NONLINEAR REGRESSION MODELING FOR ENGINEERING APPLICATIONS

Wiley-ASME Press Series List

Introduction to Dynamics and Control of Mechanical Engineering Systems	To	March 2016
Fundamentals of Mechanical Vibrations	Cai	May 2016
Nonlinear Regression Modeling for Engineering Applications	Rhinehart	August 2016
Stress in ASME Pressure Vessels	Jawad	November 2016
Bioprocessing Piping and Equipment Design	Huitt	November 2016
Combined Cooling, Heating, and Power Systems	Shi	January 2017

NONLINEAR REGRESSION MODELING FOR ENGINEERING APPLICATIONS

MODELING, MODEL VALIDATION, AND ENABLING DESIGN OF EXPERIMENTS

R. Russell Rhinehart

WILEY

ASME PRESS

This edition first published 2016

First Edition published in 2016

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Names: Rhinehart, R. Russell, 1946- author.
Title: Nonlinear regression modeling for engineering applications : modeling, model validation, and enabling design of experiments / R. Russell Rhinehart.
Description: Chichester, UK ; Hoboken, NJ : John Wiley & Sons, 2016. | Includes bibliographical references and index.
Identifiers: LCCN 2016012932 (print) | LCCN 2016020558 (ebook) | ISBN 9781118597965 (cloth) | ISBN 9781118597934 (pdf) | ISBN 9781118597958 (epub)
Subjects: LCSH: Regression analysis–Mathematical models. | Engineering–Mathematical models.
Classification: LCC TA342 .R495 2016 (print) | LCC TA342 (ebook) | DDC 620.001/519536–dc23
LC record available at https://lccn.loc.gov/2016012932

A catalogue record for this book is available from the British Library.

Set in 10/12pt, TimesLTStd by SPi Global, Chennai, India.

1 2016

Contents

Part III REGRESSION, VALIDATION, DESIGN

Series Preface

The Wiley-ASME Press Series in Mechanical Engineering brings together two established leaders in mechanical engineering publishing to deliver high-quality, peer-reviewed books covering topics of current interest to engineers and researchers worldwide.

The series publishes across the breadth of mechanical engineering, comprising research, design and development, and manufacturing. It includes monographs, references and course texts.

Prospective topics include emerging and advanced technologies in Engineering Design; Computer-Aided Design; Energy Conversion & Resources; Heat Transfer; Manufacturing & Processing; Systems & Devices; Renewable Energy; Robotics; and Biotechnology.

Preface

Utility

Mathematical models are important.

Engineers use mathematical models to describe the natural world and then rearrange the model equations to answer the question, "How do I create an environment that makes Nature behave the way I want it to?" The answer to the mathematical rearrangement of the model equations reveals how to design processes, products, and procedures. It also reveals how to operate, use, monitor, and control them. Modeling is a critical underpinning for engineering analysis, design, control, and system optimization.

Further, since mathematical models express our understanding of how Nature behaves, we use them to validate our understanding of the fundamentals about processes and products. We postulate a mechanism and then derive a model grounded in that mechanistic understanding. If the model does not fit the data, our understanding of the mechanism was wrong or incomplete. Alternately, if the model fits the data we can claim our understanding may be correct. Models help us develop knowledge.

These models usually have coefficients representing some property of Nature, which has an unknown value (e.g., the diffusivity of a new molecule in a new medium, drag coefficient on a new shape, curing time of a new concrete mix, a catalyst effective surface area per unit mass, a heat transfer fouling factor). Model coefficient values must be adjusted to make the model match the experimentally obtained data, and obtaining the value of the coefficient adds to knowledge.

The procedure for finding the model coefficient values that makes a model best fit the data is called regression.

Although regression is ages old, there seem to be many opportunities for improvements related to finding a global optimum; finding a universal, effective, simple, and single stopping criterion for nonlinear regression; validating the model; balancing model simplicity and sufficiency with perfection and complexity; discriminating between competing models; and distinguishing functional sufficiency from prediction accuracy.

I developed and used process and product models throughout my 13-year industrial career. However, my college preparation for the engineering career did not teach me what I needed to know about how to create and evaluate models. I recognized that my fellow engineers, regardless of their *alma mater*, were also underprepared. We had to self-learn as to what was needed. Recognizing the centrality of modeling to engineering analysis, I have continued to explore model development and use during my subsequent academic career.

This textbook addresses nonlinear regression from a perspective that balances engineering utility with scientific perfection, a view that is often missing in the classroom, wherein the focus is often on the mathematical analysis, which pretends that there are simple, first-attempt solutions. Mathematical analysis is intellectually stimulating and satisfying, and sometimes useful for the practitioner. Where I think it adds value, I included analysis in this book. However, development of a model, choosing appropriate regression features, and designing experiments to generate useful data are iterative procedures that are guided by insight from progressive experience. It would be a rare event to jump to the right answers on the first try. Accordingly, balancing theoretical analysis, this book provides guides for procedure improvement.

This work is a collection of what I consider to be best practices in nonlinear regression modeling, which necessarily includes guides to design experiments to generate the data and guides to interpret the models. Undoubtedly, my view of best has been shaped with my particular uses for the models within the context of process and product modeling. Accordingly, this textbook has a focus on models with continuous-valued variables (either deterministic, discretized, or probabilities) as opposed to rank or classification, nonlinear as opposed to linear, constrained as opposed to not, and of a modest number of variables as opposed to Big Data.

This textbook includes the material I wish I had known when starting my engineering career and now what I would like my students to know. I hope it is useful for you.

The examples and discussion presume basic understanding of engineering models, regression, statistics, optimization, and calculus. This textbook provides enough details, explicit equation derivations, and examples to be useful as an introductory learning device for an upper-level undergraduate or graduate. I have used much of this material in the undergraduate unit operations lab course, in my explorations of model-based control on pilot-scale units, and in modeling of diverse processes (including the financial aspects of my retirement and the use of academic performance in the first two college years to project upper-level success). A person with an engineering degree and some experience with regression should be able to follow the concepts, analysis, and discussion.

My objective is to help you answer these questions:

- How to choose model inputs (variables, delays)?
- How to choose model form (linear, quadratic, or higher order, or equivalent model structures or architectures such as dimension or number of neurons)?
- How to design experiments to obtain adequate data (in number, precision, and placement) for determining model coefficient values?
- What to use for the regression objective (vertical least squares, total least squares, or maximum likelihood)?
- How to define goodness of model (r-square, fitness for use, utility, simplicity, data-based validation, confidence interval for prediction)?
- How to choose the right model between two different models?
- What optimization algorithm should be used for the regression to be able to handle the confounding issues of hard or soft constraints, discontinuities, discrete and continuous variables, multiple optima, and so on?
- What convergence criteria should be used to stop the optimizer (to recognize when it is close enough to optimum)?
- Should you linearize and use linear regression or use nonlinear regression?

- How to recognize outliers?
- How can you claim that a model properly captures some natural phenomena?

The underlying techniques needed for the answers include propagation of uncertainty, probability and statistics, optimization, and experience and heuristics. The initial chapters review/develop the basics. Subsequent chapters provide the application techniques, description of the algorithms, and guides for application.

Access to Computer Code

Those interested can visit the author's web site, www.r3eda.com, for open access to Excel VBA macros to many of the procedures in this book.

Years back our college decided to standardize with Visual Basic for Applications (VBA) for the undergraduate computer programming course. As a result, routines supporting this text are written in VBA, which is convenient to me, and also a widely accessible platform. However, VBA is not the fastest, and some readers may not be familiar with that language. Therefore, this text also provides a VBA primer and access to the code so that a reader may convert the VBA code to some other personally preferred platform. If you understand any structured text procedures, you can understand the VBA code here.

Preview of the Recommendations

Some of the recommendations in this book are counter to traditional practice in regression and design of experiments (DoE), which seem to be substantially grounded in linear regression. As a preview, opinions offered in this textbook are:

1. If the equation is nonlinear in the coefficients, use nonlinear regression. Even if the equation can be log-transformed into a linear form, do not do it. Linearizing transformations distort the relative importance of data points within the data set. Unless data variance is relatively low and/or there are many data points, linearizing can cause significant error in the model coefficient values.
2. Use data pre-processing and post-processing to eliminate outliers.
3. Use direct search optimizers for nonlinear regression rather than gradient-based optimizers. Although gradient-based algorithms converge rapidly in the vicinity of the optimum, direct search optimizers are more robust to surface aberrations, can cope with hard constraints, and are faster for difficult problems. Leapfrogging is offered as a good optimizer choice.
4. Nonlinear regression may have multiple minima. No optimizer can guarantee finding the global minimum on a first trial. Therefore, run the optimizer for N trials, starting from random locations, and take the best of the N trials. N can be calculated to meet the user desire for the probability of finding an optimum within a user-defined best fraction. The equation is shown.
5. Pay as much attention to how constraints are defined and included in the optimization application as you do to deriving the model and objective function (OF) statement. Constraints can have a substantial influence on the regression solution.

6. The choice of stopping criteria is also influential to the solution. Conventional stopping criteria are based on thresholds on the adjustable model coefficient values (decision variables, DVs), and/or the regression target (usually the sum of squared deviations) that we are seeking to optimize (OF). Since the right choice for the thresholds requires *a priori* knowledge, is scale-dependent, and requires threshold values on each regression coefficient (DV) and/or optimization target (OF), determining right threshold values requires substantial user experience with the specific application. This work recommends using steady-state identification to declare convergence. It is a single criterion (only looking at one index – statistical improvement in OF relative to data variability from the model), which is not scale-dependent.
7. Design the experimental plan (sequence, range, input variables) to generate data that are useful for testing the validity of the nonlinear model. Do not follow conventional statistical DoE methods, which were devised for alternate outcomes – to minimize uncertainty on the coefficients in nonmechanistic models, in linear regression, within idealized conditions.
8. Design the experimental methods of gathering data (measurement protocol, number and location of data sets) so that uncertainty on the experimental measurements has a minimal impact on model coefficient values.
9. Use of the conventional least-squares measure of model quality, $\sum(y_{data} - y_{model})^2$, is acceptable for most purposes. It can be defended by idealizing maximum likelihood conditions. Maximum likelihood is more compatible with reality and can provide better model coefficient values, but it presumes knowledge of the variance on both experimental inputs and output, and requires a nested optimization. Maximum likelihood can be justified where scientific precision is paramount, but adds complexity to the optimization.
10. Akaho's method is a computationally simple improvement for the total east-squares approximation to maximum likelihood.
11. Establish nonlinear model validity with statistical tests for bias and either autocorrelation or runs. Do not use r-square or ANOVA techniques, which were devised for linear regression under idealized conditions.
12. Eliminate redundant coefficients, inconsequential model terms, and inconsequential input variables.
13. Perform both logic-based *and* data-based tests to establish model validity.
14. Model utility (fitness for use) and model validity (representation of the truth about Nature) are different. Useful models often do not need to be true. Balance perfection with sufficiency, complexity with simplicity, rigor with utility.

Philosophy

I am writing to you, the reader, in a first-person personal voice, a contrast to most technical works. There are several aspects that led me to do so, but all are grounded in the view that humans will be implementing the material.

I am a believer in the Scientific Method. The outcomes claimed by a person should be verifiable by any investigator. The methodology and analysis that led to the outcomes should be grounded in the widely accepted best practices. In addition, the claims should be tempered and accepted by the body of experts. However, the Scientific Method wants decisions to be purely rational, logical, and fact based. There should be no personal opinion, human emotion, or human bias infecting decisions and acceptances about the truth of Nature. To preserve the image of no human involvement, most technical writing is in the third person. However, an author's choice of idealizations, acceptances, permissions, assumptions, givens, basis, considerations, suppositions, and such, are necessary to permit mathematical exactness, proofs, and the consequential absolute statements. However, the truth offered is implicitly infected by the human choices. If a human is thinking it, or if a human accepts it, it cannot be devoid of that human's perspective and values. I am not pretending that this book is separate from my experiences and interpretations so I am writing in the first person.

Additionally, consider the individuals applying techniques. They are not investigating a mathematical analysis underlying the technique, but need to use the technique to get an answer for some alternate purpose. Accordingly, utility with the techniques is probably as important as understanding the procedure basis. Further, the application situation is not an idealized simplification. Nature confounds simplicity with complexity. Therefore, as well as proficiency in use, a user must understand and interpret the situation and choose the right techniques. The human applies it and the human must choose the appropriate technique. Accordingly, to make a user functional, it is important for a textbook to understand the limits and appropriateness of techniques. The individual is the agent and primary target, the tool is just the tool. The technique is not the truth, so I am writing to the user.

It is also essential that a user truly understands the basis of a tool, to use it properly. Accordingly, in addition to discussing the application situations, this text develops the equations behind the methods, includes mathematical analysis, and reveals nuances through examples. The book also includes exercises so the user can develop skills and understanding.

In the 1950s Benjamin Bloom chaired a committee of educators that subsequently published a taxonomy of Learning Objectives, which has come to be known as Bloom's Taxonomy. One of the domains is termed the Cognitive, related to thinking/knowing. There are six levels in the Taxonomy. Here is my interpretation for engineering (Table 1).

Notably most of classroom instruction has the student working in the lower three levels, where there are no user-choices. There is only one way to spell "cat," only one right answer to the calculation of the required orifice diameter using the ideal orifice equation and givens in the word problem, and so on. In school, the instructor analyzes the situation, synthesizes the exercise, and judges the correctness of the answer. By contrast, competency and success in professional and personal life requires the individual to mentally work in the upper levels where the situation must be interpreted, where the approach must be synthesized, and where the propriety of the approach and answer must be evaluated. When instruction prevents the student from working in the upper cognitive levels, it misrepresents the post-graduation environment, which does a disservice to the student and employers who have to redirect the graduate's perspective. Accordingly, my aim is to facilitate the reader's mental activity in the upper levels where human choices have to be made. I am therefore writing to the human, not just about the technology.

A final perspective, on the philosophy behind the style and contents of this book is grounded in a list of desired engineering attributes. The members of the Industrial Advisory Committee for our School helped the faculty develop a list of desired engineering attributes, which we use to shape what we teach and shape the student's perspectives. Engineering is an activity, not a body of knowledge. Engineering is performed by humans within a human environment; it is not the intellectual exercise about isolated mathematical analysis. There are opposing ideals in judging engineering and the list of Desired Engineering Attributes reveals them. The opposing ideals are highlighted in bold (Table 2).

Table 1 Bloom's taxonomy

Level	Name	Function – person does	Examples
6	Evaluation (E)	Judge goodness, sufficiency, and completeness of something, choose the best among options, know when to stop improving. Must consider all aspects	Decide that a design, report, research project, or event planning is finished when considering all issues (technical completeness, needs of all stakeholders, ethical standards, safety, economics, impact, etc.)
5	Synthesis (S)	Create something new: purposefully integrate parts or concepts to design something new that meets a function	Design a device to meet all stakeholders' approvals within constraints. Create a new homework problem integrating all relevant technology, design a procedure to meet multiple objectives, create a model, create a written report, design experiments to generate useful data
4	Analysis (An)	Two aspects related to context	
		One. Separate into parts or stages, define and classify the mechanistic relationships of something within the whole	*One*. Describe and model the sequence of cause-and-effect mechanisms: tray-to-tray model that relates vapor boil-up to distillate purity, impact of transformer start-up on the entire grid, impact of an infection on the entire body and person health
		Two. Critique, assess goodness, determine functionality of something within the whole	*Two*. Define and compute metrics that quantify measures of utility or goodness
3	Application (Ap)	Independently apply skills to fulfill a purpose within a structured set of "givens"	Properly follow procedures to calculate bubble point, size equipment, use the Excel features to properly present data, solve classroom "word problems"
2	Understanding/ comprehension (U/C)	Understand the relation of facts and connection of abstract to concrete	Find the diameter of a 1-inch diameter pipe, convert units, qualitatively describe staged equilibrium separation phenomena, explain the equations that describe an RC circuit, understand what Excel cell equations do
1	Knowledge (K)	Memorize facts and categorization	Spell words, recite equations, name parts of a valve, read resistance from color code, recite the six Bloom levels

Table 2 Desired engineering attributes

Engineering is an activity that delivers solutions that work for all stakeholders. Desirably engineering:

- Seeks **simplicity** in analysis and solutions, while being **comprehensive** in scope.
- Is **careful**, correct, self-critical, and defensible; yet is performed with a **sense of urgency**.
- Analyzes **individual mechanisms** and integrates stages to **understand the whole**.
- Uses state-of-the-art **science** and **heuristics**.
- Balances **sufficiency** with **perfection.**
- Develops **sustainable solutions** – profitable and accepted **today**, without burdening **future stakeholders**.
- Tempers **personal gain** with **benefit to others**.
- Is **creative**, yet **follows codes**, regulations, and standard practices.
- Balances probable **loss** with probable **gain** but not at the expense of EHS&LP – **manages risk**.
- Is a collaborative, **partnership activity**, energized by **individuals**.
- Is an **intellectual analysis** that leads to **implementation and fruition**.
- Is **scientifically valid**, yet **effectively communicated** for all stakeholders.
- Generates **concrete** recommendations that honestly reveal **uncertainty**.
- Is grounded in **technical fundamentals** and the **human context** (societal, economic, and political).
- Is grounded in **allegiance to the bottom line of the company** and to **ethical standards of technical and personal conduct**.
- Supports **enterprise harmony** while seeking to **cause beneficent change**.

Engineering is not just about technical competence. State-of-the-art commercial software beats novice humans in speed and completeness with technical calculations. Engineering is a decision-making process about technology within human enterprises, value systems, and aspirations, and I believe this list addresses a fundamental aspect of the essence of engineering. As a complement to fundamental knowledge and skill of the core science and technical topics, instructors need to understand the opposing ideals, the practice of application, so that they can integrate the issues into the student's experience and so that student exercises have students practice right perspectives as they train for technical competency.

A straight line is very long. Maybe the line goes between pure science on one end and pure follow-the-recipe and accept-the-computer-output on the other end. No matter where one stands, the line disappears into the horizons to the left and to the right. No matter where one stands, it feels like the middle, the point of right balance between the extremes. However, the person way to the left also thinks they are in the middle. If Higher Education is to prepare graduates for industrial careers, instructors need to understand the issues surrounding Desired Engineering Attributes from an industrial perspective, not their academic/science perspective. Therefore, I am writing to the human about how to balance those opposing ideals when using nonlinear regression techniques for applications.

Acknowledgments

My initial interest in modeling processes and products arose from my engineering experience within industry, and most of the material presented here benefited from the investigations of my graduate students as they explored the applicability of these tools, guidance from industrial advisors as they provided input on graduate projects and undergraduate education outcomes, and a few key mentors who helped me see these connections. Thank you all for revealing issues, providing guidance, and participating in my investigations.

Some of the techniques in this text are direct outcomes of the research performed by Gaurav Aurora, R. Paul Bray, Phoebe Brown, Songling Cao, Chitan Chandak, Sandeep Chandran, Solomon Gebreyohannes, Anand Govindrajanan, Mahesh S. Iyer, Suresh Jayaraman, Junyi Li, Upasana Manimegalai-Sridhar, Siva Natarajan, Jing Ou, Venkat Padmanabhan, Anirudh (Andy) Patrachari, Neha Shrowti, Anthony Skach, Ming Su, John Szella, Kedar Vilankar, and Judson Wooters.

A special thanks goes to Robert M. Bethea (Bob) who invited me to coauthor the text *Applied Engineering Statistics*, which was a big step toward my understanding of the interaction between regression, modeling, experimental design, and data analysis. Another special thank you to Richard M. Felder, always a mentor in understanding and disseminating engineering science and technology.

As a professor, funding is essential to enable research, investigation, discovery, and the pursuit of creativity. I am grateful to both the Edward E. and Helen Turner Bartlett Foundation and the Amoco Foundation (now BP) for funding endowments for academic chairs. I have been fortunate to be the chair holder for one or the other, which means that I was permitted to use some proceeds from the endowment to attract and support graduate students who could pursue ideas that did not have traditional research support. This book presents many of the techniques explored, developed, or tested by the graduate students.

Similarly, I am grateful for a number of industrial sponsors of my graduate program who recognized the importance of applied research and its role in workforce development.

Most of all, career accomplishments of any one person are the result of the many people who nurtured and developed the person. I am of course grateful to my parents, teachers, and friends, but mostly to Donna, who for the past 26 years has been everything I need.

Nomenclature

Accept	Not reject. There is not statistically sufficient evidence to confidently claim that the null hypothesis is not true. There is not a big enough difference. This is equivalent to the not guilty verdict, when the accused might have done it, but the evidence is not beyond reasonable doubt. Not guilty does not mean innocent. Accept means cannot confidently reject and does not mean correct.
Accuracy	Closeness to the true value, bias, average deviation. In contrast to precision.
AIC	Akiake Information Criterion, a method for assessing the balance of model complexity to fit to data.
A priori	Latin origin for "without prior knowledge."
Architecture	The functional form of the mathematical model.
ARL	Average run length, the average number of samples to report a confident result.
Autocorrelation	One value of a variable that changes in time is related to prior values of that variable.
Autoregressive	A mathematical description that one value of a variable that changes in time is related to prior values of that variable; the cause would be some fluctuating input that has a persisting influence.
Batch regression	The process of regression operates on all of the data in one operation.
Best-of-N	Start the optimizer *N* times with independent initializations and take the best of the *N* trials as the answer.
Bias	A systematic error, a consistent shift in level, an average deviation from true.
Bimodal	A pattern in the residuals that indicates there are two separate distributions, suggesting two separate treatments affected the data.
Bootstrapping	A numerical, Monte Carlo, technique for estimating the uncertainty in a model-predicted value from the

	inherent variability in the data used to regress model coefficient values.
Cardinal	Integers, counting numbers, a quantification of the number of items.
Cauchy's technique	An optimization approach of successive searches along the line of local steepest descent.
CDF	The cumulative distribution function, the probability of obtaining an equal or smaller value.
Chauvenet's criterion	A method for selecting data that could be rejected as an outlier.
Class	The variable that contains the name of a classification – nominal, name, category.
Coefficient correlation	When the optimizer does not find a unique solution, perhaps many identical or nearly identical OF values for different DV values, a plot of one DV value w.r.t. another reveals that one coefficient is correlated to the other. Often termed parameter correlation.
Coefficient or model coefficient	A symbol in a model that has a fixed value from the model use perspective. Model constants or parameters. Some values are fundamental such as Pi or the 2 in square root. Other values for the coefficients are determined by fitting model to data. Such coefficient values will change as new data is added.
Confidence	The probability that a statement is true.
Constraints	Boundaries that cannot be violated, often rational limits for regression coefficients.
Convergence	The optimizer trial solution has found the proximity of the optimum within desired precision.
Convergence criterion	The metric used to test for convergence – could be based on the change in DVs, change in OF, and so on.
Correlation	Two variables are related to each other. If one rises, the other rises. The relation might be confounded by noise and variation, and represent a general, not exact relation. The relation does not have to be linear.
Cross correlation	Two separate variables are related to each other. Contrast to autocorrelation in which values of one variable are related to prior values.
Cumulative sum	CUSUM, cumulative sum of deviations scaled by the standard deviation in the data.
CUSUM	Cumulative sum of deviations scaled by the standard deviation in the data.
Cyclic heuristic	CH, an optimizer technique that makes incremental changes in one DV at a time, taking each in turn. If the OF is improved, that new DV value is retained and the next increment for that DV will be larger. Otherwise, the

	old DV value is retained and the next increment for that DV will be both smaller and in the opposite direction.
Data	As a singular data point (set of conditions) or as the plural set of all data points.
Data-based validation	The comparison of model to data to judge if the model properly captures the underlying phenomena.
Data model	The calculation procedure used to take experimental measurements to generate data for the regression modeling, the method to calculate y and x experimental from sensor measurements.
Data reconciliation	A method for correcting a set of measurements in light of a model that should make the measurements redundant.
Decision variables	DVs are what you adjust to minimize the objective function (OF). In regression, the DVs are the model coefficients that are adjusted to make the model best fit the data.
Dependent variable	The output variable, output from model, result, impact, prediction, outcome, modeled value.
Design	Devising a procedure to achieve desired results.
Design of experiments	DoE, the procedure/protocol/sequence/methodology of executing experiments to generate data.
Deterministic	The model returns one value representing an average, or parameter value, or probability.
Deviation	A variable that indicates deviation from a reference point (as opposed to absolute value).
Direct search	An optimization procedure that uses heuristic rules based on function evaluations, not derivatives. Examples include Hooke–Jeeves, leapfrogging, and particle swarm.
Discrete	A variable that has discrete (as opposed to continuum) values – integers, the last decimal value.
Discrimination	Using validation to select one model over another.
Distribution	The description of the diversity of values that might result from natural processes (particle size), simulations (stochastic process, Monte Carlo simulation), or an event probability.
DoE	Design of experiments.
DV	Decision variable.
Dynamic	The process states are changing in time in response to an input, often termed transient.
EC	Equal concern – a scaling factor to balance the impact of several measures of undesirability in a single objective function. Essentially, the reciprocal of the Lagrange multiplier.

Empirical	The model has a generic mathematical functional relation (power series, neural network, wavelets, orthogonal polynomials, etc.) with coefficients chosen to best shape the functionalities to match the experimentally obtained data.
Ensemble	A model that uses several independent equations or procedures to arrive at predictions, then some sort of selection to choose the average or representative value.
Equal concern factor	The degree of violation of one desire that raises the same level of concern as a specified violation of another desire, weighting factors in a penalty that are applied as divisors as opposed to Lagrange multipliers.
Equality constraints	A constraint that relates variables in an equality relation, useful in reducing the number of DVs.
EWMA	Exponentially weighted moving average, a first-order filtered value of a variable.
EWMV	Exponentially weighted moving variance, a first-order filtered value of a variance.
Experiment	A procedure for obtaining data or results. The experiment might be physical or simulated.
Exponentially weighted moving average	EWMA, a first-order filtered value of a variable.
Exponentially weighted moving variance	EWMV, a first-order filtered value of a variance.
Final prediction error	FPE, Ljung's take on Akaike's approach to balancing model complexity with reduction in SSD. Concepts are similar in Mallows' Cp and Akaike's information criterion.
First principles	An approach that uses a fundamental mechanistic approach to develop an elementary model. A phenomenological model, but not representing an attempt to be rigorous or complete.
First-order filter	FOF – an equation for tempering noise by averaging, an exponentially weighted moving average, the solution to a first-order differential equation, the result of an RC circuit for tempering noise on a voltage measurement.
FL	Fuzzy logic – models that use human linguistic descriptions, such as: "Its cold outside so wear a jacket." This is not as mathematically precise as, "The temperature is 38 °F, so use a cover with an insulation *R*-value of 12," but fully adequate to take action.
FOF	First-order filter.
FPE	Final prediction error, which is Ljung's take on Akaike's approach to balancing model complexity with reduction in SSD. Concepts are similar in Mallows' Cp and Akaike's information criterion.

Fuzzy logic	FL – models that use human linguistic descriptions, such as: "Its cold outside so wear a jacket." This is not as mathematically precise as, "The temperature is 38 °F, so use a cover with an insulation *R*-Vvalue of 12," but fully adequate to take action.
Gaussian distribution	The bell-shaped or normal distribution.
Generalized reduced gradient	GRG, a gradient-based optimization approach that reduces the number of DVs when a constraint is encountered by replacing the inequality with an equality constraint as long as the constraint is active.
Global optimum	The extreme lowest minima or highest maxima of a function.
Gradient	The vector of first derivatives of the OF w.r.t. each DV, the direction of steepest descent. Gradient-based optimizers include Cauchy's sequential line search, Newton–Raphson, Levenberg–Marquardt, and GRG.
Gradient-based optimization	Optimization approaches that use the gradient, the direction of steepest descent. Gradient-based optimizers include Cauchy's sequential line search, Newton–Raphson, Levenberg–Marquardt, and GRG.
GRG	Generalized reduced gradient, a gradient-based optimization approach that reduces the number of DVs when a constraint is encountered by replacing the inequality with an equality constraint as long as the constraint is active.
Hard constraint	May not be violated, because it leads to an operation that is impossible to execute (square root of a negative, divide by zero) or violates some physical law (the sum of all compositions must be less than or equal to 100%).
Histogram	A bar graph representing the likelihood (probability, frequency) of obtaining values within numerical intervals.
HJ	Hooke–Jeeves, an optimization procedure that searches a minimal pattern of local OF values to determine where to incrementally move the pattern, moves the pattern center, and repeats.
Homoscedasity	Constant variance throughout a range.
Hooke–Jeeves	HJ, an optimization procedure that searches a minimal pattern of local OF values to determine where to incrementally move the pattern, moves the pattern center, and repeats.
Imputation	The act of creating missing data values from correlations to available data.
Incremental regression	The model coefficients are incrementally adjusted at each sampling, so that the model evolves with the changing process that generates the data.

Incremental steepest descent	ISD, an optimization technique that makes incremental steps in the steepest descent direction, re-evaluating the direction of steepest descent after each incremental TS move.
Independent variable	Input variable, input to the model, cause, influence.
Inequality constraints	A constraint that related variables in an inequality relation, a less than or greater than relation. Could be treated as either a hard or soft constraint.
Input variable	An influence, cause, source, input, forcing function, independent value to the model.
Inverse	The model is used "backward" to answer the question, "What inputs are required to provide a desired output?"
ISD	Incremental steepest descent, an optimization technique that makes incremental steps in the steepest descent direction, re-evaluating the direction of steepest descent after each incremental TS move.
Lag	In statistics it refers to the time interval between time-discretized data. A lag of 5 means a delay of five samples. In dynamic modeling it refers to a first-order, asymptotic dynamic response to a final value. Both definitions are used in this book.
Lagrange multipliers	Weighting factors in a penalty that are applied as multipliers as opposed to equal concern factors.
Leapfrogging	LF, an optimization technique that scatters players throughout DV space and then leaps the worst over the best, to converge on the optimum.
Levenberg–Marquardt	LM, an optimization technique that blends incremental steepest descent and Newton–Raphson.
LHS	Left-hand side, the terms on the left-hand side of an equation (either equality or inequality).
Likelihood	A measure of the probability that a model could have generated the experimental data.
Linear	The relation between two variables is a straight line.
Linearizing transforms	Mathematical operations that linearize an OF, providing the convenience of using linear regression solution methods. Be cautious about the weighting distortion that results.
LF	Leapfrogging, an optimization technique that scatters players throughout DV space, then leaps the worst over the best, to converge on the optimum.
LM	Levenberg–Marquardt, an optimization technique that blends incremental steepest descent and Newton–Raphson.
Local optimum	One of several minima or maxima of a function, but not the extreme.

Logic-based validation — Comparison of model functionality to rational, logical expectations.

MA — Moving average, the average of the chronologically most recent N data values in a time series.

Maximum — Highest or largest value.

Maximum error — An estimate of the uncertainty in a calculated value, based on all sources of uncertainty providing their maximum perturbation and influencing the outcome in the same direction.

Mean — The expected average in a list of data.

Median — The middle value in a list of data. To find it, repeatedly exclude the high and low values. If an odd-numbered list, the one that remains is the median. If an even-numbered list, average the two that remain.

Minimum — Lowest or smallest value.

Model — A mathematical representation of the human's understanding of Nature's response to the influence.

Model architecture — The mathematical structure, the functional relations within a model.

Moving average — MA, the average of the chronologically most recent N data values in a time series.

Nature — A respectful anthropomorphic representation of the mystery of the processes that generate data and which tortures us with complexity and variation.

Nelder–Mead — NM, a direct search optimization technique that uses the Simplex geometry and moves the worst local trial solution through the centroid of the others.

Neural network — NN – a modeling approach that was intended to mimic how brain neurons "calculate."

Newton–Raphson — NR, an optimization method that uses the local OF derivatives and a second-order series approximation of the OF to predict the optimum, jumps there and repeats.

NID(μ, σ) — Normally (Gaussian) and independently distributed with a mean of μ and a standard deviation of σ.

NM — Nelder–Mead, a direct search optimization technique that uses the Simplex geometry and moves the worst local trial solution through the centroid of the others.

NN — Neural network – a modeling approach that was intended to mimic how brain neurons "calculate."

Noise — Random, independent perturbations to a conceptually deterministic value.

Nominal — Latin origin for name, a class/category/string variable.

Nonlinear — Means not linear. This could be any not linear relation.

Nonparametric test — The category of statistical tests that do not presume a normal model of the residuals.

Normal equations	The set of linear equations that arise in linear regression when the analytical derivatives of the OF w.r.t. each coefficient are set to zero.
Normal SSD	The sum of squared differences between the model and data in both the x and y axes, normal to the model, perpendicular to the model, often called total SSD.
NR	Newton–Raphson, an optimization method that uses the local OF derivatives and a second-order series approximation of the OF to predict the optimum, jumps there and repeats.
Not reject	There is not statistically sufficient evidence to confidently claim that the null hypothesis is not true. There is not a big enough difference. This is equivalent to the not guilty verdict, when the accused might have done it, but the evidence is not beyond a reasonable doubt. Not guilty does not mean innocent. Not reject does not mean correct. We use accept, instead of not reject.
Null hypothesis	The supposition that two treatments are equal, that differences are due to experimental vagaries, not mechanistic cause-and-effect relations.
Objective function	The procedure or equation that provides a measure of badness to be minimized (or goodness to be maximized). Usually the OF is the sum of squared deviations of modeled output to data.
Objective function value	The value of the OF.
OF	Objective function.
Optimization	A procedure for determining best values. In regression it is applied to determine model coefficient values that make model best fit to the data.
Optimum	Either lowest or highest value, either minimum or maximum.
Ordinal	Sequence or rank from ordering things relative to some attribute.
Outlier	A data point that is expected to be part of the cluster, but deviates too far from the expected location to be accepted as "all is well."
Output variable	A response, effect, output, consequent, prediction, state variable, modeled value.
Overfitting	When there are too many adjustable model coefficients and the model begins to fit noise in the data, not just the underlying trend. In the neural network community this is termed memorization.
Parametric analysis	The exploration of trends in model output variables as coefficient values change. Used in logical validation and sensitivity analysis.

Parametric test	The category of statistical tests that presumes a normal model of the residuals.
Parity plot	A graph of the model predicted values w.r.t. the actual data. Ideally, points fall on the 1:1 line. Patterns in deviations reveal model issues.
Particle swarm optimization	PSO, an optimization technique that scatters individuals (particles, players) throughout DV space and then each explores the local area with random perturbations while being drawn to both their personal best spot and the global best of all players.
Partitioned model	A model that has an IF-THEN structure that directs which equation or procedure should be used. For instance, if laminar flow, use the Hagan–Poiseuille relation, but if turbulent flow, use the Fanning–Moody–Darcy relation.
PDF	Probability distribution function, a mathematical equation representing the normalized histogram.
Penalty	A value added to the OF representing the degree of violation of a soft constraint.
Phenomenological	The mathematical model was derived from a conceptual understanding of the behavior of Nature. Other terms include first principles, mechanistic, fundamental, theoretical, rigorous, physical, and scientific.
Post-processing	Adjustment or culling of data after data work-up, in the light of modeling outcomes.
Precision	A measure of reproducibility of replicated results, perhaps the standard deviation. This is in contrast to accuracy, which is a measure of the average deviation from true.
Pre-processing	Adjustment or culling of data prior to use in modeling.
Probable error	An estimate of the uncertainty in a calculated value, based on the multiple sources of uncertainty providing independent perturbation magnitudes and signs.
Probability model	A model that predicts the probability distribution, either PDF or CDF.
Process/product model	The nominal input–output, mechanistic, influence–response, composition–property model of the process or product.
PSO	Particle swarm optimization, a multiplayer direct search approach.
PV	Process variable, a measured value, typically from a continuously operating process.
r-lag-1	A ratio of variances used to indicate autocorrelation of the residuals.
r-square	A ratio of variances to indicate the amount of variance removed by the model.

r-statistic	A ratio of variances used in steady- and transient-state identification.
Random	Independent values in a sequence. The distribution may or may not be Gaussian.
Rank	The variable that contains the order, precedence, placement, preference.
Rational	A variable with continuous-valued values (as opposed to integer or discrete values); the value preserves the ratio of some quantification, a real number.
Real	A variable with continuous-valued values (as opposed to integer or discrete values) – rational, double precision, scientific, rational.
Realization	A particular result of a stochastic outcome, it represents one possible result, not the average.
Recursive	The identical procedure is repeated, but using the most recent outcomes. The values are iteratively updated.
Regression	The procedure of adjusting model coefficient values to minimize the SSD based on the "*y*" distance (or some alternate measure of closeness) between data and model.
Regressors	The name for input variables to the model. The user chooses these. They need to be the complete set, represent the right delays, and not contain extraneous variables.
Reject	There is statistically sufficient evidence to confidently claim that the null hypothesis is not true, that there is a difference, that the process is not a steady state, that a variable is not zero.
Replicate	Repeated, with the intent to exactly reproduce experimental conditions.
Right	The model properly captures the natural phenomena that it seeks to represent, has fidelity, is true, correct. In possible contrast to being useful.
RHS	Right-hand side, the terms on the right-hand side of an equation (either equality or inequality).
rms	Root-mean square, the square root of the average of squared deviations. It would be the standard deviation if it were normalized by $N-1$ terms in the sum.
Run	A sequence of residuals with the same sign; the pattern in the runs will change when the residuals are ordered by different variables.
Runs test	A test to see if there are too few or too many runs in the data.
Scaled variables	Values of the variables are normalized by their range. The variables are dimensionless and often scaled to be

	in the range of 0–1. The scaling might be the actual data range or an expected maximum range.
Self-tuning filter	STF, a first-order filter that adapts filter factor to the data variability.
Semi-empirical	The model is a blend of phenomenological and empirical.
Significant digits	Those digits in a measurement or calculation representing confidently known values, not uncertain values.
Simulation	Solving a mathematical procedure to mimic what Nature would do.
Soft constraint	Should not be violated, but a penalty is added to the OF proportional to the degree of violation.
SPC	Statistical process control, techniques for monitoring processes and triggering action only after a change is statistically significant.
SQ	Successive quadratic, an optimization approach that uses a surrogate model, a quadratic functionality, of the local OF response to DVs.
SS	Steady state.
SSD	Sum of squared deviations between modeled and actual values.
SSID	Steady state identification.
Static	Something that does not change in time.
Statistic	A value that represents some property of a distribution of values (average, variance, median, runs, etc.)
Statistical process control	SPC, techniques for monitoring processes and triggering action only after a change is statistically significant.
Steady state	The process has settled to a state in which it does not change in time, but sequential measurements will not have exactly reproducible values because of noise. The process may have come to thermodynamic equilibrium, but it might be off equilibrium due to persistent influences.
STF	Self-tuning filter, a first-order filter that adapts the filter factor to the data variability.
Stochastic	A process that does not return a unique value after each trial, but returns a distribution of values. Rolling a die returns a 1, 2, 3, 4, 5, or 6, a uniform distribution with six possible values. Measuring heights of people results in a normal distribution with continuous values.
Successive quadratic	SQ, an optimization technique that uses local OF values to generate a quadratic surrogate model of the function, jumps to the optimum of the surrogate model, and then repeats.

Systematic error	A bias, a consistent average deviation from true.
Total SSD	The sum of squared differences between the model and data in both the *x* and *y* axes, normal to the model, perpendicular to the model.
Transient	The process states are changing in time in response to an input, often termed dynamic.
Trial	The run of an experiment or the run of an optimization that produces results.
Trial solution	An optimizer guess for the DV values that move toward the optimum.
Truth about Nature	The impossible-to-know mechanisms within a real process. Often, it seems so simple, but each refinement of investigation takes the scientist toward greater complexity. Often idealized concepts and models are fully adequate for engineering.
TS	Transient state.
Type I error	The null hypothesis is true (there is no difference), but the data provided unusual extreme values to lead to rejecting the null hypothesis – true but rejected.
Type II error	The null hypothesis is false (there is a difference), but the data provided values that were not definitively different, that could not justify rejecting the null hypothesis – false but accepted.
Uncertainty	A measure of the possible range or a value. The value might be experimental and range could be estimated from replicate measurements. The value might be from a model, and uncertainty can be mathematically propagated through the calculations.
Useful	The model balances perfection with sufficiency, that is, provides a good-enough representation, and that it is functional (convenient, reliable, sufficiently accurate) in use. In possible contrast to being right.
Validation	A procedure that determines whether the model can be rejected by the data or by logical considerations.
Variance	The square of the standard deviation in replicate data.
Verification	A procedure that determines that the model execution has fidelity to the model concepts.
Vertical SSD	The sum of squared deviations in the model prediction, parallel to the *y* axis.
Voting	Taking the middle measurement of three independent sensors as the data that represents the process.
w.r.t.	With respect to, the equivalent of "against" or "versus" when describing a response variable graphed with respect to an influence variable.

Symbols

$a, b, c, d, \ldots$ = model coefficients
$\alpha, \beta, \gamma, \delta, \ldots$ = model coefficients
α = level of significance, Type I error probability, distance factor in steepest descent
B = measure of badness, typically the amount of a constraint violation
β = Type II error probability
c = confidence in an event, number of columns
CDF = cumulative distribution function
d = deviation between data and model, disturbance
D = equal concern deviation
DV = decision variable value
δ = infinitesimal change, measure of variance
Δ = small change
∇ = gradient operator
E = expected count, desired half-range
ε = small deviation, error, correction
f = function, fraction
F = F-statistic for variance ratio
g = function derivative
H = Hessian matrix of second derivatives
H_0 = null hypothesis
I = identity matrix
J = objective function, Jacobean vector of first derivatives
λ = weighting factor, Lagrange multiplier, filter factor
m = number of model coefficients
M = number of players in a multiplayer optimization, number of decision variables, number of function evaluations
μ = mean, true value
n = noise, random and independent perturbation to a measurement or condition, delay counter, value of an index variable, number of occurrences
N = number of data sets, number of iterations, number of trials

O	=	observed count
OF	=	objective function value
p	=	parameter, coefficient, probability of an event
P	=	penalty for a constraint violation, probability of a compound event
PDF	=	probability distribution function
q	=	probability of not-an-event $= 1 - p$
r	=	residual, UID[0,1] random variable, number of rows
r_1	=	r-lag-1 autocorrelation statistic
r^2	=	variance reduction statistic
R	=	range of a variable (high minus low values), ratio statistic
rms	=	root-mean square $= \sqrt{SSD/N}$
s	=	estimate of the standard deviation, scaling parameter in the logistic model
S	=	distance along a line
SSD	=	sum of squared deviations
σ	=	standard deviation
t	=	time, t-statistic
τ	=	time constant
u	=	input variable to a dynamic process, the forcing function
υ	=	degrees of freedom, measure of variance
w	=	weighting coefficient
x	=	input variable (influence, independent, given)
X	=	process variable (either influence or response)
χ^2	=	chi-square statistic
$\tilde{y}$	=	modeled response variable
y	=	response variable (output, dependent, effect, outcome)
ψ	=	true value of y

Subscripts, Superscripts, and Marks

i	=	counter for data set number or time or iteration
$'$	=	scaled variable, deviation variable
SS	=	steady state
*	=	optimal value