

Springer Series in Statistics

James Ramsay
Giles Hooker

Dynamic Data Analysis

Modeling Data with Differential
Equations

 Springer

Springer Series in Statistics

Series Editors

Peter Bickel, Berkeley, CA, USA

Peter Diggle, Lancaster, UK

Stephen E. Fienberg, Pittsburgh, PA, USA

Ursula Gather, Dortmund, Germany

Scott Zeger, Baltimore, MD, USA

More information about this series at <http://www.springer.com/series/692>

James Ramsay · Giles Hooker

Dynamic Data Analysis

Modeling Data with Differential Equations

 Springer

James Ramsay
Department of Psychology
McGill University
Ottawa, ON
Canada

Giles Hooker
Department of Biological Statistics and
Computational Biology
Cornell University
Ithaca, NY
USA

ISSN 0172-7397

Springer Series in Statistics

ISBN 978-1-4939-7188-6

DOI 10.1007/978-1-4939-7190-9

ISSN 2197-568X (electronic)

ISBN 978-1-4939-7190-9 (eBook)

Library of Congress Control Number: 2017941073

© Springer Science+Business Media LLC 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer Science+Business Media LLC

The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

Getting pregnant is usually easy and fun, but the gestation and delivery may be another story; messy and painful perhaps, but instructive nevertheless. So it is with this book, which began with enthusiasm and confidence, but ten or so years later the twists and turns along the way emerge as a key part of the story.

Functional data analysis leads inevitably to dynamic systems. Ramsay and Silverman (2005) emphasized the reduction in bias and sampling variance that could be achieved by incorporating even an only approximately correct model into the penalty term by using a linear differential operator, thereby extending the more usual practice of defining roughness by the size of a high-order derivative. It was a natural next step to consider how one or more parameters that were needed to define such an operator might be estimated from data. Data associating the incidence of melanoma with solar activity became a prototype problem.

Principal differential analysis—a specific case of what we here term *gradient matching*—and its close resemblance to principal components analysis was the subject of a later chapter in Ramsay and Silverman (2005). The availability of high-resolution replicated data resulted in some rather successful applications, notably to data recording complex physical motion: handwriting and juggling. These analyses estimated low-dimensional basis systems that could be used to define the linear differential operators whose kernels were spanned by these systems which largely captured the variation in the data. This leads to the somewhat clumsy attempt, judging by limited attention that it received, to introduce the reader to Wronskians and Green's functions, not to mention such ethereal topics as reproducing kernel Hilbert spaces (we still regard the later chapters of Ramsay and Silverman (2005) as highly instructive for the so-inclined reader). In contrast, chapters on function linear regression, which did seem to appeal to readers, were for us only a partial success. Our attempts to use functions as covariates in a regression equation to approximate other functions seemed to us only marginally successful and, as a premonition of trouble to come, we were struck by how hard it was to find data to illustrate the methodologies involved.

The first serious attention given to parameter estimation for differential equations began with a collaboration with N. Heckman (Ramsay and Heckman 2000) where

we noted that the large number of parameters in the smoothing function relative to the few defining the operator tended to lead to overfitting of the data and bias in the estimates of the parameters of interest. The idea of parameter cascading described in Chap. 9 came from realizing that the implicit function theorem provided a way out of this dilemma by replacing the unrestricted coefficients of the basis function expansion by a smooth function of the parameters being estimated.

We benefited enormously by a close collaboration and friendship with Kim McAuley and Jim McLellan in the Department of Chemical Engineering at Queen's University, who were able to steer us to the large literature in that field on the nonlinear least squares estimation strategy described in Chap. 7, and to pass along the nylon and refinery data used in various chapters.

We discovered again just how hard it was to find data that we could use in demonstration and test analyses in the engineering world, where data are owned by industrial concerns protected against access by competitors. When we turned to the large literature on dynamics in various fields like biology, epidemiology and physiology, we found almost no use or display of data. For example, nearly every text used the spread of disease (SIR) or the closely related Lotka-Volterra equations as a first illustration of a nonlinear system; but, if data were available at all, it was only on infected cases or predator abundances, where data fits were essentially only smooths of the data and therefore uninformative. Only recently have we discovered the invaluable archive of dynamic systems with data assembled by Klaus Schittkowski (Schittkowski 2002), which has been a great help in completing this book, and we are most grateful for his cooperation.

Why, we asked, were data-based estimations of dynamical system so hard to find? One answer seemed to be the rather restricted set of parameters yielding the solution characteristics that motivated such systems as the Lotka-Volterra, SIR, tank reactor and the FitzHugh-Nagumo, which are featured in this volume. Parameter estimation strategies, including our own, were prone to bouncing parameters into regions where they generated completely inappropriate solutions.

But another possible explanation for the paucity of data is that, in many fields, the differential equation is viewed, effectively, as data itself. That is, if there are solutions of a proposed system that exhibits the shape characteristics seen in experiments and natural settings, such as oscillations in predator-prey abundances, then these systems are considered to be demonstrations that a scientific understanding of the live system has been gained. We observed that papers on dynamics without any display of fits to data often appeared in journals like *Nature* and *Science*. Even the Hodgkin-Huxley papers, themselves exceptional examples of data-based science, were quickly followed by downgrades of their model to those like the FitzHugh-Nagumo that retained the general shape features but were not intended as data models.

Ironically, the idea of the equation as data, although rather difficult for an information scientist to warm to, dovetails beautifully with the parameter cascading algorithm that we discuss in Chaps. 9 and 10, where we allow a smooth and continuous set of compromises between data-fitting and equation-solving. This

tension between data and equation is everywhere in evidence in these pages as well as the dynamical system modeling literature. It is, in fact, why we wrote this book.

Our central concept of a dynamical system as a buffer that translates sudden changes in input into smooth controlled output responses has led us to applications to data that we have previously analyzed, such as the daily Canadian weather data and the Chinese handwriting data (Ramsay 2000). We hope to have opened up entirely new opportunities for dynamical systems where none were envisaged before, which involve extensions of the functional linear model and what we call *dynamic smoothing*.

Dear reader, if you have survived to this point in this Preface, you must be wondering how much you need to know to read further. Take heart! We have worked hard to keep the technical level as low as possible, and our first goal is to bring those with little or even no exposure to differential equations as modeling objects into this fabulous data analysis landscape. Our emphasis on linear systems reflects a belief that nature is a tough place where only rugged and stable systems exhibit adaptive behavior over a wide range of environmental conditions survive.

We thank our former graduate students David Campbell and Jiguo Cao for their own versions of dynamic systems analyses. Cornell University contributed, besides the Tennebaum and Pollard team, our colleagues Steve Ellner and John Guckenheimer whose writing, experimentation and mathematical analysis continue to inform and inspire our work. We thank our Queen's collaborators Jim McClellan and Kim McAuley, as well as their students and colleagues for their guidance through the fascinating chemical engineering world and their hospitality and support.

Ottawa, Canada
March 2017

James Ramsay

Contents

1	Introduction to Dynamic Models	1
1.1	Six Examples of Input/Output Dynamics	1
1.1.1	Smallpox in Montreal	1
1.1.2	Spread of Disease Equations	3
1.1.3	Filling a Container	4
1.1.4	Head Impact and Brain Acceleration	5
1.1.5	Compartment Models and Pharmacokinetics	7
1.1.6	Chinese Handwriting	8
1.1.7	Where to Go for More Dynamical Systems	11
1.2	What This Book Undertakes	12
1.3	Mathematical Requirements	13
1.4	Overview	14
2	Differential Equations: Notation and Architecture	17
2.1	Introduction and Chapter Overview	17
2.2	Notation for Dynamical Systems	17
2.2.1	Dynamical System Variables	17
2.2.2	Dynamical System Parameters	18
2.2.3	Dynamical System Data Configurations	19
2.2.4	Mathematical Background	19
2.3	The Architecture of Dynamic Systems	20
2.3.1	Elementary Autonomous or Unforced Systems	20
2.3.2	Forced or Non-autonomous Systems	21
2.3.3	Differential Operator Notation	23
2.4	Types of Differential Equations	23
2.4.1	Linear Differential Equations	23
2.4.2	Nonlinear Dynamical Systems	24
2.4.3	Partial Differential Equations	25
2.4.4	Algebraic and Other Equations	26

2.5	Data Configurations	26
2.5.1	Initial and Boundary Value Configurations	26
2.5.2	Distributed Data Configurations	27
2.5.3	Unobserved or Lightly Observed Variables	27
2.5.4	Observational Data and Measurement Models	28
2.6	Differential Equation Transformations	29
2.7	A Notation Glossary	29
3	Linear Differential Equations and Systems	31
3.1	Introduction and Chapter Overview	31
3.2	The First Order Stationary Linear Buffer	33
3.3	The Second Order Stationary Linear Equation	36
3.4	The m th Order Stationary Linear Buffer	39
3.5	Systems of Linear Stationary Equations	40
3.6	A Linear System Example: Feedback Control	41
3.7	Nonstationary Linear Equations and Systems	45
3.7.1	The First Order Nonstationary Linear Buffer	45
3.7.2	First Order Nonstationary Linear Systems	47
3.8	Linear Differential Equations Corresponding to Sets of Functions	49
3.9	Green's Functions for Forcing Function Inputs	50
4	Nonlinear Differential Equations and Systems	53
4.1	Introduction and Chapter Overview	53
4.2	The Soft Landing Modification	53
4.3	Existence and Uniqueness Results	54
4.4	Higher Order Equations	56
4.5	Input/Output Systems	57
4.6	Case Studies	57
4.6.1	Bounded Variation: The Catalytic Equation	57
4.6.2	Rate Forcing: The SIR Spread of Disease System	59
4.6.3	From Linear to Nonlinear: The FitzHugh-Nagumo Equations	62
4.6.4	Nonlinear Mutual Forcing: The Tank Reactor Equations	64
4.6.5	Modeling Nylon Production	66
5	Numerical Solutions	69
5.1	Introduction	69
5.2	Euler Methods	70
5.3	Runge–Kutta Methods	72
5.4	Collocation Methods	74
5.5	Numerical Problems	77
5.5.1	Stiffness	77
5.5.2	Discontinuous Inputs	78
5.5.3	Constraints and Transformations	79

- 6 Qualitative Behavior** 83
 - 6.1 Introduction 83
 - 6.2 Fixed Points 85
 - 6.2.1 Stability 85
 - 6.3 Global Analysis and Limit Cycles 89
 - 6.3.1 Use of Conservation Laws 90
 - 6.3.2 Bounding Boxes 91
 - 6.4 Bifurcations 91
 - 6.4.1 Transcritical Bifurcations 92
 - 6.4.2 Saddle Node Bifurcations 93
 - 6.4.3 Pitchfork Bifurcations 94
 - 6.4.4 Hopf Bifurcations 94
 - 6.5 Some Other Features 95
 - 6.5.1 Chaos 96
 - 6.5.2 Fast-Slow Systems 98
 - 6.6 Non-autonomous Systems 99
 - 6.7 Commentary 101
- 7 Nonlinear Least Squares or Trajectory Matching** 103
 - 7.1 Introduction 103
 - 7.2 Least Squares with Gauss–Newton Methods 105
 - 7.2.1 Sensitivity Equations 107
 - 7.2.2 Automatic Differentiation 110
 - 7.3 Inference 111
 - 7.4 Measurements on Multiple Variables 113
 - 7.4.1 Multivariate Gauss–Newton Method 113
 - 7.4.2 Variable Weighting Using Error Variance 115
 - 7.4.3 Estimating σ_i^2 115
 - 7.4.4 Example: FitzHugh–Nagumo Models 117
 - 7.4.5 Practical Problems: Local Minima 118
 - 7.4.6 Initial Parameter Values for the Chemostat Data 120
 - 7.4.7 Identifiability 122
 - 7.5 Bayesian Methods and Markov Chain Monte Carlo 125
 - 7.6 Constrained Optimization: Multiple Shooting and Collocation Methods 129
 - 7.7 Fitting Features 131
 - 7.8 Applications: Head Impacts 134
- 8 Two-Stage Least Squares or Gradient Matching** 137
 - 8.1 Introduction 137
 - 8.2 Smoothing Methods and Basis Expansions 138
 - 8.3 Fitting the Derivative 142
 - 8.3.1 Optimizing Integrated Squared Error (ISSE) 142
 - 8.3.2 Gradient Matching for the Refinery Data 144
 - 8.3.3 Gradient Matching and the Chemostat Data 144

8.4	System Mis-specification and Diagnostics	146
8.4.1	Diagnostic Plots	146
8.5	Conducting Inference	148
8.5.1	Nonparametric Smoothing Variances	149
8.5.2	Example: Refinery Data	150
8.6	Related Methods and Extensions	151
8.6.1	Alternative Smoothing Methods	151
8.6.2	Numerical Discretization Methods	152
8.6.3	Unobserved Covariates	152
8.6.4	Nonparametric Models	153
8.6.5	Sparsity and High Dimensional ODEs	154
8.7	Integral Matching	155
8.8	Applications: Head Impacts	158
9	Profiled Estimation for Linear Systems Estimated by Least Squares Fitting	161
9.1	Introduction and Chapter Overview	161
9.2	The Parameter Cascading Strategy for Estimating Parameters	163
9.2.1	Two Classes of Parameters	163
9.2.2	Defining Coefficients as Functions of Parameters	164
9.2.3	The Symmetric Relation Between the Data and the Differential Equation	165
9.2.4	Inner Optimization Criterion J	165
9.2.5	The Least Squares Cascade Coefficient Function	166
9.2.6	The Outer Fitting Criterion H	169
9.3	Choosing the Smoothing Parameter ρ	171
9.4	Confidence Intervals for Parameters	173
9.4.1	Simulation Sample Results	175
9.5	Multi-variable Systems	176
9.6	Analysis of the Head Impact Data	178
9.7	A Feedback Model for Driving Speed	181
9.7.1	Two-Variable First Order Cruise Control Model	181
9.7.2	One-Variable Second Order Cruise Control Model	184
9.8	The Dynamics of the Canadian Temperature Data	185
9.9	Chinese Handwriting	189
9.10	Exploring Complexity Through A Transformation of Basis Functions	192
9.11	Notes on Software and Computation	196
9.11.1	Rate Function Specifications	196
9.11.2	Model Term Specifications	198
9.11.3	Memoization	200

- 10 Profiled Estimation for Nonlinear Systems** 201
 - 10.1 Introduction and Chapter Overview 201
 - 10.2 Parameter Cascading for Nonlinear Systems 203
 - 10.2.1 The Setup for Parameter Cascading 203
 - 10.2.2 Parameter Cascading Computations 203
 - 10.2.3 Some Helpful Tips 204
 - 10.2.4 Nonlinear Systems and Other Fitting Criteria 205
 - 10.3 Fitting the Lotka–Volterra Model to the Chemostat Data. 206
 - 10.4 A CollocInfer Analysis of the Head Impact Data 209
 - 10.5 Fitting a Compound Model to Blood Alcohol Data 211
 - 10.6 Fitting the Catalytic Growth Model to the Root Length Data 213
 - 10.7 Chemical Reactions Among Aromatic Compounds 214
- References** 221
- Index** 227

Symbols

t	A time value
i	Usually, the index of a variable in a system of differential equations
d	The total number of variables in a system
j	Usually, the index of an observation
n or n_i	The number of observations for variable i
t	A vector of time values of length n
x_i	Used to designate a single variable among a set of d variables
$x(t)$	Used to represent the state vector of the system at time t . In the case of the SIR models above we have $x(t) = (S(t), I(t), R(t))$ and we will continue to refer to the constituent parts of the state vector by other names or by using subscripts as in $\mathbf{x}(t) = (x_S(t), x_I(t), x_R(t))$. When the state vector \mathbf{x} is viewed as a function of time, it will be described as the state <i>trajectory</i> .
ω_i	A positive weight to be applied to the fitting terms of the i th variable
D	The derivative operator that transforms a function x into its time derivative dx/dt . D^m generates the derivative of order m . D^{-1} generates the antiderivative. $Dx(t)$ is the vector of time derivatives of $x(t)$.
β	A coefficient or rate function in the homogeneous portion of a linear differential equation. This may be a function of time or the values of external variables, but may not be a function of the values of variables in the system.
L	A linear differential operator that transforms a function x into a linear combination of time derivatives $D^j x$. The coefficients in the linear combination can be functions of time, but may not be functions of values of variables.
θ	A vector of parameters that require estimation from data

ϑ	A vector of parameters that require estimation from data augmented by initial values that must also be estimated from data
x_0	Refers to the starting point of the trajectory. We distinguish this from $x(0)$ because it will often need to be estimated as additional parameters. When this is the case, we will use the augmented parameter vector $\vartheta = (\theta, x_0)$. The initial time 0 can be replaced by an suitable real number.
$f(x \theta)$	Represents a vector-valued function of x that depends on a parameter vector θ . Generally this is used to represent the right-hand side of a differential equation.
α	A coefficient or rate function in the forcing portion of a linear differential equation that multiplies an additive external input u into a linear differential equation
$u(t)$	Describes a vector of external or forcing functions that are additive inputs into an autonomous or homogeneous system and therefore render it nonautonomous or nonhomogeneous. An individual forcing function is u_ℓ .
L	The total number of forcing functions in a linear differential equation
y_{ij}	An observation or measurement value at time t_j for variable I
y_j	Represents the vector of measurements of length d of a set of d variables at time t_j . Sometimes it will be useful to write down a matrix of observation vectors over time. For this we will use Y_j in which rows denote time points and columns the dimension of the observation vector.
ϕ_κ	A known basis function used to approximate a trajectory x in a linear expansion
K	The total number of basis functions in a linear expansion
ϕ	A vector of basis functions of length K
c_k or c_{ik}	Real numbers that are the coefficients in a basis function expansion of a trajectory
c or C	Either a vector or a matrix of coefficients for one or more basis function expansions
ρ	A smoothing parameter in the half-open interval $[0, 1)$ used the analysis of linear differential equations by program Data2LD
J	The inner fitting criterion in a parameter cascading analysis in Chaps. 9 and 10
H	The outer fitting criterion in a parameter cascading analysis in Chaps. 9 and 10
λ	A non-negative smoothing parameter in used the analysis of linear or nonlinear differential equations by program CollocInfer
I	An identity matrix

- B** A matrix of rate values for a system of linear stationary differential equations
- R** The matrix defining the penalty term for the homogeneous part of a linear differential equation, defined and used in Chap. 9
- S** The matrix defining the penalty term for the forcing part of a linear differential equation, defined and used in Chap. 9

Additional symbols, primarily for variables in specific dynamical systems, are found in the index of the book.

Chapter 1

Introduction to Dynamic Models

1.1 Six Examples of Input/Output Dynamics

Science and baseball have much in common; what counts is not where the ball is but where it's going and how fast it's getting there. This places the focus on modelling change, and change is more complex than stasis. The trajectory of a baseball is greatly affected by spin, so that the three dimensions of motion combine with three spin coordinates to produce a six-dimensional system for change distributed over the single dimension of time.

It is inevitable, then, that derivatives with respect to time and space appear often in mathematical models as measures of instantaneous change, and in this book we consider how these *differential equations* can be estimated from data.

The book is also an introduction to a new horizon in statistics: an equation as a model for data, where a differential equation is only one possible sub-class, and where data-fitting, parameter estimation, model evaluation, inference and so on do not rely on finding an algebraic solution to the equation or even an accurate numerical approximation to a solution.

This chapter lets data from various settings speak to us about situations where it is the statistical description of *change* that is the primary focus. Simple dynamical models are proposed in each setting, but with no attempt to describe what we must do in subsequent chapters: estimate model parameters from noisy data, display the uncertainty in these estimates, and assess fits to the data from competing models.

1.1.1 *Smallpox in Montreal*

Nothing leads to insight faster than facing a gruesome death. In April of 1885, when a man knocked on the door of the Hotel Dieu hospital in Montreal hoping to see a doctor, the 167,000 citizens had for a few years neglected vaccination against smallpox. Within a few months 3234 would die, and about 9600 would be infected,

many of whom would be permanently disfigured. Figure 1.1 tells the story with data. We see the consequences of the introduction of compulsory vaccination in 1876 by Sir William Hingston, the Irish Roman Catholic mayor of Montreal and its chief surgeon. Working with Bishop Bourget to overcome the fear by French Canadian catholics that the real purpose of vaccination was the decimation their population by spreading the disease, it took a couple of years before most of Montreal's children were immunized, and we see in the figure the decline over the next 2 years in smallpox fatalities from around 600 per year to zero.

But by 1885 noncompliance had returned, partly due to fear-mongering by a quack physician from New York. Smallpox arrived in Montreal on the newly constructed Canadian Pacific Railroad with soldiers returning from the suppression of the Riel Rebellion in the prairies, and spread with the speed that the figure makes only too obvious. The insight that vaccination was essential spread nearly as quickly.

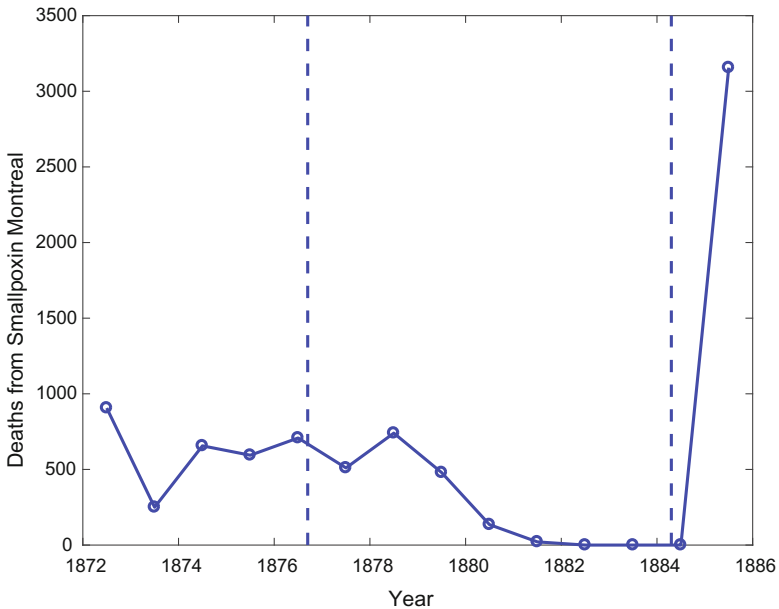


Fig. 1.1 The number of smallpox deaths per year in Montreal, Quebec. The *first vertical line* indicates the introduction of compulsory vaccination, and the second the re-introduction of vaccination following the last smallpox epidemic in North America

We are, naturally, impressed by the three levels of small pox fatalities, 600, 0 and 3000, but the data are more dramatic in terms of the two rates of change: the disappearance of the disease in 2 years or so, and its explosive return. Thus we see the structure in the data on two time scales, namely the two long-term mortality levels, and the two short-term transitions. We need to model both the levels and the rates of change. That is, we want a model that relates $x(t)$ to $Dx(t)$, where $x(t)$ is deaths per year and we use the notation Dx for the first derivative of x instead of dx/dt .

Focussing first on the decline in deaths in the 1876–1879 period, we see that, when the death rate is high, the downward slope is negative and large, but as the death rate approaches zero, the slope remains negative but declines to zero. It seems plausible that the proportional relation

$$Dx(t) = -\beta x(t) \text{ or } \tau Dx(t) = -x(t) \quad (1.1)$$

could get us somewhere. What well-known function has a derivative value negatively proportional to a function value? Introductory calculus and a little reflection leads to $x(t) = C \exp(-\beta t)$ for some arbitrary constant C as the solution to (1.1), and from there to the nonlinear model equation

$$x(t) = C e^{-\beta t} + e(t) \text{ or } x(t) = C e^{-t/\tau} + e(t) \quad (1.2)$$

where an appropriate positive constant C must be estimated and the residual $e(t)$ takes care of small and ignorable effects. Of these two equations, (1.1) is simpler and tells the story with the greater clarity, and gives us our first lesson in constructing or parsing a *differential equation* that links rate of change to level.

Parameters β or $1/\tau$ are, in effect, regression coefficients in a simple single-predictor linear relation between “covariate” x and “response” Dx . A useful rule of thumb is that $x(t)$ will decline from a level of C to practically zero in 4τ time units, leading to a rough-and-ready estimate derived from the 3-year disappearance of smallpox, so that the *time constant* $\tau \approx 3/4$ years and $\beta \approx 4/3$. That is, β controls the *speed* with which the level changes to its new level, and consequently represents what we mean by the *dynamics* of the system. Highly dynamic systems have large β 's and small τ 's, and sluggish systems have the respective converses.

This simple dynamic equation models a *buffering* process in which an impact of an event, here the introduction of compulsory vaccination, is spread over time. This concept of a differential equating defining a buffer will play a central role in this book.

We need a statistical method for using the smallpox deaths reported to the city in April of each year to estimate parameter β , along with confidence limits. The useful data are those from 1876 to 1883, with the three values over 1878–1880 being especially informative about the rate of decline in smallpox mortality. The single report in 1885 is evocative, but by itself insufficient for describing the time course of the infection.

1.1.2 Spread of Disease Equations

But what can be said about the onset of smallpox? Viewing vaccination as itself a disease, as many Montrealers did, smallpox by comparison is far more dynamic. Indeed, it is only surpassed by measles as the most communicable disease in humans. We have only a single point to record the time course of smallpox infection, but here

is a system of two equations that is often used to model the infection process:

$$\begin{aligned} DS &= -\beta SI \\ DI &= \beta SI - \delta I . \end{aligned} \tag{1.3}$$

The S variable stands for the number of individuals susceptible to infection, roughly all of the 167,000 citizens less those with immunity because of a prior infection. Variable I represents the number of infected people. Focussing on the first equation, here we have again the same relation as before between DS and S . But with a difference! The speed of the impact of the infection is now βI , since we use β here as the probability of infection of a single individual. This means that the population rate of infection gathers speed in proportion to the number of infected individuals, so that the number of uninfected persons goes down faster and faster as the epidemic spreads. Left to itself, the entire population would be infected in a year or so according to this model. But Mayor Hingston now had no trouble re-introducing compulsory vaccination, in spite of a riot or two, and this combined with severe quarantining arrested the disease by the end of the summer.

The second term in the left of the I equation also has this proportionality of DI to I , and this is due to infected persons becoming noninfective due to either recovery or death, with a speed represented by δ . But of course it also has the first term, representing the transfer of the new infections into the infected pool. As a whole, the SI equations seem to represent an *anti-buffer*, but only because the reaction speed of the buffer itself changes over time, in this case for the worse. We will return to simple first order buffers like these when we come to the chapters on how to use the data to estimate a buffer's characteristics.

1.1.3 Filling a Container

Figure 1.2 shows the level of a fluid in a tray within an oil refinery distillation column before and after a valve setting is changed. This is a simple input–output system that resembles the structure of the smallpox data following the onset of vaccination. In this case, however, we have much more detailed data on both the input (V) and output (T) processes. We see that a change in valve setting at time point 67 results in a change in fluid level in Tray 47, and this change moves virtually instantaneously. The slope of the tray level then rises steadily from this initial large value to find a new equilibrium. This is what we would expect if we increase the flow into a tank with a constant pressure-driven outflow: a rapid rise which levels off as the level rises and the pressure on the outflow increases until it matches the new inflow rate.

We will see that the simple differential equation

$$DT = -\beta T + \alpha V \tag{1.4}$$

does a fine job of representing these changes. Using our trick of estimating β as the reciprocal of the time for $2/3$ of the change to be realized, we see that τ is roughly 50, implying the value $\beta = 0.02$. The total change is about 4.5 level units, and we call this the *gain* of the system. It turns out that the gain, denoted by K , is $K = \alpha/\beta$, suggesting that $\alpha = K\beta$, or about 0.09. We will return to these data in a number of chapters as simple dynamic system furnished with abundant data.

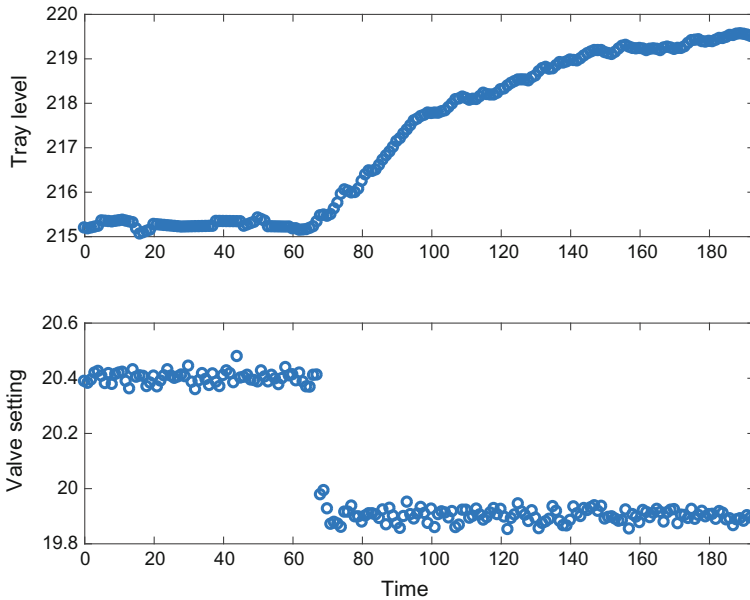


Fig. 1.2 The *upper panel* shows the fluid level in a tray that receives input from another container when a valve is opened. The *lower level* shows the setting of a valve before and after being opened

1.1.4 Head Impact and Brain Acceleration

The data in Fig. 1.3 were collected in a study to measure the effects of motorcycle accident on the driver's brain tissue. They were measurements of acceleration of brain tissue within the cranium of a corpse before and after being struck by a blunt object with a force typical in a collision by a motorcycle driver's head with a hard surface. Analyses of the data have been reported by many authors, including Härdle (1990) and Silverman (1985). This, too, is an input/output system, although we had to construct the input ourselves by representing it as an abrupt but brief pulse of force positioned at the point where the hammer encountered the cranium.

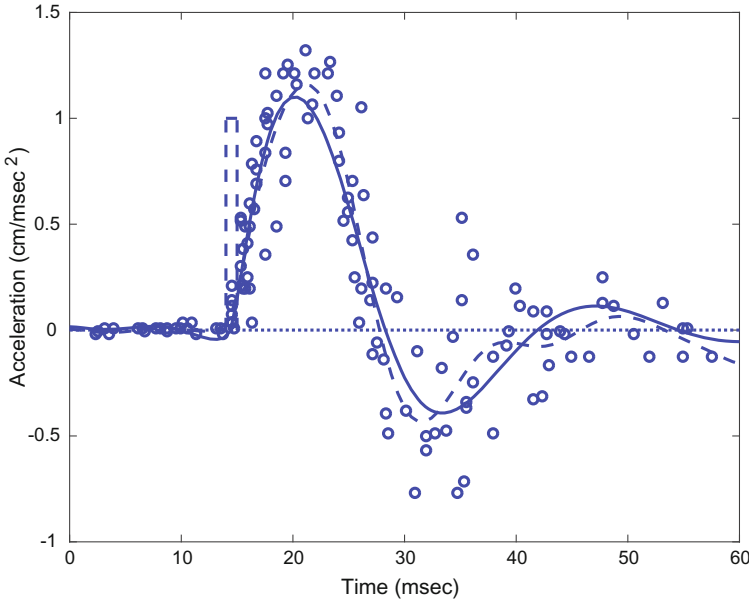


Fig. 1.3 The *circles* indicate observations of acceleration of brain tissue from five replications of an experiment involving striking the cranium of a corpse with a blunt object. The *box* was constructed to represent the impact itself spread over one time unit. The *solid line* is the fit from the data for $\rho = 0.998$ and the *dashed line* is the fit for $\rho = 0.5$

The shape of the data indicates that there is a rapidly decaying oscillation in acceleration after the impact. Anyone who dipped a paddle into a still lake early in the morning knows that this is natural viscous fluids like brain tissue that cannot be compressed, but can dissipate energy by wave action. We will learn that this requires a system that uses the second derivative. We therefore analyzed the data using a three-parameter damped harmonic equation

$$D^2x(t) = -\beta_0x(t) - \beta_1Dx(t) + \alpha u(t) \quad (1.5)$$

where u is the box function located at the impact time and with unit height and width which is displayed in the figure. Positive parameter β_0 defines the period or frequency of oscillation, and is often called the *stiffness* of the system. Parameter β_1 determines the rate of decay of the oscillations, and thus plays a role similar to that of the parameter in the smallpox equation (1.1). Impact pulse u is an external input to the system described by the first two terms on the right, and is often referred to as a *forcing* term. Parameter α is essentially a regression coefficient that determines the amplitude of the oscillations, sometimes called the *gain* in the system.

1.1.5 Compartment Models and Pharmacokinetics

The decline in smallpox deaths can be thought of as having the properties of a bath tub, where deaths are like water going down its drain. Perhaps you have noticed that when the plug is pulled, the water level at first goes down rather quickly, but that it drains out more and more slowly as the water level drops, so that you wind up waiting with shivering impatience for the dirty water to get out so that you can begin the cleanup. This happens because the pressure at the drain opening is proportional to the weight of the water above it, so that change in water level is proportional to level. First order dynamics, in short.

The uptake of drugs by the body is often a series of compartments, with the drug passing along from one to another. An injected substance passes from the tissue to the blood stream, from the blood stream to a target site such as the brain, from brain back into the blood stream, from the blood into the liver and so on. Each of these containers has its own elimination rate or drain size; and in each either all of the drug may be passed on, or some fraction may be absorbed or broken down and the balance eliminated. Simple first order dynamics seem to serve well to describe most of these systems, and pharmacologists find systems of first order constant coefficient differential equations highly useful in the discipline of pharmacokinetics (Fig. 1.4).

Suppose that at time 0 a drug is injected into a tissue site with concentration C_0 , and that its concentration $C_a(t)$ in the blood stream declines at a rate proportional to its level, that is

$$DC_a(t) = -\beta_a C_a(t) .$$

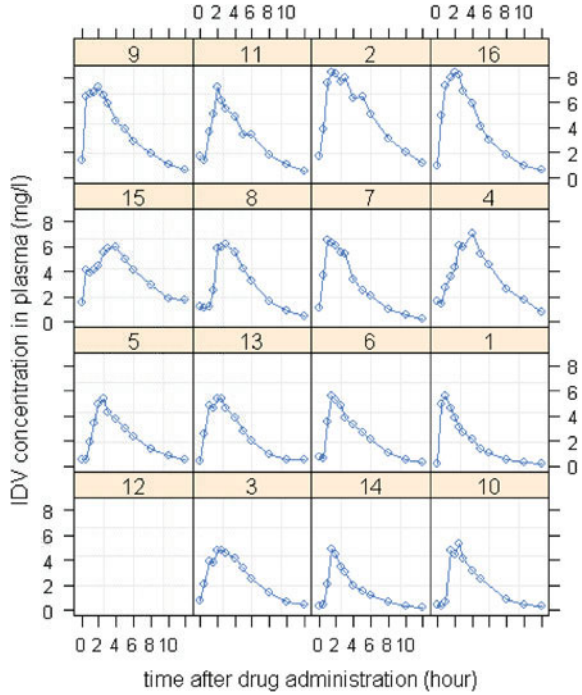
The decline in blood concentration is due to the drug being taken up by the target organs, and then eliminated by the liver with speed β_b , so that

$$DC_b(t) = -\beta_b C_b(t) + \beta_a C_a(t) .$$

Note that $C_a(t)$ is a *forcing function* for the second equation.

It is typical that the absorption dynamics are rather faster than the elimination dynamics, and Fig. 1.5 shows how blood concentrations in fifteen patients increase rapidly at first, and then decline more slowly as the elimination process begins to dominate the absorption process. The Figure also illustrates the possibility that differential equations can vary randomly. We expect that constants such as β_a and β_b will vary from patient to patient, and so will the initial dose C_0 to at least some extent. That is, we often see differential equations in *multilevel* data structures where we have to model variation across levels (patients in this case) as well as within-level variation. Differential equations having randomly varying aspects are referred to as *stochastic*, although we have to be careful with the term *stochastic differential equation* since the phrase has been co-opted by financial analysts and probability theorists to refer to something quite specialized.

Fig. 1.4 The uptake and elimination of a drug for 15 subjects as reflected in its concentration in the blood stream



These equations as well as the SIR equations are examples of *mass balance* systems where something is *conserved* in the system, so that each equation has to account for what happens to various fractions of that which remains constant. The central place occupied by differential equations in astronomy, chemistry, biology and physics is due primarily to conservation principles that imply that what goes in must come out somewhere, and which enforce smoothness of the time-varying aspects of the system. Conservation of momentum in mechanics, heat in thermodynamics, mass of chemical species in chemistry, volume in fluid dynamics and energy in all fields are examples. Even in phenomena as seemingly ephemeral as stock prices, total wealth is conserved as long as any stock can be sold for its paper value.

1.1.6 Chinese Handwriting

The dynamics of the human body are, for most of us, endlessly fascinating. We are far from the fastest or the most skilled creatures on the planet, but when we watch ourselves drive home a goal on the soccer field or produce an arpeggio on a piano, the universe seems for those few seconds to collapse. How do we do this?

When we learn a movement, such as, say, handwriting, we at first make a few clumsy strokes, and then, again and again, until the shapes that we have in mind begin to emerge. Perhaps faster and with more precision for some of us than others, but by secondary school the production of script has been mastered, and one does it without needing any appreciable concentration.

Consider the Chinese script displayed in Fig. 1.6, which is the four characters for “statistics”. This is the first of 50 replications of this script on a horizontal surface by Dr. Xiachun Li, a postdoctoral fellow working with one of us in 1996–1997. A small infra-red emitting diode was attached to the tip of the pen, and three cameras mounted on the recorded pen position 400 times a second with a typical error of about 0.5 mm. The total time for the production of the script varied slightly around 6 s. The experiment and some analyses were reported in Ramsay (2000).

The script includes these elements:

- strokes, curved or straight and in contact with the paper
- cusps, when the pen comes to a near standstill
- lifts, a stroke while the pen is more than 2.5 mm off the paper

On a hunch, we plotted the numbers from 1 to 46 at the times in the middles of 46 equal-sized time intervals, each 130.5 ms in length. It is striking that each numbered point corresponds closely to an event time, whether in the center of a curved stroke or a lift, or precisely at a cusp.

This constancy of event times for a well-learned sequence of movements has become quite familiar to us in our various investigations of human motion, although they are usually around 120–125 ms. The slight increase is no doubt due to the fact that the writing was executed while standing and on a surface substantially larger than would be used in normal circumstances.

We have conjectured that the brain achieves synchronized activation of many muscles by using a stable clock cycle of about this duration to fire a discrete pulse of neural activity over the many channels involved once each cycle. At the muscle end, the arrival of bundle of spikes resets the tension in the muscle fibre. It is well known (Kandel 2000) that the natural motion of the limb affected will be harmonic with little damping, and that the period of this oscillation will depend on the tensions in the attached muscles.

What would happen, then, if we saw the brain/arm/hand system as a single harmonic oscillator, each coordinate of which is defined by the equation $D^2x(t) = -\beta x(t)$, where D^2x is the second derivative of variable x and β defines the period of oscillation along the coordinate direction? We would need two oscillators, one for the horizontal X direction and the other for the vertical Y direction. We could add one for Z where the lifts play out, but let's not, given our two-dimensional viewing surface. This starts us off with two positive parameters, β_X and β_Y , to estimate.