OXFORD

AMERICAN SOCIETY OF **ANIMAL SCIENCE**

ASAS-NANP Symposium: Mathematical Modeling in Animal Nutrition: The progression of data analytics and artificial intelligence in support of sustainable development in animal science

Luis O. Tedeschi[*],

Department of Animal Science, Texas A&M University, College Station, TX 77843-2471, USA

*Corresponding author: luis.tedeschi@tamu.edu

Abstract

A renewed interest in data analytics and decision support systems in developing automated computer systems is facilitating the emergence of hybrid intelligent systems by combining artificial intelligence (**AI**) algorithms with classical modeling paradigms such as mechanistic modeling (**HIMM**) and agent-based models (**iABM**). Data analytics have evolved remarkably, and the scientific community may not yet fully grasp the power and limitations of some tools. Existing statistical assumptions might need to be re-assessed to provide a more thorough competitive advantage in animal production systems towards sustainability. This paper discussed the evolution of data analytics from a competitive advantage perspective within academia and illustrated the combination of different advanced technological systems in developing HIMM. The progress of analytical tools was divided into three stages: collect and respond, predict and prescribe, and smart learning and policy making, depending on the level of their sophistication (simple to complicated analysis). The collect and respond stage is responsible for ensuring the data is correct and free of influential data points, and it represents the data and information phases for which data are cataloged and organized. The predict and prescribe stage results in gained knowledge from the data and comprises most predictive modeling paradigms, and optimization and risk assessment tools are used to prescribe future decision-making opportunities. The third stage aims to apply the information obtained in the previous stages to foment knowledge and use it for rational decisions. This stage represents the pinnacle of acquired knowledge that leads to wisdom, and AI technology is intrinsic. Although still incipient, HIMM and iABM form the forthcoming stage of competitive advantage. HIMM may not increase our ability to understand the underlying mechanisms controlling the outcomes of a system, but it may increase the predictive ability of existing models by helping the analyst explain more of the data variation. The scientific community still has some issues to be resolved, including the lack of transparency and reporting of AI that might limit code reproducibility. It might be prudent for the scientific community to avoid the shiny object syndrome (i.e., AI) and look beyond the current knowledge to understand the mechanisms that might improve productivity and efficiency to lead agriculture towards sustainable and responsible achievements.

Lay Summary

Data analytics have evolved remarkably. This paper discussed the evolution of data analytics from a competitive advantage perspective within academia and illustrated the combination of different advanced technological systems in developing hybrid intelligent mechanistic models (**HIMM**). Data analytics tools are divided into 3 stages. The first stage (collect and respond) ensures that data are correct and free of influential data points, and it represents the data and information phases for which data are cataloged and organized. The second stage (predict and prescribe) results in gained knowledge from the data and comprises most predictive modeling paradigms, and optimization and risk assessment tools are used to prescribe future decision-making opportunities. The third stage (smart learning and policy making) aims to apply the information obtained in the previous stages to foment knowledge and use it for rational decisions. Although still incipient, HIMM form the forthcoming stage of competitive advantage. HIMM may not increase our ability to understand the underlying mechanisms controlling the outcomes of a system, but it may increase the predictive ability of existing models by helping the analyst explain more of the data variation. The scientific community needs to resolve the lack of transparency and reporting of artificial intelligence for code reproducibility.

Key words: analytics, animal science, computer, modeling, predictive, simulation

Abbreviations: AI, artificial intelligence; DIKW, data-information-knowledge-wisdom; DL, deep learning; DSS, decision support systems; HIMM, hybrid intelligent mechanistic models; iABM, intelligent agent-based models; IQR, interquartile range; ML, machine learning; MM, mechanistic modeling; NASEM, National Academies of Sciences; Engineering; and Medicine; NRC, National Research Council; SOS, shiny object syndrome; VIF, variance inflation factor

Introduction

The interest in data collection and analysis began with the advancements in digital computing driven by powerful computer algorithms and specialized software to manipulate and process data since the mid-1980s when they became more accessible to the public. Recently, *data analytics* gained impetus fueled by big data, cloud networking, artificial intelligence (**AI**), and increased competition among organizations (Herden, 2020) at the industry and university levels. Data analytics and decision support systems (**DSS**) are tightly connected; DSS results from building insights gained from data analytics into an automated form (i.e., a dedicated computer

tool) to solve specific problems or defined goals. Decision support systems, however, predate *data analytics*, and they have been applied for communications-driven, data-driven, document-driven, knowledge-driven, and model-driven applications when digital computing became broadly available in the 1960s (Power, 2008). Animal scientists were introduced to DSS in the 1970s when academics began developing nutrition models from accumulated scientific knowledge (Tedeschi et al., 2014; Tedeschi and Fox, 2020). Nevertheless, it has been a road full of hurdles, and many failures in DSS adoption have been documented (Newman et al., 2000) during this journey. The expectations for DSS to solve all problems within a production context were too high, in part because data had limited availability, computational processes were still rustic to the desired outcome, and there was a lack of proper training of the workforce, more specifically the next generation of students that could have made the difference between success and failures in using this technology. Given the increased availability of data through precision livestock farming initiatives (Tedeschi et al., 2021), improved data visualization (Morota et al., 2021), and AI (Wang et al., 2021), the expectations have been renewed with data analytics sparking new motivations to develop more powerful DSS by combining different tools to understand (and apply) the data.

Experimental design (Stanley, 1957) and statistical methods (Michael et al., 1957) have been around since the 1940s, after classical works by William Gossett and Sir Ronald Fisher were published to illustrate the basics of statistical inference (Cochran and Cox, 1957; Snedecor and Cochran, 1971; Steel and Torrie, 1960). The fundamental knowledge has not changed; it will forever be. What is changing is the emergence of new analytical techniques that allow us to perform faster data analysis using more extensive, dynamic databases brought about different ways to collect data, better and faster ways to store digital data, and faster computers and algorithms to analyze the data (Tedeschi, 2019; Tedeschi et al., 2021). There is nothing wrong with "old statistical methods"; digital devices and their ever-increasing processing speed have facilitated data collection, storage, compilation, and analysis. Researchers have more diverse data being collected that was unforeseen in the 1940s; thus, data analytics are changing, and existing statistical assumptions might have to be re-assessed.

Variations in the definition of *analytics* exist (Holsapple et al., 2014), and they are usually oriented toward their purpose of use, mainly for business applications. Holsapple et al. (2014) condensed *data analytics* as a technique "concerned with evidence-based problem recognition and solving that happen within the context of business situations." Tedeschi (2019) defined *data analytics* from a model-building perspective as "the process of examining data sets to obtain relationships among variables and to draw conclusions from the information therein" through the use of statistical tools. From a business perspective, *analytics* comprises one or more quantifiable approaches used to extract meaningful information from data in developing technological systems to assist in decision-making. Davenport and Harris (2017) assigned four different categories to drive decisions and actions in technological systems, as follows: descriptive, predictive, prescriptive (or forecasting), and autonomous (or self-ruling or self-learning), and their level of sophistication usually goes from basic for descriptive systems to complex for self-learning systems. Figure 1 depicts a revised evolution of different technological systems based on Davenport and Harris' (2017)

sketch of businesses' relative competitive advantage for data analytics and different levels of technological sophistication. In general, the greater the technological sophistication (i.e., knowledge formation and data analytics acuity), the greater the competitive advantage is expected (i.e., more helpful and insightful information is obtained from the data).

Although AI has been frequently assigned to the most recent group of technological tools for data analytics (i.e., learning and policy making) to increase operational efficiency, the implementation of AI can be challenging and costly. A learning period is necessary to get the most impactful and powerful benefits of AI, i.e., the monotonous, tedious, and time-consuming tasks requiring the processing of large amounts of data collected through automation and sensor technology (Tedeschi et al., 2021). Given its effective and efficient attributes to find patterns in larger data sets, AI (either machine learning, **ML**; or deep learning, **DL**) can process data to assist in finding trends to forecast outcomes, but current AI algorithms cannot still explain why and how a result was reached, i.e., AI by itself cannot provide insightful knowledge that leads to wisdom (Tedeschi, 2019). Figure 2 illustrates the relative assessment of the suitability of AI within each phase of the data-information-knowledge-wisdom (**DIKW**) structure (Ackoff, 1989; Tedeschi, 2019).

This paper aims to briefly examine the evolution of data analytics to gain insights from a competitive advantage perspective within the context of scientific research and to illustrate the combination of different advanced technological systems (i.e., model paradigms) in developing hybrid intelligent mechanistic models (**HIMM**) to support sustainable animal production systems.

Competitive Advantage in Animal Production Systems

Davenport and Harris (2017) and Herden (2020) provided comprehensive discussions about using data analytics for competitive advantage from a business perspective. Figure 1 illustrates the competitive advantage of different analytical tools for various science fields. As the world becomes more complex in analytics, the aim is to squeeze the data to get more valuable information in such a competitive environment. Figure 1 also depicts an exponential, sequential progression of analytical tools when, in fact, it might be better represented by branched advancements in which there are different branches of data analytics along with their particular dependencies, changes, and improvements over time. Typical questions that need to be answered before using such technological systems include what we can learn from the data? Do we need more data? Are there different viewpoints on the data? Which level of competitive advantage (i.e., sophistication complexity) needs to be employed?

Collect and respond

Regardless of the aimed level of analytics sophistication, the raw data cleaning up phase is the first step in data analytics. The analyst has to decide on the robustness of the data and how far (analytics progression) they can go with the data in hand. Therefore, although tiresome, determining and identifying ways to remove outliers, leverage, incorrect, and missing data are essential to establish the truth about the data. Critical information includes variable characteristics (e.g., mean, deviation, and distribution) and their relationships
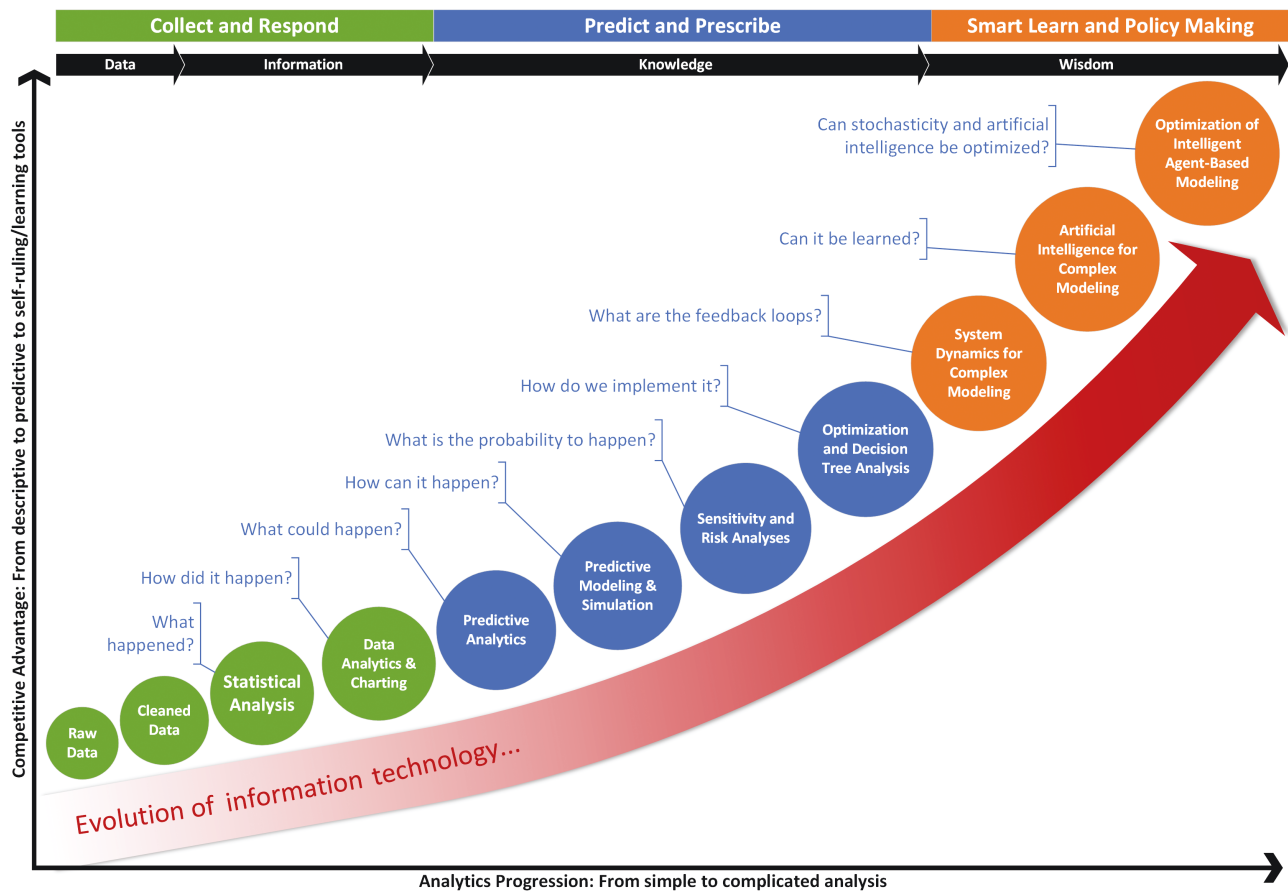
Figure 1. Evolution of technological systems based on their competitive advantage against the level of sophistication. The size of the circles is relative to the magnitude of technology and the cumulative specialized knowledge on how to acquire insights from data and information analysis. Based on Davenport and Harris (2017).

among other variables (e.g., correlation and covariance). The analyst seeks the genuine relationship among variables to discover and form new knowledge for future wise decisions. Although the DIKW concept and competitive advantage are not synonyms, they are complementary and provide vital steps in the learning process. They explore different niches of data and information analysis.

The collect and respond phase occurs in the early steps of data analytics, and it comprises the most basic technological systems of competitive advantage (Figure 1). It provides the steps necessary for the data and information in the DIKW pyramid (Figure 2). Although often neglected, data collection is the most critical step. Errors during this step can sometimes be adjusted or eliminated, but questions will constantly challenge the validity of the removal of data points that do not comply with a perceived trend or outcome. Thus, it is of utmost importance to ensure the correctness of the data during the collection phase before any data manipulation is performed. Once the correctness of the data has been confirmed and ill-positioned data points still exist, there are powerful statistical tools to deal with such data points (outliers), including leverage, studentized and semi-studentized residue, $PRESS_p$ (prediction sum of squares) criterion, and the studentized deleted residue (Kutner et al., 2005; Neter et al., 1996). Correct data are necessary not only for future analytics but also to describe and provide insights into what has happened and uncover the relationships among variables.

Extreme and influential points are usually identified with DFFITS influential statistics (i.e., the difference between fitted values with and without the ith data point), Cook's distance, and BFBETAS (i.e., the difference between regression parameters with and without the ith data point) measure (Kutner et al., 2005; Neter et al., 1996), or specialized calculations such as Rosner's test for detecting up to kth outliers (Gilbert, 1987; Rosner, 1975, 1983). However, these statistics are valid for a known statistical regression in which specific data points are tagged as outliers or influential points, assuming a particular relationship between dependent and independent variables. These methods rely almost exclusively on a given distribution, and they single out data points that do not fit the expected pattern, usually a symmetric shape. In this sense, multicollinearity diagnosis (i.e., variance inflation factor, **VIF**) is also handy to identify independent variables that are highly, mutually correlated among themselves (Kutner et al., 2005; Neter et al., 1996). From a simplistic modeling perspective, the removal of VIF might be the first step to increasing model identifiability (Boston et al., 2007; Godfrey and DiStefano, 1985; Tedeschi and Boston, 2010). Alternatives to identifying outliers and influential points exist for those that do not follow a specific distribution or have an asymmetric shape. Tukey's (1977) proposed the boxplot (i.e., box-and-whisker plot) to provide a graphical representation of the data without any formal assumption about its distribution (or depending
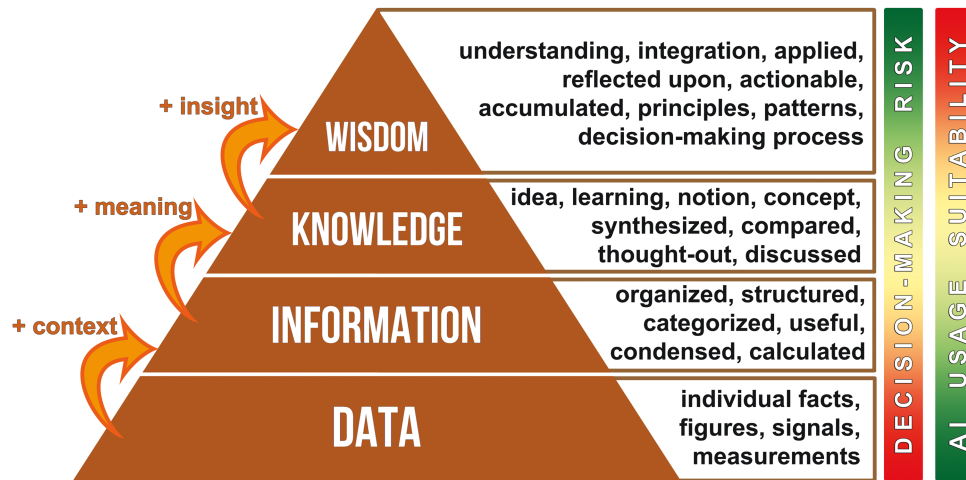
Figure 2. A relative assessment of artificial intelligence (AI) suitability for each step in the data–information–knowledge–wisdom pyramid. The color scale indicates high risk or not suitable (red) and low risk or suitable (green). The sketch was adapted and replicated with permission from Tedeschi (2019).

on statistics that assume a given distribution). It is used as an exploratory graphical tool to obtain nonparametric statistical information and analyze the distribution (Friedman and Stuetzle, 2002). Tukey (1977) insisted that "there is no excuse for failing to plot and look" because "the greatest value of a picture is when it forces us to notice what we never expected to see." Data visualization is a must in data analytics, and nowadays, there are many ways to graphically analyze data (Morota et al., 2021; Weissgerber et al., 2017). As shown in Figure 3, the boxplot is a rectangle that contains five statistics of interest: the minimum and maximum values (excluding outliers), the first and third quartile, and the median (second quartile) of the data. The minimum and the maximum values are represented by whiskers' lengths below and above the rectangle (i.e., box). The first quartile separates the lowest 25% of data from the highest 75% of data, and the third quartile divides the lowest 75% of data from the highest 25% of data. The difference between the first and third quartiles is called the interquartile range (**IQR**). Data greater (or lower) than $1.5 \times IQR$ (i.e., inner fences) are considered suspected outliers (solid circles; Figure 3), and those greater (or lower) than $3 \times IQR$ (i.e., outer fences) are outliers (empty circles; Figure 3). The average is usually shown in the boxplot as an asterisk. A normally distributed data would have the median and the average in the middle of the boxplot's rectangle sketch, and the whiskers would be of similar length, without (or few) outliers. Skewness closer to zero indicates an even distribution, and kurtosis closer to three shows a distribution equal to a normal distribution. Tukey's method is quite effective for large databases, but a transformation might be necessary for highly skewed data. There is no clear explanation why Tukey used $1.5 \times IQR$ and $3 \times IQR$ to separate potential outliers from outliers. It is possible that large datasets or datasets with different distributions might need different thresholds. Furthermore, some methods might have limitations; Rosner's test for detecting outliers is limited to up to 10 outliers. Given the assertions of R. C. Geary, E. S. Pearson, and others that normally distributed data are an illusion; it never existed, and it will never be (Tiku and Akkaya, 2004), the question remains, how effective are these methods to detect outliers when the data cannot be deemed normally distributed or when the number of data points surpasses a given threshold?

There are other methods to deal with outliers when removal is not an option. Typical robust regression methods include the median to estimate the slope and intercept (Andrews, 1974; Theil, 1992). The so-called Theil-Sen approach developed in the 1950s (Sen, 1968; Theil, 1950a, b, c) has since been used for fitting single and multiple linear regressions when outliers exist (Andrews, 1974; Siegel, 1982), but it was not likely the first mention of using the median for fitting linear regression. Wald (1940) proposed separating the data points into two groups depending on the median of X values—then computing the slope between two points: the means of X and Y for the group on the left of the median of X and another for the group on the right of the median of X. Wald's (1940) ideas likely generated sufficient interest in using the median rather than the mean for fitting linear regressions. Subsequently, other researchers have expanded on using the median to circumvent the problems caused by outliers when fitting regressions (Walters et al., 2006). However, more elaborated methods based on different measurements of scale and location (Andrews et al., 1972) with higher breakdown values were developed to *ignore* or *minimize* the impact of outliers on the parameter estimates of regressions, including quantiles, winsorized mean, trimmed mean, M-measures with diverse influence functions (e.g., Huber, Andrews, Hampel, and biweight) to list a few. Breakdown value (or point) measures the robustness of an estimator against the presence of outliers; it indicates the smallest fraction of contaminants in a sample that causes the estimator to break down (Hubert and Debruyne, 2009). Thus, the so-called robust regression analysis became essential in curbsiding outliers and extreme data points to obtain robust estimates with high breakdown values (Wilcox, 2012). G. E. P. Box firstly introduced the technical term *robust* in 1953, but only in the 1960s it gained some popularity, and yet it was deemed inexact and "dirty" (Huber and Ronchetti, 2009). Nonetheless, the efficiency of the Theil-Sen approach for small sample size datasets is still acceptable compared with different *robust regression* approaches (Wilcox, 1996), given its reasonably high breakdown point and a bounded influence function
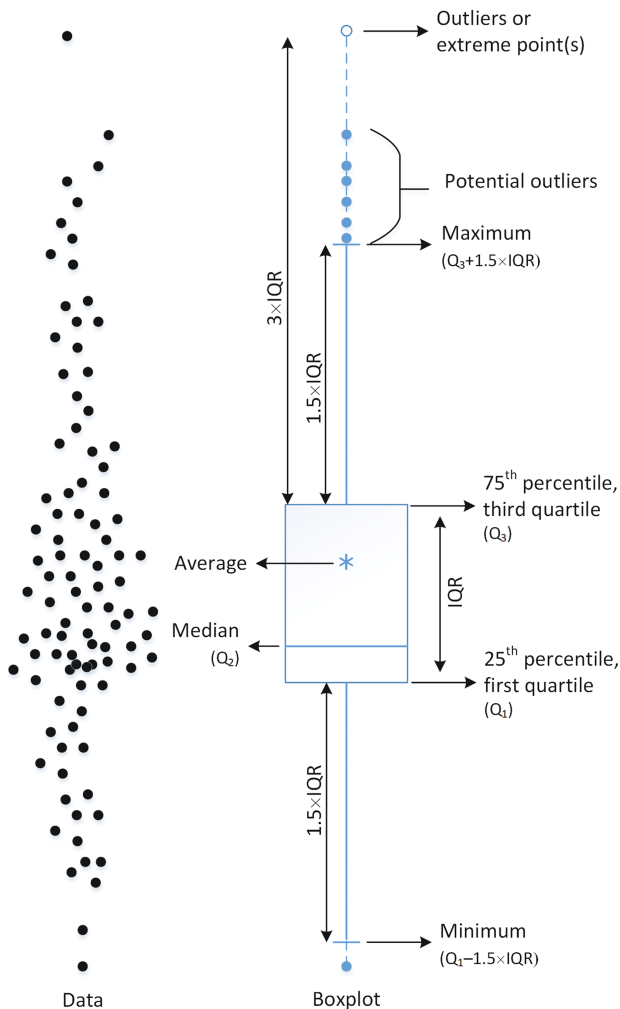
Figure 3. A graphical representation of boxplot (box-and-whisker plot) showing potential outliers and extreme points beyond the 1.5 times the interquartile range (**IQR**), i.e., end of the whiskers. Based on Tukey (1977).

(Wilcox, 1998). The Theil-Sen approach, however, might become impracticable for large datasets.

Predict and prescribe

The next phase in data analytics commences after the data have been deemed appropriate and free of known incorrect or influential points. This phase encompasses gaining knowledge from the data through prediction (i.e., modeling and simulation) and prescription (Figure 1). The prediction process should answer the question: what could happen given the data and model in hand, whereas the prescription process is related to what to do next given the most accurate and precise predictions. Together, they will provide insightful information about the data and how it can be modeled for descriptive or forecasting purposes. There are several predictive analytics, including statistical models (i.e., empirical regressions), mechanistic modeling (**MM**), dynamic versus static models, and forecasting techniques, to list a few. On the other hand, prescriptive analytics include optimization, decision tree analysis, and simulation techniques (i.e., risk analysis) that predictive models usually assist.

Diverse predictive models, sometimes simple other times complex, have been developed in many fields of science, and

animal production has vastly benefited in the last 50 years though it could be considerably expanded (Tedeschi, 2019; White et al., 2018). Specific disciplines in animal production are more suitable for predictive modeling than others, but it depends on the discipline's flexibility and acceptability to recognize and incorporate modeling as a valuable tool for data analytics. The nutrition and metabolism domains, for instance, have benefited tremendously from predictive models (Baldwin, 1995; Tedeschi and Fox, 2020), in part because of the worldwide respect of publications by the National Academies of Sciences, Engineering, and Medicine (**NASEM**) through their National Research Council's (**NRC**) Nutrient Requirement series (NASEM, 2016; NASEM, 2021; NRC, 2007a; NRC, 2007b; NRC, 2012), and in part because of the industry's need to increase the standardization and quality of their animal products as well as profits for more competitive production scenarios. In that sense, knowing when feedlot animals will achieve their most profitable point, given their carcass composition and maturity degree (Tedeschi et al., 2004), requires accurate predictive models for animal growth (Anim-Jnr et al., 2020; Hoch and Agabriel, 2004; Pettigrew, 2018; Tedeschi et al., 2004). Epidemiological models have gained considerable interest given the increasing concerns of zoonotic diseases disrupting the animal production sector (Manjoo-Docrat, 2022; Wisnieski et al., 2021) and their possible impact on humans, including the most recent concern about antimicrobial resistance (Chantziaras et al., 2014; Spicknall et al., 2013).

Comprehensive discourses about developing and evaluating predictive models abound in the scientific literature (Burnham and Anderson, 2002; Deaton and Winebrake, 2000; Dym, 2004; France and Thornley, 1984; Haefner, 2005; Hannon and Ruth, 1997; Heinz, 2011; Kuhn and Johnson, 2013). A consensus about the most critical steps in predictive model development and evaluation among these publications include (1) defining the scope and purpose of the mathematical model, (2) establishing the physical or virtual boundaries of the problem, (3) identifying endogenous and exogenous variables to the problem and their relationships, (4) developing datasets for model development and model evaluation that are representative and independent of each other, and (5) re-engineering the mathematical model after the gaining-insight step is done. These critical steps require meticulous planning and diligent thinking, but data partitioning between the development and evaluation stages has often resulted in contentious debates when data are scarce or limited. Such problems might be minimized or wholly eliminated with big data, but, even in that case, it still has to be partitioned between training, revising, and evaluation datasets (Tedeschi, 2006; White et al., 2018). Bootstrap and cross-validation techniques (Efron and Tibshirani, 1998) have frequently been employed to split the data for development/evaluation or training/revising/testing schemes.

After a predictive model is developed, calibrated, and evaluated for specific production conditions, the analyst might be interested in finding the optimum solution given the resources in hand, such as diet formulation to maximize profit, the number of animals in pen to minimize disease transmission, the combination of breeds to maximize milk protein composition. The optimum solution is usually achieved through optimization and mathematical programming, a field of study belonging to a branch of mathematics called operations research that has progressed tremendously since the mid-1990s.

The optimum solution lies in finding the ideal combination of available resources to meet specific criteria (i.e., constraints) while minimizing (or maximizing) a function (i.e., objective function). Several mathematical programming and methods to solve optimization problems exist (Floudas and Pardalos, 2009), but the most commonly used are linear or nonlinear programming, multiobjective or fractional programming, and dynamic programming (Tedeschi and Boin, 2023). The literature about optimization and mathematical programming is vast (Dryden, 2008; Karloff, 2009; Luenberger and Ye, 2008; Saigal, 1995; Sniedovich, 2010; Vajda, 1981), and many applications in livestock production have been expounded (Tedeschi and Boin, 2023).

Smart learn and policy making

This phase comprises the last stage in the DIKW concept (Ackoff, 1989; Tedeschi, 2019). It aims to apply the information obtained in the previous phase to foment knowledge and use it for rational decisions. It involves managers integrating and applying the knowledge, incorporating field experiences (combinations of success and failures), learned process delays, and expected and unintentional outcomes typically created by feedback in complex dynamic systems. From a business perspective, profits are realized based on competitive market analytics and how effectively knowledge gained by individuals within an organization is integrated (Herden, 2020). From an animal production perspective, it encompasses the fine tuning of diet formulation given the target animal's performance after the supply of primary nutrients and the logistics of diet delivery are fulfilled; the strategic selection of sires and breeds given the composition of the herd and the production objectives; and the decision about herd size and allocation of grazing animals within pasture raising conditions to satisfy sustainability concerns over time to list a few examples. However, a sustainable competitive advantage only occurs after continuous use of technological systems and acquiring new knowledge through independent data collection and its use within the DIKW context. Knowledge management (and its application towards received wisdom) is critical to the success of scientific understanding, but determining the boundaries of how far to go to get them requires persistence and resilience (Grant, 1996; Rich and Duchessi, 2004). Understanding these boundaries calls for old scientific concepts and existing data sets to be re-generated for validation and confirmation purposes (of the concepts) and enrichment of scientific knowledge, i.e., additional independent data with *new* variables, and perhaps renewed or improved methodological techniques for data collection too. Certain modeling paradigms have been used alone (Gerrits et al., 2021; Nicholson et al., 2011; Stephens, 2021; Tedeschi et al., 2011) or in combination (Kim et al., 2019) to understand knowledge management, but, so far, few have combined AI with other modeling paradigms likely because it requires complicated and continuous interactions between the paradigms that are demanding computationally. Commonly applied modeling methodologies and paradigms include system dynamics (Forrester, 1961, 1971, 1973; Sterman, 2000), agent-based modeling (Grimm and Railsback, 2005; Railsback and Grimm, 2011), discrete models (Law, 2007), and stochastic models (Birge and Louveaux, 1997; Guttorp, 1995). When developing mechanistic models, combining two or more methodologies and paradigms is also possible; it depends on the scope and purpose of the model: whether to model the trees or the forest, but without losing sight of the forest for the trees, i.e., being too focused on details and missing the big picture. Although the user can gain insights into the behavior of the problem in question, these model paradigms per se do not use any learning algorithm, i.e., it does not create rules, it simply follows the rules embedded in them.

Although the improvements in MM and the development and advent of different methodologies (e.g., system dynamics, agent-based modeling) facilitate the understanding of feedback loops and agent interactions, our ability to improve the mechanistic model's predictability is limited to the inputs availability and *de novo* conceptualization of the intricate, existing relationships among inputs (endogenous and exogenous to the problem). There are other inherent limitations to mechanistic models that we may not overcome. In ruminant nutrition, such limitations include the dependency of degradation rates on the methodology (i.e., in situ or in vitro) and how much of the gastrointestinal tract recycled nitrogen is, in fact, reused by ruminal microbes in an upcycling manner through anabolism (Eisemann and Tedeschi, 2016). But, the passage rate is perhaps the most significant limiting factor in predicting nutrient digestibility in the rumen (Allen, 2019). The scientific community continues its never-ending pursuit of new technological options to solve persistent conundrums.

Systems based on AI technology have been developed and deployed in diverse agricultural entrepreneurship worldwide. For instance, management decisions of a dairy farm can be made based on daily information (e.g., milking parlor, sensors, weather, economics, crops, genetics, and feed management) gathered from similar, representative dairy farms in the region using AI and data visualization (Ferris et al., 2020). Other applications include predicting the onset of disease in pigs given their feeding behavior, precise irrigation given soil water or crop status, or optimum nutrient management in crops (Sudduth et al., 2020), and epidemiological models to detect emerging health issues (VanderWaal et al., 2017). Computer vision associated with AI algorithms (Prince, 2013) has gained tremendous attention in the past five years, given the accessibility to high-quality cameras and speedy data collection, storage, and processing using DL algorithms, mostly based on variants of the convolutional neural network (Bezen et al., 2020; Borges Oliveira et al., 2021; Saar et al., 2022; Wang et al., 2021). But, the question then becomes, what might be the next steps in AI modeling besides improving AI methodology and algorithms?

Hybrid Intelligent Mechanistic Models

Several studies have compared different predictive analytics with AI, and the results have been mixed. For instance, Alves et al. (2019) compared multiple linear regression with ML (i.e., support vector ML, Bayesian network) to predict carcass traits and commercial meat cuts in lambs and reported that both could be used to pre-select input variables for an ML approach. Perhaps, hybrid models (mechanistic and AI) might provide better forecasting, interpretation, and comprehension of the predictions as it combines the conceptual features of MM with the speedy AI's data handling attributes (Tedeschi, 2020). The missing link to foster the development of the next generation of computer modeling (Tedeschi and Menendez, 2020) that will spur an innovative technological wave in predictive analysis (Tedeschi, 2019) might be the combination of AI (a data-driven approach) with MM (a concept-driven

approach). Figure 4 depicts two approaches to combining AI with MM in developing HIMM. In the first approach, AI is embedded in the MM, and its primary purpose is to predict variables needed by the MM. Such variables can be user-inputted or predicted by AI. In this instance, AI is assisting the user to obtain variables that might be affected by multiple factors, and one can hardly guess its value without causing the type II error, i.e., accepting an incorrect value for the variable (Dean and Voss, 1999; Tedeschi, 2006), which has often been extrapolated to the type III error, i.e., using an irrelevant model (in our case variable value) and believe the outcome is true when, in fact, the model (or the variable) answers the wrong question or has the wrong value (Kimball, 1957; Kuhn and Johnson, 2013; Sokolowski and Banks, 2009). Examples for this case include the prediction of passage rate, which has been deemed a sensitive and hard-to-get variable in nutrition MM (Allen, 2019). The second approach, shown by the green arrow in Figure 4, has the MM embedded in the neural structure of the AI model. In this case, the MM is used to estimate an input to the AI model, but multiple MM variables could be linked into different parts of the AI structure. It is based on the fact that MM are developed based on underlying natural principles that govern the natural causes. Therefore, the MM prediction is expected to be based on solid ground with a strong bio-physicochemical foundation. Examples for this case include the prediction of degraded starch in the rumen by the mechanistic model and used to predict methane of an AI model.

Given the intrinsic dependency between AI and MM variables, solving HIMM might require an interactive approach
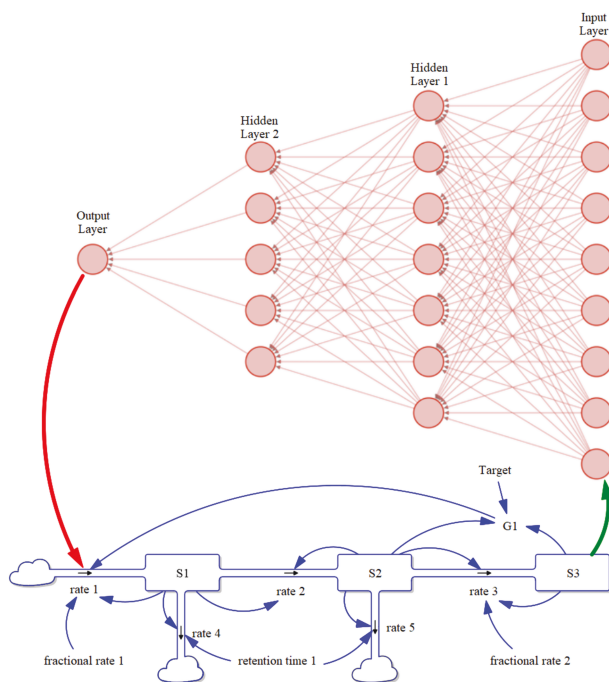
at different instances of the model until a stable solution is achieved. Furthermore, the development of HIMM may not increase our ability to understand the underlying mechanisms controlling the outcomes of a system or a problem. However, it may increase the predictive ability of existing MM by helping the analyst explain more of the data variation. It may also help to validate AI predictions when AI is employed alone. This is because AI depends heavily on the quality of the data used to train its structure; ill-conditioned data will result in biased AI predictions (Tedeschi et al., 2021). It has become customary to affirm that AI needs *big data* and that *big data* needs AI. This affirmative is not wrong; it is the basis of the existence of AI. However, the scientific community cannot use the failures of AI predictions on the lack of data, begging for more data. This mutual dependency on AI and big data and the constant criticism of AI failures due to the lack of data are not inconsequent, and it may lead to a death spiral that never reaches an end, possibly culminating with the demise of AI. It begs the question, how much more data is needed, and can it be obtained sustainably and in a timely fashion?

Mertoguno (2019) has compared sequential (stacked) and parallel (intertwined) constructions for merging AI (statistical learning) and MM (formal reasoning), using different methods such as Markov Learning Network, Bayesian Logic, and DL. For instance, removing either the red or the green arrow in Figure 4 would result in a sequential AI-MM HIMM, whereas keeping both arrows yields a parallel AI-MM HIMM. Mertoguno (2019) explored the *Learn2Reason* concept inspired by Kahneman (2013), in which the decision-making process requires cognitive (statistical learning, AI-based approach) and deliberative (formal reasoning, MM-based approach) as separate but constantly interacting entities. Similar to our expectation of HIMM described above, the synergism of integrating statistical learning (i.e., AI) with formal reasoning (i.e., MM) would enable cross-checking between these entities, allowing for a better understanding of the systems or problem in hand.

The notion of combining different paradigms (i.e., modeling methodologies) to solve problems is not new; it has been exploited in the past as a component of integrated systems. Hybrid intelligent systems have been developed using neural networks (data driven), expert systems (concept driven), fuzzy logic (association techniques to non-numeric variables), genetic algorithms (optimization), and case-based reasoning since AI was initially developed in the mid-1960s (Medsker, 1995). Case-based reasoning is an exciting element of hybrid intelligent systems that use past problems to solve new ones; it is like a database of historical problems and solutions (Medsker, 1995). Given the current processing speed of digital computers, other more demanding modeling methodologies (i.e., agent-based or individual-based models) might become more attractive than MM to be associated with AI technology.

Figure 4. A hypothetical sketch showing two approaches to hybridizing artificial intelligence (AI; red circles representing the nodes) and mechanistic models (MM; blue rectangles represent stock variables, and blue pipes represent flow variables). The red arrow illustrates the output of the AI being used as an input to the MM, and the green arrow illustrates the output of the MM being used as an input to the AI. The red and blue arrows do not necessarily co-exist in the same hybrid model.

Intelligent agent-based models

Intelligent agent-based models (**iABM**) are stochastic models (i.e., agent-based) with AI elements embedded in them. This type of modeling has been frequently used for finding solutions to a problem or a question (i.e., some sort of optimization). ABM comprises computational models that simulate the actions and interactions among unique and autonomous agents to understand the behavior and outcomes of a system or problems, using multiple agents that interact among

themselves within a specific environment (i.e., boundary) (Railsback and Grimm, 2011; Wilensky and Rand, 2015). Each agent follows a set of rules to make intrinsically stochastic decisions based on some elements of game theory. Bonabeau (2002) believes that ABM is more a mindset used to describe a system based on the interactions of its elements rather than a technology. However, the essence of all models is to represent an analyst's perception of real-life mathematics. The use of ABM to assess the impacts of climate change on species adaptation is growing among ecologists and biologists interested in conservation and management purposes, given their geographical distribution and persistence (Bioco et al., 2022).

Brearcliffe and Crooks (2021) include ML techniques (Evolutionary Computing, Q Learning, and State-Action-Reward-State-Action) into the ABM called *Sugarscape*, an artificial world game of 51 × 51 cell grid containing a renewable resource (i.e., sugar) for which agents can capture and metabolize. As a result, they can pollute, die, reproduce, inherit sources, transfer information, trade or borrow sugar, generate immunity, or transmit diseases (Epstein and Axtell, 1996). Simulations of the intelligent ABM model by Brearcliffe and Crooks (2021) (https://tinyurl.com/ML-Agents) suggested that ML methods can be integrated into ABM, but ML may not always yield the best results. Animal scientists are yet to adopt ABM to understand the grazing behavior of herbivores and how climate change can alter it, and iABM might be a more suitable approach.

Perhaps the widespread use of web-based ABM associated with AI (i.e., iABM) might expedite its adoption for a more holistic, inclusive approach to elucidate animal-plant-soil relationships within different ecosystems. However, it is not entirely clear whether iABM is developed to improve the predictability of animal impact on the environment or improve livestock feeding and management strategies to be more sustainable, or both. Bonabeau (2002) indicated that one of the benefits of ABM over other modeling paradigms is its ability to capture emergent phenomena besides being flexible and having a better interface with the nature of the systems (i.e., environment, agents, rules).

Final Remarks

There is no question that data analytics have evolved tremendously, and in some instances, the scientific community has not yet fully grasped the power (and limitations) of some tools. Given the speedy broadcast of AI and exponential interest by the scientific community in how to use AI in their specific field of study, many data-driven models (i.e., AI) have been created, but the lack of transparency and adequate reporting might have limited their reproducibility (Hutson, 2018). There is a chronic lack of open-science and open-data practices (Crüwell et al., 2019; Muñoz-Tamayo et al., 2022) that prevents widespread knowledge in agricultural sciences (Janssen et al., 2017). Thus, further development and adoption of AI might be limited to regionalized pockets and specific communities, preventing the dissemination of knowledge and impairing its reproducibility. Best practices for reporting AI research exist and should be followed (Artrith et al., 2021; Heil et al., 2021; Mateen et al., 2020; Norgeot et al., 2020).

The relatively recent development and employment of AI tools in agricultural sciences have become *en vogue*, bringing about the shiny object syndrome (**SOS**). The SOS provokes distraction from the bigger picture, causing agents to go off on tangents, searching for the most "flashy" technology rather than focusing on ready-to-be-used innovations and techniques already far down the pipeline that can provide authentic solutions to current problems. The SOS results in some attention deficit disorder at the organizational level because technologists cannot maintain a consistent direction of their perceived (desired) mission (Church et al., 2017; Roberts, 2011). In some social sciences, the recommendation has been to stay away from these shiny new objects in practice (Church and Silzer, 2016) until their practical efficiency and efficacy are proven. Thus, it seems prudent that the animal science community avoid the SOS by seeing beyond the frontiers of current knowledge to understand and control the mechanisms that govern (and limit) natural processes while improving productivity and efficiency to stewardship agriculture towards sustainable and responsible achievements. When associated with the perception that AI is today's fashionable neural network and that optimism exist regarding AI real functionality, the question becomes who is working for who? Are we developing AI methods that will benefit humankind by improving livelihoods, or will humanity work for AI by forever collecting data needed to improve its perceived predictability? The concept of AI is very powerful, but it is still under development. It is not the right time to abandon other modeling techniques yet; we still have a lot to learn from these different paradigms and how to integrate them.

Acknowledgments

Conflict of Interest Statement

The authors declare no real or perceived conflicts of interest.

Literature Cited

Ackoff, R. L. 1989. From data to wisdom. *J. Appl. Syst. Anal*. 16:3–9.

Allen, M. S. 2019. Do more mechanistic models increase accuracy of prediction of metabolisable protein supply in ruminants? *Anim. Prod. Sci*. 59(11):1991–1998. doi:10.1071/AN19337

Alves, A. A. C., A. Chaparro Pinzon, R. M. d. Costa, M. S. d. Silva, E. H. M. Vieira, I. B. d. Mendonça, V. d. S. S. Viana, R. N. B. Lôbo. 2019. Multiple regression and machine learning based methods for carcass traits and saleable meat cuts prediction using non-invasive in vivo measurements in commercial lambs. *Small Ruminant Res*. 171:49–56. doi:10.1016/j.smallrumres.2018.12.008

Andrews, D. F. 1974. A robust method for multiple linear regression. *Technometrics* 16:523–531. doi:10.1080/00401706.1974.10489 233

Andrews, D. F., P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey. 1972. *Robust estimates of location: survey and advances*. Princeton, NJ: Princeton University Press.

Anim-Jnr, S., P. C. H. Morel, P. R. Kenyon, and H. T. Blair. 2020. A mechanistic dynamic simulation model of nutrient utilization, growth and body composition in pre-weaned lambs reared artificially. *Anim. Feed Sci. Technol.* 261:114402. doi:10.1016/j.anifeedsci.2020.114402

Artrith, N., K. T. Butler, F. -X. Coudert, S. Han, O. Isayev, A. Jain, and A. Walsh. 2021. Best practices in machine learning for chemistry. *Nat. Chem.* 13:505–508. doi:10.1038/s41557-021-00716-z

Baldwin, R. L. 1995. *Modeling ruminant digestion and metabolism*. New York, NY: Chapman & Hall

Bezen, R., Y. Edan, and I. Halachmi. 2020. Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms. *Comput. Electron. Agric.* 172:105345. doi:10.1016/j.compag.2020.105345

Bioco, J., F. Cánovas, P. Prata, and P. Fazendeiro. 2022. SDSim: a generalized user friendly web ABM system to simulate spatiotemporal distribution of species under environmental scenarios. *Environ. Modell. Softw.* 147:105234. doi:10.1016/j.envsoft.2021.105234

Birge, J. R., and F. Louveaux. 1997. *Introduction to stochastic programming*. New York, NY: Springer

Bonabeau, E. 2002. Agent-based modeling: methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. USA* 99(Suppl. 3):7280. doi:10.1073/pnas.082080899

Borges Oliveira, D. A., L. G. Ribeiro Pereira, T. Bresolin, R. E. Pontes Ferreira, and J. R. Reboucas Dorea. 2021. A review of deep learning algorithms for computer vision systems in livestock. *Livest. Sci* 253:104700. doi:10.1016/j.livsci.2021.104700

Boston, R. C., P. A. Wilkins, and L. O. Tedeschi. 2007. *Identifiability and accuracy: two critical problems associated with the application of models in nutrition and the health sciences. Pages 161-193 in Mathematical modeling for nutrition and health sciences*. M. Hanigan, ed., VA: Roanoke; p 31-58

Brearcliffe, D. K., and A. T. Crooks. 2021. Creating intelligent agents: Combining agent-based modeling with machine learning. in Proceedings of the 2020 Conference of The Computational Social Science Society of the Americas. Springer Proceedings in Complexity. Z. Yang and E. von Briesen, editors. Cham: Springer. doi:10.1007/978-3-030-83418-0_3

Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed*. New York, NY: Springer-Verlag.

Chantziaras, I., F. Boyen, B. Callens, and J. Dewulf. 2014. Correlation between veterinary antimicrobial use and antimicrobial resistance in food-producing animals: a report on seven countries. *J. Antimicrob. Chemother.* 69:827–834. doi:10.1093/jac/dkt443

Church, A. H., M. Del Giudice, and A. Margulies. 2017. All that glitters is not gold: maximizing the impact of executive assessment and development efforts. *Leadership Organiz Dev J* 38:765–779. doi:10.1108/LODJ-05-2016-0127

Church, A. H., and R. Silzer. 2016. Are we on the same wavelength? Four steps for moving from talent signals to valid talent management applications. *Ind. Organ. Psychol.* 9:645–654. doi:10.1017/iop.2016.65

Cochran, W. G., and G. M. Cox. 1957. *Experimental design*. New York, NY: John Wiley & Sons

Crüwell, S., J. Doorn, A. Etz, M. C. Makel, H. Moshontz, J. C. Niebaum, A. Orben, S. Parsons, and M. Schulte-Mecklenbeck. 2019. Seven easy steps to open science. *Z. Psychol.* 227:237–248. doi:10.1027/2151-2604/a000387

Davenport, T. H., and J. G. Harris. 2017. Competing on analytics: the new science of winning. Boston, MA: Harvard Business Review Press.

Dean, A., and D. Voss. 1999. *Design and analysis of experiments*. Springer texts in statistics. New York, NY: Springer

Deaton, M. L., and J. I. Winebrake. 2000. *Dynamic modeling of environmental systems. Modeling dynamic systems*. New York, NY: Springer

Dryden, G. M. 2008. *Animal nutrition science*. Wallingford, UK: CABI.

Dym, C. L. 2004. *Principles of mathematical modeling. (2nd ed.)*. Amsterdam: Elsevier/Academic Press

Efron, B., and R. J. Tibshirani. 1998. *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC

Eisemann, J. H., and L. O. Tedeschi. 2016. Predicting the amount of urea nitrogen recycled and used for anabolism in growing cattle. *J. Agric. Sci* 154:1118–1129. doi:10.1017/S0021859616000228

Epstein, J. M., and R. L. Axtell. 1996. *Growing artificial societies; social science from the bottom up*. Washington, DC: Brookings Institution Press

Ferris, M. C., A. Christensen, and S. R. Wangen. 2020. *Symposium review*: Dairy Brain---informing decisions on dairy farms using data analytics. *J. Dairy Sci.* 103:3874–3881. doi:10.3168/jds.2019-17199

Floudas, C. A., and P. M. Pardalos. 2009. *Encyclopedia of optimization. 2nd ed*. Boston, MA: Springer. doi:10.1007/978-0-387-74759-0

Forrester, J. W. 1961. *Industrial dynamics*. Cambridge, MA: MIT Press

Forrester, J. W. 1971. *Principles of systems*. Cambridge, MA: Wright-Allen Press

Forrester, J. W. 1973. *World dynamics*. Cambridge, MA: Wright-Allen Press

France, J., and J. H. M. Thornley. 1984. *Mathematical models in agriculture: a quantitative approach to problems in agriculture and related sciences*. London, UK: Butterworths

Friedman, J. H., and W. Stuetzle. 2002. John W. Tukey's work on interactive graphics. *Ann. Statist.* 30 (6):1629–1639. doi:10.1214/aos/1043351250

Gerrits, W., M. Schop, S. de Vries, and J. Dijkstra. 2021. ASAS-NANP Symposium: digestion kinetics in pigs: the next step in feed evaluation and a ready-to-use modeling exercise. *J. Anim. Sci.* 99. doi:10.1093/jas/skab020

Gilbert, R. O. 1987. *Statistical methods for environmental pollution monitoring*. New York, NY: Van Nostrand Rainhold Company.

Godfrey, K. R., and J. J. DiStefano, III. 1985. Identifiability of model parameters. IFAC Proceedings Volumes. 18 (5):89–114. doi:10.1016/S1474-6670(17)60544-5

Grant, R. M. 1996. Toward a knowledge-based theory of the firm. *Strateg. Manage. J.* 17(S2):109–122. doi:10.1002/smj.4250171110

Grimm, V., and S. F. Railsback. 2005. *Individual-based modeling and ecology*. Princeton: Princeton University Press

Guttorp, P. 1995. *Stochastic modeling of scientific data*. New York, NY: Chapman & Hall

Haefner, J. W. 2005. *Modeling biological systems*. Principles and applications. 2nd ed. New York, NY: Springer

Hannon, B., and M. Ruth. 1997. *Modeling dynamic biological systems*. Modeling dynamic systems. New York, NY: Springer

Heil, B. J., M. M. Hoffman, F. Markowetz, S. -I. Lee, C. S. Greene, and S. C. Hicks. 2021. Reproducibility standards for machine learning in the life sciences. *Nat. Methods* 18:1132–1135. doi:10.1038/s41592-021-01256-7

Heinz, S. 2011. *Mathematical modeling*. New York, NY: Springer

Herden, T. T. 2020. Explaining the competitive advantage generated from analytics with the knowledge-based view: the example of Logistics and Supply Chain Management. *Business Res* 13:163–214. doi:10.1007/s40685-019-00104-x

Hoch, T., and J. Agabriel. 2004. A mechanistic dynamic model to estimate beef cattle growth and body composition: 1. Model description. *Agric. Syst.* 81:1–15. doi:10.1016/j.agsy.2003.08.005

Holsapple, C., A. Lee-Post, and R. Pakath. 2014. A unified foundation for business analytics. *Decis. Support Syst.* 64:130–141. doi:10.1016/j.dss.2014.05.013

Huber, P. J., and E. M. Ronchetti. 2009. *Robust statistics. 2nd ed*. New York, NY: John Wiley & Sons.

Hubert, M., and M. Debruyne. 2009. Breakdown value. *WIREs Comput. Stat*. 1:296–302. doi:10.1002/wics.34

Hutson, M. 2018. Artificial intelligence faces reproducibility crisis. *Science* 359:725–726. doi:10.1126/science.359.6377.725

Janssen, S. J. C., C. H. Porter, A. D. Moore, I. N. Athanasiadis, I. Foster, J. W. Jones, and J. M. Antle. 2017. Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. *Agric. Syst*. 155:200–212. doi:10.1016/j.agsy.2016.09.017

Kahneman, D. 2013. *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Karloff, H. 2009. Linear programming. Birkhauser. *Boston* doi:10.1007/978-0-8176-4844-2

Kim, Y., J. Son, Y. -S. Lee, M. Lee, J. Hong, and K. Cho. 2019. Integration of an individual-oriented model into a system dynamics model: an application to a multi-species system. *Environ. Modell. Softw*. 112:23–35. doi:10.1016/j.envsoft.2018.11.009

Kimball, A. W. 1957. Errors of the third kind in statistical consulting. *J. Am. Stat. Assoc*. 52:133–142. doi:10.1080/01621459.1957.10501374

Kuhn, M., and K. Johnson. 2013. *Applied predictive modeling*. New York, NY: Springer. Available at: http://appliedpredictivemodeling.com/

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2005. *Applied linear statistical models. 5th ed*. New York, NY: McGraw-Hill Irwin

Law, A. M. 2007. *Simulation modeling and analysis. 4th ed. Industrial engineering and management science*. Boston: McGraw-Hill

Luenberger, D. G., and Y. Ye. 2008. *Linear and nonlinear programming. 3rd ed*. Cham, Switzerland: Springer. doi:10.1007/978-3-319-18842-3

Manjoo-Docrat, R. 2022. A spatio-stochastic model for the spread of infectious diseases. *J. Theor. Biol*. 533:110943. doi:10.1016/j.jtbi.2021.110943

Mateen, B. A., J. Liley, A. K. Denniston, C. C. Holmes, and S. J. Vollmer. 2020. Improving the quality of machine learning in health applications and clinical research. *Nat. Mach. Intell*. 2:554–556. doi:10.1038/s42256-020-00239-1

Medsker, L. R. 1995. *Hybrid intelligent systems*. Boston, MA: Springer doi:10.1007/978-1-4615-2353-6

Mertoguno, J. S. 2019. Toward autonomy: symbiotic formal and statistical machine reasoning. Pages 210-215 In: 2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI). Los Angeles, CA. Institute of Electrical and Electronics Engineers. Available at: https://ieeexplore.ieee.org/abstract/document/8998973. [accessed February 13, 2021]. doi:10.1109/CogMI48466.2019.00038

Michael, W. B., H. F. Kaiser, and C. A. Clark. 1957. Chapter VIII: research tools: statistical methods. *Rev. Educ. Res*. 27:498–527. doi:10.3102/00346543027005498

Morota, G., H. Cheng, D. Cook, and E. Tanaka. 2021. ASAS-NANP Symposium: prospects for interactive and dynamic graphics in the era of data-rich animal science. *J. Anim. Sci*. 99. doi:10.1093/jas/skaa402

Muñoz-Tamayo, R., B. L. Nielsen, M. Gagaoua, F. Gondret, E. T. Krause, D. P. Morgavi, I. A. S. Olsson, M. Pastell, M. Taghipoor, L. Tedeschi, I. Veissier, C. Nawroth. 2022. Seven steps to enhance open science practices in animal science. *Zenodo* doi:10.5281/zenodo.5891771

National Academies of Sciences, Engineering, and Medicine. 2016. *Nutrient requirements of beef cattle. 8th ed*. Animal nutrition series. Washington, DC: National Academy Press. doi:10.17226/19014

National Academies of Sciences, Engineering, and Medicine. 2021. *Nutrient requirements of dairy cattle. 8th ed.*. Animal nutrition series. Washington, DC: National Academy Press. doi:10.17226/25806

National Research Council. 2007a. *Nutrient requirements of horses. 6th ed.*. Animal nutrition series. Washington, DC: National Academies Press. Available at: https://www.nap.edu/catalog/11653/nutrient-requirements-of-horses-sixth-revised-edition. doi:10.17226/11653

National Research Council. 2007b. *Nutrient requirements of small ruminants: sheep, goats, cervids, and New World Camelids. 7th ed*. Animal nutrition series. Washington, DC: National Academy Press. doi:10.17226/11654

National Research Council. 2012. *Nutrient requirements of swine. 11th ed.*. Animal Nutrition Series. Washington, DC: National Academy Press. doi:10.17226/13298

Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. 1996. *Applied linear statistical models. 4th ed*. Boston: McGraw-Hill Publishing.

Newman, S. T., Lynch A., A. Plummer. 2000. Success and failure of decision support systems: learning as we go. *J. Anim. Sci*. 77(E-Suppl):1–12. doi:10.2527/jas2000.77E-Suppl1e

Nicholson, C. F., L. O. Tedeschi, and A. C. F. Lellis Vieira. 2011. The application of system dynamics modeling to enhance profitability and sustainability in Latin American livestock systems. in Genómica Y *Modelación en los Nuevos Escenarios de la Ganaderia Bovina Tropical, 2o. Simposio Internacional. C. V. Durán-Castro and R. Campos-Gaona*, editors. Palmira, Colombia: Universidad Nacional de Colombia; p 132–162

Norgeot, B., G. Quer, B. K. Beaulieu-Jones, A. Torkamani, R. Dias, M. Gianfrancesco, R. Arnaout, I. S. Kohane, S. Saria, E. Topol, et al. 2020. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat. Med*. 26:1320–1324. doi:10.1038/s41591-020-1041-y

Pettigrew, J. E. 2018. Essential role for simulation models in animal research and application. *Anim. Prod. Sci*. 58:704–708. doi:10.1071/AN15794

Power, D. J. 2008. Decision support systems: a historical overview. In: F. Burstein and C. W. Holsapple, editors. Pages 121-140 in Handbook on decision support systems 1: basic themes. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-540-48713-5_7

Prince, S. J. D. 2013. *Computer visions; models, learning, and inference*. London, UK: Cambridge University Press

Railsback, S. F., and V. Grimm. 2011. *Agent-based and individual-based modeling: a practical introduction*. New Jersey: Princeton University Press.

Rich, E., and P. Duchessi. 2004. Modeling the sustainability of knowledge management programs. Page 9 in Proceedings of the Hawaii International Conference on System Sciences. Big Island, Hawaii:

Roberts, J. A. 2011. *Shiny objects: why we spend money we don't have in search of happiness we can't buy*. New York, NY: HarperCollins.

Rosner, B. 1975. On the detection of many outliers. *Technometrics* 17:221–227. doi:10.1080/00401706.1975.10489305

Rosner, B. 1983. Percentage points for a generalized ESD many-outlier procedure. *Technometrics* 25:165–172. doi:10.1080/00401706.1983.10487848

Saar, M., Y. Edan, A. Godo, J. Lepar, Y. Parmet, and I. Halachmi. 2022. A machine vision system to predict individual cow feed intake of different feeds in a cowshed. *Animal* 16:100432. doi:10.1016/j.animal.2021.100432

Saigal, R. 1995. *Linear programming: a modern integrated analysis. Operations Research & Management Science*. Boston, MA: Springer. doi:10.1007/978-1-4615-2311-6

Sen, P. K. 1968. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc*. 63:1379–1389. doi:10.1080/01621459.1968.10480934

Siegel, A. F. 1982. Robust regression using repeated medians. *Biometrika* 69:242–244. doi:10.1093/biomet/69.1.242

Snedecor, G. W., and W. G. Cochran. 1971. *Statistical methods. 6th ed.*. Ames: Iowa State University Press

Sniedovich, M. 2010. *Dynamic programming; foundations and principles. 2nd ed.*. Boca Raton, FL: CRC Press. doi:10.1201/EBK0824740993

Sokolowski, J. A., and C. M. Banks. 2009. *Principles of modeling and simulation*. New York, NY: John Wiley & Sons, Inc.

Spicknall, I. H., B. Foxman, C. F. Marrs, and J. N. S. Eisenberg. 2013. A modeling framework for the evolution and spread of antibiotic resistance: literature review and model categorization. *Am. J. Epidemiol*. 178:508–520. doi:10.1093/aje/kwt017

Stanley, J. C. 1957. Chapter III: research methods: experimental design. *Rev. Educ. Res*. 27:449–459. doi:10.3102/00346543027005449

Steel, R. G. D., and J. H. Torrie. 1960. *Principles and procedures of statistics*. New York, NY: McGraw-Hill

Stephens, E. C. 2021. ASAS-NANP Symposium: review of systems thinking concepts and their potential value in animal science research J. *Anim. Sci*. 99. doi:10.1093/jas/skab021

Sterman, J. D. 2000. *Business dynamics: systems thinking and modeling for a complex world*. New York, NY: Irwin McGraw-Hill

Sudduth, K. A., M. Woodward-Greene, B. W. Penning, M. A. Locke, A. R. Rivers, and K. S. Veum. 2020. AI down on the farm. *IT Prof*. 22:22–26. doi:10.1109/MITP.2020.2986104

Tedeschi, L. O. 2006. Assessment of the adequacy of mathematical models. *Agric. Syst*. 89:225–247. doi:10.1016/j.agsy.2005.11.004

Tedeschi, L. O. 2019. ASN-ASAS Symposium: future of data analytics in nutrition: mathematical modeling in ruminant nutrition: approaches and paradigms, extant models, and thoughts for upcoming predictive analytics. *J. Anim. Sci*. 97:1321–1944. doi:10.1093/jas/skz092

Tedeschi, L. O. 2020. The need for hybrid neural network models in precision livestock farming systems. In: Proceedings of the 71st Annual Meeting of the European Federation of Animal Science. Virtual Meeting. European Federation of Animal Science (EAX, EVT, FEZ, EAAP, EUROPA). p. 111

Tedeschi, L. O., and C. Boin. 2023. Principles of precision feed evaluation and diet formulation for ruminants. In: J. Brameld, A. Athanasiadou, D. Kenny and S. Lanham-New, editors. Animal nutrition. New York, NY: Wiley Blackwell; p. (in print).

Tedeschi, L. O., and R. C. Boston. 2010. Identifiability and accuracy: a closer look at contemporary contributions and changes in these vital areas of mathematical modelling. In: D. Sauvant, J. Van Milgen, P. Faverdin and N. Friggens, editors. Modelling nutrient digestion and utilization in farm animals. Paris, France: University of Wageningen Press; vol. 7th; p. 91–99

Tedeschi, L. O., L. F. L. Cavalcanti, M. A. Fonseca, M. Herrero, and P. K. Thornton. 2014. The evolution and evaluation of dairy cattle models for predicting milk production: an agricultural model intercomparison and improvement project (AgMIP) for livestock. *Anim. Prod. Sci*. 54:2052–2067. doi:10.1071/AN14620

Tedeschi, L. O., and D. G. Fox. 2020. *The ruminant nutrition system: volume I - an applied model for predicting nutrient requirements and feed utilization in ruminants*. *3rd ed*. Ann Arbor, MI: XanEdu

Tedeschi, L. O., D. G. Fox, and P. J. Guiroy. 2004. A decision support system to improve individual cattle management. 1. A mechanistic, dynamic model for animal growth. *Agric. Syst*. 79:171–204. doi:10.1016/S0308-521X(03)00070-2

Tedeschi, L. O., P. L. Greenwood, and I. Halachmi. 2021. Advancements in sensor technology and decision support intelligent tools to assist smart livestock farming. *J. Anim. Sci*. 99. doi:10.1093/jas/skab038

Tedeschi, L. O., and H. M. Menendez, III. 2020. Mathematical modeling in animal production. In: F. W. Bazer, G. C. Lamb, and G. Wu, editors. Cambridge, MA: Academic Press by Elsevier; p. 431–453. doi:10.1016/B978-0-12-817052-6.00025-2

Tedeschi, L. O., C. F. Nicholson, and E. Rich. 2011. Using system dynamics modelling approach to develop management tools for animal production with emphasis on small ruminants. *Small Ruminant Res*. 98:102–110. doi:10.1016/j.smallrumres.2011.03.026

Theil, H. 1950a. A rank-invariant method of linear and polynomial regression analysis. Part I. In: Proceedings of the Royal Netherlands Academy of Sciences, vol. 53: p. 386–392

Theil, H. 1950b. A rank-invariant method of linear and polynomial regression analysis. Part II. In: Proceedings of the Royal Netherlands Academy of Sciences, vol. 53: p. 521–525

Theil, H. 1950c. A rank-invariant method of linear and polynomial regression analysis. Part III. In: Proceedings of the Royal Netherlands Academy of Sciences, vol. 53: p. 1397–1412

Theil, H. 1992. A rank-invariant method of linear and polynomial regression analysis. in B. Raj and J. Koerts, editors. Henri Theil's contributions to economics and econometrics: econometric theory and methodology. Dordrecht: Springer, Netherlands; p. 345–381 doi:10.1007/978-94-011-2546-8_20

Tiku, M. L., and A. D. Akkaya. 2004. *Robust estimation and hypothesis testing*. Daryaganj, India: New Age International Ltd

Tukey, J. W. 1977. *Exploratory data analysis*. Reading, MA: Addison-Wesley Publishing Company

Vajda, S. 1981. *Linear programming; algorithms and applications*. London, UK: Chapman and Hall. doi:10.1007/978-94-011-6924-0

VanderWaal, K., R. B. Morrison, C. Neuhauser, C. Vilalta, and A. M. Perez. 2017. Translating big data into smart data for veterinary epidemiology. *Front. Vet. Sci* 4:110–110. doi:10.3389/fvets.2017.00110

Wald, A. 1940. The fitting of straight lines if both variables are subject to error. *Ann Math Stat* 11(3):284–300. doi:10.1214/aoms/1177731868

Walters, E. J., C. H. Morrell, R. E. Auer. 2006. An investigation of the median-median method of linear regression. *J Stat Educ*. 14 (2):null-null. doi:10.1080/10691898.2006.11910582

Wang, Z., S. Shadpour, E. Chan, V. Rotondo, K. Wood, and D. Tulpan. 2021. ASAS-NANP Symposium: applications of machine learning for livestock body weight prediction from digital images. *J. Anim. Sci*. 99. doi:10.1093/jas/skab022

Weissgerber, T. L., M. Savic, S. J. Winham, D. Stanisavljevic, V. D. Garovic, and N. M. Milic. 2017. Data visualization, bar naked: a free tool for creating interactive graphics. *J. Biol. Chem*. 292:20592–20598. doi:10.1074/jbc.RA117.000147

White, B. J., D. E. Amrine, and R. L. Larson. 2018. Big data analytics and precision animal agriculture symposium: data to decisions. *J. Anim. Sci*. 96:1531–1539. doi:10.1093/jas/skx065

Wilcox, R. 1998. A note on the Theil-Sen regression estimator when the regressor is random and the error term is heteroscedastic. *Biometrical J*. 40:261–268. doi:10.1002/(SICI)1521-4036(199807)40:3<261::AID-BIMJ261>3.0.CO;2-V

Wilcox, R. R. 1996. Estimation in the simple linear regression model when there is heteroscedasticity of unknown form. *Commun. Stat. Theory Methods* 25:1305–1324. doi:10.1080/03610929608831766

Wilcox, R. R. 2012. *Introduction to robust estimation and hypothesis testing. 3 ed*. Amsterdam, The Netherlands: Elsevier

Wilensky, U., and W. Rand. 2015. An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with Netlogo. Cambridge, MA: MIT Press

Wisnieski, L., D. E. Amrine, and D. G. Renter. 2021. Predictive modeling of bovine respiratory disease outcomes in feedlot cattle: a narrative review. *Livest. Sci* 251:104666. doi:10.1016/j.livsci.2021.104666