

Proceedings of Reisensburg 2010

**Harald Binder · Hans A. Kestler ·
Matthias Schmid**

Published online: 23 January 2013
© Springer-Verlag Berlin Heidelberg 2013

The conference “Statistical Computing” is held annually in Reisensburg (Germany) by the working group “Biostatistics” of the German Classification Society and by the working group “Statistical Computing” of the German Region of the International Biometric Society and of the German Society of Medical Informatics, Biometry and Epidemiology (GMDS). The conference covers recent topics in Biostatistics and Bioinformatics, with a special focus on applications regarding the computational aspects of these fields of research. In 2009, the conference organizers started publishing selected articles of the conference as a special issue of Computational Statistics (COST, see volume 26, issue 2, 2011). Being either research or tutorial papers on recent software developments, all articles featured in a Reisensburg special issue undergo the regular reviewing process of Computational Statistics.

This year’s special issue features the highlights of the 2010 Statistical Computing conference, which took place from June 20 to June 23, 2010 at Reisensburg Castle. Special topics of the conference were “Variable selection/dimension reduction”,

H. Binder (✉)

Division Biostatistics, Institute of Medical Biostatistics, Epidemiology and Informatics,
University Medical Center Johannes Gutenberg University Mainz,
Obere Zahlbacher Str. 69, 55131 Mainz, Germany
e-mail: binderh@uni-mainz.de

H. A. Kestler

Research group Bioinformatics and Systems Biology,
Institute of Neural Information Processing, University of Ulm, 89069 Ulm, Germany
e-mail: hans.kestler@uni-ulm.de

M. Schmid

Department of Medical Informatics, Biometry and Epidemiology,
Friedrich-Alexander-University Erlangen-Nuremberg, Waldstraße 6, 91054 Erlangen, Germany
e-mail: matthias.schmid@imbe.med.uni-erlangen.de

“Benchmarking” and “Systems biology/networks”. These topics form the basis of the 2010 Reisenburg special issue.

[Schels et al. \(2013\)](#), develop a multiple classifier approach to discover events in bioelectrical signals obtained from electroencephalography in neuroscience applications. The proposed approach is based on creating individual classifiers for pattern recognition that are later combined using decision fusion. Moreover, an alternative approach based on genetic algorithms is presented.

[Hopfensitz et al. \(2013\)](#) present a tutorial on Attractors in Boolean Networks, which are a popular class of models for the description of gene-regulatory networks. The tutorial presents strategies to identify attractors in Boolean networks using the BoolNet package of the statistical software environment R (cran.r-project.org).

[Reiser et al. \(2013\)](#) consider the problem of comparing groups of individuals with respect to gene expression in the presence of confounding variables. They investigate the use of boosting and matching strategies to conduct statistically meaningful comparisons of such groups. Specifically, the two strategies are compared with respect to their ability of constructing accurate risk prediction models.

While the paper of Reiser et al. deals with the construction of risk prediction models where a small set of genes with good prediction performance is to be selected, [Lausser et al. \(2013\)](#) focus on the stability of such signatures. It is well known that in particular gene expression signatures might be highly unstable, and therefore tools for assessing stability are needed. Lausser et al. introduce an index for quantifying stability and a corresponding test procedure. To complement this, they also suggest visualisation in form of a stability map, which can conveniently highlight potential problems in the signature development process.

On the same type of data the paper by [Telaar et al. \(2013\)](#) investigates the important problem of pooling RNA samples in biomarker classification studies. Pooling of RNA is an often-encountered necessity in the analysis of gene expression for reasons of cost and low extraction amount. Telaar et al. investigate a number of different classification algorithms in varying two class scenarios and give guidelines on the choice of procedures regarding biomarker search or classification performance.

Also motivated by the high dimensionality of biomedical signatures, [Unkel and Trendafilov \(2013\)](#) address the problem of adapting the exploratory factor analysis (EFA) model to data with more variables than cases. They describe a new algorithm zig-zag EFA, which is based on singular value decomposition. Their application to high-dimensional gene expression data obtains superior results for dimensionality reduction in terms of fit, making the procedure a potential competitor to PCA in these settings.

In the paper by [Glodek et al. \(2013\)](#) the idea of using ensembles of learners is transferred to a new domain. While ensemble learning has so far mostly been used for improving prediction of some response by combining several classifiers, Glodek et al. provide a promising approach for improved density estimation by ensembles of mixture models. This also addresses the difficulty of having to carefully select tuning parameters for a single mixture model, as other members of the ensemble can in principle correct for potential problems.

The paper by [Lueck et al. \(2013\)](#) illustrates an application where a wide set of computational and statistical tools are needed for adequately addressing a specific

data analysis problem. For analyzing scanning electron microscopy images, these images first have to be preprocessed and aggregated into a set of features to be accessible for further statistical analysis. Techniques from spatial statistics then allow to disentangle differences in morphology between cell types and to incorporate further domain knowledge.

References

- Glodek M, Schels M, Schwenker F (2013) Ensemble Gaussian mixture models for probability density estimation. *Comput Stat* 28. doi:[10.1007/s00180-012-0374-5](https://doi.org/10.1007/s00180-012-0374-5)
- Hopfensitz M, Müssel C, Maucher M, Kestler HA (2013) Attractors in Boolean networks—a tutorial. *Comput Stat* 28. doi:[10.1007/s00180-012-0324-2](https://doi.org/10.1007/s00180-012-0324-2)
- Lausser L, Müssel C, Maucher M, Kestler HA (2013) Measuring and visualizing the stability of biomarker selection techniques. *Comput Stat* 28. doi:[10.1007/s00180-011-0284-y](https://doi.org/10.1007/s00180-011-0284-y)
- Lueck S, Fichtl A, Sailer M, Joos H, Brenner RE, Walther P, Schmidt V (2013) Statistical analysis of the intermediate filament network in cells of mesenchymal lineage by greyvalue-oriented image segmentation. *Comput Stat* 28. doi:[10.1007/s00180-011-0265-1](https://doi.org/10.1007/s00180-011-0265-1)
- Reiser V, Porzelius C, Stampf S, Schumacher M, Binder H (2013) Can matching improve the performance of boosting for identifying important genes in observational studies? *Comput Stat* 28. doi:[10.1007/s00180-012-0306-4](https://doi.org/10.1007/s00180-012-0306-4)
- Schels M, Scherer S, Glodek M, Kestler HA, Palm G, Schwenker F (2013) On the discovery of events in EEG data utilizing information fusion. *Comput Stat* 28. doi:[10.1007/s00180-011-0292-y](https://doi.org/10.1007/s00180-011-0292-y)
- Telaar A, Repsilber D, Nürnberg D (2013) Biomarker discovery: classification using pooled samples—a simulation study. *Comput Stat* 28. doi:[10.1007/s00180-011-0302-0](https://doi.org/10.1007/s00180-011-0302-0)
- Unkel S, Trendafilov NT (2013) Zig-zag exploratory factor analysis with more variables than observations. *Comput Stat* 28. doi:[10.1007/s00180-011-0275-z](https://doi.org/10.1007/s00180-011-0275-z)