Computational Statistics und Biometrie

Wer treibt wen?

Computational statistics and biometry

Which discipline drives which?

• Lutz Edler[1]

Die Arbeit eines Biometrikers bzw. einer Biometrikerin ist bestimmt durch die biologische Problemstellung und die zu ihrer Lösung erforderlichen mathematischen und statistischen Methoden. Die biometrische Problemlösung erfordert fast immer den Einsatz rechnerischer Verfahren und damit auch von Methoden der Computational Statistics. Eine nähere Betrachtung der Geschichte der beiden Disziplinen in Deutschland zeigt aber auch eine umgekehrte Abhängigkeit. Computational Statistics hat in den vergangenen 50 Jahren wesentliche Aktivitäten der Deutschen Region (DR) der Internationalen Biometrischen Gesellschaft beeinflusst. Dazu gehören Exakte Rechenverfahren und Permutationstests, Simulationsverfahren und die Methode des Bootstrap und interaktive grafische Verfahren. Dieser Beitrag beschäftigt sich mit der wechselseitigen Beeinflussung von Biometrie und Computational Statistics in Deutschland. Eine Analyse von Beiträgen in wissenschaftlichen biometrischen Zeitschriften, von Programmen der biometrischen Kolloquien, von Aktivitäten der Arbeitsgruppe Computational Statistics der DR und von Protokollen der Arbeitstagungen auf der Reisensburg werden für eine Standortbestimmung der Computational Statistics in Deutschland herangezogen, um die Diskussion weiterer Entwicklungsmöglichkeiten dieses Faches in der Biometrie fördern.

Schlüsselwörter: Geschichte, Computational Statistics, statistische Auswertesysteme, Deutsche Region der Internationalen Biometrischen Gesellschaft (DR), Reisensburg

A biometrician's work is defined through the biological or medical problem and the mathematical and statistical methods needed for its solution. This requires in most instances statistical data analysis and the use of methods of computational statistics. At first, it seems quite obvious that the computational needs of the biometric problem determine what has to be developed by the discipline of computational statistics. However, viewing the development of biometry and computational statistics in Germany for the past decades in more details reveals an interesting interaction between the activities of the German Region of the International Biometric Society and groups engaged in computational statistics within Germany. Exact methods of statistical inference and permutation tests, simulations and the use of the Bootstrap, and interactive graphical statistical methods are examples of this fruitful reciprocal benefit. This

[1] *Abteilung Biostatistik - C060, Deutsches Krebsforschungszentrum, Heidelberg, Deutschland*

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gbp(M2005061414-1)

contribution examines therefore relationships between the historical development of biometry and computational statistics in Germany using as sources of information contributions to the scientific literature, presentations and sessions at scientific conferences on biometry and on computational statistics which influenced the development of both disciplines and exhibits a reciprocal dependency. The annual workshops organized on the Reisensburg now for more than 30 years are recognized as an outstanding factor of this interrelationship. This work aims at the definition of the present status of computational statistics in the German Region of the International Biometric Society and intends to guide and to foster the discussion of the future development of this discipline among biometricians.

Keywords: history, computational statistics, statistical analysis systems, German region, International Biometric Society, Reisensburg

1. Einleitung

Statistics may be regarded as the study of populations, the study of variation, and the study of methods of the reduction of data.

R. A. Fisher (1925) [9]

Der 50. Geburtstag der Deutschen Region (DR) der Internationalen Biometrischen Gesellschaft war ein willkommener Anlass für eine historische Betrachtung, und die Einladung zu einem Beitrag zu den Beziehungen zwischen Biometrie und Computational Statistics (CS) eine reizvolle, aber auch schwierige Herausforderung. Durch die Aufgaben einer Abteilung an einem großen Forschungsinstitut mit einer angewandten biomedizinischen Aufgabenstellung, nämlich der Krebsforschung, war und ist der Autor dieses Beitrags mit sehr verschiedenartigen Fragestellungen der Biometrie konfrontiert, vor allem aber auch mit der Notwendigkeit, rechnerische Lösungen und Auswertungen problemgerecht und zeitnah für die gemeinsam mit den Substanzwissenschaftlern durchgeführten Projekten zu liefern. Biometrie und CS werden in einem solchen Aufgabenfeld eventuell stärker als Einheit empfunden als in der universitären Forschung. Über eine historische Entwicklung zu schreiben ist für uns Biometriker keine leichte Aufgabe, da uns die Methodik der Geschichtswissenschaften fremd ist. Dieser Schwierigkeit steht jedoch der mögliche Nutzen einer Reflexion der Entwicklung unseres Fachgebietes gegenüber. Erfolgreiche Entwicklungen ebenso wie Fehleinschätzungen und Irrwege der Vergangenheit können uns sowohl bei der Bewältigung gegenwärtiger Projekte als auch bei der Formulierung neuer Ziele hilfreich sein. Die vorliegende Arbeit befasst sich mit der Frage, wie CS und Biometrie sich in den vergangenen 50 Jahren gegenseitig befruchteten und wie sie zusammenwirkten. Gab es zeitliche Perioden und Themen, bei welchen CS die Entwicklung der Biometrie vorantrieb? Haben biometrische Fragestellungen die Entwicklung von Verfahren der CS angestoßen oder beschleunigt?

Zunächst werden im folgenden Abschnitt methodisches Vorgehen und benutzte Quellen beschrieben. In Abschnitt 3 wird das Thema präzisiert und in Abschnitt 4 die Entwicklung der CS in Deutschland beschrieben. Ein Abriss der Geschichte der AG Computational Statistics der DR findet sich in Abschnitt 5. Die nächsten Abschnitte 6 und 7 widmen sich dann ausführlich den Wechselwirkungen zwischen CS und Biometrie. Abschnitt 8 fasst dann diese geschichtlichen Beobachtungen zusammen.

2. Material und Methoden

Die vorliegende Untersuchung basiert auf folgenden Quellen:

A) publizierte Arbeiten zur Geschichte der Biometrie und Statistik und der CS (z. B. [1], [5], [24])

B) Zeitschriften der Biometrie und der CS, vorrangig Biometrics, Biometrical Journal und Computational Statistics and Data Analysis (CSDA)

C) Literatursuche im Current Index of Statistics (CIS) der American Statistical Association (ASA)

D) Tagungsprogramme und Abstraktbände des Biometrischen Kolloquiums

E) Tagungsprogramme und Ankündigungen der Arbeitssitzungen der AG, Statistische Auswertesysteme der GMDS und (seit 1985) der AG Computational Statistics auf der Reisensburg, Unterlagen der AG Computational Statistics beim Schriftführer Axel Benner

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gipb0[2005]0014.1)

F) Tagungsberichte und Sammelbände der AG Mathematische Statistik der Hochschulrechenzentren der DDR aus den Jahren 1982-1986

G) Persönliche Mitteilungen von Jürgen Läuter, Egmar Rödel, Christel Richter, Volker Guiard und Rainer Koch zur Situation der CS in der DDR

Ein grundsätzliches Problem beim Schreiben eines historischen Beitrags ist die Subjektivität des Schreibers. Subjektiv wird die Darstellung durch persönliche Meinungen und Wertungen, durch die eigene Erfahrung und durch die Auswahl des Materials. Am leichtesten war für diese Arbeit der Zugang zu eigenen Dokumenten und denjenigen nahestehender Kollegen, sowie der Zugriff auf publizierte Arbeiten. Für die CS existierten Papiere bei einigen Funktionsträgern und in Aktenordnern der Arbeitsgruppen, die von Vorsitz zu Vorsitz weitergereicht werden. Aus Zeitgründen kam für diese Arbeit eine ausgedehnte Archivarbeit nicht in Betracht und entsprechend schied ein aufwändiges Sammeln von Original-Dokumenten, Aufarbeiten und Werten aus. Eine subjektive Auswahl mit der Gefahr einer selektiven Darstellung kann somit nicht ausgeschlossen werden. Den Rückzug auf einen subjektiven Standpunkt quasi als Vorsichtsmaßnahme findet man übrigens auch in der Beschreibung der CS von Norbert Victor [30] im Sammelband zur Reisensburg-Konferenz.

Im Rahmen dieser Arbeit gefundene Dokumente und neu erstellte Tabellen findet man in elektronischer Form in www.dkfz-heidelberg.de/biostatistics/History-ComputationalStatistics-DR-IBS.html.

3. Was ist Computational Statistics?

Das oben vorangestellte Zitat von R. A. Fisher [9] sollte verdeutlichen, dass Datenanalyse ein fundamentaler Bestandteil der statistischen Wissenschaft ist. In einem viel beachteten Essay hat Norbert Victor im Jahr 1984 CS als Werkzeug des Statistikers definiert [29]. Obwohl er der CS eine Eigenständigkeit als wissenschaftliches Teilgebiet abspricht, da ihr sowohl die eigene Methodik als auch der eigene Untersuchungsgegenstand fehle, weist er ausdrücklich auf die große Bedeutung der CS für die Statistik und ihre Anwendungen hin. Dabei stellt er den Wert der CS für die wissenschaftliche Erkenntnisgewinnung in der Biometrie und für die Fachgebiete biometrischer Anwendungen deutlich heraus. Victor kommt zum Schluss, dass Rechnen (computation) und Datenanalyse von Anfang an ein integraler Bestandteil der Statistik waren. CS ist die Methode, mit der sich Statistiker ihre Arbeit immer vereinfacht haben, erst durch geschicktes Kopfrechnen, dann durch die Benutzung von Tabellen und den Einsatz mechanischer Rechenmaschinen, und heute mittels Computer. In der Diskussion des Essays von Victor (1984) weisen namhafte Statistiker darauf hin, dass CS von sich aus die Zusammenarbeit mit anderen Wissenschaften sucht und findet, z. B. mit den Computerwissenschaften und der Informatik. Carlo Lauro [16] definiert CS als die Art und Weise, wie man Statistik im Zeitalter der Computer betreibt (‚as the application of computer science to statistics'). Eine ähnliche Definition gibt Andrew Westlake: ‚*CS relates to the advance of statistical theory and methods through the use of computational methods*'. John Nelder [23] definiert CS eher als Technologie und nicht so sehr als Wissenschaft, auch wenn CS von Wissenschaftlern vertreten wird. Ohne auf Definitionsprobleme einzugehen kennzeichnet John Chambers [3] CS unter Verweis auf John Tukey als Konzepte des Rechnens mit Daten. Von John Tukey selbst stammt die Definition von CS als ‚*peaceful collision of computing with statistics*', zitiert nach Chambers [3].

In seinem Vortrag auf dem Biometrischen Kolloquium 1990 in Marburg über ‚Tendenzen des Statistical Computing: Keine Zukunft für Workstations?' [27] weist Bernd Streitberg der CS eine vorantreibende Funktion in Statistik und Biometrie zu: '*Das Gebiet des Statistical Computing hat die Chance, zum Motor des Fortschritts in der gesamten Biometrie und Statistik zu werden. Dazu müssen wir freilich unsere Fixiertheit auf Programmpakete überwinden, die mit den Methoden von gestern eine Statistik von vorgestern implementieren, und uns den neuen, interdisziplinären Herausforderungen stellen.*' Bernd Streitberg prophezeite damals den PCs eine große Chance vor allem wegen ihrer höheren Innovationskraft in der Entwicklung neuer Software, und er geht mit den biometrischen Abteilungen großer Unternehmen ins Gericht, welche Software-Validation so verstünden, als ob es gleichgültig wäre, wie ineffizient, benutzerunfreundlich oder gar statistisch falsch man rechnet - 'Hauptsache, alle rechnen so'. Streitberg setzte in diesem Beitrag einen Gedanken fort, den er bereits ein Jahr zuvor beim Biometrischen Kolloquium 1989 in Celle so formulierte: ‚*Statistische Konstruktionen, die nicht auch in finiten Modellen gelten, sind falsch oder irrelevant. Der Rechner wird zum Prüfstein der Umsetzung eines statistischen Verfahrens: Wenn die Methode nicht programmierbar ist, ist sie nicht relevant. Wenn die Methode nicht programmiert wird, ist sie nutzlos.*'

4. Entwicklung der Computational Statistics in Deutschland

Die Vorgeschichte der CS ist schwer zu beschreiben, da verschiedene Strömungen in den sechziger Jahren des 20. Jahrhunderts zusammenflossen, aus denen schließlich die CS entsprang. Vorläufer und Wegberei-

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gbp(M2005061411)

ter der CS sind die Anwendungen von Rechenverfahren auf empirische Daten zur Lösung wissenschaftlicher Fragestellungen, die Verfügbarkeit von elektronischen Rechenmaschinen (Computer) hervorgegangen aus den mechanischen (Babbage, 1830 - Hollerith, 1890 - Fisher, 1930 - von Neumann, Zuse, 1940-1950), und die Entwicklung von Programmiersprachen wie z. B. FORTRAN und ALGOL. Ein ganz wichtiges Moment für die Entwicklung der CS war das Auftreten von Statistiksoftware. Das erste Programmpaket BMD (später BMDP) erschien bereits 1962, gefolgt von Genstat aus Rothamstead in 1965, SAS in 1966 und SPSS in 1967. Auch das 1973 von der AG Statistical Computing der Royal Statistical Society in Großbritannien herausgebrachte GLIM hatte mit seiner besonderen Ausrichtung auf Generalisierte Lineare Modelle einen wichtigen Einfluss auf die Entwicklung der CS und auf die Biometrie.

• 4.1 Die Anfänge der CS im Westen Deutschlands

Die Geschichte der CS im Westen Deutschlands lässt sich bis heute grob in sechs Phasen beschreiben (siehe Tabelle 1a). Die Entwicklung der CS begann in den 1970er Jahren und war - wie auch die Entwicklung der Biometrie - durch die Verwüstungen des Nazi-Regimes und des Weltkrieges gegenüber den U.S.A. und Großbritannien um mindestens 10 Jahre verzögert. So hatte in Großbritannien die Entwicklung der CS bereits im Dezember 1966 in Chilton mit einer Tagung über ‚Statistical Programming' begonnen, welcher 1967 die Gründung einer AG ‚Statistical Computing' und die Serie ‚Statistical Algorithms' im Journal of the Royal Statistical Society Teil C (Applied Statistics) folgte. Dieser Entwicklung entsprachen in Deutschland die Arbeitstagungen No. 1-7 des von Norbert Victor geleiteten DVM 107 Projekts zwischen 1973 und 1975, die Gründung der GMDS-AG Statistische Auswertungssysteme im Jahr 1975 und der Start des Statistical Software Newsletter (SSN) in München in der GSF im Jahr 1974. Fast zeitgleich zur Entwicklung in Großbritannien begann die Entwicklung der CS in den USA mit dem ersten Interface Symposium über ‚Computer Science and Statistics' am 1.2.1967 in Santa Monica, der Conference on ‚Statistical Computing' in Medicine in 1969, der Gründung eines ASA ‚Committee on Computers in Statistics' in 1968, der nachfolgenden Einrichtung einer ‚Statistical Computing Section' der ASA in 1972 und der Publikation ihrer Proceedings seit 1975. Die amerikanischen Aktivitäten waren von Beginn an stärker auf das Rechnen gerichtet, was sich auch in den Namensgebungen ‚Interface' und ‚Computing' niederschlug. Eine authentische Quelle zur 2. Phase der Entwicklung der CS in Deutschland zwischen 1970 und 1975 ist die Beschreibung von Norbert Victor im Jubiläumsband ‚25 Jahre Reisensburg-Konferenz'. Nach Victor (1994) waren es die computer-intensiven statistischen Auswertungen und die Auswertungen großer Datensätze, welche die CS im Westen Deutschlands vorantrieben. Später kamen grafische Datenanalysen, computerisierte Entscheidungsverfahren und die statistische Beratung mittels wissensbasierter Systeme hinzu.

Die folgenden Phasen der Entwicklung der CS im Westen ab Mitte der 1970er Jahre werden im Abschnitt 4.3 beschrieben. Zunächst soll im folgenden Abschnitt auf die Entwicklung der CS in der DDR eingegangen werden.

• 4.2 CS im Osten Deutschlands

Ich glaube, dass Denk-, Entscheidungs- und Produktionsprozesse durch Mathematik und Rechentechnik vorangebracht werden können.

J. Läuter (1981) [18]

Im Osten Deutschlands verlief die Entwicklung der CS grundsätzlich nicht anders als im Westen, jedoch unterscheiden sich die Organisationsformen; insbesondere war die CS stärker mit der Mathematik und der Informatik verbunden (vgl. Tabelle 1b). Der oben vorangestellte einleitende Satz aus dem Buch ‚Programmiersprache DIST' - eine Abkürzung für Daten-Interpretation, -strukturierung und -transformation - drückt dies zum Teil schon aus. Die Computerwissenschaften waren in der DDR genauso aus den Rechenzentren entstanden wie im Westen, blieben aber länger in der Tradition der Mathematik und somit waren die Anwendungen des Computers stärker durch allgemeine mathematische Prinzipien geprägt. Die Anwender - also auch die Biometriker - waren meist in der Mathematik ausgebildet. In den 1960er Jahren entstanden im Rechenzentrum der Humboldt-Universität zu Berlin einige allgemeine leistungsfähige Programme (z. B. für die mehrfaktorielle Varianzanalyse) mit variablen Steuermöglichkeiten für eine Vielzahl von Anwendungen. So erstellte Jürgen Läuter ca. 1965 ein Programm zur Mehrfaktoriellen Varianzanalyse für ein n-faktorielles und unbalanziertes Design. Ähnlich wie im DMV107 Projekt im Westen gab es ein gründliches Nachdenken über Datenstrukturen, welches Anwender wie z. B. Mediziner, Landwirte und Pädagogen einschloß, u.a. beeinflusst durch das Buch ‚Statistical Computation', herausgegeben von R.C. Milton und J.A. Nelder [20]. Aus dieser Zeit stammt auch die Dissertation des DIST-Entwicklers Läuter mit dem Titel ‚Entwicklung mathematisch-statistischer Algorithmen und ihre Realisierung auf Rechenautomaten' [17]. DIST war eine algorithmische Programmiersprache zur Spezifikation der Dateneingabe, die zwischen 1969 und 1971

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.

Tabelle 1: Wurzeln und Entwicklungsphasen der 'Computational Statistics' in Deutschland

a) Im Westen

Zeitraum	Entwicklungsphase	Treibende Kräfte
Vor 1970	Vorgeschichte, Entwicklung von Programmiersprachen, erste Statistikpakete BMD, SAS, SPSS in USA, Genstat in UK, STATSYS in Deutschland	Rechenzentren an Universitäten
1970-1975	Verbesserung, Standardisierung und Verbreitung von statistischen Auswerte-Programmen	DVM-107 Projekt „Standardisierung von Statistiksoftware auf dem Gebiet der Medizinischen Statistik“ gefördert vom BMFT 1973-1975
1976-1980	Portabilität und Erweiterung von Programm-paketen, Interface zwischen Programmen und Datenbanken	Reisensburg-Konferenz, GMDS-AG Statistische Auswertungssysteme, medis-Institut, München
1981-1984	Problemorientierte statistische Software, Qualität der Software	Reisensburg-Konferenz, GMDS-AG Statistische Auswertungssysteme, SoftStat Konferenzen
1985-1993	Behandlung aller aktuellen Probleme, Methoden und Werkzeuge der Informatik für die Statistik, 25. Reisensburg-Konferenz mit dem physica-Jubiläumsband (Hrsg: P. Dirschedl und R. Ostermann)	Reisensburg-Konferenz, GMDS-AG Statistische Auswertungssysteme, AG Computational Statistics (CS) der DR-IBS, SoftStat Konferenzen
1994-2003	Methoden und Werkzeuge der CS für aktuelle Fragestellungen der angewandten Statistik unter besonderer Berücksichtigung der Biowissenschaften und der Medizin	Reisensburg-Konferenz GMDS-AG Statistische Auswertungssysteme, AG CS der DR –IBS, AG Datenanalyse und Numerische Klassifikation der GfKL

b) Im Osten

Zeitraum	Entwicklungsphase	Treibende Kräfte
ab 1970	PP STATISTIK für ESER-Rechner	VEB Robotron Dresden
	Komplettierung des Systems, verbesserte Dateneningabe, Schaffung neuer Steuerungs-möglichkeiten	Rechenzentrum der Universitäten Institute der Akademie der Wissenschaften (AdW)
1981	Ergänzungen zum PP STATISTIK u.a. Dialogprogramme Pfadkoeffizienten Spektralanalyse Nichtlineare Parameterschätzung Mehrdimensionale Häufigkeiten Clusteranalyse Zusammenhangsanalyse	Rechenzentren der Universitäten

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.

(Fortsetzung)

Tabelle 1: Wurzeln und Entwicklungsphasen der 'Computational Statistics' in Deutschland

Zeitraum	Entwicklungsphase	Treibende Kräfte
	Entwicklung von MASTAT für PDP11-Nachbau	Forschungszentrum für Tierproduktion Dummerstorf
1983	Modernisierung der PP STATISTIK, und Ergänzung weiterer Verfahren (IV. Spezialtagung über Programme und Anwendungen der Mathematischen Statistik, Leipzig) Anforderungen an Software Programmsystem MADRAS Logistische Regression Ausreißererkennung Faktorenanalyse Confounder	VEB Robotron Dresden, Rechenzentren der Universitäten, Medizinische Akademie Dresden
1984	Programmsystem MVD, MANOVA und Diskriminanzanalyse, mit Modellwahl und Fehlerschätzung	Mathematisches Institut der AdW
1986	Programmpaket STAVE für PC	VEB Robotron Dresden
1986-1988	Behandlung von verschiedenen Themen in der Schriftenreihe ‚Informationsverarbeitung im Hoch- und Fachschulwesen' Nutzerorientierung Modellbildung mit multipler Regression, Statistik auf Kleinrechnern, Schnittstellen Grafische Verfahren Faktorenanalyse Clusteranalyse Expertensysteme	

c) Einflussreiche Gruppen im Osten mit CS Anwendungen

- Arbeitskreis der Universitätsrechenzentren und angegliederte Institutionen: Mathematische und statistische Verfahren (Algorithmen, Programme)
- Rechenzentrum der VVB Saat- und Pflanzgut in Zusammenarbeit mit der Akademie der Landwirtschaftswissenschaften und Universitätseinrichtungen: Datenspeicher Datenspeicher, Versuchsergebnisse, Pflanzenproduktion (DAVEP)
- Arbeitsgemeinschaft Statistik in Verbindung mit Akademie- und Universitätseinrichtungen: Verfahrensbibliothek, Biometrisches Wörterbuch, CADEMO
- Gesellschaft für physikalische und mathematische Biologie der DDR mit der Sektion Biomathematik, vergleichbar der Deutschen Region der Biometrischen Gesellschaft und quasi als Ersatz gegründet, da die Kontakte zur BRD abgebrochen werden mussten, vgl. Enderlein et al. (1993) [8].

von Läuter in den Grundzügen entwickelt, 1974 für die ESER-Rechner der DDR implementiert und ab 1977 im praktischen Einsatz des (statistischen) Rechnens war. Nach einer Unterbrechung wurden diese Arbeiten in den 1980er Jahren weitergeführt, da Interesse sowohl im Rechenzentrum der Humboldt-Universität als auch beim Herz- und Kreislaufinstitut in Berlin-Buch bestand. Außer kostenloser Rechenzeit gab es aber für das Paket keine wesentliche Unterstützung. Bis 1989 wurde DIST etwa 20mal an andere Institutionen verkauft. Im Rückblick ist festzuhalten, dass DIST ähnlich wie andere Systeme als praktisches Werkzeug die Entwicklung der CS in der DDR mitprägte.

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gbp0[2005]014.fr)

Die Entwicklung der CS in der DDR war zunächst einmal mit der Entwicklung von statistischer Standardsoftware und von Spezialprogrammen befasst, sowohl für den eigenen Bedarf als auch für denjenigen der Sowjetunion und andere Länder. Diese Entwicklung erfolgte vor allem im VEB Großforschungszentrum Robotron Dresden unter der Leitung von Rainer Weber. So entstand Anfang der 1970er Jahre, unter Einbeziehung von Hochschulen und Universitäten, als kommerziell vertriebenes Softwareprodukt für die ESER-Rechner (Nachbau der IBM-Serien 360-380, Einheitssystem der elektronischen Rechentechnik) das Programmpaket STATISTIK (PP STATISTIK), das in der DDR ca. 100 und im östlichen Ausland ca. 20 Installationen aufweisen konnte. Es umfasste:

• Dateneingabe und -manipulation

• Maßzahlenbestimmung und Häufigkeitsanalsye

• Anpassungstests

• parametrische und nichtparametrische Tests

• Korrelation

• Regression

• Faktoranalyse

• Varianzanalyse

• Zeitreihenanalyse

• Clusteranalyse

• Diskriminanzanalyse

• Pfadkoeffizienten

Bis in die frühen 1980er Jahre war es ein wesentliches Ziel, Software für die ESER Rechner bereitzustellen und weiter zu entwickeln (vgl. Tabelle 1b). Für den Nachbau der DEC-PDP11-Linie (K1620, K1630) wurde im Forschungszentrum für Tierproduktion in Dummerstorf im Auftrag des VEB Robotron Projekt Dresden das Programmpaket PP MASTAT für die mathematische Statistik entwickelt mit Modulen zu:

• Datentransformation

• Sortierung

• Elementare Datenaufbereitung

• Varianzanalyse mit vollständiger Kreuzklassifikation mit bis zu 6 Faktoren

• Hierarchische Varianzanalyse mit bis zu 4 Faktoren

• Quasilineare Regression

• Nichtlineare Regression

• Zeitreihen

An den Universitäten und der Akademie der Wissenschaften befasste sich die CS mit leistungsfähigen Verfahren für Anwendungen der Statistik. Als Beispiel kann das System MADRAS von Rainer Koch an der Medizinischen Akademie Dresden angeführt werden. Zeitlich vergleichbar mit der Gründung der AG Statistische Auswertungssysteme der GMDS im Westen, gab es in den 70er-Jahren einen Arbeitskreis unter Leitung von Christian Noack und in den 1980er-Jahren unter Egmar Rödel, beide von der Humboldt-Universität, der sich einmal jährlich zu einer mehrtägigen Klausurtagung (meist in Warnemünde) traf und sich bezüglich der zu bearbeitenden statistischen Methoden abstimmte. Zu erwähnen bleibt noch eine staatliche Initiative:

In den 1970er Jahren waren vom Ministerium für Hoch- und Fachschulwesen Hauptforschungsrichtungen festgelegt worden. CS war dabei ein Themenkomplex innerhalb der Richtung „Rechentechnik und Kybernetik", der jedoch dann Mitte der 1980er Jahre gegen den Widerstand ihrer Mitglieder wieder aufgelöst wurde. Es bildete sich aus den Mitgliedern des CS Themenkomplexes eine Arbeitsgruppe, die sich mit Fragen der CS beschäftigte, bis dann diese Arbeitsgruppe nach der Wende zerfiel.

• 4.3 Die Tagungen auf der Reisensburg

Die Entwicklung der CS im Westen Deutschlands ist eng verbunden mit Personen, die sich aktiv für diesen Fachbereich einsetzten, wie Norbert Victor, Allmut Hörmann und auch Peter Naeve, der für viele Jahre Mitherausgeber der Zeitschrift CSDA mit Verantwortung für Europa war. Die Entwicklung der CS ist aber auch eng verbunden mit einem reizvoll gelegenen Tagungsort, der jährlich die CS-Gemeinde zusammenführt. Die Rede ist von den Arbeitstagungen auf der Reisensburg bei Ulm, deren Anziehungskraft ungebrochen ist, und denen es gelang, in einer Art von Kontinuität im Wandel immer wieder neueste Entwicklungen der CS in ihrem Programm aufzugreifen. Eine Liste aller bisherigen Arbeitstagungen auf der Reisensburg mit ihren Themenschwerpunkten findet man auf der oben zitierten Internetseite. Für die ersten 25 Jahre hat Allmut Hörmann [15] die Errungenschaften der Reisensburgtagungen wie folgt zusammengefasst:

1975 Gründung des Statistical Software Newsletter (SSN) mit Unterstützung des medis Instituts der GSF München

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gbp(M2005061411)

1975-1978 Anforderungskatalog für Programmsysteme für die statistische Datenauswertung

1986-1988 Nichtparametrische Auswerteverfahren

1986-1990 Kriterien zur Bewertung von Software

1991-1995 Statistik Software Guide

1990-1993 Numerische Zuverlässigkeit

5. AG Computational Statistics der Deutschen Region der Internationalen Biometrischen Gesellschaft

• 5.1 Die Gründung

Es entbehrt nicht einer gewissen Ironie, dass die bisher wohl größte Fehleinschätzung der CS einer CS-Arbeitsgruppe in Deutschland zur Geburt verhalf. Auf Initiative ihres damaligen Sekretärs Heinz Hochadel gründete die Deutsche Region (DR) der Internationalen Biometrischen Gesellschaft am 13.3.1985 in Bad Nauheim eine Arbeitsgruppe mit dem späteren Namen ‚Computational Statistics' in engem Zusammenwirken mit Repräsentanten der GMDS, nämlich Reinhold Haux (Aachen) und Karl-Heinz Jöckel (Bremen). Beide waren in der GMDS AG Statistische Auswertesysteme aktiv und wurden von der DR mit der Leitung dieser neuen AG betraut (Tabelle 2).

Tabelle 2: Die AG 'Computational Statistics' der DR-IBS und ihre Leitung in 4-jährigem Zyklus

Zyklus	Leitung	Stellvertretung
1985-1988	R. Haux	K.-H. Jöckel
1988-1991	K.-H. Jöckel	R. Ostermann
1991-1994	R. Ostermann	G. Sawitzki
1994-1997	G. Sawitzki	M. Nagel
1997-2000	E. Schuster	U. Mansmann
2000-2004	U. Mansmann	M. Theus

Der Gründungsitzung war ein Aufruf Heinz Hochadels zur Gründung einer AG ‚Computing Statistics' vorausgegangen. Diese AG sollte sich des Problems der Expertensysteme annehmen. Die Biometrische Gesellschaft sollte versuchen, einen Missbrauch solcher Programmsysteme, die eine automatisierte statistische Auswertung anstreben, zu verhindern oder aktiv an der Entwicklung geeigneter Systeme arbeiten. Im Wortlaut:

Die Bedeutung des Begriffs „Expertensystem" ist nicht einheitlich. In Bezug auf Statistik wird er in folgendem Sinne verwendet: Zur Zeit sind Programmsysteme im Entstehen, die eine automatisierte statistische Auswertung anstreben. Als Eingabe dienen nur die Daten; Modellbildung und Test werden automatisch durchgeführt. Der AK soll sich z. B. folgender Fragen annehmen:

- Lässt sich aus den Daten der Versuchsplan eindeutig ableiten?

- Inwieweit lassen sich Voraussetzungen automatisch prüfen?

- Bei welchen Methoden gibt es eine eindeutige Beziehung zwischen Datenstruktur und Inferenzmodell?

- Welche Kennungen müssen den Daten beigegeben werden, damit die statistische Auswertung automatisiert werden kann?

- In welchen Grenzen und bei welchen Anwendungen kann eine automatisierte Auswertung eingesetzt werden?

Die Biometrische Gesellschaft soll versuchen, einen Missbrauch solcher Programmsysteme zu steuern oder aktiv an der Entwicklung teilzunehmen.

Die Namensgebung der neuen AG der DR war mit Hindernissen verbunden. Der Vorschlag der Gründungssitzung war ‚Informatik in der Statistik (Computational Statistics)'. Entsprechend führte die AG zu Beginn zwei Namen ‚Informatik in der Statistik' als Kurzform und ‚Informatik in der Statistik/Computational Statistics' als Langversion. Mit dieser Namensgebung waren Vorstand und Beirat der DR offenbar nicht einverstanden, zumal die DR im Jahr 1985 mit einer Geschäftsordnung für ihre AGs beschäftigt war, bei welcher auch Namensgebungen geregelt werden sollten. Dabei sollten substanzwissenschaftlich orientierte AGs nach ihrer entsprechenden Substanzwissenschaft (z. B. Biometrie in ...) und methodisch orientierte AGs nach ihren Methoden (z. B. Nichtparametrische Methoden, Generalisierte Lineare Modelle) bezeichnet werden. Die neue AG verstand sich zwar als methodische AG, hielt aber eine Umbenennung zum Zweck der Vereinheitlichung für nicht sinnvoll. In der Arbeitssitzung in Ulm 1986 gab sich die AG dann den Namen ‚Computational Statistics' und sie bat mit Schreiben vom 16.4.1986 Vorstand und Beirat der DR um ihre Zustimmung, da die AG zur Ansicht gekommen sei, dass die Bezeichnung ‚Informatik in der Statistik', die als vorläufige Lösung vom Beirat vorgeschlagen war, zu allgemein wäre. Für das Kolloquium in Ulm im Jahr 1986 war die AG noch als AG ‚Informatik in der Statistik' eingeplant, hat aber seit April 1986 ihren heutigen Namen. Dem Durchsetzungsvermögen der beiden Leiter Haux und Jöckel kam sicher auch zu Hilfe, dass die AG im Dezember 1985 eine erfolgreiche Tagung zu Expertensystemen in der Statistik mit einem Ta-

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.

gungsband beim Fischer-Verlag hatte krönen können. Die AG Computational Statistics (AG CS) hat sich dieses Selbstbewusstsein über die Jahre erhalten und konnte so zusammen mit ihrer aktiven Mitgestaltung der Reisensburgkonferenzen kontinuierlich einen selbständigen Beitrag zur Entwicklung der CS in Deutschland leisten.

• 5.2 Die ersten Jahre

Die Gründung der AG CS fiel in eine Zeit der Euphorie für Expertensysteme. Es wurden Projekte entworfen und Finanzmittel verteilt. So war die Zeit der Gründung der AG CS eine sehr aktive Zeit des Computing und der Softwareentwicklung. Reinhard Hilgers formulierte dies auf dem Biometrischen Kolloquium 1987 in Trier wie folgt:

‚Die Entwicklung bei der „Statistischen Software" der letzten Jahre hat zwar eine schier unüberblickbare Palette von Produkten kommerzieller und nicht-kommerzieller Herkunft hervorgebracht; die notwendige globale Integration der Aufgaben des Biometrikers bei Planung, Monitoring, Auswertung und Darstellung ist bisher nicht erfolgt. Da solche Art Software für den Biometriker nicht nur keine Hilfe ist, sondern ihn bei seiner Arbeit behindern kann und wird, sollte gerade die Biometrische Gesellschaft mit ihren Arbeitsgemeinschaften Anstrengungen unternehmen, vorhandene Software unter biometrischen Gesichtspunkten zu sichten, zu prüfen und für ihre Mitglieder Empfehlungen auszusprechen sowie in ihrem Namen Forderungen für zukünftige Entwicklungen aufzustellen.'

• 5.3 Die Zeit der SoftStat-Konferenz

Die SoftStat Konferenz hat mehr als 15 Jahre die Entwicklung der CS in Deutschland mit beeinflusst und international sichtbar gemacht. Als Konferenz über die wissenschaftliche Anwendung von Statistik-Software wurde sie organisiert von der ZUMA (Zentrum für Umfragen, Methoden und Analysen e.V., Mannheim) mit dem Ziel, die Anwendung von Statistik-Software in verschiedenen wissenschaftlichen Bereichen darzustellen und Softwaresysteme zu bewerten und zu vergleichen. Es wurden bewusst Software-Hersteller in die Konferenz miteinbezogen, ohne dass die Tagung in eine Produktwerbung abglitt, und es wurden spezielle Fragestellungen bearbeitet, ohne dass man sich in den technischen Details der Programmierung verlor. Die 3. SoftStat wurde ein Jahr nach Gründung der AG CS dann auch von einem Biometriker organisiert und zwar von Walter Lehmacher 1985 in München im medis Institut der GSF. Von den insgesamt 50 Beiträgen bei der 3. SoftStat kann man 14 (28%) der Biometrie zurechnen. Dabei ging es um statistische Methoden (Kontingenztafeln, Verlaufskurven, Regression, ANOVA), um Anwendungen (Zell-/Überlebenskurven, Bioassays, Tumormarker), aber auch um Fallzahlprogramme, GLIM, Versuchsplanung und Datenanalyse. Zuvor hatten die ersten beiden SoftStat-Konferenzen 1981 und 1983 in Mannheim stattgefunden. Dazwischen lag die APL 82-Konferenz in Heidelberg mit 60 Beiträgen nur zur Programmsprache APL, die damals für statistische Auswertungen eine bedeutende Rolle spielte. Am DKFZ wurde in dieser Zeit ein eigenes APL Softwaresystem für die Analyse der ersten zu dieser Zeit gerade aufkommenden DNA-Sequenzen aufgebaut zunächst von Gerd Osterburk und dann übernommen von Sandor Suhai; Entwicklungen, die in den USA erst viel später, aber dann mit hoher Durchschlagkraft stattfanden. Man kann historisch lediglich spekulieren, was gewesen wäre, hätte man diese Entwicklungen in Deutschland besser gefördert.

Die 4. SoftStat fand 1987 in Heidelberg statt mit 66 Beiträgen, von welchen 16 (24%) als biometrienahe bezeichnet werden können. Dabei ging es um multivariate Verfahren (ANOVA, Regression, Kovarianzstruktur, Korrespondenzanalyse, Faktorenanalyse), robuste Verfahren, Klassifikation, Modellbildung und kategoriale Daten. Bootstrap war ein Thema und natürlich die Expertensysteme. Von da an blieb die SoftStat in Heidelberg und sie fand zum letzten Mal im Jahr 1995 statt. Von den damaligen 101 Beiträgen kann man 34 (34%) in die Nähe der Biometrie setzen. Wieder vertreten waren multivariate Verfahren, allerdings spezialisierter als früher (Faktorenanalyse, Kovarianzanalyse, ‚multilevel' Model, mehrdimensionales Skalieren, ARMA, Rasch Modell). Weiterhin gab es ein ganz breites Spektrum von statistischen Themen: Kernschätzer und Smoothing, Repeated Measurement, inexakte Daten, fehlende Werte, Extremwerte und Clustering. Die Ära der SoftStat-Konferenz endet dann abrupt. Nachdem die DR die Planungen eines Biometrischen Kolloquiums für 1997 in Heidelberg nicht weiterführt, fehlt auch der SoftStat für 1997 die erforderliche organisatorische Grundlage, zumal der Hauptorganisator Faulbaum von seiner Stelle bei der ZUMA nach Düsseldorf gewechselt war. Für die AG CS war dies besonders schmerzlich, da sie in einer Verbindung zwischen Biometrischem Kolloquium und SoftStat eine Chance sah, ihre Aktivitäten darzustellen und ihre Funktion als Bindeglied zwischen Biometrie und Informatik deutlich zu machen. Gleichzeitig ist die AG CS ab dieser Zeit nicht mehr mit eigenständigen Themen auf dem Biometrischen Kolloquium vertreten (vgl. Tabelle 3).

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gbp(M2005061-1)

GMS Medizinische Informatik, Biometrie und Epidemiologie ISSN 1860-9171

Tabelle 3: Themen und Präsenz der CS während des Biometrischen Kolloquiums zwischen 1969 und 2003

Kolloqium	Zeit	Ort	Anzahl Beiträge	Anzahl CS Beiträge		CS Beiträge und Sitzungsthemen (In Klammern die Anzahl von Beiträgen)
16.	20.-22.2.1969	Bad Nauheim	41	9	(22%)	"Probleme der Biomathematik" mit Thema: Analogrechner
	16.-21.8.1970	Hannover				7. IBC
17.	1.-3.4. 1971	Freiburg	37	7	(20%)	"Diskriminanzanalyse"
18.	23.-25.3.1972	Bad Nauheim	32	6	(19%)	"**EDV und Biometrie**"
19.	29.-31.3.1973	Berlin	28	2	(7%)	"**EDV und Biometrie**"
20.	20.-22.2.1974	Bad Nauheim	28	1	(4%)	---
21.	5.-7.3. 1975	Hohenheim	31	2	(6%)	---
22.	10.-12.3.1976	Bad Nauheim	40	9	(22.5%)	"Exakte Tafeln"
23.	9.-11.3. 1977	Nürnberg	42	2[*1]	(5%)	"Erfahrung mit biometrischen EDV Programmen" [*2]
24.	1.-3.3. 1978	Wuppertal	24	11	(33%)	"**EDV und Biometrie**" " Kurvenschätzung"
25.	6.-9.3. 1979	Bad Nauheim	43	10	(23%)	"**EDV und Biometrie**" " Robuste Verfahren"
26.	17.-20.3. 1980	München	33	-	-	Workshop über EDV [*2]
27.	11.-13.3. 1981	Bad Nauheim	42	3	(7%)	" Explorative Datenanalyse" (5) " Nichtparametrische Verfahren" (8)
28.	16.-19.3. 1982	Aachen	38	2	(5%)	AG Sitzung Nichtparametrische Verfahren[*3] (4)

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.

(Fortsetzung)

Tabelle 3: Themen und Präsenz der CS während des Biometrischen Kolloquiums zwischen 1969 und 2003

Kolloqium	Zeit	Ort	Anzahl Beiträge	Anzahl CS Beiträge	CS Beiträge und Sitzungsthemen (In Klammern die Anzahl von Beiträgen)
29.	8.-11.3. 1983	Bad Nauheim	34	4 (12%)	AG Sitzung Nichtparametrische Verfahren (4)
30.	14.-16.3. 1984	Dortmund	39	3 (8%)	AG Sitzung Nichtparametrische Verfahren (8) " Explorative Statistik" (8)
31.	12.-15.3. 1985[*4]	Bad Nauheim	54	8 (15%)	AG Sitzung Nichtparametrische Verfahren (7) "Grafische Darstellung von Verlaufskurven" (3)
32.	18.-21.3. 1986	Ulm	67	5 (7%)	AG Sitzung **Informatik** (5) AG Sitzung Generalisierte Lineare Modelle (4) AG Sitzung Nichtparametrische Verfahren (5)
33.	16.-20.3. 1987	Trier	62	8 (13%)	AG Sitzung **Computational Statistics** (3) AG Sitzung Generalisierte Lineare Modelle (2) "Biometrie und moderne Informationstechnologie" (4)
34.	21.-25.3. 1988	Bad Nauheim	55	5 (9%)	AG Sitzung **Computational Statistics** mit AG Sitzung Generalisierte Lineare Modelle (5)
35.	27.2.-2.3. 1989	Celle	77	8 (10%)	AG Sitzung **Computational Statistics** (5)
36.	14.-16.3. 1990	Marburg	69	8 (12%)	AG Sitzung **Computational Statistics** zu "Statistische Software für nichtmetrische Daten" (4) "Arbeitsplatzrechner in der Statistik" (4)
37.	19.-22.3. 1991	Hamburg	86	9 (10%)	"Computerintensive und Graphische Modelle" (6)
38.	16.-20.3. 1992	Giessen	75	10 (13%)	AG Sitzung **Computational Statistics** (4)
39.	16.-19.3. 1993[*5]	Berlin	43	2 (5%)	----

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gbp(M2005061414.1)

(Fortsetzung)

Tabelle 3: Themen und Präsenz der CS während des Biometrischen Kolloquiums zwischen 1969 und 2003

Kolloqium	Zeit	Ort	Anzahl Beiträge	Anzahl CS Beiträge		CS Beiträge und Sitzungsthemen (In Klammern die Anzahl von Beiträgen)
40.	15.-18.3. 1994	Münster	119	20	(17%)	AG Sitzung **Computational Statistics** zu „Aspekte der Lehre und Didaktik (4) "Neuronale Netze" (5) "Sampling and Reseampling" (3)
41.	14.-17.3. 1995	Hohenheim	93	10	(11%)	"Statistische Methoden und Software-Praxis" (5)
42.	12.-15.3. 1996	Magdeburg	100	13	(13%)	"Softwaredemonstrationen" (4) *6
43.	18.-21.3. 1997	München	106	12	(11%)	-----
44.	16.-19.3. 1998	Mainz	95	7	(7%)	-----
45.	16.-19.3. 1999	Dortmund	97	8	(8%)	
46.	20.-23.3. 2000	Rostock	94	7	(7%)	"Softwaredemonstration" (1)
47.	20.-23.3. 2001	Homburg	92	10	(11%)	-----
48.	21.-26.7. 2002	Freiburg mit IBC	234	15	(6%)	„Bioinformatics“

*1) Die Tagung war zu einem wesentlichen Teil in 6 Workshops organisiert, für welche die Beiträge nicht im Programm standen. Damit ist der Anteil der CS in dieser Tagung sicher höher als 5%.
*2) Organisator: H. Geidel
*3) Für das 28.- 33. Biometrische Kolloquium werden die AGs Nichtparametrische Methoden und Generalisierte Linear Modelle in der Tabelle aufgeführt.
*4) Zusätzlich 5.-7.12.1985, Tagung der AG CS in Aachen zu Expertensystemen mit Herausgabe eines Tagungsbandes ‚Expert Systems in Statistics‘ im S. Fischer Verlag
*5) Zusätzlich 8.-10.11.1993, Tagung in Stift Keppel zu Statistische Programmiersprachen und Interaktive Datenanalyse
*6) nicht bei der Zählung der Beiträge berücksichtigt

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.

6. CS und Biometrie

The power of the computer to assist the development of biometry, both in breadth and depth, will continue, limited as much by our ability to make effective use of that power as by any limitations of the power itself.

J. A. Nelder (1996) [23]

Vier wesentliche Entwicklungen der CS, die auf die Biometrie Einfluss nahmen, werden in diesem Abschnitt besprochen.

• 6.1 Hardware-Entwicklung

Das wesentliche Element, das die Entwicklung der CS in der Biometrie weltweit vorantrieb, war und ist die fortwährende Steigerung der Schnelligkeit und Speicherkapazität der Hardware, die gewährleistet, Datensätze realistischer Größe in akzeptabler Zeit auszuwerten. Die CS hat damit die Biometrie in die Lage versetzt auch komplexere Fragestellungen anzugehen. So konnten Regressionsaufgaben vor 1970 in der Größenordnung von 10-10^2 experimentellen Einheiten ausgewertet werden, in den Zeiten der „Großrechner" zwischen 1970 und 1980 in der Größenordnung von 10^3-10^4, und heute sind Auswertungen mit mehr als 10^6 Einheiten möglich.

• 6.2 Software-Entwicklung

Bis in die frühen 1960er Jahre erfolgte die Statistikprogrammierung in Standardprogrammiersprachen und im Stapelbetrieb (batch processing). Umfassend konzipierte Statistik-Programmpakete gab es seit ca. 1960. Allerdings beschränkte sich das Angebot für die Biometrie auf einige wenige Pakete wie SAS oder SPSS. Kleinere zweck- und problemorientierte Spezialpakete hatten ihre Blütezeit zwischen 1970-1985. Objekt-orientierte Programmierung (z. B. mit der Sprache S) gibt es seit ca. 1985. Vor diesem Software-Hintergrund sah die CS in Deutschland eine wichtige Aufgabe darin, den Anwendern eine Übersicht über existierende Statistiksoftware zu geben. Neben zwei internationalen Übersichten von Schucany et al. [25] und Francis [10] wurden so verschiedene Übersichten in den AGs der DR und der GMDS erstellt. Unter der Regie von Uwe Haag und Armin Koch erschien der Software Guide im SSN mehrmals in der Zeitschrift CSDA, zum letzten Mal Anfang der 1990er Jahre. Es wurde danach immer schwieriger, eine objektive und umfassende Information über statistische Software zusammen zu tragen, da sich Softwaresysteme zunehmend überlappten. Gleichzeitig wurde es immer leichter, sich im Internet über ein Produkt umfassend zu informieren.

• 6.3 Algorithmen und rechenintensive Verfahren

Simulationsstudien, iterative numerische Verfahren und der Gebrauch von Zufallszahlen sind inzwischen ein fester Bestandteil vieler biometrischer Arbeiten geworden. Ganz offensichtlich hat die Anzahl der Publikationen mit Simulationsergebnissen deutlich zugenommen, seitdem PCs einzelnen Biometrikern ab Mitte der 1980er Jahre zur Verfügung standen. Nachdem die Software benutzerfreundlicher wurde, konnten auch größere Datenmengen leichter ausgewertet werden. Desweiteren wurden iterative Algorithmen entwickelt, um komplexe Modelle wie z. B. das Generalisierte Lineare Modell von Nelder und Wedderburn [21] oder das Generalisierte Additive Modell von Hastie und Tibshirani [14] anzupassen. Auswertungen von zufälligen Effekten in Gemischten Linearen Modellen gehören heute zum biometrischen Alltag.

Rückblickend auf die letzten 20 Jahre lassen sich eine Reihe von algorithmischen Verfahren benennen, die für den Biometriker von großem Nutzen waren:

• Splines, ‚Smoothing' und Nichtparametrische Regression

• EM Algorithmus

• Bootstrap und Resampling Verfahren

• Permutationstests und Exakte Verfahren

• Symbolisches Rechnen

• Anpassung nichtlinearer komplexer Modelle

• Markov Chain Monte Carlo (MCMC) Methode

• 6.4 Auswerteverfahren und -strategien

Lineares Modell und Regression sind biometrische Verfahren, welche die CS nachhaltig herausgefordert haben, z. B. die Entwicklung von Matrixkalkulationen und effiziente Rechentechniken für die Generalisierte Inverse. Die CS hat der Biometrie kontinuierlich Auswertemethoden zur Verfügung gestellt. In diesem Zusammenhang ist auch der Ansatz der Expertensysteme zu nennen, der die CS und zum Teil auch die Biometrie ca. ein Jahrzehnt (1980-1990) in Atem hielt, viele Hoffnungen weckte, viele Projektförderungen herauslockte, aber letztendlich seine Versprechungen nicht erfüllte. Es ist ein Verdienst der Biometrie in Deutschland, dass einer ihrer Vertreter schon früh auf die Problematik der Expertensysteme hinwies, nämlich Bernd Streitberg z. B. in seinem Beitrag im SSN [27]. John Nelder [23] brachte das Scheitern der Expertensysteme auf den Punkt: '*The primary problem in any*

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gbp(M2005061414-1)

implementation of an expert system for statistics is that it presupposes that we can write down the rules of inference in a computable form. In the area of model selection and model checking I do not think this is true; if it were, we should not see so many incompetent analyses in the published literature.'

7. Der Einfluss der CS oder Wer treibt wen?

• 7.1 Verfahren der CS für biometrische Anwendungen

Victor [29] charakterisiert CS als '*on the one hand a collection of tools which must be improved by troublesome work over and over again, and on the other hand these tools permit the penetration into new and unknown fields of statistics*'. Für den Zeitraum von ca. 1965 bis 1985 nennt er sieben Felder, in denen CS statistische Methoden und Theorie stimulierte:

a) multivariate Verfahren, einschließlich der Analyse mehrerer Kovariablen und die Variablenselektion

b) exakte Tests bzw. Permutationstests

c) nichtparametrische Verfahren einschließlich (robuste Verfahren, smoothing, Kernschätzer und splines)

d) Statistische Tafelwerke und Tabellen, Tabellierung von Verteilungen

e) Simulation, Monte Carlo Verfahren

f) explorative Datenanalyse (EDA)

g) integriertes Arbeiten an Problemen unter Einfluss integrierter Softwarepakete und Datenanalyse

Seither sind weitere Felder hinzugekommen:

h) EM Algorithmus und GEE und EE mit Anwendungen auf longitudinale Daten

i) Bootstrap und Resampling Methoden

j) MCMC Verfahren und Gibbs Sampling

k) Methoden der Computational Bayes Verfahren

l) interaktive und grafische Datenanalyse.

• 7.2 Verfahren der CS für biometrische Anwendungen

Unter den oben genannten Verfahren haben die exakten Verfahren und die Methoden der Permutationsverteilungen, die nicht-parametrischen Verfahren und die Anwendung des Simulationsprinzips in der Biometrie seit ihren ersten Jahren nichts von ihrer Aktualität eingebüßt. Multivariate Verfahren wurden von der MCMC-Methodik quasi überrollt. Statistische Tabellen spielen fast keine Rolle mehr und die EDA ist zusammen mit dem was unter integriertem Arbeiten an Problemen gemeint war, in die Praxis der interaktiven und grafischen Datenanalyse übergegangen.

Beschäftigt man sich intensiver mit den Arbeitsthemen der CS, wie sie z. B. auf den Arbeitstagungen auf der Reisensburg diskutiert wurden, so kann man zusätzlich zu den zwölf oben aufgeführten Feldern a)-l) eine Gruppe biometrischer Anwendungsbereiche ausmachen, die von der CS und ihren Auswerteverfahren profitieren:

m) Verfahren für die Berücksichtigung fehlender Werte und unvollständige Beobachtungen einschließlich der Auswertung zensierter Daten

n) Goodness-of-fit und Residuenanalysen

o) Klassifikationsverfahren, Regressionsbäume (CART)

p) Meta-Analysen

q) Sequentielle Verfahren und multiples Testen

r) Räumliche Statistik

s) rechnerunterstützte Planung und Fallzahlprogramme.

• 7.3 Wechselwirkung und gegenseitige Durchdringung

Eine Beurteilung des wechselseitigen Gehalts von Biometrie in CS-Publikationen oder von CS in Biometriearbeiten ist schwierig. Die Publikationen der CS zählen meist nicht oder nur unvollständig die möglichen Anwendungsmöglichkeiten auf und meist fehlt eine genauere Beschreibung des Nutzens für das biometrische Beispiel. Auf der anderen Seite benutzt die Biometrie Methoden der CS, um ihre Arbeit durchzuführen und stellt dann die CS-Methode nicht in den Vordergrund ihrer Publikationen.

Um einen Eindruck zu erhalten, wie CS Methoden in biometrische Arbeiten Einzug fanden und wie CS-Arbeiten mit Anwendungen der Biometrie umgehen, wurden zwei Jahrgänge der Biometrics, des Biometrical Journal und der CSDA ausgewählt und miteinander verglichen. In den beiden biometrischen Zeitschriften wurden diejenigen Arbeiten bestimmt, welche in einer Überschrift einen der folgenden Begriffe aufwiesen:

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.

numerical analysis, computational aspects, Monte Carlo, simulation, implementation, algorithms.

In den CS-Arbeiten wurde nachgesehen, ob eine biometrische Anwendung beschrieben wurde. Das Ergebnis dieser Suche ist auf der oben zitierten Internetseite im Einzelnen dargestellt. Während im gesamten Band der Biometrics von 1985 26/91 (28.6%) Arbeiten eine CS-Methode explizit zitierten, waren es in 2000 im ersten von vier Heften 27/39 (69.2%). Der Unterschied ist signifikant (Exakter Fisher Test: $p<10^{-5}$, OR=5.6 mit 95%, CI 2.5-12.8). Zum Vergleich sind es im Jahr 1985 im Biometrical Journal 12/98 (12.2%) und im gesamten Band von 2000 19/76 (25.0%). Dieser Unterschied ist ebenfalls statistisch signifikant (Exakter Fisher Test: p=0.045, OR=2.4, 95%, CI 1.1-5.3), jedoch deutlich geringer. Es fiel dabei auch auf, dass Autoren der DR zu einem weitaus größeren Anteil im Biometrical Journal als in den Biometrics publizierten. Obwohl Biometrics lange Zeit das einzige offizielle Organ der DR war, tauchen Beiträge aus der DR dort nur vereinzelt auf. Biometrische Beispiele sind in CS-Arbeiten selten. Im CSDA findet man im Band von 1986 4/20 Arbeiten mit einer biometrischen Auswertung und im Band von 2000 11/45 (24.4%).

Eine andere Möglichkeit, die gegenseitige Durchdringung von Biometrie und CS zugeschnitten auf Deutschland zu untersuchen, bieten die Vortragsthemen des alljährlichen Biometrischen Kolloquiums. Dazu wurden für die Jahre 1969 bis 2001 alle Programmhefte und teilweise auch die Abstractbände nach Beiträgen zur CS und nach Beiträgen mit einem wesentlichen CS Anteil gesichtet (Tabelle 3). Offensichtlich gibt es in der DR eine lange Tradition der CS, vgl. das Thema „EDV und Biometrie" unter Initiative von H. Geidel aus dem Jahr 1970. Vielfältige Aktivitäten der CS erkennt man für die 1980er Jahre. Ab 1990 gibt es einen Rückgang von CS-Themen verbunden mit einem Zuwachs von biometrischen Fragestellungen in verschiedenen Anwendungsbereichen. Man kann aber auch feststellen, dass ab 1995 die CS im Programm des Biometrischen Kolloquiums keine nennenswerte Rolle mehr spielt, zumindest fehlen eigene wissenschaftliche Sitzungen der AG CS, was oben im Abschnitt 5.3 schon erwähnt wurde.

8. Diskussion und Schlussfolgerungen

Die Aktivitäten und die Ergebnisse der CS haben biometrische Modellbildung und biometrische Auswertungen entscheidend unterstützt. Beispiele dafür sind die Generalisierten Modelle und die ‚Mixed-Effects' Modelle. Umgekehrt haben die Erfordernisse aus den Anwendungen der Biometrie zumindest in Deutschland die Arbeiten der CS auf wichtige Fragestellungen fokussiert und somit zu einer Anwendungsorientierung der CS beigetragen. Beispiele sind die Implementierung von iterativen Verfahren und die Residuenanalysen in Programmpaketen, aber auch die Entwicklung von einheitlich strukturierten Verfahren. Dies ist ein Ansatz, den Statistiker und Biometriker von Anfang an verfolgt haben. Man erinnere sich an das DMV107 Projekt oder die Entwicklung und Intention von DIST.

Verfolgt man die Entwicklung der CS in Deutschland, so kann man feststellen, dass

1. statistisches Rechnen unter Verwendung des Computers schon sehr früh in der DR ein Thema war und auf dem Biometrischen Kolloquium behandelt wurde. Im Rückblick wird deutlich, dass eine Reihe namhafter deutscher Biometriker aus den Erfordernissen der biometrischen Datenanalyse heraus die Notwendigkeit der Entwicklung der CS erkannten und vorantrieben u.a. die Gründung der AG CS.

2. in der Frühphase der Anwendungen der CS in der Biometrie in Deutschland sowohl im Westen als auch im Osten umfassende und wohldurchdachte Konzepte für eine statistische Datenanalyse diskutiert und zum Teil umgesetzt wurden. Im Umbruch vom Großrechner zu PCs und Workstations stockte diese Arbeit und vieles ist bestenfalls unerfüllte Anforderung geblieben. Die Gemeinsamkeiten der Entwicklung im Westen und im Osten scheinen weitgehend aus dem heutigen Bewusstsein verschwunden zu sein. In der DR wurde schon sehr früh die Notwendigkeit der Bearbeitung bestimmter CS Themen erkannt und angestoßen, aber dann nicht immer mit der für eine nachhaltige Wirkung (z. B. in Form von Publikationen) notwendigen Planung, Unterstützung und Energie weitergeführt.

3. die Entwicklung der CS in der Biometrie durch die rechnerischen Anforderungen der statistischen Modellbildung und durch die Anforderungen an die Genauigkeit und Gültigkeit statistischer Aussagen vorangetrieben wurden. Bei neueren rechenintensiven Verfahren wurde zum Teil sehr spät reagiert, so dass auf die Entwicklung der Verfahren und ihrer Umsetzung in Software nur bedingt Einfluss genommen werden konnte. Als Beispiel sind hier die Bootstrap Verfahren zu nennen.

4. die CS durch ihre Vermittlung von Rechenverfahren und Software auf die Entwicklung der Biometrie starken Einfluss genommen hat, vor allem dadurch, dass größere und komplexere Datensätze schneller und gründlicher auswertbar wurden. Der Anteil der CS in Arbeiten der Biometrie ist substanziell und betrifft deutlich mehr als 50% aller Arbeiten. Oft wird dies aber nicht mehr als Folge der Leistungen der CS wahrgenommen.

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gbp(M2005061414-1)

5. Biometrie als notwendiger Partner der Substanzwissenschaften für die Lösung biologischer, biomedizinischer und agrarwissenschaftlicher Probleme auch deswegen anerkannt wird, weil sie mittels Verfahren der CS wesentlich zur Lösung der substanzwissenschaftlichen Probleme beiträgt.

6. es in Deutschland vor allem der Kontinuität der Reisensburgtagungen zu verdanken ist, dass die CS die Veränderungen im Hardware- und Softwarebereich meistern und ihren Einfluss auf die Biometrie erhalten konnte. Im Gegensatz zur Kontinuität der Reisensburgtagungen, die wesentlich auch von der GMDS AG Statistische Auswertesysteme mit getragen werden, zeigen die Aktivitäten der AG ‚Computational Statistics' keine Kontinuität in ihrer Präsenz beim Biometrischen Kolloquium.

7. CS für die Biometrie in Deutschland weiterhin eine wichtige Funktion hat bzw. haben sollte, da sich die Biometrie in Deutschland stärker als anderswo auf die Anwendungen und die Kooperation in biologischen und biomedizinischen Projekten konzentriert. Diese Anwendungsorientierung der Biometrie in der DR sollte zu einer adäquaten Publikationsstrategie der DR und der GMDS in Deutschland führen, bei der beide, Biometrie und CS, nachhaltig gefördert werden.

Danksagung und Anmerkung

Bedanken möchte ich mich bei Rüdiger Ostermann für den Anstoß zu dieser Arbeit, bei Jürgen Läuter für sehr wertvolle Hinweise zur Entwicklung in der DDR und die zeitweise Überlassung von Berichten der Arbeiten einer AG der Hochschulrechenzentren der DDR, ebenso bei Egmar Rödel, Volker Guiard, Rainer Koch und Christel Richter, und schließlich bei Norbert Victor, der mir durch seine vorbildliche Ablage den Zugang zu Material bis zurück in das Jahr 1969 ermöglichte. Den Professoren Läuter und Victor bin ich weiter für eine kritische Durchsicht und Verbesserungsvorschläge zum Manuskript verbunden.

Dieser Beitrag gründet auf einem Vortragsmanuskript anlässlich des 50. Geburtstags der Deutschen Region (DR) der Internationalen Biometrischen Gesellschaft, gehalten am 19. März 2003 in Wuppertal beim Biometrischen Kolloquium in der Sitzung 'Historische Entwicklung der Biometrie in Deutschland organisiert von Rüdiger Ostermann.

Der Beitrag ist gewidmet dem Andenken an meine Mutter, die am 19. März 2003 an den Folgen einer Diabetes-Erkrankung verstarb.

Nachtrag

Nach Fertigstellung des Manuskripts erschienen zwei Werke, die in diesem Zusammenhang erwähnt werden müssen.

Auf dem COMPSTAT-Symposium im August 2004 haben Wilfried Grossmann, Michael Schimek und Peter Paul Sint die Geschichte der COMPSTAT Symposien mit den wesentlichen Schritten der Entwicklung des Statistischen Rechnens (Statistical Computing) der vergangenen 30 Jahre in einer umfangreichen Arbeit dargestellt und kommentiert [12].

Ein absolutes ‚Highlight' für die CS ist das Erscheinen des Handbuchs der Autorengruppe James Gentle, Wolfgang Härdle und Yuichi Mori ebenfalls in 2004 [11]. In 35 Kapiteln stellen auf 1053 Seiten 49 namhafte Autoren das gesamte Fachgebiet vor und demonstrieren damit eindrucksvoll seine Aktualität.

Schließlich ist anzumerken, dass die im Abschnitt 8 geforderte verbesserte Publikationsstrategie der DR durch die herausgeberische Kooperation der DR mit Wiley-VCH für das Biometrical Journal auf einem guten Weg ist.

Korrespondenzadresse:

• Dr. Lutz Edler, Abteilung Biostatistik - C060, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, Tel.: 06221-42 2392, Fax: 06221-42 2397
edler@dkfz.de

Literatur:

[1] Armitage P, David HA. Advances in Biometry. 50 Years of the International Biometric Society. New York: Wiley; 1996.

[2] Becker RA, Chambers JM. S: An Interactive Environment for Data Analysis and Graphics. Belmont, Calif: Wadsworth; 1984.

[3] Chambers J. Computing with data: concepts and challenges. The American Statistician. 1999;53:73-84.

[4] Chambers JM, Gale WA, Pregibon D. On the existence of expert systems. Statistical Software Newsletter. 1988;14:63-6

[5] Eddy WF, Gentle JE. Statistical Computing: What's past is prologue. In: Atkinson AC, Fienberg SE. A Celebration of Statistics. The ISI Centenary Volume. New York: Springer; 1985.

[6] Edler L. Statistical computing with the IASC at the end of 2000: Continuation and consolidation - innovation and invention. CSDA. 2001;33:113-9.

[7] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. London: Chapman and Hall; 1993.

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gbp(M2005014.1)

[8] Enderlein G, Grimm H, Geidel H, Lorenz RJ. Die Entwicklung der Biometrie in der DDR. In: Geidel H, Lorenz RJ. 40 Jahre Biometrie in Deutschland. Biometrische Berichte; Band 1. Münster-Hiltrup: Landwirtschaftsverlag; 1993. p. 27-36.

[9] Fisher RA. Statistical Methods for Research Workers. Edinburgh: Oliver & Boyd; 1925

[10] Francis I. Statistical Software. A Comparative Review. Amsterdam: North Holland; 1981

[11] Gentle JE, Härdle W, Mori Y. Handbook of Computational Statistics. Concepts and Methods. Berlin: Springer; 2004.

[12] Grossmann W, Schimek MG, Sinn PP. The history of COMPSTAT and key-steps of statistical computing during the last 30 years. In: Antoch J et al., eds. COMPSTAT 2004 Symposium. Heidelberg: Springer; 2004. p. 1-35.

[13] Hand DJ. Expert systems in statistics. The Knowledge Engineering Review. 1986;1:2-10.

[14] Hastie TJ, Tibshirani RJ. Generalized Additive Models. London: Chapman and Hall; 1990.

[15] Hörmann A. Twenty-five working conferences on statistical computing - Reflections on twenty years of Reisensburg meetings. In: Dirschedl P, Ostermann R. Computational Statistics. Heidelberg: Physica Verlag; 1994. p. 17-36.

[16] Lauro NC. Computational Statistics or statistical computing, is that the question? CSDA. 1996;23:191-3.

[17] Läuter J. Entwicklung mathematisch-statistischer Algorithmen und ihre Realisierung auf Rechenautomaten [Dissertation]. Berlin: Akademie der Wissenschaften; 1973.

[18] Läuter J. Programmiersprache DIST. Dateneingabe und Datenstrukturierung. Berlin: Adaemie Verlag; 1981.

[19] McCullagh P, Nelder JA. Generalized Linear Models. 2. ed. London: Chapman and Hall; 1989.

[20] Milton RC, Nelder JA. Statistical Computation. New York: Academic Press; 1969.

[21] Nelder JA, Wedderburn RHW. Generalized Linear Models. J Roy Statist Soc A. 1972;135: 370-84.

[22] Nelder JA. How should the statistical expert system and its user see each other? Compstat. 1988:107-16.

[23] Nelder JA. Statistical Computing. In: Armitage P, Hand A. Advances in Biometry. New York: Wiley; 1996. p. 201-12.

[24] Ostermann R, Dirschedl P. Computational Statistics. Heidelberg: Physica Verlag; 1994.

[25] Schucany WW, Minton PD, Shannon BS. A survey of statistical packages. Computing Surveys. 1972;4:65-79.

[26] Sint PP. Remarks on the history of computational statistics. In: Dirschedl P, Ostermann R. Computational Statistics. Heidelberg: Physica Verlag; 1995. p. 17-36.

[27] Streitberg B. On the nonexistence of expert systems - critical remarks on artificial intelligence in statistics. Statistical Software Newsletter. 1988;14:55-62.

[28] Vahle H. Mathematisch-statistische Programmsammlung Statistik. In: Autorenkolektiv Mathematische Statistik. Weiterentwicklung von Programmen für ESER Teil 1. Strukturen und Steuerprobleme. Schriftenreihe Informationsverarbeitung im Hoch- und Fachschulwesen. Berlin: Ministerium für Hoch- und Fachschulwesen; 1981. p. 11-31.

[29] Victor N. Computational Statistics - Science or Tool? (with discussion). Stat Soft Newsl. 1984;10:105-25.

[30] Victor N. The roots of computational statistics in Germany. In: Dirschedl P, Ostermann R. Computational Statistics. Heidelberg: Physica Verlag; 1994. p. 17-36.

Artikel online frei zugänglich unter http://www.egms.de/en/journals/mibe/2005-1/mibe000010.shtml
Bitte zitieren als: Edler L. Computational Statistics und Biometrie: wer treibt wen?. GMS Med Inform Biom Epidemiol. 2005;1(2):Doc10.
gipi(M2005061414.1)