



# **K** Briefe zur **LASSIFIKATION**

## **Mitgliederbrief der Gesellschaft für Klassifikation e.V.**

**Januar 2016**

### **Inhalt:**

Grußwort des Präsidenten der GfKI.....	2
Vorstands-/Beiratssitzung am 14. März 2016 .....	3
Mitgliederversammlung am 15. März 2016.....	3
Protokoll der 36. Mitgliederversammlung.....	3
Adressänderungen, Umzüge, Bankwechsel .....	3
75. Geburtstag von Prof. Dr. Dr. h.c. Hans-Hermann Bock.....	4
70. Geburtstag von Prof. Dr. Wolfgang Gaul .....	7
3 <sup>rd</sup> European Conference on Data Analysis (Colchester, 2015) .....	9
Bericht AG BIB und DK .....	10
LIS 2016.....	11
Bericht AG BT .....	11
Bericht AG DANK.....	12
ADAC – Advances in Data Analysis and Classification.....	14
Archives of Data Science (Series A).....	15
BRCA Data Analysis Award.....	16

## Grußwort des Präsidenten der GfKI

Liebe Mitglieder und Interessierte,

erst einmal möchte ich Ihnen/Euch ein gutes und erfolgreiches neues Jahr 2016 wünschen und mein Grußwort mit einem Rückblick auf das vergangene Jahr und einen Ausblick auf das neue Jahr verbinden.

Ich möchte Ihnen allen danken für Ihre Beiträge und Ihr Mitwirken in unserer Gesellschaft. Auch als lokaler Organisator möchte ich unsere **Jahrestagung 2015, European Conference on Data Analysis 2015**, 2.-4. September an der University of Essex in Colchester hervorheben, die als gemeinsame Tagung mit der Britischen und Polnischen Klassifikationsgesellschaft durchgeführt wurde. Das Programm der Tagung war erfolgreich, vielschichtig und insbesondere mit vielen Beiträgen zu **Data Science** hochaktuell und für unsere Gesellschaft für Klassifikation e.V. (GfKI) richtungsweisend. Auf der Mitgliederversammlung haben wir die in Bremen 2014 begonnene **Diskussion des Namens der GfKI** weitergeführt und beschlossen, dass wir zukünftig in der Außendarstellung der Gesellschaft den Namen **Gesellschaft für Klassifikation Data Science Society** verwenden. Die Mitgliederversammlung verlieh unserem langjährigen Mitglied Herrn **Professor Dr. Gunter Ritter** einstimmig die **Ehrenmitgliedschaft**.

Weitere wesentliche Beiträge im vergangenen Jahr waren die Tagungen und Workshops durchgeführt von unseren Arbeitsgruppen: Die Herbsttagung der **AG-DANK** wurde im November auf Einladung von Andreas Geyer-Schulz (KIT) in Karlsruhe durchgeführt. **AG Biostatistik** veranstaltete im Juli die jährliche Tagung 'Statistical Computing' auf Schloss Reisenburg (Günzburg). Die **AG-BIB** und **AG-DK** organisierten wie jedes Jahr eine besondere Veranstaltung im Rahmen der Jahrestagung.

Im Laufe des Jahres feierten unsere Ehrenmitglieder **Herr Professor Bock (75.)** und **Herr Professor Gaul (70.)** runde Geburtstage. Unser Ehrenvorsitzender Professor Bock wurde bei IFCS Konferenz in Bologna (Juli 2015) mit der **IFCS Research Medal for outstanding research achievements** ausgezeichnet.

Die **kommende Jahrestagung** findet im Rahmen der **DAGStat2016** vom **14.-18. März 2016 in Göttingen** statt (Tagungsleiter: Prof. Dr. Thomas Kneib). Wir hoffen, dass viele von Ihnen an unserer **40. Jahrestagung** teilnehmen. Anlässlich unserer 40. Jahrestagung laden wir die Tagungsteilnehmer der DAGStat2016 zu einem Empfang am Mittwoch 16. März 2016 um 18 Uhr ein.

Vom 10.-13. Mai 2016 findet die **10-th International Scientific Conference on Modelling and Forecasting of Socio-Economic Phenomena** in Zakopane statt. Vom 12.-13. September 2016 planen wir den nächsten **Deutsch-Japanischen Workshop** auf Schloss Reisenburg (Günzburg). Vom 8.-10. August 2017 findet die **IFCS-2017** in Tokio statt. In 2017 ist unsere Jahrestagung wieder im Rahmen ECDA Konferenz geplant, die als **ECDA2017** von der Polnischen Klassifikationsgesellschaft in der Woche 25.-28. September 2017 in Wrocław (Breslau) ausgerichtet wird.

Mit herzlichen Grüßen

Ihr/Euer

Berthold Lausen

Vorsitzender Gesellschaft für Klassifikation e.V. (GfKI)

Autor:

B. Lausen

### Vorstands-/Beiratssitzung am 14. März 2016

Die diesjährige Vorstands- und Beiratssitzung findet am Montag, 14.03.2016, 17:30–20:00 Uhr, am Tagungsort der 40. Jahrestagung im Rahmen der DAGStat2016 in Göttingen statt.

Die offizielle Einladung und die Tagesordnung werden rechtzeitig per E-Mail versandt.

Autor: B. Lausen

### Mitgliederversammlung am 15. März 2016

Die diesjährige Mitgliederversammlung findet am Dienstag, 15.03.2016, 16:40–18:00 Uhr, am Tagungsort der 40. Jahrestagung im Rahmen der DAGStat2016 in Göttingen statt.

Die offizielle Einladung und die Tagesordnung werden rechtzeitig per E-Mail versandt.

Autor: B. Lausen

### Protokoll der 36. Mitgliederversammlung

Das Protokoll der 36. Mitgliederversammlung wird separat verschickt.

Autor: B. Lausen

### Adressänderungen, Umzüge, Bankwechsel

Da es für die Verwaltung des Vereins sehr wichtig ist, dass Ihre Daten (E-Mail-Adresse, Telefonnummer, Postanschrift, ggf. Kontoverbindung) immer auf dem aktuellen Stand sind, bitten wir Sie Änderungen an

- [Schatzmeister@gfkl.de](mailto:Schatzmeister@gfkl.de) bzw. an
- Prof. Dr. Andreas Geyer-Schulz ([Andreas.Geyer-Schulz@kit.edu](mailto:Andreas.Geyer-Schulz@kit.edu)) und an
- Frau Victoria-Anne Schweigert ([v.schweigert@kit.edu](mailto:v.schweigert@kit.edu))

zu senden.

Wir behandeln Ihre Daten selbstverständlich vertraulich und nutzen diese nur für die typische Vereinsangelegenheiten der GfKI. Insbesondere benötigen wir Ihre Daten zur Zusendung der Advances of Data Analysis and Classification (ADAC), des Newsletters und um zusätzliche Bankkosten beim Einzug des Mitgliedsbeitrags zu vermeiden.

Leider entstehen der GfKI jedes Jahr Kosten und zusätzliche Arbeit durch veraltete Mitgliederinformationen. Durch die Mitteilung von Änderungen könnten wir gemeinsam diese Kosten vermeiden.

Wir danken Ihnen für Ihre Mithilfe und Ihr Verständnis!

Andreas Geyer-Schulz (Dezember 2015, KIT, Karlsruhe)

Autor: A. Geyer-Schulz

## 75. Geburtstag von Prof. Dr. Dr. h.c. Hans-Hermann Bock



Im September dieses Jahres feierte unser Ehrenpräsident und Beiratsmitglied H.-H. Bock seinen 75ten Geburtstag. Dazu nachträglich herzlichen Glückwunsch und alles Gute.

Herr Bock studierte in Karlsruhe, Paris und Freiburg und promovierte 1968 während seiner Assistententätigkeit bei Prof. Morgenstern in Freiburg mit einer Arbeit zur Clusteranalyse. Aber darüber später. Im Jahre 1978 wurde er auf eine Professur für angewandte Statistik am Institut für Statistik und Wirtschaftsmathematik der RWTH Aachen berufen, wo er bis zu seinem Ruhestand 2005 lehrte. Herrn Bock ist eine Anzahl von Ehrungen zuteil geworden. Er ist Ehrendoktor der Universität

Krakau, Ehrenmitglied der Belarussischen Statistischen Vereinigung, Ehrenpräsident der GfKI und erster Träger der Forschungsmedaille der IFCS.

Um Herrn Bock zu würdigen, muss man auf seine Forschung, seine internationale Vortragstätigkeit, sein organisatorisches Talent und seine internationalen Beziehungen eingehen. Er war ein aktiver Gründer und Gründungsmitglied von wissenschaftlichen Gesellschaften und Medien, die sich als nachhaltig erwiesen haben. Er ist Gründungsmitglied der "International Federation of Classification Societies" (IFCS) und war von 1985 bis 1989 ihr erster Präsident. Im Jahre 1977 gründete er die Gesellschaft für Klassifikation (GfKI) und diente ihr von 1986 bis 1995 als Präsident. Innerhalb der GfKI gründete er die Arbeitsgruppe "Datenanalyse und Numerische Klassifikation" (AG-DANK), die er von 1979 an leitete, bis ich ihn 2001 ablöste. Es ist nicht übertrieben zu sagen, dass es GfKI und AG-DANK ohne Herrn Bock nicht gäbe. Mit ihm zusammen könnte man sicher klären, ob sich diese Aussage auch auf die IFCS übertragen lässt. Er war Mitgründer und ist noch Herausgeber der wissenschaftlichen Reihe "Studies in Classification, Data Analysis, and Knowledge Organization" und der inzwischen hochangesehenen Springerzeitschrift "Advances in Data Analysis and Classification".

Herr Bock war und ist ein prominenter Teilnehmer, Redner und Sektionsleiter auf Konferenzen. Seine Zusammenfassungen und Fragen nach Vorträgen anderer Konferenzteilnehmer sind erhellend. Besonders auf den zweijährlichen Konferenzen der IFCS pflegt er seine internationalen Beziehungen mit Kollegen aus Europa und der ganzen Welt. Immer wieder war er in Forschungssemestern und zwischen den Semestern ein gern gesehener Gast an verschiedenen internationalen Universitäten. Jeder profitiert von seinen umfassenden Kenntnissen in der Statistik und der Datenanalyse.

Herr Bock hat fundamentale Beiträge zur Clusteranalyse und anderen Bereichen der Statistik geliefert. Es begann in der zweiten Hälfte der 60er Jahre, als er sich für das Gebiet der Clusteranalyse zu interessieren begann und seine Doktorarbeit "Statistische Modelle für die einfache und doppelte Klassifikation von normalverteilten Beobachtungen" schrieb. Dieses Thema war ein Glücksfall und, wie man heute feststellt, damals sehr weitsichtig. Das erkennt man jetzt ein halbes Jahrhundert später an der heutigen Bedeutung der Clusteranalyse, ihren Anwendungen auf die meisten Wissensgebiete, an ihren Ramifikationen "Data Mining" und "Unsupervised Classification" sowie an ihrer wirtschaftlichen Bedeutung. Der damalige Kenntnisstand beschränkte sich auf geometrische und algorithmische Sachverhalte: Einige hierarchische Verfahren, das bekannte, von Ward heuristisch und geometrisch begründete und nach ihm benannte Quadratsummenkriterium, der üblicherweise MacQueen zugeschriebene aber schon viele Jahre zuvor von anderen gefundene k-means-Algorithmus und ein Jahr vor seiner Arbeit das Kriterium von Friedman und Rubin, ebenfalls heuristisch begründet.

Fortsetzung nächste Seite

Was damals mit Ausnahme einer singulären Arbeit von Karl Pearson aus dem Jahre 1894 fehlte, war eine probabilistische Theorie des Clusters, wie es damals in anderen Gebieten der Statistik schon klassisch war. Herr Bock packt den Stier bei den Hörnern und untersucht, was herauskommt, wenn man ein probabilistisches Modell der Daten annimmt und die statistische Entscheidungstheorie darauf anwendet. Bei Annahme von sphärischen Clustern diskutiert er systematisch eine ganze Reihe von Verlustfunktionen und a-priori-Maßen. Interessanterweise bekommt er in einem Fall das Wardsche Kriterium zurück. Damit war die Katze aus dem Sack – dieses Kriterium hat einen probabilistischen Hintergrund. Es reiht sich nahtlos in die Statistik ein, wie sie von ihren Vätern Ende des 19ten und anfangs des 20sten Jahrhunderts konzipiert worden war. Das war auch die Begründung dessen, was man heute auf einen Vorschlag von McLachlan und Basford etwas ungenau "model-based clustering" nennt. Ein Meilenstein, denn nun konnte man durch Anwendung allgemeinerer Datenmodelle allgemeinere Kriterien einfach berechnen, man musste sie nicht mehr mit heuristischen Ideen ermitteln. Das geschah schon drei Jahre später 1971 durch A.J. Scott und M.J. Symons. Unter der Annahme allgemeiner Normalverteilungen erhielten sie ihr Determinantenkriterium. Scott und Symons hätten Herrn Bocks Arbeit eigentlich zitieren müssen. Dass dies nicht geschah, hat zwei Gründe, die man Scott und Symons aber nicht anlasten kann. Zum einen ist Herrn Bocks Arbeit auf Deutsch abgefasst. Das ist im deutschsprachigen Raum natürlich und auch heute noch in der deutschen Statistik gängig. Sogar in manchen internationalen Journalen war Deutsch eine akzeptierte Sprache. Für ein internationales Publikum war die Arbeit aber nicht leicht zugänglich. Obwohl die Wissenschaft schon viel früher global war als die Wirtschaft, fehlte doch damals noch eine gemeinsame Sprache. Wer heute weltweit gelesen werden will, muss Englisch schreiben.

Für den anderen Grund muss ich Ihnen eine kleine Geschichte erzählen. Ich war vom Vorstand der GfKI eingeladen worden, auf der vergangenen AG-DANK-Tagung bei Herrn Geyer-Schulz in Karlsruhe anlässlich Herrn Bocks Geburtstag einen Vortrag zu halten. Dort wollte ich eigentlich die Titelseite seiner Doktorarbeit zeigen. Natürlich hätte ich sie mir von Herrn Bock besorgen können. Das konnte ich aber nicht, denn ich hatte von Herrn Lausen die Auflage, vorher nichts zu sagen. Ich bat also vier Wochen vor dem Vortrag einen Bibliothekar unserer Bibliothek, mir diese Arbeit zu besorgen. Interessanterweise konnte er sie in den einschlägigen Datenbanken nicht finden. Er schickte aber trotzdem ein Bestellformular weiter. Der Vortrag fand ohne eine Seite aus der Arbeit statt. Einige Tage später kam die Doktorarbeit tatsächlich, allerdings durfte ich sie nicht mitnehmen sondern nur im Lesesaal reinschauen. Auch durfte ich nichts kopieren. Die Kopie hätte ich also sowieso nicht auf diesem Wege bekommen. Des Rätsels Lösung lernte ich kurz danach von Herrn Bock: Er musste damals 1968 nur drei Exemplare der Arbeit abliefern. Es herrschte also in Freiburg damals für Doktorarbeiten eine Art "Geheimhaltung". Herr Bock durchbrach sie allerdings bald von sich aus, denn 1972 erschien in Metrika ein größerer Auszug aus der Arbeit. Außerdem erschien seine Theorie nochmals 1974 in seinem Buch "Automatische Klassifikation" in der Reihe "Studia Mathematica" bei Vandenhoeck & Ruprecht.

Herr Bock hat sich in den 40 Jahren nach dem Buch mit vielen statistischen Fragestellungen befasst und Lösungen veröffentlicht. Eine Arbeit, die mich besonders beeindruckt, erschien 1985 in Journal of Classification mit dem Titel "On Significance Tests in Cluster Analysis". Die Arbeit ist ein Beitrag zu den schwierigen Problemen der Clusteranzahl und der Clustervalidierung. Er schlägt mehrere Signifikanztests zur Schätzung dieser Zahl und der Qualität der Lösung vor. Man erwartet ein kleineres Clusterkriterium, wenn die Daten tatsächlich die gefundene Clusterzahl aufweisen. Deswegen verwendet er Clusterkriterien als Teststatistiken und bestimmt ihre Verteilungen unter der Nullhypothese. Dazu benutzt er ein tiefes Ergebnis von Pollard aus dem Jahr 1982.

Fortsetzung nächste Seite

Herr Bock hat viele Übersichtsarbeiten geschrieben und sich mit vielen anderen Themen befasst: Mit symbolischen Daten (zusammen mit F.A.T. de Carvalho und Paula Brito), Zweiwegclusterung (mit Iven Van Mechelen und Paul De Boeck), mit Neuronalen Netzen, Zeitreihen, Multivariater Skalierung, Fuzzy Clustering, Explorativer Datenanalyse und Unähnlichkeitsmatrizen. Viel davon ist in der "Studies"-Reihe erschienen. Ich wünsche ihm noch viele interessante Erkenntnisse.

Gunter Ritter

Autor:

G. Ritter

## 70. Geburtstag von Prof. Dr. Wolfgang Gaul



Am 3. Juli 2015 vollendete Prof. Dr. Wolfgang Gaul, langjähriger Vorstand, Präsident und seit einiger Zeit auch Ehrenmitglied der Gesellschaft für Klassifikation e.V., sein 70. Lebensjahr. Die Gesellschaft für Klassifikation gratuliert auch auf diesem Wege ganz herzlich!

Prof. Gaul feierte den runden Geburtstag im Kreise seiner Familie, Freunde, Kollegen und Schüler in einem ausgesuchten Ambiente, verbunden mit einer exzellenten Bewirtung in zeitlichem Anschluss an die diesjährige IFCS-Konferenz in Bologna. Bei der Feier in Karlsruhe blickten zahlreiche Redner auf sein interessantes Leben und sein bisheriges, ebenso umfangreiches wie hochwertiges wissenschaftliches Lebenswerk zurück. Es wurden viele denkwürdige, gemeinsame Erlebnisse rekapituliert, die neben den rein zwischenmenschlichen Dingen gerade aber auch deutlich werden ließen, wie Prof. Gaul einer ganzen Generation von Kollegen und akademischen Schülern und Enkeln vorgelebt hat, was gute Forschung und Lehre von schlechter unterscheidet. Sein ausgeprägtes Bewusstsein für eine qualitativ hochwertige Forschung konnte jeder schon einmal auf der einen oder anderen internationalen Konferenz erleben, sei es in Form eines ebenso innovativen wie gehaltvollen Vortrags oder in Form engagierter, häufig kritischer, immer aber konstruktiver Rede- und Diskussionsbeiträge. Ganz in dieser Tradition verhaftet, ließ es sich der Jubilar auch nicht nehmen, die Erinnerungen der Redner in humorvoller Weise zu kommentieren.

Prof. Gaul wurde 1945 in Stolp in Pommern geboren, hat an der Universität Bonn Mathematik studiert, wurde 1974 promoviert und hat dort im Jahre 1980 auch habilitiert. Nach seinem Ruf an die Universität Karlsruhe (TH) im gleichen Jahr übernahm er dort, gemeinsam mit Prof. Göppl, die Leitung des Instituts für Entscheidungstheorie und Unternehmensforschung. Diese Funktion hatte er bis zum Ausscheiden aus dem Hochschuldienst inne.

Seine Forschungs- und Lehrschwerpunkte waren stets geprägt von einer tiefen Theorieorientierung – mit besonderen Schwerpunkten in Graphentheorie, Netzwerkmodellierung, Zuverlässigkeitstheorie, stochastischer Optimierung, Wahrscheinlichkeitstheorie und Statistik sowie in der Stichprobenplanung und – vor allem natürlich – in der Datenanalyse. Gleichzeitig war ihm aber auch immer die Anwendungsorientierung seiner Forschung und Lehre ein zentrales Anliegen (z.B. in den Feldern Kaufverhaltensanalyse, Marktforschung, Computergestütztes Marketing, Wissensorientiertes Marketing, Operations Research, Recommender-Systeme oder Data und Web Mining).

Prof. Gaul kann auf eine umfangreiche Publikationsliste mit mehr als 150 hochkarätigen Veröffentlichungen verweisen, die in den unterschiedlichsten, renommierten Zeitschriften erschienen sind, so z.B. in *Advances in Data Analysis and Classification*, *Decision Support Systems*, *International Journal of Research in Marketing*, *Journal of Classification*, *Journal of Econometrics*, *Journal of Marketing Research*, *Studies in Classification, Data Analysis, and Knowledge Organization*, *Zeitschrift für Betriebswirtschaft* und *Zeitschrift für betriebswirtschaftliche Forschung*.

Fortsetzung nächste Seite

Neben seinen grundlegenden Neu- und Weiterentwicklungen methodischer Ansätze, etwa in der Clusteranalyse, der Conjointanalyse oder im Data und Web Mining, ist Prof. Gaul insbesondere aber auch für seine richtungsweisende Forschung zum Thema Computergestütztes Marketing bekannt geworden. Bereits Ende der 1980er Jahre sah er voraus, wie wichtig die Entwicklungen in der Kommunikations- und Informationstechnik für das Marketing in der Zukunft sein werden, und er hat mit grundlegenden Artikeln und Büchern zu diesem Thema (z.B. dem gemeinsam mit Martin Both veröffentlichten Springer-Buch "Computergestütztes Marketing") der modernen Marketingforschung den Zugang zur Kommunikations- und Informationstechnik geebnet.

Große DFG-Projekte wie das WIMDAS-Projekt, das er mit dem Kollegen Martin Schader initiierte und leitete, zeigten, wie mit Expertensystemen im Kontext von Datenanalyse und Marktforschung die Zukunft des datengestützten resp. quantitativen Marketings gestaltet werden kann. Heute – in einem Zeitalter der schieren Allgegenwärtigkeit von Digitalisierung – wird klar, wie richtungsweisend seine Forschungen in den Achtziger- und Neunzigerjahren des letzten Jahrhunderts zu diesen Themen waren. Hiervon profitieren insbesondere auch heute noch seine akademischen Schüler und Enkel, zu denen auch die beiden Verfasser dieses kleinen Berichtes zählen.

Sein interdisziplinärer Forschungsansatz an der Schnittstelle zwischen Mathematik/Statistik, Informatik und Technik sowie den Sozial- und Wirtschaftswissenschaften war auch für die Gesellschaft für Klassifikation maßgeblich: Prof. Gaul wirkte für die Gesellschaft langjährig erfolgreich als Präsident, Vorstand und Beirat. Er hat zahlreiche hochkarätig besetzte internationale Konferenzen organisiert und als Vorsitzender und Mitglied von Programmkomitees auch eine beträchtliche Zahl von Tagungsbänden in der von ihm mitinitiierten, sehr erfolgreichen Springer-Reihe "Studies in Classification, Data Analysis, and Knowledge Organization" herausgegeben. Daneben war und ist er Mitherausgeber der mit der Gesellschaft eng verbundenen Springer-Zeitschrift *Advances in Data Analysis and Classification* und eine der treibenden Kräfte hinter dem intensiven und erfolgreichen Austausch der GfKI mit anderen Klassifikationsgesellschaften, etwa in Japan, Italien oder Polen, um nur einige jener zu nennen, zu denen er intensive Verbindungen aufgebaut und gepflegt hat.

Die Gesellschaft für Klassifikation verdankt Prof. Gaul sehr viel. Sie freut sich auf eine auch in Zukunft gute und erfolgreiche Zusammenarbeit und wünscht ihm alles Gute für den weiteren Lebensweg!

Daniel Baier und Reinhold Decker

Autor:

D. Baier / R. Decker



### 3<sup>rd</sup> European Conference on Data Analysis (Colchester, 2015)

The *European Conference on Data Analysis 2015 (ECDA2015)*, co-hosted by the *British Classification Society (BCS)*, *Gesellschaft für Klassifikation e.V. (GfKI)* and *Sekcja Klasyfikacji i Analizy Danych (SKAD)* and held at the University of Essex, Colchester, UK from 1st to 4th September 2015 (conference webpage [www.ecda2015.eu](http://www.ecda2015.eu)). After the successful ECDA conferences in Luxembourg (2013) and Bremen (2014) participants came from 19 different countries: Austria, Belgium, Canada, Cyprus, Germany, Finland, Greece, Italy, Luxembourg, Morocco, Mexico, New Zealand, Poland, Portugal, South Africa, Tunisia, United Kingdom, United States of America and Switzerland. The conference included the *39th Annual Meeting of the German Classification Society* and the *Librarians and Information Science (LIS) Workshop*. Keynotes of the conference were

- Hendrik Blockeel, Leuven: *Declarative data analysis*
- Ulf Brefeld, Darmstadt: *Capturing user behaviour*
- Andrzej Dudek, Wroclaw: *Cluster analysis in the XXI century, new algorithms and tendencies*
- Arthur Gretton, London: *Kernel nonparametric tests of homogeneity, independence and multi-variable interaction*
- Janette McQuillan, London: *Crowdsourcing classifications to accelerate cancer research*
- Christine Müller, Dortmund: *Data depth*
- Iris Pigeot, Bremen: *Challenges in the statistical analysis of longitudinal data*
- Claus Weihs, Dortmund: *Efficient global optimization: motivation, variation, and application*
- Stefan Wrobel, Bonn: *Data analytics in a networked world*

The conference had invited and contributed sessions under the headings *Data Analysis, Data Science, Clustering, Classification, Machine Learning and Knowledge Discovery, Library and Information Science, Finance and Economics, Digital Humanities and Social Sciences, Geosciences and Archaeology, Marketing, Musicology, Machine Learning and Knowledge Discovery, Outliers in Classification, Life Sciences, Mathematical Foundations of Data Science, Education, Big Data, Engineering, Logistics and Optimisation*.

The GfKI Best Paper Award 2014 was awarded for the paper at ECDA2014 *Incremental Generalized Canonical Correlation Analysis* by Angelos Markos (Democritus University of Thrace) and Alfonso Iodice D'Enza (University of Cassino and Southern Lazio). Professor Hans Kestler was the laudator and Professor Angelo Markos gave a short presentation on the awarded paper in opening session of the conference.

Sponsors of the ECDA2015 were the data science company *Profusion Ltd, London* and the *Institute for Analytics and Data Science, University of Essex*. The conference dinner was at Colchester Castle, the biggest Keep ever built in the United Kingdom and the largest that remains in existence throughout Europe.

At the general annual meeting the members of GfKI decided to add *Data Science Society* to the name of the Society. The members awarded Professor Gunter Ritter, Passau, the honorary membership of GfKI. As honorary member Gunter Ritter is a member of the GfKI advisory board together with the honorary president Hans Bock and the honorary members Wolfgang Gaul, Otto Opitz and Claus Weihs. The other members of the advisory board were elected – as usual – by the board of the society for a term of three years (2013-16): Christian Hennig (London), Lars Schmidt-Thieme (Hildesheim), Frank Scholze (Karlsruhe), Myra Spiliopoulou (Magdeburg) and Alfred Ultsch (Marburg).

The conference and abstract booklet of ECDA2015 is available under <http://repository.essex.ac.uk/15600/>

Autor:

B. Lausen

## Bericht AG BIB und DK

Die AG Dezimalklassifikationen und die AG Bibliothekswesen haben im Rahmen der ECDA-Jahrestagung in Colchester wiederum eine gemeinsame Veranstaltung zu Fragen aus dem Bereich von LIS (Library and Information Science) durchgeführt. Von besonderer Bedeutung für die Bibliothekarische Klassifikation im Allgemeinen und die Dezimalklassifikationen im Besonderen seien hier folgende Beiträge genannt:

- Franke-Maier, Michael: A choice of nice web-features for subject cataloguing
- Peichl, Gerald: Bibliographic Report 2015. A choice of relevant classification literature
- Balakrishnan, Uma: Cocoda "Colibri Concordance Database" – A mapping tool for library classification schemes

Grundlagen des Projekts sind DDC (mehr Daten im GVK) und RVK (mehr Daten in SWB und BVB) als die Klassifikationen mit der größten Zahl an Erschließungsdaten. Erfreulicherweise ist in den vergangenen Jahren auch die notwendige Arbeit an Konkordanzen in Fahrt gekommen; auch hier sind konsequenterweise RVK und DDC am meisten genannt neben u.a. BK sowie den Fachsystemen MSC und PACS.

- Pieper, Dirk – Summann, Friedrich: The Role of Classification Information in Open Access Repositories. Current status and future directions

Thema der Darstellung ist hier die Arbeit an und mit BASE (Bielefeld Academic Search Engine).

Angefügt seien noch Notizen zur Entwicklung von DDC und UDC:

- a) DDC (vgl. <http://www.oclc.org/dewey>)

Derzeit gehört mit Berndt Dugall ein deutscher Bibliothekar zum "Board of Trustees". Doch wird Deutschland bei den "countries with libraries that use the DDC system" nicht genannt (Stand: 10.11.2015); die DDC-Anwendung außerhalb der Deutschen Nationalbibliographie wird wohl als zu sporadisch angesehen.

- b) UDC (vgl. <http://www.udcc.org>)

Die Ausweitung und Aktivierung der UDC-Arbeit im vergangenen Jahrzehnt ist bemerkenswert:

Allein für das laufende Jahr 2015 sind Kroatisch, Slowakisch, Spanisch und Tschechisch zur Anwendung in UDC online vorgesehen.

Die "Translation Database" mit der Basissprache Englisch sieht weitere neun Sprachen vor, darunter – endlich wieder – Deutsch.

Die 11. Ausgabe des MRF (Master Reference File) umfasst 70.000 UDC-Klassen.

Sehr hilfreich ist auch die umfangreiche Rubrik "Major changes to the UDC 1993-2013", ebenso die "Cancellations".

Autor:

B. Lorenz

## LIS 2016

Der Workshop „Classification and Subject Indexing in Library and Information Science“ (LIS 2016) findet in Verbindung mit dem 4. Joint Statistical Meeting der Deutschen Arbeitsgemeinschaft Statistik (DAGSTAT) und dem 6. Deutschen Bibliothekskongress am 16. März 2016 von 14.00-18.00 Uhr auf der Leipziger Messe statt. Er wird veranstaltet von der Arbeitsgruppe Bibliotheken in der Gesellschaft für Klassifikation (GfKI) und widmet sich aktuellen Fragestellungen der verbalen und klassifikatorischen Sacherschließung in automatisierten oder manuellen Ansätzen. Der LIS Workshop 2016 ist das Forum für Sacherschließung beim Bibliothekskongress. Der Bibliothekskongress ist die größte Veranstaltung im Bereich Bibliotheks- und Informationswesen im deutschsprachigen Raum.

<http://www.professionalabstracts.com/bid2016/iplanner/>

Autor: F. Scholze

## Bericht AG BT

Die traditionelle Arbeitstagung der AG Biostatistik auf Schloss Reisenburg („Statistical Computing“) fand vom 19.7. bis 22.7.2015 statt. Sie wurde in Zusammenarbeit mit der AG Statistical Computing (AG SC) der Biometrischen Gesellschaft DR (GMDS/IBS-DR) durchgeführt. An der diesjährigen Veranstaltung nahmen insgesamt 42 Personen teil. Gestaltet wurde das Programm durch 23 Vorträge aus verschiedensten methodischen Bereichen der Klassifikations-, Regressions- und Survivalanalyse. Diskutiert wurden Themen der Merkmals- und Modellselektion über die Parallelisierbarkeit von Lernalgorithmen bis hin zur visuellen und plastischen Darstellung von Datenmodellen (3D-Druck). Hervorzuheben sind hierbei vor allem die eingeladenen Vorträge: Andreas Ziegler (Lübeck): Random forests: Current concepts and implementations; Holger Fröhlich (Bonn): Probabilistic Modeling of biological high throughput data; und Joachin Vanschoren (Eindhoven): OpenML: Networked science in machine Learning. Bereichert wurde das Treffen zudem durch eine Poster-Session und zwei Hands-on Tutorials, in denen die Teilnehmer durch eigene Experimenten die Anwendung aktueller R-Pakete erproben konnten: Marvin Wright (Lübeck): Implementation of random forests in the R package Ranger; Bernd Bischl und Giuseppe Casalicchio (München): OpenML with R and mlr. Abgerundet wurde das Treffen durch einen Besuch des „Schwörmontags“ im nahegelegenen Ulm.

Ausgewählte Beiträge werden derzeit zusammen mit den Beiträgen der Vorjahresveranstaltung zu einer Special Issue des Journals Computational Statistics zusammengefasst.

Die nächste Statistical Computing findet vom 17.7.2016 bis 20.7.2016 statt. Schon jetzt bestätigte eingeladene Vortragende sind Eyke Hüllermeier und Peter Chronz (Göttingen), der u.a. ein Tutorial der relativ neuen Programmiersprache Julia (als möglichen R Ersatz) geben wird.

Autor: H. Kestler

## Bericht AG DANK

Arbeitsgemeinschaft Datenanalyse und numerische Klassifikation (AG DANK)

### 1) Herbsttagung 2015

Die Herbsttagung 2015 der AG DANK wurde am 20. und 21. November auf Einladung von Herrn Prof. Andreas Geyer-Schulz in Karlsruhe durchgeführt. Gastgeber war das KIT (Karlsruhe Institute of Technology) und hier speziell das Institute of Information Systems and Marketing. Der lokale Veranstalter hatte nicht nur eingeladen, er hat vielmehr engagiert versucht, lokale Teilnehmer und Vorträge einzuwerben und insbesondere die Personen, die potenziell am Thema interessiert sind, gezielt und persönlich angesprochen. Das mündete schließlich in der sehr guten Programmgestaltung einschließlich Versammlungsleitung. Wir bedanken uns hiermit ganz herzlich bei Prof. Geyer-Schulz und seinem Team, insbesondere bei Frau Victoria-Anne Schweigert sowie Herrn Thomas Hummel, für die professionelle und perfekte Organisation.

Das Schwerpunktthema „Herausforderung Data Science und Big Data“ wurde durch mehrere Vorträge und die Sessions „Vorstellung und Diskussion über den entsprechenden Studiengang“ sowie „Talks zur Lehre“ zum Studiengang „Data Science“ herausragend repräsentiert. Letztere wurde durch die Professoren Claus Weihs (Erfahrungsbericht von der Universität Dortmund aus mehr als ein Dutzend Jahren) und Berthold Lausen (Bericht von der University of Essex) sowie Andreas Geyer-Schulz als Moderator gestaltet.

Ein weiterer Höhepunkt war die Würdigung der wissenschaftlichen Arbeiten von Herrn Prof. Hans-Hermann Bock durch Herrn Prof. Gunter Ritter anlässlich seines 75igsten Geburtstages am 8. September 2015. Herr Bock hatte u.a. die Entwicklung und Ausrichtung der AG DANK maßgeblich geprägt und einen hohen Maßstab gesetzt, und er wird hoffentlich, wie gerade in Karlsruhe erlebt, auch weiterhin sehr aktiv mitwirken. Ihnen, lieber Herr Bock wünschen wir eine gute Zeit.

Es wurden folgende Vorträge gehalten:

- Ali Ünlü – Educational Test Mastery Hierarchy Mining with ITA
- Irmela Herzog – Auswertung von digitalen topographischen Daten aus Ecuador, um die Lage von archäologischen Fundstellen und Altwegen zu analysieren
- Gunter Ritter – Alles Gute zum 75ten, Herr Bock!
- M. Schweizer - Vorstellung der studentischen MLKGruppe
- Andreas Geyer-Schulz – And the Future is Data Science
- T. Setzer und S. Nickel – Analytical Enterprise Processes - Designing the Interplay of Data Science, Operations Research and Behavioral Decision Theory
- Achim Mucha – A fast Pre-clustering Method by Iterative Binary Binning
- Christian Hennig – Some Thoughts on Data Preprocessing Including Dimension Reduction with Application to Football Player Performance Data
- Claus Weihs – Data Science (Talks zur Lehre: 1)
- Berthold Lausen – Data Science (Talks zur Lehre: 2)
- Claus Weihs – Big Data Classification - Simulations in the many Features Case
- S. Vökler – Computational Intelligence in Product-line Optimization: Simulations and Applications
- Ulrich Müller-Funk – Sklar's Kopulas versus Hoeffding's normalisierte Verteilungsfunktionen
- G. Steinbuß – Predictive Maintenance: Basierend auf Ersatzdaten Alarme Vorhersagen

Fortsetzung nächste Seite

Es wurde der Bogen von der Universitätsausbildung (Studiengang Data Science und dessen vieljährige Erfahrung in Deutschland und Großbritannien) bis hin zu Anwendungen (echtzeitnahe Big-Data-basierte Datenanalyse) gespannt.

Neben den oben aufgeführten Vorträgen gab es, mit den inzwischen schon traditionellen Datensatzanalysen, weitere Präsentationen. Die zu analysierenden Big Data-Datensätze wurden von Herrn Peter Kurz (TNS Infratest, München) und Herrn Prof. Geyer-Schulz den Teilnehmern vor der Herbsttagung verfügbar gemacht. Unser herzlicher Dank gilt diesen Herren für die Bereitstellung der Datensätze. Zum ersten Datensatz gab es zwei zu prämierende Analysen von Herrn Dr. Reinhard Schachtner (Infineon AG, Regensburg) und Dr. Marvin Schweizer (KIT). Die Analysen des zweiten Datensatzes werden Gegenstand der nächsten Herbsttagung sein.

Die Folien zu den Vorträgen sowie die zur Analyse zur Verfügung gestellten Datensätze finden sich im Web unter <https://www.em.uni-karlsruhe.de/ag-dank2015/>.

## 2) Herbsttagung 2016

Zur kommenden Herbsttagung der AG DANK in 2016 lade ich alle herzlich nach Berlin an das WIAS ein. Schwerpunktthemen sind „Data Acquisition and Pre-Processing“ und „Big Data Clustering and Classification“. Es sind, wie immer, auch interessante und inspirierende Vorträge aus dem gesamten Gebiet der statistischen Datenanalyse willkommen. Es besteht die Möglichkeit der Publikation ausgewählter Beiträge in einem Buch. Als vorläufiger (wahrscheinlicher) Termin zeichnet sich der 18/19.11.2016 ab, da dafür bereits Zusagen eingeladener Sprecher vorliegen. Erstmals sollen bei der Programmplanung mindestens 10 Minuten je Vortrag als Diskussionszeit angesetzt werden, um noch mehr Diskussionen je Vortrag zu ermöglichen. Das geht auf eine Anregung von Herrn Dr. Christian Hennig zurück, der in Karlsruhe über sehr positive Erfahrungen aus Seattle berichtete, wo Vortrags- und Diskussionszeit hälftig geplant sind.

Berlin, Dezember 2015, Hans-Joachim Mucha

Autor:	H.-J. Mucha
--------	-------------

## ADAC – Advances in Data Analysis and Classification

Das neue Heft 4 von Band 9 (2015) ist ein Sonderheft zum Thema ‚New trends in model-based classification and clustering‘, das von den Gasteditoren Salvatore Ingrassia, Geoffrey J. McLachlan, Gérard Govaert zusammengestellt wurde. Es enthält sechs Arbeiten, und zwar zu: ML-Schätzung bei Mischungsmodellen ohne Matrixoperationen, Wahl der Anfangspartition bei Gauss’schen Mischungsmodellen, Bayes’sche Methoden zur Bewertung von Clusterergebnissen, robuste Clustermethoden bei schiefen t-Verteilungen, Simulation von multivariaten Mischungen mit vorgegebenem Klassen-Overlap und die Analyse von latenten Übergängen bei speziellen Hidden Markov Models. Im Einzelnen:

Hien D. Nguyen, Geoffrey J. McLachlan:

- Maximum likelihood estimation of Gaussian mixture models without matrix operations

Lucca Scrucca, Adrian E. Raftery:

- Improved initialisation of model-based clustering using Gaussian hierarchical partitions

Xuwen Zhu, Volodymyr Melnykov:

- Probabilistic assessment of model-based clustering

Wan-Lun Wang, Tsung-I Lin:

- Robust model-based clustering via mixtures of skew-t distributions with missing information

Marco Riani, Andrea Cerioli, Domenico Perrotta, Francesca Torti:

- Simulating mixtures of multivariate data with fixed cluster overlap in FSDA

M. Francesca Marino, Marco Alfò:

- Latent drop-out based transitions in linear quantile hidden Markov models for longitudinal responses with attrition

Eine vollständige Liste aller bislang (seit 2007) in ADAC erschienenen Artikel ist auf der GfKI-Website zu finden.

Hans-Hermann Bock

Autor:	H.-H. Bock
--------	------------

## Archives of Data Science (Series A)

Seit Ende August 2015 sind die ersten Beiträge zum neuen Journal Archives of Data Science Series A <http://www.ArchivesOfDataScience.org/> online. Der erste Band zur GPSDAA-Konferenz 2013 in Polen steht knapp vor der Fertigstellung. Die Beiträge für die ECDA 2015-Konferenz werden im zweiten Band voraussichtlich ab Mai 2016 erscheinen.

Dem Design haben wir das Motiv von Piet Mondrians Composition with Red, Yellow, and Blue zugrunde gelegt, um Assoziationen mit konkreten Anwendungen von Data Science zu vermeiden und dem abstrakten und interdisziplinären Charakter von Data Science gerecht zu werden.

Vor fast 50 Jahren haben Allen Newell, Alan J. Perlis und Herbert A. Simon die Frage "Was ist Computer Science?" kurz und knapp als die Wissenschaft von Computern definiert (Science, 157(3795), S. 1374-1375).

Wo es Phänomene gibt, kann es eine Wissenschaft geben, die diese Phänomene beschreibt und erklärt. Ergo, Data Science ist die Erforschung von Daten. Die Phänomene von Daten sind vielgestaltig, komplex und reich. Und dabei hat die Zeit der Daten gerade begonnen.

Die Archives of Data Science sind diesen Phänomenen und ihrer wissenschaftlichen Erforschung gewidmet.

Wir bitten Sie uns Feedback bei Problemen mit der Website bzw. mit den LaTeX-Styles zu geben.

Andreas Geyer-Schulz,  
Eyke Hüllermeier,  
Hans A. Kestler

Dezember 2015, Karlsruhe, Paderborn, Jena

Autor:	A. Geyer-Schulz
--------	-----------------

## BRCA Data Analysis Award

### *Data Analysis in Inherited BRCA Cancer Research*

Under the auspices of: Gesellschaft für Klassifikation e.V. – Data Science Society (GfKI)  
International Federation of Classification Societies (IFCS)  
International Biometric Society, German Region (IBS-DR)  
Deutsches Konsortium Familiärer Brust- und Eierstockkrebs  
HeritX, Sparking Solutions for Inherited Cancer

### Call for Submissions

The **BRCA Data Analysis Award** will be conferred for original and innovative work that uses data-analytic and statistical methods in order to support the development of strategies for the prevention of inherited cancer caused by mutations in the BRCA (Breast Cancer) gene.

- **Data analysis techniques**, combined with computational algorithms, are broadly and routinely applied in many fields like medicine, bioinformatics, economics, marketing, pattern recognition, sociology, etc. Typical approaches include statistical modeling, classification and clustering, factorial analysis, exploratory tools, and visualization. Representative scientific societies are, e.g., the *International Federation of Classification Societies*, the *German Classification Society*, the *Biometric Society*, and others.
- This award concentrates on the **domain of inherited cancer research**. Inherited cancer is affecting millions of people worldwide. Its physical, emotional and often financial burden devastates entire families across genders and generations. The most prominent type of inherited cancer is caused by a genetic defect (mutation) in the BRCA gene. A carrier of this gene defect will, with large probability, develop one or more cancers including breast cancer, ovarian cancer, pancreatic cancer, prostate or skin cancer. Fifty percent of the children will inherit the mutated gene and the same staggering cancer risk as well.

Inherited cancer can and must be prevented before it starts rather than treated after it has occurred. This appears achievable because – unlike in other cancers - we already know the first step in the development of inherited BRCA cancer: the inherited gene defect. However, there is a lack of research and innovation in transforming this knowledge into an available preventive therapy. To change this situation, novel approaches are needed and might be supported by data-analytic methods. The **BRCA Data Analysis Award** pursues this important goal, supported by the *Deutsches Konsortium Familiärer Brust- und Eierstockkrebs* and a global research initiative for preventing inherited BRCA cancer led by *HeritX*, a patient-driven, nonprofit research and development organization in the USA.

- *The BRCA Data Analysis Award* has been introduced to recognize a compelling investigation that is related to inherited BRCA genes and may benefit the early diagnosis and/or medical prevention of BRCA-related cancer by using data-analytic methods. The award includes a honorarium of 1000 € (for personal use) and the presentation (in person) of the author's work at one of the major conferences on data analysis and its applications.
- The award not only honors relevant research in the BRCA field, but also is intended to motivate researchers across the data analysis community in the early stage of their careers to focus on solutions on this field. Qualified research needs to be completed in 2014-2016, and its author must be  $\leq 40$  years. The award winner will be selected through an expert committee including experts in BRCA research and data analysis, orchestrated by Hans-Hermann Bock.

Fortsetzung nächste Seite



Interested researchers are invited to submit a nomination package including:

1. a brief summary (two pages or less) of the work or investigation on which the candidacy is based, which explains a) its relevance to the field of BRCA, b) the proposed data analytic approach, and c) its potential benefit for the early diagnosis and/or medical prevention of BRCA-related cancer
2. up to three published papers, an accepted thesis, or an accepted manuscript
3. a curriculum vitae, and
4. the names and addresses of two possible referees

to [BRCA-Award@gfkl.de](mailto:BRCA-Award@gfkl.de), until September 30, 2016. Candidates will be notified by the end of 2016. For further information see the website [www.isw.rwth-aachen.de/BRCA-Award.php](http://www.isw.rwth-aachen.de/BRCA-Award.php)

November 2015

*Hans-Hermann Bock*

*Institute of Statistics, RWTH Aachen University, Aachen, Germany*

*bock@stochastik.rwth-aachen.de*

Autor:

H.-H. Bock