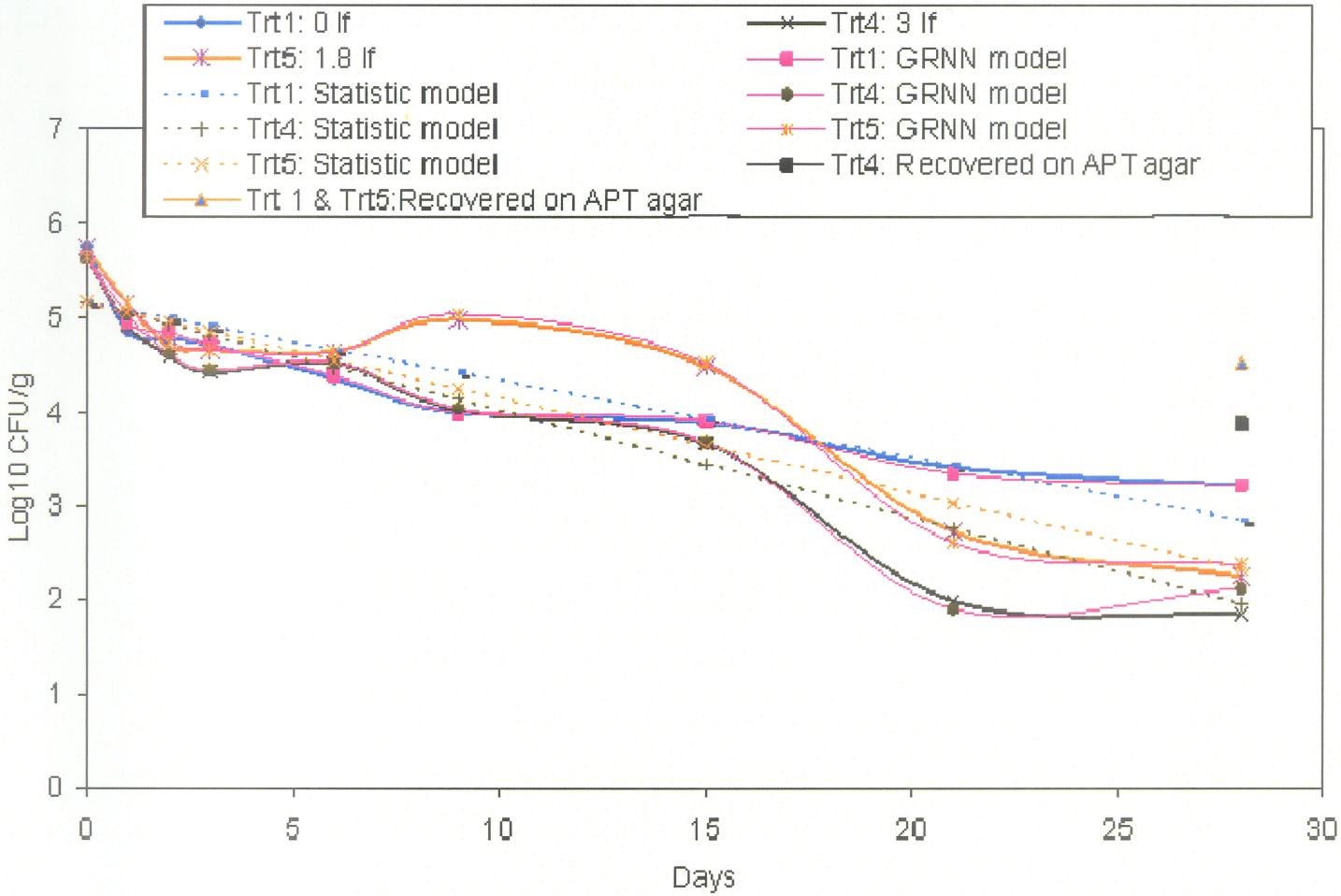

Fig. 2: Observation, GRNN and Statistical trends of E.coliO157:H7 survival curves against Lactoferrin treatments 1,4, and 5

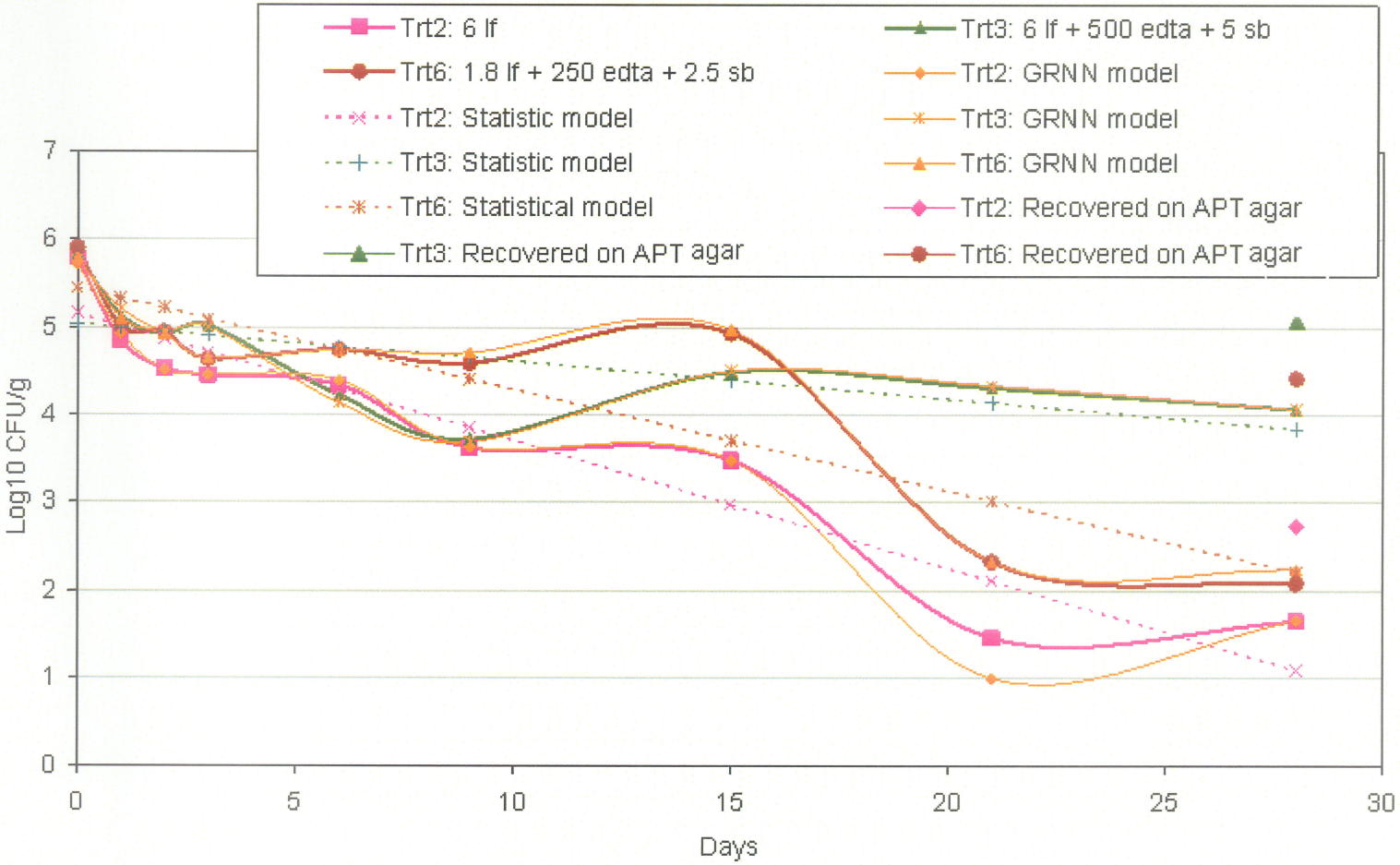

Fig. 3: Observation, GRNN and Statistic trends of E.coliO157:H7 survival curves against Lactoferrin treatments 2,3, and 6

The result also indicates that the factor SB (X2) and / or EDTA(X3) contributed to an increase in number of cells during processing. Both the factors were confounding in the model. But, the real contribution of each of the factors in either increasing or reducing the cells could not be understood. This was due to the experimental design in which the levels of both the factors were present in a linear proportion in the treatments concerned. Al Nabulsi and Holley (2006) found that EDTA enhanced the antimicrobial activity of LF in reduction of *E.coli* O157:H7 strains 0627 and 0628, the same strains used for this experiment. They also showed that when SB was added to LF and EDTA a 4 log reduction was obtained that may have been due to the enhancement of LF stability by bicarbonate. Also, it was evident that treatment 6 (1.8 LF + 250 EDTA + 2.5SB) performed better than treatment 5 (1.8 lf). But treatment 2 (6 LF) performed far better than treatment 3 (6 LF + 500 EDTA + 5 SB) (Fig. 4). This might have been due to the use of SB in treatment 3, which maintained high pH values in the meat because of its buffering capacity. So, there was an ambiguity present which prevented to comment whether EDTA or SB or both contributed to an increase in number of cells. This could be solved only if we have the experimental data were available without linear proportionate levels of EDTA and SB.

The treatment 2 at a higher level of LF (unencapsulated LF) has performed better than all other treatments with the largest reduction of 4.2 log. But, the results from samples at the 28^{th} day (only) plated on All Purpose Tween (APT) broth overlaid with ct-SMAC showed that part of this reduction occurred as a result of

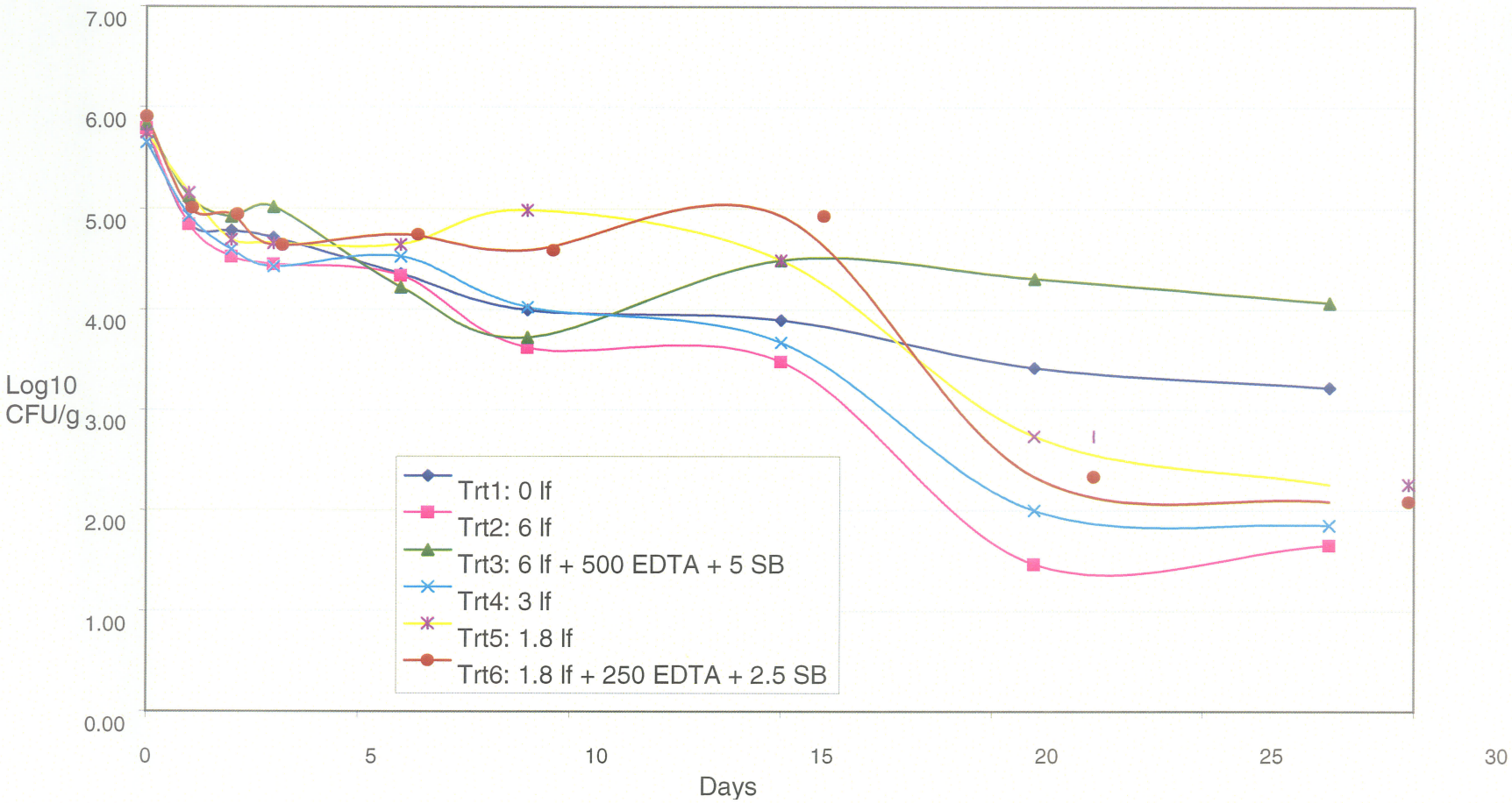

Fig. 4: Observation trends of *E.coli* O157:H7 survival curves for Lactoferrin treatments.

EDTA - ethvlene diamine tetraacetic acid. SB-sodium bicarbonate

cell injury. The recovered cell increase is shown in Fig. 2 and Fig.3. So, the lethality of treatment 2 was only 3 log (Al-Nabulsi and Holley, 2006).

4.2 Comparison of Models' Performance using Statistical Indices

Chacon's data: The results of GRNN and statistical model predictions in terms of statistical indices are shown in Fig.5 and Table 6 for the *E.coli* O157:H7 count. They show the bias and the residual plots for the GRNN predictions for training and test data sets as well as the statistical predictions for whole data sets. The GRNN predictions for the training data set are very close to having no bias from the observed values, whereas the same predictions for the test data set are slightly biased from observed values. Statistical model predictions for the whole data set are comparatively far biased from the observed values. The R-square value 0.964 shows that the statistical model is well developed. Even then, the GRNN prediction for the training data set is far superior to the statistical model predictions, while the prediction for the test set is slightly superior to the statistical model predictions. The test set prediction was similar or slightly worse than statistical prediction obtained by Jeyamkondan et al. (2001). The number of negative residual points was greater than the number which was positive (Fig. 5). This means that the predictions by both the GRNN and statistical models are under-predictions, which was again confirmed by the MRPR values, which were positive for both the GRNN and statistical models. Under-estimation of bacterial

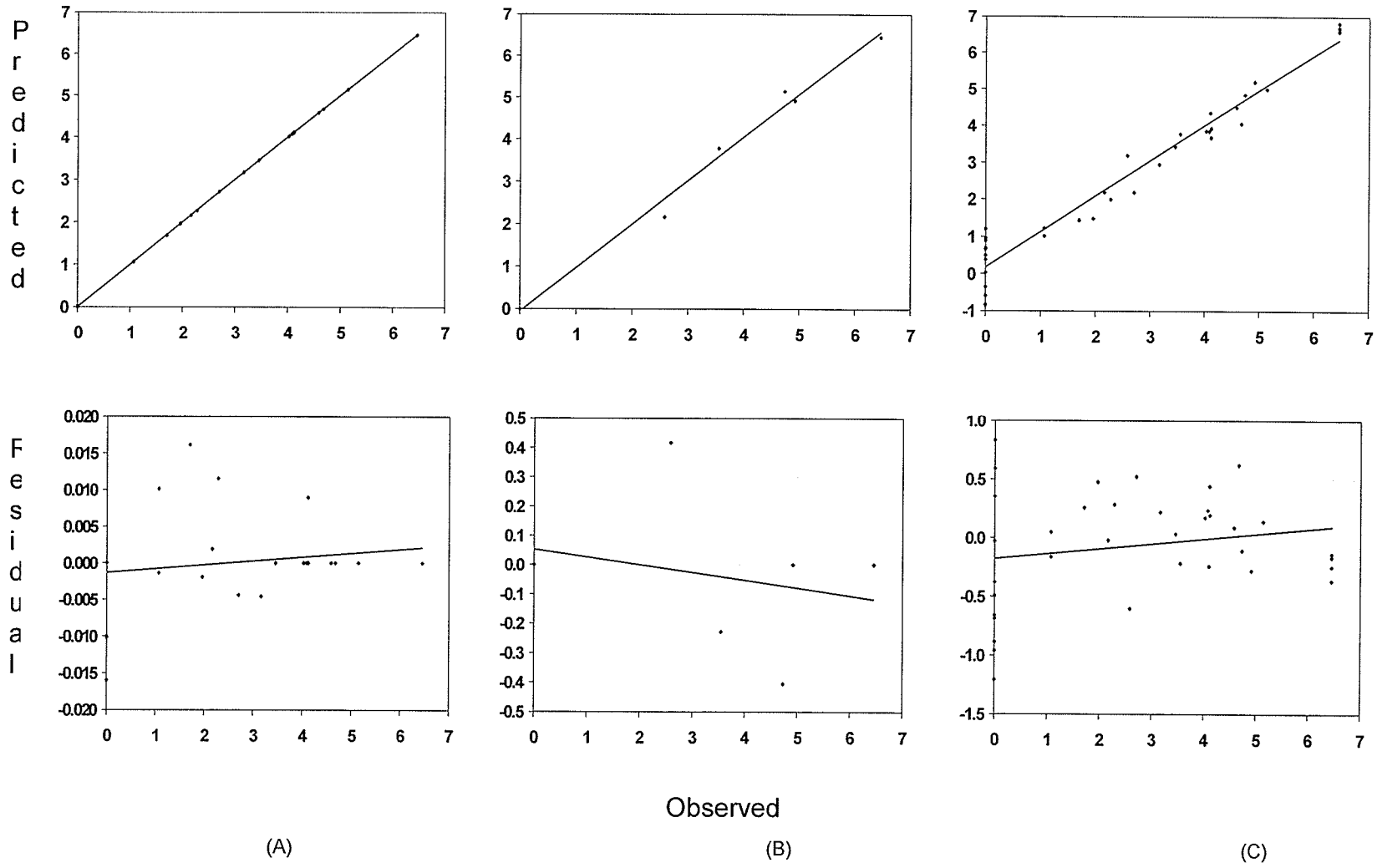

Fig.5: Comparison of predictions of neural network model with statistical model for *E.coli O157:H7* population (log_{10} CFU/g) for the observed experimental observations from Chacon (2006). (A) GRNN prediction for training data set, (B) GRNN prediction for test data set, and (C) Statistical model prediction for full data set.

Table 6: Comparison of GRNN and Statistical models of Data Set 1 (Chacon) and Data Set 2 (Al-Nabulsi)

Statistical indices	Mathematical Expressions	Model	Data set	For AIT (Data 1) treatments	For LF +EDTA+SB (Data 2) treatments
Mean Relative Percentage Residual (MRPR)	$1/N * \sum(O-P)*100 / O$	GRNN	Training	0.075*	-0.033**
			Test	0.168*	2.14*
		Statistical	Full	1.500*	0.281*
Bias Factor	$A.\log10 \sum\log10(P/O)/N$	GRNN	Training	0.999*	0.995*
			Test	0.995*	1.001**
		Statistical	Full	0.981*	1.002**
Mean Absolute Percentage Residual (MARR)	$1/N * \sum\lvert O-P\rvert *100 / O$	GRNN	Training	0.112	8.079
			Test	4.456	10.942
		Statistical	Full	5.600	12.181
Accuracy Factor	$A.\log10 \sum\lvert \log10(P/O)\rvert /N$	GRNN	Training	1.001	1.070
			Test	1.047	1.107
		Statistical	Full	1.060	1.120
Root Mean Square Residual	$\sqrt{\sum(O-P)^2/N}$	GRNN	Training	0.006	0.447
			Test	0.236	0.391
		Statistical	Full	0.470	0.607

* indicates that the model has under-predicted;

** indicates that the model has over-predicted *E.coli* survival.

numbers by both methods indicates that the models are fail-dangerous. This suggests that care is needed while applying these models.

From the MRPR values in Table 6, it was found that the GRNN has under-estimated the *E.coli O157:H7* numbers by 0.075 % for the training set, and 0.168 % for the test set. All other indicators, including the bias factor, predicted and residual graphical plots also confirm the under-prediction. The same was the case for the statistical model. MRPR, MARR and RMSR values clearly differentiate the predicting power of the models, but since the bias and accuracy factor values are very close for all the models, they could not be used to clearly evaluate the predictability of these models. This is because the residuals are normalized in the MRPR, MARR and RMSR, whereas the predicted values are normalized in the bias and accuracy factors.

Al-Nabulsi's data: Fig. 6 and Table 6 show the comparison of the prediction performance of GRNN and Statistical models for the Al-Nabulsi data, using statistical indices. The predictions of the GRNN model for the training set was modestly better than that of the statistical model while considering all the indices. The predictions of the GRNN model for the test set was worse than that of the statistical model when MRPR was considered, whereas it was slightly better when other indices were considered.

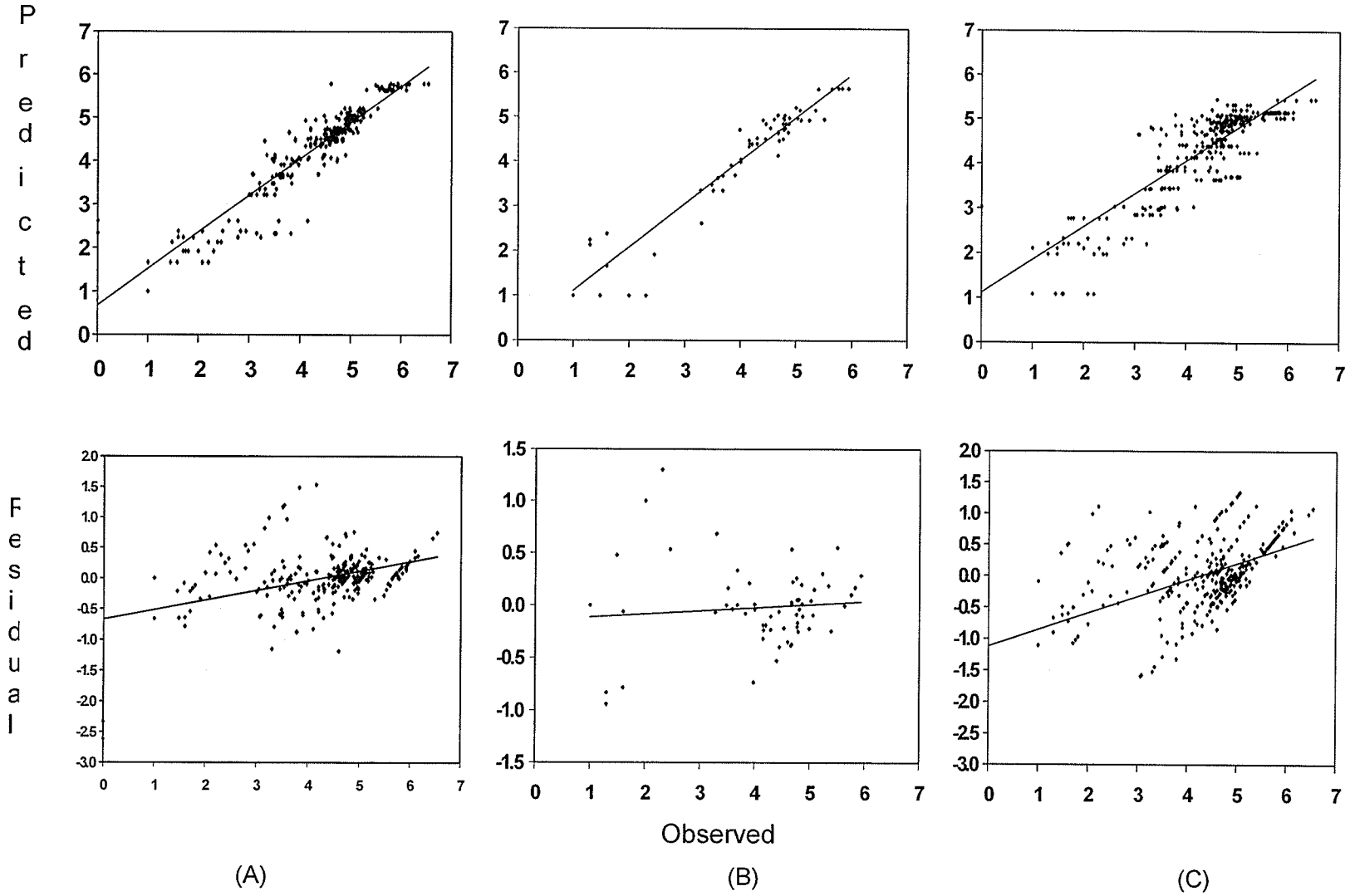

Fig.6: Comparison of predictions of neural network model with statistical model for *E.coli O157:H7* population (log_{10} CFU/g) for the observed experimental observations from Al-Nabulsi (2006). (A) GRNN prediction for training data set, (B) GRNN prediction for test data set, and (C) Statistical model prediction for full data set.

GRNN prediction for training vs. test sets: The predictions of the GRNN model for test set was also slightly better than that of the GRNN model for training set, and when the bias factor and RMSR were considered it was worse for other indices. The test set was randomly chosen by GRNN. It may vary for another run and change the validation of model and there by the predictions. So, GRNN predictions for the training set was always better than that of the test set while considering MRPR, MARR (the total deviation) and the accuracy factor.

Also, the indices MRPR and Bias factor are misleading when assessing whether the models have over- or under-predicted. The GRNN test and statistical models have under predicted and GRNN training model has over predicted based on MRPR. It was opposite when considering the bias factor. The index 'graphical plot' has helped clear this ambiguity. The graphs in Fig. 6 show that all the Predicted vs. Observed plots (have more points in the upper side of the diagonal) and Residual vs. observed plots (have more negative residual points) confirm the over-prediction performance of all three models. This is 'fail-safe'.

4.3 ANN vs. Statistical Model Predictions for Unseen Data

The ANN model predictions were best for the training data set and reasonably good for the test data set. Similar results were obtained by Hajmeer et al. (1997) and Jeyamkondan et al. (2001). Unless the number of data points is much larger than the number of variables of the polynomial equations, there will be little

assurance that the statistical model predictions for unseen data will be accurate (Specht, 1991). The response surfaces produced by statistical models developed using partial data were quite different and quite erratic. Thus, the predictions for unseen data by this type of statistical model may not be accurate (Bratchell et al., 1989, 1990).

From Table 4 and Fig. 1 it can be seen that the GRNN has predicted the AIT level should be 580 ppm in order to reduce the *E.coli* O157:H7 count 1.45 $\log_{10}$CFU/g at day 28, whereas the statistical model predicted a similar result at 500 ppm. However, at 500 ppm, the observed value was 1.70 $\log_{10}$CFU/g. In order to reduce the *E.coli* O157:H7 a further 0.25 $\log_{10}$CFU/g, the AIT level should be greater than 500 ppm as noted above. Therefore, the GRNN produced an acceptable result, whereas the statistical model was unsuccessful. To achieve a count of 1.45 $\log_{10}$CFU/g (or a 5 $\log_{10}$CFU/g reduction) at day 35, the GRNN generated an inaccurate result. It predicted a number of 1.06 $\log_{10}$CFU/g for all concentrations of AIT from 500 ppm to 400 ppm at 10 ppm intervals. The GRNN prediction of 1.45 $\log_{10}$CFU/g for AIT levels from 400 to 500 ppm appears incorrect. We suspect this has occurred because the GRNN has seen the data set which has only one AIT level below 500 ppm (0 ppm only). There was insufficient data available to train the GRNN between 0 and 500 ppm AIT. However, the statistical model predicted that the required 1.45 $\log_{10}$CFU at day 35 can be achieved at 460 ppm AIT, which is a reasonable estimate. Results indicate that if sufficient data for training and prediction are available, prediction by the GRNN model will be very accurate. The statistical model may be used when the data set is not as complete as desirable

and still generate a useful model. However, under these conditions, the predictions may or may not be accurate (Bratchell et al., 1989). Both the models have under-predicted the numbers of bacterial survivors. This problem could be easily addressed by modifying the criterion for stopping the training period. An introduction of an offset of 5% over-prediction in the program as a safety margin instead of minimizing the residual toward zero is possible with ANN. As the performance of ANN is based on training, the greater accuracy may be achieved by altering training and selecting suitable criteria (Jeyamkondan et al., 2001).

5.0 CONCLUSIONS

There is no significant difference in model predictions since ANN is randomly choosing the test set for every run. The best split ratio between training and test set for ANN model development is 80:20.

GRNN may perform better for future predictions for production set when the mean values were used for model development.

GRNN predictions for the training data set were always best and those for test set were slightly best or similar to those of the statistical model predictions.

The performance of the GRNN for future predictions (production data sets) was accurate when sufficient data were used for training. The GRNN model may not perform well for future predictions when incomplete data are available for training. The statistical model for future predictions was not as good for production data sets as the ANN model when sufficient data were used for developing the model.

It is uncertain whether the statistical model would be more accurate than the ANN model when only partial data sets are available. ANN-based models are proposed as the best option for use by the meat industry for predicting results from new sets of parameters which fall into an interpolation region of the data used for training because of better model performance, accuracy and speed.

The meat industry can use commercial neural network software as it is easy to handle and does not require expert knowledge of the basic operation of neural networks to develop models. With basic computer skills, an operator can predict bacterial survival or any other output variable from a new set of input variables during processing. ANN-based modeling could be a valuable tool for predictive microbiology to ensure food safety in the meat industry and for preventing further foodborne illness outbreaks from meat products. Unfortunately, profiles of experimental data available for use in this study for bacterial response to 0 to 500 ppm AIT were not available and this did not allow complete evaluation of a comparison of ANN and statistical model performance.

There is an opportunity for further food safety engineering research to test the feasibility of ANN use in the meat industry for online processing during manufacture of dry fermented sausages.

6.0 REFERENCES

Anon. 1995. Food borne outbreak of *Escherichia coli* O157-H7 infections from hamburgers - western United-States, *Morbidity and Mortality Weekly report* 44, 157-160.

Al-Nabulsi, A.A. 2006. Use of the milk protein lactoferrin as a natural antimicrobial in meat products. PhD thesis. Department of Food Science, Winnipeg, MB: University of Manitoba.

Al-Nabulsi, A.A., Holley, R.A. 2006. Activity of bovine lactoferrin against *Escherichia coli* O157:H7 and meat starters in broth and during dry sausage manufacture following its microencapsulation. Submitted to *International Journal of Food Microbiology.*

Baranyi, J., Ross, T., McMeekin, T.A., Roberts, T.A. 1996. Effects of parameterization on the performance of empirical models used in 'predictive microbiology'. *Food Microbiology* 13, 83-91.

Barbosa,W.B., Cabedo, L., Wederquist, H.J., Sofos, J.N., Schmidt, G.R. 1994. Growth variation among species and strains of *Listeria* in culture broth. *Journal of Food Protection* 57, 765-769.

Belehradek, J. (1930) as cited in McKellar, R.C., Lu, X. (eds.). 2004. Temperature co-efficients in biology. *Biological Reviews of the Cambridge Philosophical Society* 5, 30.

Bishop, C. (1995) as cited in Statsoft (2006). Neural Networks for Pattern Recognition. Oxford: University Press.

Bratchell, N., Gibson, A.M., Truman, M., Kelly, T.M., Roberts, T.A. 1989. Predicting microbial growth: the consequence of quantity of data. *International Journal of Food Microbiology* 8, 47–58.

Bratchell, N., McClure, P.J., Kelly, T.M., Roberts, T.A., 1990. Predicting microbial growth: graphical methods for comparing models. *International Journal of Food Microbiology.* 11, 279–288.

Brocklehurst, T. M. 2004. Challenge of food and the environment. In: Modeling Microbial Responses in Food. McKellar, R.C., Lu, X. (eds.), 1-20, New York, NY: CRC Press.

Buchanan, R.L. 1993. Developing and distributing user-friendly application software. *Journal of Industrial Microbiology* 12, 251-255.

Cerf, O., Davey, K.R., Sadoudi, A.K. 1996. Thermal inactivation of bacteria – a new predictive model for the combined effect of three environmental factors: temperature, pH and water activity. *Food Research International* 29, 219-226.

Chacon, P., 2006. Application of allyl isothiocyanate to control *Escherichia coli* O157-H7 in chopped beef and dry fermented sausages. M.Sc. thesis. Department of Food Science, Winnipeg, MB: University of Manitoba.

Cheroutre-Vialette, M., Lebert, A. 2000. Modeling the growth of *Listeria monocytogenes* in dynamic conditions. *International Journal of Food Microbiol*ogy 55, 201-207.

Cheroutre-Vialette, M., Lebert, A. 2002. Application of recurrent neural networks to predict bacterial growth in dynamic conditions, *International Journal of Food Microbiology* 73, 107-118.

Esnoz, A., Periago, P.M., Conesa, R., Palop, A. 2006. Application of artificial neural networks to describe the combined effect of pH and NaCl on the heat resistance of *Bacillus stearothermophilus*. *International Journal of Food Microbiology* 106(2): 153-158.

FAO (Food and Agriculture Organisation). 1996. Rome declaration of World food security and Action plan. http://www.fao.org/documents/show_cdr.asp?url_file=/docrep/003/w3613e/w3613e/w3613e00.html (2005/07/14).

Gibson, A.M., Bratchell, N., and Roberts, T.A., 1987. The effect of sodium chloride and temperature on the rate and extent of growth of *Clostridium botulinum* type A in pasteurized pork slurry. *Journal of Applied Bacteriology* 62, 479-490.

Garcia-Gimeno, R.M., Hervas-Martinez, C., Barco-Alcaia, E., Zurera-Cosano, G., Sanz-Tapia, E. 2003. An artificial neural network approach to *Escherichia coli O157:H7* growth estimation. *Journal of Food Science* 68, 639-645.

Geeraerd, A.H., Herremans, C.H., Cenens, C., Van Impe, J.F. 1998a. Application of artificial neural networks as a non-linear modular modeling technique to describe bacterial growth in chilled food products. *International Journal of Food Microbiology* 44, 49–68.

Geeraerd, A.H., Herremans, C.H., Ludikhuyze, L.R., Hendrickx, M.E., Van Impe, J.F. 1998b. Modeling the kinetics of isobaric-isothermal inactivation of *Bacillus subtilis* alpha-amylase with artificial neural networks. *Journal of Food Engineering* 36, 263-379.

Hajmeer, M.N., Basheer, I.A., Najjar, Y.M. 1997. Computational neural networks for predictive microbiology II. Application to microbial growth. *International Journal of Food Microbiol*ogy 34, 51–66.

Hajmeer,M.N., Basheer, I.A. 2003. A hybrid Bayesian-neural network approach for probabilistic modeling of bacterial growth/no-growth interface. *International Journal of Food Microbiology* 82, 233-243.

Health Canada. 2000. *Escherichia coli O157:H7* infections associated with ground beef and their control in Canada. *Canadian Communicable Disease Report,* July, 26, 13.

Jay, J.M., Loessner, M.J, Golden, D.A, 2005. The HACCP and FSO systems for food safety In: Modern Food Microbiology-7th Ed., 497-515, Springer, USA.

Jeyamkondan, S., Jayas D.S., Holley R.A. 2001. Microbial growth modeling with artificial neural networks. *International Journal of Food Microbiology.* 64, 343-354.

Lebert, I., Robles-Olvera, V., Lebert, A. 2000. Application of polynomial models to predict growth of mixed cultures of *Pseudomonos* spp. and *Listeria* in meat. *International Journal of Food Microbiology* 61, 27-39.

Lou, W., Nakai, S. 2001. Application of artificial neural networks for predicting the thermal inactivation of bacteria: a combined effect of temperature, pH and water activity. *Food Research International* 34, 573-579.

McClure, P.J., Blackburn, C. de W., Cole, M.B., Curtis, P.S., Jones, J.E., Legan, J.D., Ogden, I.D., Peck, M.W., 1994a. Modeling the growth, survival and death of micro organisms in foods: the UK food micromodel approach. *International Journal of Food Microbiology* 23, 265-275.

McClure, P.J., Cole, M.B., Davies, K.W. 1994b. An example of the stages in the development of a predictive mathematical model for microbial growth: the effects of NaCl, pH and temperature on the growth of *Aeromonas hydrophila*. *International Journal of Food Microbiology* 23 (3-4):359-375.

McKellar, R.C., Lu, X., Delaquis, P.J. 2002. A probability model describing the interface between survival and death of *E. coli* O157:H7 in a mayonnaise model system. *Food Microbiology* 19, 235–247.

McKellar, R.C., Lu, X. (eds.). 2004. Modeling Microbial Responses in Food, New York, NY: CRC Press.

McMeekin, T.A., Olley, J., Ratkowsky, D.A., Ross, T. 1993. Predictive microbiology: Theory and application. 340pp.Taunton, UK: Research Studies Press Ltd.

McMeekin, T. 2004. An essay on the unrealized potential of predictive microbiology In: *Modeling Microbial Responses in Food.* McKellar, R.C., Lu,X. (eds.), 321-335, New York, NY :CRC Press.

Mittal, G.S., Zhang, J. 2002. Prediction of food thermal process evaluation parameters using neural networks. *International Journal of Food Microbiology* 79, 153-159.

Najjar, Y.M., Basheer, I.A., Hajmeer, M.N. 1997. Computational neural networks for predictive microbiology I. Methodology. *International Journal of Food Microbiol*ogy 34, 27–49.

NeuroShell®2. 1993. Online tutorial, NeuroShell®2, Release 4.0. Ward Systems Group, Inc., 252 pp. Frederick, MD.

Olley, J., Ratkowsky, D.A. 1973. The role of temperature function integration in monitoring fish spoilage. *Food Technology N.Z.* 8:13-17.

Patterson, D. (1996) as cited in Statsoft. 2006. Artificial Neural Networks. Singapore: Prentice Hall.

Pond, T.J., Wood, D.S., Mumin, I.M., Barbut, S., Griffiths, M.W. 2001. Modeling the survival of *E. coli* O157:H7 in uncooked, semidry, fermented sausage. *Journal of Food Protection* 64, 759-766.

Ratkowsky, D.A.1990. *Handbook of Nonlinear Regression Models*. 241 pp. New York, NY: Marcel Dekker Inc.

Ratkowsky, D.A.1993. Principles of nonlinear regression modeling. *Journal of Industrial Microbiology* 12, 195-199.

Ratkowsky, D.A. 2004. Model fitting and Uncertainity In: Modeling microbial responses in Food. McKellar, R.C., Lu, X. (eds.) 151-196, New York, NY: CRC Press.

Ratkowsky, D.A., Olley, J., McMeekin, T.A., Ball, A. 1982. Relationship between temperature and growth rate of bacterial cultures. *Journal of Bacteriology* 149, 1-5.

Ratkowsky, D.A., Ross, T., McMeeking, T.A., Olley, J., 1991. Comparison of arrhenius-type and belehradek-type models for prediction of bacterial growth in foods. *Journal of Applied Bacteriology* 71, 452–459.

Ray, B. 2004. Predictive modeling of microbial growth in food In *Fundamental Food Microbiology*. 3rd ed. New York, NY: CRC press. 541-544.

Reed, C., 1995. Letter to plant managers, April 28. United States Department of Agriculture, Food Safety and Inspection Service, Washington, D.C.

Ross,T. 1999. *Predictive Microbiology for the Meat Industry*, Meat and Livestock Australia, Sydney, 196 pp.

Ross, T., Dalgaard, P., 2004. Secondary models In: *Modeling Microbial Responses in Food*, McKellar, R.C., Lu, X. (eds.), 64-135. New York, NY: CRC Press.

Ross, T. Baranyi, J., McMeekin, T.A. 2000. Predictive microbiology and food safety. In: Encyclopedia of Food Microbiology, Robinson, R., Batt, C.A., Patel, P. (eds.), 1699 -1710. London, UK: Academic Press.

Scott, W.J. (1936) as cited in McKellar, R.C., Lu, X. (eds.). 2004. The growth of micro organisms on ox muscle. I. The influence of water content of substrate on rate of growth at -1°C. *Journal of Council Science and Industrial Research (Australia)* 9,177.

Scott, W.J. (1937) as cited in McKellar, R.C., Lu, X. (eds.). 2004. The growth of micro organisms on ox muscle. II. The influence of temperature. *Journal of Council Science and Industrial Research (Australia)* 10, 338.

Specht, D.F. 1991. A general regression neural network. *IEEE Transactions on Neural Networks* 2, 568–576.

Statsoft. 2006. Generalized Regression Neural Networks. 28pp. http://www.statsoft.com/ textbook/ stneunet.html. (2006/02/23).

Tomicka, A., Chen, J., Barbut, S., Griffiths, M.W. 1997. Survival of bioluminescent *Escherichia coli* O157:H7 in a model system representing fermented sausage production. *Journal of Food Protection* 60(12): 1487-1492.

Tu, J.V. 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting microbial outcomes. *Journal of Clinical Epidemiology* 11, 1225-1231.

WHO. 1997. *E.coli* infections appear to be increasing-Experts say. Press Release, WHO, 41, 21-May.

Whiting, R.C., Buchanan, R.L. 1993. Letter to the editor: a classification of models in predictive microbiology. A reply to K.R. Davey. *Food Microbiology* 10, 175-177.

Wijtzes, T., Rombouts, F.M., Kant-Muermans, M.L., van't Riet, K., Zwietering, M. 2001. Development and validation of a combined temperature, water activity, pH model for bacterial growth rate of *Lactobacillus curvatus. International Journal of Food Microbiology* 63, 57-64.

Yu, C., Davidson, V.J., Yang, S.X. 2006. A neural network approach to predict survival/death and growth/no-growth interfaces for *Escherichia coli* O157:H7. *Food Microbiology* 23 (6): 552-560.

Zhao, L., Chen, Y., Schaffner, D.W. 2001. Comparison of logistic regression and linear regression on modeling percentage data. *Applied Environmental Microbiol*ogy 67, 2129-2135.

APPENDIX I

DATA Set 1 (Chacon, 2006)

pH values								
Treatment	Day 0	Day 2	Day 3	Day 6	Day 9	Day 16	Day 21	Day 28
T1	5.56	4.90	4.88	4.72	4.76	4.77	4.99	4.88
T1	5.57	4.85	4.88	4.69	4.79	4.80	4.89	4.94
T1	5.57	4.88	4.84	4.70	4.80	4.79	4.96	4.96
	5.57	**4.88**	**4.87**	**4.70**	**4.78**	**4.79**	**4.95**	**4.93**
T2	5.56	4.84	4.81	4.81	4.73	4.77	4.94	4.90
T2	5.57	4.90	4.80	4.78	4.80	4.82	4.90	4.96
T2	5.57	4.89	4.83	4.83	4.78	4.82	4.95	4.93
	5.57	**4.88**	**4.81**	**4.81**	**4.77**	**4.80**	**4.93**	**4.93**
T3	5.56	4.88	4.79	4.70	4.81	4.82	5.02	5.00
T3	5.57	4.88	4.80	4.72	4.82	4.76	5.02	5.02
T3	5.57	4.84	4.76	4.72	4.81	4.83	4.96	5.00
	5.57	**4.87**	**4.78**	**4.71**	**4.81**	**4.80**	**5.00**	**5.01**
T4	5.56	4.90	4.86	4.80	4.89	4.87	4.98	4.99
T4	5.57	4.91	4.87	4.80	4.89	4.88	4.98	5.02
T4	5.57	4.87	4.88	4.80	4.80	4.85	5.03	5.03
	5.57	**4.89**	**4.87**	**4.80**	**4.86**	**4.87**	**5.00**	**5.01**

Bolded numbers are mean values

Water Activity									
Treatment	Day 0	Day 3	Day 6	Day 9	Day 16	Day 28	Day 35	Day 40	Day 45
T1	0.948	0.941	0.936	0.920	0.908	0.880	0.878	0.874	0.870
T2	0.943	0.938	0.939	0.918	0.910	0.890	0.877	0.873	0.870
T3	0.946	0.936	0.925	0.900	0.902	0.886	0.873	0.870	0.866
T4	0.941	0.940	0.923	0.911	0.905	0.874	0.870	0.868	0.864

Source: Chacon (2006)

T1 – Control treatment

T2 - 500 ppm AIT treatment

T3 - 750 ppm AIT treatment

T4 - 1000 ppm AIT treatment

Appendix I – DATA Set 1 continued

	E.coli O157:H7 population									
Treatment	Day 0	Day 3	Day 6	Day 9	Day 16	Day 21	Day 28	Day 35	Day 40	Day 45
T1	6.58	5.19	4.99	4.87	4.56	4.65	4.68	4.13	4.05	4.01
T1	6.30	4.75	5.19	4.86	4.79	4.78	4.37	4.07	4.08	4.08
T1	6.48	4.81	5.24	4.48	4.66	4.55	4.70	4.15	4.11	3.98
	6.45	**4.92**	**5.14**	**4.73**	**4.67**	**4.66**	**4.58**	**4.12**	**4.08**	**4.02**
T2	6.58	4.56	3.90	2.38	2.08	1.60	1.60	1.60	0.00	0.00
T2	6.30	3.92	4.36	2.64	2.38	1.90	1.60	1.60	0.00	0.00
T2	6.48	3.82	4.07	2.72	2.38	0.00	1.90	0.00	0.00	0.00
	6.45	**4.10**	**4.11**	**2.58**	**2.28**	**1.17**	**1.70**	**1.07**	**0.00**	**0.00**
T3	6.58	3.35	3.15	2.08	1.60	0.00	0.00	0.00	0.00	0.00
T3	6.30	3.58	3.45	2.20	1.60	0.00	0.00	0.00	0.00	0.00
T3	6.48	3.72	2.90	2.20	0.00	0.00	0.00	0.00	0.00	0.00
	6.45	**3.55**	**3.17**	**2.16**	**1.07**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
T4	6.58	3.30	2.51	1.60	0.00	0.00	0.00	0.00	0.00	0.00
T4	6.30	3.59	2.75	2.08	0.00	0.00	0.00	0.00	0.00	0.00
T4	6.48	3.48	2.86	2.20	0.00	0.00	0.00	0.00	0.00	0.00
	6.45	**3.45**	**2.70**	**1.96**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

Bolded numbers are mean values

Source: Chacon (2006)

T1 – Control treatment

T2 - 500 ppm AIT treatment

T3 - 750 ppm AIT treatment

T4 - 1000 ppm AIT treatment

APPENDIX II

Is there significant difference in model predictions because the test set is randomly chosen by ANN?

GRNN models were developed from the data sets (Chacon, 2006) of 5 different trials of each split ratio 80:20, 70:30, and 90:10 between training and test sets. T-tests were done using SAS t-test procedure for determining if the predicted values significantly differed from the observed values.

Results of GRNN predictions and of t-tests for five different trials of each split ratio 80:20, 70:30' and 60:40

Observed value	Predicted Values														
	80:20 [1]	80:20 [2]	80:20 [3]	80:20 [4]	80:20 [5]	70:30 [1]	70:30 [2]	70:30 [3]	70:30 [4]	70:30 [5]	60:40 [1]	60:40 [2]	60:40 [3]	60:40 [4]	60:40 [5]
6.45	6.45	4.94	6.45	6.45	5.71	6.45	6.45	6.44	6.45	5.66	6.45	4.90	4.95	6.45	6.43
4.92	5.14	4.95	4.92	4.92	5.23	5.14	4.92	4.94	4.92	5.25	5.14	4.91	4.97	4.92	4.94
5.14	5.14	4.93	5.14	5.14	5.06	5.14	5.14	5.10	5.14	5.06	5.14	4.90	4.96	5.14	4.82
4.73	5.14	4.88	4.73	5.14	4.96	5.14	4.73	4.76	5.14	4.95	5.14	4.87	4.90	5.14	4.73
4.67	4.67	4.73	4.73	4.67	4.74	4.67	4.67	4.62	4.67	4.73	5.14	4.77	4.78	4.67	4.67
4.58	4.58	4.44	4.58	4.58	4.44	4.58	4.58	4.58	4.58	4.44	4.58	4.45	4.53	4.58	4.58
4.12	4.12	4.19	4.12	4.12	4.20	4.12	4.12	4.12	4.12	4.30	4.12	4.32	4.18	4.12	4.12
4.08	4.08	4.09	4.08	4.07	4.12	4.12	4.08	4.07	4.12	4.23	4.12	4.26	4.12	4.12	4.12
4.02	4.02	4.05	4.02	4.02	4.09	4.12	4.02	4.02	4.12	4.18	4.12	4.21	4.12	4.12	3.38
6.45	6.45	6.21	6.45	4.10	5.98	4.10	6.45	6.44	4.10	5.60	4.11	6.21	6.36	4.10	6.45
4.10	4.10	4.42	4.10	4.10	4.51	4.10	3.55	4.11	4.10	4.40	4.11	3.90	4.90	4.10	5.13
4.11	4.11	4.07	4.11	4.11	3.64	3.94	4.10	4.01	4.11	3.88	4.11	3.65	3.16	3.35	3.10
2.58	2.58	3.60	2.58	2.58	3.05	2.78	3.35	2.68	2.58	3.39	2.58	3.35	2.62	2.58	2.33
2.28	2.28	2.50	2.28	2.28	2.49	2.28	2.26	2.28	2.28	2.52	2.28	2.62	2.32	2.28	1.07
1.70	1.70	1.47	1.70	1.70	1.43	1.70	1.70	1.68	1.70	1.26	1.70	1.38	1.61	1.07	1.69

Appendix II - Table Continued

	Predicted Values														
Observed value	80:20 [1]	80:20 [2]	80:20 [3]	80:20 [4]	80:20 [5]	70:30 [1]	70:30 [2]	70:30 [3]	70:30 [4]	70:30 [5]	60:40 [1]	60:40 [2]	60:40 [3]	60:40 [4]	60:40 [5]
1.07	1.07	0.80	1.07	1.07	0.80	1.06	1.05	1.07	1.07	0.77	1.07	0.93	1.12	1.07	1.05
0.00	0.00	0.35	0.00	1.07	0.35	0.01	0.00	0.05	1.07	0.44	1.07	0.68	0.53	1.07	0.01
0.00	0.00	0.11	0.00	0.96	0.11	0.00	0.00	0.00	0.00	0.22	1.07	0.50	0.02	0.01	0.00
6.45	6.45	5.87	6.45	6.45	4.87	6.45	6.45	6.44	6.45	4.49	6.45	6.02	5.51	6.45	6.43
3.55	3.55	3.56	3.55	3.55	3.65	3.51	3.55	3.54	3.55	3.54	3.55	3.21	3.92	3.55	3.51
3.17	3.17	3.27	3.17	3.17	3.18	3.13	3.17	3.20	2.86	3.18	3.17	2.98	3.50	2.87	3.08
2.16	2.16	2.88	3.17	2.16	2.76	2.25	1.96	3.08	2.16	2.84	2.16	2.77	3.31	2.16	2.32
1.07	1.17	1.38	1.14	1.07	1.59	2.16	1.08	1.15	1.07	1.83	2.16	2.06	1.92	1.07	1.07
0.00	0.00	0.05	0.00	0.00	0.05	0.00	0.85	0.02	0.00	0.15	0.00	0.93	0.02	0.00	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.51	0.00	0.01	0.00	0.13	0.03	0.00	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6.45	6.45	5.86	6.45	6.45	5.14	3.45	6.45	6.43	6.45	4.56	3.45	5.88	5.35	6.45	6.39
3.45	3.45	2.90	3.46	3.45	3.50	3.37	3.55	3.43	3.45	3.30	3.45	2.68	3.79	3.46	2.73
2.70	2.70	2.38	2.70	2.70	2.75	2.70	2.71	2.76	2.70	2.73	2.70	2.48	3.22	2.70	2.59
1.96	1.96	1.99	2.70	1.96	2.17	2.05	1.96	2.71	1.96	2.27	1.96	2.40	2.60	1.96	2.08
0.00	0.00	0.78	0.00	0.00	0.79	0.00	1.07	0.00	0.00	1.22	0.00	1.78	0.05	0.00	1.09
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.10	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
t-value	-1.58	0.13	-1.53	-0.02	0.22	0.86	-1.22	-1.78	0.39	0.14	0.20	-0.66	-0.45	0.86	0.84
Pr > t	0.13	0.90	0.13	0.98	0.82	0.39	0.23	0.08	0.70	0.89	0.84	0.51	0.65	0.39	0.40

The probability is higher than 5 % of t-value (from 0.98 to 0.08) for all the 15 trials of GRNN models. It was concluded that there is **no significant difference** in predicted values of the models developed for different test sets randomly chosen by GRNN for 20 %, 30%, 40% test sets.

APPENDIX III

What is the best split ratio between training and test data sets for GRNN model development?

GLM procedure was used to determine if there is a significant difference among the observed values, and the values predicted by the models developed with 80:20, 70:30, and 60:40 split ratios between training and test sets (Appendix II). The mean and standard deviation for observed values of three replications (Appendix I) and for GRNN predicted values with five trials for each ratio is given for each pattern (combination of parameters) in the following table. All the four values were compared one another and grouping was done according to the SAS program results (given here after the table). The values grouped as 'a' and 'b' in the table are significantly differing one another for the corresponding pattern.

For the pattern12, the GRNN predicted value for 60:40 ratio, is significantly differing with other values. For the pattern 17, the GRNN predicted value for 60:40 ratio is significantly differing with observed value only. But the GRNN predicted value for 80:20 and 70:30 ratios are falling under both the groups 'a' and 'b' which means they are significantly same to both observed value and the predicted value for 60:40.

E.coli population for 36 patterns

Data Patterns	Observed	80:20 predictions	70:30 predictions	60:40predictions
	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD
P1	6.45 ± 0.14 [a]	6.00 ± 0.67 [a]	6.29 ± 0.35 [a]	5.84 ± 0.83 [a]
P2	4.92 ± 0.24 [a]	5.03 ± 0.14 [a]	5.03 ± 0.15 [a]	4.98 ± 0.09 [a]
P3	5.14 ± 0.13 [a]	5.08 ± 0.09 [a]	5.12 ± 0.04 [a]	4.99 ± 0.14 [a]
P4	4.74 ± 0.22 [a]	4.97 ± 0.18 [a]	4.94 ± 0.20 [a]	4.96 ± 0.18 [a]
P5	4.67 ± 0.12 [a]	4.71 ± 0.03 [a]	4.67 ± 0.04 [a]	4.81 ± 0.19 [a]
P6	4.58 ± 0.19 [a]	4.52 ± 0.08 [a]	4.55 ± 0.06 [a]	4.54 ± 0.06 [a]
P7	4.12 ± 0.04 [a]	4.15 ± 0.04 [a]	4.16 ± 0.08 [a]	4.17 ± 0.09 [a]
P8	4.08 ± 0.03 [a]	4.09 ± 0.02 [a]	4.12 ± 0.06 [a]	4.15 ± 0.06 [a]
P9	4.02 ± 0.05 [a]	4.04 ± 0.03 [a]	4.09 ± 0.07 [a]	3.99 ± 0.34 [a]
P10	6.45 ± 0.14 [a]	5.84 ± 0.99 [a]	5.34 ± 1.18 [a]	5.45 ± 1.23 [a]
P11	4.10 ± 0.40 [a]	4.25 ± 0.20 [a]	4.05 ± 0.31 [a]	4.43 ± 0.55 [a]
P12	4.11 ± 0.23 [a]	4.01 ± 0.21 [a]	4.01 ± 0.10 [a]	3.47 ± 0.42 [b]
P13	2.58 ± 0.18 [a]	2.88 ± 0.45 [a]	2.96 ± 0.38 [a]	2.69 ± 0.39 [a]
P14	2.28 ± 0.17 [a]	2.37 ± 0.12 [a]	2.32 ± 0.11 [a]	2.11 ± 0.60 [a]
P15	1.70 ± 0.17 [a]	1.60 ± 0.14 [a]	1.61 ± 0.19 [a]	1.49 ± 0.27 [a]
P16	1.07 ± 0.92 [a]	0.96 ± 0.15 [a]	1.00 ± 0.13 [a]	1.05 ± 0.07 [a]
P17	0.00 ± 0.00 [a]	0.35 ± 0.44 [ab]	0.31 ± 0.46 [ab]	0.67 ± 0.44 [b]
P18	0.00 ± 0.00 [a]	0.24 ± 0.41 [a]	0.04 ± 0.10 [a]	0.32 ± 0.47 [a]
P19	6.45 ± 0.14 [a]	6.02 ± 0.69 [a]	6.06 ± 0.88 [a]	6.17 ± 0.41 [a]
P20	3.55 ± 0.19 [a]	3.57 ± 0.04 [a]	3.54 ± 0.02 [a]	3.55 ± 0.25 [a]
P21	3.17 ± 0.28 [a]	3.19 ± 0.04 [a]	3.11 ± 0.14 [a]	3.12 ± 0.24 [a]
P22	2.16 ± 0.07 [a]	2.63 ± 0.45 [a]	2.46 ± 0.48 [a]	2.54 ± 0.50 [a]
P23	1.07 ± 0.92 [a]	1.27 ± 0.21 [a]	1.46 ± 0.50 [a]	1.66 ± 0.54 [a]
P24	0.00 ± 0.00 [a]	0.02 ± 0.03 [a]	0.20 ± 0.37 [a]	0.19 ± 0.41 [a]
P25	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.11 ± 0.23 [a]	0.03 ± 0.06 [a]
P26	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.01 ± 0.01 [a]
P27	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]
P28	6.45 ± 0.14 [a]	6.07 ± 0.58 [a]	5.47 ± 1.39 [a]	5.50 ± 1.23 [a]
P29	3.46 ± 0.15 [a]	3.35 ± 0.25 [a]	3.42 ± 0.09 [a]	3.22 ± 0.49 [a]
P30	2.71 ± 0.18 [a]	2.65 ± 0.15 [a]	2.72 ± 0.03 [a]	2.74 ± 0.28 [a]
P31	1.96 ± 0.32 [a]	2.16 ± 0.32 [a]	2.19 ± 0.32 [a]	2.20 ± 0.29 [a]
P32	0.00 ± 0.00 [a]	0.31 ± 0.43 [a]	0.46 ± 0.63 [a]	0.58 ± 0.81 [a]
P33	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.01 ± 0.03 [a]	0.02 ± 0.04 [a]
P34	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]
P35	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]
P36	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]	0.00 ± 0.00 [a]

Data Source: Chacon, 2006.

a and *b* are significantly different

It is concluded that either 80:20, 70:30 and 60:40 ratios could be considered as acceptable split ratio between training and test sets for GRNN model development as far as the 'seen' data is concerned.

Comparing the GRNN model predictions for unseen data for the split ratios 80:20, 70:30, and 60:40 between training and test sets.

Aw	AIT	pH	Day	Predicted values		
				80:20	70:30	60:40
0.89	500	4.93	28	1.69	1.68	1.56
0.89	510	4.93	28	1.68	1.67	1.56
0.89	520	4.93	28	1.67	1.66	1.56
0.89	530	4.93	28	1.66	1.64	1.56
0.89	540	4.93	28	1.64	1.62	1.56
0.89	550	4.93	28	1.61	1.59	1.56
0.89	560	4.93	28	1.57	1.54	1.56
0.89	570	4.93	28	1.51	1.49	1.56
0.89	580	4.93	28	1.44	1.41	1.56
0.877	500	4.93	35	1.06	1.07	1.14
0.877	490	4.93	35	1.06	1.07	1.14
0.877	480	4.93	35	1.06	1.07	1.14
0.877	470	4.93	35	1.06	1.07	1.14
0.877	460	4.93	35	1.07	1.07	1.14
0.877	450	4.93	35	1.07	1.07	1.14
0.877	440	4.93	35	1.07	1.07	1.14
0.877	430	4.93	35	1.07	1.07	1.14
0.877	420	4.93	35	1.07	1.07	1.14
0.877	410	4.93	35	1.07	1.07	1.08

The GRNN model developed from 80:20 and 70:30 split data were able to predict the *E.coli* population successfully for 28 day sets. It could give different values for each set. But, the model of 60:40 split data has predicted the same values for all the corresponding sets. Of 10 sets for 35 days, the model of 80:20 split has predicted two different values for 4 and 6 sets. The model of 70:30 split has predicted only one value for all 10 sets. The model of 60:40 split data has predicted one value for first 9 sets and another value for the last set.

It is concluded that the split ratio between training and test sets should be 80:20 for developing a best model for sufficient training and validation and for future predictions for unseen data.

SAS Program results for Appendix III

```
Data one;
      Input Trt Rep EcoliP1;
Datalines;
Trt   Rep   EcoliP1
1     1     6.58
1     2     6.3
1     3     6.48
2     1     6.45
2     2     4.94
2     3     6.45
2     4     6.45
2     5     5.71
3     1     6.45
3     2     6.45
3     3     6.44
3     4     6.45
3     5     5.66
4     1     6.45
4     2     4.9
4     3     4.95
4     4     6.45
4     5     6.43
Proc GLM Data=one;
Class Trt;
Model EcoliP1=Trt/solution;
Lsmeans Trt/stderr;
Estimate 'Trt1 vs Trt2' Trt 1 -1 0 0;
Estimate 'Trt1 vs Trt3' Trt 1 0 -1 0;
Estimate 'Trt1 vs Trt4' Trt 1 0 0 -1;
Estimate 'Trt2 vs Trt3' Trt 0 1 -1 0;
Estimate 'Trt2 vs Trt4' Trt 0 1 0 -1;
Estimate 'Trt3 vs Trt4' Trt 0 0 1 -1;
quit;
```

The GLM Procedure

Class Level Information

Class	Levels	Values
Trt	4	1 2 3 4

Number of Observations Read	19
Number of Observations Used	18

The GLM Procedure

Dependent Variable: EcoliP1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.95150778	0.31716926	0.87	0.4811
Error	14	5.11958667	0.36568476		
Corrected Total	17	6.07109444			

R-Square	Coeff Var	Root MSE	EcoliP1 Mean
0.156728	9.896297	0.604719	6.110556

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Trt	3	0.95150778	0.31716926	0.87	0.4811

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Trt	3	0.95150778	0.31716926	0.87	0.4811

Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		5.836000000 B	0.27043844	21.58	<.0001
Trt	1	0.617333333 B	0.44162413	1.40	0.1839
Trt	2	0.164000000 B	0.38245772	0.43	0.6746
Trt	3	0.454000000 B	0.38245772	1.19	0.2550
Trt	4	0.000000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely

estimable.

The GLM Procedure

Least Squares Means

Trt	EcoliP1 LSMEAN	Standard Error	Pr > |t|
1	6.45333333	0.34913453	<.0001
2	6.00000000	0.27043844	<.0001
3	6.29000000	0.27043844	<.0001
4	5.83600000	0.27043844	<.0001

The GLM Procedure

Dependent Variable: EcoliP1

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.45333333	0.44162413	1.03	0.3221
Trt1 vs Trt3	0.16333333	0.44162413	0.37	0.7170
Trt1 vs Trt4	0.61733333	0.44162413	1.40	0.1839
Trt2 vs Trt3	-0.29000000	0.38245772	-0.76	0.4609
Trt2 vs Trt4	0.16400000	0.38245772	0.43	0.6746
Trt3 vs Trt4	0.45400000	0.38245772	1.19	0.2550

The GLM Procedure

Dependent Variable: EcoliP2

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.11533333	0.11126745	-1.04	0.3175
Trt1 vs Trt3	-0.11733333	0.11126745	-1.05	0.3095
Trt1 vs Trt4	-0.05933333	0.11126745	-0.53	0.6022
Trt2 vs Trt3	-0.00200000	0.09636043	-0.02	0.9837
Trt2 vs Trt4	0.05600000	0.09636043	0.58	0.5704
Trt3 vs Trt4	0.05800000	0.09636043	0.60	0.5569

The GLM Procedure

Dependent Variable: EcoliP3

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.05800000	0.07725776	0.75	0.4652
Trt1 vs Trt3	0.02400000	0.07725776	0.31	0.7606
Trt1 vs Trt4	0.14800000	0.07725776	1.92	0.0761
Trt2 vs Trt3	-0.03400000	0.06690719	-0.51	0.6192
Trt2 vs Trt4	0.09000000	0.06690719	1.35	0.2000
Trt3 vs Trt4	0.12400000	0.06690719	1.85	0.0850

The GLM Procedure

Dependent Variable: EcoliP4

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.23333333	0.13915482	-1.68	0.1158
Trt1 vs Trt3	-0.20733333	0.13915482	-1.49	0.1584
Trt1 vs Trt4	-0.21933333	0.13915482	-1.58	0.1373
Trt2 vs Trt3	0.02600000	0.12051161	0.22	0.8323
Trt2 vs Trt4	0.01400000	0.12051161	0.12	0.9092
Trt3 vs Trt4	-0.01200000	0.12051161	-0.10	0.9221

The GLM Procedure

Dependent Variable: EcoliP5

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.03800000	0.08464604	-0.45	0.6604
Trt1 vs Trt3	-0.00200000	0.08464604	-0.02	0.9815
Trt1 vs Trt4	-0.13600000	0.08464604	-1.61	0.1304
Trt2 vs Trt3	0.03600000	0.07330562	0.49	0.6310
Trt2 vs Trt4	-0.09800000	0.07330562	-1.34	0.2026
Trt3 vs Trt4	-0.13400000	0.07330562	-1.83	0.0889

The GLM Procedure

Dependent Variable: EcoliP6

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.05933333	0.06777741	0.88	0.3961
Trt1 vs Trt3	0.03133333	0.06777741	0.46	0.6510
Trt1 vs Trt4	0.03933333	0.06777741	0.58	0.5709
Trt2 vs Trt3	-0.02800000	0.05869696	-0.48	0.6407
Trt2 vs Trt4	-0.02000000	0.05869696	-0.34	0.7384
Trt3 vs Trt4	0.00800000	0.05869696	0.14	0.8935

The GLM Procedure

Dependent Variable: EcoliP7

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.03333333	0.05024385	-0.66	0.5178
Trt1 vs Trt3	-0.03933333	0.05024385	-0.78	0.4468
Trt1 vs Trt4	-0.05533333	0.05024385	-1.10	0.2893
Trt2 vs Trt3	-0.00600000	0.04351245	-0.14	0.8923
Trt2 vs Trt4	-0.02200000	0.04351245	-0.51	0.6210
Trt3 vs Trt4	-0.01600000	0.04351245	-0.37	0.7186

The GLM Procedure

Dependent Variable: EcoliP8

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.00800000	0.03655654	-0.22	0.8299
Trt1 vs Trt3	-0.04400000	0.03655654	-1.20	0.2487
Trt1 vs Trt4	-0.06800000	0.03655654	-1.86	0.0840
Trt2 vs Trt3	-0.03600000	0.03165890	-1.14	0.2746
Trt2 vs Trt4	-0.06000000	0.03165890	-1.90	0.0789
Trt3 vs Trt4	-0.02400000	0.03165890	-0.76	0.4610

The GLM Procedure

Dependent Variable: EcoliP9

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.01666667	0.13800575	-0.12	0.9056
Trt1 vs Trt3	-0.06866667	0.13800575	-0.50	0.6265
Trt1 vs Trt4	0.03333333	0.13800575	0.24	0.8126
Trt2 vs Trt3	-0.05200000	0.11951649	-0.44	0.6701
Trt2 vs Trt4	0.05000000	0.11951649	0.42	0.6820
Trt3 vs Trt4	0.10200000	0.11951649	0.85	0.4078

The GLM Procedure

Dependent Variable: EcoliP10

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.61533333	0.77033442	0.80	0.4378
Trt1 vs Trt3	1.11533333	0.77033442	1.45	0.1697
Trt1 vs Trt4	1.00733333	0.77033442	1.31	0.2121
Trt2 vs Trt3	0.50000000	0.66712917	0.75	0.4660
Trt2 vs Trt4	0.39200000	0.66712917	0.59	0.5662
Trt3 vs Trt4	-0.10800000	0.66712917	-0.16	0.8737

The GLM Procedure

Dependent Variable: EcoliP11

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.14600000	0.28084770	-0.52	0.6113
Trt1 vs Trt3	0.04800000	0.28084770	0.17	0.8667
Trt1 vs Trt4	-0.32800000	0.28084770	-1.17	0.2624
Trt2 vs Trt3	0.19400000	0.24322124	0.80	0.4384
Trt2 vs Trt4	-0.18200000	0.24322124	-0.75	0.4667
Trt3 vs Trt4	-0.37600000	0.24322124	-1.55	0.1444

The GLM Procedure

Dependent Variable: EcoliP12

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.10200000	0.19594751	0.52	0.6108
Trt1 vs Trt3	0.10200000	0.19594751	0.52	0.6108
Trt1 vs Trt4	0.63600000	0.19594751	3.25	0.0059
Trt2 vs Trt3	0.00000000	0.16969553	0.00	1.0000
Trt2 vs Trt4	0.53400000	0.16969553	3.15	0.0071
Trt3 vs Trt4	0.53400000	0.16969553	3.15	0.0071

The GLM Procedure

Dependent Variable: EcoliP13

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.29800000	0.28060343	-1.06	0.3062
Trt1 vs Trt3	-0.37600000	0.28060343	-1.34	0.2016
Trt1 vs Trt4	-0.11200000	0.28060343	-0.40	0.6958
Trt2 vs Trt3	-0.07800000	0.24300970	-0.32	0.7530
Trt2 vs Trt4	0.18600000	0.24300970	0.77	0.4567
Trt3 vs Trt4	0.26400000	0.24300970	1.09	0.2957

The GLM Procedure

Dependent Variable: EcoliP14

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.08600000	0.24745254	-0.35	0.7334
Trt1 vs Trt3	-0.04400000	0.24745254	-0.18	0.8614
Trt1 vs Trt4	0.16600000	0.24745254	0.67	0.5132
Trt2 vs Trt3	0.04200000	0.21430019	0.20	0.8474
Trt2 vs Trt4	0.25200000	0.21430019	1.18	0.2592
Trt3 vs Trt4	0.21000000	0.21430019	0.98	0.3438

The GLM Procedure

Dependent Variable: EcoliP15

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.10000000	0.14793564	0.68	0.5101
Trt1 vs Trt3	0.09200000	0.14793564	0.62	0.5440
Trt1 vs Trt4	0.21000000	0.14793564	1.42	0.1776
Trt2 vs Trt3	-0.00800000	0.12811602	-0.06	0.9511
Trt2 vs Trt4	0.11000000	0.12811602	0.86	0.4050
Trt3 vs Trt4	0.11800000	0.12811602	0.92	0.3726

The GLM Procedure

Dependent Variable: EcoliP16

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.10466667	0.26782700	0.39	0.7018
Trt1 vs Trt3	0.06266667	0.26782700	0.23	0.8184
Trt1 vs Trt4	0.01866667	0.26782700	0.07	0.9454
Trt2 vs Trt3	-0.04200000	0.23194499	-0.18	0.8589
Trt2 vs Trt4	-0.08600000	0.23194499	-0.37	0.7164
Trt3 vs Trt4	-0.04400000	0.23194499	-0.19	0.8523

The GLM Procedure

Dependent Variable: EcoliP17

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.35400000	0.30153386	-1.17	0.2600
Trt1 vs Trt3	-0.31400000	0.30153386	-1.04	0.3154
Trt1 vs Trt4	-0.67200000	0.30153386	-2.23	0.0427
Trt2 vs Trt3	0.04000000	0.26113598	0.15	0.8804
Trt2 vs Trt4	-0.31800000	0.26113598	-1.22	0.2434
Trt3 vs Trt4	-0.35800000	0.26113598	-1.37	0.1920

The GLM Procedure

Dependent Variable: EcoliP18

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.23600000	0.24606929	-0.96	0.3538
Trt1 vs Trt3	-0.04400000	0.24606929	-0.18	0.8606
Trt1 vs Trt4	-0.32000000	0.24606929	-1.30	0.2144
Trt2 vs Trt3	0.19200000	0.21310226	0.90	0.3828
Trt2 vs Trt4	-0.08400000	0.21310226	-0.39	0.6994
Trt3 vs Trt4	-0.27600000	0.21310226	-1.30	0.2162

The GLM Procedure

Dependent Variable: EcoliP19

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.43533333	0.46549295	0.94	0.3655
Trt1 vs Trt3	0.39733333	0.46549295	0.85	0.4077
Trt1 vs Trt4	0.28133333	0.46549295	0.60	0.5553
Trt2 vs Trt3	-0.03800000	0.40312872	-0.09	0.9262
Trt2 vs Trt4	-0.15400000	0.40312872	-0.38	0.7082
Trt3 vs Trt4	-0.11600000	0.40312872	-0.29	0.7778

The GLM Procedure

Dependent Variable: EcoliP20

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.43533333	0.46549295	0.94	0.3655
Trt1 vs Trt3	0.39733333	0.46549295	0.85	0.4077
Trt1 vs Trt4	0.28133333	0.46549295	0.60	0.5553
Trt2 vs Trt3	-0.03800000	0.40312872	-0.09	0.9262
Trt2 vs Trt4	-0.15400000	0.40312872	-0.38	0.7082
Trt3 vs Trt4	-0.11600000	0.40312872	-0.29	0.7778

The GLM Procedure

Dependent Variable: EcoliP20

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.02200000	0.11256998	-0.20	0.8479
Trt1 vs Trt3	0.01200000	0.11256998	0.11	0.9166
Trt1 vs Trt4	0.00200000	0.11256998	0.02	0.9861
Trt2 vs Trt3	0.03400000	0.09748846	0.35	0.7325
Trt2 vs Trt4	0.02400000	0.09748846	0.25	0.8091
Trt3 vs Trt4	-0.01000000	0.09748846	-0.10	0.9198

The GLM Procedure

Dependent Variable: EcoliP21

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.02533333	0.13372704	-0.19	0.8525
Trt1 vs Trt3	0.05866667	0.13372704	0.44	0.6676
Trt1 vs Trt4	0.04666667	0.13372704	0.35	0.7323
Trt2 vs Trt3	0.08400000	0.11581101	0.73	0.4802
Trt2 vs Trt4	0.07200000	0.11581101	0.62	0.5441
Trt3 vs Trt4	-0.01200000	0.11581101	-0.10	0.9189

The GLM Procedure

Dependent Variable: EcoliP22

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.46600000	0.32178786	-1.45	0.1696
Trt1 vs Trt3	-0.29800000	0.32178786	-0.93	0.3701
Trt1 vs Trt4	-0.38400000	0.32178786	-1.19	0.2526
Trt2 vs Trt3	0.16800000	0.27867646	0.60	0.5562
Trt2 vs Trt4	0.08200000	0.27867646	0.29	0.7729
Trt3 vs Trt4	-0.08600000	0.27867646	-0.31	0.7622

The GLM Procedure

Dependent Variable: EcoliP23

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.20333333	0.39430550	-0.52	0.6141
Trt1 vs Trt3	-0.39133333	0.39430550	-0.99	0.3378
Trt1 vs Trt4	-0.58933333	0.39430550	-1.49	0.1572
Trt2 vs Trt3	-0.18800000	0.34147858	-0.55	0.5906
Trt2 vs Trt4	-0.38600000	0.34147858	-1.13	0.2773
Trt3 vs Trt4	-0.19800000	0.34147858	-0.58	0.5712

The GLM Procedure

Dependent Variable: EcoliP24

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.02000000	0.21570703	-0.09	0.9274
Trt1 vs Trt3	-0.20400000	0.21570703	-0.95	0.3603
Trt1 vs Trt4	-0.19200000	0.21570703	-0.89	0.3885
Trt2 vs Trt3	-0.18400000	0.18680777	-0.98	0.3414
Trt2 vs Trt4	-0.17200000	0.18680777	-0.92	0.3728
Trt3 vs Trt4	0.01200000	0.18680777	0.06	0.9497

The GLM Procedure

Dependent Variable: EcoliP25

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.00000000	0.09076185	-0.00	1.0000
Trt1 vs Trt3	-0.10600000	0.09076185	-1.17	0.2624
Trt1 vs Trt4	-0.03400000	0.09076185	-0.37	0.7136
Trt2 vs Trt3	-0.10600000	0.07860207	-1.35	0.1989
Trt2 vs Trt4	-0.03400000	0.07860207	-0.43	0.6719
Trt3 vs Trt4	0.07200000	0.07860207	0.92	0.3752

The GLM Procedure

Dependent Variable: EcoliP26

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.00000000	0.00349149	0.00	1.0000
Trt1 vs Trt3	0.00000000	0.00349149	0.00	1.0000
Trt1 vs Trt4	-0.00600000	0.00349149	-1.72	0.1077
Trt2 vs Trt3	0.00000000	0.00302372	0.00	1.0000
Trt2 vs Trt4	-0.00600000	0.00302372	-1.98	0.0672
Trt3 vs Trt4	-0.00600000	0.00302372	-1.98	0.0672

The GLM Procedure

Dependent Variable: EcoliP27

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0	0	.	.
Trt1 vs Trt3	0	0	.	.
Trt1 vs Trt4	0	0	.	.
Trt2 vs Trt3	0	0	.	.
Trt2 vs Trt4	0	0	.	.
Trt3 vs Trt4	0	0	.	.

The GLM Procedure

Dependent Variable: EcoliP28

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.38333333	0.76088336	0.50	0.6222
Trt1 vs Trt3	0.98533333	0.76088336	1.29	0.2163
Trt1 vs Trt4	0.94933333	0.76088336	1.25	0.2326
Trt2 vs Trt3	0.60200000	0.65894432	0.91	0.3764
Trt2 vs Trt4	0.56600000	0.65894432	0.86	0.4048
Trt3 vs Trt4	-0.03600000	0.65894432	-0.05	0.9572

The GLM Procedure

Dependent Variable: EcoliP29

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.10466667	0.22269575	0.47	0.6456
Trt1 vs Trt3	0.03666667	0.22269575	0.16	0.8716
Trt1 vs Trt4	0.23466667	0.22269575	1.05	0.3098
Trt2 vs Trt3	-0.06800000	0.19286018	-0.35	0.7296
Trt2 vs Trt4	0.13000000	0.19286018	0.67	0.5112
Trt3 vs Trt4	0.19800000	0.19286018	1.03	0.3220

The GLM Procedure

Dependent Variable: EcoliP30

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.06066667	0.13531855	0.45	0.6608
Trt1 vs Trt3	-0.01333333	0.13531855	-0.10	0.9229
Trt1 vs Trt4	-0.03133333	0.13531855	-0.23	0.8202
Trt2 vs Trt3	-0.07400000	0.11718931	-0.63	0.5379
Trt2 vs Trt4	-0.09200000	0.11718931	-0.79	0.4455
Trt3 vs Trt4	-0.01800000	0.11718931	-0.15	0.8801

The SAS System 14:05 Wednesday, August 9, 2006 128

The GLM Procedure

Dependent Variable: EcoliP31

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.19600000	0.22542404	-0.87	0.3993
Trt1 vs Trt3	-0.23000000	0.22542404	-1.02	0.3249
Trt1 vs Trt4	-0.24000000	0.22542404	-1.06	0.3050
Trt2 vs Trt3	-0.03400000	0.19522295	-0.17	0.8642
Trt2 vs Trt4	-0.04400000	0.19522295	-0.23	0.8249
Trt3 vs Trt4	-0.01000000	0.19522295	-0.05	0.9599

The SAS System 14:05 Wednesday, August 9, 2006 132

The GLM Procedure

Dependent Variable: EcoliP32

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	-0.31400000	0.43546232	-0.72	0.4827
Trt1 vs Trt3	-0.45800000	0.43546232	-1.05	0.3107
Trt1 vs Trt4	-0.58400000	0.43546232	-1.34	0.2012
Trt2 vs Trt3	-0.14400000	0.37712143	-0.38	0.7083
Trt2 vs Trt4	-0.27000000	0.37712143	-0.72	0.4858
Trt3 vs Trt4	-0.12600000	0.37712143	-0.33	0.7432

The SAS System 14:05 Wednesday, August 9, 2006 136

The GLM Procedure

Dependent Variable: EcoliP33

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.00000000	0.02035869	0.00	1.0000
Trt1 vs Trt3	-0.01200000	0.02035869	-0.59	0.5650
Trt1 vs Trt4	-0.02000000	0.02035869	-0.98	0.3426
Trt2 vs Trt3	-0.01200000	0.01763114	-0.68	0.5072
Trt2 vs Trt4	-0.02000000	0.01763114	-1.13	0.2757
Trt3 vs Trt4	-0.00800000	0.01763114	-0.45	0.6570

The SAS System 14:05 Wednesday, August 9, 2006 140

The GLM Procedure

Dependent Variable: EcoliP34

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0.00000000	0.00174574	0.00	1.0000
Trt1 vs Trt3	0.00000000	0.00174574	0.00	1.0000
Trt1 vs Trt4	-0.00200000	0.00174574	-1.15	0.2711
Trt2 vs Trt3	0.00000000	0.00151186	0.00	1.0000
Trt2 vs Trt4	-0.00200000	0.00151186	-1.32	0.2071
Trt3 vs Trt4	-0.00200000	0.00151186	-1.32	0.2071

The GLM Procedure

Dependent Variable: EcoliP35

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0	0	.	.
Trt1 vs Trt3	0	0	.	.
Trt1 vs Trt4	0	0	.	.
Trt2 vs Trt3	0	0	.	.
Trt2 vs Trt4	0	0	.	.
Trt3 vs Trt4	0	0	.	.

The GLM Procedure

Dependent Variable: EcoliP36

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Trt1 vs Trt2	0	0	.	.
Trt1 vs Trt3	0	0	.	.
Trt1 vs Trt4	0	0	.	.
Trt2 vs Trt3	0	0	.	.
Trt2 vs Trt4	0	0	.	.
Trt3 vs Trt4	0	0	.	.

APPENDIX IV

Developing model with the data set of mean of replicated values

Need: It was observed that the GRNN model developed with the data set of means of 3 replicated values (Chacon's data) was able to predict the *E.coli* population for the all production set patterns. Whereas the model developed with the entire data of 3 replicated values was not able to predict the same. So, there needs to be justification that the model developed with mean data will be able to predict the values for the whole size range of each mean value.

The GRNN models were separately developed from the data sets having replication 1, replication 2, and replication 3. The GLM procedure of SAS program was used to compare the means of these three model predictions with the mean of observed values (See the SAS program results). Scheffe's test was used to group all these means. The results showed that all the mean values 2.800, 2.749, 2.700, 2.666 were coming under a same group 'A' with a minimum significant difference of 1.4161. It was concluded that there is no significant difference among the means of mean values and the three individual replicated values. So, the model developed with the mean observed values will be applicable to replicated values too.

Comparing the means – Scheffe's test using Proc GLM of SAS

Prorgam & Results

```
Options linesize=110;
Data one;
Input rep $ @;
Do Plant = 1 to 36;
      Input cp @;
      Output;
End;
Cards;
R1    6.525 5.000 4.955 4.920 4.782 4.409 4.267 4.201 4.155 6.480 3.304
      3.190 3.036 2.684 1.821 1.398 1.154 0.946 6.468 3.015 2.970 2.626
      1.881 0.295 0.093 0.040 0.016 6.392 2.205 2.048 1.848 1.339 0.331
      0.103 0.040 0.015
R2    6.125 4.978 4.967 4.960 4.878 4.271 4.130 4.087 4.080 6.055 3.826
      3.634 3.382 2.652 1.632 1.430 0.846 0.239 6.093 3.610 3.540 3.537
      2.747 0.000 0.000 0.000 0.000 5.947 3.277 2.936 2.413 0.526 0.001
      0.000 0.000 0.000
R3    6.219 4.893 4.862 4.840 4.736 4.384 4.250 4.183 4.132 5.906 3.443
      3.336 3.191 2.762 1.462 0.826 0.516 0.306 6.184 3.306 3.190 3.009
      2.026 0.189 0.032 0.008 0.002 5.827 2.727 2.452 2.215 1.525 0.231
      0.035 0.008 0.002
Rm    6.453 4.916 5.140 4.735 4.670 4.583 4.117 4.079 4.023 6.453 4.102
      4.112 2.580 2.280 1.702 1.068 0.000 0.000 6.453 3.550 3.165 2.162
      1.068 0.000 0.000 0.000 0.000 6.453 3.455 2.704 1.962 0.000 0.000
      0.000 0.000 0.000
Proc GLM Data=one;
class rep;
Model cp=rep/solution;
Contrast 'All vs. Rm' rep .33 .33 .33 -1;
Means rep/Scheffe hovtest=bartlett;
/*Lsmeans rep/stderr Pdiff adjust=Scheffe;
Output out=oneOut Residual=cpRes Predicted=cpPred;
Proc Univariate Plot Normal Data=oneOut;
Var cpRes;*/
quit;
```

The SAS System 16:36 Wednesday, July 26, 2006 3

The GLM Procedure

Class Level Information

Class	Levels	Values
rep	4	R1 R2 R3 Rm

Number of Observations Read	144
Number of Observations Used	144

The SAS System 16:36 Wednesday, July 26, 2006 4

The GLM Procedure

Dependent Variable: cp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.3664182	0.1221394	0.03	0.9940
Error	140	631.0315965	4.5073685		
Corrected Total	143	631.3980147			

R-Square	Coeff Var	Root MSE	cp Mean
0.000580	77.80108	2.123056	2.728826

Source	DF	Type I SS	Mean Square	F Value	Pr > F
rep	3	0.36641819	0.12213940	0.03	0.9940

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rep	3	0.36641819	0.12213940	0.03	0.9940

Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		2.666250000 B	0.35384274	7.54	<.0001
rep	R1	0.082416667 B	0.50040920	0.16	0.8694
rep	R2	0.133722222 B	0.50040920	0.27	0.7897
rep	R3	0.034166667 B	0.50040920	0.07	0.9457
rep	Rm	0.000000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

The GLM Procedure

Bartlett's Test for Homogeneity of cp Variance

Source	DF	Chi-Square	Pr > ChiSq
rep	3	0.3559	0.9492

The GLM Procedure

Scheffe's Test for cp

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	140
Error Mean Square	4.507369
Critical Value of F	2.66926
Minimum Significant Difference	1.4161

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	rep
A	2.8000	36	R2
A			
A	2.7487	36	R1
A			
A	2.7004	36	R3
A			
A	2.6663	36	Rm

APPENDIX V – SAS program Results for modeling DATA Set 1

The SAS System Results 1 16:42 Friday, January 27, 2006

Statistical model 1: for Chacon's data

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read	36
Number of Observations Used	36

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	170.88513	14.24043	51.25	<.0001
Error	23	6.39136	0.27789		
Corrected Total	35	177.27649			

Root MSE	0.52715	R-Square	0.9639
Dependent Mean	2.66556	Adj R-Sq	0.9451
Coeff Var	19.77630		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	256.84602	259.80547	0.99	0.3331
X1	1	-0.08195	0.04637	-1.77	0.0905
X2	1	0.77759	4.00953	0.19	0.8479
X3	1	-127.43547	369.42398	-0.34	0.7333
X4	1	-69.95072	43.34235	-1.61	0.1202
X1X2	1	-0.00123	0.00270	-0.46	0.6518
X1X3	1	0.07060	0.04834	1.46	0.1577
X1X4	1	0.00277	0.00133	2.08	0.0487
X2X3	1	-2.14422	2.22076	-0.97	0.3443
X2X4	1	0.20210	0.44341	0.46	0.6528
X3X4	1	19.74483	76.17431	0.26	0.7978
X4X4	1	5.02615	4.62203	1.09	0.2881
X1X2X3X4	1	0.00029	0.00064	0.45	0.6557

Appendix VI – DATA Set 2 (Al-Nabulsi's data)

Experimental design used for challenge studies of lactoferrin against *E. coli* O157:H7 in dry fermented sausages

Treatments	*E. coli* O157:H7[2]	Lactoferrin	mg lactoferrin /g sausage batter	Ppm EDTA	mM SB
Control	-	-	-	-	-
1	+	-	-	-	-
2	+	LF[3]	6.0	-	-
3	+	LF + EDTA+ SB[4]	6.0	500	5.0
4	+	Paste-like microcapsules of LF [5]	3.0	-	-
5	+	Dried powder microcapsules of LF [6]	1.8	-	-
6	+	Dried powder microcapsules of LF with EDTA and SB [7]	1.8	250	2.5

'+' and '-'signs mean respectively that the component was present and absent in the treatment

1- Meat starter cultures (7.2 log cfu/g *L. curvatus* and 6.6 log cfu/g *S. carnosus*) were added to all treatments.

2- 5.8 log cfu/g sausage batter

3- lactoferrin dissolved in distilled water and added to sausage batter.

4- lactoferrin , ethylene diamine tetraacetic acid (EDTA) and sodium bicarbonate (SB) dissolved in distilled water and added to sausage batter.

5- Water-in-oil emulsion; lactoferrin in distilled water was encapsulated in oil (78% corn oil + 22% butter fat with 0.1% polyglycerol polyricinoleate, PGPR).

6- Water-in-oil-in-water emulsion; lactoferrin in distilled water was encapsulated in oil (78% corn oil + 22% fat butter containing 0.1% PGPR) in 30% (w/v) whey protein isolate (WPI).

7- Water-in-oil-in-water emulsion; lactoferrin, ethylene diamine tetraacetic acid (EDTA) and sodium bicarbonate (SB) in distilled water were encapsulated in oil (78% corn oil + 22% fat butter with 0.1% PGPR) in 30% (w/v) WPI.

Source: Al-Nabulsi (2006)

Appendix VI – Data Set 2 - continued

	E.coli O157:H7 cells recovered on									
	on ct-SMAC agar									on APT/ ct-SMAC
Treat ment	time d0	time d1	time d2	time d3	time d6	time d9	time d15	time d21	time d28	time d28
Control	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Control	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Control	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Control	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Control	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Control	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	0.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
T1	5.93	5.50	4.86	4.86	4.40	4.72	4.47	3.20	3.00	
T1	5.78	5.03	4.96	4.96	4.20	3.86	3.68	3.51	3.15	
T1	5.30	4.94	4.88	4.58	4.57	3.92	3.83	3.31	3.29	
T1	6.09	4.60	4.45	4.48	4.58	3.48	3.62	3.53	3.34	
T1	5.82	4.78	4.76	4.65	4.20		3.97	3.68	3.50	
T1	5.64	4.30	4.81	4.77	4.20		3.81	3.28	3.04	
	5.76	**4.86**	**4.79**	**4.72**	**4.36**	**4.00**	**3.90**	**3.42**	**3.22**	**4.52**
T2	5.91	5.19	4.84	4.84	4.16	3.60	3.48	2.00	1.60	
T2	5.60	5.23	4.69	4.69	4.28	3.62	3.45	2.30	2.08	
T2	5.53	5.06	4.82	4.82	4.29	3.45	3.45	1.00	1.58	
T2	5.75	5.08	4.15	4.30	4.30	3.81	3.20	1.00	1.00	
T2	5.94	4.40	3.90	4.15	4.57	3.66	3.83	1.48	1.45	
T2	5.99	4.10	4.76	3.90	4.44	3.59		1.00	2.20	
	5.79	**4.84**	**4.53**	**4.45**	**4.34**	**3.62**	**3.48**	**1.46**	**1.65**	**2.74**
T3	5.84	5.52	4.95	4.94	4.67	3.90	4.59	4.15	3.50	
T3	6.09	4.99	4.76	5.11	4.89	3.60	4.72	4.20	3.73	
T3	5.48	4.87	4.91	5.05	3.34	4.36	4.48	4.40	4.60	
T3	5.91	5.23	5.12	5.24	3.50	4.35	4.40	4.50	4.44	
T3	5.92	4.97	4.88	5.09	4.48	3.08	4.44	4.10	3.97	
T3	5.76	5.17		4.70	4.46	3.06	4.32	4.50	4.17	
	5.83	**5.13**	**4.92**	**5.02**	**4.22**	**3.73**	**4.49**	**4.31**	**4.07**	**5.09**

Bolded numbers are average values of six replications

Source: Al-Nabulsi (2006)

Appendix VI– Data Set 2 - continued

	E.coli O157:H7 cells recovered on									
	on ct-SMAC agar									on APT/ ct-SMAC
Treatment	time d0	time d1	time d2	time d3	time d6	time d9	time d15	time d21	time d28	time d28
T4	5.83	5.08	4.66	4.68	4.52	4.00	3.58	2.45	2.45	
T4	5.69	4.99	4.47	4.60	4.56	4.00	3.68	1.80	1.30	
T4	5.63	5.06	4.64	4.56	4.52	4.15	3.53	2.00	2.38	
T4	5.72	5.01	4.61	4.71	4.68	3.83	3.66	1.70	2.20	
T4	5.66	4.79	4.72	4.72	4.58	3.45	3.84	1.75	1.48	
T4	5.39	4.66	4.51	3.30	4.31	4.73	3.75	2.30	1.30	
	5.65	**4.93**	**4.60**	**4.43**	**4.53**	**4.03**	**3.67**	**2.00**	**1.85**	**3.88**
T5	5.75	5.13	4.93	4.68	4.50	4.47	4.68	2.60	2.94	
T5	5.73	5.34	4.80	4.82	4.60	4.67	4.29	2.78	2.83	
T5	5.56	5.06	4.75	4.67	4.67	5.13	4.58	3.58	2.08	
T5	5.83	4.98	4.53	4.88	4.64	5.38	4.64	4.15	1.60	
T5	5.82	5.26	4.95	5.13	4.76	5.22	4.54	3.30	2.48	
T5	5.79	5.18	4.20	3.78	4.71	5.05	4.20	0.00	1.60	
	5.75	**5.16**	**4.69**	**4.66**	**4.65**	**4.99**	**4.49**	**2.74**	**2.26**	**4.53**
T6	5.78	5.25	5.07	4.85	4.68	4.67	5.02	3.15	1.30	
T6	6.53	4.95	5.06	4.88	4.72	4.88	4.77	3.53	1.90	
T6	5.92	4.91	5.18	5.04	4.66	4.83	5.05	3.82	1.70	
T6	6.15	5.19	4.96	4.69	4.53	4.99	4.89	3.50	1.60	
T6	6.44	4.96	4.91	4.61	5.10	3.98	4.99	0.00	2.78	
T6	4.60	4.80	4.50	3.78	4.79	4.20	4.87	0.00	3.23	
	5.90	**5.01**	**4.95**	**4.64**	**4.75**	**4.59**	**4.93**	**2.33**	**2.09**	**4.42**

Bolded numbers are average values of six replications

Source: Al-Nabulsi (2006)

Appendix VII – SAS program Results for modeling DATA Set 2

The SAS System results 17:38 Sunday, May 28, 2006 1

The REG Procedure

Statistical Model 2: MODEL FOR Al-Nabulsi's data

Dependent Variable: Yo

Number of Observations Read	324
Number of Observations Used	324

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	333.08985	47.58426	126.09	<.0001
Error	316	119.25657	0.37739		
Corrected Total	323	452.34642			

Root MSE	0.61432	R-Square	0.7364
Dependent Mean	4.24355	Adj R-Sq	0.7305
Coeff Var	14.47666		

NOTE: Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased.

NOTE: The following parameters have been set to 0, since the variables are a linear combination of her variables as shown.

ED = 100 * SB

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.17027	0.09473	54.58	<.0001
LF	1	-0.00221	0.02728	-0.08	0.9355
SB	B	0.13686	0.06365	2.15	0.0323
ED	0	0	.	.	.
Day	1	-0.08261	0.00715	-11.56	<.0001
LFEDSB	1	-0.00005340	0.00002403	-2.22	0.0270
LFDay	1	-0.01044	0.00206	-5.07	<.0001
EDSBDay	1	-0.00005051	0.00002333	-2.17	0.0311
LFEDSBDay	1	0.00001525	0.00000399	3.82	0.0002