Jukka Remes

METHOD EVALUATIONS IN SPATIAL EXPLORATORY ANALYSES OF RESTING-STATE FUNCTIONAL MAGNETIC RESONANCE IMAGING DATA
JUKKA REMES

METHOD EVALUATIONS IN SPATIAL EXPLORATORY ANALYSES OF RESTING-STATE FUNCTIONAL MAGNETIC RESONANCE IMAGING DATA

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Technology and Natural Sciences of the University of Oulu for public defence in OP-sali (Auditorium L10), Linnanmaa, on 18 October 2013, at 12 noon

UNIVERSITY OF OULU, OULU 2013
Remes, Jukka, Method evaluations in spatial exploratory analyses of resting-state functional magnetic resonance imaging data.
University of Oulu Graduate School; University of Oulu, Faculty of Technology, Department of Computer Science and Engineering, Faculty of Medicine, Institute of Diagnostics, Department of Diagnostic Radiology; Oulu University Hospital
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

Resting-state (RS) measurements during functional magnetic resonance imaging (fMRI) have become an established approach for studying spontaneous brain activity. RS-fMRI results are often obtained using explorative approaches like spatial independent component analysis (sICA). These approaches and their software implementations are rarely evaluated extensively or specifically concerning RS-fMRI. Trust is placed in the software that they will work according to the published method descriptions. Many methods and parameters are used despite the lack of test data, and the validity of the underlying models remains an open question. A substantially greater number of evaluations would be needed to ensure the quality of exploratory RS-fMRI analyses.

This thesis investigates the applicability of sICA methodology and software in the RS-fMRI context. The experiences were used to formulate general guidelines to facilitate future method evaluations. Additionally, a novel multiple comparison correction (MCC) method, Maxmad, was devised for adjusting evaluation results statistically.

With regard to software considerations, the source code of FSL Melodic, popular sICA software, was analyzed against its published method descriptions. Unreported and unevaluated details were found, which implies that one should not automatically assume a correspondence between the literature and the software implementations. The method implementations should rather be subjected to independent reviews.

An experimental contribution of this thesis is that the credibility of the emerging sliding window sICAs has been improved by the validation of sICA related preprocessing procedures. In addition to that, the estimation accuracy regarding the results in existing RS-fMRI sICA literature was also shown not to suffer even though repeatability tools like Icasso have not been used in their computation. Furthermore, the evidence against conventional sICA model suggests the consideration of different approaches to analysis of RS-fMRI.

The guidelines developed for facilitation of evaluations include adoption of 1) open software development (improved error detection), 2) modular software designs (easier evaluations), 3) data specific evaluations (increased validity), and 4) extensive coverage of parameter space (improved credibility). The proposed Maxmad MCC addresses a statistical problem arising from broad evaluations.

Large scale cooperation efforts are proposed concerning evaluations in order to improve the credibility of exploratory RS-fMRI methods.

Keywords: analysis, BOLD, brain, data mining, evaluation, exploratory, FastICA, fMRI, Icasso, independent component, medical imaging, quality assurance, research quality, resting-state, sICA, signal processing, spontaneous, validity
Remes, Jukka, Menetelmäevaluaatiot lepotilan funktionaalisen magneetti-
kuvantamisdataan arvuudellisissa eksploratiivisissa analyysissa.
Oulun yliopiston tutkielma; Oulun yliopisto, Teknillinen tiedekunta, Tietotekniikan osasto, Lääketieteellinen tiedekunta, Diagnostiikan laitos, Radiologia; Oulun yliopistollinen sairaala
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Tämä väitöskirja tutki sICA-menetelmien ja -ohjelmistojen soveltuvuutta lepotilan fMRI-tutkimuksiin. Kokemuksien perusteella luotiin yleisiä ohjenuoria helpottamaan tulevaisuuden menetelmäevaluaatioita. Lisäksi väitöskirjassa kehitettiin uusi monivertailukorjusmenetelmä, Maxmad, evaluaatiotulosten tilastolliseen korjaukseen.

Tunnetun sICA-ohjelmiston, FSL Melodicin, lähdekoodi analysoitiin suhteessa julkaistuihin menetelmäkuvauksiin. Analyysissä ilmeni aiemmin raportoimattomia ja evaluoimattomia menetelmäyksiköitä, mikä tarkoittaa, ettei kirjallisuudessa olevien menetelmäkuvausten ja niiden ohjelmistototeutusten välille pitäisi automaattisesti olettaa vastaavuutta. Menetelmätoteutukset pitäisi katsella riippumattomasti.

Väitöskirjan kokonaisena parannetaan liukuvassa ikkunassa suoritettavan sICA:n uskottavuutta varmistamalla sICA:n esikäsittelyjen oikeellisuus. Lisäksi väitöskirjassa näydetään, että aiempien sICA-tulosten tarkkuus ei ole käsinnyt, vaikka niiden estimoinnissa ei ole käytetty toistettavuustyökaluja, kuten Icasso-ohjelmistoa. Väitöskirjan tulokset kyseenalaistavat myös perinteisen sICA-mallin, minka vuoksi tulisi harkita sitä poikkeavia lähtökohtia lepotilan fMRI-datan analyysiin.

Evaluaatioiden helpottamiseksi kehitetty ohjeet sisältävät seuraavia periaatteita: 1) avoin ohjelmistokehitys (parantunut virheiden havaitseminen), 2) modulaarinen ohjelmistosuunnittelu (nykyistä helpommin toteutettavat evaluaatiot), 3) datatyppiksiertaset evaluauat (parantunut validiteetti) ja 4) parametriarvaukseen laaja kattavuus evaluaucioissa (parantunut uskottavuus).

Evaluaatioiden helpottamiseksi kehitetty ohjeet sisältävät seuraavia periaatteita: 1) avoin ohjelmistokehitys (parantunut virheiden havaitseminen), 2) modulaarinen ohjelmistosuunnittelu (nykyistä helpommin toteutettavat evaluaatiot), 3) datatyppiksiertaset evaluauat (parantunut validiteetti) ja 4) parametriarvaukseen laaja kattavuus evaluaucioissa (parantunut uskottavuus).

Monivertailukorjus tarjoaa ratkaisun ratkaisuvaihdehdollista laajenevaksi evaluaucioiden tilastollisiin haasteisiin.

Jotta lepotilan fMRI:ssä käytettävien eksploratiivisten menetelmiä uskottavuus paranisi, väitöskirjassa ehdotetaan laaja-alaisia yhteistyötä menetelmiä evaluaatiorikkeisiin.

Asiakas: aivot, analyysi, BOLD, eksploratiivinen, evaluauato, FastICA, fMRI, Icasso, itsenäinen komponentti, laadunvarmistus, lepotila, lääketieteellinen kuvantaminen, sICA, signaalinkäsittely, spontaani, tiedonrikastus, tutkimuksen laatu, validiteetti
Science is the belief in the ignorance of experts.
– Richard Feynman

To my family, to my parents, to lifelong learning and to critical thinking
Acknowledgements

The work for this thesis was done in the fMRI research group during my employment as a research engineer at Oulu University Hospital and simultaneously also as a part of Machine Vision Group (MVG) in the Department of Computer Science and Engineering (CSE) at the University of Oulu. This work has been supported by the Academy of Finland (the EDEN grant in the NEURO program) and by the Ethel E.E. Remes Fund.

First of all, I would like to thank my thesis supervisors Professors Olli Silvén, PhD, and Osmo Tervonen, MD, PhD. Back in 2004, I was leaving my previous employer and could not find available positions with personally interesting topic for graduate studies in Oulu. Professor Silvén saw that I was potentially suitable for my current position, and Professor Tervonen as the head of department at the hospital gave me the opportunity to start there. Having worked both in industry and in applied research, but only in engineering workplaces before, it has been a great opportunity for me to learn about basic research, the medical application domain and cross-disciplinary work communities. My warm thanks go also to my Master’s thesis supervisor, Professor Juha Röning, who supported my admission to PhD program in 2002 thereby enabling me to pursue this path in the first place.

I wish to thank Professor Silvén and Professor Tervonen also for their guidance and support over the years. In the hospital, I have had the possibility to participate in all aspects of research from funding application and imaging project preparations to technical details of data processing and analyses, and building an fMRI analysis infrastructure. On the other hand, within Professor Silvén’s Machine Vision Group I was able to utilize the vast computing resources of CSE, which were crucial in carrying out the many evaluations presented in this thesis. In the MVG premises at the Linnanmaa campus, I also found the possibility to better concentrate on thesis research by having a secondary work location there. Concerning later times, I would also like to thank my current supervisor at the hospital, Chief Physicist, Professor Miika Nieminen, PhD, for arrangements enabling me to finalise my thesis.

I would also like to thank the Principal fMRI Investigator at the hospital, Docent Vesa Kiviniemi, MD, PhD, whose own thesis work on resting-state fMRI made it possible for me (as well as for many others) to pursue a thesis regarding this topic. Without Docent Kiviniemi’s seminal work and passionate interest in spontaneous brain
activity and advocating of the topic, I doubt there would be current fMRI research traditions in Oulu. I also want to thank him for all the interest he has had in my thesis, and for all the advice he has given over the years. I have learned the most at the hospital through interactions with him and through the technical challenges his research problems have presented for me to solve.

Other long time members and collaborators in our fMRI research also deserve warm thanks from me. From the beginning Tuomo Starck, MSc (Tech), has been an important peer, with whom I have been able to discuss research matters from an engineering point of view, especially given the cross-disciplinary nature of our fMRI research. Docent Juha Nikkinen, PhD, too has been a crucial member in our group; his efforts along the way have enabled also my work for this thesis. It has been a pleasure working with you both, learning together fMRI, developing our research practices, and imaging volunteers. In addition to Tuomo and Juha, I would like to thank our freelance member Jussi Kantola, whose efforts in system administration and analysis development has also freed my time for other things, such as doing this thesis.

Concerning learning about spatial ICA and ICA in general, I express my gratitude to Professor Christian Beckmann, PhD, from University of Twente, Netherlands and to Docent Esa Ollila, PhD, from the Department of Signal Processing and Acoustics in Aalto University School of Science and Technology. Their expert advice and instructions on these topics, delivered both personally and through public forums, have been extremely helpful in understanding methodological details.

Besides the studies included to this thesis, I have also been privileged to participate in various other fMRI studies. In addition to all the aforementioned co-authors, I give my thanks to Ahmed Abou Elseoud, MD, Marianne Haapea, PhD, Tuija Hiltunen, MSc, Harri Littow, MD, Jyri Paakki, MD, Salla-Maarit Kokkonen, MD, PhD, and Minna Silfverhuth, PhD, as well as to the many others who have made the studies and corresponding publications possible. I would also like to thank Professors Juha Veijola, MD, PhD, and Matti Isohanni, MD, PhD, from the Department of Psychiatry at the University of Oulu for allowing me to participate in their Northern-Finland birth cohort studies and to gather imaging data during them, some of which is used also in this thesis. The lack of fMRI data with state-of-the-art quality was a big challenge when I first started at the hospital. Many thanks go also to Leila Salo, Arja Väisänen and other secretariat who have taken care of various things during my years at the hospital, as well as to many radiographers, radiologists, hospital physicists and other staff who have given their advice and help on various subjects.
I also wish to thank the official reviewers of this thesis, Professor Koen Van Leemput, PhD, from the Technical University of Denmark, and Docent Ricardo Vigário, PhD, from Aalto University School of Science and Technology, for their valuable comments that were very helpful in improving the thesis. Similarly, I want to thank Mr. Gordon Roberts for performing the language review.

Even though this thesis is critical of current software tools and their development, I want to recognize and emphasize their importance. Without efforts such as the development of FSL Melodic, Icasso and other repeatability tools, or GIFT, it would be much more laborious to perform any explorative analyses. Moreover, the reported results would probably suffer from much additional variability. As a former software designer, and knowing now the practicalities of research work, I also acknowledge the challenges of developing software with limited time alongside various other research tasks. Also the evaluations experiments carried out in this thesis would have been much harder if the current tools had not been available.

I would like to acknowledge the significance of my other friends and acquaintances, many of whom I know from society activities. During my years at the hospital especially activities and friends in the Technical Society of Oulu have provided me with escapes from science which at times can become a little wearisome. I especially want to thank my long term friend Markus and my cousin Juho for refreshing conversations and company. It has also been interesting to compare notes with Juho who made his PhD in forestry.

My warmest thanks go to my family. My parents Martti and Heljä have always supported me, and I think something crucial happened concerning the path I have taken when they encouraged me to spend my last school year in a program containing advanced mathematics and physics. They have also set quite an example themselves concerning life-long professional development with their various studies continuing almost up to their retirement years.

I am very thankful for all my relatives with whom it is always straight-forward (yet sometimes intensive;) to discuss any matter. I want to thank especially my uncles Seppo and Heikki, engineers and former researchers themselves, for all their encouragement and interesting (and sometimes comforting) stories that they have told about their own experiences in the academic and engineering worlds.

Finally, more than anything else I am grateful for and to my wife Päivi and our three adorable daughters, the best kind of products two engineers can envision. I would not
have been able to finish this thesis without opportunities to work on it also at home. On the other hand, family time has shown me the real perspective of life.

Oulu, July 2013

Jukka Remes
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>ALFF</td>
<td>amplitude of low frequency fluctuations</td>
</tr>
<tr>
<td>ANN</td>
<td>artificial neural network (a class of machine learning methods)</td>
</tr>
<tr>
<td>ARABICA</td>
<td>Adaptive Robust Additions to Bagging ICA, a tool similar to Icasso</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian information criterion</td>
</tr>
<tr>
<td>BOLD</td>
<td>blood-oxygen-level-dependent</td>
</tr>
<tr>
<td>BSS</td>
<td>blind source separation</td>
</tr>
<tr>
<td>CDF</td>
<td>cumulative density function</td>
</tr>
<tr>
<td>DMN</td>
<td>default mode network</td>
</tr>
<tr>
<td>ECG</td>
<td>electrocardiography</td>
</tr>
<tr>
<td>EEG</td>
<td>electroencephalography</td>
</tr>
<tr>
<td>EPI</td>
<td>echo planar imaging</td>
</tr>
<tr>
<td>fALFF</td>
<td>fractional ALFF</td>
</tr>
<tr>
<td>FastICA</td>
<td>ICA estimation method based on fixed-point iteration scheme and maximization of source estimate non-Gaussianity</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>FN</td>
<td>false negative</td>
</tr>
<tr>
<td>fNIRS</td>
<td>functional near infrared spectroscopy</td>
</tr>
<tr>
<td>FOBI</td>
<td>Fourth-Order Blind Identification</td>
</tr>
<tr>
<td>FP</td>
<td>false positive</td>
</tr>
<tr>
<td>FSL</td>
<td>FMRIB Software Library, a software suite published by Oxford Center for Functional MRI of the Brain</td>
</tr>
<tr>
<td>FWE</td>
<td>family-wise error</td>
</tr>
<tr>
<td>GE</td>
<td>gradient echo</td>
</tr>
<tr>
<td>GICA</td>
<td>Group ICA</td>
</tr>
<tr>
<td>GIFT</td>
<td>The Group ICA Of fMRI Toolbox, published by Medical Image Analysis Lab</td>
</tr>
<tr>
<td>GLM</td>
<td>General Linear Model</td>
</tr>
<tr>
<td>IC</td>
<td>independent component</td>
</tr>
<tr>
<td>ICA</td>
<td>independent component analysis</td>
</tr>
</tbody>
</table>
Icasso software for investigating the reliability of ICA estimates by clustering and visualization

ICN intrinsic connectivity network, same as RSN

Infomax ICA estimation method based on principle of information maximization (an optimization for ANN algorithms)

JADE Joint Approximate Diagonalization of Eigenmatrices

KCC Kendall’s coefficient of concordance

LAP Laplacian evidence for model order

LFP local field potential

Maxmad Maximum of maximum statistic distribution

MCC multiple comparison correction

MDL minimum description length

MEG magnetoencephalography

Melodic Multivariate Exploratory Linear Optimized Decomposition into Independent Components, sICA tool in FSL

MI mutual information

MREG MR-Encephalography

MRI magnetic resonance imaging

NMR nuclear magnetic resonance

PC principal component

PCA principal component analysis

PDF probability density function

PET positron emission tomography

PICA probabilistic ICA

PPCA probabilistic PCA

RAICAR ranking and averaging independent component analysis by reproducibility, a tool similar to Icasso

ReHo regional homogeneity

RF radio frequency

RS resting-state

RSN resting-state network, same as ICN

RS-fMRI resting-state fMRI

SAD seasonal affective disorder

SCA seed correlation analysis

SE spin echo
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sICA</td>
<td>spatial ICA</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>sPCA</td>
<td>spatial PCA</td>
</tr>
<tr>
<td>SPM</td>
<td>Statistical Parametric Mapping</td>
</tr>
<tr>
<td>T</td>
<td>Tesla, unit of magnetic flux density</td>
</tr>
<tr>
<td>TFCE</td>
<td>Threshold Free Cluster Enhancement</td>
</tr>
<tr>
<td>tICA</td>
<td>temporal ICA</td>
</tr>
<tr>
<td>VN</td>
<td>voxel-wise variance normalization</td>
</tr>
<tr>
<td>ε</td>
<td>stopping (convergence) threshold in FastICA and Infomax estimations</td>
</tr>
<tr>
<td>κ</td>
<td>learning rate in Infomax ICA estimation</td>
</tr>
<tr>
<td>A</td>
<td>mixing matrix in ICA model</td>
</tr>
<tr>
<td>B</td>
<td>unmixing matrix in ICA model, inverse or pseudoinverse of A</td>
</tr>
<tr>
<td>d_{cw}</td>
<td>subspace similarity computed between PCA subspaces of consecutive temporal windows</td>
</tr>
<tr>
<td>d_{wf}</td>
<td>subspace similarity computed between PCA subspaces of temporal window and full data</td>
</tr>
<tr>
<td>e</td>
<td>vector of error terms in ICA models and in GLM</td>
</tr>
<tr>
<td>g</td>
<td>first derivative of $G$</td>
</tr>
<tr>
<td>$G$</td>
<td>contrast function used to approximate negentropy</td>
</tr>
<tr>
<td>$H$</td>
<td>differential entropy</td>
</tr>
<tr>
<td>$J$</td>
<td>negentropy</td>
</tr>
<tr>
<td>$p$</td>
<td>PDF</td>
</tr>
<tr>
<td>$s$</td>
<td>vector of latent variables (source signals) in ICA model</td>
</tr>
<tr>
<td>$t_{max}$</td>
<td>the number of time points in sICA models</td>
</tr>
<tr>
<td>$V$</td>
<td>whitening matrix</td>
</tr>
<tr>
<td>w</td>
<td>column vector of $W$</td>
</tr>
<tr>
<td>$W$</td>
<td>whitened unmixing matrix in ICA model</td>
</tr>
<tr>
<td>$x$</td>
<td>vector of observed variables in ICA model</td>
</tr>
<tr>
<td>$y$</td>
<td>vector of source signal estimates in ICA model</td>
</tr>
</tbody>
</table>
List of original publications

This thesis is based on the following publications and their respective studies, which are referred to in the text by their Roman numerals (I–IV):


# Contents

Abstract

Tiivistelmä

Acknowledgements

Abbreviations

List of original publications

Contents

## 1 Introduction

1.1 Background and motivation

1.2 Research questions

1.3 Focus and scope of the thesis

1.4 Research methods

1.5 Novel contributions of the thesis

1.6 Author’s contributions

1.7 Organization of the thesis

## 2 Background theories and literature

2.1 Functional magnetic resonance imaging

2.1.1 Magnetic resonance imaging

2.1.2 Measuring brain activity with BOLD fMRI

2.1.3 Signal artefacts

2.1.4 Spontaneous brain activity, resting-state fMRI and resting-state networks

2.2 Spatial exploratory analyses

2.2.1 Spatial independent component analysis

2.2.2 Other exploratory approaches

2.3 Method evaluations regarding spatial exploratory analyses

2.4 Assessing the significance of analysis results

2.4.1 Statistical testing in analyses of image data

2.4.2 Corrections for multiple comparisons
3 The contributions

3.1 Investigations into current methodology used in exploratory analysis of RS-fMRI data

3.1.1 Analysis of FSL Melodic implementations of probabilistic sICA and the temporal concatenation group ICA

3.1.2 Experimental studies

3.1.3 Key findings and evidence

3.2 Developmental proposals

3.2.1 Guidelines for improving method evaluations (DP1)

3.2.2 A concatenation method for multiple comparison correction through permutation testing

3.3 The summary of the answers provided to the research questions

3.3.1 The methodological research questions and results

3.3.2 The developmental research questions and proposals

3.3.3 Other contributions

4 Discussion

4.1 Implications and relevance of the contributions

4.1.1 Aspects of academic software reliability

4.1.2 Sliding-window sICA for RS-fMRI

4.1.3 Validity of RS-fMRI sICA research with respect to estimation robustness

4.1.4 The model and method development for exploratory analyses in RS-fMRI

4.2 Methodological considerations

4.3 Future directions

5 Summary and conclusions

References

Original publications
1 Introduction

1.1 Background and motivation

Functional magnetic resonance imaging (fMRI) has become an established method for studying brain function (Pekar 2006, Bandettini 2012). Since the beginning of fMRI research, approximately 20 years ago, most studies have focused on brain responses with respect to pre-defined cognitive or other types of challenges that participants have been subjected to. During the last 10 years, however, the so called resting-state fMRI (RS-fMRI) has emerged as a widely accepted means of studying the cognitive processes of the brain in a less restricted manner (Biswal et al. 2010, Snyder & Raichle 2012).

RS-fMRI research investigates how the brain functions spontaneously when no structured stimuli is present and an experiment contains a minimal amount of control over the participant. It has been successful in describing both the macro-architecture of the brain and the structure of endogenous signals which are spatially and temporally omnipresent in the brain and which make modelling in more structured fMRI studies difficult (Fox & Raichle 2007a, Snyder & Raichle 2012). RS-fMRI have also been seen to hold promise of a breakthrough for clinical applications of fMRI (Fox & Greicius 2010). On the other hand, the analysis of RS-fMRI data has several challenges (Cole et al. 2010, Smith 2012).

During the last 5-6 years, exploratory analyses, such as independent component analysis (ICA), have become popular methods for finding structure within RS-fMRI data sets (Beckmann 2012). They offer means for studying data without any a priori hypotheses, compared to other type of methods such as computing temporal correlations with a seed voxel (Biswal et al. 1995). Different variants of spatial ICA (sICA; McKeown et al. 1998b) seem by far the most widely used exploratory approaches for RS-fMRI.

Ideally, exploratory analyses of RS-fMRI data would be used to formulate hypotheses that could be verified by employing more structured experiments and corresponding methodology. In practice, however, exploratory results from RS-fMRI studies are also used as end points of analyses and neuroscientific inference is drawn directly from them. This imposes notable expectations on the reliability of exploratory methods and the models and assumptions which they are based on.
Concerning the development and deployment of exploratory methods in RS-fMRI research, it is common that methods are evaluated to a very limited extent concerning their applicability in RS-fMRI analyses. For example, the use of approaches such as probabilistic ICA (PICA; Beckmann & Smith 2004), Icasso (Himberg et al. 2004) (and other repeatability measures) and group-level sICAs (Calhoun et al. 2009) rely on ideas derived from and evaluations performed mainly in contexts other than RS-fMRI. Evaluations may even be performed only once in an initial publication which introduces a method and cover only a very limited range of method parameters and options. Moreover, fMRI analyses in general consist of many independent, but mutually linked phases (Strother 2006, Churchill et al. 2012b,a), each having multiple options to select from, and the resulting explosion in possible parameter combinations further increases the disparity between the amounts of evaluation efforts and requirements.

The limitedness of method evaluations constitutes a blind spot concerning the quality of RS-fMRI research. On the one hand, the limited extent of evaluations is very understandable as the research community employing exploratory methods is small (Beckmann 2012) compared to the rest of RS-fMRI community (Snyder & Raichle 2012) (and the number of method developers even smaller); on the other hand, there is a fundamental contradiction with the need for reliability. The problem is exacerbated by the fact that there are many variants of the widely used approaches.

In addition to purely methodological considerations, also software implementations of methods are a potential source of errors for analyses. Most research groups and individuals developing novel methods also do their software implementations by themselves. However, there seem not to be any reported reviews concerning software even though some of them, like Melodic in the FSL software suite (Jenkinson et al. 2012) and GIFT (Correa et al. 2005), are quite widely used by both RS-fMRI researchers and the rest of the fMRI community. For example, Melodic is a much less formally tested FastICA (Hyvärinen 1999) implementation than the original FastICA source code, which has been subjected to a multitude of numeric and other evaluations in several independent studies, and used for various data domains in addition to fMRI.

Both methods and their software implementations are ultimately based on some hypothetical model and related assumptions. In this sense, the employed models set the baseline validity for analyses based on them. Incorrect theory may lead to non-optimal performance, if not to erroneous analysis results. For example, statistical independence has been challenged regarding common approaches to fMRI sICA (Daubechies et al. 2009). On the other hand, you might well ask how applicable general
analysis frameworks, such as ICA, are in different data domains, and how much the analysis approaches might have to be customized, for example in the case of RS-fMRI.

This thesis addresses questions about the reliability of methods, software and models in spatial exploratory analyses of RS-fMRI data. The subject is treated from both specific (sICA) and general perspectives. The lessons from the analyses performed in the thesis are used to formulate a proposal on how to improve method evaluations in RS-fMRI research. These contributions both complement current understanding about the tools and approaches that are commonly employed, and offer new directions for development.

1.2 Research questions

The main methodological research questions of the thesis are:

MQ1 Can the widely used exploratory analysis software tools be expected to always implement the analysis steps according to the method descriptions in the literature?
MQ2 Are the existing probabilistic sICA related pre-processing procedures applicable also for the emerging sliding window analyses of RS-fMRI data?
MQ3 Does the validity of single-subject RS-fMRI sICA results require repeatability analysis based estimations?
MQ4 Are the data models assumed in exploratory sICA of RS-fMRI data still plausible or should they be considerably revised?

In addition, the thesis investigates the following developmental questions:

DQ1 How could method evaluations be improved in order to further improve the quality of methodology used in exploratory analyses of RS-fMRI data?
DQ2 How could the quality of method evaluations themselves be improved from a statistical point of view?

The methodological research questions are later addressed through respectively numbered methodological results (referred to as MR1-MR4 throughout the thesis). The developmental questions, on the other hand, are approached by respective developmental proposals (denoted by DP1 and DP2 later in this thesis).
1.3 Focus and scope of the thesis

While there are a number of other methods that are, or could be, used in exploratory analyses of RS-fMRI data, sICA is the most widely used approach. Regarding the results of analyses, this thesis focuses on sICA, and does not focus on specific questions related to other methods. SICA is studied in this thesis specifically from subject-level point of view, omitting group-analysis related additional considerations and complexity. Furthermore, the examples and experiments in this thesis largely focus on FastICA, even though Infomax sICA is another widely used method in fMRI sICA. FastICA focus was chosen as it has been the method of choice in the research group within which the research for this thesis was carried out.

Pre-processing and model related considerations in this thesis are applicable regardless of the sICA variant used. The pre-processing research is limited to a sliding window context, and specifically to the effects of principal component analysis (PCA) and voxel-wise variance normalization (VN). Other, more conventional pre-processing subjects, such as slice timing or head motion correction, are not treated in this thesis, apart from some of them being normal parts of the pre-processing pipelines applied to the data.

Concerning data, the experiments in this thesis address the research questions from the viewpoint of conventional echo planar imaging (EPI) blood-oxygen-level-dependent (BOLD) 1.5 T measurements. As such, the scope of the thesis does not include newer imaging sequence or 3 T, 7 T or other higher field strength related imaging and analysis considerations, nor the advantages that they may provide.

The proposals and approaches formulated in this thesis provide viewpoints that are not limited to subject-level sICA and 1.5 T BOLD EPI only, but also address exploratory analysis of RS-fMRI in general. The proposals are focused on testing and evaluations rather than on the development of new methods. However, directions for method development are discussed in the context of the sICA model related results.

1.4 Research methods

To answer research question MQ1, software source code was analysed with respect to corresponding method descriptions in the literature (mainly in Study I but also in Studies II and III). Questions MQ2 and MQ3 were investigated through quantitative
method evaluation experiments (Studies II and III respectively). The findings of these evaluations also provide case examples for answering research question MQ4.

Based on the experiences from the analyses, new guidelines were formulated as a solution to question DQ1. Finally, concerning question DQ2 a novel extension to statistical testing developed for a specific sICA setting (in Study IV) was generalized for the purpose of method evaluations.

1.5 Novel contributions of the thesis

The thesis includes a case analysis of software that is widely used in the exploratory analysis of RS-fMRI (Studies I, II and III). The findings (MR1) expose the need for evaluations of method implementations.

The experimental method evaluations in this thesis provided novel answers to the applicability of sICA related pre-processing procedures (Study II) and methods (Study III). The pre-processing procedures were evaluated in advance before wider deployment of the sliding window sICA. Therefore the results regarding pre-processing (MR2) contribute to the credibility of using the sliding window sICA in later analyses, and to making meaningful methodological choices regarding them. On the other hand, evaluations of sICA methods were done retrospectively, several years after they had been applied to several data sets. These results (MR3) complement previous evaluations that are few in number, and contribute to re-assessing the validity of the sICA results reported in existing literature.

The results of the evaluations were also used to provide a novel viewpoint on the plausibility of the data models that currently underlie RS-fMRI sICA methods (MR4). These observations provide a suggestion for the direction that could be beneficial for method development regarding RS-fMRI exploratory analysis. Furthermore, the observations made in Studies I, II and III were used to formulate a set of guidelines about how method evaluations could be improved (DP1). The changes proposed to current practices provide possibilities for the development of more thoroughly tested methods and their implementations, which increase the credibility and validity of exploratory results concerning RS-fMRI.

Finally, a generalized version of the statistical correction method presented in Study IV was proposed (DP2). The proposed method provides a novel means of accounting for multiple comparisons that may increase concomitantly with the number of parameters in method evaluations.
1.6 Author’s contributions

The contributions by the author of this thesis (later "the author") are as follows concerning the original publications I - IV. In the studies reported in Publications I, II and IV the author participated in data acquisition (planning and imaging) and pre-processing. For the study reported in Publication III the author performed all the pre-processing.

Regarding Studies I, II and III, the author performed the source code analysis concerning the tools used in them. The author performed the spectral estimations used in Studies I and II, and devised and implemented the spatial overlap computations used for the quantitative anatomical characterizations of findings in Study I. He also planned and implemented all the experimental setups and related estimations and comparisons in Studies II and III.

The author was the main writer of Publications II and III and participated substantially to the writing of Publications I (concerning methodology) and IV (concerning the multiple comparison correction method). Finally, the author developed the novel multiple comparison correction approach presented in Publication IV.

1.7 Organization of the thesis

The organization of the thesis is as follows. First, theories are presented and the literature is reviewed in relation to RS-fMRI and its exploratory analysis with sICA (Chapter 2). Second, the contributions of this thesis are presented regarding the software analysis case, experimental studies, their results and methodological proposals (Chapter 3). Finally, in Chapter 4, the contributions are discussed with respect to the research questions, and regarding the limitations and the implications, and in relation to the future of RS-fMRI exploratory analyses.
2 Background theories and literature

The theoretical background relevant to the subject of this thesis is presented in this chapter, both generally and concerning specific details and viewpoints pertaining to the thesis contributions. Section 2.1 introduces relevant measurement related theories and considerations, but also describes data domain specific short-comings which motivate improvements to method reliability.

Different types of exploratory analyses are reviewed in 2.2 with emphasis on spatial ICA, the many aspects of which are covered extensively in 2.2.1. Section 2.2.1 also points out the many factors that have not been evaluated before, concerning RS-fMRI or even fMRI, and demonstrates further the relevance of the contributions delivered by this thesis.

Section 2.3 describes the modest state of method evaluations in RS-fMRI, which emphasizes the need for contributions (evaluations and guidelines) provided in this thesis. Finally, 2.4 reviews the theory regarding the aspects of statistical testing in fMRI research, and shows the background of multiple comparison correction developed in this thesis.

2.1 Functional magnetic resonance imaging

This section introduces the basics of the concepts that underlie fMRI and RS-fMRI. The main points leading to typically acquired fMRI data are shortly illustrated in Figure 1. Further details are discussed in the following subsections.

As shown in Figure 1A, the subject lies along the field lines of a strong magnetic field generated with a MRI scanner. During data acquisitions, the target tissue of interest (head) is positioned inside the scanner (Figure 1B). The whole head can be probed by performing slice selections at multiple positions along the z-axis (an example slice selection in Figure 1C).

Imaging can provide slices with high spatial resolution and an image contrast which helps distinguish between tissue types in analyses (Figure 1D). When brain activity is measured with fMRI, low resolution images with the blood-oxygen level dependent (BOLD) contrast (Figure 1E) can be acquired rapidly for multiple positions and at multiple time points (Figure 1F) resulting in a 3-dimensional video showing BOLD contrast changes in brain.
Fig 1. The fMRI setup, the data organization and axes, and the origin of observed signals. A) The subject aligned along magnetic ($B_0$)-field lines within an MRI scanner. B) Head positioning during fMRI session and the axes of the transverse plane. C) Slice selection along z-axis. D) A high resolution slice with anatomical image contrast. E) A low resolution fMRI slice with BOLD contrast. F) FMRI data: stacks of BOLD images (or “volumes”) acquired by multiple slice selections at multiple points of time (t). G) 'Baseline' levels of blood flow and oxygen (blue balloons) bound to blood haemoglobins (red balloons). Low BOLD value. H) Blood vessel dilation (grey arrow), increased blood flow and an excess (an overshoot) of oxygen in blood during higher electrical brain activity. High BOLD value. I) BOLD value time evolution in a pixel (the red point in D-F) of RS-fMRI data.
The image intensity of a particular pixel depends on the relative level of oxygen in blood (Figure 1G-H) in the corresponding anatomical point (the red dots in Figure 1D-F). When electrical brain activity in neurons at or near the point is at a 'baseline' level (e.g. at the blue point in Figure 1I), much of oxygen, bound to haemoglobins carried by blood flow, is consumed by neurons (Figure 1G). When electrical activity increases (e.g. at the red point in Figure 1I), metabolic demand increases but the blood vessel also dilates and blood flow increases which results in excess oxygen in blood (Figure 1H), and consequently in elevated BOLD contrast.

In RS-fMRI, relative BOLD contrast changes of a few percent (Figure 1I, vertical axis) are observed for each pixel and measured at tens, hundreds or thousands of time points (Figure 1I, horizontal axis). The resulting 4-dimensional (x, y, z, t) RS-fMRI data (Figure 1F&I) can be analyzed with spatial exploratory analyses to provide neuroscientific insights into brain function.

2.1.1 Magnetic resonance imaging

Magnetic resonance imaging (MRI) (Lauterbur 1973), based on nuclear magnetic resonance (NMR) (Rabi et al. 1938) provides the means of producing tomographic images from the human body non-invasively and without ionizing radiation. NMR experiments utilize the fact that many types of atomic nuclei have a non-zero "spin". The spin, an intrinsic quantum property of elementary particles that constitute the nuclei, cause a nucleus to act as a magnetic dipole whose orientation precesses about an axis. The orientations of precession axes and the precession frequencies can be manipulated and used to produce measurable responses from a target of interest.

In NMR and MRI measurements, the subjects of the experiment are placed in strong static magnetic fields (e.g. in the case of MRI 1.5 T, 3 T, 7 T or 11 T or even higher). This affects atomic nuclei (that have a non-zero spin quantum number) by aligning their otherwise randomly oriented precession axes along the field lines of the external high field ("B0-field", Figure 1A) which they are subjected to, and by adjusting their precession frequencies (Larmor frequencies) to ones characteristic of them for that field strength. The frequencies vary between different types of nuclei. As a result, all the nuclei in the target sample reach certain uniformity in terms of axis alignment and for mutually similar nuclei also in terms of their precession frequencies.

The phases of precession remain random, and a part of the nuclei align their precession axes in the direction of the external field, whereas others align in the opposite
direction. This leads to magnetic spins of different nuclei approximately canceling each other out. However, in certain conditions, there can be an excess of nuclei precessing about axes in one direction, thus producing a net magnetization along that direction ("longitudinal magnetization"), which can be utilized for measurements. Concerning MRI, in room temperature, the hydrogen nuclei in water molecules, abundant in tissue, provide such a system of nuclei.

When this small excess of hydrogen nuclei, providing a net magnetization, are subjected to a radio frequency (RF) wave transmission which matches their precession frequency for the given $B_0$-field (and is orthogonal to it), they absorb energy from the transmission, the phases of precession synchronize, and a part of them change alignment. As a result, the net magnetization diminishes in the direction of the $B_0$-field, but due to phase synchronizations the net magnetization grows perpendicular to the $B_0$-field ("transverse magnetization"), i.e. in x- and y-directions in Figure 1B. Effectively, this new two-component net magnetization eventually rotates partly or completely onto the plane perpendicular to the $B_0$-field if enough RF signal is transmitted to the studied sample. If the transmission is still continued further, the rotation will also continue back in the direction parallel with the external field, but in the opposite direction compared to the original alignment.

The system is not in its stable state after such excitation and it starts to reverse (relax) back to the original state when RF transmission ceases. The phases of precession desynchronize ("$T_2$-relaxation") between nuclei, which decreases transverse magnetization, and the spin alignments become normalized ("$T_1$-relaxation"), which makes longitudinal magnetization grow back to its original intensity. The rates of these relaxations are characterized with the time constants $T_2$ and $T_1$, that correspond to a 63% reduction in transverse magnetization and a 63% recovery of longitudinal magnetization, respectively. Also, $T_2 \leq T_1$. $T_2$ and $T_1$ both depend on the nuclei type and vary in tissue from one location to another as a function of local magnetic field variations which depend on the chemical composition of the tissue.

Due to relaxation, the system of nuclei also emit back an RF signal which can be detected and utilized to measure local properties in the subject of the experiment. The strength of the received signal reflects the density of the spins, which is determined by tissue type, but it is also affected by imaging parameters. For example, the interval between an excitation and the corresponding signal acquisition (reflected by "echo time" or "TE") and the interval between excitations performed in consecutive measurements (defined by "repeat time" or "TR") determine how sensitive signal intensity is to the
and $T_1$ of different tissue types. Also, phase desynchronizations occur faster in the presence of field inhomogeneities caused, for example, by certain types of compounds in the vicinity of the nuclei. Thus the signal can be used to characterize and indirectly measure molecular properties near the nuclei.

When additional smaller magnetic field "gradients" (Lauterbur 1973) along different spatial axes (x, y and z in Figure 1) are added on top of the $B_0$-field, the strength of the total field slowly grows along these axes. In this way, different locations in the subject of the experiment experience a slightly unique magnetic field and, consequently, the nuclei at different places also precess at slightly different frequencies (as they are dependent on the field strength). During imaging, usually a preferred slice (or a thicker slab) of tissue is first selected by applying a gradient in a direction perpendicular to the selected plane (e.g. along z-axis). After this, only the spins in the plane are excited by RF transmission that contains frequencies corresponding to the spin precession frequencies in that plane.

After this "slice selection", gradient fields in in-plane (or in-slab) directions (along x- and y-axes) can be used to perform further spatial encoding. When a gradient is turned on in one direction, while acquiring a signal sent back from spins, precession speeds of spins vary along the direction of the gradient. Consequently, different locations emit RF signals with unique frequencies ("frequency encoding"). As a result, the amplitudes for different frequencies in the received signal correspond to different locations along one in-plane axis (e.g. x-axis). Thus the received signals can be used to determine a tomographic image with each pixel showing an image intensity characteristic of a similar (x-axis) location (volume element or voxel) in the tissue.

However, this "readout" for the received signal can be performed concerning location information along only one spatial axis (x-axis). For other 1-2 spatial directions (y-axis; or y- and z-axes if a thick slab was selected), the level of spin phase synchrony is modulated by using the other magnetic field gradients prior to acquisition by deliberate dephasing of spins ("phase encoding"). This changes signal amplitudes from one signal readout to another. Individual signal readout lines with these mutual phase synchrony differences are then used to fill a "k-space" matrix. The elements in the matrix correspond to spatial frequencies in an image of the tissue. Effectively, both frequency encoding and phase encoding probe the tissue for different spatial frequencies in two or three directions.

In addition to applying gradients and transmitting RF signals, their timing and intensities have to be suitably set, which can be done in various ways. A set of these parameters constitutes an "imaging sequence". Since the conception of MRI, numerous
sequences have been developed, each of them having their own characteristics in terms of practicality and the quality of data.

After data acquisition, a 2- or 3-dimensional Fourier transform is applied to the k-space matrix in order to reconstruct pixel data for an individual image slice (Figure 1C-E) or volume, depending on the acquisition scheme (slice selection and y-axis phase encoding, or slab selection and y- and z-axis phase encoding). If only two dimensions (e.g. x and y) are used, i.e. single slice imaged (Figure 1C-E), multiple slice selections are used to cover the whole volume of interest in z-direction (Figure 1F). In this case, the frequencies of RF transmissions used for excitations are slightly changed when signals from a different slice are preferred as the slice selection gradient correspondingly makes the nuclei in different slices sensitive to slightly different frequency bands. In addition to these conventional procedures, various other methods for gathering k-space data and for performing reconstruction do exist.

MRI acquisitions can be parametrized in various ways to produce image contrasts that are complementary to each other concerning the information they provide about tissue. For example, as TE and TR control how sensitive a received RF signal is to the $T_2$ and $T_1$ of different tissue types, adjusting TE and TR can also be used to make pixel values reflect differences between the composition and content of different voxels. Images with ’$T_1$-weighted’ contrast (Figure 1D) allow for easy distinction of different tissue types concerning the evaluation of anatomy, whereas ’$T_2$-weighted’ images emphasize locations with fluids, which is of use in the detection of many pathologies.

Conventionally, MRI has been most used for providing these sort of views into static or slow changing features in the body. Slow changes can be studied in longitudinal comparisons of the same type of images from two or more time points. However, sufficiently fast imaging sequences, such as echo planar imaging (EPI; Mansfield 1977), have made tracking of faster changes feasible. For example, currently EPI is used in fMRI (c.f. EPI image with BOLD contrast in Figure 1E) to measure the whole brain about every two seconds (Figure 1F), a pace which enables useful experiments within suitably short imaging sessions. Moreover, emerging new imaging approaches, e.g. ”MR-Encephalography” (MREG; Hennig et al. 2007, Zahneisen et al. 2012) or multiplexed EPI (Feinberg et al. 2010), can further reduce sampling intervals to the order of hundreds or even just tens of milliseconds.
2.1.2 Measuring brain activity with BOLD fMRI

FMRI measures brain activity indirectly through changes in the magnetic properties of the blood, which are reflected in image contrast (Figure 1E-I). In this way it differs from techniques measuring electrical brain activity such as electroencephalography (EEG) (Swartz & Goldensohn 1998) and magnetoencephalography (MEG) (Cohen 1968). The first fMRI experiments involved using injected contrast agents which changed the relaxation properties of blood (Belliveau et al. 1991). However, soon after most studies started using blood oxygen level dependent (BOLD) image contrast (Ogawa et al. 1990a,b), which is still the most utilized approach. In BOLD imaging, the blood itself serves as an endogeneous contrast agent (Figure 1G-H), which excludes the adverse reactions which are a possibility when using external contrast agents. The tradeoff is that the signal-to-noise ratio (SNR) is not as high.

BOLD contrast reflects changes in the relative amounts of haemoglobin that carries oxygen, and haemoglobin that does not. The former compound (oxyhaemoglobin) is diamagnetic, whereas the latter (deoxyhaemoglobin) is paramagnetic (Pauling, Linus & Coryell, Charles D. 1936). Being paramagnetic, the deoxyhaemoglobin distorts the magnetic field locally, which leads to faster desynchronization of spin precession phases, and consequently to a more rapidly decaying RF signal received from tissue and to low BOLD values (Figure 1G). As a result of locally increased electrical brain activity, cellular metabolism is also increased locally, which then leads locally to an increase of blood flow after a small delay. This haemodynamic response produces a local overshoot of oxyhaemoglobin (with respect to levels of deoxyhaemoglobin) due to oxygen not being extracted to tissue at the same rate as its concentration rises in capillaries (Figure 1H). As there is relatively less of deoxyhaemoglobin during the overshoot, the magnetic field experienced by local hydrogen nuclei becomes less distorted compared to the situation before. Consequently, the phases of precession desynchronize at a slower pace, which leads to a stronger received RF signal from or nearby the site of neural activation.

The change of strength in the signal due to a change of oxygenation levels provides the BOLD contrast between different time points. As image acquisitions for the same slice are repeated at a constant interval, this leads to a temporal signal for each image pixel (Figure 1I). The intensity of an individual pixel varies in time, reflecting the changes due to oxygenation levels in the corresponding voxel. Another way of looking at the obtained data is that a 3-dimensional video is obtained about the oxygenation changes since the tomographic images are acquired from multiple positions (Figure 1F).
In state-of-the-art fMRI, it is common to cover the whole brain by using tens of image slices.

There are several limitations concerning interpretation of BOLD signals as they only measure brain activity indirectly (i.e. they do not directly measure electrical activity itself). For example, the neocortex is built mainly of microcircuits containing internally both excitatory and inhibitory feed-forward and feed-back connections (Douglas & Martin 2004). The net electrical outputs of these microcircuits to other areas may be insignificant even when there is significant electrical activity within them leading to increased metabolism, thus to increased blood flow and consequently to detectable BOLD signal elevation (Logothetis 2008). Thus concurrent elevation of the BOLD signal at different locations may falsely imply electrical signaling between them even when there is none.

A BOLD signal elevation related to electrical neural activity also originates from a location nearby but not exactly at the site of electrical activation. Some of the signal comes from and near the capillaries providing the oxygen and nutrients to the neural tissue, but a part of the signal comes from veins downstream that drain the capillaries (Logothetis 2008). This can place BOLD signal variation in images also to pixels which correspond to locations that do not contain concurrently activating neurons. The effect is, however, dependent on resolution. The imaging sequence used also affects the situation. Spin echo (SE) based imaging can compensate for these venous artefact signals, but gradient echo (GE) based sequences detect them (Logothetis 2008). As a tradeoff, SE acquisitions have overall lower SNR, and do not provide any advantage when imaging at lower field strengths (e.g. 3 T, or 1.5 T such as in the experiments of this thesis).

Finally, a BOLD signal has both rather low temporal and spatial (Figure 1E) resolution compared to the timing and spatial extent of the electrical neural activity that is being (indirectly) studied with it. Current temporal sampling rates of 1 Hz or less (used when imaging the whole brain in a typically used 1.5 T or 3 T environment) are very low compared to the electrical 0-80 Hz oscillations detected with EEG concerning local field potentials (LFPs), which have been associated the most with concurrent BOLD signal elevations (Logothetis 2008). However, novel faster imaging approaches, such as the aforementioned MREG, can take BOLD measurements closer to the timeframe of electrical activity in the brain.

Regarding spatial resolution, BOLD imaging, and fMRI in general, offer the best possible spatial distinction between activation sites, compared to other non-invasive brain measurement modalities. FMRI can also cover the whole brain quite uniformly
(in contrast to the limitations of other measurements, for example EEG being mainly suitable for measuring superficial cortical signals). Still, currently the typically used sizes of voxels ($9 - 16 mm^2$ inplane and $5 - 7 mm$ slice thickness) contain millions of neurons making a BOLD signal in an individual image pixel only a gross-measurement of neuronal events (Logothetis 2008). On the other hand, there are practical limitations to improving spatial resolution power as the reduction of voxel size decreases SNR.

In addition to individual data set related considerations, the comparability of two or more BOLD data sets imposes further challenges which hinder analyses. First, MRI in general gives variable data when performed at different imaging sites, even for same subject, which is due to the differences in imaging hardware and software. Second, in practice BOLD signals are driven and affected by a multitude of factors (c.f. Section 2.1.3) that do not remain exactly the same between imaging sessions or even during one session. Third, there are inter-subject differences in physiology, anatomy and behaviour which all affect both the quality and the content of data.

There is no way to compensate definitely for these BOLD signal origin and variability-related issues as they stem from the intrinsic features of BOLD measurements. Instead, the interpretations of BOLD signal findings should be adjusted accordingly, and reviewed with a critical mind (Logothetis 2008), possibly also relying on other complementary measurements. In this thesis, this notion of cautiousness is extended from BOLD fMRI measurement shortcomings to some of the methods that are used to analyze BOLD data. The examples presented in Chapter 3 concerning analysis software (MR1) warrant examination of how method implementations further affect the results acquired from BOLD fMRI data. The evaluation experiments (MR2 and MR3, Studies II and III) show how parameters can affect the results and how their selection deserve careful consideration. Finally, the model related considerations (MR4) imply that the results may be affected throughout a class of methods, meaning that interpretations of related BOLD fMRI results should be adjusted accordingly.

### 2.1.3 Signal artefacts

In addition to restrictions caused by the aforementioned fundamental properties of BOLD imaging, BOLD signals are also confounded with artefact signals and noise that are not related to the BOLD effect itself. For example, motion of the head during imaging can be corrected by spatial co-registration of volumes (image stacks in Figure 1F) obtained at different time points (c.f. e.g. Jenkinson et al. 2002). However, even
after that, significant variations remain in observed BOLD measurements (Maxim et al. 2005) due to the "spin (excitation) history" (Friston et al. 1996) which refers to nuclei experiencing a change in magnetic field due to head motion and which consequently affects temporal BOLD signal in each voxel. Even though inflow effects (blood flow in and out of slice while imaging it) are insignificant for the currently widely used fMRI sampling intervals of several seconds (Gao & Liu 2012), they may affect the new fast imaging methods which use subsecond temporal resolution (as in e.g. MREG). BOLD contrast variations also contain changes due to global physiological effects, such as cardiac pulsation and the oxygenation cycle of blood in lungs (Glover et al. 2000, Birn et al. 2006, Shmueli et al. 2007), which are not directly related to local neural phenomena.

These shortcomings, however, can be addressed to a certain extent by applying imaging-time or retrospective techniques. Head motion can be compensated for by applying prospective readjustment of magnetic field gradient directions during imaging, as in, for example, the PACE technique (Thesen et al. 2000). Alternatively, spin history corrections (c.f. e.g. Friston et al. 1996) can be applied retrospectively by removing motion estimates from the data, and the time frames containing the most motion can be scrubbed from the data (Power et al. 2012). "Motion scrubbing" is reported to be necessary even when using PACE (Power et al. 2012). Similarly, estimates of global physiological signals can be modeled, and the corresponding correlated portions of data nullified as in for example Glover et al. (2000), Birn et al. (2006). Also singular gross estimates of global signal confounds, e.g. global volume-wise mean BOLD, have been used but they have given rise to a still-continuing controversy (Murphy et al. 2009, Cole et al. 2010) about whether they actually induce further errors in the analyses.

Overall, fMRI and BOLD measurements require multiple pre-processing to improve the data for analyses (Strother 2006). Analyses require also additional preparation steps, depending on the method used. This leads, as a whole, to quite complex networks of processing stages where for each step there are often several options. In the end, it is important to know how each processing affects the final results of analyses. Good research practices involve manual inspections of how different steps have succeeded.

On the other hand, pre-processing related steps, mostly common to all types of fMRI studies, are relatively well studied and recently also evaluated as a whole (Churchill et al. 2012a,b). These latest evaluation results suggest that pre-processing settings customized for individual data may be beneficial compared to using a fixed set of operations for all data, which is the current practice. This thesis adds to these considerations by extending
the existing evaluations of pre-processing related effects with evaluations regarding the effects of the analysis methods themselves in the RS-fMRI context. The results obtained from Study II demonstrate the viability of analysis specific pre-processing steps in different scenarios (MR2), and the results from Study III whether certain analysis options are relevant and under which conditions (MR3).

2.1.4 **Spontaneous brain activity, resting-state fMRI and resting-state networks**

"Resting-state" (RS) is a misnomer for describing brain activity. However, RS-fMRI has become an established term referring to experimental setups which aim at studying "spontaneous" (or "intrinsic") brain activity through spontaneous changes observed in fMRI data. This is in contrast to the "stimulus", "task" and "challenge" based fMRI studies in which, usually, BOLD contrast levels in controlled experiment conditions with pre-defined timing are compared with each other. For example, in task studies, voxel values may be compared between repeated periods of finger tapping and pause to localize in which brain areas BOLD levels elevate with respect to such a motor function (Biswal *et al.* 1995).

Brains are in a constant state of high-level metabolic activity. They make up only 2% of body weight, but consume about 20% of energy, of which at least up to 80% have been indicated to be related to cycling (i.e. restoration) of neurotransmitters (Raichle & Mintun 2006) consumed in neural signal propagation. Moreover, task induced local increases of energy consumption can be as little as 1% (Raichle & Mintun 2006). In this respect, ‘rest’ in RS-fMRI should be understood as the studied subjects being as minimally stimulated and as minimally cognitively engaged (to any specific thing) as possible during data acquisitions.

Spontaneous, continual changes both in electrical brain activity and in blood oxygenation levels have been observed and studied for decades before the conception of RS-fMRI (Snyder & Raichle 2012) - however, largely separately from each other. Work by Biswal *et al.* (1995) was the first to show that low frequency components of different BOLD RS-fMRI voxel time series, which were obtained from motor cortex, correlate within and between the anatomically disjoint areas in both hemispheres. Furthermore, these results were anatomically mostly in agreement with locations revealed by the fMRI finger tapping experiment designed for mapping cortical areas significant to motor functions.
Prior to RS-fMRI becoming a mainstream interest, these spontaneous, mainly relatively slowly occurring (Cordes et al. 2000, 2001) temporal changes (or "fluctuations") in BOLD contrast levels (Figure 1I) were regarded only as noise, signal artefacts and nuisance factors. The main concern was to remove their effect (Friston et al. 1995, Worsley & Friston 1995) as efficiently as possible because otherwise they distort statistics in stimulus fMRI studies (Xiong et al. 1996, Zarahn et al. 1997, Purdon & Weisskoff 1998). However, interest in analyzing RS-fMRI data gradually increased (Lowe 2010) both from the perspective of data mining to detect structured patterns that are potentially of neuroscientific relevance, and concerning modelling the noise in stimulus studies in a more rigorous manner. Concerning the neuroscientific and cognitive relevance of spatio-temporal activity patterns detected in RS-fMRI, a seminal demonstration of credibility was given by RS-fMRI detection of the "default mode network" (DMN) (Greicius et al. 2003) which had been previously determined with more quantitative positron emission tomography (PET) imaging and areas of which were known to deactive during a cognitive task (i.e. they are more active during RS) (Raichle et al. 2001).

Detecting these "resting-state networks" (RSNs) (also called "intrinsic connectivity networks" or ICNs), i.e. anatomically and cognitively plausible sets of brain areas with synchronous BOLD level variations (such as DMN), has become the most prominent neuroscientific application of RS-fMRI (c.f. e.g. Beckmann et al. 2005, Fox & Raichle 2007a, Biswal et al. 2010, Cole et al. 2010, and also Study I in this thesis). These kinds of "connectivity analyses" are also the main objective and application of spatial exploratory analyses of RS-fMRI data that are discussed in this thesis. The ability to detect similar types of RSNs consistently between and within subjects has been demonstrated through numerous studies (van de Ven et al. 2004, Damoiseaux et al. 2006, Shehzad et al. 2009, Biswal et al. 2010, Zuo et al. 2010b, Chou et al. 2012, Guo et al. 2012).

Due to accumulated evidence from multiple studies and experimental setups, RSNs are perceived to relate to core perceptual and cognitive functions and processes in the brain (Cole et al. 2010). Namely, in addition to being localized to grey matter (Beckmann et al. 2005, Damoiseaux et al. 2006), which is responsible for most of the complex neural information processing in brain, the brain activities detected with RS-fMRI have also been shown to correspond anatomically to functional specialization determined with stimulus and task fMRI (Smith et al. 2009). Recently, the research has even reached a point where all the grey matter in the brain has been parcellated into
functional units based on temporal correlations in BOLD RS-fMRI (Yeo et al. 2011, Buckner et al. 2011, Choi et al. 2012). In addition to studies of normal brain function, also pathology related changes in RS have become very popular topics (for reviews c.f. Greicius 2008a, Zhang & Raichle 2010, Fox & Greicius 2010).

These results from RS- and RSN fMRI research provide important information which complements findings obtained from more structured fMRI (stimulus, task etc.), other measurement modalities and structural brain analyses (Cole et al. 2010). Compared to RS imaging, the highly structured studies are often limited in their coverage of various brain functions. Other functional measurements, such as functional near infrared spectroscopy (fNIRS) (Chance et al. 1993, Laguë-Beauvais et al. 2012), are often limited in their anatomical coverage. One exception concerning structured fMRI are the natural stimuli (Bartels & Zeki 2004, Malinen et al. 2007, Ylipaavalniemi et al. 2009, Pamilo et al. 2012) which can be used to provide data regarding multiple cognitive processes with a single imaging session.

However, despite their advantages and potential, RS and RS-fMRI remain challenging research areas due to the ambiguity of how spontaneous brain activity should be measured and studied. For example, data acquisition varies. A common denominator for most studies is the minimization of auditory perception and head motion, and instructing the subject to relax and not to think of anything particular. On the other hand, there are three main approaches related to visual stimuli and focus: some studies prefer imaging while eyes are closed to minimize visual input and others gather data while eyes are open, while a third kind of experiment contains a visual fixation task where visual input and the subjects’ focus are controlled by instructing them to look at a crosshair, for example. These sort of differences are known to impact the data significantly (Marx et al. 2004).

The assumptions about relevant data contents may not be as well-founded as they may appear. For example, the original observations about RSN frequencies being low (Cordes et al. 2000, 2001) has recently been challenged (Niazy et al. 2011). In addition, most existing studies assume that the properties of RS-fMRI data do not change significantly over the course of an imaging session. Only rather recently has effort been focused on temporal non-stationarities in the data (Chang & Glover 2009, Grigg & Grady 2010, Kiviniemi et al. 2011, Majeed et al. 2011, Jones et al. 2012, Smith et al. 2012).

The aforementioned assumptions and experimental choices regarding RS-fMRI significantly affect RSNs and other findings observed with it. There are also great
differences in results due to the choice of analysis approach (Ma et al. 2007, Biswal et al. 2010, Joel et al. 2011, Guo et al. 2012), which further decreases their comparability and interpretability. As there are many potential sources of error in RS-fMRI studies, many of which cannot even be counteracted, it is especially important to increase the quality of analyses wherever it is possible (in addition to the more conventional efforts of improving data acquisition). The guidelines formulated (DP1, 3.2.1) and future suggestions given in this thesis aim at reducing the unwanted variability of RS-fMRI results regarding exploratory analysis approaches discussed in the next section.

2.2 Spatial exploratory analyses

Exploratory analyses can be seen as a collection of approaches which probe data for phenomena, explaining it without incorporating pre-defined explanatory hypotheses to be tested. In this thesis, spatial exploratory analyses are used to refer to methods which produce, in an unsupervised manner, spatial maps that show locations for some phenomena of interest with respect to anatomy, e.g. maps describing RSNs as illustrated in Figure 2. "Exploratory analysis" is less common terminology in fMRI research, unlike "connectivity analysis" which is often used to refer to some of the same analysis approaches. However, use of the former seems more suitable with respect to referring also to analyses which do not produce connectivity maps per se.

It has been recently noted and projected (Smith 2012) that resting-state research, along with fMRI connectivity studies in general, may be changing focus from determining functionally relevant brain areas to analyzing their interactions. Analyses of those interactions require much more than just spatial descriptions, but providing exploratory results still continues to be an important basis for such experiments. For example, RS-fMRI based parcellations can provide function related regions-of-interest (ROIs) with associated temporal signals (Figure 3) for such analyses (compared to using time courses extracted using anatomical ROIs, possibly based on generic brain atlases). There are also evident practical and complementary aspects regarding the continuing utility of spatial analyses (Smith 2012). In order to serve such purposes, also the methods used for spatial exploratory analyses require further studying and development, to an even larger extent than before. This thesis delivers just such research results (MR1-MR4). On the other hand, the developmental proposals provided (DP1, DP2) support future studies concerning methodology.
Fig 2. Example brain maps of activity/connectivity/BOLD effects given by sICA in Study III: RSN-related maps with statistically significant brain activation (red and blue) overlaid on subject anatomy (T1-weighted image). Top-down: 1) executive brain functions related map, 2) medial and lateral visual cortex map, 3) posterior default-mode network map, and 4) and anterior default-mode network map. Coordinates are millimetres in the Montreal Neurological Institute (MNI) standard space. (Published with permissions from Elsevier.)
The following sections describe in detail spatial ICA, which is both the most prominent approach to spatial exploratory analyses of RS-fMRI data and as such the focus of interest in this thesis. Also other spatial exploratory methods are discussed briefly.

2.2.1 Spatial independent component analysis

ICA and blind source separation in fMRI

Independent component analysis (ICA) is a class of methods in a superset of blind source separation (BSS) approaches. BSS aims at determining underlying and unobserved source signals (latent variables) that lead to observed data when mixed in some unknown manner. The BSS problem can be modelled and formulated as follows (Comon & Jutten 2010): let

\[ \mathbf{x}(i) = \mathbf{A}(i) \mathbf{s}(i), \]

where \( \mathbf{x}(i) = (x_1(i), \ldots, x_P(i))^T \in \mathbb{C}^P \) are the observed, stationary signal samples having zero mean and indexed with \( i \), correspondingly \( \mathbf{s}(i) = (s_1(i), \ldots, s_N(i))^T \in \mathbb{C}^N \) are the source signal samples and \( \mathbf{A} \) is an unknown transformation (which is not
necessarily linear); determine \( s(i) \), and consequently also \( s \). In order for BSS not to remain an ill-posed problem, BSS setups make assumptions about the set of source signals, the transformation or both.

The BSS problem was formulated in the early 1980’s and the method development started around the mid-1980’s, and currently there are numerous approaches and applications of BSS in various fields (Comon & Jutten 2010). These include (among other things) biomedical applications in general (EEG, MEG, fMRI and electrocardiography or ECG), and audio (speech and music) and telecommunication related applications.

The concept of ICA was introduced in the late 1980’s and crystallized in the early 1990’s (Comon 1994). ICA, as the name suggests, approaches solving the BSS problem by assuming that \( s_j \) are mutually statistically independent. The assumption of statistical independence among the variables means that their joint probability density function (PDF), given that sources admit one, can be factorized to marginal PDFs of individual sources (Comon & Jutten 2010):

\[
p_s(s_1, \ldots, s_n) = p_1(s_1)p_2(s_2) \ldots p_n(s_n).
\] (2)

This requirement of statistical independence between variables is much stricter than that of linear uncorrelatedness only, even though the former implies also the latter. As a result of this, ICA is forced to search for non-linearly decorrelated estimates of \( s_j \), which enables it to separate also sources that are inseparable through correlation (Hyvärinen et al. 2001).

In an fMRI context, ICA offers a multivariate approach to modelling which complements other, primarily univariate methodology. Through this, ICA offers one way of utilizing structure in data to a fuller extent. It also produces results in a data-driven manner compared to traditional, hypothesis based analyses, and thus provides a way of performing data mining on (and consequently exploratory analysis of) fMRI data. This comparison is slightly analogous with what RS-fMRI offers compared to more structured fMRI paradigms (e.g. conventional stimulus fMRI), and corresponds to ICA utilization being more prevalent in the RS-fMRI context than in fMRI in general.

Spatial ICA (or sICA) was first introduced for fMRI in McKeown et al. (1998a,b). In a manner analogous to how RS-fMRI was accepted very gradually, also sICA was initially adopted very slowly (Beckmann 2012), perhaps due to it being a less established analysis paradigm compared to hypothesis based analyses (Friston 1998).
Conceptually, spatial ICA can be seen for example as an extension to spatial PCA (sPCA). SPCA (or alternatively "eigenimage analysis") was used first on PET data (Friston et al. 1993), and later adopted to fMRI analyses, as many original fMRI methods have been. Both sICA and sPCA decompose fMRI data to spatial components or "spatial modes" (i.e. ICs or principal components, PCs) which serve as brain maps of voxels that have mutually coherent temporal changes in their values (e.g. the voxels shown with red and blue in Figure 2 and their respective coherent BOLD signals shown in Figure 3).

The idea is that the maps, when related to brain functions, can be interpreted to reflect "functional connectivity" (Friston 1994) between different locations in the brain (e.g. between and within the red and blue areas shown in each row of Figure 2). In RS-fMRI analyses, some of these maps may be interpreted as RSNs.

However, in contrast to sICA, which builds on statistical independence, sPCA provides only linearly decorrelated data components. Furthermore, sPCA has not been perceived to be very accurate compared to sICA with regard to either stimulus fMRI (McKeown et al. 1998a,b) or studies of spontaneous BOLD fluctuations (Kiviniemi et al. 2003), which indicates that non-linear decorrelation of components, used in sICA, is a more suitable criterion concerning fMRI data. In line with these considerations, the use of sICA surpassed the use of sPCA at an early stage.

**ICA model in fMRI**

A conventional sICA model can be defined as a special case of the generative model in (1) that incorporates a linear transformation for real numbers:

\[ \mathbf{x}(i) = \mathbf{A} \mathbf{s}(i), \]

where \( \mathbf{x}(i) = (x_1(i), \ldots, x_{t_{\text{max}}}(i))^T \in \mathbb{R}^{t_{\text{max}}}, \mathbf{s}(i) = (s_1(i), \ldots, s_n(i))^T \in \mathbb{R}^n, i \) is the index for spatial samples (i.e. the voxels/pixels included in the analysis which usually entails all non-zero entries in the image stack like the one in Figure 1F), \( t_{\text{max}} \) is the number of time points and \( \mathbf{A} = [a_{c,e}]_{t_{\text{max}} \times n} \in \mathbb{R}^{t_{\text{max}} \times n} \) is the "mixing" matrix. The analysis then proceeds by estimating \( \mathbf{A} \) and \( \mathbf{s} \) through the estimation of inverse (or pseudoinverse) for \( \mathbf{A} \), i.e. \( \mathbf{B} = \mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \), referred to as the "unmixing" (or "demixing") matrix. In an ICA context \( n \) source signals \( s_j \) are referred to as "independent components" (ICs).

The number of estimated components \( n \), i.e. the sICA model order, can be determined in various ways. For example, RS-fMRI group-level sICAs (Abou Elseoud et al. 2010)
show that a high model order produces more detailed descriptions of signal components, whereas a low model order offers more general descriptions combining several data features into the same ICs. In this respect, the model order may be set according to the type of (neuroscientific) research question that is being addressed (Beckmann 2012).

Besides choosing an appropriate description level, one may question the number of signal components that we can reasonably estimate due to SNR considerations. There are many methods for estimating a meaningful model order in this respect. For example, Laplacian evidence for model order ('LAP') (Tipping & Bishop 1999, Minka 2000), the Bayesian information criterion (BIC) (Schwartz 1978), minimum description length (MDL) (Rissanen 1978) and the Akaike information criterion (AIC) (Akaike 1969) and some derivations based on them are common alternatives that have been considered in the fMRI context. The subject of dimensionality estimation has been previously studied concerning fMRI data in Beckmann & Smith (2004), Cordes & Nandy (2006), Li et al. (2007a), Xie et al. (2008, 2009), Varoquaux et al. (2010), Zuo et al. (2010b), Yourganov et al. (2011), Hui et al. (2011), Pendse et al. (2011). Study II in this thesis adds a new perspective to considerations about dimensionality by assessing its temporal changes.

As can be seen from the aforementioned list of BSS and ICA applications, most signals analyzed are in the temporal domain, i.e. $i$ in (1) indexes time. This is in contrast to using spatial samples, as in sICA, which is in general rarer. Consequently, there is less information available concerning ICA with respect to spatial decompositions. This further emphasizes the need for spatial ICA related method evaluations such as the ones in this thesis.

Given the data dimensions, the spatial version of ICA is more feasible for fMRI data than ICA computed for temporal signals. For example, in the data sets used in this thesis, the number of voxels is in the order of tens of thousands, whereas there are only a few hundred temporal samples for each voxel. From the temporal ICA (tICA) point of view, there would be as many channels/variables as there are voxels ($n_{vox}$), and the problem of estimating source signals would be severely underdetermined since $t_{max} \ll n_{vox}$. Consequently very radical data reduction prior to applying tICA would be needed, possibly leaving out important portions of data. With sICA, the data can be utilized to a fuller extent.

Another reason for using sICA instead of tICA is that, in general, the localization of fMRI signals is more specific in the spatial domain, and especially concerning RS-fMRI (Beckmann et al. 2005). Namely, as the statistical independence assumption in the ICA model implies that $s_j$ must be also mutually uncorrelated, estimation approaches usually
also ensure this. Consequently, due to enforcing the uncorrelatedness of source signal estimates, the estimates may not accurately present the whole content of actual sources, if those original sources contain overlap (Beckmann et al. 2005). Apart from perhaps event-related fMRI studies, and some very specific effects, signals in fMRI data cannot be considered to be very temporally local and therefore are prone to being mutually overlapping in time. Especially temporal signals related to spontaneous brain activity are continuously present in the data. Hence the use of tICA is not well-founded.

The spatial extent of most effects is, however, anatomically limited, comprising only a fraction of all voxels included in the analysis. This partly alleviates the problem of mutual overlap of signals in the spatial domain. Even if there is an anatomical overlap between areas related to different source signals, the random values in other voxels contribute to estimations, so that uncorrelatedness between source estimates is achieved without compromising the separability of the sources (Beckmann et al. 2005).

The aforementioned contraindications to the use of temporal ICA in fMRI research may change in the future. Fast imaging (Hennig et al. 2007, Feinberg et al. 2010, Zahneisen et al. 2012), discussed in a previous section, provides many more time points within feasible imaging durations. Also ICA decompositions limited to particular brain areas, such as grey matter (Formisano et al. 2004), can reduce the number of voxels and corresponding time courses. In addition, temporal ICA can be used in combination with spatial ICA (Smith et al. 2012).

The conventional ICA model (3) is a generative one, i.e. mixing values (in A) describe how observations are generated from ICs (Hyvärinen et al. 2001). This restriction in the model restricts also estimation practices, as final estimates are often conditioned so that they reproduce the observations. It is, however, an open question to which extent such restrictions are plausible concerning each application area, for example regarding how processes behind fMRI data should be studied. As seen later from the results in this thesis (Study II), forcing the estimates to satisfy a generative model can lead to distorted descriptions of the biological phenomena being modelled.

Furthermore, it is also generally an open question how good a model ICA is for fMRI data. There is no doubt in the fMRI research field that ICA produces interesting and physically, biologically and even cognitively plausible results, especially concerning RS-fMRI. However, these results can be rarely explained in a comprehensive manner, and they contain also artefactual confounds. As such, the role of fMRI sICA results remain as ad hoc descriptions of data with the need for separate validation or further processing which could be a sign of ICA being a rather suboptimal data model for...
fMRI. Some concerns (Daubechies et al. 2009) along these lines have been proposed concerning whether the whole notion of mutual independence between source signals is actually the key factor driving the estimations. Also, even though the aforementioned description level based model order determination is practical, and neuroscientists seem to be perceive multiple description levels as useful, such an ambiguity in setting model parameter also raises preliminary questions about whether some other models (different from conventional ICA) would be more complete in describing RS-fMRI signal components. The question of sICA model validity is studied also in this thesis. The results (MR4) point out several observations that do not support the use of the conventional sICA model for RS-fMRI data.

**Practicalities of spatial IC determination**

Practical IC estimation methods are often based on criteria different from (2) as formulations based on it would require explicit and accurate information about sources (Comon & Jutten 2010). Such information is not usually available, especially concerning application domains such as RS-fMRI in which ICA is used for data exploration. However, also the ICA model (3) itself sets fundamental constraints on how well sources can be determined. Because the scale and the sign of each column of $A$ can be changed arbitrarily as long as compensating change is applied to the corresponding $s_j$, ICA methodology in general can determine sources only up to sign and scale. Scales and signs must be determined with additional criteria.

Many practical ICA algorithms contain more or less accurate approximations of information theoretic quantities that are used to justify their applicability as BSS approaches. Moreover, as described in following sections, some algorithms, such as the ones used in fMRI sICA, contain several parameters for which some feasible values must be chosen. The performance of different algorithms and their parametrizations may also depend on data. Thus it is ultimately dependent on the number of empirical evaluations whether optimality of the IC estimated can be argued. This thesis provides such information with regard to sICA performed on RS-fMRI data.

**Spatial PCA pre-processing for spatial IC estimation**

Estimation in ICA can be simplified with certain pre-conditioning of data. Even though fMRI sPCA is not used to obtain results per se, as discussed previously, PCA is often
used to pre-process the data prior to ICA. It is also commonplace in the case of fMRI sICA, where PCA is both a default assumption in the employed ICA models and a default option in the utilized software implementations. On the one hand, the results of PCA computation can be used to whiten the data, which enables the use of simpler formulas and fewer estimated parameters in ICA. On the other hand, PCA can also be, and often is, used for reducing the data set to a subspace containing most of the signal, which both improves SNR and further reduces the number of parameters that ICA needs to estimate. However, as PCA is also data-driven, subspace constraints placed on ICA estimation by PCA become a crucial element affecting final IC estimates. The PCA subspace comparisons in this thesis (Study II, MR2 and comparisons between windowed and full data) provide viewpoints to such considerations.

When PCA is incorporated as a pre-processing step the sICA model in (3) becomes:

\[ z(i) = Vx(i) = VAs(i) = W^T s(i), \]  

where \( z(i) = (z_1(i), \ldots, z_n(i))^T \in \mathbb{R}^n \) and \( V = [v_{ij}]_{n \times t_{max}} \in \mathbb{R}^{n \times t_{max}} \) is the whitening matrix. \( D_{1 \ldots n} = \text{diag}(d_1, \ldots, d_n) \) and \( E_{1 \ldots n} = (e_1, \ldots, e_n) \in \mathbb{R}^{t_{max} \times n} \) are the largest \( n \) eigenvalues and corresponding eigenvectors of data covariance matrix \( E\{xx^T\} = ED^TE^T \), which are computed in PCA step. \( W^T = VA \) is the mixing matrix in whitened space (and possibly reduced subspace if \( n < t_{max} \)). \( x(i), A, s(i), n \) and \( i \) are same as those in (3).

Instead of \( B = A^T \) in (3), \( W \in \mathbb{R}^{n \times n} \) is estimated by ICA in (4). When PCA is utilized with \( n \ll t_{max} \), the number of parameters to estimate is reduced significantly for ICA. Furthermore, the uncorrelatedness between signal sources \( s_j \) (implied by the assumption of statistical independence between them) leads to \( W \) being orthogonal (Hyvärinen et al. 2001). As such, only \( n(n-1)/2 \) parameters (matrix elements) need to be estimated in ICA instead of the full \( n^2 \).

**Spatial IC estimation in fMRI**

Concerning sICA estimation itself, there are two approaches that are widely used for sICA in the fMRI context, "Infomax" (Bell & Sejnowski 1995) and FastICA (Hyvärinen 1999), first applied for fMRI sICA in McKeown et al. (1998a,b) and (Suzuki et al. 2000, Kiviniemi et al. 2003) respectively. FastICA is evaluated in this thesis while Infomax is not.

48
Both methods are iterative optimization approaches and have been originally derived as learning rules for artificial neural networks (ANNs). This is in contrast to tensorial methods based on algebraic origins such as Fourth-Order Blind Identification (FOBI Cardoso 1989) or Joint Approximate Diagonalization of Eigenmatrices (JADE Cardoso & Souloumiac 1993). JADE and also many other algorithms are available in the GIFT analysis package (Correa et al. 2007) readily available for fMRI analysis, but they seem not to be utilized in most reported studies. The use of only FastICA or Infomax in most studies makes comparisons between results more plausible, but on the other hand that also places extra expectations on their feasibility and performance concerning fMRI data. FastICA evaluations in this thesis (MR3 from Study III) add to previous generic and fMRI specific evaluations.

The operation principles behind Infomax and FastICA can be understood through the minimization of mutual information (MI) between estimates of $s_j$ (Hyvärinen et al. 2001). If the estimates are denoted by $y_j$, then MI between them is defined as follows

$$MI(y) = MI(y_1, \ldots, y_n) = \sum_{i=1}^{n} H(y_j) - H(y), \quad (5)$$

where $H$ is the differential entropy. $MI(y)$ is nonnegative and $MI(y) = 0$ iff $y_j$ are statistically independent. When $W$ in $y(i) = Wz(i)$ is estimated so that $MI(y)$ is minimized, statistical independence between $y_j$ is maximized. Full statistical independence may not be realized, though.

Infomax refers to maximization of "information transmission" between ANN inputs (in this case observations $x_j(i)$) and outputs when mapped through a nonlinear function $g$ parametrized by $W$ (Bell & Sejnowski 1995). $g$ is chosen to approximate the cumulative density function (CDF) for distributions of assumed signal sources $s_j$ (McKeown et al. 1998a). Optimization of $W$ corresponds to finding the optimal slope parameters for $g$ so that output entropy $H(g(x))$ is maximized, which corresponds to maximization of information transmission (Bell & Sejnowski 1995). If $g$ approximates CDF for $s_j$, then the maximization of $H(g(x))$ leads to maximum log-likelihood estimation of sources $s_j$ which, on the other hand, is equivalent to minimization of $MI(y)$ (Hyvärinen et al. 2001).

The original algorithm (Bell & Sejnowski 1995) is feasible for estimation of supergaussian signal sources. Some sources, such as those corresponding to BOLD fMRI signal artefacts, may be subgaussian instead. An extended Infomax algorithm (Lee et al. 1999) can estimate both sub- and supergaussian sources but, concerning which
version researchers may be using, at least in GIFT software, the original algorithm is the default option (according to GIFT version 1.3 documentation).

The form for update rule in both the original and the extended Infomax algorithm is

$$\delta W = \kappa (I - f(y(k))) W,$$  \hspace{1cm} (6)

where $y(k) = Wz(i)$, $k$ belongs to a subset of all sample indices (drawn randomly without substitution) and $\kappa$ is a learning rate which is gradually reduced until overall changes to $W$ become smaller than some predefined stopping threshold $\epsilon$.

Accurate approximation of true probability densities of $s_j$ is very difficult. The intuitive notion behind FastICA is the maximization of non-gaussianity of signal source estimates $y$ (Hyvärinen et al. 2001), which does not require very accurate probability density approximations. In the ICA model (3), the observations $x_j(i)$ are weighted sums of signal sources $s_j(i)$, and due to the central limit theorem, the former must be more gaussian than the latter. Non-gaussianity can be measured through negentropy

$$J(y) = H(y_{gauss}) - H(y),$$  \hspace{1cm} (7)

where $H$ is entropy and $y_{gauss}$ is a gaussian random variable with the same covariance matrix as $y$. $J(y)$ is nonnegative and $J(y) = 0$ iff $y_j$ are gaussian (Hyvärinen et al. 2001).

However, as computation of $H$ depends on knowing $p_j(y_j)$, in practice $J$ needs to be approximated. Approximations can be derived through probability density expansions based on polynomials of high-order cumulants, such as skewness or kurtosis, or "non-polynomial moments", such as formulations of hyperbolic cosine or exponential function (Hyvärinen et al. 2001). Non-polynomial moments provide robustness to estimation as they are not as sensitive to extreme values as high-order polynomial elements in cumulants. Preferably, both an odd and an even function should be included to approximations in order to separate underlying source signals $s_j$ both according to possible asymmetry and spikedness/sparseness in their distributions. In practice, however, only single function based approximations seem to be available in the widely used software implementations of FastICA (official FastICA Matlab and C source code releases, FSL Melodic, GIFT). The approximation functions are commonly referred to as "contrast functions".

$MI$ in (5) can be expressed as a sum of $-J(y_j)$ and a constant when $y_j$ are set to unit variance (Hyvärinen et al. 2001). Thus maximization of $J(y)$ through maximization of
source estimate non-gaussianities leads also to minimization of $MI(y)$, and consequently to maximization of statistical independence between source estimates $y_j$.

In FastICA, the optimization of $W$ is based on fixed-point update rules derived from an approximation of Newton’s method (Hyvärinen et al. 2001). In its vector form and for whitened data the FastICA update rule can be represented as follows

$$
\delta w_j = w_j (1 + E \{ g'(w_j^T z(i)) \}) - E \{ zg(w_j^T z(i)) \},
$$

(8)

where $w_j$ is the jth column vector of $W$ and $g$ is the first derivate of some chosen contrast function $G$ that is used as an approximation of negentropy $J$. After each iteration, the $w_j$ need to be re-orthogonalized with respect to each other, as statistical independence implies uncorrelatedness between $s_j(i)$ corresponding to orthogonality of $W$ (Hyvärinen et al. 2001). As in the case of Infomax, iterations are repeated until changes to $w_j$ become smaller than some predefined stopping threshold $\varepsilon$. Concerning software implementations, however, the results in this thesis (MRI) show that in practice FSL Melodic uses update formula different from (8). Namely, this FastICA variant implemented in some versions of Melodic does not strictly center the data, even though derivations of update formula in (8) assume that. The evaluations in Study III further prove the relevance of this difference.

If all $w_j$ are updated at each iteration, $W$ is re-orthonormalized symmetrically. FastICA also offers a "deflation" approach (referring to shrinkage of the searched data subspace) where $w_j$ are estimated separately and later ones are re-orthogonalized with respect to earlier ones. This possibility to estimate even only a single $w_j$ links FastICA to general projection pursuit (Hyvärinen et al. 2001). Even though Infomax can also estimate single $w_j$, BSS functionality in Infomax is achieved only through simultaneous estimation of whole $W$ (Bell & Sejnowski 1995). However, in practice mainly symmetric estimation seems to be used in most fMRI sICAs, also regarding FastICA. When evenly distributed errors are preferred, symmetric orthogonalization is a more suitable approach because ordered estimation in the deflatory case makes preceding IC estimates more accurate than later ones. In addition to the accuracy aspect, the rare use of a deflatory approach may be also partly due to the fact that the symmetric approach is the default option, for example in FSL Melodic. The official FastICA release by Hyvärinen uses a deflatory approach by default, but the official release does not seem to be used in fMRI sICA as much as fMRI specific analysis software, such as FSL Melodic.
Another notable difference between FastICA and Infomax is the update step size. In FastICA it is one, whereas in Infomax $\kappa$ must be adjusted. If there are convergence issues regarding FastICA, it can be changed from the Newton method approximation to be closer to gradient methods (such as Infomax) by using a stabilized version with adjustable step size $\kappa$ (Hyvärinen 1999). However, the stabilized version does not seem to be used in fMRI sICA, perhaps for the same reasons as in the case of deflationary approach. It is not implemented in FSL Melodic, it is optional in the official FastICA release, and GIFT uses the official code. Moreover, the tradeoff of deploying gradient methods is an increased convergence time, but concerning fMRI sICA literature there does not seem to be relevant issues regarding either convergence time or lack of convergence. Instead, a more relevant issue might be differences affecting IC estimates (Correa et al. 2005, 2007).

FastICA is usually used in "batch"/"block" mode (also referred to as offline learning) utilizing at once many data points, in fMRI sICA practically all of them. In constrast, $k$ in (6) indexes only small randomly drawn batches of data concerning Infomax iterations. Since the learning rate $\kappa$ is adapted in the course of estimation, different data points contribute more randomly to final estimates compared to standard use of FastICA, where averages in (8) are over all samples. The use of small batches enables quick adaptation relevant, for example, to "on-line" adaptation of estimates in a non-stationary environment (Hyvärinen 1999) where only a limited sample may also be available at a given time, but fMRI sICA assumes spatially stationary samples, and analyses are usually done after complete data acquisition ("off-line" learning). In this sense, there does not seem to be a good argument for utilizing online learning approaches in an fMRI sICA context. Using it also raises a question of how much this additional variability during iterations affects the final estimates. It has been reported (Duann et al. 2005) that Infomax provides stable estimates in this regard, at least concerning visual stimulus fMRI. In this thesis, an analogical stability analysis was performed regarding FastICA and RS-fMRI (MR3, Study III).

In addition to their mutual differences, both FastICA and Infomax use the stopping threshold ($\varepsilon$ concerning (6) and (8)) which may significantly affect the final estimates. Most of the reported fMRI sICA studies do not report $\varepsilon$ and let believe that in general a software provided default value for $\varepsilon$ has been used in them. As convergence of algorithms is very specific to data properties, any evaluations with respect to what is a small enough threshold should be performed in a data domain specific manner. The
effect of $\varepsilon$ is studied in this thesis concerning RS-fMRI FastICA (along with MR3, Study III).

Yet another dimension concerning variability of estimates in FastICA, is the choice of contrast function (concerning $g$ in (8)). Some data concerning fMRI sICA is available in Suzuki et al. (2002), Correa et al. (2005, 2007). In Correa et al. (2005, 2007) notable or conclusive differences cannot be found among the usual contrast functions, but Suzuki et al. (2002) emphasizes the role of data-based contrast function selection.

**Regression analyses in fMRI and interpretation of sICA results**

RS-fMRI sICA results are often compared to those acquired with seed correlation analyses (SCA; Biswal et al. 1995, Greicius et al. 2003, Fox et al. 2005, Margulies et al. 2007, Biswal et al. 2010). Compared to the data-driven detection of RSNs and other effects in sICA, SCA relies on pre-defined hypotheses. It is performed by computing correlation coefficients for time series in each voxel against time series extracted from a pre-defined voxel (such as the red point in Figure 1E) or a set of voxels. These correlations are then tested for significant findings by trying to refute the null hypothesis of no correlation with the seed location. This produces an activity/connectivity map similar to the ones shown in Figure 2.

SCA can be seen as a simplified case of Statistical Parametric Mapping (SPM; Friston et al. 1994) which is a general framework for detecting voxels with temporal signals that match some pre-defined regressor variable or variables. In stimulus and task fMRI studies, these regressor variables are set based on event timing in experimental setup whereas in SCA seed based time courses are used. Additionally some regressor variables may contain temporal signals related to confounding factors such as the signal artefacts discussed earlier concerning fMRI.

SPM describes data with the General Linear Model (GLM; Friston et al. 1994) which resembles that of sICA presented in (3):

$$x(i) = As(i) + e(i),$$

(9)

where $e$ are residual error terms and other notations are the same as those in (3). However, estimation and interpretation are quite different when compared to sICA.

In GLM, each column vector of $A$ contains a pre-defined and fixed regressor (a time course) and $s_j$ represent "parameter maps" each corresponding to a regressor
variable and estimated from the data. In the sICA model (3) both $A$ and $s_j$ are estimated from the data that is being analysed. Furthermore, in GLM individual values $s_j(i)$ are determined voxel-wise in a massive univariate manner as least-squares estimates by computing $B = A^\dagger = (A^TA)^{-1}A^T$, and by projecting observations to a corresponding ‘parameter space’: $s(i) = Bx(i)$. I.e. each $s_j(i)$ is determined using measurements from only one voxel, $x(i)$, and not utilizing data from other anatomical locations. This is fundamentally different compared to sICA in which data from multiple (often all) voxels is used simultaneously in the computation of results.

In the sICA model (3), each column vector in mixing matrix $A$ (mixing vector) contains values associated with one sIC. As the rows of $A$ correspond to time, these values show how prominently the corresponding sIC explains the observations at each time point. As a result, column vectors in $A$ are usually interpreted as temporal dynamics (Figure 3) associated with the same effects which spatial maps (sICs, Figure 2) describe anatomically (i.e. timing of activity or events in the case of signals with neural origin). In this sense, the results of SCA/SPM and sICA are treated in the same way even though they are computed very differently.

It is challenging to compare sICs and SCA/SPM results, given the estimation differences between the sICA and regression approaches. Due to the data-driven nature of ICA, ICs and the corresponding time courses can be expected to deviate from regression results in a sufficiently noisy estimation environment. In addition, concerning specifically RS-fMRI, sICA is sensitive to model order selection and seed selection strongly determines SCA results, which both affect the comparability of RSN estimates between the methods (Cole et al. 2010).

Spatial confounding factors, such as the overlapping RSNs or fMRI noise mentioned earlier, may also distort SCA maps (Cole et al. 2010), whereas sICA, which processes spatial instead of voxel-wise signals, is probably not as much affected. Specifically, sICA provides separate ICs for signal artefacts (Thomas et al. 2002, Kochiyama et al. 2005, Perlberg et al. 2007, Tohka et al. 2008, Beckmann 2012), thus separating structured noise (in varying degrees) from the effects of interest, an important fMRI sICA application. ICA builds on the notion that data is a mixture of various factors. SCA, on the other hand, implicitly assumes that the seed and the rest of data are sufficiently representative of only the effects of interest. Consequently, SCA requires separate data de-noising prior to analysis or by incorporation of artefact related confound regressors within GLM. The advantage of SCA is being able to test a specific hypothesis about functional
connectivity, compared to the ambiguity of how ICs should be interpreted, a difference that fundamentally complicates comparisons between SCA and sICA results.

**Variants and extensions of sICA**

Both fMRI specific and general additions have been proposed concerning the original fMRI sICA model in (3). Given the relatively complex nature of fMRI data, data domain specific customizations of explorative analyses are especially important additions to purely mathematical improvements to the methods. However, given the limited amount of research effort on sICA and on RS-fMRI, there are issues concerning the use of various method variants.

On the one hand, the major research lines concerning some methods can become divergent if different development groups fail to find consensus on best practices (e.g. as in the case of group sICA analyses which are described later). Consequently, the neuroscientific results derived with different flavors of methodology can become incomparable with each other. On the other hand, smaller method development efforts (despite their potential) can simply go unnoticed if they are not incorporated into the widely used tools (e.g. FSL Melodic or GIFT). Their use may remain limited to the research groups that developed them, their results incomparable to those computed with more commonly used methods, and their validity less tested than that of widely deployed methods.

Furthermore, the parameters and other options used for different methodological variants may not be selected carefully with a full understanding of their significance and suitability to a particular analysis. Default options may be used because that convention has the potential to reinforce itself over time, as many, especially smaller, research sites may not have the resources to look into method details very closely. This emphasizes the need to validate that default options are applicable to different scenarios; this is also studied in this thesis. Study II and MR2 provide information concerning sliding window applications of sICA and Study III, and MR3 regarding conventional RS-fMRI sICA.

The following sections describe the most common variants and extensions suggested and used in fMRI sICA.
Probabilistic ICA

Probabilistic ICA or PICA (Beckmann & Smith 2004) is a noisy ICA model and an analysis framework built on top of a noiseless sICA model in (3). It has become a widely used method for performing fMRI sICA in general and one of two widely used lines of sICA based group inference methods is also built on it. The popularity of PICA (and its group analysis extensions) is probably partly because it is readily available as freely distributed FSL Melodic software. However, it also offers significant improvements to the basic fMRI sICA in terms of fMRI specific pre-conditioning of data prior to sICA and statistically meaningful interpretation of sICs.

On the other hand, in practice this customization of sICA for the needs of fMRI research means that PICA has also more parameters and options which need to be decided. There is limited information concerning the impact of those choices and consequently widespread use of PICA and its FSL Melodic implementation make related evaluations potentially very useful. For these reasons, PICA and FSL Melodic were chosen as test cases of exploratory analyses in evaluations and other studies in this thesis (MR1-MR3, Studies I, II and III).

PICA model can be formulated as follows

$$x(i) = As(i) + \mu + e(i),$$

where $\mu = E\{x\}$, $e(i) \sim \mathcal{N}(0, \sigma^2 \Sigma(i))$ are voxel-wise Gaussian noise terms with covariance given by variance $\sigma^2$ and correlation matrices $\Sigma(i)$ and other notations are the same as those in (3). This form of sICA resembles that of GLM in (9) even more than the original sICA model in (3). However, in the form used for GLM in (9), the mean ($\mu$) may be defined as a regressor (column vector) in $A$, whereas in (10) it is defined explicitly through the term $\mu$ because the actual ICA algorithms are normally used only for estimation of $A$ and $s$.

Compared to the sICA model in (3), PICA explicitly defines that ICA estimation is performed in the presence of noise. The estimation is constrained to a "signal+noise" subspace by performing PCA, which leads to operating under the assumption of isotropic noise during ICA estimation in the whitened subspace (Beckmann & Smith 2004). In PICA and its FSL Melodic implementation, the voxel-wise normalization of time series variance is then used as a preprocessing step prior to sICA. It preconditions data so that variance due to non-gaussian signal sources relate to assumed gaussian noise similarly.
across voxels and consequently also standardizes level of contribution to IC estimates from different voxels (Beckmann & Smith 2004) removing unintentional spatial bias in analyses.

Original PICA evaluations (Beckmann & Smith 2004) demonstrate that voxel-wise variance normalization (VN) makes the estimation of the number of data components (PCA cut-off dimension $n$ in (4)) more feasible. These evaluations, however, are very limited in their extent. Moreover, the subject has not been explicitly treated in the literature, even though VN is enabled by default in FSL Melodic, and could affect comparisons between results derived with it and other tools. For these reasons, the effect of VN has been studied in this thesis. Study II shows the validity of applying VN in various sliding window scenarios (MR2).

**Constrained and weighted sICA**

Information from other MRI sub-modalities, such as anatomical imaging, can also be utilized in sICA. For example, as in Formisano et al. (2004) sICA can be anatomically constrained only to voxels in the brain cortex by spatially masking fMRI data with cortical grey matter images obtained from segmentation of anatomical T1-weighted images like the one shown in Figure 1D. Feasibility of spatial constraint then depends on the quality of segmentation, and appropriate masking with respect to it. For example, the relevant effects in BOLD fMRI data are mostly localized to grey matter and coarse spatial resolution cause partial volume effect due to signals from grey and white matter being mixed to the same voxels in border regions. Leaving such voxels out of analysis, or in it, may distort the results, as they may either contain essential data or add unnecessary noise to the analysis. Moreover, from an sICA estimation point of view, spatial constraints reduce the number of spatial samples (voxels) which potentially reduces the estimation quality, even though, for fMRI sICA, there is usually a quite good ratio of voxels to variables.

Also PICA contains the possibility of utilizing spatial information, but through data weighting instead of hard masking. Contributions from some voxels can be weighted more in the calculation of the data covariance matrix used in the computation of whitening transform $V$ in (4). PCA then selects relevant data subspace emphasizing signals from high weight areas. E.g. grey matter maps with probabilistic values for each voxel may be used for weighting. Thus, no actual masking is done, and consequently the sample size in sICA estimation remains the same. However, concerning literature
neither cortical sICA, nor the weighting feature in PICA seem to have been used many times. This may again indicate use of default options when utilizing FSL Melodic.

Constrained versions of sICA have also been proposed concerning the actual estimation steps. Update rules for $W$ can be interleaved with additional steps that, for example, increase the similarity between temporal reference signals and IC time courses in $A$ concerning stimulus fMRI (Calhoun et al. 2005), or between some anatomical references and sICs contained in $s$ (Lin et al. 2010). In practice, also these alternatives seem to be used rather rarely.

Non-linearity alternatives and data transformations

Also non-linearities different from conventional options have been proposed concerning fMRI sICA. Alternatives aiming for ICs with skewed PDFs have been introduced for both Infomax (Stone et al. 2002) and FastICA (Suzuki et al. 2002). Contrast functions can also be adapted (Karvanen & Theis 2004) using the Pearson ICA (Karvanen & Koivunen 2002). These suggestions have not become popular either. Nor do they seem to be part of any popular tools.

Other less used developments include sICA in frequency (Calhoun et al. 2003, Damoiseaux et al. 2006), wavelet (Khullar et al. 2011) and high-dimensional feature (Gruber et al. 2009) domains. In the first approach, voxel time courses are transformed to corresponding temporal signal features prior to applying the sICA used. In the second example, the de-noised coefficients of a spatial 3-dimensional wavelet transform of fMRI data are spatially concatenated prior to sICA. In Gruber et al. (2009) kernel methods (and genetic algorithm based base reduction) are used for extending sICA with the potential to separate also nonlinear mixing. The results, however, do not allow for conclusions about whether a non-linear sICA model would be more suitable than the linear one in (3).

Temporally windowed sICA

Another interesting notion concerning sICA of fMRI data is whether temporal non-stationarities should be accounted for. The conventional sICA model in (3), and the derivations thereof, assume that the same underlying spatial signal sources lead to observations at every time point. Only temporal changes are reflected in the amount of mixing per time point, i.e. between consecutive rows of $A$. Respectively ICs in
the conventional fMRI sICA are effectively ‘spatial averages’ of what signal source estimates might be best for explaining observations through the whole temporal length of observations.

This assumption of temporal ‘stationarity’ with respect to the spatial characteristics of ICs might be invalid concerning the evidence regarding RS-fMRI non-stationarities mentioned in a previous section. Even in the case of structured fMRI (such as stimulus fMRI) background activity or effects observed in data could be changing throughout the imaging session, raising a question on whether any model with fixed parameters is highly optimal for detecting interesting effects.

To address non-stationarity related issues, temporally windowed variants of sICA can be employed. In general, the temporally windowed sICA model can be formulated as follows considering some window length $h$

$$x^{(k)}(i) = A^{(k)}s^{(k)}(i), \quad (11)$$

where $x^{(k)}(i) = (x_k(i), \ldots, x_{k+h-1}(i))^T \in \mathbb{R}^h, s^{(k)}(i) = (s_1(i), \ldots, s_{n_k}(i))^T \in \mathbb{R}^{n_k}, A^{(k)} = [a_{rk}]_{h \times n_k} \in \mathbb{R}^{h \times n_k}, k \in \{1, 2, \ldots, t_{\text{max}} - h + 1\}$ is the temporal position of the data window, $n_k$ is the number of ICs estimated in the $k$th window and other notations are the same as those in (3). I.e. conventional sICA is just performed in temporal segments of data as window position is slid through the length of data, though possibly varying the number of ICs, $n_k$, between consecutive windows. The width of temporal window, $h$, becomes an additional parameter with respect to the conventional sICA model in (3), and also a varying model order $n_k$ may affect analysis outcomes. The effects of varying them remain an uninvestigated subject which is why they have been studied in this thesis (Study II: the applicability of PCA and VN, i.e. MR2 and dimensionality estimation results being presented in Chapter 3).

The windowed approach was first introduced in Esposito et al. (2003) as a means for on-line monitoring of brain responses with respect to stimulus during imaging sessions. Soon after, it was suggested for offline analysis of stimulus fMRI data (Karvanen & Theis 2004), and recently for RS-fMRI data (Kiviniemi et al. 2011). However, these analyses have yet to become mainstream approaches. In order to establish their validity in advance, prior to their possibly wider deployment, some prerequisites for their successful application were studied in this thesis (Study II: the applicability of PCA and VN, i.e. MR2, and the aforementioned dimensionality and window width issues).
Group-level sICA and related analyses

Group ICAs or GICAs are likely the most used extensions of sICA. In these methods, fMRI data sets from several subjects or imaging sessions are combined in some manner prior to the actual sICA step so as to benefit from increased SNR due to the pooling together of low SNR measurements, especially in the case of RS-fMRI. Some assumptions about correspondence between different data sets have to be made and feasible. The relevance and popularity of GICAs also stem from the need to draw neuroscientific inferences within a set of subjects, i.e. on a group-level, instead of being constrained to studies of single data sets.

Originally, sICA results were examined on the group-level by combined ICs after applying the sICA separately for each data set (Calhoun et al. 2001a). In GICAs, data pooling prior to sICA has the advantages of conditioning sICA so that estimated ICs have a similar relation to all data sets involved. In approaches based on individual sICAs, the data driven nature of ICA can make components related to the same underlying effects vary across different data sets, also in a manner not pertaining to the effects of interest. This leads to both interpretability issues and IC matching challenges (Tohka et al. 2008, Zuo et al. 2010b) between data sets. However, some automated solutions based on, for example, IC clustering have been suggested (Esposito et al. 2005) concerning the latter.

Data can be combined in GICAs, for example by concatenating individual data sets temporally (Calhoun et al. 2001b) or spatially (Svensén et al. 2002). Temporal concatenation approaches to GICA are probably the most used sICA methods within RS-fMRI research.

Temporal concatenation assumes that data sets correspond spatially to each other, i.e. that each spatial sample \( x(i) \), as defined in (3), in individual data sets corresponds, at least approximately, to the same anatomical location in the brain. To meet the requirement of spatial correspondence, individual fMRI data sets from different imaging sessions have to be "spatially normalized" to some common coordinate space through spatial co-registration between individual data sets and a template, and resampled to new image matrix prior to concatenation. Otherwise, individual differences between brain anatomy and varying head positioning between sessions will distort the group-level results. However, the spatial normalization will always have residual errors due to variations in image contrast caused by basic MRI artefacts, anatomical differences between subjects and inherent limitations of spatial co-registration approaches (e.g.
linear vs. non-linear). Furthermore, these residual errors may cause systematic errors to analysis results acquired from concatenated data.

Temporal concatenation gives a new $\mathbf{x}(i)$ which contains a multiple of time points compared to an individual data set. Consequently, the number of variables for sICA increases but the number of spatial samples remains the same. As a result, two separate PCA steps are performed prior to sICA to reduce the number of variables to a feasible level with respect to the number of spatial samples available. "First level" ("subject-level") PCA is performed on individual data sets which are then concatenated, and "second level" ("group-level") PCA is used to further reduce the combined data to a significantly smaller subspace within which (group-level) sICA is then performed.

Spatial concatenation is based on the assumption that time points between data sets approximately correspond to the same phenomena. This is the case for example in stimulus fMRI with respect to stimulus timing, but does not apply to RS-fMRI where the effects observed in different imaging sessions are not synchronized. In spatial concatenation, the number of spatial samples increase but the number of variables (time points) with respect to sICA remain the same.

Group-level averaging of data sets (Schmithorst & Holland 2004) or use of tensorial GICA (Beckmann & Smith 2005) make assumptions about both temporal and spatial correspondence across data sets. The number of variables or spatial samples does not change. Tensorial GICA offers three-way decomposition of data sets into common spatial (ICs), common temporal (IC time courses) and subject-specific factors (group-level effects). In practice, the decomposition involves a temporal concatenation scheme and an additional update rule which enforces temporal similarity. Group-level data averaging is also performed in tensorial GICA (Beckmann & Smith 2005), but only to compute on the group-level a common whitening matrix for PCA performed on individual data sets prior to concatenation and sICA. FSL Melodic implementing tensorial GICA, however, also offers ordinary temporal concatenation GICA for example to RS-fMRI studies. However, the results of the software source code analysis in this thesis (MR1) show some deviations between method descriptions and implementation concerning temporal concatenation GICA in Melodic.

GICAs are also used in deriving data set specific effect maps that are statistically tested on the group-level. This enables testing for differences between, for example, normal subjects and patient groups as in Study IV. ICs produced by GICAs describe common features within the combined data set, but do not allow for inferences about differences between individual subjects or subgroups. As a result, an additional layer of
processing has been developed on top of GICA. In Calhoun et al. (2001b) and in later derivations thereof, precise "back-reconstructions" are used to form "subject-level" components and corresponding time courses for individual data sets from the summary ICs produced by GICA on the group-level. These subject-level components can then be pooled on a group-level using GLM in (9) to enable group-level statistical tests.

Another popular approach (Filippini et al. 2009), used also in this thesis (Study IV), uses "dual regression" to compute subject-level effect maps in a less constrained manner. Dual regression entails using GLM twice in a consecutive manner, and separately, for each subject. In this setting $x$ in (9) refers to the fMRI data of an individual subject. For the first GLM estimation, $s$ in (9) are set equal to GICA estimated components $s_{GICA}$ after which subject-level time courses $A$, corresponding to GICA ICs, are estimated by $A^T = (S^T)^\dagger X^T = (S_{GICA}^T)^\dagger X^T$, where $S = (s_i)$ and $X = (x_j)$. In the second GLM estimation, $A$ and $X$ are kept from the first phase, whereas $S$, now signifying the subject-level maps $S_{subj}$ related to original GICA ICs, are estimated by $S_{subj} = A^\dagger X = (XS_{GICA}^\dagger)^\dagger X$.

Again, the resulting subject-level maps can be pooled on group-level using GLM. Using the earlier notation, the group-level modelling of subject-level components/maps means that, for each GICA IC $s_k$ separately, $x_l(i)$ concerning (9) is set to be equal to $S(k,i)_{subj}$ of the $l$th subject and the regressors, that determine different subject groups and other factors, are defined in $A$ in (9). After estimation, $s_j$ in (9) then contain group-level effect maps which can be statistically tested.

As described, use of group-level sICAs introduce additional factors for consideration, such as spatial normalizations, multilevel PCAs or a type of subject-level map reconstruction. These possibly affect the final results of exploratory analyses in addition to the already vast number of parameters and options regarding underlying sICA methodology. Also, it is not obvious how the advantages of combining data sets could be translated to improving analyses of individual data sets concerning, for example, clinical interests or sliding window analyses. For these reasons, group-level methodology and related issues have been excluded from the evaluations performed in this thesis. Instead, the focus of experiments has been on factors affecting sICA on the subject-level (MR2 and MR3) and GICA is only used to provide reference information (frequency spectra from Study I) and as a part of the case study into software implementations of methods regarding MR1.
Repeatability approaches for robust sICA

The optimization step in ICA estimation can lead to varying results, depending on the starting points, and it can also lead to overly fitted estimates due to noise in the data (Himberg et al. 2004, Yang et al. 2008, Ylipaavalniemi & Vigario 2008). In addition to possibly affecting the validity of neuroscientific inferences based on fMRI sICs, this inherent variability may also interfere with comparisons concerning the effects of other factors and parameters in fMRI sICA.

Icasso (Himberg et al. 2004), RAICAR (Yang et al. 2008) and ARABICA (Ylipaavalniemi & Soppela 2009) are frameworks which have been suggested for reducing unwanted variability in the final estimates. They allow for selecting ICs appearing repeatedly during different runs of the ICA method. One can account for effects of different initializations, overfitting or both. An improved IC can be acquired through selecting an estimate from a single ICA repetition that corresponds to a cluster ’center’ ("centrotype") when estimates from all repetitions are pooled and clustered (Icasso and ARABICA). Alternatively, estimates in a cluster can be averaged (RAICAR).

To obtain the clusters, Icasso uses different flavors of hierarchical clustering, an approach similar to hierarchical clustering (complete linkage) has been demonstrated (Ylipaavalniemi & Vigario 2008) concerning ARABICA, and RAICAR utilizes a special algorithm. The original motivation behind these frameworks has been to study the variability of IC estimates. Thus the frameworks serve as useful tools for studying data whether or not the improved IC estimates are used.

Repeatability measures have been demonstrated in the context of stimulus fMRI studies and comparable simulated effects (Himberg et al. 2004, Yang et al. 2008, Ylipaavalniemi & Vigario 2008) but comparisons concerning how different repeatability measures (initialization randomization, bootstrapping or both) affect the results are missing. Moreover, there are no studies about whether they are really necessary in the RS-fMRI context. The evaluations in this thesis (Study III, MR3) include such comparisons.

2.2.2 Other exploratory approaches

Besides different versions of sICA, there are a number of other exploratory methods which are not based on solving the BSS problem. Some of these methods produce brain maps of connectivity (concurrent activity), like sICA and SCA do, while others provide
brain maps of other characteristics. Nevertheless, the results of these methods can be used to explore RS-fMRI data in a data-driven manner. However, usually, unlike sICA, they do not produce estimates of temporal BOLD signals leading to their map estimates (like the ones in Figure 3), which may be one additional reason for the popularity of sICA.


In these approaches, fMRI time courses are transformed voxel-wise to some (typically) multidimensional feature space that is then processed further with some clustering method. Detected clusters of voxels with similarities are then highlighted anatomically and used to make inferences based on the data.

Exploring data with clustering is thus performed in a multivariate manner, as in sICA, but the features are computed prior to that voxel-wise uni- or bivariately. The features may be based on the time course itself, such as the time course spectra used in Mezer et al. (2009), or the mutual relationship between time courses from different locations, such as partial correlations (Cordes et al. 2002) or spectral coherence (Thirion et al. 2006), and potentially additional subspace reduction methods as in Thirion et al. (2006).

In addition to clustering, some straightforward methods have been formulated explicitly for detecting low frequency fluctuations observed in RS-fMRI data. The amplitude of low frequency fluctuations (ALFF, Zang et al. 2007) and fractional ALFF (fALFF, Zou et al. 2008) are purely univariately computed voxel-wise indices that do not utilise relationships between measurements from different anatomical locations. In ALFF, amplitude spectra of voxel-wise time series are averaged below a predefined cut-off frequency to produce the index. 0.08 Hz has been reported as the selected cut-off frequency. In fALFF, the same is done, but the ALFF value is scaled by dividing it with a similar value computed over the whole frequency range, i.e. fALFF describes the relative amount of data in low temporal frequencies.

Compared to ALFF and fALFF, a method known as regional homogeneity or ReHo (Zang et al. 2004) does account for relationships between different anatomical locations but only locally, as the name suggests. In ReHo, the Kendall’s coefficient of concordance
(KCC) is used to compute voxel-wise a statistic describing how mutually similar the time courses are in a local neighborhood around each voxel.

Similarly to SCA, clustering approaches, ALFF, fALFF and ReHo all need careful preprocessing of data prior to analyses. Otherwise signal artefacts and structured noise will dominate data variance too much, and structure in data related to the effects of interest may go undetected. The popularity of sICA over these other choices for explorative analysis may be due to a decreased need for using preprocessing, as applying sICA automatically separates at least some of structured noise from the effects of interest.

The lack of a need to do extensive preprocessing or explicit signal feature selection enables more straightforward exploration of data through sICA. On the other hand, sICA contains more parameters that need to be chosen and considered. Automatic clustering is primarily dependent only on algorithm choice, initialization and selecting a number of clusters, which is similar to determination of the model order in sICA. The only parameter to consider in ALFF-based analyses is the cut-off frequency, and in ReHo the size of the neighborhood for which KCC is computed. Consequently these methods require fewer evaluations regarding method parametrizations themselves compared to sICA, but perhaps more studies concerning the effects of different fMRI preprocessing and feature selections. However, this does not imply that results produced by these methods could not be sensitive to the choice of parameters.

### 2.3 Method evaluations regarding spatial exploratory analyses

The use of sICA variants and extensions, whether default options or not, adds to the need for method evaluations. As most of variants and extensions are fMRI specific, regarding fMRI sICA only relatively little can be learned from for example the ICA evaluations performed outside the fMRI domain. Additional needs concerning RS-fMRI may rise from the fact that quantitative method tests are often performed with stimulus fMRI or comparable simulated data. Moreover, many evaluations seem to compare notably different approaches, such as sICA and SCA or whole sICA frameworks and implementations such as those in FSL Melodic and those in GIFT. Tests with actual RS-fMRI data are mostly qualitative in nature.

Existing fMRI sICA evaluations in the literature include various kind of setups. Specific contrast functions have been studied concerning FastICA (Schmithorst & Holland 2004). Single parametrizations of FastICA and Infomax have been compared
on the subject-level (Esposito et al. 2002, Meyer-Baese et al. 2004) as well as more extensively on the group-level between JADE, Infomax and FastICA with multiple contrast alternatives (Correa et al. 2005, 2007). Also single evaluations regarding the effects of preprocessing have been performed (Churchill et al. 2012b). In addition, GICAs have been compared (Schmithorst 2009).

The reports of evaluations seem limited. Some amount of test results may have been restricted for internal evaluations within research sites and groups. It is unfortunate if there are a lot of unpublished evaluations results. Another viable explanation may be the previously mentioned fact that, given the cross-disciplinary nature of RS-fMRI research, the focus of most studies is applied research seeking primarily novel neuroscientific insights. The problem, however, is that the quality of such results depends on the reliability of the used methods.

Also, despite providing important comparisons between methodological alternatives, the aforementioned evaluations are far from exhaustive concerning the overall size of the parameter space when all the method options and variants described in the previous sections are considered. Consequently, the developmental proposals (DP1) in this thesis include a guideline (GL4) that aims at reducing this shortcoming in method evaluations regarding future of RS-fMRI analyses.

The aforementioned reported evaluations have also been primarily performed with simulated or stimulus fMRI data. Consequently, further evaluation efforts are needed, especially regarding RS-fMRI data. The evaluations presented in this thesis provide some viewpoints on how to perform method evaluations in a more RS-fMRI specific manner.

In Study III and concerning MR3, simulated signals were generated with a strength comparable to RSN IC findings from real RS-fMRI data and added to the real data. This provided a more realistic reference with which to compare IC estimates from different RS-fMRI sICA evaluations compared, for example, to using predefined timing in stimulus fMRI or arbitrary simulated signals as the reference signals. In Study II and regarding MR2 and related results, the PCA subspaces of overlapping parts between consecutive data windows were compared. This demonstrates the possibility to perform evaluations using a relative measure that does not even need static predefined reference signals such as simulated effects or stimulus timings. These kinds of approaches have not been generally used in prior literature concerning method evaluations. The developmental proposals (DP1) in this thesis also address (GL3) this aspect of improving RS-fMRI related method evaluations.
The current literature also implies very limited concerted efforts in method evaluations. This thesis provides practical points concerning increasing community involvement. The proposed (DP1) guidelines (GL1 and GL2) in this thesis suggest principles for constructing analysis software in ways that facilitate participation and carrying out evaluations. The possibilities to advance joint evaluations are discussed also in 4.3.

2.4 Assessing the significance of analysis results

The results of fMRI analyses, especially due to data being noisy, need to be assessed statistically in order to infer which findings are significant (e.g. the red and blue voxels in Figure 2) and which are observed by chance. This is especially true for low SNR RS-fMRI data, and particularly if the results of exploratory analyses are used to make neuroscientific inferences without separate validation experiments. However, assessment may also be needed concerning method evaluations in order to detect which options and factors are relevant and which are not. This issue is explored further in Section 3.2.2 concerning the development proposal in this thesis (DP2).

The following sections review the usual statistical practices that are used to determine significant findings in imaging research.

2.4.1 Statistical testing in analyses of image data

As in the case of spatial exploratory analyses, the results of analyses are usually in the form of brain maps containing a multitude of anatomical locations, i.e. voxels, which need to be assessed for significant findings in them. The amount of the studied effect detected concerning each voxel, e.g. an estimate of $s_j(i)$ in (3) or (9) or the fALFF/ALFF/ReHo value, is transformed to some statistical score, e.g. t-statistic or Z-score. These scores are then thresholded with respect to some chosen criteria on how big false positive (FP) and false negative (FN) detection rates are accepted.

Often the results are tested against a null hypothesis of no effect, in which case only the FP rate can be controlled by computing a score threshold corresponding to the chosen significance level $\alpha$ for test p-values. Usually an $\alpha$ of 0.05 or 0.01 is used, as in many other fields of research.

Traditionally parametric tests, such as the t-test or the F-test, have been used, as in many other fields of research. Nowadays, however, also non-parametric approaches, such as a permutation test (Nichols & Holmes 2002), are available. Permutation approaches
provide increased sensitivity compared to parametric tests because possibly invalid parametric models of data are not assumed. They are used, for example, in conjunction with performing dual regression based on the GICA maps discussed in the previous sections and have been used in this thesis regarding Study IV.

Permutation approaches create an estimate of the test statistic PDF experimentally. The selected statistic, for example the t-score (or some of the cluster-based measures introduced in the next section), is computed repeatedly in a Monte Carlo experiment where classes or labels of data are randomly sampled. The threshold value for the statistic corresponding to the preferred $\alpha$ can then be determined from the PDF estimate acquired regarding score values or alternatively p-values can be assigned to scores in all voxels.

The probabilistic ICA framework (Beckmann & Smith 2004) described in earlier sections has introduced also a possibility to model modes in PDF, corresponding to both null hypothesis and an alternative hypothesis. By adjusting the score threshold according to how these modelled sub-PDFs change with respect to each other, FP and FN rates can be controlled with respect to each other. PICA ICs are often (and by default in FSL Melodic) thresholded with score values leading to equal FP and FN rates.

PICA also provides meaningful statistical scoring for IC values (voxel values $s_j(i)$ in (3) and (10)). Residuals $e(i)$ in (10) are used for computing voxel-wise noise estimates and IC values are divided by them. Prior to this approach, concerning sIC, the only convention was to transform IC values to Z-scores by dividing them with standard deviation within corresponding IC maps $s_j$ but that leads to ambiguous FP rates with respect to the choice of $\alpha$ (Beckmann & Smith 2004).

### 2.4.2 Corrections for multiple comparisons

The large amount of voxels to be tested causes problems with regard to multiple comparisons when null hypothesis testing is used. The effective FP rate concerning the whole map of significant voxels, also called the family-wise error rate or FWE, inflates if the statistical threshold level $\alpha$ is set on a per-voxel basis instead of considering all tested voxels at once. Thus multiple comparison correction (MCC) methods are needed.

The p-value concerning FWE can be formulated as follows with respect to $\alpha$ (Brett et al. 2004)

$$P_{FWE} = 1 - (1 - \alpha)^n,$$  \hspace{1cm} (12)
where \( n \) is the number of voxels being tested for significance, and this leads to the following approximation when \( \alpha \) is small

\[
P_{\text{FW E}} \leq n\alpha.
\]  

Consequently, \( \alpha \) needs to be adjusted so that the preferred \( P_{\text{FW E}} \) is acquired.

Adjustment according to FWE sets \( \alpha \) so that the frequency of having any FPs in any voxels corresponds to \( P_{\text{FW E}} \). An alternative, less strict approach utilising the false discovery rate of FDR (Genovese et al. 2002) has also been introduced for controlling, and is used in fMRI studies. FDR focuses on the number of FPs with respect to only the number of all detections, whereas FWE is based on the FP rate regarding all tests. \( \alpha \) is then set according to the p-value which yields on average these false discoveries (FPs per all discoveries) at some predefined rate between 0 and 1. This thesis focuses on FWE related formulations of multiple comparison corrections.

The classical Bonferroni correction for multiple comparisons solves the FWE adjustment problem by setting \( \alpha = P_{\text{FW E}} / n \) (Brett et al. 2004). In principle, the correction would need to be applied over all voxels, of which there can be tens of thousands in regular fMRI data, and consequently \( \alpha \) would be set at an extremely strict level, especially concerning low SNR in fMRI and RS-fMRI. In practice, however, formulas in (12) and (13) assume that tests performed in different voxels are all mutually independent. This assumption is invalid due to correlation between adjacent voxels because of spatial smoothness both inherent in the data and produced by data preprocessing (Brett et al. 2004).

To address MCC issues, functional imaging data specific MCC approaches have been developed. They typically consider the overall effect computed from detected clusters of voxels (for example the spatially joint red and blue blobs in Figure 2), instead of considering voxel-wise statistics (each voxel/pixel considered separately). These approaches utilize the spatial extent of clusters as well as their "mass" (in terms of effect size), and more recently also the Threshold Free Cluster Enhancement (TFCE; Smith & Nichols 2009) approach. Concerning parametric tests, Random Field Theory (Brett et al. 2004) has been developed to enable disciplined thresholding of cluster-based statistics.

Permutation tests offer a straightforward solution to MCC. If the maximum is taken over PDF estimates in all voxels (i.e. maximum over voxels during each permutation), the resulting aggregate PDF estimate corresponds to score values which, at a chosen
significance level $\alpha$, guarantee that the corresponding FP rate is achieved in every voxel (Nichols & Holmes 2002).

If fMRI analyses are performed with various alternatives, such as dual regression with multiple GICA maps obtained with different sICA model orders, as in Study IV, additional MCC would also be needed between them. This inter-analysis aspect of MCC seems not to have been studied before, and is addressed in Study IV and specifically regarding performing evaluations of exploratory analysis methods. The novel method proposed (DP2) in this thesis offers potential solutions to both of these scenarios (inter-analysis and between-evaluation multiple comparison adjustments).
3 The contributions

This chapter presents the original contributions of this thesis which are used to answer the research questions presented in Section 1.2. The first part addresses the methodological research questions, and the second part the developmental questions. Also, the topics are presented in an order whereby prior subjects give foundations to later ones. This way the train of thought regarding method evaluations proceeds from rationale to experiments, from results and experiences to developmental suggestions, and from emerging statistical issues to novel solutions to them.

First, a case analysis of widely used sICA software (FSL Melodic) is presented (3.1.1) in relation to whether the published method descriptions and their implementations match (MQ1). The conclusions from this software code analysis provide further basis for performing evaluations.

Then, the performed experiments are described (3.1.2) and the subsequent sections describe the key findings based on the results (3.1.3). A sliding window experiment, and the corresponding findings address the question of pre-processing applicability (MQ2). They also provide views on model order and on comparability between windows and normal analyses. The need for repeatability analyses (MQ3) is also investigated in a corresponding evaluation experiment. In addition, both repeatability and sliding window experiments also contribute to answering whether RS-fMRI sICA data models are still valid premises for analyses (MQ4).

Regarding developmental aims, guidelines, based on the software case analysis and the experiences from the evaluation experiments, are formulated (3.2.1) to provide the means of improving the quality of the methodology used in the exploratory analyses of RS-fMRI (DQ1). These guidelines, however, raise a new statistical consideration concerning evaluations. In relation to that emerging issue, a novel correction method is proposed (3.2.2) as an approach for improving the quality of method evaluations from a statistical point of view (DQ2).

3.1 Investigations into current methodology used in exploratory analysis of RS-fMRI data

The analyses that were performed range from manual inspection (software source code review) to numerical computations and further evaluations. Concerning the following
sections, Table 1 summarises the objectives and methodology of different analyses and their relation to Studies I - III.

Table 1. The analyses and corresponding studies.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Aims</th>
<th>Approaches</th>
<th>Results</th>
<th>Study I</th>
<th>Study II</th>
<th>Study III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis of FSL Melodic implementations</td>
<td>Identify deviations from descriptions in literature</td>
<td>Code review</td>
<td>MR1</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Power spectra of RSN dynamics</td>
<td>Estimates for RSN IC time courses</td>
<td>GICA, Spectrogram averaging</td>
<td>Reference information</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluations of sliding window dimensionality, PCA and VN</td>
<td>Evaluate stability of in-window PCA, VN and dimensionality</td>
<td>Temporal windowing, PCA subspace analysis, Dimensionality estimation, Real RS-fMRI data</td>
<td>MR2, MR4</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluations regarding FastICA and repeatability measures</td>
<td>Evaluate the need for repeatability measures in estimation</td>
<td>FastICA, lasso; Simulated sources embedded in real data</td>
<td>MR3, MR4</td>
<td></td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

3.1.1 Analysis of FSL Melodic implementations of probabilistic sICA and the temporal concatenation group ICA

As part of Studies I, II and III FSL Melodic source code (Melodic versions from 3.05 to 3.09) was analyzed with respect to how PICA and temporal concatenation GICA are implemented in practice. This was originally done to provide accurate method descriptions for ICA study reports, such as Study I, but the analyses revealed also some methodological steps and choices which had not been documented in the literature. It was noted that Melodic implementation diverges from PICA and sICA concerning three aspects (later denoted by Issues 1-3), and from temporal concatenation GICA regarding one step (Issue 4).
Implementational considerations regarding PICA

As described in Section 2.2.1, the variance of time courses in each voxel is normalized as a pre-processing step prior to analysis so that similar noise levels can be assumed across voxels. In Melodic, however, this voxel-wise variance normalization is not performed iteratively as described in the published method description in Beckmann & Smith (2004). Instead, it is done only once in the beginning and with data thresholded in whitened space, as described in Study I.

The exclusion of iterative normalization (Issue 1) is an implementational choice consistent with the reported observations (Beckmann & Smith 2004) that repeated normalization+sICA -rounds do not offer much additional value regarding results. However, since Beckmann & Smith (2004) remains the main scientifically published method description concerning PICA analyses that Melodic provides, the reader can easily come to the conclusion that Melodic performs iterative variance normalization.

Data thresholding (in whitened space) is not documented in Beckmann & Smith (2004) either (Issue 2). It is used in Melodic under the hypothesis that it excludes the signal from variance estimates used in normalization. In this thresholding scheme, data is first whitened similarly to what is done by sPCA, i.e. by using whitening transform $V$ described in Section 2.2.1. The values of the resulting variables are then set to zero when they exceed a chosen threshold. Finally, the thresholded data is reconstructed in the original space after which voxel-wise variance estimates are computed for time courses. Melodic then performs the variance normalization by dividing original time courses voxel-wise with standard deviations corresponding to corresponding voxel-wise variance estimates.

Thresholding data, prior to the computation of variance estimates mentioned above, is in itself a sensible and necessary step. Having large contributions in the estimates due to the signals of interest would inflate the estimates, which are supposed to reflect primarily only noise, and consequently the signals would be scaled down during variance normalization.

However, these parameters regarding thresholding constitute another factor which is not documented in the published method description in Beckmann & Smith (2004). As documented for the first time in Study I, Melodic uses two thresholding levels in whitened space, 2.3 for single-subject sICA and 3.1 for group-level sICA. Also documented for the first time in Study II, in single-subject analyses, Melodic performs dimensionality reduction, in addition to simple whitening, using PCA to retain at
maximum 30 principal components from which the signal is excluded. Thus the remaining data components are assumed not to contain a signal. These choices for thresholds or the number of retained dimensions (30) seem not to have been documented anywhere besides in the source code itself and the aforementioned contributions of this thesis.

A third difference between Melodic and the original PICA description in Beckmann & Smith (2004) is that, in later versions of Melodic, data is not forced to have zero-mean prior to sICA estimation (Issue 3), which was documented for the first time in Study III. Namely, the PCA step prior to sICA estimation in Melodic retains spatial mean values in the data. As a result, the update formulas for contrast optimization in FastICA, that are described in 2.2.1 and used also in Melodic, do not strictly hold, as they have been originally derived with the assumption of data having zero-mean. In practice, the problem of non-zero mean is alleviated by the fact that voxel-wise time courses are temporally demeaned before PCA, which makes also spatial mean values close to zero. The original PICA description (Beckmann & Smith 2004), however, refers to the use of the standard FastICA algorithm which is not the case concerning Melodic.

**Implemented default options regarding temporal concatenation GICA**

As described in Section 2.2.1, temporal concatenation GICA only assumes anatomical correspondence between multiple pooled data sets. No correspondence between time points (variables) from different data sets is assumed. However, not all the default options implemented in Melodic concerning temporal concatenation are set according to that (Issue 4).

Namely, "first level" (subject-level) PCAs, that are performed before the temporal concatenation step, are supposed to be based on each data set alone (c.f. Section 2.2.1 concerning temporal concatenation GICA). However, unless the "group-level whitening" option is disabled by the user, temporal concatenation GICA in Melodic by default forces a subject-level PCA to be performed based on the whitening matrix $V$ that has been computed from data averaged across all data sets (group-level data averaging). In this respect, the default temporal concatenation setting in Melodic is actually similar to tensorial GICA (described in Section 2.2.1), which assumes temporal correspondence between data sets (in addition to spatial correspondence).

When RS-fMRI data sets are analysed, this convention in Melodic will make subject-level PCAs sensitive to potential temporal similarities between data sets from
different imaging sessions. As there is (or should be) no correspondence between brain activity timing between different RS-fMRI sessions, the most probable origin of temporal similarities should be artefactual. As a result, in some scenarios, subject-level principal components that are not related to the effects of interest may be selected prior to performing "second level" (group-level) PCA and sICA (c.f. Section 2.2.1).

If the automatic model order estimation is used in Melodic temporal concatenation GICA, it will be based on group-level average data as well. This, however, is a feature that cannot be disabled through available options.

Temporal concatenation GICA regarding Melodic seems not to have been used, and at least its implementational details were not documented in the published literature, prior to Study I. According to the official FSL Melodic web pages containing the method description "this approach does not assume that the temporal response pattern is the same across the population". This is true in the sense that one should expect spatial PCA based on group-level averaging of temporally unsynchronized data, such as RS-fMRI data, to have evenly random effects among individual fMRI data sets. Nevertheless, this disparity between an implementation and corresponding method descriptions still serves as an example of how seemingly irrelevant implementational details may not all end up in the documentation.

Conclusion: Method descriptions and implementations should not be expected to match (MR1)

Issues 1-4 identified in the previous sections show how even a widely used tool like FSL Melodic may not strictly implement the methods in the way they are described in the published literature (or even in the documentation of the software). Due to these kinds of realities in academic software development, it is reasonable to assume that any analysis software may contain potentially significant steps that may not be revealed unless the corresponding software source code is analyzed. With respect to research question MQ1, the findings suggest that method descriptions and their corresponding implementations should not be expected to match.

With respect to these aforementioned implementation details that deviate from the originally proposed PICA model (Beckmann & Smith 2004), it was also noted that there were no reports in the literature concerning their evaluations. Because of this, the subsequent evaluation experiments (Studies II and III) were used to study also the effects of using variance normalization and the effects of using Melodic FastICA.
3.1.2 Experimental studies

The following sections describe the methodological details that are necessary for understanding the results presented later. For further details, the reader is referred to the original, extensively detailed articles. All the RS-fMRI data used in the studies cover the whole brain, and were acquired using 4 mm x 4 mm x 4 mm voxels and sampling interval of 1.8 s between consecutive time points.

Power spectrum profiles of RSN temporal dynamics (in Study I)

Power spectrum profiles were computed to describe temporal dynamics related to RSNs identified with sICA. The results are used as supporting information concerning the results of the sliding window study described in the next section.

Profile computation utilised group-level sICA results (temporal concatenation GICA, c.f. Section 2.2.1) computed from real BOLD RS-fMRI data (visual fixation task, c.f. Section 2.1.4) of 55 healthy subjects using a fixed model order of 70 components. The group-level back-reconstructed mixing matrix containing IC related time courses was divided to subject-specific segments (having 250 time points each like the original data sets). The power spectrum was computed for all IC time courses of each subject by averaging the squared spectrogram (128 time point window) of each IC time course. Group-level mean of power spectra were computed for each RSN related IC by averaging the corresponding IC time course power spectra across subjects.

Evaluations of sliding window dimensionality, PCA and VN (Study II)

The stability of PCA retained subspaces, voxel-wise variance normalization, and data dimensionality were studied concerning pre-processing in temporally windowed sICA (c.f. Section 2.2.1). The results are used for assessing whether PCA and VN are applicable as pre-processing steps in sliding temporal window sICAs, and whether model order estimation is needed in them.

Real RS-fMRI BOLD data (280 time points; an 'eyes closed' setting, c.f. Section 2.1.4) of 13 men and 13 women were divided to temporally consecutive and overlapping segments (1-time-point difference between consecutive windows) using 10 different window widths ranging from 20 to 200. The dimensionality of data in each segment was estimated with Laplacian evidence for model order ('LAP') (Tipping & Bishop 1999,
Minka 2000), the Bayesian information criterion (BIC) (Schwartz 1978), minimum description length (MDL) (Rissanen 1978) and the Akaike information criterion (AIC) (Akaike 1969) within the probabilistic PCA (PPCA) framework as proposed in Beckmann & Smith (2004). The methods provide estimates that range from strict (MDL and BIC) to liberal (AIC) covering thus a wide range of model orders that could be realistically employed.

Bases of PCA retained subspaces (defined by $E_1,...,E_n$ concerning (4)) were computed for windowed and full data using the four dimensionality estimates, as $n$ in (4). Furthermore, to study the effects of voxel-wise variance normalization (c.f. Section 3.1.1), this was done separately in three different cases of VN: 1) no variance normalization (later denoted by "No VN"), 2) variance normalization utilizing signal exclusion threshold 2.3 (later referred to as "low threshold" VN), and 3) variance normalization utilizing signal exclusion threshold 3.1 (later referred to as "high threshold" VN).

Temporal evolution of dimensionality estimates were analyzed with respect to the existence of linear trends and regarding power distribution between frequencies. Spectral estimation was performed as described in the previous section, but with a 64-time-point window.

PCA retained subspaces were compared between consecutive windows with regard to their temporally overlapping segments, and similarly between windowed and full data. These comparison were performed using the aforementioned bases $E_1,...,E_n$ and the Frobenius norm-based subspace similarity measure, denoted later by $d_{cw}$ (consecutive windows) and $d_{wf}$ (windowed and full data), respectively. This similarity measure takes into account both the size difference between compared subspaces and their mutual misalignment. The values vary between 0 and 1 being largest when the compared subspaces are similar and smallest when they are completely different.

The subspace similarity values were compared across window widths and variance normalizations. In order to assess the contribution of misalignment between compared subspaces, the similarity values were also converted to angles describing alignment, necessary for the observed similarities.

Evaluations regarding FastICA and repeatability measures (Study III)

The effects of repeatability measures and converge thresholds ($\epsilon$ concerning sICA estimation in Section 2.2.1) on the accuracy of sICA estimation were studied. The results are used for assessing their necessity in sICA of RS-fMRI data.
The evaluations were performed primarily with simulated signal sources embedded into RS-fMRI data and secondarily regarding real effects observed in RS-fMRI data. To produce the simulated resting-state effects, two random spatial ICs and their corresponding mixing vectors were generated with variance corresponding to both low and high SNR RSNs that were estimated with sICA from real single-subject RS-fMRI BOLD data (280 time points; an "eyes closed" setting, c.f. Section 2.1.4). This generated random data was then superpositioned on the same RS-fMRI data.

IC mixing vector estimates for the data components in composite data were acquired repeatedly (25 times) with original FastICA, FSL4 Melodic FastICA, and FastICA based Icasso repeatability analyses. Icasso, with random initializations, was performed both with and without bootstrapping, and with and without pseudo-inversion based estimates (later referred to as original and 'modified' Icasso, respectively).

The difference between the original and modified Icasso approaches is that the original method uses the pseudoinverse of the unmixing matrix to compute mixing vectors in adherence to generativity of the ICA model in (3), whereas the modified version determines mixing vectors through clustered IC estimate centrotypes (in the same way that both original and modified Icasso determine unmixing vectors). The modified version does not produce generative decomposition of data to mixing vectors, and spatial IC maps as centrotypes are not necessarily orthogonal in the whitened space and may present partly overlapping portions of the original data variance.

All the estimations were performed for several FastICA convergence thresholds $\varepsilon$ ranging from a typical (0.0005) to a very strict (0.0000001) value, and with four model orders (LAP based PPCA estimate which was 51, the estimate $\pm 10$, and 48). The model orders were chosen to represent the default option in Melodic (LAP), relatively large deviations from it ($\pm 10$) and a rule of thumb (48) that has been proposed earlier.

Four RSNs were identified from IC estimates. Power spectra were computed for the corresponding time series (mixing vectors). Spectral estimation was performed as described in the previous section, but with a 256-time-point window.

The mixing vector estimates for simulated ICs were compared with the originally generated mixing vectors by computing inner-products (later referred to as "similarity" values) between them. As there are no such references for other ICs related to real RS effects in the data, all IC estimates of each analysis were compared similarly (i.e. through computation of inner-products), with the IC estimates obtained from the modified Icasso analyses performed without bootstrapping. The results of modified Icasso analyses were chosen as the comparison point because they gave the most stable ICs concerning...
simulated effects. Maximally similar ICs were identified between modified Icasso and other analyses automatically through maximal values of the inner-product, separately for each value of $\epsilon$. To summarize the findings concerning both simulated and all ICs, sample means of inner product values over the 25 repetitions were computed at each convergence threshold level for each sICA approach along with 99% level confidence intervals for the actual means.

3.1.3 **Key findings and evidence**

The following sections present the major (MR2-4) and minor results regarding experimental studies described in the previous section.

*Changes in dimensionality suggest a need for model order estimation in the sliding window sICA*

Dimensionality estimates computed in the sliding window experiments changed throughout the length of the data (examples in Figure 4). The temporal evolution and the variability of estimates are dependent on pre-processing and data length, and they vary from subject to subject. It can be seen that there can be, for example, linear trends in the data of subject #9 (e.g. at window widths 120 and 140 for low and high threshold VN data) but not as clearly in the data of subject #10. Similarly there may be step-like changes, such as with subject #10 (e.g. at window widths 60 and 80 for low and high threshold VN data), but not with subject #9. Also, slow fluctuations of dimensionality estimates can be seen for various different scenarios in the data of both subjects (e.g. regarding subject #9 at window width 60 for no VN and low threshold cases, and concerning subject #10 at window width 40 for high threshold VN data and at window width 120 for no VN case).

Dimensionality can drift by several units in either direction concerning trends alone (Figure 5) which can be important in subject-level sICA as the expected dimensionality in RS-fMRI is lower than in group-level analyses.

The observed rapid, slow and periodical changes suggest that estimation is the only straight-forward approach to determining model order in each window. Consequently, a fixed model order is not suitable and should not be employed in sliding-window sICA.

Low frequency fluctuations dominated the power spectra of dimensionality estimate time series (Figure 6) despite detrending. This indicates that in addition to the observed
drifts and steps, the dimensionality has also periodical structure within an RS-fMRI data set. The frequencies of these fluctuations (≤ 0.1 Hz) correspond to the frequency band commonly associated with resting-state brain activity in fMRI. In general, power becomes increasingly concentrated on the lowest frequency, which peaks at window width 80 (144 seconds). However, the level of power at the lowest frequency may vary subject-wise considerably (Figure 5).

Low frequency profiles of dimensionality variation resemble closely those of IC time courses obtained from full data sICA. The similarity is evident when one compares group-level frequency profiles obtained in Study I to those presented in Figure 6. This suggests that sliding window sICA, which utilises model order varying with SNR, may capture temporal variability of RSNs in its spatial maps in contrast to full data studies (such as Study I). As discussed under section 2.2.1, in full data approaches sICA estimation produces static spatial maps, and corresponding time courses reflect when map activity is present in the data. This prevents detecting if RSNs are, for example, anatomically morphing or migrating back and forth in addition to activating and deactivating in time.
Subject #9, woman, age 33 years
Subject #10, man, age 34 years

Data length:
20 time points
40 time points
60 time points
80 time points
100 time points
120 time points
140 time points
160 time points
180 time points
200 time points

Window offset / time points (wrt scan timing)

Fig 4. Dimensionality estimates as functions of time: LAP (dark blue), BIC (green), MDL (red) and AIC (cyan). Example subjects 9 and 10. (Study II, published with permissions from Elsevier.)
Fig 5. Subject-wise temporal evolution in time series of dimensionality estimates (dotted lines) and group-level means (solid lines). Left panel: effects of linear trend on dimensionality change. Right panel: normalized power at the lowest frequency. Subject-level sample means computed over all estimation methods (LAP, BIC, MDL and AIC) and VN options. (Study II, published with permissions from Elsevier.)

Fig 6. Normalized power at different frequencies in time series of dimensionality estimates: sample means and bootstrap confidence intervals (bars) computed over all estimation methods (LAP, BIC, MDL and AIC), subjects and VN options. (Study II, published with permissions from Elsevier.)

*In-window PCA and variance normalization can be used as pre-processing procedures in sliding-window sICA (MR2)*

Concerning the sliding window experiments, PCA subspaces retained for overlapping parts of consecutive windows of 60 time points or more were very similar and aligned. Namely, subspace similarities $d_{cw}$ were very close to 1, and the corresponding angles...
describing subspace misalignment were in the order of 5 degrees or less (Table 2, Figure 7). Due to this, the results of subsequent analyses in sliding time windows based on PCA subspaces, such as sICA, will probably not be confounded by subspace differences. As such, the results of subsequent analyses in consecutive time windows must instead reflect differential changes found by those analyses (e.g., sICA) between the two different temporal segments of data. It seems safe to conclude that, from this PCA subspace stability point of view, in-window PCA can be used as a preprocessing step in sliding window sICA.

There were no effective differences between variance normalizations concerning subspace alignments and similarities (Figure 7; other material in Study II). Thus, from a PCA subspace stability point of view, also VN or the lack of it should not affect the comparability of subsequent sICA results acquired in the retained subspaces. Consequently, VN is viable also as in-window pre-processing in sliding window analyses and, concerning VN, PICA and its Melodic implementation are applicable for sliding window analyses.

Subspace dissimilarities render comparisons between windowed and full data sICA ill-founded

d_{w,f} were substantially lower than d_{c,w} (Tables 3 and 2, the sliding window experiments). The low value corresponds to both misalignment and large rank differences between subspaces. In general, the angles were in the order of 10 degrees or more (Study II). Also the ranges of d_{w,f} and angles were wider. Low levels of d_{w,f} demonstrate that PCA retained very different variance from data windows for further analyses (such as sICA). The bigger subspaces in the full data case also provide a very different SNR environment for estimations.
Table 2. Percentiles for the subspace similarities $d_{cw}$ measured between consecutive windows and corresponding angles. Computed over all methods (LAP, BIC, MDL and AIC), all time points, subjects and VN options. (Study II, published with permissions from Elsevier.)

<table>
<thead>
<tr>
<th>Window (TPs)</th>
<th>VN</th>
<th>Subspace similarity (d)</th>
<th>Percentiles</th>
<th>Angle (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>20</td>
<td>No</td>
<td>0.964</td>
<td>0.972</td>
<td>0.986</td>
</tr>
<tr>
<td>20</td>
<td>Low thr</td>
<td>0.899</td>
<td>0.940</td>
<td>0.959</td>
</tr>
<tr>
<td>20</td>
<td>High thr</td>
<td>0.959</td>
<td>0.969</td>
<td>0.987</td>
</tr>
<tr>
<td>40</td>
<td>All</td>
<td>0.979</td>
<td>0.986</td>
<td>0.996</td>
</tr>
<tr>
<td>60</td>
<td>All</td>
<td>0.985</td>
<td>0.993</td>
<td>0.997</td>
</tr>
<tr>
<td>80</td>
<td>All</td>
<td>0.988</td>
<td>0.996</td>
<td>0.998</td>
</tr>
<tr>
<td>100</td>
<td>All</td>
<td>0.990</td>
<td>0.997</td>
<td>0.998</td>
</tr>
<tr>
<td>120</td>
<td>All</td>
<td>0.991</td>
<td>0.998</td>
<td>0.999</td>
</tr>
<tr>
<td>140</td>
<td>All</td>
<td>0.993</td>
<td>0.998</td>
<td>0.999</td>
</tr>
<tr>
<td>160</td>
<td>All</td>
<td>0.994</td>
<td>0.998</td>
<td>0.999</td>
</tr>
<tr>
<td>180</td>
<td>All</td>
<td>0.995</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>200</td>
<td>All</td>
<td>0.996</td>
<td>0.999</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Fig 7. The subspace similarities $d_{cw}$ measured between consecutive windows. The values from all methods (LAP, BIC, MDL and AIC), all time points and subjects. (Study II, published with permissions from Elsevier.)

Because of the SNR and misalignment issues, the results of subsequent analyses (e.g. sICA) from windowed and full data PCA subspaces will likely be affected. Correspondingly, comparisons between windowed and full data results will be confounded by the effects due to PCA. Thus, in contrast to the case of consecutive windows, sliding
analyses performed in PCA subspaces of full and windowed data are not comparable. This also implies that, in general, it may be ill-founded to use any matching between full and windowed data based ICs to help automatic identification of the latter (e.g. in the case when an IC is tracked across consecutive temporal windows by trying to identify it similarly from arbitrarily ordered sICA results computed in those windows).

Table 3. Percentiles for the subspace similarities $d_w/f$ measured between windowed and full data, and corresponding angles. Computed over all methods (LAP, BIC, MDL and AIC), all time points, subjects and VN options. (Study II, published with permissions from Elsevier.)

<table>
<thead>
<tr>
<th>Window (TPs)</th>
<th>VN</th>
<th>Percentiles</th>
<th>Subspace similarity (d)</th>
<th>Angle (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>No</td>
<td>0.816</td>
<td>0.842</td>
<td>0.866</td>
</tr>
<tr>
<td>20</td>
<td>Low thr</td>
<td>0.816</td>
<td>0.842</td>
<td>0.888</td>
</tr>
<tr>
<td>20</td>
<td>High thr</td>
<td>0.756</td>
<td>0.788</td>
<td>0.816</td>
</tr>
<tr>
<td>40</td>
<td>All</td>
<td>0.772</td>
<td>0.803</td>
<td>0.837</td>
</tr>
<tr>
<td>60</td>
<td>All</td>
<td>0.801</td>
<td>0.835</td>
<td>0.861</td>
</tr>
<tr>
<td>80</td>
<td>All</td>
<td>0.834</td>
<td>0.861</td>
<td>0.881</td>
</tr>
<tr>
<td>100</td>
<td>All</td>
<td>0.860</td>
<td>0.881</td>
<td>0.897</td>
</tr>
<tr>
<td>120</td>
<td>All</td>
<td>0.881</td>
<td>0.897</td>
<td>0.910</td>
</tr>
<tr>
<td>140</td>
<td>All</td>
<td>0.899</td>
<td>0.911</td>
<td>0.922</td>
</tr>
<tr>
<td>160</td>
<td>All</td>
<td>0.914</td>
<td>0.924</td>
<td>0.933</td>
</tr>
<tr>
<td>180</td>
<td>All</td>
<td>0.927</td>
<td>0.936</td>
<td>0.944</td>
</tr>
<tr>
<td>200</td>
<td>All</td>
<td>0.940</td>
<td>0.947</td>
<td>0.954</td>
</tr>
</tbody>
</table>

FastICA based RS-fMRI results do not require strict convergence thresholds or repeatability measures (MR3)

Increasing the convergence threshold used by optimization in FastICA increased the similarity between estimates and references only when unestimated dimensionality and non-bootstrapping original Icasso were used (supplementary materials of Study III). The main results concerning estimated dimensionality remained unaffected (Figures 8 and 9).

On the other hand, there was only moderate improvement as a result of repeatability measures and only in the bootstrapping case - excluding original Icasso, which provided very low quality estimates (Figure 8). Bootstrapping attenuated power from time courses of resting-state network related ICs at frequencies higher than 0.1 Hz and made subsets of low frequency oscillations more emphasized IC-wise (Study III). These findings are based on two simulated sources and four RSN-related ICs (Figure 2), but the results concerning all IC estimates (Figure 9) show that the differences between IC estimates
of conventional FastICA and Icasso are also in general small (similarities between corresponding mixing vector estimates are consistently in the order of 0.9 or more for almost all 51 components, as shown in Figure 9).

Fig 8. Similarity of estimates of different approaches to original mixing. Solid lines and markers are mean values over 25 repetitions, and dotted lines are confidence intervals for the actual mean of similarity value distributions. (Study III, published with permissions from Elsevier.)
Fig 9. Similarities between the 51 estimates produced by the original FastICA and the reference (matching modified ICASSO estimates). Solid lines and markers are mean values over 25 repetitions, and dotted lines are confidence intervals for the actual mean of the similarity value distribution. (Study III, published with permissions from Elsevier.)

The lack of differences between FastICA and repeatability analysis implies that FastICA analyses of RS-fMRI do not require repeatability based estimation improvements. These results also give further evidence on the validity of the results in plenty of existing sICA studies of both stimulus and resting-state fMRI data which have not used repeatability measures. Despite the results concerning other fMRI data, the necessity of repeatability based estimation in RS-fMRI has not been evaluated before. Also, a similar lack of effect concerning the use of strict convergence thresholds can be observed, and a similar conclusion made on their necessity and the validity of existing studies not deploying them.

Plausibility of data models that underlie resting-state fMRI sICA methods need re-examinations (MR4)

Potential problems were identified concerning stationarity assumptions, the role of the spatial mean and the generative and 1-dimensional nature of ICA models.
First, dimensionality estimates were shown in the sliding window experiments to change substantially in time within data from imaging sessions (Figure 4). This suggests that analyses of individual subjects should not rely on assumptions of data being statistically similar throughout the imaging sessions. In contrast, however, with full data sICA one may assume that it analyses variance representative of the whole imaging session, whereas temporal changes in dimensionality (and consequently SNR) suggest that PCA may emphasize whole sections of time series compared to others. As discussed earlier concerning the dimensionality results and in 2.2.1, additional problems are introduced concerning the full data sICA model if it turns out that RSNs do not follow the assumption of being anatomically static.

Second, the lack of global mean subtraction (as in Melodic FastICA) violates common FastICA-based estimations, as discussed in Section 3.1.1, but it may be a sensible methodological choice concerning aspects specific to fMRI data, as discussed in Study III. There was a decrease in accuracy compared to normal FastICA in which the global spatial mean is subtracted as part of pre-whitening (Figure 8). Melodic FastICA estimates in general also differ more from robust estimates (Icasso) compared with normal FastICA (Figures 10 and 9): compared to the original FastICA case (Figure 9), that shows inner-product (similarity) values consistently over 0.9, Melodic FastICA results (Figure 10 are much more spread with many ICs having similarity values less than 0.9, several even less than 0.8 and a few as low as 0.7 or less.

Third, the repeatability experiments revealed that the restriction of ICA models to be generative causes problems with the original Icasso. As described in Section 3.1.2, the original Icasso uses pseudoinversion of an unmixing matrix to obtain mixing vector estimates. However, the pseudoinversion of those matrices in the case of RS-fMRI can be ill-conditioned, which is reflected in the condition numbers of the matrices being large (as shown in Figure 4 in Study III). This is the explanation for the decreased performance in the case of bootstrapping (Figure 8). As modified Icasso provides better results, it is debatable if the ICA model (in (3)) is optimal for RS-fMRI data since it forces generativity with respect to the estimation of ICs.

Fourth, the observations in Study III suggest that some sICs may be more invariantly detectable with different model orders than others. This discrepancy could be related to the existence of multidimensional independent subspaces (Cardoso 1998, Ylippavalniemi & Vigario 2008) in the data instead of all IC estimates representing 1-dimensional signal sources that are actually independent. The RSN-related ICs detected (Figure 2) might be just forced 1D-projections of such subspaces.
These examples suggest that current models used in RS-fMRI sICA may be too simple to accommodate and describe the effects that comprise RS-fMRI data and the spatio-temporal variations therein. Instead, more relaxed but higher dimensional models may provide better means for exploratory analysis. The current models should be re-assessed in this respect.

3.2 Developmental proposals

3.2.1 Guidelines for improving method evaluations (DP1)

The lessons from analysing the FSL Melodic software and the experiences from the experiments suggest some approaches to improve method evaluations. Consequently the quality of spatial exploratory analysis of resting-state fMRI can be improved. The proposal consists of four guidelines denoted later by GL1-GL4.
Exploratory analyses of RS-fMRI data are mainly applied research which relies on methods implemented in a few readily available software tools. These tools, such as Melodic or GIFT, are authored and developed mainly by single research groups, or even by individual researchers. The lack of larger community involvement in the development is likely to increase the probability of residual errors and undocumented features, as was shown earlier concerning Melodic and Icasso.

The closed nature of the method and corresponding software development is also a likely factor explaining the limited extent of current method and software evaluations in RS-fMRI research. Regarding findings presented previously in this thesis, it is evident that two features of original software development would have facilitated method evaluations such as the ones performed in this thesis. Consequently, the evaluations might have been done at much earlier stage of method deployment.

First, public projects could have been established to facilitate explicitly open development. These projects would have enabled discussion and testing of methodological choices prior to their deployment as features in tools or before the wider deployment of tools within the research community. For example, features such as variance normalization thresholds in Melodic, retaining the spatial mean in data prior to FastICA in Melodic and the use of the pseudo-inverse in Icasso have been used as such since the availability of the respective tools.

Second, higher modularity of software would have enabled testing individual analysis steps in an easier manner. The evaluations performed in this thesis required execution of the whole software (Melodic, Icasso). This is unnecessarily computationally complex, but currently the necessary way to carry out many evaluation experiments. Also, changing the original source code could have been omitted (Icasso) and the evaluations would have been practically simpler to manage, compile and analyse if the analysis software would have consisted of modules more readily separable from each other.

It is proposed here that these features of software development, i.e. explicit openness (GL1) and a highly modular design (GL2), should be incorporated into common method practices in the field. Furthermore, it is suggested that on a practical level they could be implemented through migrating from closed projects and tool packaging to more truly open source software approaches, and a highly modular and atomic structure of method implementations. Tools such as FSL Melodic, GIFT or Icasso are already distributed as source code, and offer in varying degrees independence between different steps of analyses. However, these features could be made more explicitly available for the whole research field. Establishment of new method implementation endeavours,
and re-establishment of existing ones as open source software projects, like in the case of processing pipeline development (e.g. the development of Nipype described in Gorgolewski et al. 2011), would allow for more direct contributions from the whole community while still offering means to control the changes to the software.

Concerning modularity in practice, it is suggested that all the analysis steps would be implemented as independent executables (whether binaries or scripts) instead of designing monolithic whole analysis software. Also, all analysis parameters should be included as options to those executables. These approaches would allow for relatively easy evaluation of individual steps by the whole research community. They would also allow better integration between different tools and analysis steps within pipeline frameworks such as LONI (Dinov et al. 2010) or Nipype (Gorgolewski et al. 2011). The integration would enable both easier exploration of method combinations, and more convergence of RS-fMRI explorative analysis results, as the best performing established aspects of different tools could be collected to common state-of-the-art tool ensembles.

The observations in this thesis suggest also two general practices which would be beneficial to method evaluations themselves. Concerning the original Icasso, bootstrapping was found to produce distorted mixing estimates. However, the report by the Icasso authors suggests that a similar setup works for magnetoencephalography (MEG) data (Himberg et al. 2004). Based on this, it is proposed that there should be evaluations of methods considering specifically the data domain that is to be explored with them (GL3), i.e. RS-fMRI. Otherwise their applicability may well be questionable, even if tests based on simulations or, for example, stimulus fMRI data suggest good performance. RS-fMRI can often lack the reference information needed for such evaluations, but simulated data comparable to RS-fMRI can be used (for example as in Study III). The problems of not having reference information could also be alleviated by performing evaluations using relative measures (such as the subspace similarity computed between consecutive windows in Study II or the mixing vector similarity computed between ICs from different analyses in Study III).

Concerning RS-fMRI data, there were no reported evaluations regarding VN thresholds prior to Study II and no reports concerning default convergence threshold prior to Study III, even though they have been deployed in many RS-fMRI analyses. The evaluation results also showed that the issue of issue of optimal window width is a complex subject regarding windowed sICA (Study II), and that too low a FastICA convergence threshold can impact the results, depending on the ICA model order (Study III). Identification of these factors required searching through a wide range of parameter
values. Based on this, it is proposed that the parameters of the methods should be evaluated much more extensively than has been reported so far (GL4).

A wide range of evaluations requires resources. However, increases in the availability of affordable computation capabilities has already resulted both locally and globally in considerable computer clusters. These centers and grid services can be utilized for massive evaluations within reasonable execution time as it is straightforward to parallelize parts of evaluation studies data-wise (as was done also in this thesis in Studies II and III). Also the previously suggested community participation offers solutions to resource problems. On the other hand, increasing the amount of statistical testing in evaluation of multiple parameters may require a means of correction for multiple comparisons. The following section describes one solution to such problems.

### 3.2.2 A concatenation method for multiple comparison correction through permutation testing

As described in Section 2.2.1 regarding group-level sICAs, the results of the temporal concatenation GICA can be used in combination with dual regression to compute brain maps of within- and between-group effects. These maps are then usually tested statistically for significant findings using permutation testing, as in Study IV. As was explained in Section 2.4.2, performing tests in multiple voxels against null hypotheses requires controlling of the family-wise false positive rate by correcting for multiple comparison.

Permutation testing can be used to account for multiple comparisons within an individual map when the experimental test statistic PDF is constructed by taking the maximum statistic over all tests (voxels) during each permutation (c.f. Section 2.4.2). However, when multiple effect maps are tested, as in the case of multiple RSNs being analysed, the correction should be applied also between maps, which seems not to have been the case.

In Study IV, a novel extension to the aforementioned conventional permutation testing was developed to enable also correction for multiple comparisons between multiple maps. The method is explained in the next section. However, the applicability of the developed correction is likely not limited to statistical testing in such settings. Concerning method evaluations, the correction procedure could be used to account for false positives, if the effects of parameter values on sICA results were to be statistically
tested. Without the corrections, the conclusions based on such evaluation results remain unreliable and, consequently, of low significance.

The proposed method (DP2): Maximum of maximum statistic distribution for multiple comparison correction (Maxmad MCC)

In general, the proposed method, or Maxmad MCC, simply joins different sets of variables that are to be tested, before applying on them the normal permutation testing and corresponding multiple comparison correction (testing against maximum statistic null distribution). Joining means that previously separate sets of tested variables are pooled into a single set of variables. After this, the maximum of test statistic is taken over all variables during permutations. Hence we end up with a ’maximum of maximum’ with respect to the test statistic maxima of the original separate variable sets.

In a dual regression context, this means that the effect maps that dual regression has produced for each IC do not have to be tested separately, but they can be spatially concatenated subject-wise (as described in Study IV) before permutations tests are performed. This leads to a maximum of statistical scores in each permutation being computed over voxel values in all spatially concatenated effect maps. Thresholding of statistical scores against the resulting null distribution of maximum scores not only corrects for multiple comparisons concerning findings in single maps but simultaneously in all maps concatenated prior to test statistic computation.

In the case of method evaluations, it may be possible to use this method to correct for multiple comparisons over parameter variations which are not included as explaining variables in the model. For example, sliding window sICA could be performed with all the combinations of three variance normalizations and 10 window widths (as in Study II) and 4 sICA variants and 7 FastICA converge thresholds (as in Study III). The effects of window width and convergence thresholds could then be chosen to be statistically tested. They would be placed as explaining variables in the model and regressed voxel-wise against maps produced by sICA. False positives due to performing sICA also with multiple other options (different variance normalizations and sICA methods) could then be accounted for by spatially concatenating, prior to permutation testing, all the sICA maps produced with a specific combination of window width and convergence threshold.
Preliminary evaluations in a group-ICA dual regression setting (Study IV)

Study IV involved RS-fMRI experiments on seasonal affective disorder (SAD), which is considered to be a subtype of either recurrent major depressive disorder, or bipolar affective disorder. The fMRI parameters in Study IV were the same as those described in Section 3.1.2 concerning Study II (the latter used data from healthy controls that were recruited regarding the former).

SAD patients and their healthy controls were evaluated through psychiatric assessment, as described in Study IV. Their RS-fMRI data sets were pooled together and subjected to temporal concatenation GICA for the estimation of ICs related to RSNs. This was followed by dual regression, permutation tests for differences in RSN maps between SAD patients and their healthy controls, and the corresponding multiple comparison correction. Permutation testing and corresponding MCC were performed both conventionally and using the proposed method Maxmad MCC (referred to in Study IV as "1st level correction" and "Inter-RSN correction", respectively).

The test results demonstrate how Maxmad MCC decreases the number of RSNs that have detectable between-group differences, as well as the number of voxels with detected differences concerning each RSN (Table 4). As can be seen from the results, the decrease is not plainly associated with low mean t-scores, nor with the small spatial extent of the findings (number of voxels). Instead, as discussed in Study IV, the findings remaining after applying Maxmad MCC concerned primary sensory networks and networks related to attention, which are known from independent and different types of studies to be likely affected in SAD. On the other hand, some of the RSNs excluded by the correction are associated with affective processing, which is also assumed to be affected, but for which there is less evidence in the literature.

Given this evidence, Maxmad MCC seems to be behaving in a well-grounded manner. However, further studies are needed to establish more exactly and quantitatively how conservative the method is. In any case, Maxmad MCC is much less restrictive than Bonferroni correction, which was found to remove all the findings.
Table 4. The effects of conventional maximum statistic based multiple comparison correction within permutation testing (“1st level correction”, applied separately concerning each RSN map) and the proposed method Maxmad (“Inter-RSN correction” in Study IV), which corrects across all maps simultaneously. (Study IV, published with permissions from John Wiley and Sons.)

<table>
<thead>
<tr>
<th>RSN</th>
<th>1st level correction P &lt; 0.05</th>
<th>Inter-RSN correction P &lt; 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Voxels</td>
<td>Mean t-score</td>
</tr>
<tr>
<td>1</td>
<td>534</td>
<td>2.9</td>
</tr>
<tr>
<td>2</td>
<td>373</td>
<td>3.24</td>
</tr>
<tr>
<td>3</td>
<td>261</td>
<td>3.45</td>
</tr>
<tr>
<td>4</td>
<td>675</td>
<td>3.31</td>
</tr>
<tr>
<td>5</td>
<td>1036</td>
<td>2.89</td>
</tr>
<tr>
<td>6</td>
<td>1067</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>1205</td>
<td>2.75</td>
</tr>
<tr>
<td>8</td>
<td>723</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>1504</td>
<td>2.65</td>
</tr>
<tr>
<td>10</td>
<td>3903</td>
<td>2.94</td>
</tr>
<tr>
<td>11</td>
<td>2217</td>
<td>2.72</td>
</tr>
<tr>
<td>12</td>
<td>482</td>
<td>2.81</td>
</tr>
<tr>
<td>13</td>
<td>123</td>
<td>2.96</td>
</tr>
<tr>
<td>14</td>
<td>225</td>
<td>2.9</td>
</tr>
<tr>
<td>15</td>
<td>1662</td>
<td>2.56</td>
</tr>
<tr>
<td>16</td>
<td>54</td>
<td>3.42</td>
</tr>
<tr>
<td>17</td>
<td>12</td>
<td>3.65</td>
</tr>
<tr>
<td>18</td>
<td>150</td>
<td>3.1</td>
</tr>
<tr>
<td>19</td>
<td>213</td>
<td>3.34</td>
</tr>
<tr>
<td>20</td>
<td>322</td>
<td>3.84</td>
</tr>
<tr>
<td>21</td>
<td>227</td>
<td>3.23</td>
</tr>
<tr>
<td>22</td>
<td>949</td>
<td>2.96</td>
</tr>
</tbody>
</table>

3.3 The summary of the answers provided to the research questions

3.3.1 The methodological research questions and results

Regarding the question about the match between method descriptions and their software implementations (MQ1), the analysis of FSL Melodic software (3.1.1) demonstrated that there can be discrepancies between method descriptions and implementation (MR1). Consequently, one should use readily available method implementations, even those commonly used, with caution.

Sliding window experiments showed (MR2) that the current pre-processing steps in probabilistic sICA can be used also in sliding window analyses (MQ2). On the other hand, concerning estimation accuracy, the use of repeatability measures proved a probably unnecessary analysis step (MQ3) in the corresponding evaluation experiments (MR3). The results from the two experiment sets also suggested that sICA models need
re-assessment (MQ4) because of model violations observed in the experimental data (MR4).

### 3.3.2 The developmental research questions and proposals

Guidelines (DP1) were formulated to enable improved method evaluations (DQ1). They include four principles proposed, based on observations presented in this thesis:

- **GL1** Explicitly open method and software development to enable more extensive community involvement.
- **GL2** Modular software design to increase the practicality of evaluations and integration of single analysis steps to other frameworks.
- **GL3** Data domain specific method evaluations to guarantee the feasibility of the methods and their parameters.
- **GL4** Extensive exploration of parameter space to determine the applicable ranges and sets of parameter options.

In addition, a novel multiple comparison correction approach (DP2), Maxmax MCC, was presented as a potential solution to the probable increase of multiple comparison issues with the increasing number of evaluations. As such DP2 provides one solution to DQ2.

### 3.3.3 Other contributions

In addition to answering the main research questions, the contributions also offer other insights. The data from the sliding window experiments showed the complexity of determining the sICA model order in an RS-fMRI sliding window context, and the need for dimensionality estimation in it. Furthermore, the results suggested that variation of dimensionality may well be related to fluctuations associated with resting-state brain activity, and that full data based sICA may miss detecting anatomical changes of RSNs in time.

Also, it was shown that there may be no sense in trying to use full data RS-fMRI sICA results in conjunction with sliding window results (e.g. for IC tracking purposes). Finally, in addition to answering MQ3, the level of the FastICA convergence threshold was found to be irrelevant concerning RS-fMRI sICA.
4 Discussion

RS-fMRI and fMRI in general are cross-disciplinary and evolving fields of research, encompassing both neuroscientific and technical aims. Development of new methods is limited to relatively few researchers and research groups compared to the number of people who only apply the methods for neuroscientific discovery. Because of this, the developers assume a lot of responsibility for the quality of the research with respect to the expectation that the developed and proposed methods produce results in a reliable manner.

The responsibility becomes emphasised due to the fact that the same method developers are also developing the methods into software that most researchers then use. In the case of sICA and the popular method implementations, for example FSL Melodic, an important aim is probably to maximize productivity and ease-of-use, including setting meaningful default options that would suite most users who may not explore the significance of different method parametrisations. This may account for deviations between method descriptions and corresponding software development, such as in the case of Melodic (c.f. the methodological views on global mean removal described in Study III).

Development of methods often requires a certain mathematical background, and as such it is only natural that development is limited to a relatively few researchers within a cross-disciplinary field. Evaluations of methods, however, could be perhaps performed by a significantly larger number of researchers and made into a more integrative community effort than currently.

This thesis has focused on method evaluations. As an example concerning exploratory methods in RS-fMRI, sICA related methodology has been experimentally evaluated and the software reviewed. The results have been used to formulate a proposal for improving the quality of methods, and consequently the quality of research through wider involvement of the research community.
4.1 Implications and relevance of the contributions

4.1.1 Aspects of academic software reliability

An evident implication of the software analysis results (MR1, research question MQ1) is that the software tools applied should be either independently reviewed or, as a challenging alternative, end-users should check them against corresponding method descriptions. Independent reviews of software code would be easy to perform as a part of community development projects as proposed in 3.2.1. In any case, a reasonable amount of doubt might be warranted when employing software where the inner workings are unfamiliar.

These suggestions and considerations increase the awareness of software implementation related effects on analysis results. The author feels this is especially important with fMRI, it being a cross-disciplinary research field in which most people use software tools developed by others. A single systematic software error or deviation from method specification, if left in academic software for a long time, has the potential to bias, confound or corrupt results of many analyses. If software with such an anomaly is widely used, self-correction in the research continuum may not so easily account for these errors, and a large amount of valuable and often publicly funded research effort may be lost.

Interestingly, also method evaluations and comparisons, suggested in this thesis, may be distorted if the evaluations are performed without a review of the implementational details. For example, concerning the findings in Study III, the results in Daubechies et al. (2009), according to which Infomax outperforms FastICA, might be partly due to using Melodic implementation of FastICA instead of the original one.

It is hard to know exactly how much RS-fMRI, fMRI or MRI researchers blindly rely on readily available software tools and what kind of quality assurance practices they employ concerning the software. However, the example offered by this thesis regarding FSL Melodic suggests that the situation may not be as good as one might want to believe, and that the source code review approach adopted in this thesis deviates from normal research practices.

Actually, it seems in general very difficult to locate much literature of any kind regarding this aspect of academic research. In this sense, the academic software inspection related viewpoint in this thesis presents itself as being quite original also in a larger extent. The subject may be unintentionally ignored or handled only within the
framework of software development and related unacademic communication forums. In any case, a more explicit approach with systematic research and corresponding publications would seem beneficial to any research field that employs complex software in their analyses. In principle, the matter should not be any different from quality assurance requirements regarding physical equipment used in scientific research, especially in biomedical research.

In addition to the hazards of blind utilization of software, the possibility of mismatches between method descriptions and their software implementations affect also the reproducibility of methods. For example, one would fail in re-implementing PICA as software in a manner comparable to Melodic if one would rely on method descriptions alone. Thus, from also this viewpoint, software source code itself is a more comprehensive description of methods and should be utilized as documentation when re-implementing methods in new tools. Ignoring this aspect may lead to false impressions about which tools offer the same methods and consequently may lead to, for example erroneous method evaluations or distorted comparison of RS-fMRI results derived with different tools. However, formal method descriptions are needed to provide a baseline description against which source code and the implementational considerations therein can be interpreted and evaluated.

4.1.2 Sliding-window sICA for RS-fMRI

As a result of PICA pre-processing being applicable in sliding window (MR2, research question MQ2) RS-fMRI sICA sliding window analyses can be performed under the statistical inference framework that PICA and its implementation in Melodic provide. This enables the use of well-found inference and current software versions also in the sliding window context. Consequently, baseline quality is also ensured when performing sliding window sICA on RS-fMRI data by using PICA and Melodic (compared to reverting back to using pre-PICA sICA models).

This extension of applicability regarding PICA also complements previous method validation efforts (Zuo et al. 2010b) which found PICA reliable in group-level analysis context. As such, applicability of PICA and FSL Melodic in RS-fMRI is established further in this thesis.

The potential connection between RS-fMRI effects of interest and estimated dimensionality fluctuations (3.1.3) complements previous studies (Chang & Glover 2010, Kiviniemi et al. 2011) that support sliding analyses as a promising approach for future
RS-fMRI research. However, the window width producing most power on the lowest frequency varied notably from subject to subject. This implies that sliding analyses may have to be optimised subject-wise concerning sensitivity. This would be in contrast to full data analyses in which one can more easily compensate through increasing temporal samples (longer session or concatenated data sets). On the other hand, this notion of subject-wise optimizations is in line with recent results which suggest that even pre-processing pipelines should be fitted individually for each subject (Churchill et al. 2012a,b).

An important possibility identified in this thesis, which complements similar recent findings concerning correlation analyses (Majeed et al. 2011), is that current full data based sICAs (both single-subject and group-level analyses) may be inadequate for describing potential temporal changes of RSN anatomy. Such a shortcoming could lead to current views of RSNs being substantially revised (Smith 2012).

4.1.3 Validity of RS-fMRI sICA research with respect to estimation robustness

Repeatability measures probably being unnecessary (MR3, research question MQ3), and strict sICA convergence threshold having no effect (3.1.3) imply that the results in current RS-fMRI sICA literature are valid even if they have not been computed using repeatability measures or strict thresholds. Consequently this also suggests that repeatability measures and a strict threshold do not necessarily have to be used in future studies either, which makes performing the studies simpler. In addition, FastICA related evaluations in this thesis complement stability results concerning Infomax-based fMRI sICA (Duann et al. 2005) as FastICA and Infomax are the two most used algorithms in fMRI sICA.

The findings concerning convergence threshold complement the few FastICA comparisons that are available concerning fMRI data (Suzuki et al. 2002, Esposito et al. 2002, Correa et al. 2005). Concerning repeatability measures no estimation advantages were found in contrast to what one might expect from the original method publications (Himberg et al. 2004, Yang et al. 2008, Ylipaavalniemi & Vigario 2008). Instead, the Icasso results in this thesis are supported by Icasso measurements in Abou Elseoud et al. (2010) and in Study I. In those studies, cluster compactness ("Iq"), even though it is a relative criterion, was found to be high for most RS-fMRI ICs, especially RSN related ICs. Moreover, Pendse et al. (2011) reports high reproducibility of single-subject
RS-fMRI ICs, and the method described in Ylipaavalniemi & Vigario (2008) seems to have never been publicly tested in an RS-fMRI context.

On the other hand, repeatability frameworks provide other features besides ‘averaging’ over algorithmic induced variability of sICA results. All the frameworks provide visualization to assist in assessing variability component-wise. Li et al. (2007b) propose Icasso for dimensionality estimation on the basis of IC reproducibility. In Pendse et al. (2011) RAICAR (Yang et al. 2008) has been developed further to provide a means of handling subject-to-subject variability, which affects ICs in group analyses. RAICAR has also been proposed for finding homogeneous subgroups of data (Yang et al. 2012). It has been also suggested that clustering involved in repeatability analyses would be used to aim at identifying independent subspace instead of independent components (Ylipaavalniemi & Vigário 2007, Vigario & Oja 2008, Ylipaavalniemi & Vigario 2008). In this respect, the repeatability analyses provide many interesting opportunities. They just seem to concern aspects of data analysis other than improving estimation accuracy in conventional sICA of RS-fMRI data, as shown in this thesis.

4.1.4 The model and method development for exploratory analyses in RS-fMRI

The need for re-assessing and perhaps revising sICA models (MR4, research question MQ4), which was demonstrated in 3.1.3, may lead to a variety of novel candidates for RS-fMRI exploratory analysis methods. There are many possibilities which could be tried singly or in combination.

Daubechies et al. (2009) suggests developing new algorithms that optimize particularly for sparsity instead of aiming and assuming independence as in current approaches. Outside the sICA context, for example clustering, discussed in 2.2.2, provides the means for exploratory analyses not constrained by all the same issues as sICA. In addition to these examples, it could be highly beneficial to acquire new ideas from research fields detached from RS-fMRI or fMRI.

Independent subspace analyses (Ylipaavalniemi & Vigário 2007, Vigario & Oja 2008, Ylipaavalniemi & Vigario 2008) offer ways to operate within subsets of original data variance and search for meaningful data projections in them. With the ideas in Cardoso (1998), this could be taken further so that those subspaces would be used as a whole to describe the phenomena being analysed. This could, for example, alleviate the problem of not being able to track anatomical changes of RSN with conventional sICA.
(discussed in Section 3.1.3) if multidimensional subspaces are temporally sufficiently more stationary than ordinary one-dimensional ICs. Sliding window sICAs (Esposito et al. 2003, Karvanen & Theis 2004, Kiviniemi et al. 2011) also offer a potential alternative to characterising anatomical changes in RSNs, but the existing approaches are restricted to one-dimensional projections.

Additional supporting data could help both in designing new data models and performing the exploratory analyses. Rich stimuli based on movie viewing (Bartels & Zeki 2004, Malinen et al. 2007, Ylipaavalniemi et al. 2009, Pamilo et al. 2012) seem to give good coverage of brain functions within single imaging sessions, but the timing of movie elements can be used as reference information, similarly to how analyses are performed in traditional stimulus studies. Multimodal data acquisition and analyses provide the means to match between RS-fMRI and other data modalities (e.g. concerning EEG-fMRI Martínez-Montes et al. 2004, Yuan et al. 2012) or to drive ICA decompositions (Groves et al. 2011).

While the sICA model related findings in this thesis encourage pursuit of other models and analyses, such progression fragments method development further. The proposals for improving method evaluations formulated in this thesis (3.2.1, DP1, developmental question DQ1) thus offer important guidelines so that possibly numerous different approaches can be adequately tested. If RS-fMRI exploratory analysis results are derived with a number of different methods, the risk of systematic errors is decreased, as results can be averaged on meta-analysis level to increase the credibility of the findings. The averaging, in that case, eliminates method induced errors if they are unrelated, whereas the information of interest, if it is detectable with multiple methods, is probably retained. However, the number of results not common to several analyses (and therefore possibly discarded by meta-analyses) can be trusted more if the methods have been properly validated.

The proposal on method evaluations has also potential in educating researchers about the methods they use, given the cross-disciplinary nature of fMRI and RS-fMRI. If taken into practice, the shared evaluation would thus imply an improved capability of the research field to also give feedback to method developers and to provide fresh viewpoints from those outside the core of method development.

In addition to method evaluations, a statistical correction method was proposed for the problem of multiple comparisons (DP2, developmental question DQ2). It seems that at the moment there are no alternatives besides adjusting threshold levels, for example,
by the classical Bonferroni approach. As such, this thesis provides the means to both current group-level analyses and to potential future method evaluations.

4.2 Methodological considerations

A general limitation of this thesis is that feasible data has been available only from 1.5 T measurements locally, and therefore the evaluations lack comparisons to 3 T data, which has become increasingly more state-of-the-art in fMRI research. Existing reports show that, for example, physiological noise from cardiac and respiratory cycles affects ICA results more concerning 3 T data than regarding 1.5 T data (Starck et al. 2010, Beall & Lowe 2010). The evaluations in this thesis being constrained only to the case of conventional EPI BOLD data may also limit the interpretations allowed concerning future research. Namely, for example the emerging high speed imaging (c.f. e.g. Hennig et al. 2007, Feinberg et al. 2010, Zahneisen et al. 2012), offering an order of magnitude faster temporal sampling, will probably change radically the current fMRI and RS-fMRI modelling and analysis conventions and issues. A third notable limitation in this thesis is the small number of subjects used in the experiments, which was practical necessity considering the dimensionalities of the evaluated parameter spaces.

However, this thesis still provides quite good premises for the generalization of the conclusions presented in it. The results presented herein cannot be used to describe how extensively neuroscience is in practice affected by the method related shortcomings identified in this thesis, but the examples and cases allow for deriving a generally cautious view about the methodological state-of-the-art. The software related concerns in this thesis (MR1, research question MQ1) are evidently altogether independent of data. Also the data model reassessment related issues raised (MR4, research question MQ4) are mostly generic, despite the evidence originating from specific or limited data. Furthermore, the proposals regarding method evaluations (DP1, developmental question DQ1) and statistical testing (DP2, developmental question DQ2) are applicable, regardless of the data and even independent of the data domain considered (RS-fMRI).

Validation of previous RS-fMRI sICA research concerning sICA parametrizations (MR3, research question MQ3) is also justified due to the use of simulated data embedded in real measurements, and as the older studies have been largely based on 1.5 T field strength that was in the past the de facto standard for fMRI measurements. On the other hand, the sliding window evaluations employ a great number of both parameter combinations and temporal subsamples, which increases the probability
of discovering various phenomena, in spite of being restricted to 1.5 T and a limited number of subjects. The simulation and combinatorial rationales also apply, despite the fact that the experiments have been limited to EPI BOLD data only. However, the sICA specific results from both the sliding window and repeatability experiments may be to a certain degree limited to the given data domain (i.e. 1.5 T EPI BOLD RS-fMRI). Further studies would be required to test this possibility.

Finally, as an aspect different from many previous studies, the evaluation experiments done in this thesis consider all the data components or the whole signal containing subspace, instead of focusing on a few or individual components only (such as default-mode network and its nodes). This generality provides additional utility for the results presented in this thesis.

4.3 Future directions

Concerning RS-fMRI research, the ideas of community efforts have previously been promoted (Fox & Raichle 2007a) and realized (Biswal et al. 2010) in the context of data sharing. There would also be a chance to expand that co-operation to method development endeavours.

For example, it could be beneficial to formulate in detail an agreement among method developing research groups on how all software and methods should be tested, arguably much might be learned from industrial practices. Corresponding documentation could then be disseminated to the whole field, much like the efforts concerning education about thorough reporting of experiments (Poldrack et al. 2008).

The first guidelines could be a coarse-grained sketch gradually updated with time as practical experiences are obtained from actual testing. The key idea would be to have some sort of process that provides structure to how all tools, some of which are potentially used by the whole research community, would be evaluated at the very least. As new major tool projects in RS-fMRI research may be initiated relatively rarely, the process could be started by re-viewing and re-evaluating existing ones.

Ultimately standards could also be developed concerning method and parameter reliability. In this regard, practices analogous to manuscript peer reviews or technological standardization could be utilized. For example, in standardization of Internet-related technologies, two independent implementations are required. Similarly, two independent thorough evaluations could be required before considering new methods and parameters standard approaches (in addition to validations done and published by method authors).
Standardization would also help in making peer reviews. In cross-disciplinary research, the background of a single reviewer rarely covers all the aspects relevant to the subject. In this sense, the situation regarding RS-fMRI and also fMRI in general deviates, for example, from more traditional medical or engineering research.

Evaluations, and especially their commensurability, could be further facilitated by standardization of test data. The aforementioned data sharing endeavours offer large amounts of real data to be used as a basis for test sets concerning both 1.5 T and 3 T measurements. Also simulations, for example through POSSUM (Drobnjak et al. 2006) or SimTB (Erhardt et al. 2012) software, enable the creation of RS-fMRI test data with various effects and levels of details. Also, the other way around, methods could be validated against various data sets from different research sites by arranging challenges analogous to competitions, and other similar events, that are occasionally arranged for example in conjunction with conferences. However, in the method validation context, the purpose might not be (only) to evaluate certain methods with a single data set, but also at different research sites with the proprietary data sets they possess.

In addition to the lack of common rules and test data for method research, there could be a structural problem within the RS-fMRI community concerning the nature of research. FMRI itself is a very cross-disciplinary research field requiring everything from cognitive sciences and neurophysiological understanding to statistical and engineering expertise, and one could argue that the unconstrained nature of RS-fMRI makes data even harder to analyze in a way that produces unambiguous results. Namely, the blind use of analysis tools might be even less warranted in the case of RS-fMRI, where one ultimately aims at explaining 'background effects', the practical modelling of which is sufficient in SNR improvement scenarios of conventional stimulus fMRI.

Despite the potential of community efforts to improve the tools, it would seem beneficial - if not necessary - to increase, in general, the amount of in-depth technical understanding in research groups which utilize the tools. In addition to enabling improved RS-fMRI analyses, such a change would facilitate also the community co-operation discussed earlier in this thesis.

With or without of aforementioned improvements, it would also be highly useful to carry out and report ordinary RS-fMRI analyses with multiple parameter alternatives, instead of using just some fixed set of analysis options. Generally, a lack of computing resources should not be a problem nowadays and especially parallel computing facilitates the obtaining of results in a reasonable timeframe, as in the case of evaluations reported in this thesis. On the other hand, multiple comparison problems related to multiple
analyses could be mitigated by using correction methods such as the one formulated in this thesis or for example by changing from voxel-wise statistics to more compact, higher level characterizations of results (such as overlap percentages per anatomy that were used in Study I). However, one remaining problem, regarding at least multiple sICA analyses from single data, is the explosion of results that have to be ordered and matched in order to make meaningful comparisons about the effects of the parameters. Some effort would be needed to automate these procedures that are still currently manual or semi-manual.

Finally, it would seem that the remaining credibility and plausibility problems with RS-fMRI and its exploratory analyses could be largely mitigated if controlled experiments would be used more to confirm the hypotheses suggested by the results. This is not to claim that such experiments would be easy to construct, or that they are undemanding to perform, but much more effort could be invested into such endeavours.
5 Summary and conclusions

Evaluating methodology can be seen as a compulsory, yet laborious and unrewarding task compared to developing new methods and making seminal proof-of-concept demonstrations regarding their potential. Especially in the context of cross-disciplinary and applied research, the methodological details may seem secondary to some other primary objectives pertaining to the field of study (such as neuroscientific findings in the case of RS-fMRI).

Still, methodological choices can affect results to a large extent, and the sufficiency of evaluations deserves appropriate attention. Whatever the level of abstraction on which validity is assessed - the underlying model, practical method or software implementation - the role of evaluations is analogous to the role of product testing in industry. No one would use, for example, a spreadsheet application to process important data if they could not rely on it to produce error-free computations. Unfortunately, errors can be much less trivial to observe in the case of complex scientific computations. Unless thorough evaluations are reported, the use of analysis tools remain a matter of faith instead of evidence.

It was shown in this thesis how there are numerous unevaluated methodological choices and options concerning explorative spatial analyses of RS-fMRI data. Even in the case of the default tools and options, the applicability of many approaches remain untested. Also the popular tools can suffer from methodological details that are ill-fitted for analysing RS-fMRI data, and remain undetected due to the lack of evaluations, as was shown in this thesis regarding FSL Melodic and Icasso. Consequently, there is a growing possibility of systematic errors in the reported neuroscientific results, especially if current research practices regarding tool development and deployment continue.

To remedy this current situation, this thesis provided a number of answers regarding sICA, and also proposals about how method evaluations could be facilitated and improved. The main conclusions regarding current methodology can be summarized as follows (with regard to the original research questions MQ1-MQ4):

MR1 Software implementations of deployed analysis methods should be checked first hand, prior to analyses, unless there is evidence that the implementations adhere to the reported method descriptions. Preferably source code reviews, independent from development, would be made available to the whole community.
MR2 PCA and voxel-wise variance normalization pre-processing related to the probabilistic ICA model and the FSL Melodic implementation can be used also in sliding window analyses. Thus, PICA and Melodic offer the means to perform emerging non-stationarity related RS-fMRI analyses with sICA in addition to using, for example, sliding window seed correlation approaches.

MR3 Given the results obtained in this thesis, sICA estimations based on repeatability analyses are not necessary in the case of RS-fMRI. This also further validates earlier studies in this regard. However, the repeatability approaches and related tools may be valuable concerning other applications such as seeking interesting projections of data without following a strict ICA model.

MR4 The suitability of spatial ICA models should be re-assessed concerning RS-fMRI data. Statistical dissimilarity through time, mean substraction issues, incompatibility between most stable IC estimates and the generative model, and the possibility of multidimensional ICs imply at minimum sub-optimality in data modelling when using currently assumed sICA models and interpretations based upon them.

The guidelines proposed in this thesis (DP1) for improving method evaluations - and consequently the quality of RS-fMRI research - can be summarized as follows:

GL1 Carry out open method and related software development instead of closed projects confined to a single research site or group. Community involvement increases the likelihood of earlier error detection.

GL2 Design research software in a highly modular fashion. Modularity makes evaluations more practical, as single steps in the analysis can be evaluated separately and as stand-alone modules enable easy integration between different analysis frameworks. The integration also enables easier construction of new tool ensembles incorporating all the latest method developments and more options than any of current tools.

GL3 Evaluate methods and their options within the data domain in which the analyses are going to be performed. As demonstrated within this thesis, in the case of RS-fMRI analysis methods, generic evaluations and evaluations with different kind of data, such as stimulus fMRI based test data, may not alone suffice.

GL4 Perform evaluations extensively with respect to the size of parameter space. Evaluations with a narrow range of parameters or too few parameter dimensions can fail to reveal suboptimal method options (as shown in this thesis concerning the
convergence threshold). Deploying massive parallelism within a cluster computing infrastructure and utilizing community involvement can help resource-wise in covering a large parameter space.

Performing extensive evaluations and statistical analysis of parameters and the options therein results in an explosion of multiple comparisons. Consequently, corresponding corrections are needed. The method developed and proposed in this thesis for such corrections (DP2), Maxmad MCC, offers one such correction method within the framework of permutation testing.

RS-fMRI offers an extensive view to brain functions with relatively simple measurements, but in return requires demanding processing and analyses to utilize the data. The opportunities concerning both basic neuroscience and potential clinical applications seem abundant. The research field has matured greatly in a short period of time, and joint data collecting and dissemination efforts have been carried out to facilitate research from a neuroscientific point of view. The next logical step would be to jointly develop and evaluate analysis methods and their implementations. Otherwise the efforts of a still relatively small research field are not utilized to their full potential, nor will the results be properly validated. Less competition and more co-operation is good a recipe in present-day science involving complex problems and complex tools. We are beyond toy problems.
References


Original publications


Reprinted with permissions from Elsevier (II & III) and John Wiley and Sons (I & IV).

Original publications are not included in the electronic version of the dissertation.
458. Lauri, Janne (2013) Doppler optical coherence tomography in determination of suspension viscosity
459. Kukkola, Jarmo (2013) Gas sensors based on nanostructured tungsten oxides
460. Reiman, Arto (2013) Holistic work system design and management : a participatory development approach to delivery truck drivers’ work outside the cab
461. Tammela, Simo (2013) Enhancing migration and reproduction of salmonid fishes : method development and research using physical and numerical modelling
462. Yadav, Animesh (2013) Space-time constellation and precoder design under channel estimation errors

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
Jukka Remes

METHOD EVALUATIONS IN SPATIAL EXPLORATORY ANALYSES OF RESTING-STATE FUNCTIONAL MAGNETIC RESONANCE IMAGING DATA