

Computer algebra in the life sciences

Michael P. Barnett

Department of Chemistry, Princeton University, Princeton, NJ 08544-1009

michaelb@princeton.edu

Abstract

This note (1) provides references to recent work that applies computer algebra (CA) to the life sciences, (2) cites literature that explains the biological background of each application, (3) states the mathematical methods that are used, (4) mentions the benefits of CA, and (5) suggests some topics for future work.

INTRODUCTION

The use of computer algebra (CA) in biology and related subjects is increasing rapidly, both in variety and extent. Most of the work is reported for expert audiences in a burgeoning array of specialist journals. This note lists a selection of papers that mention applications of CA in the life sciences. These impact virtually every area of biology and medicine that uses mathematics and statistics.

In health care alone, CA has been used in work that bears on

1. cancer [1, 2, 11, 113, 114, 138, 181, 182, 188, 204, 219, 220, 235, 237],
2. the cardiovascular system [49, 50, 106, 172],
3. drug design and testing [69, 72, 84, 152, 153],
4. EEG scans [234],
5. food poisoning [41],
6. genetic counseling [64, 188],
7. the immune system [2, 46, 55, 100],
8. medical imaging [163, 183, 190, 208],
9. mental health [11, 108],
10. prosthetic devices [122, 162] and
11. trauma care [54].

The most popular CA software platforms are MAPLE [79] and MATHEMATICA [225]. Other systems that have been used in life science work include AXIOM [120] (formerly called SCRATCHPAD), MACSYMA [109], REDUCE [107], and the specialized polynomial system SINGULAR [96]. Very intense symbolic calculations are supported by FORM [212] and SACLIB [53], but I have not found applications of these to biology. The numerical system MATLAB [89] has an add-on that accesses MAPLE. Combinatorial and graph theoretic applications are coded in a variety of languages.

CA simplifies algebraic expressions and performs symbolic integration and differentiation. Also, it supports unrestricted precision arithmetic. This, in turn, supports the use of numerical algorithms which lose too much accuracy in fixed precision calculations. The symbolic results of CA are often fed into procedures that evaluate formulas and solve equations numerically. Symbolic calculations are combined with statistical operations and massive file handling in applications to data analysis. Several features of CA programming languages, used in conjunction with graphics and character string operations, make it easy to construct lists of chemical formulas and other diagrams. MAPLE, MATHEMATICA and other CA systems are evolving, accordingly, into “full service” software that interfaces all these kinds of operation smoothly. As a result, these systems are used increasingly for numerical and other work that does not include actual symbolic operations.

Besides supporting the methods of classical algebra and calculus, CA has stimulated work on the mathematics of Gröbner bases, differential and Clifford algebras, quantifier elimination, geometric visualization and other novel techniques. Basic CA and these more advanced methods bear on topics in the life sciences that range from the structure and dynamics of biomolecules, through the anatomy and physiology of complete organisms, to the short and long term behavior of entire populations of humans and other species. Both users and developers will benefit from a shared awareness of the breadth of CA and its applications.

Sections 1–10 of this survey cite applications to

1. clinical medicine, which includes the control of equipment, the analysis of data and other diagnostic methods,
2. bioengineering, which includes prosthetic robotics, computer vision and ergonomic design,
3. public health issues, which include risk analysis, survival analysis, drug testing, epidemiology and related work that is studied probabilistically,
4. population dynamics — how the abundance and spatial distribution of populations varies with time — and some related problems,
5. population and evolutionary genetics based on classical and molecular theories of heredity (these include population issues deferred from Section 4),
6. physiology and related work on the behavior of cellular substructures and complete cells and organs, determined by the laws of physics and chemistry,
7. biochemical kinetics, which deals with the rates of metabolic processes — primarily systems of enzyme reactions that are in a pseudo-steady state,
8. molecular kinematics, which deals with the movement of atoms in the molecules,
9. molecular and crystal structure, which deals with the arrangement of atoms in crystals, biological molecules and supramolecules,
10. some general statistical methods that provide infrastructure for several earlier sections.

The final sections comment on

11. the benefits of using CA, and
12. the development of the survey.

I have tried to follow the organization of major textbooks by, *e.g.*, Bulmer [37], Edelstein-Keshet [68] and Murray [175] where possible. None of these covers the entire range of topics that apply CA in the life sciences, however. And, while there is considerable overlap in the topics addressed by the major texts, they “slice the pie” in different ways and sometimes take different approaches. The recent text by Yeagers, Shonkwiler and Herod [227] provides

broad coverage that touches on most sections of this survey. The explanations are very clear. They are supported by suggestions for further reading and MAPLE code [111] that produces numerical and graphical output and is maintained.

The bibliography that I have compiled refers to individual applications, to explanations of the underlying natural science and mathematics, and to related CA examples from chemistry and physics. The range of material is enhanced by the varied origins of the mathematical methods. These include

1. biometrics — dominated by statistics, probability and stochastic theory,
2. computational physics and chemistry — dominated by differential equations and linear algebra,
3. control systems — topics similar to those in computational physics and chemistry, but often couched in different terminology and with different assumptions of prior knowledge,
4. signal processing — the preceding comment about control systems applies here, too,
5. structural mechanics and fluidics — dominated by linear algebra and differential equations.

Work that uses CA systems for numerical calculation and/or graphical display, without actual symbolic calculation, was omitted. This includes many MAPLE “worksheets” and MATHEMATICA “notebooks”. The MAPLE Application Center website [241] and the MathsourcE Applications site [243] provide extensive information on these topics and related matters. Many more worksheets and notebooks support individual textbooks and papers that describe research and teaching.

While computationally less efficient than languages such as FORTRAN and C for highly repetitive work, the CA languages are much easier to use for “one-shot” investigations, *e.g.* the effect of a parameter change, particularly when there is no closed form solution, and for program development. Because it is extraneous to the objectives of the present survey, the topic will not be pursued here, but the increasing popularity of CA systems for non-symbolic calculation is an important consideration in the design of future programming languages. It requires further discussion.

Examples of particular mathematical methods are distributed through the sections that deal with applications as follows.

1. Clifford algebras — crystal structure in §9.2 [90],
2. counting methods — population genetics in §5 [64].
3. differential algebra — identifiability problems of population dynamics, physiology and enzyme kinetics in §§4–7 [154].
4. differential equations — the entirety of population dynamics in §4 and much of the physiology and enzyme kinetics in §§6, 7.
5. differential geometry — computer vision in §3 [190].
6. discrete and computational geometry — computer vision in §3 [208], phylogenesis in §5 [66], crystal structure in §9.2 [77, 199].
7. graph theory and combinatorics — phylogenesis in §5 [71, 250], crystal structure in §9.2 [77].
8. Gröbner bases — robotics [122] and computer vision [183] in §3, almost all the enzyme kinetics in §7, molecular kinematics in §8 [69, 152, 153]. and mathematical statistics in §10 [65, 184].
9. quantifier elimination — epidemiology in population dynamics in §4 [46].

10. stochastic methods — medical diagnosis in §1 [49, 50, 108], prognosis in §2 [1, 2, 181, 182], evolutionary processes in §5 [210, 238] and models of cell behavior in §6 [235, 237].

The use of algebraic simplification is taken for granted and mentioned only when it is extremely lengthy. Most of the specialized mathematical terms that are used can be found in the encyclopedia [118], and a readable overview of many of the terms used in statistics, probability and stochastic theory is given by [209].

This survey is intended

1. to alert specialists in CA to the applications of their work in the life sciences, in part to help them determine resources that will be needed in the future,
2. to establish mutual awareness of problems that different branches of the life sciences have in common, and solutions that are transferable,
3. to provide ideas for curriculum development.

In this last regard, a recent “careers” article [110] in one of the most widely read popular science magazines discussed the demand for “bioinformaticians . . . comfortable with both biology and computer science” and on the need for them to maintain currency in the CS component. Academic programs that feature “applied” CA seem more numerous in Europe than in the U.S.A at present. New biological topics that I think CA can help are also emerging, *e.g.* “metabonomics”, that applies nuclear magnetic resonance to the study of drug action and gene function [177].

1 CLINICAL MEDICINE

1. Clark [49, 50] considers the use of probabilistic decision trees and logistic regression in medical decision making. He discusses phlebitis during pregnancy as an example. He uses Taylor series and differentiation, and provides MATHEMATICA code. The background is in [112].
2. Cook *et al* [54] discuss the monitoring of the oxygen uptake of a severely burned patient. They analyze the results of indirect calorimetry using a first order autoregressive model. Symbolic differentiation leads via inversion of a Hessian to a contour plot of a joint posterior distribution.
3. Heckerling [108] uses three-way receiver operating characteristics of signal detection theory to resolve ambiguous diagnoses. The calculations include matrix operations and differentiation. He provides MATHEMATICA code. The background is in [94].
4. Rebbeck *et al* [188] consider improvements in the ability to detect or reject genetic linkage in human cancer, based on the model of recessive oncogenesis and the use of tumor genotype data. They explore a modified lod score method of linkage analysis and use MATHEMATICA to differentiate and simplify the maximum likelihood estimates of the score.
5. Stampanoni [204] uses unrestricted precision arithmetic in the reduction of clinical X-ray data.

A Bayesian approach to modeling in medical decision making is discussed by Parmigiani [178]. It involves derivatives and integrals for which CA is helpful.

Automated medical imaging is discussed in §3.

2 PUBLIC HEALTH AND SOCIAL PSYCHOLOGY

CA has been used in statistical studies of the causes, prevalence, prevention and treatment of disease. Some authors now describe work of this kind as “risk analysis” [58, 217]. The older term “survival analysis” is used for stochastic modeling of these issues [145].

1. Aalen [1] considers (i) the incidence of cancer in a population (Armitage-Doll model), (ii) the spread of an epidemic and (iii) the interval between the birth of the first and second child in a family as examples, in the context of a homogeneous time continuous Markov chain model based on phase type distributions used in queueing theory. The computation uses Taylor series to construct survival functions.
2. Aalen [2] considers (i) diseases such as breast cancer, where the rate of progression is related to the age at which it begins, and (ii) the incubation time of AIDS, as examples in a related model that involves mixing (frailty). He calculates functions of an intensity matrix, Laplace transforms, eigenvalues of an acyclic Markov chain and survival distribution functions for particular models, symbolically.
3. Baglivo *et al* [11] consider (i) prostate cancer (to determine the effectiveness of five preoperative variables in predicting the extent of the disease), (ii) a rare type of sarcoma (to determine possible relevance of sex and age to occurrence), and (iii) response of psychotic patients to a neuroleptic drug (to compare the relative effectiveness of different injection schedules), using a generating function approach to handle permutation distributions. They use recursive algorithms related to Fast Fourier Transforms to manipulate polynomials. The method facilitates sensitivity analysis.
4. Burmaster [39] conducts risk analysis on the health effects of fish consumption and numerous other interactions of people with their environment. He uses symbolic differentiation and integration.
5. Campanella and Peleg [41] model the destruction of botulin producing bacteria. The calculation includes some symbolic substitution, followed by numerical integration.
6. Gittins and Pezeshk [84] discuss the determination of optimal sample size in pharmaceutical quality control and clinical trials.
7. Hunka [116] and Hunka and Leighton [117] discuss a statistical problem that the later paper illustrates by an academic proficiency study. They consider regions of significance in three-covariate ANCOVA problems using MATHEMATICA. The work includes the manipulation of design matrices and polynomial operations.
8. Kolassa [138] uses an adjusted profile likelihood and the saddlepoint approximation in statistical studies of bladder tumors and endometrial cancer. The calculations include symbolic differentiation, array manipulation and numerical minimization. MATHEMATICA code to handle cumulant generating functions is included.
9. Pérez-Ocón *et al* [181, 182] develop both homogeneous and non-homogeneous time continuous Markov chain models to study the relative efficacy of different treatments for cancer. They use differentiation and integration to deal with the forward Kolmogoroff equation.
10. Waller *et al* [219, 220] consider the detection of epidemic leukemia, using focused tests of clustering. They derive analytic power functions for three of these tests. Also, they use numerical inversion of characteristic functions that uses fast Fourier transforms and requires high precision.

Several other papers on risk analysis mention MATHEMATICA, but just use it for numerical computation and plotting. Recent texts on the use of CA in statistics at large show a trend that risk analysis is likely to follow (see §10). The general background of stochastic epidemic models of infectious diseases is discussed in [3]. Work on epidemiology that uses the methods of population dynamics includes studies of AIDS by Chauvin *et al* [46], Corless [55] and Gupta *et al* [100], discussed in §4.

The recent introductory text on MAPLE for environmental scientists by Scott [198] contains a variety of elementary examples relevant to the field.

3 BIOENGINEERING

Medical prosthesis is a major objective of robotics research. CA has been applied extensively to the inverse kinematics of robot limbs. The standard approach follows the Denavit-Hartenburg formulation that is described in many texts, *e.g.* [176]. CA is used in simulated human movement by several authors.

1. Ju and Mansour [122] consider human gait and foot action, using Kane's method and MACSYMA.
2. Megahed [162] considers human arms, using a tree structure and the homogeneous transformation method.

The problem has also been modeled using stochastic differential equations that are amenable to CA ([135], Chapter 7) — see §10.

Computer vision is of paramount importance in the development of mechanical devices to aid the sight-impaired. The automated inspection of clinical data produced by medical imaging and microscopy, both off-line and in real-time, should lead to substantial savings in the time and, hopefully, the cost of medical diagnosis and treatment. Automated inspection of satellite images assists many ecological studies. Several computer vision projects have been reported that use CA.

1. Petitjean [183] used algebraic geometry, Gröbner bases and resultants.
2. Romeny [190] reports an extensive project that is in progress. He uses differential geometry, algebraic invariance, topology, group theory and probability. Also, he stresses the necessity to unify the computation of formulas and numbers and the processing of large bodies of data. He gives MATHEMATICA code. This uses differentiation, integration, matrix and other algebraic operations and pattern matching. The background material includes [73, 189, 147].
3. Theobald *et al* [163, 208] consider visibility problems, using the methods of discrete geometry. They used the SINGULAR system [95] to manipulate and factor polynomials.

The last item provides a link, that is easy to follow, between the life sciences and the overall field of computational and discrete geometry. There is an abundance of work on the latter subject. Goodman [91] provides encyclopedic coverage of the mathematics. Greuel [95] provides SINGULAR code for a selection of algorithms. Much of the work uses the term CA as a descriptor of mathematical methods but does not mention life science applications explicitly. These algebraic methods have had a major impact on the geometry because many of the algebraic notions have geometrical meaning, *e.g.* term orders in Gröbner bases correspond to polyhedral operators.

Visibility considerations and mathematically equivalent problems can affect the life sciences in many ways, that include

1. computer vision, as mentioned above, *e.g.* when orienting a perception device by reference to marker objects in a known scenario,
2. tomography, *e.g.* to construct two-dimensional views of three-dimensional structures,
3. penetrability in radiation therapy,
4. accessibility by linear extension of a prong or equivalent object in systems ranging from the sub-cellular to the skeletomuscular,
5. concealment in models of animal behavior,
6. steric hindrance in conformational studies of molecules.

Three dimensional modeling from X-ray and other imaging data has an extensive literature that includes considerable work on “algebraic reconstruction techniques” that are actually numerical — see, *e.g.* [124]. In other work on bioengineering

1. Toet [34] used automatic integration in relation to the distribution of light on the cornea of aircraft pilots subject to laser glare,
2. Moore [106, 172] studied the fluidics of cardiovascular action in a MATHEMATICA calculation that can be extended symbolically.

4 POPULATION DYNAMICS

Population dynamics deals with the abundance and spatial distribution of one or more species in an ecological environment. Spatial considerations include invasion, diffusion, dispersal, arrangement and the study of “metapopulations” [103]. Many different models are used to describe the combined effects of reproduction, migration, and death, and the factors that affect these. The texts by Edelstein-Keshet [68] and Murray [175] develop these deterministically, within a more general context. Kloeden and Platten present stochastic models ([135], Chapter 7) — see §10. The work is dominated by differential and difference equations. Also, integro-difference equations are being applied to discrete time-continuous spatial models [139]. Many of the models are dynamical systems. Several models of population dynamics are similar to models of biochemical and physiological systems. For example, cell populations are treated as dynamical systems in some models of cancer. The text by Bulmer [37] deals with deterministic and probabilistic aspects of population growth. It is supported by extensive MATHEMATICA notebook material [240]. Modern methods of geostatistics [48] are also relevant.

Key issues in population dynamics include

1. continuous population models for a single species that reproduces continuously (this includes the logistic model that assumes exponential growth is curtailed),
2. population growth that is treated on a generation by generation basis, *e.g.* because there is a discrete breeding season, treated by difference equations,
3. population growth of competing species — the predator-prey models that are formalized as Lotka-Volterra systems,
4. biological oscillators and waves,
5. population movement.

Another major approach to population dynamics uses matrix models [37, 43]. Applications of “Leslie-Lefkowitz” models include conservation ecology and management, in which the effects of changing mortality rates, *e.g.* by harvesting, on different age classes can have radically different effects.

Miller and Wethey give instructional MAPLE prototypes for

1. integrating the basic equations of population and ecological modeling [164] — the background is in [156, 157],
2. a competition model (two species compete for the same food or other resource) [165] — the background is in [151],
3. the dynamics of arthropod predator-prey systems [166], the background is in [68],
4. the game theory approach to evolutionarily stable strategies [167] — the background is in [158],
5. the analysis of the chemostat system [168]— the background is in [68],

6. population growth and the earth's human carrying capacity [169], the background is in [51].

Further MAPLE prototypes are given for

1. a model of AIDS epidemiology by Corless [55],
2. Hopf bifurcation in a predator-prey model by Forwalter-Friedman [75],
3. the logistic model of population growth by Schwalbe [197].

In the journal literature

1. Chauvin *et al* [46] modeled the dynamics of a disease, *e.g.* AIDS, applying quantifier elimination theory to a parameterized system of non-linear PDEs. This leads to the manipulation of polynomial equations. The authors provide a full explanation of the underlying methodology.
2. Gupta *et al* [46] also modeled the dynamics of AIDS.
3. Margaria *et al* [154] apply compartmental analysis and identifiability methods (see §6) to microbial growth in a batch reactor,
4. Murray [174] surveyed krill in Antarctica acoustically — although this paper only applies MATHEMATICA numerically, he uses unrestricted precision arithmetic in related work.
5. Pecelli [180] considers prey-predator systems with delay and obtains Hopf bifurcations and stable oscillations.
6. Woodward [226] uses a dynamic nutrient model to help farmers in New Zealand plan their fertilizer expenditure.
7. Viscido *et al* [213, 214, 215] discuss problems of the “selfish herd” and foraging by animals, based on field observation of fiddler crabs.

Work that applies CA to genetic aspects of population dynamics is covered in the section that follows.

5 GENETICS

Within this overall field,

1. population genetics and quantitative genetics use ideas that predated the discovery of DNA,
2. molecular genetics is based on the present understanding of the double helix.

Applications of population genetics include the breeding of plants and animals and some forms of genetic counseling. Animals and plants consist of cells. Each cell contains a nucleus. Each nucleus contains chromosomes. These can be seen through a microscope. A large amount of experimental data is explained by assuming that certain objects, called genes, are arranged at loci along the length of each chromosome. The genes at a particular locus can take a small number of variant forms, called alleles. Chromosomes are paired (with an exception that can be skipped for present purposes).

Certain biological traits with discrete values, such as blood type, are determined by the pair of alleles at the same locus on two paired chromosomes. During reproduction there is an equal chance for either of two paired chromosomes in the father to be replicated in the child. The same is true of a pair of chromosomes in the mother. As a result, if the father's chromosome contains two A alleles at a particular locus (*i.e.* the father is A/A in a standard notation), and the mother is B/B at the same locus, then each child will be A/B. When a large number of pairs of

just these children mate (which can happen for certain plants and animals) the next generation will be A/A, A/B and B/B, in the ratio 1:2:1.

In a more general situation, this simple model predicts that if a fraction p of the alleles in a large, randomly mating population are A, and $q = 1 - p$ are B, and these traits are not subject to selection, then in ensuing generations, the population will be in proportions $p^2 : 2pq : q^2$ for A/A, A/B and B/B (“Hardy-Weinberg” equilibrium). Also, in the very simple model, alleles on the same chromosome accompany each other through successive generations. This is called “linkage”. In the actual course of events, however, a portion of one chromosome often gets replaced by the corresponding portion of its pair. This is called “recombination” (related terms include “crossover” and “Hardy-Weinberg disequilibrium”).

As mentioned earlier, there are two large areas of genetics that do not deal directly with DNA structure. Population genetics deals with questions of change in allele frequency by factors that include migration, mutation, population/subpopulation interactions, drift, selection, effective population size and inbreeding effects. Quantitative genetics shows how traits which vary continuously can be analyzed in terms of classical Mendelian genetics when multiple genes affect a single trait. Typically, a census is taken of some genetic traits in a sample of an animal or a plant population, and statistical methods (especially analysis of variance) are used to determine the frequencies of different alleles, and the extent of recombination.

Genetics expanded vastly in further directions as a result of the work of Franklin, Wilkins, Crick and Watson on the molecular structure of the DNA that constitute chromosomes. Molecular genetics focuses on the sequences of building blocks that make up the chromosome, at a vastly greater level of detail than the individual gene. It should be noted that although genes occur in a particular sequence along a chromosome, “gene sequencing” (genome reconstruction) deals with the much more detailed issue of DNA building blocks. (see *e.g.* [78]). Forensic DNA testing is yet another matter [52].

Mutations are changes in the molecular structure of the DNA. The causes include irradiation. Some mutations are extremely localized, consisting of a change in a group of less than 20 atoms of carbon, hydrogen and other chemical elements, called a nucleotide, that constitutes part of the fundamental building block. Other changes involve the insertion of new material or deletions. Most are harmful, but others play a vital role in evolution. The similarity of the sequences of the building blocks in the DNA of two species gives a measure of their “phylogenetic distance”, *i.e.* their closeness within a common line of descent or within collateral lines.

Weir [221] describes the background of a considerable body of work on evolutionary and population genetics. His book presents the mathematics and many practical applications. It mentions some unpublished applications of CA to the statistics of recombination, specifically

1. variances of maximum likelihood estimates of gametic disequilibrium at four loci by Weir, Golding and Lewis (see p. 119),
2. variances for the trigenic and quadrigenic coefficients by Brooks (see pp. 124, 127).

CA work is mentioned in several journal and web articles.

1. Balloux *et al* [12] discussed the choice of markers (reference points along a chromosome) when determining the probabilities of genetic diversity. They used the common shrew for illustration. Mutation rates are computed by symbolic integration. The background is in [221].
2. Chang and Rausher [44] discussed outcrossing (cross pollination) in the Morning Glory. They used symbolic matrix manipulation and differentiation. The background is in [221].
3. Denniston [64] discussed equivalence by descent. This is of concern *e.g.* in deciding the probability that particular genes would occur in a child, given certain information about the relatives of the potential parents. He uses enumeration methods. The background is in [57]. MAPLE code is available.

4. Kirkpatrick *et al* [131, 132, 133] discussed sexual selection in relation to the concept of “good genes” and a particular mutation model. They used CA to construct recurrence schemes for counting operations, to construct Taylor series and to evaluate covariate and fitness function integrals. They also discussed cattle breeding and used surd arithmetic, matrix operations and orthogonal polynomials.
5. Peccoud and Ycart [179] discussed a Markovian model for gene induction. They computed the distribution of protein number for transient and stationary states and showed how to estimate parameters of the model. MATHEMATICA was used to construct a linear system of differential equations with constant coefficients and to find symbolic solutions.
6. Rebbeck *et al* [188] used the model of recessive oncogenesis in statistical studies of genetic linkage in cancer, mentioned in §1.
7. Turelli and Barton [210] discussed statistical and multilocus population genetic analyses of the effects of selection on polygenic traits. They used MATHEMATICA to apply recurrence schemes and to compute cumulants from moments symbolically.
8. Zheng [238] discussed the dispersion index of the molecular clock. This is a genetic process that is used to measure phylogenetic distance and the rate of evolution. The model is a continuous time four-state Markov chain. Zheng manipulated integrals and inequalities to solve coupled first order differential equations. The background is in [207]. Several related enumeration problems were mentioned. These are suited to CA studies.
9. Miller and Wethey [171] provided a MAPLE worksheet for the “r- and K-” approach to the problem of natural selection [194].

The books by Lange [144] and by Lynch and Walsh [150] provide detailed accounts of the basic mathematics that complement the Weir text cited earlier. All three books mention numerous topics that seem amenable to CA methods. The recent text on mathematical statistics using CA [193] contains two chapters on maximum likelihood estimation — a topic of major concern in quantitative genetics.

A large, rapidly expanding body of work uses combinatorics and graph theory to construct phylogenetic trees. Some authors place their work within the field of discrete and computational geometry. “Up to date information on present knowledge of (belief in) the detailed branching structure of species evolution” (Dress and Terhalle [66]) is maintained in the “Tree of life” web site [250] that links to many other sites in turn. Felsenstein provides links to nearly 200 phylogeny programs [71]. Pursuing this literature here would swamp the survey and could not be attempted.

PERL seems to be the most popular language for coding genome reconstruction algorithms at present. The power of some of the CA systems to manipulate strings suggests their use as alternatives, that can be read and adapted by much larger audiences.

6 PHYSIOLOGY AND RELATED FIELDS

1. Grotendorst *et al* [98, 99] used CA to analyze NMR work on transport and diffusion across living cell membranes (see §9.1).
2. In the same group, Zhao *et al* [234] modeled electroencephalography (EEG) measurements of the brain, as a problem in electrical potential theory. They manipulated surface harmonics, infinite series and derivatives.
3. Huggins [115] applied statistical methods to cell lineage data. He discussed identifiability of measurement error in the bifurcating autoregressive model, that he explained. He used automatic differentiation and matrix manipulation.

4. Kuchel *et al* used NMR to study erythrocytes in the living cell (see §9.1).
5. Williams and Jeffrey [224] studied the growth of cells in the geniculate nucleus — the part of the middle brain where the optic nerve fibres from the two eyes meet. They used integration and differentiation.
6. Zheng [235, 237] discussed two models of carcinogenesis as stochastic problems involving the population of cells in the body. In [235] he constructed the cumulant from a probability distribution function by expanding a series and equating terms to create an ODE. In [237] he converted the PDE for the cumulant generating function to a system of ODEs for the cumulants. In both cases he solved the ODEs numerically. He gives MATHEMATICA code. The background and analogous problems that are amenable to similar treatment are described in [155].

A growing body of work expresses physiological and metabolic processes as studies in compartmental analysis [42, 119], *i.e.*

1. systems where substances flow between different compartments, that may be sites or regions of a cell (as in sequestered reactions), or entire cells, or even larger units of a living plant or animal, and
2. systems involving sequences of reactions that occur within the same physical space (*i.e.* the compartmentalization is temporal).

This approach is well suited to the tensegrity models of living structures that are of current interest [47].

Engineering control theory provides the background for much of the work on compartmental analysis. This work serves, in turn, as the starting point for discussions of the identifiability problem. This deals with models which assume that the overall rate and the rates of a particular selection of intermediate steps can be measured in a metabolic process. A model is called structurally identifiable if the set that is measured can determine the rates of all the other steps. Global and local identifiability are relatively strong and weak forms of structural identifiability, that depend on further assumptions concerning boundary conditions. A compartmental system is linear when the overall rate of output of the end product is linear with respect to the input. It is non-linear, otherwise.

1. Cobelli's group has contributed extensively to this field — their CA work is cited in §7.
2. Margaria *et al* [154] provided a full and rigorous explanation of the issues mentioned above. They discussed the mathematical and computational approaches. These include differential algebra, Gröbner bases and Taylor series expansions. Case studies include models of bovine mastitis and cattle immunization.
3. Zheng [236] discussed two relatively simple compartmental analysis problems as prototypes of pharmacokinetic and carcinogenesis modeling. He used differentiation and algebraic manipulation without the need for Gröbner bases.

7 ENZYME KINETICS AND RELATED TOPICS

The metabolism of animals and plants is sustained by sequences of reactions that involve certain substances called enzymes. The rate of a reaction depends on the concentrations of the reactants and a “rate constant”. A vast amount of work is directed to

1. determining the metabolic pathways, *i.e.* identifying the compounds that are used and produced in the successive reactions,
2. determining the rate constants of the individual steps.

Most of this work uses a steady state approximation, that assumes an adequate supply of the metabolites. These come directly or indirectly from the nutrients in the life supporting environment.

Chemical kinetics is governed by the law of mass action. In the simplest situation, this calls for a rate that is proportional to the concentration of each reactant. In other situations, the concentrations are raised to integer powers. Some reactions are reversible, and give rise to equilibrium situations. A catalyst assists a reaction without being consumed by it. The basic model of a prototypical two-step enzyme reaction is described by the Michaelis-Menten equation. This is explained in [175] and in many standard texts on physical chemistry. It is also used to describe predator-prey interactions when the predator may become satiated (“type II functional response”).

In most of the work on enzyme kinetics that is handled by CA, the law of mass action is applied to each of the biochemical equations in a metabolic process that has reached the steady state. In the model, each forward and backward biochemical step contributes a polynomial equation that contains rate constants and the concentrations of one or more molecular species. These equations are coupled by the principle of conservation of matter. Boundary conditions are set on the concentrations of the initial reactants and the final products. This turns the problem into a set of polynomial equations in the rate constants. Increasingly, this is treated by the method of Gröbner bases, that constructs an equivalent system of equations of lower computational complexity. Then the rate constants of individual steps are found by solving the reduced equations numerically, using measured values of the rates of (1) the overall process and (2) a sufficient number of subsidiary (possibly multi-step) processes. Generally, the use of Gröbner bases introduces some equations of higher degree but simpler structure (*e.g.* fewer variables). At the extreme, the method mimics Gaussian elimination for linear systems. Recent books that present the method in the context of applications and practical computations include [211].

The following work spans a progression from very simple CA to the limits of present resources. It is reported in journals that deal with general biochemistry, control theory, and the relatively novel field of biothermokinetics [159].

1. Kleene and Cejtun [134] discussed an elementary problem that concerns buffered chemical solutions (*i.e.* that maintain constant acidity). A small set of polynomial equations is converted to a cubic. The authors provide MATHEMATICA code.
2. Geil [83] derived the Michaelis-Menten equation, using MAPLE, for instructional purposes.
3. Jumars [123] modeled animal digestion by using MATHEMATICA to combine the kinetic equations for hydrolysis and absorption in three kinds of chemical reactor, and found the optimal ingestion rate by differentiation.
4. Bennett, Dewar *et al* [32] reported early Gröbner basis work on simple enzyme systems.
5. Bayram [26]–[28] developed this further.
6. Yildirim [228]–[232] is continuing Bayram’s work, and he provides prototype MAPLE code in the first of these papers.
7. Grinfeld, Bennett and Hubble [33, 97] considered the related problem of affinity binding relations.
8. Margaria *et al* [154] applied compartmental analysis and identifiability methods (see §6) to a problem in pharmacokinetics,
9. Raksanyi *et al* [187], Chappell, Godfrey and Vajda [45] and Ljung and Glad [148] reported early work that applies compartmental analysis to enzyme kinetics.
10. Cobelli’s group discuss global identifier problems (see §6) in enzyme kinetics. Results for the linear and non-linear problems were reported by Audoly *et al* in [9, 10] respectively. This is one of the most computationally intense CA applications to the life sciences that I have seen.

11. Schulz and Südi [196] used a graph theoretic approach in work that deals with abortive complexes and random substrate binding.
12. Farza and Chéryu [70] discussed the use of CA in their BIOESTIM system for on-line estimation in bioprocess engineering. It helped the automatic design of state and parametric estimators from minimal knowledge of process kinetics.
13. Queeney *et al* [186] considered chemical oscillations in an enzyme system, treated as an eigenvalue problem. They showed steady states, damped oscillations and stable limit cycles.

Both Edelstein-Keshet [68] and Murray [175] develop systematic treatments of a considerable range of biochemical processes, that are not restricted to the steady state. These are based on the underlying differential equations and can be viewed as part of modern dynamical theory. Almost every section of these books can be the starting point for CA applications.

8 BIOMOLECULAR DYNAMICS/ KINEMATICS

Structural formulas and molecular models are probably the most familiar sights associated with organic chemistry. The planar depictions on paper actually represent three-dimensional structures. Molecular dynamics is concerned with the vibrations and other internal movements of molecules. These are controlled by the force constants that limit departures from the configurations which the molecules would have if they were completely at rest. The movements determine the infra-red and some related spectra of the molecules. CA has been applied to the basic theory of the normal modes of molecular vibrations in [102] and papers cited therein. Movements in proteins are of major interest and are studied by several experimental methods and elaborate molecular dynamic simulations [101]. CA is applicable in this area.

Within the constraints of

1. fixed bond length and
2. limits on possible variations in the bond angles and dihedral angles,

the molecules also can be regarded as linkages. Many large molecules exist in several alternative conformations. A change between these conformations usually provides a net release of energy or requires a net provision of energy. Often, an energy barrier must be overcome to effect the change. However, the bond lengths between two atomic species tend to stay fairly constant among a wide range of molecules. So do bond angles formed by pairs of bonds which join one atom to two neighbors. Although energy considerations are paramount, often it is important to know whether a particular conformation would be consistent with the approximate bond lengths and angles in a molecule (even if the conformation is not the lowest in energy). This kind of conformation question occurs in relation to

1. the puckering of certain cyclic molecules, *i.e.* molecules that contain atoms joined in a ring,
2. the possibility that two groups of atoms in a molecule of a drug can attach simultaneously to two parts of a molecule in a patient's tissue, *i.e.* whether the drug molecule can "dock",
3. the folding of proteins — of fundamental importance in the life process,
4. the behavior of tensegrity structures within larger systems.

Docking is important in non-therapeutic processes, too, *e.g.* it plays a role in a theory of the mechanism of odor [74]. Several authors discuss the conformation of organic molecules by combining

1. the Denavit-Hartenburg formulation of kinematic linkages (see *e.g.* [176]) with

2. the chemical model of an organic molecule as a linkage proposed by Go and Scheraga [88].

The problem reduces to a set of polynomial equations. All the authors use Gröbner bases to solve these.

1. Emiris and Mourrain [69] also focus on drug design and compare the Gröbner basis method with classical resultant methods.
2. Finn and Kavradi [72] consider the issue in relation to drug design. They discuss the shapes of molecular surfaces, conformational search and pharmacophore identification.
3. von zur Gathen and Gerhardt [81] discuss the Gröbner basis treatment of the conformation of cyclohexane as a pedagogic example.
4. Manocha *et al* [152, 153] consider cyclo-octane and peptide chains.

9 BIOMOLECULAR STRUCTURE

9.1 Nuclear magnetic resonance

CA is helping to elucidate the structure of biological molecules in work that uses nuclear magnetic resonance (NMR). This physical phenomenon is also at the heart of medical resonance imaging (MRI). CA is used

1. in the theory that underlies NMR,
2. to analyse NMR data, and
3. to design NMR equipment.

The background is explained in several recent books *e.g.* [29, 76, 146]. NMR is based on the fact that the nuclei of hydrogen (and many other) atoms behave as if they can only spin at certain speeds. These “speeds” depend on the electrical fields produced by the electrons near the nuclei. In an NMR experiment, a radio signal is passed through a sample of the material that is being studied. At a certain frequency, the radio waves give up the energy that is needed to increase the nuclear spins of a particular element. Often, this frequency also depends on the neighbors of the atoms that have their spins changed. The absorption of energy is measured very accurately over a range of frequencies, to produce an NMR spectrum. The results provide extensive information about the structure and identity of the molecules that make up the sample. The experiments are carried out in strong magnetic fields, that align the molecules to make the effect easier to detect.

Recently,

1. Gasparovic *et al* [80] discussed the design of equipment (a shielded gradient probe for high resolution work *in vivo*). They used integration.
2. Grotendorst *et al* [98, 99] computed rate constants for transport and diffusion across living cell membranes from NMR measurements. They solved a linear inhomogeneous system of ODEs with constant coefficients (the McConnell equations) by use of matrix exponentials. This reduced the problem to symbolic matrix manipulation.
3. Kuchel *et al* [125, 141, 142, 143, 173] used NMR to explore the biochemistry and physiology of erythrocytes (red blood cells). Differentiation and integration was used in the context of classical field calculations.
4. Levitt [146] produced hundreds of diagrams, that depict spin states, algorithmically.
5. Straubinger *et al* [205] developed the theory of a novel experimental approach (pulse angle dependence of double-spin-echo proton NMR). They manipulated an abstract operator representation of the problem, to construct formulas to interpret the results.

9.2 Crystal structure

For most of the twentieth century, crystallography dealt with crystals that have a periodic structure — a “unit cell” of atoms that repeats in three dimensions. The arrangement of atoms is found primarily by X-ray diffraction — the lattice of atoms acts as a grid that diffracts a beam of X-rays to produce a pattern of spots on a photographic plate. Determining the structure of the unit cell from these patterns was one of the first major scientific uses of automatic calculating machines, even before stored program computers were invented. The recently discovered aperiodic crystals require new analytic techniques, to determine their structure and to understand the formation and growth of facets. Use must be made of an algebra that can express geometrical relationships in hyperbolic and other spaces, sometimes of dimension greater than 3. The problem is discussed in the literature of

1. discrete and computational geometry: as studies in tilings, *e.g.* by Friedrichs *et al* [77] using graph theoretic methods, and by Senechal [199] who mentions tilings and other topics under the heading of “mathematical crystallography”,
2. Clifford algebras by Gómez *et al* [90], who give MATHEMATICA code.

Most of the aperiodic crystals that have been studied so far are non-biological, but recent papers discuss materials of this kind that include

1. bitumenous hydrocarbons of interest in relation to “the crystallization of life” [233], and
2. biological helices such as collagen and the α -helix, in connection with the extremely important issue of protein folding [195].

9.3 Non- and semi-empirical calculations

These develop approximate solutions of the Schrödinger equation to predict chemical behavior. CA is used increasingly in this field. It includes most of my own recent research. Many quantum chemical calculations deal with molecules of biological interest. CA has not been applied directly to biological quantum chemistry so far, however, although there is considerable potential for such work. The Schrödinger equation for the electronic structure of a molecule contains an operator that specifies

1. the number of nuclei,
2. the charges on the nuclei,
3. the number of electrons, and
4. the putative geometry of the molecule.

It is satisfied by a “wave function” that depends on the coordinates of all the electrons in the molecule. Each eigenvalue is an allowed energy of the system. Minimizing the energy with respect to the geometrical parameters of the molecule gives the optimum structure(s). Differentiating the energy with respect to these parameters gives the force constants that determine vibrational behavior. Differences between energy levels determine the optical and ultra-violet spectra. Field gradients affect NMR spectra. Solving the Schrödinger equation “*ab initio*” for systems containing more than a few electrons requires elaborate approximations and vast computation. A variety of mathematical techniques are used. Almost all of these require the computation of large numbers of “molecular integrals” that comprise the matrix elements of an eigenvalue calculation. The degree of the matrix may exceed $O(10^6)$. Recent texts that explain the background include [56, 121].

Quantum chemistry has been linked to computer algebra since the start of electronic computing. Gray, Pritchard and Sumner [93, 185] and Turner and Boys (see [35]) mechanized the construction of formulas for some relatively

simple molecular integrals in 1956, using array manipulation. I started to develop and use CA software for more of these molecular integrals in 1961 [13, 218] as a direct result of discussions with Boys. My continuation of this work in recent years, and the development of associated software methods [14]–[21] led me to write this note, as an extension of a survey of CA applications in quantum chemistry.

My main concerns, in applying CA to my research, have been to ensure demonstrable accuracy of the results and clarity of derivations. All the formulas that I have developed and reported over the past eight years are part of a growing corpus produced from control files that are input to MATHEMATICA sessions that use my MATHSCAPE procedures. These generate complete documents as well as individual formulas. Each theorem that is copied from a reference source and each new definition is typed just once. Extensive checks are applied to the formulas that are copied from monographs and to the results that are derived. New files refer to earlier files by simple naming and addressing conventions. The files serve an extra role as mathematical audit trails. While these measures cannot exclude all error, they provide far greater confidence than I find possible otherwise, and make inconsistencies much easier to track down and correct. A collegial use of this approach would lead to a coalescence of rigorously dependable material for use in computations, much as networks of standards have emerged in the laboratory.

10 PROBABILITY, STATISTICS AND STOCHASTIC THEORY

For work that uses CA in the application of mathematical statistics and stochastic theory to specific problems in the life sciences see the earlier comments in this survey:

1. in §1, apropos Clark [49, 50], Cook *et al* [54], Heckerling [108] and Rebbeck *et al* [188],
2. in §2, apropos Aalen [1, 2], Baglivo *et al* [11], Burmaster [39], Gittins and Pezeshk [84], Hunka [116], Hunka and Leighton [117], Kolassa [138], Pérez-Ocón *et al* [181, 182] and Waller *et al* [219, 220],
3. in §5, apropos most of the work that is cited,
4. in §6, apropos Huggins [115] and Zheng [235, 237].

The application of CA to statistical problems is helped by a number of teaching aids. These include

1. the early book by Heller [109] which uses MACSYMA,
2. the encyclopedia article by Kendall [128],
3. the book by Tanis and Karian [206] which uses MAPLE, and the accompanying set of MAPLE worksheets by Karian [245],
4. the APPL language of Leemis and his coworkers [87],
5. the recent text by Rose and Smith [193] and its supporting mathStatica software package [244] — these provide a systematic introduction to mathematical statistics, using MATHEMATICA, that contains chapters on continuous random variables, discrete random variables, distributions of functions of random variables, systems of distributions, multivariate distributions, moments of sampling distributions, asymptotic theory, statistical decision theory, unbiased parameter estimation and maximum likelihood estimation.

In the research literature, CA is being used increasingly to develop statistical methods that are reported in general terms without any mention of the specific natural and social science topics to which the results can be applied. The process of connecting these solutions with “real life” problems defines an interesting problem in information transfer, that I am exploring.

Risk analysis provides a rich source of examples that can be used to illustrate the benefits of the new methods, for potential users throughout the life sciences. Risk analysis addresses many issues of public concern, and a large

part of the methodology consists of the traditional components of statistics and probability — see *e.g.* [58, 217]. The `mathStat` package [244] of Rose and Smith contains modules to deal analytically with these building blocks, that include

1. the analytical derivation of moments, variances and other statistics from probability density functions (pdfs) that model individual problems,
2. the analytic construction of partial derivatives, scores, Hessians and other functions of the pdfs, to facilitate parameter fitting and the optimization of maximum likelihood estimates,
3. moment based estimation methods to fit density curves,
4. integration of the convolutions in multivariate models, enabling analytic exploration of parameter variation,
5. inversion of cumulative distribution functions to expedite Monte Carlo calculations.

The modeling methods for medical decision making in [178] require a variety of derivatives and integrals. CA can help evaluate these.

A major area of current research in mathematical statistics concerns the expansion of statistical properties as infinite series. In some cases, these are based on derivatives of the density functions, in Taylor series or in ways that are loosely analogous (as in the saddlepoint approximation — a simple explanation is given in [92]). Expansions are used, *e.g.*, when the integral that represents a particular statistic cannot be evaluated in closed form analytically, even using CA. In other cases, the behavior of a variable is approximated by the normal distribution, and the series expansion provides correction terms for the finite population size. The use of such “Edgeworth expansions” was discussed extensively in the major texts of the mid-twentieth century *e.g.* [126], but it is not mentioned in many of the recent survey texts on probability and statistics. Recent books that deal with the subject include [136] by Kolassa. This contains the MATHEMATICA code that generated some of his examples.

Within the general area of series expansions and related topics

1. Andrews and Stafford [5] discuss asymptotic expansions of maximum likelihood estimates, associated deviances and Bartlett identities. They cite earlier, related work by other authors.
2. Andrews, Stafford *et al* report further work on likelihoods [202, 203], symbolic operators for multiple sums [30], iterated full partitions [6], score functions [63] and intersection matrices [201]. Their recent book [7] extends the work still further, and describes novel mathematical formulations designed to facilitate mechanization. Some of their papers include MATHEMATICA code. They have also developed several MATHEMATICA notebooks. Their websites [252, 251] provide further information.
3. Bellio and Brazzale [31] discuss approximate conditional inference.
4. Butler [40] discusses reliabilities of feedback systems and their saddlepoint approximations.
5. Gatto [82] discusses saddlepoint approximations for the distributions, the likelihood ratio, exponential score and Wald-Wolfowitz tests.
6. Harvill and Newton [104, 105] discuss saddlepoint approximations for difference of order statistics.
7. Heller [109] applies MACSYMA (i) to Edgeworth expansions in §4.2.3 (see §6.18 of [126]), and (ii) to augmented symmetric functions in §6.5.1 (see §12.5 of [126]).
8. Ronchetti and Ventura [191] discuss CA in relation to the impact of model misspecification on higher order expansions and suggest applications of CA to extensions of their work.

9. Rose and Smith [193] convert augmented symmetric functions to power sums and *vice versa* (pp. 272–276) to take advantage of the fundamental expectation theorem that relates the statistical expectation of an augmented symmetric function to the raw moments of a random variable. They discuss the conversion of moments to cumulants and *vice versa* and extend the capability to perform these conversions to include, *e.g.* h-statistics, k-statistics, poly-k statistics and multivariate generalizations (chapter 7). They also discuss several other aspects of asymptotic theory (chapter 8).
10. Zheng [239] (and earlier papers cited therein) discusses the conversion of moments to cumulants and *vice versa*. He follows the classical treatment [126] in his MATHEMATICA package MomCumConvert.

Another major area of current research concerns stochastic models of dynamical systems that are affected by noise or other random influences. In this area

1. W. S. Kendall reports extensive work on the application of the Itô calculus to stochastic differential equations, on the web site [255], from which many papers can be downloaded. Recent developments are reported in [129].
2. Cyganowski, Grüne and Ombach develop the topic pedagogically, using MAPLE in [62].
3. Cyganowski [249] gives instructions to download the MAPLE package `stochastic` that supports this work. He gives a short account of the background in [60]. Cyganowski, Grüne and P. E. Kloeden provide a long account in [61].
4. Kloeden's web site [246] includes ongoing information about work on this topic.

Kloeden and Platten describe several diverse applications of the underlying methods from a numerical standpoint ([135], Chapter 7) that I think can be carried over to symbolic work. These include

1. population dynamics, involving the Verhulst equation and stochastic extensions of Volterra-Lotka systems,
2. protein kinetics, involving the Brusselator equations,
3. genetic models of natural selection and geographical variations in allele distribution,
4. experimental psychology, in models of the coordination of human movement,
5. neuronal firing activity,
6. biological waste treatment,
7. hydrological studies of outflow from lakes and reservoirs,
8. air quality control using filtration equipment,
9. blood clotting dynamics,
10. cellular energetics.

Several early applications of CA in statistics are surveyed in [127, 128]. Further recent applications address a variety of topics.

1. Barroso *et al* [25] discuss test scores for von Mises regression models.
2. Burbea and Delcastillo [38] discuss geodesic submanifolds of statistical models.

3. Currie [59] gives MATHEMATICA code that helps teach maximum likelihood estimation.
4. Diaconis and Sturmfels [65] construct Markov chain algorithms, that involve computations in polynomial rings using Grobner bases, for sampling from conditional distributions.
5. Drew *et al* [67] give an algorithm to compute the cumulative distribution function of the Kolmogorov-Smirnoff statistic and discuss its CA implementation.
6. Leemis, Glen and their coworkers discuss a change of variable technique [86] and order statistics in goodness-of-fit testing [85] as well as their APPL programming system [87] that was mentioned above. They provide extensive MAPLE code.
7. Huffer and Lin [113, 114] discuss the theoretical basis of cluster probabilities. This is used *e.g.* to detect disease clustering. They generate polynomials by a recursive procedure, that is limited by the length of the expressions that the CA system can handle. They use MAPLE.
8. Huggins [115] discusses identifiability of measurement error in the bifurcating autoregressive model.
9. Kennedy and Lennox [130] apply non-classical orthogonal polynomials to the problem of moments, in an engineering context, with results that are transferable.
10. Kolassa [137] discusses bounding convergence rates for Markov chains. He provides MATHEMATICA code.
11. Lourens and Smit [149] provide some introductory examples of series expansion, integration and differentiation of statistical objects. Then they discuss the symbolic computation of the probability function, the moments and some generating functions for a commonly used runs test statistic. They provide MATHEMATICA code.
12. McDonald *et al* [160] consider exact tests for two-way contingency tables. They discuss the benefits of rational arithmetic.
13. McLeod and Quenneville [161] discuss mean likelihood estimators and cite a downloadable MATHEMATICA notebook that they developed [253].
14. Pistone, Riccomagno and Wynn develop the applications of computational commutative algebra to statistics in their monograph [184]. These include the use of Gröbner bases to identify design points in a large factorial experiment and to identify structural zeroes in a large contingency table. They use MAPLE.
15. Rose and Smith [192] derive formulas for maximum likelihood estimators using MATHEMATICA pattern matching operations.
16. Smith and Field [200] discuss cumulants for frequency domain time series using an operational approach.
17. van der Wiel *et al* [223] discuss exact distributions of nonparametric test statistics.

11 GENERAL CONSIDERATIONS

The benefit of unrestricted precision arithmetic is mentioned in [87, 160, 174, 204] that deal with statistical methods, population dynamics and x-ray data analysis, respectively. Rational arithmetic appears in papers on mathematical statistics [67, 87, 113, 114], nuclear magnetic resonance [146], and elsewhere. Surd arithmetic is used in papers on genetics [133]. These resources have been essential in my own work for the past decade, both for the production of formulas and for their evaluation. I think that during the next few years they will impact some areas of scientific computing with the dramatic effect that floating point had, half a century ago.

Several authors mention the use of CA to derive formulas correctly that had been derived incorrectly by hand in the past. For example, an error in a paper by R.A. Fisher that has been propagated since the 1930s is corrected by Rose and Smith ([193], Example 14 of Chapter 7). Emphasis is placed, at times, on the construction of general formulas that can be used to produce large numbers of formulas for individual problems of a particular kind. Many papers deal with calculations that are too tedious for manual work, but do not challenge the computer resources. Calculations that include molecular structure prediction and global identification in metabolic pathways will always be limited by the capabilities of available resources, and provide incentives for still more powerful CA systems. Several authors replace manual typesetting by mechanized publication of formulas that are output by CA software. I mentioned my own efforts to build a dependable body of material in §9.3. Several authors manipulate operators as well as expressions that have numerical values.

The representations of mathematical objects and mathematical processes play a key role in my own research that uses CA [15]–[24]. Andrews and Stafford develop notations, that are matched by internal representations, to express certain statistical problems concisely [7]. Also, for statistical work, the APPL system [87] includes a concise representation of piecewise continuous functions.

Most CA calculations considered in this survey lead directly to numerical results, often within the same paper, or to formulas that are saved for later numerical calculations. Several calculations benefit from the mutual annihilation of sub-expressions of equal magnitude and opposite sign. Mechanical code generation produces scripts in mainstream numerical programming languages from formulas that were produced by CA. In some circumstances, CA can be used to produce code that is optimal. I use formulas produced by CA to calibrate the accuracy of numerical algorithms. The opportunities provided by CA stimulate the development of new mathematical notations and algorithms. A substantial number of the applications which are listed in this survey produce graphical output. This includes several animations.

A number of MATHEMATICA notebooks and MAPLE worksheets have been mentioned in earlier sections, that include several by Miller and Wethey. These authors provide many more, accessed from the web site [254] that, while focusing on numerical and graphical solutions, also support the investigation of dependence on parameters symbolically and analytically, *e.g.* dependence of equilibria and stability of these as functions of the parameter values. Some parameters in biological models are generally difficult to measure with precision, and investigators seek qualitatively distinct solutions in terms of “parameter space”. In the best models the number of parameters is reduced to the least possible, through a very tedious process of “non-dimensionalization”. This amounts to algebraic manipulation and some symbolic calculus. The resulting set of parameters is interpreted biologically, and the regions of parameter space are mapped analytically, rather than by trial and error runs with sampled parameter values. Edelstein-Keshet [68] provides a very helpful discussion of this issue.

Most of the work described here uses “general purpose” CA. The visibility work [163, 208] mentioned in §3 uses the specialized system SINGULAR [96]. Brumberg [36] discusses the trade-off between general and specialized CA systems with reference to planetary dynamics, but the issues are the same in the other sciences.

12 CONSTRUCTING THE SURVEY

The core of the bibliography that follows was taken from an on-line literature search of Chemical Abstracts and Science Citation Index. The search was run for the Salt Lake City workshop on Computational Science sponsored by the National Science Foundation last December. The life sciences portion has been expanded by web searches, general browsing and email correspondence. The working procedures and software that I developed to extract the relevant information from the data bases, and to organize it within explanatory text can be applied to the construction of surveys in other fields of study. I am writing an account of these aspects of work. This will include suggestions for new resources to facilitate interdisciplinary information exchange that matches problems with solutions.

I will be glad to receive information that should have been included in this survey but was missed, new material and corrections. References to the numerical and graphical applications of CA systems and to the MAPLE worksheets and MATHEMATICA notebooks that I collected but did not cite are available on request.

ACKNOWLEDGMENTS

This work was conducted as part of the program of the L. C. Allen Theoretical Chemistry Group. Thanks are due to L. C. Allen for his support and encouragement, to W. Krandick of Drexel University for asking the question that got this survey started, and to M. Miller of the University of South Carolina for numerous additions to the draft. Clarification of papers that were cited, leads to further relevant material and other comments affecting the draft were provided in generous “private communications” from O. O. Aalen, D. E. Burmaster, R. W. Butler, M. R. Chernick, P. Cook, C. Denniston, A. Dress, R. Gardner, J. E. Goodman, P. Gritzmman, J. Grotendorst, F. W. Huffer, P. E. Kloeden, J. E. Kolassa, P. W. Kuchel, J. de Loera, G. Margaria, J. Moore, W. F. Morris, A. Murray, R. P. Ocón, T. R. Rebbeck, B. H. Romenij, M. D. Smith, J. Stafford, T. Theobald, L. Toet, E. O. Voit, A. L. Williams, W. Whiteley, N. Yildirim, and Q. Zheng.

References

- [1] O. O. AALEN, *Phase type distributions in survival analysis*, Scand. J. Stat. 22 (4) 447–463, 1995.
- [2] O. O. AALEN, *Phase type distributions: computer algebra and a simple mixing model*, Technical report, Section of Medical Statistics, University of Oslo, 1999.
- [3] H. ANDERSSON AND T. BRITTON, *Stochastic epidemic models and their statistical analysis*, Lecture notes in statistics, 151, Springer, 2000.
- [4] D. ANDREWS, *Asymptotic expansions of moments and cumulants*, Stat. Comp. 11 (1) 7–16, 2001.
- [5] D. F. ANDREWS AND J. E. STAFFORD, *Tools for the symbolic computation of asymptotic expansions*, J. Roy. Stat. Soc. B 55 (3) 613–627, 1993.
- [6] D. F. ANDREWS AND J. E. STAFFORD, *Iterated full partitions*, Stat. Comp. 8 (3) 189–192, 1998.
- [7] D. F. ANDREWS AND J. E. STAFFORD, *Symbolic computation for statistical inference*, Oxford University Press, 2001.
- [8] P. ARMITAGE AND T. KOLTON, *Encyclopedia of Biostatistics*, 6 volumes, Wiley, 1998.
- [9] S. AUDOLY, L. D’ANGIÒ, M. P. SACCOMANI AND C. COBELLI, *Global identifiability of linear compartmental models — a computer algebra algorithm*, IEEE Trans. Biomed. Eng. 45 (1) 36–47, 1998.
- [10] S. AUDOLY, G. BELLU, L. D’ANGIÒ, M. P. SACCOMANI AND C. COBELLI, *Global identifiability of nonlinear models of biological systems*, IEEE Trans. Biomed. Eng. 48 (1) 55–65, 2001.
- [11] J. BAGLIVO, M. PAGANO AND C. SPINO, *Permutation distributions via generating functions with applications to sensitivity analysis of discrete data*, J. Am. Stat. Assoc. 91 (435) 1037–1046, 1996.
- [12] F. BALLOUX, H. BRÜNNER, N. LUGON-MOULIN, J. HAUSSER AND J. GOUDET, *Microsatellites can be misleading: an empirical and simulation study*, Evolution, 54 (4), 1414–1422, 2000.
- [13] M. P. BARNETT, *The evaluation of molecular integrals by the zeta-function method*, in B. ALDER, S. FERNBACH AND M. ROTENBERG, eds., *Methods of Computational Physics*, Academic Press, New York, vol 2, 95–153, 1963.
- [14] M. P. BARNETT, *Molecular integrals over Slater orbitals*, Chem. Phys. Letters 166 (1) 65–70, 1990.
- [15] M. P. BARNETT, *Summing $P_n(\cos\theta)/p(n)$ for certain polynomials $p(n)$* , Computers Math. Applic. 21 (10) 79–86, 1991.
- [16] M. P. BARNETT, *Implicit rule formation in symbolic computation*, Computers Math. Applic. 26 (1) 35–50, 1993.
- [17] M. P. BARNETT, *Combining Mathematica and TeX*, TUGboat, 19 (2) 147–158, 1998.
- [18] M. P. BARNETT, *Symbolic calculation of auxiliary functions for molecular integrals over Slater orbitals*, Int. J. Quantum Chem. 76 (3) 464–472, 2000.
- [19] M. P. BARNETT, *Two-center non-exchange integrals over Slater orbitals*, J. Chem. Phys. 113 (21) 9419–9428, 2000.
- [20] M. P. BARNETT, *Digital erosion in the evaluation of molecular integrals*, Theor. Chem. Accounts, 107 (4) 241–245, 2002.
- [21] M. P. BARNETT, *Symbolic tabulation of overlap integrals over Slater orbitals*, Theor. Chem. Accounts, in press, 2002.
- [22] M. P. BARNETT, T. DECKER AND W. KRANDICK, *Power series expansion of the roots of a secular equation containing symbolic elements: computer algebra and Moseley’s law*, J. Chem. Phys. 114 (23) 10265–10269, 2001.
- [23] M. P. BARNETT AND K. R. PERRY, *Hierarchical addressing in symbolic computation*, Computers Math. Applic. 28 (8) 17–35, 1994.
- [24] M. P. BARNETT AND K. R. PERRY, *Symbolic computation for electronic publishing*, TUGboat, 15 (3) 285–292, 1994.
- [25] L. BARROSO, G. CORDEIRO AND K. VASCONCELLOS, *Improved score tests for von Mises regression models*, Commun. Stat.-Theory Methods 30 (7) 1295–1315, 2001.
- [26] M. BAYRAM, J. P. BENNETT AND M. C. DEWAR, *Using computer algebra to determine rate constants in biochemistry*, Acta Biotheoretica, 41 (1-2) 53-62, 1993.
- [27] M. BAYRAM, *Automatic analysis of the control of metabolic networks*, Comput. Biol. Med. 26 (5) 401–408, 1996.
- [28] M. BAYRAM, *Automatic derivation of steady state rate laws using a computer algebra system*, Proc. Indian Natl. Sci. Acad., A. 63 (3), 241-249, 1997.
- [29] E. D. BECKER, *High resolution NMR, theory and chemical applications*, 3rd ed. Academic Press, 2000.

- [30] D. BELLHOUSE, R. PHILIPS AND J. STAFFORD, *Symbolic operators for multiple sums*, *Comput. Stat. Data Anal* 24 (4) 443–454, 1997.
- [31] R. BELLIO AND A. BRAZZALE, *A computer algebra package for approximate conditional inference*, *Stat. Comput.* 11 (1) 17–24, 2001.
- [32] J. P. BENNETT, J. H. DAVENPORT, M. C. DEWAR, D. L. FISHER, M. GRINFELD AND M. SAURO, *Computer algebra approaches to enzyme kinetics*, *Lect. Notes in Cont. Inf. Sci.* 165, 23–30, 1991.
- [33] J. P. BENNETT, M. GRINFELD AND J. HUBBLE, *Computer algebra techniques in affinity binding equations: the dimer case*, *J. Symb. Comp.* 15 (1) 79–83, 1993.
- [34] D. W. BLICK, J. M. A. BEER, W. D. KOSNICK, S. TROXEL, A. TOET, J. WALRAVEN AND W. MITCHELL, *Laser glare in the cockpit: psychophysical estimates versus model predictions of veiling luminance distribution*, *App. Optics*, 40 (10) 1715–1726, 2001.
- [35] S. F. BOYS, G. B. COOK, C. M. REEVES, AND I. SHAVITT, *Automatic fundamental calculations of molecular structure*, *Nature (London)*, 178 (4544) 1207–1209, 1956.
- [36] V. A. BRUMBERG, *Analytical techniques of celestial mechanics*, Springer, 1995.
- [37] M. BULMER *Theoretical evolutionary biology*, Sinauer, 1994.
- [38] J. BURBEA AND J. DELCASTILLO, *Geodesic submanifolds of statistical-models with location parameters*, *Comput. Stat. Data Anal* 14 (3) 301–313, 1992.
- [39] D. E. BURMASTER, <http://www.Alceon.com>.
- [40] R. BUTLER, *Reliabilities for feedback systems and their saddlepoint approximation*, *Stat. Sci* 15 (3) 279–298, 2000.
- [41] O. H. CAMPANELLA AND M. PELEG, *Theoretical comparison of a new and the traditional method to calculate Clostridium botulinum survival during thermal inactivation*, *J. Sci. Food Agric.* 81 (11) 1969–1076, 2001.
- [42] E. R. CARSON, C. COBELLI AND L. FINKELSTEIN, *The mathematical modeling of metabolic and endocrine systems*, Wiley, 1983.
- [43] H. CASWELL *Matrix population models: construction, analysis and interpretation*, 2nd ed. Sinauer, 2001.
- [44] S.-M. CHANG AND M. D. RAUSHER, *Frequency-dependent pollen discounting contributes to maintenance of a mixed mating system in the common Morning Glory Ipomoea purpurea*, *Am. Naturalist*, 152 (5), 671–683, 1998.
- [45] M. J. CHAPPELL, K. R. GODFREY AND S. VAJDA, *Global identifiability of the parameters of nonlinear systems with specified inputs: a comparison of methods*, *Math. Biosci.* 102 (1) 41–73, 1990.
- [46] C. CHAUVIN, M. MÜLLER AND A. WEBER, *An application of quantifier elimination to mathematical biology*, in J. FLEISCHER, J. GRABMEIER, F. W. HEHL AND W. KÜCHLIN, eds. *Computer Algebra in Science and Engineering*, 287–298, World Scientific, 1995.
- [47] C. S. CHEN, M. MRKSICH, S. HUANG, G. WHITESIDES AND D. E. INGBER, *Geometric control of cell life and death*, *Science*, 276 (5317) 1425–1428, 1997.
- [48] G. CHRISTAKOS, *Modern spatiotemporal geostatistics*, Oxford University Press, 2000.
- [49] D. E. CLARK, *Computational models for probabilistic decision trees*, *Comp. Biomed. Res.* 30 (1) 19–33, 1997.
- [50] D. E. CLARK AND M. EL-TAHA, *Generation of correlated logistic-normal random variates for medical decision trees*, *Methods Inf. Med.* 37 (3) 235–238, 1998.
- [51] J. E. COHEN, *Population-growth and the earth’s human carrying-capacity*, *Science*, 269, (5222) 341–348, 1995.
- [52] J. E. COHEN, M. LYNCH AND C. E. TAYLOR, *Forensic DNA tests and Hardy-Weinberg equilibrium*, *Science*, 253 (5023) 1037–1038, 1991.
- [53] G. E. COLLINS, M. J. ENCARNACION, H. HONG, J. R. JOHNSON., W. KRANDICK, A. M. MANDACHE, A. NEUBACHER, AND H. VIELHABER, *SACLIB User’s Guide*, Technical Report 93–19, RISC–Linz, Johannes Kepler University, Linz, Austria, 1993.
- [54] P. COOK AND L. BROEMLING, *Bayesian statistics using MATHEMATICA*, *Am. Stat.* 49 (1) 70–76, 1995.
- [55] R. M. CORLESS, *HIV and antiviral therapy*, accessed from [242].
- [56] C. J. CRAMER, *Essentials of computational chemistry: theories and models*, Wiley, (in press) 2002.
- [57] J. F. CROW AND M. KIMURA, *An introduction to population genetics*, Burgess, 1970.
- [58] A. C. CULLEN AND H. C. FREY, *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Plenum, 1999.
- [59] I. CURRIE, *Maximum-likelihood-estimation and MATHEMATICA*, *Appl. Stat.-J. R. Stat. Soc* 44 (3) 379–394, 1995.
- [60] S. CYGANOWSKI, *A MAPLE package for stochastic differential equations*, in R. L. MAY AND A. K. EASTON, eds. *Proceedings of the seventh biennial computational techniques and applications conference: CTAC95*, 223–230, World Scientific, 2006; download from [247], page 2, item 2.
- [61] S. CYGANOWSKI, L. GRÜNE AND P. E. KLOEDEN, *MAPLE for stochastic differential equations*, in J. F. BLOWEY, J. P. COLEMAN AND A. W. CRAIG eds. *Theory and numerics of differential equations*, 127–178, Springer, 2001; download from [247], page 1, maplesde.pdf.
- [62] S. CYGANOWSKI, P. E. KLOEDEN AND J. OMBACH, *From elementary probability to stochastic DEs with MAPLE*, Springer, 2001.
- [63] A. DAVISON AND J. STAFFORD, *The score function and a comparison of various adjustments of the profile likelihood*, *Can. J. Stat.-Rev. Can. Stat* 26 (1) 139–148, 1998.

- [64] C. DENNISTON, *Equivalence by descent*, Ann. Hum. Genet. 64 (1) 61–82, 2000.
- [65] P. DIACONIS AND B. STURMFELS, *Algebraic algorithms for sampling from conditional distributions*, Annals Stat. 26 (1) 363–397, 1998.
- [66] A. DRESS AND W. TERHALLE, *The tree of life and other affine buildings*, Documenta Mathematica, extra vol. ICM (III) 565–574, and <http://www.mathematik.uni-bielefeld.de/documenta/xvol-icm/16/16.html>, 1998.
- [67] J. DREW, A. GLEN AND L. M. LEEMIS, *Computing the cumulative distribution function of the Kolmogorov-Smirnov statistic*, Comput. Stat. Data Anal 34 (1) 1–15, 2000.
- [68] L. EDELSTEIN-KESHET, *Mathematical models in biology*, McGraw-Hill, 1988.
- [69] I. Z. EMIRIS AND B. MOURRAIN, *Computer algebra methods for studying and computing molecular conformations*, Algorithmica, 25 (2–3) 372–402, 1999.
- [70] M. FARZA AND A. CHÉRY, *BIOESTIM: SOFTWARE FOR AUTOMATIC DESIGN OF ESTIMATORS IN BIOPROCESS ENGINEERING*, Computer Applications in the Biosciences, 10 (5) 477–488, 1994.
- [71] J. FELSENSTEIN, *Phylogeny programs*, <http://evolution.genetics.washington.edu/phylip/software.html>.
- [72] P. W. FINN AND L. E. KAVRAKI, *Computational approaches to drug design*, Algorithmica, 25 (2–3) 347–371, 1999.
- [73] L. M. J. FLORACK, *Image structure*, Kluwer, Dordrecht, 1997.
- [74] W. B. FLORIANO, N. VAIDEHI, W. A. GODDARD III, M. S. SINGER, AND G. M. SHEPHERD, *Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor*, Proc. Natl. Acad. Sci. 97 (20) 10712–10716, 2000.
- [75] P. FORWALTER-FRIEDMAN, *Hopf bifurcation in a predator-prey model*, accessed from [242].
- [76] R. FREEMAN, *Spin choreography: basic steps in high resolution NMR*, Oxford University Press, 1998.
- [77] O. D. FRIEDRICHS, A. W. M. DRESS, D. HUSON, J. KLINOWSKI AND A. L. MACKAY, *Systematic enumeration of crystalline networks*, Nature, 400 (6745) 644–647, 1999.
- [78] D. J. GALAS, *Making sense of the sequence*, Science, 291 (5507) 1257–1260, 2001.
- [79] F. GARVAN, *The Maple book*, Chapman and Hall, London, 2001.
- [80] C. GASPAROVIC, M. CABAÑAS AND C. ARÚS, *A simple approach to the design of a shielded gradient probe for high-resolution in vivo spectroscopy*, J. Mag. Res. B 109 (2) 146–152, 1995.
- [81] J. VON ZUR GATHEN AND J. GERHARDT, *Modern computer algebra*, Cambridge University Press, 1999.
- [82] R. GATTO, *Symbolic computation for approximating distributions of some families of one and two-sample nonparametric test statistics*, Stat. Comp. 11 (1) 89–95, 2001.
- [83] R. GEIL, *Determination of rate law for enzyme kinetics*, accessed from [242].
- [84] J. GITTINS AND H. PEZESHK, *A behavioral Bayes method for determining the size of a clinical trial*, Drug Inf. J. 34 (2) 355–363, 2000.
- [85] A. G. GLEN, L. M. LEEMIS AND D. R. BARR, *Order statistics in goodness-of-fit testing*, IEEE Trans. Reliability, 50 (2) 209–213, 2001.
- [86] A. G. GLEN, L. M. LEEMIS AND J. H. DREW, *A generalized univariate change-of-variable transformation technique*, INFORMS Journal on Computing, 9 (3) 288–295, 1997.
- [87] A. G. GLEN, L. M. LEEMIS AND D. L. EVANS, *APPL: A probability programming language*, Am. Stat 55 (2) 156–166, 2001.
- [88] N. GO AND H. A. SCHERAGA, *Ring-closure and local conformation deformations of chain molecules with C_n , I or S_{2n} symmetry*, Macromol. 6 (2), 273–281, 1973.
- [89] M. S. GOCKENBACH, *A practical introduction to MATLAB*, <http://www.math.mtu.edu/msgocken/intro/intro.html>.
- [90] A. GÓMEZ, J. L. ARAGÓN, O. CABALLERO AND F. DÁVILA, *The application of Clifford algebras to crystallography using MATHEMATICA*, in R. ABLAMOWICZ, P. LOUNESTO AND J. M. PARRA, *Clifford algebras with numeric and symbolic computations*, 250–266, Birkhauser, 1996.
- [91] J. E. GOODMAN AND J. O’ROURKE (eds.) *Handbook of discrete and computational geometry*, CRC Press, 1997.
- [92] C. GOUTIS AND G. CASELLA, *Explaining the saddlepoint approximation*, Am. Stat. 53 (3) 216–224, 1999.
- [93] B. F. GRAY, H. O. PRITCHARD AND F. H. SUMNER, *Hybridization in the ground state of the hydrogen molecule-ion*, J. Chem. Soc. 2631–2635, 1956.
- [94] D. M. GREEN AND J. A. SWEETS, *Signal detection theory and psychophysics*, Kresge, New York, 1974.
- [95] G.-M. GREUEL, *Computer algebra and algebraic geometry — achievements and perspectives*, J. Symb. Comp. 30 (3) 253–289, 2000.
- [96] G.-M. GREUEL, G. PFISTER AND H. SCHÖNEMANN, *SINGULAR 2.0. A computer algebra system for polynomial computations*, Centre for Computer Algebra, University of Kaiserslautern, <http://www.singular.uni-kl.de>, 2001.
- [97] M. GRINFELD, J. BENNETT AND J. HUBBLE, *Application of computer algebra techniques to affinity binding equations*, IAM J. Math. App. Bus. Ind. 8, 157–166, 1997.
- [98] J. GROENDORST, P. JANSEN AND S. M. SCHOBERTH, *The McConnell equations in biochemistry*, accessed from [242].

- [99] J. GROTEENDORST, P. JANSEN AND S. M. SCHOBERTH, *On solving the McConnell equations in biochemistry*, MAPLE Technical Newsletter, 1, 55–62, 1994.
- [100] S. GUPTA, R. M. ANDERSON AND R. M. MAY, *Mathematical models and the design of public-health policy — HIV and antiviral therapy*, SIAM Rev. 35 (1) 1–16, 1993.
- [101] J. M. HAILE, *Molecular dynamics simulation: elementary methods*, Wiley, 1997.
- [102] N. C. HANDY, S. CARTER AND S. M. COLWELL, *The vibrational energy levels of ammonia*, Mol. Phys. 96 (4) 477–491, 1999.
- [103] I. HANSKI, *Metapopulation Ecology*, Oxford University Press, 1999.
- [104] J. L. HARVILL AND H. J. NEWTON, *Saddlepoint approximations for the difference of order statistics*, Biometrika, 82 (1) 226–231, 1995.
- [105] J. L. HARVILL AND H. J. NEWTON, *Using symbolic math to evaluate saddlepoint approximations for the difference of order statistics*, Comm. Stat.-Simul. Comp. 24 (3) 781–791, 1995.
- [106] X. HE, D. KU AND J. E. MOORE, *Simple calculation of the velocity profiles for pulsatile flow in a blood vessel using MATHEMATICA*, Ann. Biomed. Eng. 21 (1) 45–49, 1993.
- [107] A. C. HEARN, *REDUCE user's manual, version 3.6*, RAND Corporation, Santa Monica, CA, 1995.
- [108] P. S. HECKERLING, *Parametric three-way receiver operating characteristic surface analysis using MATHEMATICA*, Med. Dec. Making. 21 (5) 409–417, 2001.
- [109] B. HELLER, *MACSYMA for statisticians*, Wiley, 1991.
- [110] C. M. HENRY, *Careers in bioinformatics*, Chemical and Engineering News, 80, 47–53, 2002.
- [111] <http://www.math.gatech.edu/herod/MathBio/textcode.html>.
- [112] D. W. HOSNER AND S. LEMESHOW, *Applied Logistic Regression*, Wiley, 1989.
- [113] F. W. HUFFER AND C.-T. LIN, *Computing the exact distribution of the extremes of sums of consecutive spacings*, Comput. Stat. Data Analysis, 26 (2) 117–132, 1997.
- [114] F. W. HUFFER AND C.-T. LIN, *Computing the joint distribution of general linear combinations of spacings or exponential variates*, Stat. Sinica 11 (4) 1141–1157, 2001.
- [115] R. HUGGINS, *On the identifiability of measurement error in the bifurcating autoregressive model*, Stat. Probab. Lett. 27 (1) 17–23, 1996.
- [116] S. HUNKA, *Identifying regions of significance in ANCOVA problems having nonhomogeneous regressions*, Br. J. Math. Stat. Psychol. 48 (1) 161–188, 1995.
- [117] S. HUNKA AND J. LEIGHTON, *Defining Johnson-Neyman regions of significance in the three-covariate ANCOVA using MATHEMATICA*, J. Educ. Behav. Stat. 22 (4) 361–387, 1997.
- [118] K. ITÔ AND N. SÛGAKKAI, *Encyclopedic dictionary of mathematics*, 2 volumes, MIT Press, 1993.
- [119] J. A. JACQUEZ, *Compartmental analysis in biology and medicine*, BioMedware, Ann Arbor, 1996.
- [120] R. D. JENKS AND R. S. SUTOR, *AXIOM: the scientific computation system*, Springer, 1992.
- [121] F. JENSEN, *Introduction to computational chemistry*, Wiley, 1999.
- [122] M. S. JU AND J. M. MANSOUR, *Simulation of the double limb support phase of human gait*, J. Biomech. Eng. 110 (3) 223–9, 1998.
- [123] P. A. JUMARS, *Animal guts as ideal chemical reactors: maximizing absorption rates*, Am. Naturalist, 155 (4) 528–543, 2000.
- [124] A. C. KAK AND M. SLANEY, *Principles of Computerized Tomographic Imaging*, www.slaney.org/pct/pct-toc.html, SIAM, 2001.
- [125] R. KEMP-HARPER, D. J. PHILP AND P. W. KUCHEL, *Nuclear magnetic resonance of J-coupled quadrupolar nuclei: use of the tensor operator basis*, J. Chem. Phys. 115 (7) 2908–2916, 2001.
- [126] M. G. KENDALL, *Advanced theory of statistics*, 6th ed. A. STUART AND J. K. ORD. eds. Halsted Press, New York, 1994.
- [127] W. S. KENDALL, *Computer algebra in probability and statistics*, Statistica Neerlandica, 47, 9–25, 1993.
- [128] W. S. KENDALL, *Computer algebra*, in [8].
- [129] W. S. KENDALL, *Symbolic Itô calculus in AXIOM: an ongoing story*, Stat. Comp. 11 (1) 25–35, 2001.
- [130] C. KENNEDY AND W. LENNOX, *Solution to the practical problem of moments using non-classical orthogonal polynomials, with applications for probabilistic analysis*, Probab. Eng. Eng. Mech 15 (4) 371–379, 2000.
- [131] M. KIRKPATRICK, *Good genes and direct selection in the evolution of mating preferences*, Evolution, 50 (6), 2125–2140, 1996.
- [132] M. KIRKPATRICK, W. G. HILL AND R. THOMPSON, *Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle*, Genet. Res. (Camb.) 64 (1) 57–69, 1994.
- [133] M. KIRKPATRICK AND M. R. SERVEDIO, *The reinforcement of mating preferences on an island*, Genetics, 151 (2) 865–884, 1999.
- [134] S. J. KLEENE AND H. C. CEJTIN, *Solving buffering problems with Mathematica software*, Analytical Biochem. 222 (2) 310–314, 1994.
- [135] P. E. KLOEDEN AND E. PLATEN *Numerical solution of stochastic differential equations*, Springer, 1992.
- [136] J. E. KOLASSA, *Series approximation methods in statistics*, 2nd ed., Springer, 1997.

- [137] J. E. KOLASSA, *Bounding convergence rates for Markov chains: An example of the use of computer algebra*, Stat. Comp. 11 (1) 83–87, 2001.
- [138] J. E. KOLASSA, *Approximate multivariate conditional inference using the adjusted profile likelihood*, Can. J. Stat. submitted, 2002.
- [139] M. KOT, *Elements of mathematical ecology*, Cambridge University Press, 2001.
- [140] S. KOTZ AND N. L. JOHNSON, eds., *Encyclopedia of Statistical Sciences*, 9 volumes, Wiley, 1988.
- [141] P. W. KUCHEL AND C. J. DURRANT, *Permeability coefficients from NMR q -space data: models with unevenly spaced semi-permeable parallel membranes*, J. Mag. Res. 139 (2) 258–272, 1999.
- [142] P. W. KUCHEL AND E. F. FACKERELL, *Parametric-equation representation of biconcave erythrocytes*, Bull. Math. Bio. 61 (2) 209–220, 1999.
- [143] P. W. KUCHEL, A. L. LENNON AND C. J. DURRANT, *Analytical solutions and simulations for spin-echo measurements of diffusion of spins in a sphere with surface and bulk relaxation*, J. Mag. Res. B 112 (1) 1–17, 1996.
- [144] K. LANGE, *Mathematical and statistical methods for genetic analysis*, Springer, 1997.
- [145] C. T. LE, *Applied Survival Analysis*, Wiley, 1997.
- [146] M. H. LEVITT, *Spin dynamics: basic principles of NMR spectroscopy*, Wiley, 2001.
- [147] T. LINDBERG, *Scale-space theory in computer vision*, Kluwer, Dordrecht, 1994.
- [148] L. LJUNG AND T. GLAD, *On global identifiability for arbitrary model parameterization*, Automatica, 30 (2) 265–276, 1994.
- [149] A. LOURENS AND C. SMIT, *Symbolic computation used to solve a Markov chain-based problem in statistical distribution theory*, South Afr. Stat. J 33 (1) 71–93, 1999.
- [150] M. LYNCH AND B. WALSH, *Genetics and analysis of quantitative traits*, Sinauer, 1998.
- [151] R. H. MACARTHUR, *Geographical ecology*, 35–58, Harper and Row, 1972.
- [152] D. MANOCHA AND Y. ZHU, *Kinematic manipulation of molecular chains subject to rigid constraints*, Proceedings of the 2nd International Conference on Molecular Biology, 285–194, 1994.
- [153] D. MANOCHA, Y. ZHU AND W. WRIGHT, *Conformational analysis of molecular chains using nano-kinematics*, Computer application of biological sciences (CABIOS), 11 (1) 71–86, 1995.
- [154] G. MARGARIA, E. RICCOMAGNO, M. C. CHAPPELL AND H. P. WYNN, *Differential algebra methods for the study of the structural identifiability of state-space models in the biosciences*, Math. Biosci. 174 (1) 1–26, 2001.
- [155] J. H. MATIS AND T. R. KIFF, *Stochastic population models — a compartmental perspective*, Springer, 1999.
- [156] R. M. MAY, *Biological populations with nonoverlapping generations: stable points, stable cycles and chaos*, Science, 186 (4164) 645–647, 1974.
- [157] R. M. MAY AND G. F. OSTER, *Bifurcations and dynamic complexity in simple ecological models* Am. Naturalist, 110 (974) 573–579, 1976.
- [158] J. MAYNARD SMITH, *Evolution and the theory of games*, Cambridge University Press, 1982.
- [159] J. P. MAZAT, R. OUHABI, M. RIGOLET AND S. SCHUSTER, *Modern trends in biothermokinetics*, Kluwer Academic, 1994.
- [160] J. McDONALD, D. DEROURE AND D. MICHAELIDES, *Exact tests for two-way symmetric contingency tables*, Stat. Comp. 8 (4) 391–399, 1998.
- [161] A. MCLEOD AND B. QUENNEVILLE, *Mean likelihood estimators*, Stat. Comp. 11 (1) 57–65, 2001.
- [162] S. M. MEGAHED, *Automatic algebraic computation of basic kinematic equations of tree structure robot arms: application to human arms*, Proc. Inst. Mech. Eng. H, J. Eng. Med. 205 (3) 135–43, 1991.
- [163] G. MEGYESI, F. SOTTILE AND T. THEOBALD, *Common transversals and tangents to two lines and two quadrics in P^3* , arXiv:math.AG/0206044, v1, 2002.
- [164] M. MILLER AND D. S. WETHEY, *Biological populations and ecological models*, accessed from [242].
- [165] M. MILLER AND D. S. WETHEY, *Competition models*, accessed from [242].
- [166] M. MILLER AND D. S. WETHEY, *The dynamics of arthropod predator-prey systems*, accessed from [242].
- [167] M. MILLER AND D. S. WETHEY, *Evolutionarily stable strategies*, accessed from [242].
- [168] M. MILLER AND D. S. WETHEY, *Graphical analysis of the chemostat system*, accessed from [242].
- [169] M. MILLER AND D. S. WETHEY, *Population growth and the earth's human carrying capacity*, accessed from [242].
- [170] M. MILLER AND D. S. WETHEY, *Predator-prey models*, accessed from [242].
- [171] M. MILLER AND D. S. WETHEY, *r - and K -selection*, accessed from <http://marine.geol.sc.edu/BIOL/Courses/BIOL765/f96/RK.html>.
- [172] J. E. MOORE, N. GUGGENHEIM, A. DELFINO, P.-A. DORIOT, P.-A. DORSAZ, W. RUTISHAUSER AND J.-J. MEISTER, *Preliminary analysis of the effects of blood flow patterns in the coronary arteries*, J. Biomech. Eng. 116 (3) 302–306, 1994.
- [173] P. J. MULQUINNEY AND P. W. KUCHEL, *Model of 2,3-biphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement*, Biochem. J. 342 (3) 581–596, 1999.

- [174] A. MURRAY, J. WATKINS AND D. BONE, *A biological acoustic survey in the marginal ice-edge zone of the Bellinghausen Sea*, Deep-Sea Res. II -Top. Stud. Oceanog. 42 (4-5) 1159–1175, 1995.
- [175] J. D. MURRAY, *Mathematical Biology*, 2nd ed. Springer, 1998.
- [176] R. M. MURRAY, Z. LI AND S. S. SASTRY, *A mathematical introduction to robotic manipulation*, CRC Press, Boca Raton, 1994.
- [177] J. K. NICHOLSON, J. CONNELLY, J. C. LINDON AND E. HOLMES, *Metabonomics: a platform for studying drug toxicity and gene function*, Nature Reviews/Drug Discovery, 1 (2) 153–162, 2002.
- [178] G. PARMIGIANI, *Modeling in medical decision making: a Bayesian approach*, Wiley, 2002.
- [179] J. PECCOUD AND B. YCART, *Markovian modeling of gene-product synthesis*, Theo. Pop. Bio. 48 (2) 222–234, 1995.
- [180] G. PECELLI, *Prey-predator systems with delay: Hopf bifurcation and stable oscillations*, Math. Comp. Model. 25 (10) 77–98, 1997.
- [181] R. PÉREZ-OCÓN, J. E. RUIZ-CASTRO AND M. L. GÀMIZ-PÉREZ, *A multi-variate model to measure the effect of treatments in survival to breast cancer*, Biomet. J. 40 (6) 703–715, 1998.
- [182] R. PÉREZ-OCÓN, J. E. RUIZ-CASTRO AND M. L. GÀMIZ-PÉREZ, *Non-homogeneous Markov models in the analysis of survival after breast cancer*, Appl. Stat. 50 (1), 111–124, 2001.
- [183] S. PETITJEAN, *Algebraic geometry and computer vision: polynomial systems, real and complex roots*, J. Math. Imaging Vis. 10 (3) 191–220, 1999.
- [184] G. PISTONE, E. RICCOMAGNO, H. P. WYNN, *Algebraic statistics: computational commutative algebra in statistics*, CRC Press, Boca Raton, 2001.
- [185] H. O. PRITCHARD, *Computational chemistry in the 1950s*, J. Mol. Graphics Mod. 19 (6) 623–627, 2001.
- [186] K. L. QUEENEY, E. P. MARIN, C. M. CAMPBELL AND E. PEACOCK-LOPEZ, *Chemical oscillations in enzyme kinetics*, Chem. Educ. [Electronic Publication] 1 (3), 1996.
- [187] A. RAKSANYI, Y. LECOURTIER, E. WALTER AND A. VENOT, *Identifiability and distinguishability testing via computer algebra*, Math. Biosci. 77 (1–2) 245, 1985.
- [188] T. R. REBBECK, E. L. LUSTBADER AND K. H. BUTOW, *Somatic allele loss in genetic linkage analysis of cancer*, Genetic Epidem. 11 (5) 419–429, 1994.
- [189] B. M. TER HAAR ROMENY, ed. *Geometry-driven diffusion in computer vision*, Kluwer, Dordrecht, 1994.
- [190] B. M. TER HAAR ROMENY, *Computer vision and Mathematica 4*, Comp. Vis. Sci. in press, 2002.
- [191] E. RONCHETTI AND L. VENTURA, *Between stability and higher-order asymptotics*, Stat. Comp. 11 (1) 67–73, 2001.
- [192] C. ROSE AND M. D. SMITH, *Symbolic maximum likelihood estimation with MATHEMATICA*, J. Roy. Stat. Soc. Series D, 49 (2) 229–240, 2000.
- [193] C. ROSE AND M. D. SMITH, *Mathematical Statistics with MATHEMATICA*, Springer, 2002.
- [194] J. ROUGHGARDEN, *Theory of Population Genetics and Evolutionary Ecology*, MacMillan, London, 1979.
- [195] J. F. SADC AND N. RIVIER, *Boerdijk-Coxeter helix and biological helices as quasicrystals*, Mat. Sci. Eng. A — Struct. 294 (Sp. Iss. Dec. 15) 397–400, 2000.
- [196] A. R. SCHULZ AND J. SÜDI, *Control analysis of single enzyme sequences with abortive complexes and random substrate binding*, J. Theor. Biol. 182 (3) 397–403, 1996.
- [197] D. SCHWALBE, *The logistic model of population growth*, accessed from [242].
- [198] B. SCOTT, *MAPLE for environmental sciences: a helping hand*, Springer, 2001.
- [199] M. SENECHAL, *Crystals and quasicrystals* in [91] 933–949.
- [200] B. SMITH AND C. FIELD, *Symbolic cumulant calculations for frequency domain time series*, Stat. Comp. 11 (1) 75–82, 2001.
- [201] J. STAFFORD, *Using intersection matrices to identify graphical structure*, Stat. Comp. 11 (1) 47–55, 2001.
- [202] J. E. STAFFORD AND D. F. ANDREWS, *A symbolic algorithm for studying adjustments to the profile likelihood*, Biometrika, 80 (4) 715–730, 1993.
- [203] J. E. STAFFORD, D. F. ANDREWS AND Y. WANG, *Symbolic computation — a unified approach to studying likelihood*, Stat. Comp. 4 (4) 235–245, 1994.
- [204] M. STAMPANONI, M. FIX, P. FRANÇOIS AND P. RÜEGSEGG, *Computer algebra for spectral reconstruction between 6 and 25 mev*, Med. Phys. 28 (3) 325–327, 2001.
- [205] K. STRAUBINGER, F. SCHICK AND O. LUTZ, *Pulse angle dependence of double-spin-echo proton NMR spectra of citrate — theory and experiments*, J. Mag. Res. B 110 (2) 188–194, 1996.
- [206] E. A. TANIS AND Z. A. KARIAN, *Probability and statistics: explorations with MAPLE*, 2nd ed. Prentice-Hall, 1999.
- [207] S. TAVARÉ, *Calibrating the clock: using stochastic processes to measure the rate of evolution*, in E. S. LANDER AND M. S. WATERMAN, eds., *Calculating the secrets of life*, National Academy Press, 1995.
- [208] T. THEOBALD, *Visibility computations: from discrete algorithms to real algebraic geometry*, DIMACS Series in Discrete Math. Theor. Comp. Sci. in press (2002).

- [209] G. L. TIETJEN, *A topical dictionary of statistics*, Chapman and Hall, 1986.
- [210] M. TURELLI AND N. H. BARTON, *Genetic and statistical analyses of strong selection on polygenic traits: what, me normal?* *Genetics*, 138 (3) 913–941, 1994.
- [211] W. V. VASCONCELOS, *Computational methods in commutative algebra and algebraic geometry*, Springer, 1998.
- [212] J. A. M. VERMASEREN, *How Useful is FORM?* J. FLEISCHER, J. GRABMEIER, F. W. HEHL AND W. KÜCHLIN, eds. *Computer Algebra in Science and Engineering*, 67–76, World Scientific, 1995.
- [213] S. V. VISCIDO, M. MILLER AND D. S. WETHEY, *The response of a selfish herd to an attack from outside the group perimeter*, *J. Theor. Biol.* 208 (3) 315–328, 2001.
- [214] S. V. VISCIDO, M. MILLER AND D. S. WETHEY, *The paradox of the selfish herd: the search for a realistic movement rule*, *J. Theor. Biol.* to appear, 2002.
- [215] S. V. VISCIDO, M. MILLER AND D. S. WETHEY, *Group foraging: coincidental gathering or local enhancement*, submitted, 2002.
- [216] E. O. VOIT, *Computational analysis of biochemical systems*, Cambridge University Press, 2000.
- [217] D. VOSE, *Risk Analysis* 2nd ed., Wiley, 2000.
- [218] H. D. WACTLAR AND M. P. BARNETT, *Mechanization of tedious algebra — the e coefficients of theoretical chemistry*, *Comm. ACM*, 7 (12) 704–710, 1964.
- [219] L. WALLER AND A. LAWSON, *The power of focused tests to detect disease clustering*, *Stat. Med.* 14 (21-22) 2291–2308, 1995.
- [220] L. WALLER, B. TURNBULL AND J. HARDIN, *Obtaining distribution-functions by numerical inversion of characteristic functions with applications*, *Am. Stat.* 49 (4) 346–350, 1995.
- [221] B. S. WEIR, *Genetic Data Analysis*, II, Sinauer, 1996.
- [222] I. D. WICKHAM, G. C. WAKE, S. J. R. WOODWARD AND B. S. THORROLD, *Dynamical systems modeling of the interactions of animal stocking density and soil fertility in grazed pasture*, *J. App. Math. Decision Sci.*, 1(1) 27-43, 1997.
- [223] M. A. VAN DE WIEL, A. DI BUCCHIANICO AND P. VAN DER LAAN, *Symbolic computation and exact distributions of nonparametric test statistics*, *Statistician*, 48 (4) 507–516, 1999.
- [224] A. WILLIAMS AND G. JEFFERY, *Growth dynamics of the developing lateral geniculate nucleus*, *J. Comp. Neurol.* 430 (3) 332–342, 2001.
- [225] S. WOLFRAM, *The Mathematica Book*, 4th ed. Cambridge University Press, 1999.
- [226] S. J. R. WOODWARD, *A multi-variable optimal control problem in fertilizer economics*, accessed from [242].
- [227] E. K. YEARGERS, R. W. SHONKWILER AND J. V. HEROD, *An introduction to the mathematics of biology: with computer algebra models*, Birkhäuser, 1996.
- [228] N. YILDIRIM, *Application of Gröbner bases theory to derive steady state rate equations in biochemical kinetic theory*, accessed from [242].
- [229] N. YILDIRIM AND M. BAYRAM, *Analysis of the kinetics of unstable enzymatic systems using MAPLE*, (Abstract). *App. Math. Comp.* 112 (1), 41-48, 2000.
- [230] N. YILDIRIM AND M. BAYRAM, *Derivation of conservation relationships for metabolic networks Using MAPLE*, (Abstract). *App. Math. Comp.* 112 (2-3) 255-263, 2000.
- [231] N. YILDIRIM, M. CIFTCI AND O. I. KUFREVIOLU, *Kinetic analyses of multi-enzyme systems: a case study of the closed system of creatine kinase, hexokinase and glucose 6-phosphate dehydrogenase*, *J. Math. Chem.* 31 (1) 121–130, 2002.
- [232] N. YILDIRIM, D. YILDIRIM AND F. AKCAY, *Application of Gröbner basis theory to derive rate equations for reactions with more than one substrate or product*, *Applied Math. Comp.* in press, 2002.
- [233] N. P. YUSHKIN, *Natural polymer crystals of hydrocarbons as models of prebiological organisms*, *J. Cryst. Growth*, 167 (1-2) 237-247, 1996.
- [234] S.-R. ZHAO, J. GROTENDORST AND H. HALLING, *Calculation of the potential distribution for a three-layer spherical volume conductor*, *MAPLE Technical Newsletter*, 2 (1) 59-66, 1995.
- [235] Q. ZHENG, *On the MVK stochastic carcinogenesis model with Erlang distributed cell life lengths*, *Risk Analysis*, 15 (4), 495–502, 1995.
- [236] Q. ZHENG, *Computer algebra is indispensable in some problems of mathematical biology*, *Math. Biosci.* 151 (2) 219–225, 1998.
- [237] Q. ZHENG, *Automating the computation of cumulants of stochastic population processes*, in K. BERK AND M. POURAHMADI, eds. *Computing science and statistics, Proceedings of the 31st symposium on the interface — models, predictions and computing*, 328–333, Interface Foundation of North America, 1999.
- [238] Q. ZHENG, *On the dispersion index of a Markovian molecular clock*, *Math. Biosci.* 172 (2) 115–128, 2001.
- [239] Q. ZHENG, *Computing relations between moments and cumulants*, *Comp. Stat.* in press, 2002.
- [240] see [37] p. 8.
- [241] <http://www.mapleapps.com>.

- [242] link “biology” under “science” in [241].
 [243] <http://www.mathsource.com/Content/Applications>.
 [244] <http://www.mathStatica.com>.
 [245] http://www.mapleapps.com/powertools/statisticssupplement/stats_supplement.shtml.
 [246] <http://www.math.uni-frankfurt.de>
 [247] <http://www.math.uni-frankfurt.de/~grueneden/papers/maplesde.html>.
 [248] <http://www.math.uni-frankfurt.de/~numerik/kloeden/maplsede>.
 [249] <http://www.math.uni-frankfurt.de/~numerik/maplestoch>.
 [250] <http://www.phylogeny.arizona.edu/tree/phylogeny.html>.
 [251] <http://www.utstat.utoronto.ca/~stafford/papers.html>.
 [252] <http://www.utstat.toronto.edu/david/home.html>.
 [253] <http://www.stats.uwo.ca/mcleod/epubs/mele>.
 [254] www.math.sc.edu/~miller.
 [255] <http://www.warwick.ac.uk/statsDept/Staff/WSK>.

ANNOUNCEMENT AND CALL FOR PARTICIPATION

EAST COAST COMPUTER ALGEBRA DAY 2003

Saturday, April 5, 2003, Clemson University

Conference website: <http://www.math.clemson.edu/~sgao/ECCAD03/>

The 10th annual East Coast Computer Algebra Day (ECCAD'2003) will be held on Saturday, April 5, 2003, at Clemson University, South Carolina, USA. Up-to-date information regarding accommodations, directions, registration, etc. can be found at the conference website, given above. The themes of the conference are:

Algebraic Algorithms; Hybrid Symbolic-Numeric Computation; Computer Algebra Systems and Generic Programming; Mathematical Communication; Complexity of Algebraic Problems.

INVITED SPEAKERS

Prof. Richard Brent, Oxford University, England: *Primitive and almost primitive trinomials over $GF(2)$*

Prof. Jeremy Johnson, Drexel University, USA: *Computer Algebra and Signal Processing*

Prof. Wolfgang Schreiner, Johannes Kepler University, Austria: *Distributed Maple - Lessons Learned on Parallel Computer Algebra in Distributed Environments*

POSTER SESSIONS

In keeping with tradition, there will be two poster sessions offering all participants an opportunity to present timely research in an informal environment. The conference auditorium is well equipped with video projectors, computers and internet access. Software demonstrations are welcome.

TRAVEL SUPPORT

Subject to successful funding, a limited number of participants will be supported to attend ECCAD 2003. Support may cover partially your travel expenses and lodging for up to two nights. Graduate students and junior faculty are particularly encouraged to apply.

REGISTRATION

There is no registration fee for the conference; You can register at the conference website.

IMPORTANT DATES

Registration begins: February 1, 2003

Deadline for submitting poster abstracts: March 22, 2003

Conference: April 5, 2003

Questions about the conference may be addressed to the organizing committee:

Shuhong Gao, sgao@math.clemson.edu, or David Jacobs, dpj@cs.clemson.edu