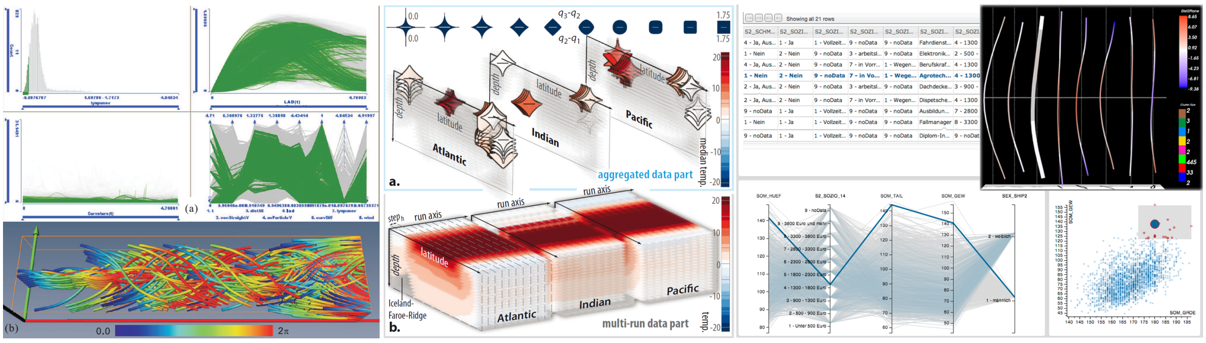
Interactive Visual Analysis of Scientific Data

Steffen Oeltze[1], Helwig Hauser[2], and Johannes Kehrer[3]

[1]Department of Simulation and Graphics, Otto-von-Guericke University of Magdeburg, Germany
[2]Department of Informatics, University of Bergen, Norway
[3]Institute of Computer Graphics and Algorithms, Vienna University of Technology, Austria



In a growing number of application areas, a subject or phenomenon is investigated by means of multiple datasets being acquired over time (spatiotemporal), comprising several attributes per data point (multi-variate), stemming from different data sources (multi-modal) or multiple simulation runs (multi-run/ensemble) [KH13]. Interactive visual analysis (IVA) comprises concepts and techniques for a user-guided knowledge discovery in such complex data. Through a tight feedback loop of computation, visualization and user interaction, it provides new insight into the data and serves as a vehicle for hypotheses generation or validation. It is often implemented via a multiple coordinated view framework where each view is equipped with interactive drill-down operations for focusing on data features. Two classes of views are integrated: *physical views*, such as direct volume rendering, show information in the context of the spatiotemporal observation space while *attribute views*, such as scatter plots and parallel coordinates, show relationships between multiple data attributes. The user may drill-down the data by selecting interesting regions of the observation space or attribute ranges leading to a consistent highlighting of this selection in all other views (*brushing-and-linking*). Three patterns of explorative/analytical procedures may be accomplished by doing so. In a *feature localization*, the user searches for places in the 3D/4D observation space where certain attribute values are present. In a *multi-variate analysis*, relations between data attributes are investigated, e.g., by searching for correlations. In a *local investigation*, the user inspects the values of selected attributes with respect to certain spatiotemporal subsets of the observation space.

In this tutorial, we discuss examples for successful applications of IVA to scientific data from various fields: climate research, medicine, epidemiology, and flow simulation / computation, in particular for automotive engineering. We base our discussions on a theoretical foundation of IVA which helps the tutorial attendees in transferring the subject matter to their own data and application area. In the course of the tutorial, the attendees will become acquainted with techniques from statistics and knowledge discovery, which proved to be particularly useful for a specific IVA application. The tutorial further comprises an overview of off-the-shelf IVA solutions, which may be be particularly interesting for visualization practitioners. It is concluded by a summary of the gained knowledge and a discussion of open problems in IVA of scientific data.

Organizer and Contact Address

Steffen Oeltze
Otto-von-Guericke University of Magdeburg
Department of Simulation and Graphics
P.O. Box 4120
D-39016 Magdeburg, Germany

e-mail: stoeltze@isg.cs.uni-magdeburg.de
www: www.vismd.de

Phone: +49 (0) 391 67 52527

Tutorial Title

Interactive Visual Analysis of Scientific Data

Tutorial Length

Half-day tutorial

Why is it worthwhile?

A review of all IEEE VisWeek submissions from 2010 to 2012 showed that many of them describe software or approaches which combine several views on the investigated data. In various cases, in particular for VAST and BioVis submissions, these views are also linked and equipped with brushing facilities. From application papers it becomes obvious that different datasets showing the same subject or phenomenon need to be concurrently processed employing a set of linked views. Three workshops and one tutorial dedicated to visual analytics and interactive visual analysis were held during the past three years at VisWeek: "Visual Analytics in Healthcare" (multiple times), "Extreme Scale Visual Analytics", "Interactive Visual Text Analytics: Task-Driven Analysis of Social Media Content", and our first edition of this tutorial "Interactive Visual Analysis of Scientific Data". All of them received a good attendance. Altogether, this indicates a growing demand for interactive visual analysis (IVA) in the community and motivates our tutorial.

The tutorial encompasses a theoretical foundation of IVA which supports the tutorial attendees in transferring the subject matter to their own data and application area and helps them in engineering their own IVA solutions. Based on the theoretical part, successful applications of IVA to scientific data from various fields are discussed: climate research, medicine, epidemiology, as well as flow simulation and computation, in particular for automotive engineering. These discussions go beyond the direct application of techniques that were introduced in the theoretical part. They also cover new techniques, e.g., from statistics and knowledge discovery (printed in bold in the tutorial syllabus section), which proved to be particularly useful for a specific application. Hence, the tutorial attendees will continuously broaden their theoretical IVA knowledge.

The tutorial also provides an overview of off-the-shelf IVA solutions which may be particularly interesting for visualization practitioners. Each topic is presented by a leading researcher with excellent communication skills, and considerable experience in giving lectures or talks. The talks will be stimulating by also discussing open problems. All tutorial speakers, as well as their topics, do not only aim at pure academic progress but instead at contributing to solving real application problems. Their experiences and lessons-learned in interdisciplinary collaborations are included. The entire tutorial material will be made available to the attendees on a website.

Tutorial Speakers

- Steffen Oeltze, Otto-von-Guericke University of Magdeburg, Germany
- Helwig Hauser, University of Bergen, Norway
- Johannes Kehrer, Vienna University of Technology, Austria

Tutorial History

The tutorial has been held once before at VisWeek 2012. The tutorial notes of last year are available at: `http://www.vismd.de/doku.php?id=teaching_tutorials:start`. This year, the tutorial has been significantly changed to present new application fields of IVA to the audience: IVA of epidemiological and flow data. In that course, the tutorial parts on IVA of biological and very large data have been removed. The topic "IVA of engineering data", which used to be a stand-alone part, is partially covered by the IVA of flow data since engineering data are employed here for demonstrating the proposed solutions. The tutorial part on IVA of climate data has been updated by the new speaker Johannes Kehrer. The part on IVA tools now covers two instead of three tools in detail, which facilitates a more thorough introduction and leaves time for more examples.

Level of Difficulty and Prerequisites

Intermediate/Advanced. We assume a basic understanding of visualization techniques in physical (e.g., surface and volume rendering, rendering modalities, glyphs) and attribute (e.g., histograms, scatter plots, parallel coordinates) space.

Intended Audience

Potential attendees are researchers interested in visualization and visual analytics who consider adopting the interactive visual analysis (IVA) methodology and engineer an according solution. Furthermore, visualization practitioners who are interested in off-the-shelf IVA solutions are addressed by the tutorial. In general, scientists who are aiming at hypotheses generation or validation based on complex spatiotemporal, multi-variate, multi-modal or multi-run data represent the target audience.

Tutorial Syllabus

The tutorial is designed as a textbook-like presentation. Besides the presentation of interactive visual analysis (IVA) solutions in different application fields, it also includes a theoretic introduction to IVA and a wrap-up comprising the lessons learned as well as open IVA problems. The proposed tutorial schedule is as follows:

- I Introduction, *Oeltze* (10 min)
- II Basics of Interactive Visual Analysis, *Hauser* (35 min)
- III Interactive Visual Analysis of Climate Data, *Kehrer* (35 min)
- IV Interactive Visual Analysis of Medical Data and Epidemiological Data, *Oeltze* (20 min)
- — Break
- IV Interactive Visual Analysis of Medical Data and Epidemiological Data (contd), *Oeltze* (20 min)
- V Interactive Visual Analysis of Flow Data, *Hauser* (35 min)
- VI Interactive Visual Analysis Tools, *Kehrer, Oeltze* (35min)
- VII Lessons Learned and Open Problems, *Oeltze* (10 min)

II This part constitutes the backbone of the tutorial. It equips the attendees with the theoretical background of IVA. The subject matter supports them in engineering their own IVA solutions and also in following the subsequent application part of the tutorial. Basic IVA concepts such as physical views, attribute views, features, and brushing-and-linking are introduced. Further, the three patterns of explorative/analytical procedures in IVA are explained: feature localization, local investigation, and multi-variate analysis. Extensions of the basic IVA concepts are also discussed, for instance, *smooth brushing*, which facilitates the definition of a gradual course between "interesting data" and "data not of interest" [DH02]. This smooth *degree of interest* facilitates the use of generalized focus plus context (F+C) methods which reduce cluttering in the resulting visualizations and draw a user's attention to the most important details. The theoretical part of the tutorial has evolved over years in university lectures and was continuously fine-tuned and updated.

III In climate research, massive amounts of multi-variate, multi-dimensional and time-dependent data are generated by means of climate observations and model-based simulations. IVA concepts can be employed in order to explore the data and to generate new hypotheses for subsequent statistical evaluation [KLM*08, LSL*10]. One of the most prominent topics in climate research is the investigation, detection, and allocation of climate change. In this tutorial part, we demonstrate how IVA helps identifying climate parameters and regions in the atmosphere (e.g., certain height layers) which can act as sensitive and robust indicators for climate change [KLM*08]. For multi-run simulation data, we further show that integrating and brushing **statistical moments** (mean, variance, skewness, and kurtosis) in an IVA framework helps the climate researcher in studying data distributions, data trends and outliers [KFH10]. Climate data may consists of different parts given on different spatial grids over time, e.g., the atmosphere and ocean part from a coupled climate model. We discuss an interface for IVA of such heterogeneous scientific data [KMDH11]. The interface establishes a relation between grid cells in different data parts, it defines how feature specifications are transferred between the parts, and it integrates an update mechanism which keeps the feature specification in both data parts consistent during the visual analysis.

IV In medicine, the generation and diagnostic evaluation of multi-field and multi-modal data are gaining importance. Instead of a single 3D dataset, several datasets are acquired at different points in time and with different imaging modalities. Furthermore, additional data may be derived from the measured data. The complexity of the data requires visual analysis approaches that guide the user to interesting portions of the data by incorporating his/her a priori knowledge and by providing interactive filtering mechanisms. In this tutorial part, we present an interactive visual analysis approach [ODH*07], which comprises, among others, a **data dimension reduction** step (Principal Component Analysis) and a **data clustering and classification** step. The approach has been applied to perfusion data in breast cancer [GPTP10] and brain tumor diagnosis [GOP*13] and in the diagnosis of ischemic stroke [OHR*09] and Coronary Heart Disease [ODH*07]. Perfusion data represent a special instance of medical multi-field data. After the injection of a contrast agent (CA), several volumes covering the object of interest are acquired over time during a single imaging scan and additional parameters are derived from the data which characterize the temporal course of the CA accumulation.

Large-scale longitudinal epidemiological studies such as the Study of Health in Pomerania (SHIP) [VAS*11] investigate groups of people with common characteristics or experiences (a cohort) including socio-demographic and biological factors. Their goal is the characterization of health by identifying risk factors and their relations to diseases and the indication of a per subject risk of developing a disease. Carried out in waves over many years, they comprise thousands of individuals and ten thousands of variables. Unique for SHIP is the inclusion of medical image data acquired via an extensive full-body MRI protocol [HKV*09]. Current SHIP-research projects are the characterization and differentiation of age-dependent and pathologic variance in spinal curvature and the investigation of influences of different lifestyle factors on the shape of the liver. A major goal is the visual analysis of inter- and intra-group statistical variance in curvature and organ shape. In this tutorial, we present an IVA approach that fits into the epidemiological workflow and facilitates hypotheses validation and generation [KOH*12]. The approach supports a bi-directional visual analysis of the non-image and the image data. (1) Groups of individuals and control groups can be defined by the epidemiologist or a **data clustering algorithm** based on the SHIP variables, i.e. the non-image data. Next, features characterizing curvature and shape are derived from the image data, e.g., the centerline of the spinal canal is computed. Finally, intra-group variance as well as inter-group difference with regard to these features are computed and visualized. (2) Features are derived from the image data and groups are generated based on feature differences, e.g., by a clustering technique grouping similar spinal canal centerlines. The groups are then investigated with respect to the associated SHIP variables.

V Computational fluid dynamics (and also experimental flow imaging to a certain degree) lead to large and multivariate datasets, representing flows, which increasingly often also are time-dependent, i.e., unsteady. IVA is a powerful methodology for the exploration and analysis of such datasets and in this part of the tutorial, we focus on IVA of unsteady flows. In detail, we discuss four solutions, dealing with (1) unsteady vortex extraction, a (2) time-dependent analysis, (3) pathlines, and (4) a statistical approach to extract representative factors from multivariate data. (1) In [FPH*08], we presented new criteria for vortex extraction and vortex core line computation in unsteady flow and their embedding in an IVA framework. Since there is no fully satisfying approach to extract only the relevant vortex core lines automatically, the user can interactively specify the flow of interest in attribute views via a set of attribute selections that control the vortex regions. (2) To access unsteady aspects of flows in an IVA framework, we integrate the computation and inspection of temporal attribute changes, the derivation of time-step-relative **normalization**, the interpolation of selections over time, a measure of how stationary an attribute is, and another measure to capture local extrema [DHGK06]. (3) Unsteady flow is often visualized via pathlines. In order to characterize the flow, we compute pathlines and various pathline attributes (scalar and time-dependent) describing their local and global behavior and we use IVA to explore the attribute space [LZM*11]. For interaction and a simplified selection of physical subspaces, we provide projections of pathlines to 2D planes in physical views. (4) The main problem with having multiple pathline attributes is the selection of relevant ones. We employ **Exploratory Factor Analysis** for dimension reduction in six datasets and identify a universal set of attributes, which may serve as a starting point for IVA of unsteady flow [PLMH12]. All presented solutions (1-4) are demonstrated based on engineering data, among others from applications in the automotive engineering industry.

VI In this part, an overview of off-the-shelf IVA solutions will be given and two solutions will be presented in detail: `ParaView` (`www.paraview.org`, Kitware, Clifton Park, NY, U.S.) and `SimVis` [Dol07]. They have been chosen since they incorporate not only attribute but also physical 2D/3D views which are crucial in depicting spatial relations. The open-source software `ParaView` is built on top of the Visualization Toolkit (VTK) and is widely used in the visualization community. It comprises basic IVA concepts such as brushing and linking. As a professional, tailor-made IVA solution, `SimVis` will be presented. `SimVis` is used by various researchers and visualization practitioners working in different scientific and industrial areas. It has also been employed in several research publications, among which 8 IEEE TVCG papers are listed. This tutorial part will be completed by a listing and brief comparison of further IVA solutions, such as `VisIt`, `Tableau`, `Improvise`, and `Tulip`.

Presenters Bio's

Steffen Oeltze
Otto-von-Guericke University of Magdeburg
Department of Simulation and Graphics
P.O. Box 4120
Magdeburg, D-39106, Germany

Phone: +49 (0) 391 67 52527
e-mail: stoeltze@isg.cs.uni-magdeburg.de
www: www.vismd.de

Steffen Oeltze is a postdoctoral researcher at the Computer Science Department at the Otto-von-Guericke University of Magdeburg (UoM). His research interests are in medical visualization, in particular the visualization of vasculature and the visual analysis of 4D-perfusion and blood flow data, and in biological data visualization, for instance the visual analysis of protein colocalization studies. In 2004 and 2008 he received the "Karl-Heinz-Höhne" award (2nd and 1st place) for his scientific work on the visualization of vasculature with convolution surfaces and the interactive visual analysis of perfusion data, respectively. He won the 2005 and 2008 annual research awards of the computer science faculty at the UoM for his work on vessel visualization, published in IEEE Transactions on Medical Imaging, and his scientific overall achievement, respectively. Steffen organized the tutorials "Interactive Visual Analysis of Scientific Data" (2012), "Advanced Visual Medicine: Techniques, Applications and Software" (2007), and "Visual Medicine: Techniques, Applications and Software" (2006) at IEEE Vis/VisWeek. He was a speaker at the tutorial "Advanced Visual Medicine: Techniques for Visual Exploration & Analysis" at IEEE Vis in 2008. Steffen received the diploma (M.Sc.) degree in computational visualistics from the UoM in 2004. In 2010, he received his Ph.D. in computer science from the UoM.

Helwig Hauser
University of Bergen
Department of Informatics
P.O.Box 7803
N-5020 Bergen, Norway

Phone: +47 55584380
e-mail: Helwig.Hauser@UiB.no
www: http://www.ii.UiB.no/vis/

Helwig Hauser graduated in 1995 from Vienna University of Technology (TU Wien), Austria. In 1998, he finished his PhD project on the visualization of dynamical systems. In 2003, he finished his Habilitation at TU Wien entitled "Generalizing Focus+Context Visualization" - in 2006 this work was awarded with the prestigious Heinz-Zemanek Preis. In 2013, H. Hauser received the Dirk Bartz Prize for Visual Computing in Medicine from Eurographics. One of his main activities, more recently, is to chair visualization conferences, including EuroVis 2011, PacificVis 2012, and IEEE InfoVis 2013. H. Hauser is member of the EuroVis Steering Committee, the TopoInVis Steering Committee, and has served / is serving on the Editorial Boards of Computers & Graphics, Computer Graphics Forum, and IEEE Transactions on Visualization and Computer Graphics. After first working for TU Wien as assistant (since 1994) and later as assistant professor, he changed to the new VRVis Research Center in 2000. There, he led the basic research group on interactive visualization (until 2003) before he became the scientific director of VRVis (until 2007). Since then, 2007, he is a full professor in visualization at the University of Bergen in Norway. Helwig Hauser is promoting interactive visual analysis for about thirteen years (with dozens of publications and a larger number of talks, also including several invited/keynote talks).

Johannes Kehrer
Vienna University of Technology
Institute of Computer Graphics and Algorithms
Favoritenstrasse 9-11 / E186
A-1040 Wien, Austria

Phone: +43-1-58801-18645
e-mail: kehrer@cg.tuwien.ac.at
www: http://www.cg.tuwien.ac.at

Johannes Kehrer is a postdoctoral researcher at the Institute of Computer Graphics and Algorithms at the Vienna University of Technology, Austria. He received his MSc degree in visual computing from the same university in 2007. He then worked as a research fellow at the Department of Informatics, University of Bergen, Norway, and received his PhD degree from there in 2011. He spent the year 2012 as a researcher at the VRVis Research Center in Vienna, Austria. His research interests include the visual analysis of multi-variate, spatiotemporal, multi-modal, and multi-run/ensemble data (primarily from the application fields of climate research and engineering). His recently published survey article on these topics [KH13] was the spotlight paper for the March 2013 issue of TVCG. He has cooperated with several climate and meteorological institutes.

References

[DH02] DOLEISCH H., HAUSER H.: Smooth Brushing For Focus+Context Visualization of Simulation Data in 3D. *Journal of WSCG 10*, 1 (2002), 147–154.

[DHGK06] DOLEISCH H., HAUSER H., GASSER M., KOSARA R.: Interactive focus+context analysis of large, time-dependent flow simulation data. *Simulation 82*, 12 (2006), 851–865.

[Dol07] DOLEISCH H.: SIMVIS: Interactive Visual Analysis of Large and Time-Dependent 3D Simulation Data. In *Winter Simulation Conference* (2007), pp. 712–720.

[FPH*08] FUCHS R., PEIKERT R., HAUSER H., SADLO F., MUIGG P.: Parallel Vectors Criteria for Unsteady Flow Vortices. *IEEE Transactions on Visualization and Computer Graphics 14*, 3 (5 2008), 615–626.

[GOP*13] GLASSER S., OELTZE S., PREIM U., BJORNERUD A., HAUSER H., PREIM B.: Visual Analysis of Longitudinal Brain Tumor Perfusion. In *Proc. of the SPIE Medical Imaging* (2013).

[GPTP10] GLASSER S., PREIM U., TÖNNIES K. D., PREIM B.: A Visual Analytics Approach to Diagnosis of Breast DCE-MRI Data. *Computer and Graphics 34(5)* (2010), 602–611.

[HKV*09] HEGENSCHEID K., KÜHN J., VÖLZKE H., BIFFAR R., HOSTEN N., PULS R.: Whole-Body Magnetic Resonance Imaging of Healthy Volunteers: Pilot Study Results from the Population-Based SHIP Study. *Fortschr Røntgenstr 181*, 8 (2009), 748–59.

[KFH10] KEHRER J., FILZMOSER P., HAUSER H.: Brushing Moments in Interactive Visual Analysis. *Computer Graphics Forum 29*, 3 (2010), 813–822.

[KH13] KEHRER J., HAUSER H.: Visualization and visual analysis of multi-faceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (2013), 495–513.

[KLM*08] KEHRER J., LADSTÄDTER F., MUIGG P., DOLEISCH H., STEINER A., HAUSER H.: Hypothesis Generation in Climate Research with Interactive Visual Data Exploration. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1579–1586.

[KMDH11] KEHRER J., MUIGG P., DOLEISCH H., HAUSER H.: Interactive Visual Analysis of Heterogeneous Scientific Data across an Interface. *IEEE Transactions on Visualization and Computer Graphics 17*, 7 (2011), 934–946.

[KOH*12] KLEMM P., OELTZE S., HEGENSCHEID K., VÖLZKE H., TÖNNIES K. D., PREIM B.: Visualization and Exploration of Shape Variance for the Analysis of Cohort Study Data. In *VMV 2012 - Vision, Modeling and Visualization* (2012), pp. 221–222.

[LSL*10] LADSTÄDTER F., STEINER A. K., LACKNER B. C., PIRSCHER B., KIRCHENGAST G., KEHRER J., HAUSER H., MUIGG P., DOLEISCH H.: Exploration of climate data using interactive visualization. *J. Atmospheric and Oceanic Technology 27*, 4 (2010), 667–679.

[LZM*11] LEZ A., ZAJIC A., MATKOVIC K., POBITZER A., MAYER M., HAUSER H.: Interactive Exploration and Analysis of Pathlines in Flow Data. In *Proc. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2011)* (2011), pp. 17–24.

[ODH*07] OELTZE S., DOLEISCH H., HAUSER H., MUIGG P., PREIM B.: Interactive Visual Analysis of Perfusion Data. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1392–1399.

[OHR*09] OELTZE S., HAUSER H., RORVIK J., LUNDERVOLD A., PREIM B.: Visual Analysis of Cerebral Perfusion Data – Four Interactive Approaches and a Comparison. In *Proc. of the 6th International Symposium on Image and Signal Processing and Analysis (ISPA)* (2009), pp. 588–595.

[PLMH12] POBITZER A., LEZ A., MATKOVIC K., HAUSER H.: A Statistics-based Dimension Reduction of the Space of Path Line Attributes for Interactive Visual Flow Analysis. In *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis 2012)* (2012), pp. 113–120.

[VAS*11] VÖLZKE H., ALTE D., SCHMIDT CARSTEN OLIVER ET AL.AND RADKE D., LORBEER R., FRIEDRICH N., AUMANN N., LAU K., PIONTEK M., BORN G., HAVEMANN C., ITTERMANN T., SCHIPF S., HARING R., BAUMEISTER S. E., WALLASCHOFSKI H., NAUCK M., FRICK S., ARNOLD A., JÜNGER M., MAYERLE J., KRAFT M., LERCH M. M., DÖRR M., REFFELMANN T., EMPEN K., FELIX S. B., OBST A., KOCH B., GLÄSER S., EWERT R., FIETZE I., PENZEL T., DÖREN M., RATHMANN W., HAERTING J., HANNEMANN M., RÖPCKE J., SCHMINKE U., JÜRGENS C., TOST F., RETTIG R., KORS J. A., UNGERER

S., HEGENSCHEID K., KÜHN J.-P., KÜHN J., HOSTEN N., PULS R., HENKE J., GLOGER O., TEUMER A., HOMUTH G., VÖLKER U., SCHWAHN C., HOLTFRETER B., POLZER I., KOHLMANN T., GRABE H. J., ROSSKOPF D., KROEMER H. K., KOCHER T., BIFFAR R., JOHN U., HOFFMANN W.: Cohort Profile: The Study of Health in Pomerania. *International Journal of Epidemiology 40*, 2 (2011), 294–307.