



OTTO VON GUERICKE  
UNIVERSITÄT  
MAGDEBURG

**MATH**

FAKULTÄT FÜR  
MATHEMATIK

# Forschungsbericht 2020

Institut für Mathematische Stochastik

# INSTITUT FÜR MATHEMATISCHE STOCHASTIK

Universitätsplatz 2, 39106 Magdeburg  
Tel. 49 (0)391 67 58651, Fax 49 (0)391 67 41172  
imst@ovgu.de

## 1. LEITUNG

Prof. Dr. Alexandra Carpentier - geschäftsführende Leiterin  
Prof. Dr. Anja Janßen (seit 01.09.2020)  
Prof. Dr. Claudia Kirch  
Prof. Dr. Rainer Schwabe (bis 31.03.2020)

Dr. Heiko Großmann  
Priv.-Doz. Dr. Martin Wendler

## 2. HOCHSCHULLEHRER/INNEN

Prof. Dr. Alexandra Carpentier  
Prof. Dr. Anja Janßen  
Prof. Dr. Claudia Kirch  
Prof. Dr. Rainer Schwabe (bis 31.03.2020)

apl. Prof. Dr. Berthold Heiligers (extern)  
Priv.-Doz. Dr. Martin Wendler

Professoren im Ruhestand:  
Prof. em. Dr. Otfried Beyer  
Prof. Dr. Gerd Christoph  
Prof. Dr. Norbert Gaffke  
Prof. Dr. Rainer Schwabe (seit 01.04.2020)

## 3. FORSCHUNGSPROFIL

Mathematische Stochastik (Mathematische Statistik und Maschinelles Lernen): Prof. Dr. Alexandra Carpentier

- High or Infinite-Dimensional Adaptive Inference
- Uncertainty Quantification and Adaptive Confidence Sets
- Composite-Composite Testing Theory
- Sequential Sampling, Bandit Theory
- Optimisation of Computational Resources
- Inverse Problems and Compressed Sensing
- Applications in Statistical Problems (like regression/non-parametric estimation/matrix completion/extreme value theory/anomaly detection, etc)

Mathematische Stochastik (Stochastische Prozesse): Prof. Dr. Gerd Christoph; apl. Prof. Dr. Waltraud Kahle

- Asymptotische Methoden in der Stochastik
- Edgeworth und Cornish-Fisher Entwicklungen
- Statistik in Abnutzungsprozessen mit unvollständiger Reparatur
- Optimale unvollständige Instandhaltung in Abnutzungsprozessen
- Optimale Instandhaltung in allgemeinen Ausfall-Reparatur-Prozessen bei diskreten Lebensdauerverteilungen

Mathematische Stochastik (Mathematische Statistik): Prof. Dr. Norbert Gaffke

- Statistische Regressionsmodelle
- Experimental Design: Theorie und Algorithmen
- Tests und Konfidenzschranken
- Statistische Modellierung interdisziplinär

Mathematische Stochastik (Angewandte Mathematische Stochastik): Prof. Dr. Anja Janßen

- Extremwerttheorie
- Nicht- und semiparametrische Extremwertstatistik
- Abhängigkeitsmodellierung
- Zeitreihenanalyse, insbesondere in Bezug auf das Extremwertverhalten
- Grenzwertsätze
- Anwendungen im Bereich der Wirtschaftswissenschaften, insbesondere im Risikomanagement und der Modellierung von Finanzzeitreihen

Mathematische Stochastik (Mathematische Statistik und Anwendungen): Prof. Dr. Claudia Kirch; Priv.-Doz. Dr. Martin Wendler

- Zeitreihenanalyse und Signalverarbeitung
- Changepoint-Analyse und Daten-Segmentierung
- Probabilistische Unsicherheitsquantifizierung
- Computationelle und Machine-Learning-Methoden
- Funktionale/Hochdimensionale Daten
- Sequentielle Methoden
- Anwendungen in den Neurowissenschaften
- Nichtparametrische statistische Methoden

Mathematische Stochastik (Statistik und ihre Anwendungen): Prof. Dr. Rainer Schwabe; Dr. Heiko Großmann

- Planung und Auswertung statistischer Experimente
- Conjoint-Analyse (Psychologie, Marktforschung)
- Intelligenzforschung (Psychologie)
- Populationspharmakokinetik (Arzneimittelforschung)
- Adaptive und gruppensequenzielle Verfahren
- Diagnostische Studien mit räumlicher Datenstruktur und zeitlicher Verlaufskontrolle (Perimetrie in der Augenheilkunde)
- Klinische Dosisfindungsstudien
- Statistik in industriellen Anwendungen
- Multivariate Äquivalenz und Nichtunterlegenheit
- Multizentrische Studien
- Lineare, verallgemeinert lineare und nichtlineare gemischte Modelle

- Optimale Auswahl von Teilstichproben in großen Datenmengen

## 4. SERVICEANGEBOT

### **Beratung und Unterstützung bei allen statistischen Fragestellungen**

Das Institut für Mathematische Stochastik bietet Beratung zur Planung und statistischen Auswertung von Experimenten an, insbesondere:

- zur Unterstützung von Abschlussarbeiten bei der Konzeption und Durchführung von Studien
- bei der Stichproben-/ Versuchsplanung, Datengewinnung und Sicherstellung der Datenqualität
- bei der Auswahl und Anwendung geeigneter Analysemethoden
- bei der Interpretation und Präsentation der Untersuchungsergebnisse

### **Dieses Angebot richtet sich an ...**

- Studierende und Promovierende der Otto-von-Guericke-Universität Magdeburg (OVGU)
- Ausgenommen sind Personen, die mit dem Fachbereich Medizin assoziiert sind. (Das Universitätsklinikum bietet über das Institut für Biometrie und Medizinische Informatik Statistikberatungen an.)

[http://www.statistik.ovgu.de/Statistische Beratung.html](http://www.statistik.ovgu.de/Statistische_Beratung.html)

## 5. KOOPERATIONEN

- Annika Betken, Ruhr-Universität Bochum
- Dr. Celine Duval, Universite Paris Descartes, France
- Dr. Daniel Vogel
- Dr. Debarghya Ghoshdastidar, Universitaet Tuingen, Germany
- Dr. Etienne Roquain, Universite Paris VI, France
- Dr. Frenkel, Beer Sheva, Israel Sami Shamoon College of Engineering, Israel
- Dr. Fritjof Freise, TU Dortmund
- Dr. Maureen Cerc, INRIA Sophia Antipolis, France
- Dr. Michal Valko, INRIA Lille Nord Europe, France
- Dr. Nicolas Verzelen, INRA Montpellier, France
- Dr. Olga Klopp, ESSEC Business School, France
- Dr. Patricio Maturana Russel, Auckland University, New Zealand
- Dr. Sylvain Delattre, Universite Paris VI, France
- Juliette Achdou, HEC and Telecom Paris, France
- Oleksandr Zadorozhnyi, Universitaet Potsdam, Germany
- Priv.-Doz. Dr. Ekkehard Glimm, Novartis Pharma AG, Basel
- Priv.-Doz. Dr. Norbert Benda, BfArM, Bonn
- Prof. Dr. Andreas Greven, Universität Erlangen-Nürnberg
- Prof. Dr. Arlene K.H. Kim, Sungshin Women's University, Korea
- Prof. Dr. Bharath Sriperumbudur, Penn State University, USA
- Prof. Dr. Christian Paroissin, Universität Pau, Frankreich
- Prof. Dr. Gilles Blanchard, Universitaet Potsdam, Germany
- Prof. Dr. Haeran Cho, University of Bristol
- Prof. Dr. Heinz Holling, Westfälische Wilhelms-Universität Münster
- Prof. Dr. Herold Dehling
- Prof. Dr. Holger Drees, Universität Hamburg
- Prof. Dr. Idris Eckley, Lancaster University

- Prof. Dr. John Aston, University of Cambridge
- Prof. Dr. Laura Gibson, University of Massachusetts Medical School, USA
- Prof. Dr. Luc Pronzato, Université de Nice, CNRS-13R
- Prof. Dr. Olimjon Sh. Sharipov, National University of Usbekistan
- Prof. Dr. Radoslav Harman, Comenius-Universität Bratislava
- Prof. Dr. Renate Meyer, University of Auckland, New Zealand
- Prof. Dr. Richard Nickl, University of Cambridge, UK
- Prof. Dr. Samory Kpotufe, Princeton University, USA
- Prof. Dr. Sidney Resnick, Cornell University
- Prof. Dr. Sophie Mercier, Universität Pau, Frankreich
- Prof. Dr. Thomas Kahle, FMA-IAG
- Prof. Dr. Timothy Kowalik, University of Massachusetts Medical School, USA
- Prof. Dr. Ulrike von Luxburg, Universität Tübingen, Germany
- Prof. Dr. Vladimir Ulyanov, Moskauer Staatliche Lomonosov-Universität, Russische Föderation

## 6. FORSCHUNGSPROJEKTE

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Kooperationen:** Anne Maneugueu; Gilles Blanchard; Oleksandr Zadorozhnyi  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.11.2018 - 30.11.2021

### Participation in the SFB 1294 on Data Assimilation in Potsdam

The group is also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) on the SFB 1294 Data Assimilation "Data Assimilation - The seamless integration of data and models" on Project A03 together with Prof. Gilles Blanchard.

This project is concerned with the problem of learning sequentially, adaptively and in partial information on an uncertain environment. In this setting, the learner collects sequentially and actively the data, which is not available before-hand in a batch form. The process is as follows: at each time  $t$ , the learner chooses an action and receives a data point, that depends on the performed action. The learner collects data in order to learn the system, but also to achieve a goal (characterized by an objective function) that depends on the application. In this project, we will aim at solving this problem under general objective functions, and dependency in the data collecting process - exploring variations of the so-called bandit setting which corresponds to this problem with a specific objective function.

As a motivating example, consider the problem of sequential and active attention detection through an eye tracker. A human user is looking at a screen, and the objective of an automatized monitor (learner) is to identify through an eye tracker zones of this screen where the user is not paying sufficient attention. In order to do so, the monitor is allowed at each time  $t$  to flash a small zone  $a_t$  in the screen, e.g. light a pixel (action), and the eye tracker detects through the eye movement if the user has observed this flash. Ideally the monitor should focus on these difficult zones and flash more often there (i.e. choose more often specific actions corresponding to less identified zones). Therefore, sequential and adaptive learning methods are expected to improve the performances of the monitor.

---

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.09.2018 - 31.10.2021

### Teilnahme an dem GK Daedalus 2433 mit der TU Berlin

The main goal of DAEDALUS is the analysis of the interplay between incorporation of data and differential equation-based modeling, which is one of the key problems in model-based research of the 21st century. DAEDALUS focuses both on theoretical insights and on applications in life sciences (brain-computer interfaces and biochemistry) as well as in fluid dynamics. The projects cover a scientific range from machine learning, mathematical theory of model reduction and uncertainty quantification to respective applications in turbulence

theory, simulation of complex nonlinear flows as well as of molecular dynamics in chemical and biological systems. In our group, we cover mathematical statistics and machine learning aspects.

This project is in the context of Daedalus, and is concerned with uncertainty quantification in complex cases.

---

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Projektbearbeitung:** M.Sc. Joseph Lam  
**Kooperationen:** Prof. Dr. Bharath Sriperumbudur, Penn State University, USA; Joseph Lam, FMA-IMST  
**Förderer:** Haushalt - 01.11.2017 - 31.10.2020

### **Adaptive two sample test in the density setting**

We consider the problem of testing between two samples of (non necessarily uniform) density. While minimax signal detection in the case where the null hypothesis density is uniform is well understood, recent works in the case of multinomial distributions have highlighted the amelioration in the minimax rate that can come when considering non uniform null hypothesis density. We want to study this problem in the two sample testing case, which is significantly more complex, and extend it to smooth densities.

---

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Kooperationen:** Prof. Dr. Gilles Blanchard, Universitaet Potsdam, Germany; Oleksandr Zadorozhnyi, Universitaet Potsdam, Germany; Anne Maneugueu; Dr. Claire Vernade; Dr. Michal Valko, INRIA Lille Nord Europe, France  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.10.2017 - 31.08.2020

### **Projekt on Data Assimilation**

This project is concerned with the problem of learning sequentially, adaptively and in partial information on an uncertain environment. In this setting, the learner collects sequentially and actively the data, which is not available before-hand in a batch form. The process is as follows: at each time  $t$ , the learner chooses an action and receives a data point, that depends on the performed action. The learner collects data in order to learn the system, but also to achieve a goal (characterized by an objective function) that depends on the application. In this project, we will aim at solving this problem under general objective functions, and dependency in the data collecting process exploring variations of the so-called bandit setting which corresponds to this problem with a specific objective function.

As a motivating example, consider the problem of sequential and active attention detection through an eye tracker. A human user is looking at a screen, and the objective of an automatized monitor (learner) is to identify through an eye tracker zones of this screen where the user is not paying sufficient attention. In order to do so, the monitor is allowed at each time  $t$  to flash a small zone  $a_t$  in the screen, e.g. light a pixel (action), and the eye tracker detects through the eye movement if the user has observed this flash. Ideally the monitor should focus on these difficult zones and flash more often there (i.e. choose more often specific actions corresponding to less identified zones). Therefore, sequential and adaptive learning methods are expected to improve the performances of the monitor.

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Kooperationen:** Emmanuel Pilliat; Dr. Nicolas Verzelen, INRA Montpellier, France  
**Förderer:** Haushalt - 01.01.2019 - 01.10.2021

### **Minimax change point detection in high dimension**

The objective is to establish the minimax rates for sparse change point estimation in high dimension. We want in particular to investigate in a refined way intermediary regimes. Joint project with Emmanuel Pilliat and Dr. Nicolas Verzelen.

---

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Kooperationen:** James Cheshire; Prof. Dr. Sebastian Sager  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.01.2019 - 30.09.2021

### **Participation in the GK 2297 Mathcore**

The objective of this GRK is to investigate the problem of complexity reduction across the different areas of mathematics. In our group, we bring to this project some expertise on the field of sequential learning, in order to reduce the complexity of given problems by adapting the sampling strategies.

---

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Kooperationen:** Dr. Maurilio Gutzeit; Andrea Locatelli; James Cheshire; Anne Maneugueu; Joseph Lam  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.10.2017 - 31.08.2020

### **MuSyAD on Anomaly Detection**

Anomaly detection is an interdisciplinary domain, borrowing elements from mathematics, computer science, and engineering. The main aim is to develop efficient techniques for detecting anomalous behaviour of systems. In the classical scenario a monitor receives data from a system and compares this data to a reference system with some single normal behaviour. Ideally no strong assumptions are made on the nature of anomalous behaviours, so the problem of anomaly detection is by essence a non parametric problem. Here I propose to study a more complex scenario, which will be referred to as multisystem anomaly detection. In this setting, reference systems can have a variety of normal behaviours, and moreover, there are many systems under the monitor's surveillance, and the monitor must allocate its resources wisely among them. In this situation new theoretical and computational challenges arise. The overall objective of this proposal is to find efficient methods to solve the problem of multi-system anomaly detection. This aim will be reached by addressing the following sub-objectives. First, we will generalise the theoretical framework of anomaly detection to the broader setting of multi-system anomaly detection. Second, multi-system anomaly detection methods will be developed, by taking ideas from the non parametric testing field and applying them to the new framework. Third, we will study optimal monitoring strategies for cases where the multiple systems cannot be monitored simultaneously. Here, it is important that the monitor allocates its resources among the systems in a way that is as efficient as possible. To this end, sequential and adaptive sampling methods that target the anomaly detection problem will be designed. Since anomaly detection is a non parametric problem, elements in the theory of non parametric confidence sets will be used. Finally, the newly developed methods will be applied to practical problems: a methodological example in extreme value theory, an economic application for speculative bubble detection and two applications in a Brain Computer Interface framework.

---

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Kooperationen:** Dr. Olivier Collier; Dr. Laetitia Comminges; Prof. Dr. Alexandre Tsybakov; Yuhaho Wang  
**Förderer:** Haushalt - 01.01.2019 - 31.10.2021

### **Minimax testing rates in linear regression**

In this project we focus on finding the minimax testing rates in  $L_2$  norm for the linear regression model. We also investigate the problem of estimating optimally the  $L_2$  norm for the parameter. We close some gaps in linear regression.

---

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Kooperationen:** Prof. Dr. Cristina Butucea; Julien Chhor; Prof. Dr. Rajarshi Mukherjee  
**Förderer:** Haushalt - 01.01.2019 - 01.10.2021

### **One sample local test in the Graph model**

In this project we aim at finding minimax rates for the problem of local testing in the graph model, in  $L_q$  norm. We focus particularly on local rates, and aim also at the multinomial tetsig model, which can be seen as a special case.

---

**Projektleitung:** Prof. Dr. Alexandra Carpentier  
**Kooperationen:** Dr. Pierre Menard; Joseph Lam  
**Förderer:** EU - ESF Sachsen-Anhalt - 01.06.2020 - 30.06.2022

### **Risk Estimation for Brain-Computer Interfaces**

The project RE-BCI was awarded in the beginning of 2020 by the Land Sachsen Anhalt, more precisely by the Sachsen-Anhalt WISSENSCHAFT Spitzenforschung/Synergien. The objective of RE-BCI is to prepare preliminary results supporting the BCI (Brain-Computer Interfaces, i.e. a technology for connecting a human user with a computer through the lectrical impulses emitted by her/his brain) application to shared authority situations.

---

**Projektleitung:** Prof. Dr. Gerd Christoph  
**Projektbearbeitung:** Prof. Dr. Vladimir V. Ulyanov  
**Förderer:** Sonstige - 01.04.2020 - 31.12.2021

### **Analysis of the quality of approximations in the statistical analysis of multivariate observations**

We consider high-dimension low-sample-size data taken from the standard multivariate normal distribution under assumption that dimension of the vectors is a random variable. Three geometric statistics of the normal vectors are investigated, the length of a observation vector, the difference between any two independent observation vectors and the angle between these vectors at the population mean. The second order Chebyshev-Edgeworth expansions for distributions of these geometric statistics are constructed with error bounds. The poject continues studies on approximation of statistics for random size samples.

---



**Projektleitung:** Prof. Dr. Norbert Gaffke  
**Kooperationen:** Prof. Dr. Rainer Schwabe, OVGU, FMA-IMST  
**Förderer:** Sonstige - 01.10.2020 - 30.09.2022

### **Quasi-Newton algorithmus zum optimalen Design**

Im Rahmen der approximativen Design-Theorie für lineare Regressionsmodelle sollen optimale Designs algorithmisch berechnet werden (insbesondere D-optimale und I-optimale Designs). Ein universell einsetzbarer Algorithmus existiert nicht.

Unsere Quasi-Newton Methoden (s. Gaffke; Schwabe, 2019) sollen auf den Fall eines endlichen Versuchsbereichs angewendet und als R-Programm implementiert werden.

Literatur:

Gaffke, N.; Schwabe, R.: Quasi-Newton algorithm for optimal approximate linear regression design: Optimization in matrix space. *Journal of Statistical Planning and Inference* 198 (2019), 62-78.

---

**Projektleitung:** Prof. Dr. Anja Janßen  
**Kooperationen:** Sebastian Neblung, M.Sc., Universität Hamburg; Prof. Dr. Holger Drees, Universität Hamburg  
**Förderer:** Haushalt - 01.09.2020 - 31.12.2021

### **Asymptotic analysis of tail index estimation for network data**

Real-life networks like online communities or citation relationships are often observed to be approximately scale-free, meaning that the empirical degree distribution has a regularly varying behavior. In theory, this fact is reflected for example by the class of preferential attachment models. The index of regular variation is an important parameter to describe the extremal behavior of those models and thus an important quantity to estimate from network data, often estimated by methods developed for i.i.d. data like the famous Hill-estimator. However, since network data is non-i.i.d., in particular the largest nodes have a strong asymptotic dependence structure, the asymptotic behavior of those estimators is largely unknown, with so far only a few consistency results available. The aim of this project is to develop concentration inequalities tailored to the tails of the joint distribution of largest nodes that allow us to derive finer results like asymptotic normality of estimators.

---

**Projektleitung:** Prof. Dr. Anja Janßen  
**Kooperationen:** Sebastian Neblung, M.Sc., Universität Hamburg; Prof. Dr. Holger Drees, Universität Hamburg  
**Förderer:** Haushalt - 01.09.2020 - 30.09.2021

### **Cluster based inference for extremes of time series**

This work is part of the Ph.D.-project of Sebastian Neblung, for whom I am the second supervisor.

In this project we introduce a new type of estimator for the spectral tail process of a regularly varying time series. The approach is based on a characterizing invariance property of the spectral tail process which has been derived in Janßen (2019) and is incorporated into the new estimator via a projection technique. Based on the limit results for empirical tail processes developed in Drees & Neblung (2019), we show uniform asymptotic normality of this estimator both in the case of known and unknown index of regular variation. A simulation study illustrates that the new procedure provides an often more stable alternative to previous estimators.

---

**Projektleitung:** Prof. Dr. Claudia Kirch  
**Kooperationen:** Dr. Haeran Cho, University of Bristol, UK  
**Förderer:** Haushalt - 01.01.2018 - 31.12.2020

### **Multiscale MOSUM procedure with localised pruning**

In this work, we investigate the detection and estimation of multiple change-points in the mean of univariate data. A localised methodology is proposed for pruning down possibly conflicting change-point estimators computed from any change-point procedure that supplies the information about the local interval in which they are detected. We establish the theoretical consistency of the proposed localised pruning method in combination with the multiscale extension of the MOving SUM (MOSUM) procedure by Eichinger and Kirch (2018). Extensive simulation studies show the computational efficiency and good finite sample performance of the combined methodology.

---

**Projektleitung:** Prof. Dr. Claudia Kirch  
**Kooperationen:** Dr. Christina Stöhr (Ruhr-Universität Bochum)  
**Förderer:** Haushalt - 01.01.2019 - 31.12.2020

### **Sequential change point tests based on U-statistics**

We propose a general framework of sequential testing procedures based on U-statistics which contains as an example a sequential CUSUM test based on differences in mean but also includes a robust sequential Wilcoxon change point procedure. Within this framework, we consider several monitoring schemes that take different observations into account to make a decision at a given time point. Unlike the originally proposed scheme that takes all observations of the monitoring period into account, we also consider a modified moving-sum-version as well as a version of a Page-monitoring scheme. The latter behave almost as good for early changes while being advantageous for later changes. For all proposed procedures we provide the limit distribution under the null hypothesis which yields the threshold to control the asymptotic type-I-error. Furthermore, we show that the proposed tests have asymptotic power one. In a simulation study we compare the performance of the sequential procedures via their empirical size, power and detection delay and give a data example.

Keywords: structural breaks, Wilcoxon statistics, CUSUM statistics, data monitoring, control charts

---

**Projektleitung:** Prof. Dr. Claudia Kirch  
**Projektbearbeitung:** M.Sc. Felix Gnettner  
**Kooperationen:** Prof. Dr. Alicia Nieto-Reyes (University of Cantabria, Santander)  
**Förderer:** Haushalt - 01.10.2019 - 30.09.2022

### **Change Point Tests based on Depth Functions**

Depth functions provide measures of the deepness of a point with respect to a given set of observations. This non-parametric concept can be applied in spaces of any dimension and entails a center-outward ordering for the given data. In 1993 Liu and Singh published a new idea for a Wilcoxon-type two-sample test considering generalised depth-based ranks and in 2006 Zuo and He proved the test statistic to be asymptotically normal. Our aim is to construct change point tests by means of this Liu-Singh statistic and to investigate their asymptotic properties. Those tests that prove beneficial should be implemented such that a performant evaluation is enabled. In particular, we are interested in the behaviour of tests for high-dimensional or functional data.

---

**Projektleitung:** Prof. Dr. Claudia Kirch  
**Kooperationen:** Dr. Christina Stöhr, Ruhr-Universität Bochum  
**Förderer:** Haushalt - 01.01.2019 - 31.12.2020

### Stopping times for sequential U-Statistics

To understand the speed of detection is of particular importance in sequential change point analysis as, for example, monitoring patient or machine data requires a quick intervention as soon as possible after a structural break has occurred. Therefore, we derive the limit distribution of the delay time for a general framework of sequential change point procedures based on U-statistics for early as well as late change points. The asymptotic delay time for late changes has not been considered in the literature before, not even for the classical sequential CUSUM procedure, and requires different asymptotic considerations. Based on the asymptotic behavior of the delay time we provide a theoretical comparison of the sequential CUSUM procedure and a more robust Wilcoxon-type procedure in terms of their speed of detection. The approximation of the stopping time for finite samples via the limit distribution is evaluated by some simulations.

---

**Projektleitung:** Prof. Dr. Claudia Kirch  
**Kooperationen:** Prof. Dr. Renate Meyer, University of Auckland, New Zealand; Dr. Alexander Meier  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.01.2019 - 31.12.2021

### Bayessche semiparametrische Modelle mit Zeitreihenfehlern

Die Bayessche Zeitreihenanalyse erfreut sich zunehmend wachsender Beliebtheit in der Fachliteratur. Oft geht man hierbei in der Modellierung von einer stationären zentrierten Zeitreihe aus. In vielen relevanten Fällen stellt eine solche Zeitreihe jedoch nicht das primäre Objekt von Interesse dar, sondern wird lediglich als Fehlerterm in einem Modell mit zusätzlichem (endlichdimensionalem) "Parameter von Interesse" zugrunde gelegt. Beispiele hierfür reichen von linearen Modellen (mit Modelkoeffizienten als Parameter von Interesse) über Strukturbruch-Modelle (mit den Strukturbrüchen als Parameter von Interesse) bis hin zur nichtlinearen Regression (mit Regressionsfunktion als Parameter von Interesse). Wenn man sich für den Fehlerterm nicht auf ein endlichdimensionales Zeitreihenmodell beschränken möchte, besteht die Möglichkeit, diesen nichtparametrisch zu modellieren – man spricht in diesem Fall von einem semiparametrischen Modell.

Obwohl es einige Arbeiten zu Bayesschen semiparametrischen Modellen in der Fachliteratur gibt, sind dennoch wenig semiparametrische Ansätze im Zeitreihen-Kontext entwickelt worden. Insbesondere mit Blick auf asymptotische Betrachtungen gibt es zudem kaum theoretische Erkenntnisse.

Wir betrachten ein Bayessches semiparametrisches lineares Modell, mit Fehlerterm bestehend aus einer stationären zentrierten Zeitreihe, welche nichtparametrisch mit einem Bernstein-Hpd-Gamma Prior für die Spektraldichtematrix im Zusammenspiel mit der Whittle Likelihood modelliert wird. Die Resultate des Verfahrens werden in einer vergleichenden Simulationsstudie evaluiert. Für den wichtigen Spezialfall des Erwartungswert-Modells werden zudem Kontraktionsraten der gemeinsamen a posteriori Verteilung sowie ein Bernstein-von-Mises Resultat für die marginale a posteriori Verteilung des Erwartungswerts hergeleitet.

---

**Projektleitung:** Prof. Dr. Claudia Kirch  
**Kooperationen:** Dr. Kerstin Reckrühm  
**Förderer:** Haushalt - 01.04.2015 - 31.03.2021

### Die Detektion multipler Strukturbrüche basierend auf dem MOSUM-Verfahren

Es existieren zwei grundlegende Verfahren zur Erkennung multipler Strukturbrüche in Zeitreihen im klassischen Modell der Erwartungswertänderung, die binäre Segmentierung und das MOSUM-Verfahren. Das

Segmentierungsverfahren ist eine iterative Methode, die ausnutzt, dass Tests für Ein-Changepoint-Alternativen weiterhin Macht im Fall von multiplen Änderungen besitzen. Die zweite Methode hingegen basiert auf Statistiken, die gleitende Summen verwenden. Ein Vorteil des MOSUM-Verfahrens besteht darin, dass das Gesamtsignifikanzniveau kontrolliert werden kann. Tests und statistische Eigenschaften von Changepoint Schätzern, die auf derartige Statistiken gleitender Summen basieren, wurden von Kirch und Muhsal (2015+) im klassischen Erwartungswert-Modell detailliert untersucht. Diese Resultate sollen nun für verschiedene Changepoint Situationen verallgemeinert werden. Durch die Verwendung von MOSUM-Statistiken basierend auf Schätzfunktionen können Modelle verschiedener Parameteränderungen in ein Erwartungswert-Modell der Schätzfunktion transformiert werden. Dazu muss lediglich der globale Schätzer ermittelt werden, was einen großen Vorteil in Bezug auf den Rechenaufwand darstellt. Wir konstruieren eine entsprechende Teststatistik und analysieren ihr asymptotisches Verhalten unter der Nullhypothese und Alternativen. Weiterhin werden die zugehörigen Changepoint Schätzer hinsichtlich ihrer Konsistenzeigenschaften näher untersucht.

Das Hauptproblem des MOSUM-Verfahrens besteht darin, dass die Güte dieser Methode im Wesentlichen von der Wahl der Bandbreite  $G$  abhängt. Dies erweist sich insbesondere dann als sehr problematisch, wenn die Abstände zwischen den Changepoints stark variieren. So eignen sich große Bandbreiten zur Detektion kleiner Änderungen und kleine Bandbreiten zur Erkennung großer Änderungen. Eine Lösungsmöglichkeit wurde kürzlich im Zusammenhang mit Änderungen in Punktprozessen von Messer et al. (2014) vorgeschlagen. Ein Multiskalenverfahren basierend auf MOSUM-Statistiken soll dementsprechend konstruiert und untersucht werden. Da es für dieses Verfahren bisher noch keinerlei theoretische Untersuchungen gibt, wollen wir hier zunächst bei dem einfachen Erwartungswert-Modell bleiben.

---

**Projektleitung:** Prof. Dr. Claudia Kirch  
**Projektbearbeitung:** Dr. Marco Meyer, M.Sc. Philipp Klein  
**Kooperationen:** Prof. Dr. Claudia Redenbach, TU Kaiserslautern; Prof. Dr. Evgeny Spondarev, Universität Ulm; Dr. Katja Schladitz, Fraunhofer ITWM; Sowie diversen Industriepartnern  
**Förderer:** Bund - 01.05.2020 - 30.04.2023

### **Detektion von Anomalien in großen räumlichen Bilddaten**

Ziel des Forschungsvorhabens ist es, Methoden für die Detektion von Abweichungen/Auffälligkeiten in großen Bilddatenmengen zu entwickeln. Bei diesen Abweichungen kann es sich z.B. um Mikrorisse in Betonträgern, Materialverdichtungen in textiler Bahnware oder lokale Faserfehlorientierungen in Bauteilen aus faserverstärktem Kunststoff handeln. Dazu sollen Methoden des maschinellen Lernens, Modellierung der Strukturen und der Bildgebung sowie statistische Methoden für die Detektion von Auffälligkeiten kombiniert werden. Hierbei sollen insbesondere asymptotische Methoden aus der Changepoint-Analyse verallgemeinert werden, um Anomalien in Zufallsfeldern erkennen zu können.

---

**Projektleitung:** Prof. Dr. Claudia Kirch  
**Projektbearbeitung:** M.Sc. Felix Gnettner  
**Kooperationen:** Prof. Dr. Renate Meyer, University of Auckland, New Zealand; Dr. Patricio Matu-rana Russel, Auckland University, New Zealand  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.01.2019 - 31.03.2022

### **Bayessche nichtparametrische Zeitreihenanalyse für lokal-stationäre Zeitreihen**

In den letzten Jahren haben nichtparametrische Bayessche Verfahren stark an Aufmerksamkeit und Bedeutung gewonnen. Dennoch sind nur wenige Ansätze für die Zeitreihenanalyse entwickelt worden. Eine zusätzliche Schwierigkeit besteht darin, dass Bayessche statistische Verfahren der vollständigen Spezifikation einer Likelihood-Funktion bedürfen, was einer nichtparametrischen Herangehensweise zunächst entgegen steht. Mehrere Autoren haben das Problem mit Hilfe der Whittle-Likelihood gelöst, einer Approximation der wahren Likelihood, die von der Spektraldichte als der wichtigsten nichtparametrischen Kenngröße von Zeitreihen abhängt.

Moderne nichtparametrische Bootstrap-Verfahren für Zeitreihen setzen sich mit den gleichen Schwierigkeiten auseinander und verwenden implizit ebenfalls Approximationen der wahren Likelihood-Funktion. In diesem Projekt werden wir für die Bayessche nichtparametrische Analyse Approximationen moderner Resampling-Verfahren für lokal-stationäre Zeitreihen, d.h. Zeitreihen mit sich langsam ändernder Abhängigkeitsstruktur, die zwar nicht global wohl aber in einer Umgebung jeden Punktes approximativ stationär sind.

Hierzu definieren und analysieren wir eine neue Likelihood-Approximation für lokal stationäre Zeitreihen, die auf gleitenden lokalen Fourier-Koeffizienten basiert, deren globale statistische Eigenschaften denen von globalen Fourier-Koeffizienten im stationären Fall ähneln.

---

**Projektleitung:** Prof. Dr. Claudia Kirch  
**Projektbearbeitung:** M.Sc. Philipp Klein  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.10.2018 - 30.09.2021

### **Ein Verfahren zur Erkennung multipler Strukturbrüche in Erneuerungsprozessen**

Die Erkennung von Strukturbrüchen spielt für die Analyse von stochastischen Punktprozessen eine wichtige Rolle. Allerdings gibt es nur wenige Verfahren zur Erkennung und Lokalisierung von Strukturbrüchen.

Eine Möglichkeit hierfür ist, MOSUM-Teststatistiken zu verwenden. MOSUM-Teststatistiken eignen sich in der Regel sehr gut zur Erkennung von Strukturbrüchen, besitzen aber das Problem der geeigneten Wahl der Bandweite, da die Art der detektierten Strukturbrüche ganz wesentlich von der Bandweite abhängt. Messer et al. (2014) haben für Erneuerungsprozesse ein Verfahren entwickelt, welches Strukturbrüche mithilfe von verschiedenen (symmetrischen) Bandweiten detektiert. Dabei wird ein MOSUM-basiertes Verfahren verwendet, um die Strukturbrüche bei für eine fixe Bandweite zu detektieren. Anschließend werden die Strukturbrüche mithilfe eines Bottom-Up-Algorithmus zusammengefasst.

Eine ganz wesentliche Fragestellung hierbei ist die Qualität der Teststatistiken und Schätzer. Wir wollen dabei in diesem Projekt insbesondere Aussagen über die Konsistenz der Strukturbruchschätzer zu treffen und Aussagen über die Größenordnung der Abweichungen zu den "wahren" Strukturbrüchen treffen.

Darüber hinaus geht es darum, das Verfahren auf verschiedene Situationen z. B. die Verwendung asymmetrischer Bandweiten oder Bandweiten kleinerer Größenordnungen zu erweitern und ebenfalls Konsistenzaussagen für die Schätzer zu treffen.

Außerdem sollen die Verfahren auf reale Daten, wie z. B. neuronale Spike-Trains angewandt werden.

---

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** Dr. Maryna Prus  
**Kooperationen:** Dr. Norbert Benda, Bundesinstitut für Arzneimittel und Medizinprodukte; Prof. Radoslav Harman, Comenius-Universität, Bratislava; Prof. Norbert Gaffke, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik; Prof. Luc Pronzato, Université de Nice, Sophia Antipolis; Dr. Heiko Großmann, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik; Prof. Dr. Hans-Peter Piepho, Institute of Crop Science, Biostatistics, Faculty of Agricultural Sciences, University of Hohenheim  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 16.02.2019 - 15.10.2021

### **Generierung optimaler und effizienter Experimentaldesigns zur individualisierten Vorhersage in hierarchischen Modellen (II)**

Das Ziel des vorliegenden Projektes ist die Entwicklung analytischer Ansätze zur Gewinnung optimaler Designs für die Vorhersage in hierarchischen linearen Modellen sowie in verallgemeinerten linearen und nichtlinearen gemischten Modellen mit zufälligen Parametern. Derartige Modelle wurden ursprünglich in den Bio- und Agrarwissenschaften entwickelt und werden heutzutage in den unterschiedlichsten statistischen Anwendungsgebieten vielfältig eingesetzt.

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** Dipl.-Math. Martin Radloff  
**Förderer:** Sonstige - 01.10.2019 - 31.03.2021

### **Optimales Design für Sphärische Versuchsbereiche (II)**

Die Gültigkeit statischer Modelle ist oft auf einen lokalen Bereich der erklärenden Variablen beschränkt. Dieser wird in vielen Anwendungsbereichen als rechteckig angenommen, d.h. die erklärenden Variablen können unabhängig voneinander variieren. In manchen Situationen sind jedoch sphärische Bereiche sinnvoller, die durch einen beschränkten Euklidischen oder Mahalanobis-Abstand zu einem zentralen Punkt für die Versuchseinstellungen beschrieben werden können.

Ziel der Versuchsplanung ist es, optimale oder zumindest effiziente Einstellungen für die erklärenden Variablen zu bestimmen, um die Qualität der statistischen Analyse zu optimieren. Beim Vorliegen klassischer linearer Regressionsmodelle sind Charakterisierungen optimaler Designs für sphärische Versuchsbereiche mit Hilfe von Invarianzen und Symmetrien schon seit längerem bekannt. Fragestellung dieses Projekts ist es, für die in der statistischen Praxis zunehmend verwendeten verallgemeinerten linearen Modelle bzw. nichtlinearen Modelle optimale Designs auf derartigen sphärischen Versuchsbereichen zu bestimmen. Erste Ergebnisse für Poisson-verteilte Zählraten zeigen deutliche Abweichungen der hierfür benötigten optimalen Designs von denjenigen für klassische lineare Modelle.

---

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** MSc Arnab Sarkar  
**Kooperationen:** Priv.-Doz. Dr. Ekkehard Glimm, Novartis Pharma AG, Basel  
**Förderer:** Sonstige - 01.04.2018 - 31.03.2022

### **Analyse rekurrenter Ereignisprozesse mit einem terminalen Ereignis (informative Zensierung) - Überlegungen zum Studiendesign**

Das Konzept rekurrenter Ereignisse bezieht das wiederholte zeitliche Auftreten von Ereignissen ein und derselben Art im Kontext klinischer Studien ein. Beispiele umfassen das Auftreten von Anfällen in Epilepsiestudien, Aufflammen in Gichtstudien oder Hospitalisierung bei Patienten mit chronischen Herzleiden.

Eine wichtige Herausforderung bei der Analyse rekurrenter Ereignisse tritt auf, wenn informative Zensierung vorliegt. In klinischen Studien können beispielsweise Patienten aus einer Behandlung ausscheiden, weil sich ihre Verfassung so verschlechtert hat, dass eine alternative Behandlung notwendig wird. In dieser Situation kann die reine Tatsache, dass ein Patient ausscheidet, anzeigen, dass das interessierende Ereignis voraussichtlich eher oder häufiger auftritt, als unter der Annahme unabhängiger Zensierung zu erwarten wäre. Informative Zensierung kann dabei auch in Kombination mit einem terminalen Ereignis auftreten, das den rekurrenten Ereignisprozess beendet. Zum Beispiel kann in einer Studie zu chronischen Herzerkrankungen das Eintreten des Todes den Prozess der Hospitalisierung abbrechen. Da die Einflussfaktoren für Hospitalisierung bei Herzerkrankungen mit den Risikofaktoren für das Eintreten des Todes einhergehen, darf dieser Zusammenhang nicht vernachlässigt werden, da die resultierende Datenanalyse andernfalls verfälscht werden kann.

Zur Planung von Studien zur Aufdeckung und Bestimmung von Behandlungseffekten bei derartigen Endpunkten gibt es eine Reihe von Erweiterungen klassischer Überlebenszeitmodelle. Von besonderem Interesse ist dabei das Modell gemeinsamer Schwächung mit korrelierten Schwächungen, wobei separate marginale Modelle für die Intensität der beiden Ereignisprozesse unter Berücksichtigung korrelierter zufälliger Effekte, die subjektspezifische Schwächungen untersucht werden können.

Dieses Projekt umfasst sowohl methodologische Aspekte als auch Simulationsstudien und die Analyse realer Daten.

---

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** MSc Helmi Shat  
**Kooperationen:** Prof. Norbert Gaffke, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik  
**Förderer:** Deutscher Akademischer Austauschdienst e.V. (DAAD) - 01.10.2017 - 30.09.2021

### **Optimale Planung multi-variabler Accelerated-Degradation-Tests**

Die rasante Entwicklung moderner Fertigungstechniken zusammen mit den Bedürfnissen der Verbraucher nach hochqualitativen Produkten dienen als Motivation für Industrieunternehmen, Produkte zu entwickeln und herzustellen, die ohne Ausfall über Jahre oder gar Jahrzehnte funktionieren können. Für derartig langlebige Produkte ist es jedoch eine nicht einfache Aufgabe, innerhalb kurzer verfügbarer Zeit Zuverlässigkeitsaussagen zu treffen, da nicht genügend Daten für eine akkurate Schätzung der Lebensdauer gewonnen werden können. Dementsprechend ist eine Lebensdauerprüfung unter Normalbedingungen nicht sinnvoll. Daher werden Ermüdungstests mit wiederholte Messungen ("repeated measures accelerated degradation tests") häufig in der produzierenden Industrie angewendet, um Lebensdauerverteilungen hochzuverlässiger Produkte zu bestimmen, die bei traditionellen oder beschleunigten Lebensdauerests nicht ausfallen würden. In diesen Experimenten werden Beobachtungen bei hohen Belastungsstufen (z.B. Temperatur, Stromspannung oder Druck) mit Hilfe eines physikalisch sinnvollen statistischen Modells extrapoliert, um Schätzungen der Lebensdauer für niedrigere Belastungen unter Normalbedingungen zu erhalten. Zusätzlich ist zu beachten, dass verschiedene Faktoren wie die Häufigkeit der Messungen, die Stichprobengrößen und die Dauer des Experiments Einfluss auf die Kosten und die Genauigkeit der Schätzung haben.

Im Rahmen dieses Projektes werden zuerst adäquate und relevante Computerexperimente identifiziert und robuste Methoden der Regressionsanalyse entwickelt. Danach werden Optimalitätskriterien für experimentelle Designs definiert, die auf der Qualität der ausgewählten robusten Methoden basieren, und Simulationsbasierte Designs werden entwickelt, um einen einheitlichen Zugang zur Generierung optimaler oder zumindest effizienter Designs für die robuste Analyse in Computerexperimenten zu erhalten.

---

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** Torsten Reuter  
**Kooperationen:** Prof. Dr. Alexandra Carpentier, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.04.2020 - 31.03.2023

### **Optimales Sampling Design für Big Data**

Dank moderner Informationstechnologie besteht heutzutage die Möglichkeit, riesige Datenmengen zu sammeln, die sowohl im Hinblick auf die Anzahl der Beobachtungseinheiten (Umfang des Datensatzes) als auch hinsichtlich der Anzahl der Merkmale (multivariate Beobachtungen) von immenser Dimension sind und die häufig als massive Daten oder "Big Data" bezeichnet werden. Die reine Verfügbarkeit derartiger Big Data führt jedoch nicht zwangsläufig zu neuen Erkenntnissen über kausale Zusammenhänge innerhalb der Daten. Stattdessen kann die schiere Masse an Daten ernsthafte Probleme bei der statistischen Analyse verursachen. Zudem sind in vielen Situationen Teile (gewisse Merkmale) in den Daten einfach oder kostengünstig zu beobachten, während die Ausprägungen anderer, besonders interessierender Merkmale nur schwierig oder mit großen Kosten zu erhalten sind. Daher sind Vorhersagen für die Ausprägungen kostenintensiver Merkmale wünschenswert. Dieses kann mit klassischen statistischen Methoden erreicht werden, wenn für eine geeignete Teilstichprobe sowohl die Ausprägungen für die einfach als auch für die schwierig zu beobachtenden Merkmale verfügbar sind. Um Kosten zu reduzieren und/oder die Genauigkeit der Vorhersage zu verbessern, besteht ein Bedarf an optimalen Auswahlverfahren für Stichproben. In diesem Kontext können Konzepte aus der ursprünglich für technische Experimente entworfenen Theorie optimaler Designs unkonventionell genutzt werden, um effiziente Strategien für die Stichprobenauswahl zu entwickeln. Grundlegende Konzepte wie Relaxation auf stetige Verteilungen der Daten und Symmetrieeigenschaften können dabei zu einer wesentlichen Reduktion der Komplexität und somit zu praktikablen Lösungen führen. Ziel des vorliegenden Projekts ist es, diese allgemeinen Ideen zu konkretisieren und sie auf ein solides theoretisches Fundament zu stellen, um sie somit für die Auswertung realer Datensätze verwertbar zu machen.

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** Parisa Parsamaram  
**Kooperationen:** Prof. Dr. Heinz Holling, Universität Münster, Institut für Psychologie IV; Prof. Norbert Gaffke, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik  
**Förderer:** Land (Sachsen-Anhalt) - 01.07.2019 - 30.06.2022

### **Quasi-Likelihood und Quasi-Information für nicht-lineare und verallgemeinert lineare gemischte Modelle**

Nicht-lineare und verallgemeinert lineare gemischte Modelle werden effizient in der statistischen Datenanalyse in einem weiten Feld von Anwendungen in Bio- oder Sozialwissenschaften eingesetzt, wenn die grundlegenden Annahmen eines üblicherweise angesetzten linearen Modells nicht erfüllt sind. Derartige Situationen treten dann auf, wenn die Daten entweder aus einem intrinsisch nicht-linearen Zusammenhang stammen wie beispielsweise in der Pharmakokinetik, bei Wachstums- und Dosis-Wirkungs-Kurvens oder die Zielvariable auf einer nicht-metrischen Skala gemessen wird wie beispielsweise Zählraten und nominale oder ordinale Antworten. Zusätzlich treten gemischte Effekte auf, wenn Messwiederholungen an ein und denselben statistischen Einheiten beobachtet werden. Dies führt zu einer Verletzung der üblichen Annahme statistisch unabhängiger Beobachtungen. Die Nicht-linearität in Kombination mit der Modellierung mit gemischten Effekten macht eine explizite Berechnung der Likelihood und damit der Fisher-Information unmöglich. Als Ersatz kann die Quasi-Likelihood und die daraus resultierende Quasi-Information genutzt werden, die einfacher zu bestimmen sind und zu ausrechenbaren Schätzungen und deren Unsicherheitsquantifizierung führen. Dieser Ansatz erlaubt zudem die Konstruktion zuverlässiger Experimentaldesigns, die die Qualität der durchzuführenden Experimente im Vorhinein optimiert. In diesem Sinne vereinfacht dieser Ansatz die Komplexität des vorliegenden Schätz- und Planungsproblems und kann einfach mit anderen, häufig in der Statistik verwendeten Reduktionsprinzipien wie Invarianz und Äquivarianz kombiniert werden. Ziel des vorliegenden Projekts ist es, handhabbare Lösungen für die zuvor beschriebene Problemstellung zu entwickeln und diese in praktischen Situationen umzusetzen.

---

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** MSc Frank Röttger  
**Kooperationen:** Prof. Dr. Thomas Kahle, FMA-IAG  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 01.04.2017 - 31.03.2020

### **Geometrie optimaler Designs für nichtlineare Modelle in der Statistik**

Geometrische Beschreibungen optimaler Designbereiche sind in Zeiten zunehmender Komplexität statistischer Modelle von wachsendem Interesse. Das Ziel dieses Projektes besteht in der Suche von Optimalitätsbereichen von experimentellen Designs für derartige statistische Modelle, insbesondere für verallgemeinerte lineare Modelle mit Poisson- oder logistisch verteilten Zielvariablen. Diese Bereiche können durch Systeme von polynomialen Ungleichungen im Parameterraum beschrieben werden, was bedeutet, dass sie nichts anderes als semialgebraische Mengen sind. Somit können Methoden der algebraischen Geometrie benutzt werden, um die Eigenschaften dieser Optimalitätsbereiche zu studieren. Als Beispiel können im Paarvergleichsmodell nach Bradley-Terry, das ein statistisches Modell für den Vergleich verschiedener Alternativen auf der Basis logistischen Antwortverhaltens ist, die Optimalitätsbereiche für sogenannte saturierte Designs, d.h. Designs mit einer minimalen Anzahl von Trägerpunkten, bestimmt werden.



**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** M.Sc. Osama Idais  
**Kooperationen:** Prof. Norbert Gaffke, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik  
**Förderer:** Sonstige - 01.10.2019 - 31.03.2020

### Optimales Design für multivariate verallgemeinerte lineare Modelle mit stetigen Zielfunktionen (II)

In vielen Anwendungssituationen, in denen Daten gesammelt werden, werden nicht nur eine einzelne, sondern mehrere Zielvariablen gleichzeitig beobachtet, die miteinander korreliert sein können. Derartige multivariate Beobachtungen werden oft mit einer multivariaten Normalverteilung modelliert. In einigen Situationen ist dies jedoch nicht angebracht, insbesondere wenn die beobachteten Merkmale nicht stetig sind. Für diese Situationen ist das Konzept der verallgemeinerten linearen Modelle entwickelt worden, die sich speziell bei binären Daten (z.B. logistische Regression) oder Zählraten (z.B. Poisson-Regression) bewährt haben. Jedoch kann auch bei stetigen Merkmalen statt der Normalverteilungsannahme eine andere Verteilungsannahme angemessener sein, die sich über ein verallgemeinertes lineares Modell mit nichtlinearer Linkfunktion beschreiben lässt. Ziel des Projektes ist es, für derartige Modelle asymptotische Eigenschaften unter verschiedenen Korrelationsstrukturen zu bestimmen und auf dieser Basis optimale Designs zu generieren, die zu einer Verbesserung der Datenanalyse führen.

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** Dipl.-Math. Marius Schmidt  
**Kooperationen:** Priv.-Doz. Dr. Steffen Uhlig, Quo Data, Dresden; Dr. Tobias Mielke, Aptiv Solutions, Köln; Dr. Thomas Schmelter, Bayer, Berlin; Dr. Hermann Kulmann, Bayer, Berlin; Prof. Dr. Heinz Holling, Universität Münster, Institut für Psychologie IV  
**Förderer:** Sonstige - 01.10.2019 - 31.01.2020

### Optimales Design für verallgemeinerte lineare gemischte Modelle (II)

Gemischte Modelle spielen zunehmend eine wichtige Rolle nicht nur in Biowissenschaften sondern auch bei wirtschafts- und sozialwissenschaftlichen Fragestellungen, um individuelle Effekte der verschiedenen Beobachtungseinheiten als Repräsentanten einer größeren Grundgesamtheit bei der statistischen Datenanalyse berücksichtigen und erfassen zu können. Mit verallgemeinerten linearen gemischten Modellen werden Zusammenhänge für binäre ("Erfolg - Misserfolg") und diskrete Zielgrößen ("Anzahlen") beschrieben, die nicht sinnvoll durch standardmäßige lineare gemischte Modelle für metrische Daten dargestellt werden können. Für die zufälligen Effekte können dann neben normalverteilten individuellen Einflüssen auch solche aus konjugierten Familien angenommen werden, die eine explizitere Analyse erlauben. Wie in allen statistischen Analysen hängt auch hier die Qualität der Ergebnisse wesentlich vom Beobachtungs- oder Experimentaldesign, d.h. der Wahl der Beobachtungseinheiten und Beobachtungszeitpunkte, ab. Ziel dieses Projektes ist es, optimale oder zumindest effiziente Designs für verallgemeinerte lineare gemischte Modelle zu entwickeln, die sowohl normalverteilte als auch Effekte aus konjugierten Verteilungen beinhalten können, und diese zu validieren.

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** Prof. Dr. Norbert Gaffke  
**Kooperationen:** Dr. Fritjof Freise, TU Dortmund  
**Förderer:** Haushalt - 01.01.2019 - 31.03.2021

### Sequenziell-adaptives Design

Nicht-lineare Regression spielt eine wichtige Rolle zur adäquaten statistischen Modellierung von Daten, wenn der Einfluss erklärender Variablen auf die interessierende Zielvariable nicht durch einen einfachen linearen Ursache-Wirkungs-Zusammenhang beschrieben werden kann. In derartigen Modellen hängt die Informationsmatrix eines Versuchsplans (Design) vom Parametervektor ab, dessen wahrer Wert unbekannt ist. Häufig verwendete Ansätze der optimalen Versuchsplanung in dieser Situation sind lokal optimale Designs, Bayes-optimale Designs oder auch Minimax-Designs. Diese Konzepte benötigen und verwenden jedoch a-priori Kenntnisse über den wahren

Parameterwert. Sequenziell-adaptive Designs hingegen sind lernende Verfahren. Sie sammeln Informationen über den wahren Parameterwert aus bereits gemachten Beobachtungen in einem sequenziellen Prozess und können daher auf a-priori Informationen verzichten. Dabei werden sequenziell adaptive Updates der Parameterschätzung auf Basis der bereits gemachten Beobachtungen berechnet, und mit Hilfedieser wird das Design entsprechend um weitere Beobachtungen ergänzt. Ein populärer Algorithmus dieser Art ist der adaptive Wynn-Algorithmus zur asymptotischen Generierung eines D-optimalen Designs. In der gemeinsamen Arbeit von Freise, Gaffke und Schwabe (2019a) ist es gelungen, das seit Langem offene Problem der Konvergenz dieses Algorithmus zumindest für die in den Anwendungen wichtige Klasse der verallgemeinerten linearen Modelle (positiv) zu lösen. In der zweiten Arbeit von Freise, Gaffke und Schwabe (2019b) konnte dies auch auf eine weitere Klasse von nicht-linearen Modellen und auf andere Schätzverfahren erweitert werden. Gegenwärtig arbeiten die Autoren an der Analyse eines neuen Algorithmus zur asymptotischen Generierung D-optimaler Designs, bei dem gleichzeitig mehrere Beobachtungen hinzugefügt werden. Weitere Ziele des Projekts sind zum einen die Ausweitung der Untersuchungen auf weitere Klassen nicht-linearer Modelle sowie auf weitere Optimalitätskriterien. Zum anderen soll das praktische Konvergenzverhalten der Algorithmen erprobt und beurteilt werden.

Freise, F.; Gaffke, N.; Schwabe, R. (2019a). The adaptive Wynn-algorithm in generalized linear models with univariate response. Preprint arXiv:1907.02708

Freise, F.; Gaffke, N.; Schwabe, R. (2019b). Convergence of least squares estimators in the adaptive Wynn algorithm for a class of nonlinear regression models. Preprint. arXiv:1909.03763

---

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** Dr. Osama Idais  
**Kooperationen:** Prof. Norbert Gaffke, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik  
**Förderer:** Haushalt - 01.10.2020 - 31.03.2021

### **Äquivarianz und IMSE-Optimalität für Designs in verallgemeinert linearen Modellen mit stetigen Zielfunktionen (II)**

In vielen Anwendungssituationen, in denen Daten gesammelt werden, ist die Annahme der Normalverteilung nicht angebracht, insbesondere wenn die beobachteten Merkmale nicht stetig sind. Für diese Situationen ist das Konzept der verallgemeinerten linearen Modelle entwickelt worden, die sich speziell bei binären Daten (z.B. logistische Regression) oder Zähldaten (z.B. Poisson-Regression) bewährt haben. Jedoch kann auch bei stetigen Merkmalen statt der Normalverteilungsannahme eine andere Verteilungsannahme angemessener sein, die sich über ein verallgemeinertes lineares Modell mit nichtlinearer Linkfunktion beschreiben lässt. Ziel des Projektes ist es, für derartige Modelle unter Verwendung von Symmetrieeigenschaften dieser Modelle optimale Designs zu generieren, die zu einer Verbesserung der Datenanalyse führen.

---

**Projektleitung:** Prof. Dr. Rainer Schwabe  
**Projektbearbeitung:** Dr. Osama Idais  
**Kooperationen:** Prof. Norbert Gaffke, Otto-von-Guericke-Universität Magdeburg, Institut für Mathematische Stochastik  
**Förderer:** Sonstige - 01.04.2020 - 30.09.2020

### **Äquivarianz und IMSE-Optimalität für Designs in verallgemeinert linearen Modellen mit stetigen Zielfunktionen**

In vielen Anwendungssituationen, in denen Daten gesammelt werden, ist die Annahme der Normalverteilung nicht angebracht, insbesondere wenn die beobachteten Merkmale nicht stetig sind. Für diese Situationen ist das Konzept der verallgemeinerten linearen Modelle entwickelt worden, die sich speziell bei binären Daten (z.B. logistische Regression) oder Zähldaten (z.B. Poisson-Regression) bewährt haben. Jedoch kann auch bei stetigen Merkmalen statt der Normalverteilungsannahme eine andere Verteilungsannahme angemessener sein, die sich über ein verallgemeinertes lineares Modell mit nichtlinearer Linkfunktion beschreiben lässt. Ziel des Projektes ist

es, für derartige Modelle unter Verwendung von Symmetrieeigenschaften dieser Modelle optimale Designs zu generieren, die zu einer Verbesserung der Datenanalyse führen.

---

**Projektleitung:** Doz. Dr. Martin Wendler  
**Kooperationen:** Annika Betken, Ruhr-Universität Bochum  
**Förderer:** Haushalt - 01.01.2018 - 30.06.2020

### **Rank based change-point analysis under long range dependence**

The aim of the project is to develop robust change-point test for long range dependent time series which have a good efficiency under normality. For this, the empirical process of ranks is studied in a function space equipped with a wheighted norm.

---

**Projektleitung:** Doz. Dr. Martin Wendler  
**Projektbearbeitung:** M.Sc. Lea Wegner  
**Förderer:** Deutsche Forschungsgemeinschaft (DFG) - 15.09.2019 - 15.11.2022

### **Analysis of Functional Data without Dimension Reduction: Tests for Covariance Operators and Changepoint Problems**

Functional data arises in many applications and the main strategy for statistical inference is dimension reduction: The data is projected on a finite-dimensional space with techniques such as functional principal components. After this, it is possible to use statistical test for finite-dimensional data. In contrast, there are recent proposals to base the statistical tests on the full functional information, typically modeld as Hilbert-space-valued time series. These methods have been investigated in the context of sample means and simple changepoints. The aim of this project is to develop fully functional methods in more complicated data situations: We will investigate test for hypothesis not on the functional mean, but on the covariance operator. Furthermore, we plan to develop test for changepoints in data including extreme outliers, which might lead to false negatives and false positive results of standard methods. The last part will deal with segmentation of functional time series or detection of multiple changepoints. To get critical values, we will extend nonparametric methods like bootstrap to these challenging data situations.

---

**Projektleitung:** Dr. Heiko Großmann  
**Projektbearbeitung:** Dr. Elizabeth Cottrell  
**Kooperationen:** Keele University, UK  
**Förderer:** Sonstige - 01.04.2019 - 31.03.2021

### **Explaining osteoarthritis: development and implementation of a multimedia Patient Explanation Package (PEP-OA)**

Grant number: NIHRDH-PB-PG-0817-20031. Osteoarthritis (OA) is a common, debilitating and painful condition, particularly when patients move the affected joint. Core-management approaches (exercise and weight control) reduce pain and improve function, but exercise-induced pain creates anxiety and confusion about such self-management. Common, unhelpful, misconceptions about OA exist and currently professionals do not have the language to explain OA in a way that reflects current scientific understanding. The overarching aim of the project is to improve OA explanations through the development and implementation of a multimedia Patient Explanation Package (PEP-OA). A partial-profile conjoint analysis study with patients will estimate the extent to which new, prioritised, explanation statements are preferred over currently used/available statements. Suitable OA explanations identified in this study will be used in the further development of the multimedia package. The corresponding work package requires the development of an efficient experimental design for the choice experiment which will be carried out at the University of Magdeburg.

## **7. EIGENE KONGRESSE, WISSENSCHAFTLICHE TAGUNGEN UND EXPONATE AUF MESSEN**

Maryna Prus, Rainer Schwabe: Organisation des Workshops "PODE20", Magdeburg, 15.06.2020. **Leider musste der Workshop wegen der Coronapandemie ausfallen.**

## 8. VERÖFFENTLICHUNGEN

### BEGUTACHTETE ZEITSCHRIFTENAUFsätze

**Braun, Markus; Schneider, Thomas**

Ein generelles Verbot würde dazu führen, dass alle Diesel illegal wären  
Motortechnische Zeitschrift: MTZ ; die technisch-wissenschaftliche Fachzeitschrift für Verbrennungsmotor und Gasturbine - Wiesbaden: Vieweg, 1998, Bd. 81.2020, S. 26-29;

**Cho, Haeran; Kirch, Claudia**

Discussion of Detecting possibly frequent change-points - wild binary segmentation 2 and steepest-drop model selection

Journal of the Korean Statistical Society / Han'guk-T'onggye-Hakhoe - Singapore : Springer Singapore, insges. 5 S., 2020 ;

[Online first]

[Imp.fact.: 0.556]

**Christoph, Gerd; Monakhov, M. M.; Ulyanov, V. V.**

Second-order ChebyshevEdgeworth and CornishFisher expansions for distributions of statistics constructed from samples with random sizes

Journal of mathematical sciences - New York, NY: Consultants Bureau, 1973, Bd. 244.2020, 5, S. 811-839;

**Christoph, Gerd; Ulyanov, Vladimir V.**

Second order expansions for high-dimension low-sample-size data statistics in random setting

Mathematics: open access journal - Basel: MDPI, 2013, Volume 8 (2020), issue 7, article 1151, 28 Seiten;

[Imp.fact.: 1.747]

**Dehling, H.; Fried, R.; Wendler, Martin**

A robust method for shift detection in time series

Biometrika: a journal for the statistical study of biological problems - London: Biometrika Trust, 1901, Bd. 107.2020, 1, S. 647-660;

[Imp.fact.: 1.632]

**Eckley, Idris; Kirch, Claudia; Weber, Silke**

A novel change-point approach for the detection of gas emission sources using remotely contained concentration data

The annals of applied statistics : an official journal of the Institute of Mathematical Statistics - Beachwood, Ohio : Inst. of Mathematical Statistics (IMS), Bd. 14.2020, 3, S. 1258-1284

[Imp.fact.: 1.675]

**Ghoshdastidar, Debarghya; Gutzeit, Maurilio; Carpentier, Alexandra; Luxburg, Ulrike**

Two-sample hypothesis testing for inhomogeneous random graphs

The annals of statistics : an official journal of the Institute of Mathematical Statistics - Hayward, Calif. : IMS Business Off., Bd. 48.2020, 4, S. 2208-2229

**Graßhoff, Ulrike; Holling, Heinz; Röttger, Frank; Schwabe, Rainer**

Optimality regions for designs in multiple linear regression models with correlated random coefficients

Journal of statistical planning and inference: JSPI - Amsterdam: North-Holland Publ. Co., 1977, Bd. 209.2020, S. 267-279;

[Imp.fact.: 0.756]

**Graßhoff, Ulrike; Holling, Heinz; Schwabe, Rainer**

Optimal design for the Rasch counts model with multiple binary predictors

The British journal of mathematical and statistical psychology - Hoboken, NJ [u.a.]: Wiley, 1965, Bd. 73.2020, 3, S. 541-555;

[Imp.fact.: 2.388]

**Großmann, Heiko**

On the meaning of block effects in paired comparison choice experiments and a relationship with blocked  $2(K)$  main effects plans

Journal of statistical planning and inference: JSPI - Amsterdam: North-Holland Publ. Co., 1977, Bd. 209.2020, S. 76-84;

[Imp.fact.: 0.756]

**Großmann, Heiko; Weinbauer, Gerhard F.; Baker, Ann; Fuchs, Antje; Luetjens, C. Marc**

Enhanced normograms and pregnancy outcome analysis in nonhuman primate developmental toxicity studies

Reproductive toxicology - Amsterdam [u.a.]: Elsevier Science, 1987, Bd. 95.2020, S. 29-36;

[Imp.fact.: 3.2]

**Guillotini-Plantard, Nadine; Pène, Françoise; Wendler, Martin**

Empirical processes for recurrent and transient random walks in random scenery

Probability and statistics - Les Ulis: EDP Sciences, 1997, Bd. 24.2020, S. 127-137;

[Imp.fact.: 0.766]

**Idais, Osama**

A note on locally optimal designs for generalized linear models with restricted support

Statistics & probability letters - Amsterdam: Elsevier Science, 1982, Volume 159(2020), article 108692;

**Idais, Osama**

Locally optimal designs for multivariate generalized linear models

Journal of multivariate analysis: JMVA - Orlando, Fla.: Acad. Press, 1971, Volume 180(2020), article 104663;

[Imp.fact.: 1.029]

**Idais, Osama**

On local optimality of vertex type designs in generalized linear models

Statistical papers - Berlin: Springer, 1988 . - 2020;

[Online first]

[Imp.fact.: 1.433]

**Kahle, Thomas; Röttger, Frank; Schwabe, Rainer**

Geometrie optimaler Versuchspläne

Mitteilungen der Deutschen Mathematiker-Vereinigung/ Deutsche Mathematiker-Vereinigung - Berlin: DMV, 1993, Bd. 28.2020, 2, S. 71-76;

**Meier, Alexander; Kirch, Claudia; Meyer, Renate**

Bayesian nonparametric analysis of multivariate time series: A matrix Gamma Process approach

Journal of multivariate analysis - Orlando, Fla.: Acad. Press, 1971, Volume 175 (2020), article 104560;

[Imp.fact.: 1.029]

**Nyarko, Eric**

Optimal  $2 \times K$  paired comparison designs for third-order interactions

Statistical papers - Berlin: Springer, 1988 . - 2020;

[Online first]

[Imp.fact.: 1.345]

**Prus, Maryna; Benda, Norbert; Schwabe, Rainer**

Optimal design in hierarchical random effect models for individual prediction with application in precision medicine

Journal of statistical theory and practice - Cham: Springer International Publishing, 2007, Volume 14(2020), issue 2, article 24, 12 Seiten;

**Rakauskas, Alfredas; Wendler, Martin**

Convergence of U-processes in Hölder spaces with application to robust detection of a changed segment

Statistical papers - Berlin: Springer, 1988, Bd. 61.2020, 4, S. 1409-1435;

[Imp.fact.: 1.433]

**Rottengruber, Hermann**

Thermofenster lassen sich nicht schließen

Motortechnische Zeitschrift: MTZ - Wiesbaden: Springer Vieweg, Springer-Fachmedien Wiesbaden GmbH, 1939, Bd. 81.2020, 10, S. 74

**Sharipov, Olimjon Sh.; Wendler, Martin**

Bootstrapping covariance operators of functional time series

Journal of nonparametric statistics - Abingdon: Taylor & Francis, 1991 . - 2020;

[Online first]

[Imp.fact.: 0.706]

**Sofronov, Georgy; Wendler, Martin; Liebscher, Volkmar**

Editorial for the special issue: Change point detection

Statistical papers - Berlin: Springer, 1988, Bd. 61.2020, 4, S. 1347-1349;

[Imp.fact.: 1.433]

**Stoehr, Christina; Aston, John A. D.; Kirch, Claudia**

Detecting changes in the covariance structure of functional time series with application to fMRI data

Econometrics and statistics - Amsterdam [u.a.]: Elsevier B.V, 2017 . - 2020;

[Online first]

**Wendler, Martin; Wu, Wei Biao**

Central limit theorems for nearly long range dependent subordinated linear processes

Journal of applied probability - Cambridge: Cambridge University Press, 1964, Bd. 57.2020, 2, S. 637-656;

[Imp.fact.: 0.577]

**NICHT BEGUTACHTETE ZEITSCHRIFTENAUFsätze**

**Cheshire, James; Menard, Pierre; Carpentier, Alexandra**

The influence of shape constraints on the thresholding bandit problem

De.arxiv.org - [S.l.]: Arxiv.org - 2020, article 2006.10006

**Christoph, Gerd; Ulyanov, Vladimir V.; Bening, Vladimir E.**

Second order expansions for sample median with random sample size

Magdeburg: Otto-von-Guericke-Universität, Fakultät für Mathematik, 2020, 18 Seiten - (Preprint; Fakultät für Mathematik, Otto-von-Guericke-Universität Magdeburg; 2020, Nr. 01);

[Literaturangaben: Seite 16-18]

**Freise, Fritjof; Graßhoff, Ulrike; Röttger, Frank; Schwabe, Rainer**

D-optimal designs for Poisson regression with synergetic interaction effect

De.arxiv.org - [S.l.]: Arxiv.org, 1991, 2020, article 2006.04656, insgesamt 21 Seiten;

**Graßhoff, Ulrike; Großmann, Heiko; Holling, Heinz; Schwabe, Rainer**

Optimal design for probit choice models with dependent utilities

De.arxiv.org - [S.l.]: Arxiv.org, 1991, 2020, article 2001.09036, insgesamt 26 Seiten;

**Idais, Osama; Schwabe, Rainer**

In- and equivariance for optimal designs in generalized linear models - the gamma model

De.arxiv.org - [S.l.]: Arxiv.org - 2020, article 2011.09745, insgesamt 22 Seiten

**Manegueu, Anne Gael; Vernade, Claire; Carpentier, Alexandra; Valko, Michal**

Stochastic bandits with arm-dependent delays

De.arxiv.org - [S.l.]: Arxiv.org - 2020, article 2006.10459, insgesamt 19 Seiten

**Prus, Maryna**

Equivalence theorems for compound design problems with application in mixed models

De.arxiv.org - [S.l.]: Arxiv.org, 1991, 2020, article 2007.14971, 15 Seiten;

**Prus, Maryna; Piepho, Hans-Peter**

Optimizing the allocation of trials to sub-regions in multi-environment crop variety testing  
De.arxiv.org - [S.l.]: Arxiv.org, 1991, 2020, article 2004.05925, 12 Seiten;

**Vernade, Claire; Carpentier, Alexandra; Lattimore, Tor; Zappella, Giovanni; Ermis, Beyza; Brueckner, Michael**

Linear bandits with stochastic delayed feedback  
De.arxiv.org - [S.l.]: Arxiv.org - 2020, article 2011.09745, insgesamt 22 Seiten

**DISSERTATIONEN**

**Röttger, Frank; Kahle, Thomas [AkademischeR BetreuerIn]; Schwabe, Rainer [AkademischeR BetreuerIn]**

Geometry of optimal design and limit theorems  
Magdeburg, 2020, IX, 105 Seiten, Diagramme, 30 cm;  
[Literaturverzeichnis: Seite 99-105]

**Schmidt, Marius; Schwabe, Rainer [AkademischeR BetreuerIn]**

Optimale Versuchsplanung für Zähldaten mit zufälligen Blockeffekten  
Magdeburg, 2020, IV, 119 Seiten, Diagramme, 30 cm;  
[Literaturverzeichnis: Seite 116-119]