Linking Logical and Coordinate-based Resources for Interoperability of Primate Brain Mapping and Connectivity Data

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Gleb Bezgin

aus Tuapse, Russland

Düsseldorf, Dezember 2009

aus dem Institut für Informatik

der Heinrich-Heine Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. R. Kötter

Koreferent: Prof. Dr. E. Wanke

Tag der mündlichen Prüfung:

Dedicated to my precious grandmother, Anna F. Sokolova

Contents

Abstract

Nowadays, the most comprehensive information about the human brain's structure and function is provided by non-invasive neuroimaging experiments, such as fMRI or DTI. Given that reliability and tractability of data provided by these studies are still highly questionable, the utility of results from animal brain research is needed. Such a utility is a hard problem to maintain: whereas neuroimaging data are represented by specific properties assigned to the three-dimensional space, the previous animal studies are typically stored in textual databases where entities are linked by rules, as retrieved from original research reports. In this thesis, two different tools for linking spatial with ontological neuroscientific data are presented. One of them, SORT, described in the Chapter 3, is extracting database-compatible logical statements from spatial data. Another tool, CP3D, described in the Chapter 4, links sectional stereotaxic atlas data with a connectivity database, organizing a plain list of structures in a hierarchical manner, according to a neurological taxonomy. Both tools are freely available and being used by the neuroinformatics research community; here, they are further systematized and described as one continuous pipeline.

Acknowledgements

The period of my life marked by doing my Ph.D. study was an experience I can hardly compare to anything I had before. Sometimes uneasy and frustrating, very rewarding and exciting in general, this experience gave me plentiful insights so precious for my future career. There are, indeed, numerous people who made this possible – those who shared great moments and supported me during hard times.

First of all, I would like to thank my supervisors. Rolf has been very flexible and understanding research father, offering the best ratio between freedom and control. Egon has carried out an exciting collaboration on research issues provided in this thesis, also on extensive discussions of the latter.

Secondly, I would like to thank my beloved parents, sister and grandmother for their care and moral support; also, a big thanks to my very first mentor, Vasiliy A. Further, I have been systematically boosted up by my cousins Tetyana, Oleg, Andrey and Victoriya during my study. Also, thanks goes to all their wonderful daughters, whose ability to learn so fast is highly inspiring! Thanks also to Viktor Y. and Klara G. who supported me spiritually from time to time.

Many of my colleagues, indeed, deserve mentioning and great appreciation. Andrew has been a helpful labmate and fellow research conspirator, but also a friendly roommate for some tough periods of searching affordable accommodation around Nijmegen, and also a great friend. In addition, I highly appreciate his linguistic analysis of this thesis. Rembrandt has greatly helped me to assimilate in Dutch culture, and has also been a very helpful and flexible (albeit informal) co-supervisor. He is one of those (and some others would be Marieke and Oliver from his family) whom I would call a true friend. Moreover, I would like to thank Rembrandt also for helping with several technical issues of the section 3.5, related to Fig. 3.5.2 and Table 3.5.2. Big thanks to Eva and Martijn who thoroughly went through the German and Dutch versions of my summary, respectively. Dirk and Ingo have been great companions in discussing research and non-research issues, pursuing our joint projects, and also having relaxing events like movie evenings and barbeque parties, frequently with their wonderful inquisitive-minded sons. Our secretaries – Judith, Margriet and Irene – have been great helpers. Also, I am indebted to Frau Loesch, Frau Sommerkorn, Vladimir Vorobiev and Milenko Kujovic who helped me a lot during my very first days in Düsseldorf in a stage of getting acquainted to a completely new environment. Thanks goes also to Antje K. whose contribution to several issues discussed in this thesis is very important. Since I am out of capacity for any further extended acknowledgements, I would like to thank all my friends, and here is the list of other people in our department (and not only) who made my days during this period: Stan Gielen, Jan van Gisbergen, Thom Oostendorp, Bert Kappen, Eva Ludowig, John van Opstal – and for the entire department for such a friendly and fun, yet highly intellectual atmosphere, so greatly needed for doing a fruitful research!

Abbreviations used in this thesis

10 area 10 of cortex

10D area 10 of cortex, dorsal part

10M area 10 of cortex, medial part

10V area 10 of cortex, ventral part

23 area 23 of cortex (posterior cingulate cortex)

8A area 8A (frontal eye field)

A1 primary auditory cortex

A2 secondary auditory cortex

API application programming interface

BOLD blood oxygen level dependent (signal)

CCa anterior cingulate cortex

CCp posterior cingulate cortex

CCr retrosplenial cingulate cortex

CCs subgenual cingulate cortex

CoCoMac Collations of Connectivity data on Macaque brain

CP3D CoCoMac-Paxinos-3D tool

D disjoint (CoCoMac mapping relationship statement)

DSI diffusion spectrum imaging

DTI diffusion tensor imaging

ECoG electrocorticography

EEG electroencephalography

FEF frontal eye field

fMRI functional magnetic resonance imaging

GUI graphical user interface

I identity (CoCoMac mapping relationship statement)

IA anterior insula

INCF International Neuroinformatics Coordinating Facility

Ip posterior insula

JDBC Java Database Connectivity

L larger area than (CoCoMac mapping relationship statement)

M1 primary motor cortex

MRI magnetic resonance imaging

NIF Neuroscience Information Framework

O overlap (CoCoMac mapping relationship statement)

ODBC Open Database Connectivity

ORT objective relational transformation

PCi inferior parietal cortex

PCip inferior posterior parietal cortex

PCm medial parietal cortex

PCs superior parietal cortex

PDF probability density function

PFCcl centrolateral prefrontal cortex

PFCdl dorsolateral prefrontal cortex

PFCdm dorsomedial prefrontal cortex

PFCm medial prefrontal cortex

PFCorb orbital prefrontal cortex

PFCpol pole of the prefrontal cortex

PFCvl ventrolateral prefrontal cortex

PG parietal area PG (caudal inferior parietal lobule)

PGa area PG associated region in the superior temporal sulcus

PHC parahippocampal area

PMCdl dorsolateral premotor cortex

PMCm medial premotor cortex

PMCvl ventrolateral premotor cortex

RC relation code (in mapping between brain sites)

RM "Regional Map" of the primate cerebral cortex

S subarea (CoCoMac mapping relationship statement)

S1 primary somatosensory cortex

S2 secondary somatosensory cortex

SORT spatial objective relational transformation

SQL structured query language

SuMS Surface Management System

SVG Scalable Vector Graphics

TCc central temporal cortex

TCi inferior temporal cortex

TCpol temporal cortex pole

TCs superior temporal cortex

TCv ventral temporal cortex

TPO temporal parietooccipital associated area in the superior temporal sulcus

V1 area V1 (primary visual cortex)

V2 area V2 (secondary visual cortex)

V3A visual area V3A

V4 visual area 4

V4D visual area 4, dorsal part

V4V visual area 4, ventral part

VACd anterior visual cortex, dorsal part

VACv anterior visual cortex, ventral part

X an existing connection with unknown density ranking

XML Extensible Markup Language

Chapter 1

Introduction

By means of investigating the brain of any species, the neuroscientist tries to find an answer to the ultimate question of brain science: how does the *human* brain work? This neuroscientist might be engaged directly in human brain research, or he or she can be performing animal experiments – the final goal remains the same. How can animal research and its data mining be exploited in this paradigm? There are indeed some striking resemblances between the human brain and brains of some other species (particularly primates) in structural and functional senses, albeit investigation of such similarities (homologies) may take very long time and entails very extensive research. Despite these complications, the neuroscientific research community keeps performing animal experiments, theoretical researchers exploit extensive non-human mammal databases – and all these efforts are a tribute to understanding of the human brain structure and function.

From the point of view of theoretical neuroscience, this goal statement can be rephrased in a following way: having an invasive animal experiments database available on the one hand, and non-invasive human research data on the other, are there any properties resembling each other in a statistical sense? To narrow down this problem statement even further, are these different datasets compatible with each other so that the researcher can easily take one dataset, make a statistical comparison with the other and say whether they are similar in particular aspects or not?

These datasets must be, first of all, compatible with each other. What does compatibility between human and non-human datasets imply? The first issue to be considered is that the most of relevant human brain data are provided by neuroimaging experiments, including functional magnetic resonance imaging (fMRI), and diffusion tensor imaging (DTI); these data are typically represented in volumetric units (voxels) located in Euclidean three-dimensional space. By contrast, data obtained from invasive animal studies (such as tract tracing or electrophysiological techniques (Bakker et al., 2009)) are usually described by text and supporting figures obtained from publications. Occasionally, such studies are collated in extensive databases (Kötter, 2001), allowing the researcher to obtain insights from not just one, but several studies on the same field. Therefore, the problem of linking human with non-human data resources can be to a certain extent reduced to providing the means of linking spatial (geometric) with ontological (textual) data.

One of the most illustrative examples showing the relevance of linking spatial with textual neuroscientific data is the utility of anatomical templates in fMRI experiments. So far, no

reliable and at the same time detailed human parcellation of the entire cortex is provided yet, and neither is the human connectome (Sporns, Tononi & Kötter, 2005). Hence the existing atlases (templates) are either very coarse or based on purely topographic criteria (taking into account information on sulci and gyri, but not on cell architecture or other structural or functional criteria). In such an absence of a reliable parcellation scheme of the cerebral cortex, the neuroimaging research community typically relates its data to anatomical landmarks, or to the map of Brodmann created exactly hundred years ago (Brodmann, 1909). The latter subdivision is based on the cytoarchitectural criteria and contains 52 distinct regions, whereas corresponding investigations in the macaque cortex, for instance, provide nowadays parcellations including up to 300 regions (Paxinos et al., 2000). Given that the human cortex is much more complex that of macaque, the problem becomes instantaneously evident (Fig. 1.1).

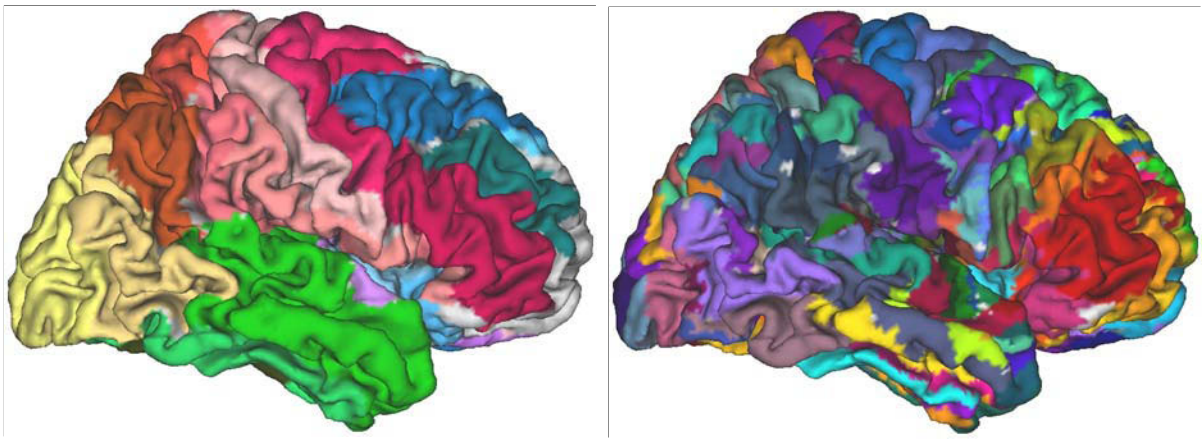


Figure 1.1. Coarseness of existing human anatomical references. Left: the map of Brodmann (Brodmann, 1909), commonly used as an anatomical template for neuroimaging studies. Right: macaque parcellation after (Paxinos et al., 2000), deformed to the same human template.

The issue of the utility of animal studies is complicated by the fact that they are numerous and heterogeneous in terms of nomenclature and other description criteria. For example, the CoCoMac database on the primate tract tracing studies, described in detail in Chapter 3, contains more than 400 studies providing the anatomical parcellations and wiring schemes of the primate brain. These studies describe various subsystems, and their authors use different terminologies – therefore the issue of combining them is a huge effort in itself, currently tackled by CoCoMac. In our framework, we address the issue of linking these studies as a whole to neuroimaging investigations of the human brain.

The organization of this thesis is hence as follows.

In the second chapter, some general paradigms used throughout this thesis are discussed. On the one hand, this chapter describes several neuroinformatics concepts relevant for the issue. One of the main notions in this regard is the fact that this field is constantly changing and, as a result, demanding new requests which are largely different from those addressed in the same field several years ago. The subject of that subchapter is therefore marking out these development trends aiming to extrapolate the needs of a common researcher in the neuroinformatics field arising in coming years, given the current state of this field. In the second subchapter, the particular case of the neuroinformatics field is described; the question is whether a human brain researcher might consider soon that animal studies are not of any relevance at all given the advances in modern neuroimaging, or these investigations will remain pertinent for years ahead.

Describing an even more specific case, the third subchapter gives a brief overview of various brain atlases which have been developed until now, together with computational techniques used for their management, extension and general accessibility. Particular emphasis is put into computational approaches, which are not merely a means of providing the user access to a particular atlas, but also a powerful instrument for extending the atlas, handling its vocabulary in an advanced way, and providing interoperability with other neuroinformatics datasets. These aspects provide a necessary flexibility for a researcher, so that adaptation of the atlas dataset and associated tools to a particular experiment becomes convenient and intuitive.

Chapter 3 provides a description of the SORT procedure (Bezgin et al., 2008), a machine learning approach for extraction of logical statements from spatial data. This approach provides a kind of “backward compatibility” so that any newly generated neuroimaging study can be attributed to a particular animal study (or several ones) produced in the past. Two particular datasets are taken as an example for such interoperability: CoCoMac (Kötter, 2004) and SuMS (Dickson et al., 2001).

Depending on the data modality used for such a conversion, the model can be trained differently, and therefore three different approaches are provided in this regard:

- hypothesis-driven, assuming that the data have a particular distribution (e.g. normal)
- theory-driven, taking into account the knowledge about the domain of interest
- data-driven, the most objective method based fully on data.

It is also shown that all three approaches produced similar results, and slight dissimilarities are discussed.

Chapter 4 describes the software tool called CP3D (Kötter et al., 2008; Bezgin et al., 2009), representing a three-dimensional graphical user interface for retrieval, visualization and analysis of connectivity data between regions in stereotaxic space, sorted according to multi-level hierarchical representation. The rationale for such a tool is provided by answering some common example questions generally asked by the research community, particularly by clinicians. Development stages are described in detail, allowing the approach to be generalized to other datasets, and techniques to be used in other paradigms also in fields different from neuroscience. In particular, performance of SQL database queries is independent of a platform and has only a very slight variability between different databases. For that reason, the concepts of query implementations are provided here as well.

Altogether, this thesis provides a mature framework for interlinking various resources of brain research by means of extensible software development. On the one hand, relating modern human neuroimaging research to previous invasive animal studies allows one to employ the wealth of information obtained within decades of primate brain research; on the other hand, the accuracy of geometric representation of delineations in stereotaxic space is complemented by an access to the connectivity database and flexible three-dimensional rendering. Moreover, organization of structures in a hierarchical fashion delivers a comprehensive neuroanatomical description to atlas volumes. Since the development methods can be also of considerable interest for neuroscientific research community, they are provided alongside, allowing one to extend the proposed tools with self-developed features, or to create new tools on the basis of introduced methods (e.g., Bohland et al., 2009).

Chapter 2

Main concepts discussed in this thesis

The field of neuroinformatics is broad and heterogeneous. Different areas of this field subserve various needs of a general neuroscientific community. Hence whenever a researcher pronounces the word “neuroinformatics”, the issue has to be clarified – be it databasing, computational modelling, computer-based analysis, etc. Terminology can also differ substantially in each of these subfields.

This chapter will therefore provide a description of issues relevant exclusively to the current study, but taking into consideration those subjects appropriate for the entire field of neuroinformatics – and, as a result, for the broad field of neuroscience.

2.1 General neuroinformatics introduction

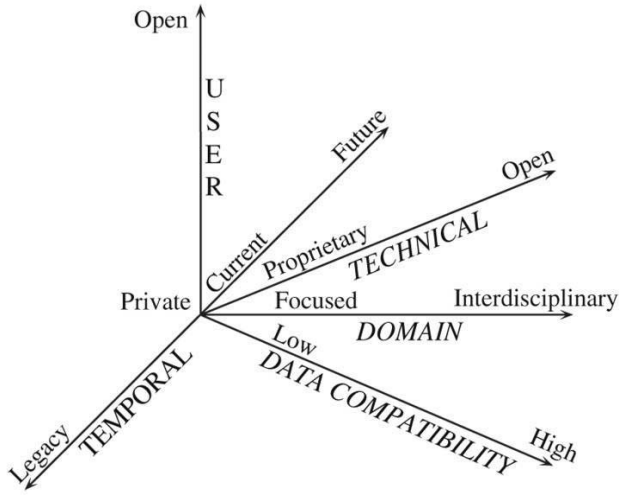


Figure 2.1.1. Various aspects comprising tool development issues in the field of neuroinformatics. Taken from Gardner et al., 2008.

Neuroinformatics as a separate branch of neuroscience appeared in late 80’s when the need to systematize data generated from neurological studies became increasingly urgent. It needed several years to become formally a distinct branch of science; this process was aided by the initiative called The Human Brain Project (Koslow and Huerta, 1997).

This enterprise involved such diverse areas as development of human brain atlases, neuronal activity simulators, applying informatics to learning health and disease states, together with policies for involving various neuroscientists in a process of sharing their data for efficient collaboration and flexible management (Amari et al., 2002; Gardner et al., 2008). In this paradigm, user interactivity, data interoperability and other criteria are intimately linked to each other (Fig. 2.1.1).

For comparison, molecular biology databases appeared about a decade earlier. This time gap can be explained not by the lack of necessity for such a development in a field of neuroscience, but rather by the increased intricacy of neuroscientific data management. In early 80's tools were merely not able to deal with data of such a high complexity. Indeed, gene sequence databases are technically easier to maintain, since nucleotide and amino-acid sequences are representable by relatively small sets of letters. Neuroscientific data, by contrast, are distinguished by its extreme diversity and hence high complexity in management, representation, quality control and acquisition (Kötter, 2001; Chicurel, 2000). Therefore, providing a neuroinformatics tool implies not only a scrutinized scientific development, but also the issue of convenient data representation and interoperability with other resources.

2.2 Relevance of databases on animal research for human neuroimaging studies

Whereas sharing newly produced data from neuroimaging experiments is generally considered important and performed with a certain degree of success, the issue of utilizing data on invasive animal studies from past years tends to be much more complicated and for that reason frequently ignored. A common question is: are those studies of a particular relevance anyway, or will expected advances in non-invasive human measurement techniques eventually provide enough reliability so that the reference to animal studies will appear merely superfluous?

Recent evidence suggests that, despite extremely rapid advances in neuroimaging techniques, the data they produce will remain hardly tractable in coming years. As an example, fMRI data, despite having high spatial precision, features very low time resolution; on the other hand, even spatial resolution does not speak to a high reliability: hemodynamics and BOLD response have very complex relationship with neuronal activity, which is far from being fully understood at present. Taking another measuring modality, electroencephalography (EEG), we face a different problem: whereas temporal resolution is high, the observed activity cannot be attributed to the region of interest with a sufficient degree of precision. Patient studies involving direct measurements of brain surface potentials, i.e. electrocorticography (ECoG), have other set of limitations: first, the area taken into account is commonly spanning just a small part of the cortex (neurological diseases treatable by surgery are reasonably localized in space), and subjects typically exhibit abnormal neuronal activity which is apparently different from the normal brain state (example study: He et al., 2008). Considering in vivo techniques providing structural descriptions of the human brain (e.g.,

DTI (Le Bihan et al., 2003), or its advanced version, DSI (Schmahmann et al., 2007)), one can also notice that current limitations are still extremely prominent: fiber tracts depicted by such techniques are represented with no information about their extent and directionality. These limitations are complemented to some extent by invasive tract tracing studies on the non-human primate brain, which provide such information. Moreover, such animal studies frequently provide additional details on connectivity – e.g., laminar patterns of connections, which can be used to determine hierarchical organization of brain areas related to a particular sensory modality (for example, visual – see Fig. 2.2.1).

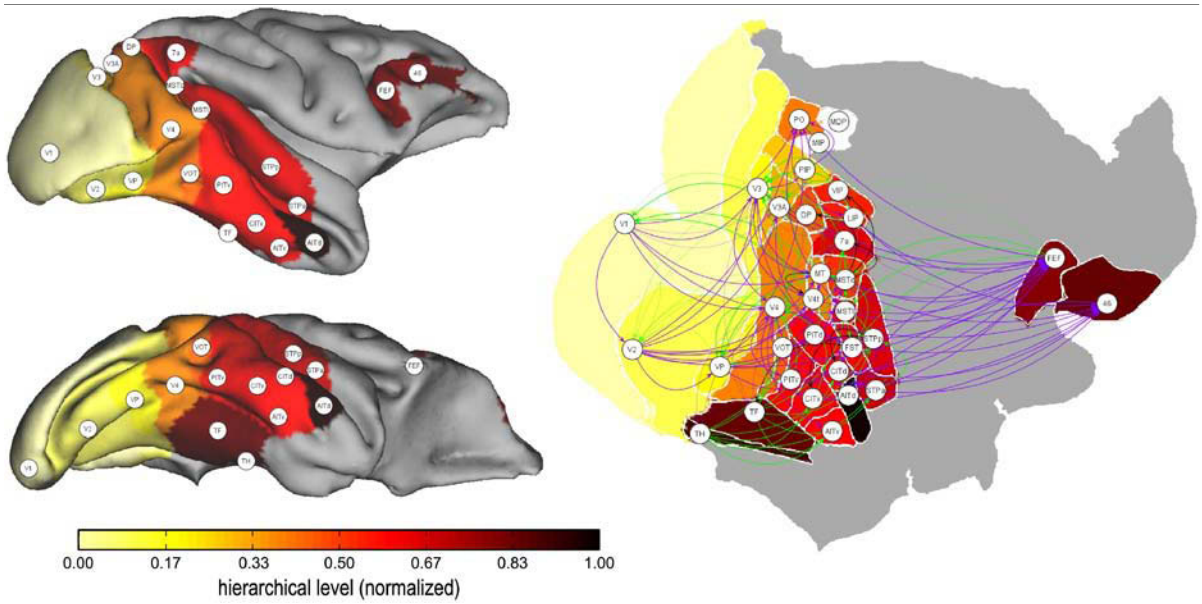


Figure 2.2.1. Classification of connections between visual areas on the basis of their laminar patterns. Here, magenta arrows refer to "ascending" projections, defined by termination in layer IV, and green arrows show "descending" projections, defined by termination in infra- and/or supragranular layers. Taken from (Reid et al. 2009).

Therefore, the general conclusion on the question "should we utilize animal studies in non-invasive human data analysis", would be the positive answer.

2.3 Brain atlases and their computational advancement

Development of neuroinformatics tools is commonly associated with the concept of the brain atlas. In a similar way to which a geographic map helps a traveler to reach the goal,

maps of the brain help neuroscientists to navigate within such a complex organ. The brain atlas, as a concept, appeared ages ago. The first human brain atlas that is still widely exploited by the research community is the aforementioned Brodmann map – a light microscope cytoarchitecture-based parcellation. Many subsequent studies did not rely upon the Brodmann partitioning, since no original data from that study were preserved. Therefore, consequent studies (Vogt and Vogt, 1919; von Economo and Koskinas, 1925) developed new terminologies that resulted in widespread confusion which still persists. In those early years of brain atlas development, such inconsistency did not cause much trouble, since there were just a few partitioning schemes available at all. In the middle of twentieth century, techniques started to become more sophisticated, which resulted in the appearance of more advanced anatomical studies on the primate brain (e.g., von Bonin and Bailey, 1947). As a result, the complexity of data also increased, and relating such heterogeneous datasets became a problem of particular difficulty, virtually impossible to accomplish without the aid of computer.

Early macaque atlases were generally limited to brain stem structures, only seldom describing gyral and sulcal patterns in cerebral cortex (Müssen 1967; Smith et al. 1972; Szabo and Cowan 1984; Martin and Bowden 1996). The first macaque atlas with a full cortical parcellation was published by Paxinos et al. in its first edition in 2000 (Paxinos et al. 2000; see Chapter 4 for description of the tool associated with this atlas). At present, there are at least three macaque atlases which include parcellations of cerebral cortex and corresponding information from several data modalities available (Black et al. 2004; Van Essen et al. 2005; Saleem and Logothetis 2006; already mentioned Paxinos et al., 2000).

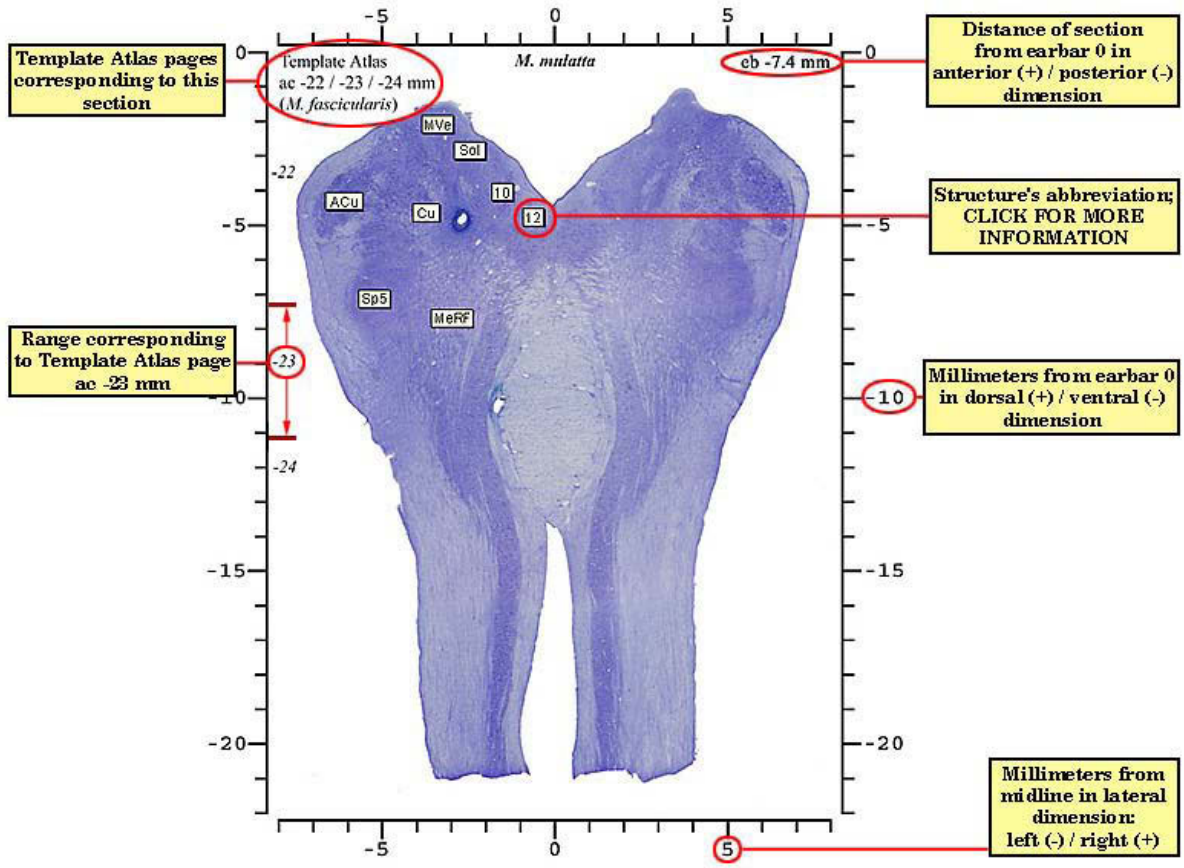


Figure 2.3.1. The web version of the (Smith et al. 1972) macaque brain stem atlas. Source: Brainstem Atlas of the Rhesus Macaque (Macaca mulatta) in Sitting Posture, BrainInfo, Regional Primate Research Center, University of Washington (2001).

In regard to the human brain, in the late 80's there has been an attempt to superimpose Brodmann's parcellation onto a three-dimensional map obtained from a single subject (Talairach and Tournoux, 1988); this atlas is also used nowadays, despite numerous limitations including coarseness and uncertainty of the Brodmann map.

Such an explosion of various atlases precipitated an urgent need to handle them in an appropriate way, enable interoperability between them, and provide a flexible user interface suitable for a broad community of researchers. For these purposes, the Talairach and Tournoux atlas was highly acclaimed, since it provided the option of attributing activations observed in neuroimaging experiments to the Brodmann areas directly. Alongside this, user-independent probabilistic techniques for development of human parcellation appeared – on the basis of cell architecture (Schleicher et al., 1998), as well as anatomical landmarks derived from MRI data (Tzourio-Mazoyer et al., 2002). Hence, one can note that the paradigm of the primate brain atlas development started to shift gradually towards the development of informatics tools supporting these atlases and enabling flexible data exchange.

Chapter 3

SORT: an approach for extraction of logical statements from spatial data

The abbreviation *SORT* stands for *Spatial Objective Relational Transformation* and represents an automated approach for obtaining logical relationships between cortical areas in different brain maps registered in the same Euclidean space. Given the recent abundance of voxel-based neuroimaging data that provide us with coordinate-based information about the location of differently defined areas in the brain on the one hand, and coordinate-independent, parcellation-based mapping data from animal studies on the other hand, such a mapping becomes more and more important.

The approach makes an objective conversion of spatial relationships to logical statements, among them are identity, inclusion, expansion or general overlap, between areas that were delineated by different methods, in different individuals, or mapped to three-dimensional space using diverse warping algorithms. The concept is implemented in the Java 2 programming language, allowing for multiplatform use and hence providing additional flexibility. An example of practical use of this technique is given; the SORT approach is validated against a database of the coordinate-independent statements that have been inferred using other methods.

The overall approach was devised in collaboration with Antje Krumnack at the Institute of Computer Science of the Heinrich Heine University of Düsseldorf.

3.1 Problem statement

In recent years, a vast amount of data has been produced from studies of the mammalian brain. Alongside, the number of methods to store, process, and analyze these data increased even more dramatically. Most of these data are unimodal measurements in individuals or small groups with no explicit reference to spatial characteristics. With the introduction of spatially registered neuroimaging data and the possibility of multi-modal techniques, in vivo measurements have become more precise and valid. On the other hand, the data from preceding invasive studies on non-human primates are sufficiently valuable to perform comparisons between coordinate-independent data from parcellation schemes, and those data registered in geometric space. In this endeavour, we develop methods providing objectivity, reliability, and validity in conversion of geometric data to parcellation schemes taken

according to a user preference. We demonstrate that this technique can also be utilized for a comparison between data from different modalities of neuroimaging.

Coordinate-based methods usually provide a relatively high degree of spatial precision. However, a substantial degree of variability and uncertainty in surface- and volume-based registration remains (Amunts et al., 1999; Bjaalie, 2002; Cachia et al., 2003; Mangin et al., 2004a, 2004b; Roland et al., 1997; Zhou, Thompson, & Toga, 1999). The problems in both surface- and volume-based registration arise from the variability of brain shape, because data are collected from different individuals with highly variable cortical structure; apart from that, methodologies employed in different experiments are also highly variable. Considering a particular approach for warping between individual brains and population templates, for instance, sulcal landmarks and surrounding gyri often appear to be inconsistent (Fischl, Sereno, Tootel, & Dale, 1999; Zilles et al., 1997). In addition, important information appears frequently lost as a consequence of utility of standard templates.

3.2 Coordinate-based versus coordinate-independent data

Volumetric data from neuroimaging experiments are represented in voxels (volumetric units, analogous to pixels in two-dimensional images), to which particular properties such as intensity or time series are assigned. Such data can be translated to surface (2-manifold) representations to make further manipulations more convenient (Van Essen et al., 2003; see 3.2.1). By contrast, coordinate-independent mapping typically represents entities by text supported with figures (Petrides & Pandya, 2007). Therefore, it does not supply localization parameters, but instead offers descriptive vocabulary where entities are linked by rules. This kind of representation offers additional advantages that serve as an important aid for neuroimaging methods. Although parcellation-based studies also include data on individuals with a number of variability patterns, the knowledge base provided by them is highly tolerant for being applied to other brain mapping data. Up to this point of time, there has been a limitation imposed by the fact that both of aforementioned datasets should be either coordinate-based or coordinate-independent; stepping out of these limits, here I propose an approach for linking both classes of entities together.

3.2.1 Spatial dataset example: SuMS database

Spatial data for exemplifying the SORT approach was taken from the SuMS (Surface Management System) database, which contains data on cortical structure and function regis-

tered to surfaces (Dickson et al., 2001; Van Essen et al., 2003). For taking this approach into account, one has to consider the cortex as a folded sheet (or 2-manifold), so that the spatial data representation can be reduced from three to two dimensions. This makes data processing, visualization and analysis easier in several aspects – namely, the original (fiducial) surface can be smoothened, unfolded or represented as a flat map, so that deeper structures buried inside the sulci become visible. One drawback of this approach is that it obviously reduces the information on the cortical profile, which can differ significantly across layers. Moreover, most of subcortical structures are either hard or impossible to render in such a way. Nevertheless, for validating our SORT approach, these disadvantages do not pose a problem – moreover, a major benefit was the fact that the SuMS dataset included a large amount of data from our textual database, CoCoMac, which is described in the next subsection. Indeed, these data were represented in different formats (spatial versus non-spatial) before application of the SORT procedure.

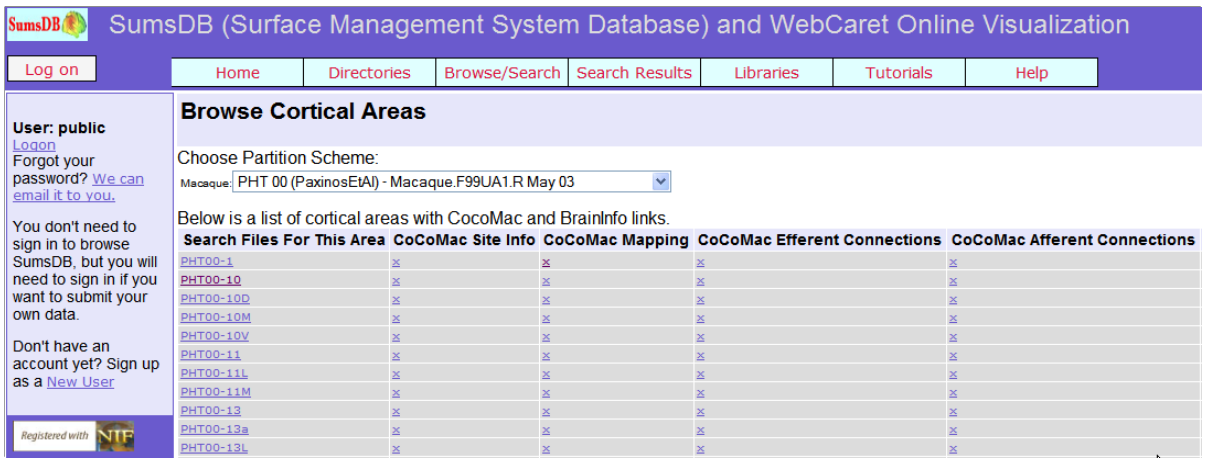


Figure 3.2.1.1. A snapshot of the SuMS database interface for browsing areas related to a particular partitioning scheme (here: Paxinos et al., 2000).

3.2.2 Textual database example: CoCoMac

The database called *the Collations of Connectivity data on the Macaque brain* (as an acronym: CoCoMac; www.cocomac.org; Stephan et al. 2001; Kötter 2004) consists of a systematic collation of macaque tracing and mapping data from more than 400 studies. Nowadays, invasive tracing studies on the primate brain still provide the most explicit and comprehensive information about anatomical wiring, as compared to neuroimaging data. These reports usually give descriptions of microstructural delineation methods, tracer application and transport sites, tracer description (anterograde versus retrograde), relative amount of

label (connection strength), and further characteristics of label (whether it appears at soma or axonal terminals) (Kötter 2007). Sometimes these studies provide additional information, including laminar patterns of connections (for both origin and termination sites), or even the number of neurons with relative contribution of surrounding connections to such quantity; these occasions were exceptionally rare in previous studies, although will be expectedly ubiquitous in forthcoming investigations given the advancement of modern tracing techniques. To make sure that the database can handle such a diversity of information and be adapted to newly introduced concepts, while preserving the general structure developed originally, there are further metadata available such as precision of description code (PDC) for each data entry. The PDC provides a convenient handling of contradictory statements and gives higher weight for more reliable entries. This concept is very important since the database is being constantly updated with recently introduced tracing and parcellation studies (Lewis and Van Essen 2000a, b; Takada et al. 2004; Morel et al. 2005, Gamberini et al., 2009). The relation of different tracing studies to each other is of particular relevance, given that comparisons between available studies (within or between representation modalities, spatial vs. non-spatial) provide more significant outcomes than each of these investigations taken separately: one of them might describe fine-grained partitioning, another one could provide details on connectivity on large scale interactions or within a particular functional subsystem – relating them both allows for a greater wealth of important information. For relating these studies within a modality (non-spatial data), a procedure called ORT has been devised (Stephan et al., 2000) – this approach has been previously utilized by the CoCoMac database (Stephan et al., 2001). Now we propose a link between spatial and non-spatial modalities.

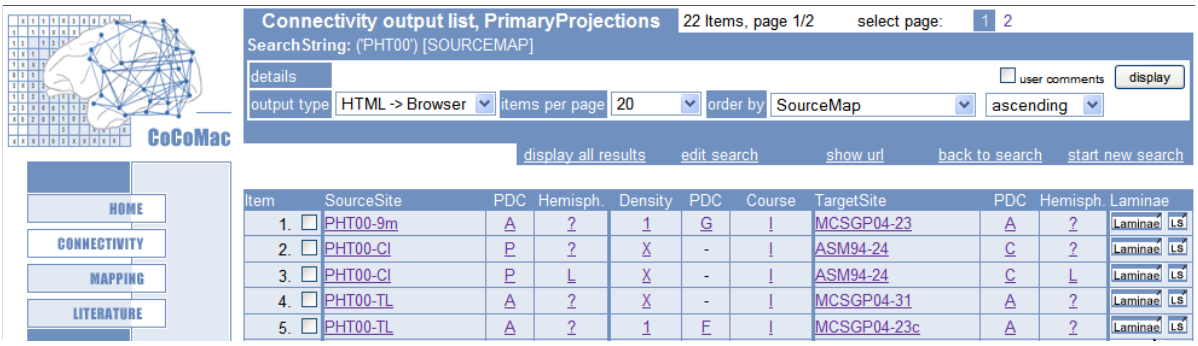


Figure 3.2.2.1. A snapshot of the CoCoMac database interface showing connectivity information for Paxinos et al. atlas regions.

3.3. Approach description

In the context of surface-based data representation, one treats a brain area as a contiguous set of nodes on the surface representing the cortical sheet. Different parcellation schemes can be overlaid on such surfaces, and each of these schemes will have non-overlapping (in the spatial sense) areas. Each of these areas would then be merely a subset of nodes, connected by edges and covered by triangular tiles, making up the entire cortical sheet. Therefore the question of identifying relationships between areas registered in space is a set theory problem.

Given two cortical maps *X* and *Y* registered on the surface, our procedure can be summarized in such a way:

(1) find all the areas in the map *Y* that share common nodes with a specified area in the map *X*;

(2) compute relative sizes of the intersections between the specified area in the map *X* and all coextensive areas in the map *Y*;

(3) associate the measurements obtained in the step (2) with the relationship statements between corresponding pairs of areas from the CoCoMac database; in the latter dataset, both the ones taken directly from the literature and those inferred with the ORT algorithm (Stephan et al., 2000) are utilized;

(4) on the basis of this association, infer a reasoning gauge to be learnt by the SORT algorithm for producing relationship statements between a specified area in the map *X* and all related areas in the map *Y*.

Here follows a more detailed description of these steps.

1. Each node in the surface template is assigned an ID; therefore finding out whether an area from the map *X* has common nodes with some area in the map *Y* is simply checking whether these areas have common IDs in their vertex lists.

2. The computation of areal relationships must satisfactorily tackle irregularities in surface topology and parcellation scheme overlays. To ensure that, we computed area sizes and degree of their intersection; the ratio between the latter and the former was an intuitive measure to consider: it captures the extent of overlap between areas. To validate this measurement, we took the subset of all representations of primary visual cortex in different cor-

tical maps. Since each of them represents the same structure, these areas should possess the "Identity" statement in relationship to each other.

To the area *A* from the map *X* and area *B* from the map *Y*, the following equations are applied:

$$R_A = \frac{\#(A \cap B)}{\#(A)}$$
$$R_B = \frac{\#(A \cap B)}{\#(B)} \qquad (3.3.1)$$

where the notation #(*X*) means the size of the set *X* – or, in other words, the number of nodes in the set *X*; this notation is compatible with that used elsewhere (e.g., Ohser and Mücklich, 2001). R_A signifies the fraction of nodes of A that are contained in the intersection of *A* and *B*, and R_B stands for fraction of nodes of *B* in the intersection of *A* and *B*.

These measures, applied to the data on primary visual cortex taken from different maps, gave results that are shown in the Table 3.3.1 and Fig. 3.3.1. The closer all these values approaching R_X = 1, the closer areas resemble each other in terms of size and shape. Given that none of those values equals one, one can conclude that there are no pairs of areas sharing exactly the same subset of nodes with each other, even though these areas are considered to be identical in a nomenclatural sense.

Area A	Area B	R_A	R_B
RM-V1	LV00-V1	0.689126	0.985533
RM-V1	UD86-V1	0.638728	0.989645
RM-V1	B05-17	0.665824	0.99166
RM-V1	BB47-OC	0.676662	0.929759
RM-V1	FV91-V1	0.6875	0.992438
RM-V1	PHT00-V1	0.567197	0.947781
LV00-V1	UD86-V1	0.906742	0.982368
LV00-V1	B05-17	0.917334	0.95534
LV00-V1	BB47-OC	0.908809	0.87317
LV00-V1	FV91-V1	0.968742	0.977836
LV00-V1	PHT00-V1	0.800827	0.935708
UD86-V1	B05-17	0.959138	0.92198
UD86-V1	BB47-OC	0.958578	0.850087
UD86-V1	FV91-V1	0.975371	0.908735
UD86-V1	PHT00-V1	0.815281	0.879264
B05-17	BB47-OC	0.939736	0.866965
B05-17	FV91-V1	0.953457	0.92412
B05-17	PHT00-V1	0.787194	0.883187
BB47-OC	FV91-V1	0.870688	0.914733
BB47-OC	PHT00-V1	0.750558	0.912768

Table 3.3.1. R_A and R_B values for pairs of primary visual cortex representations in different maps. R_A and R_B are defined by the formula (3.3.1).

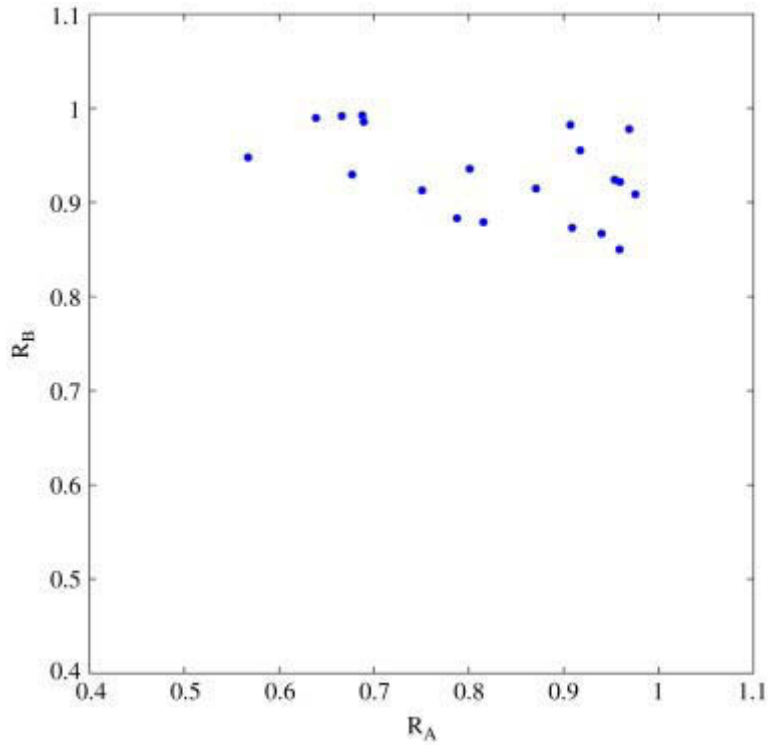


Figure 3.3.1. Visualization of data points represented in the Table 3.3.1

At this point we face the first occurrence of uncertainty concerning the degree of "closeness" of these values to 1. If we assume that two area representations from different maps have fully identical borders (or, in other words, share the same set of nodes), it would be the only instance when all values are equal to 1. Given all the uncertainties of delineations and transformations from one brain surface to another, from which the areas are derived, such perfect correspondence is highly improbable. By mere visual inspection of the dataset plot, we can conclude that there is quite a considerable scatter of the values, some of which are seemingly much less than one and therefore lie somewhat outside of the main point cloud. Data points related to the intersection between areas V1 in the newly defined Regional Map (RM-V1, (Kötter & Wanke, 2005)) and V1 as defined by Ungerleider and Desimone (1986) (UD86-V1) provide an illustrative example. Technically, this discrepancy is to a large extent due to irregular topology of the surface (see Fig. 3.3.2; Hilgetag and Barbas, 2005, 2006).

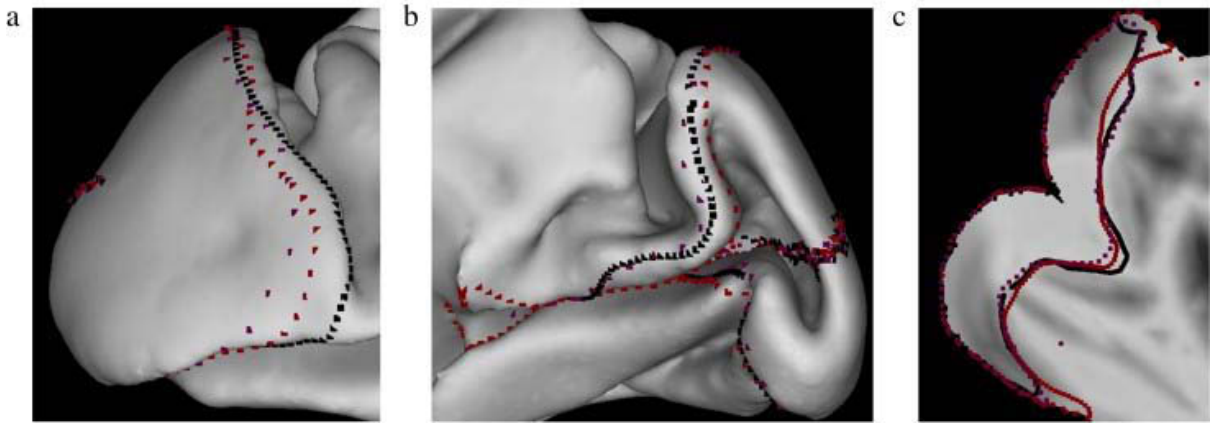


Figure 3.3.2. Variability of spatial positions of V1 borders as defined by different authors and stored in the SuMS database: Brodmann, 1905 – magenta; Bonin and Bailey, 1947 – red; Lewis and Van Essen, 2000 – black. On (a) and (b) plates, there are lateral and medial aspects representations, respectively; on (c), the same regions are displayed on flat surface.

One can notice that such inconsistencies mostly occur in those situations where the original surface data was converted from one template/individual to another. For instance, the primary visual cortex as delineated by (Paxinos et al., 2000) (PHT00-V1) is derived from a study in one individual, whose surface was initially represented in the SuMS database. Further, these data were converted to a general template to enable comparisons with other datasets – and in this process apparently lost fidelity to a certain extent. Interestingly, inconsistency with the Brodmann representation of V1 is not as prominent, although the data were originally represented in the human template and then converted to macaque surface. One possible explanation for this is the fact that the Caret software used for processing SuMS data exploits calcarine sulcus as one of the landmarks for interspecies (human-monkey) deformations; primary visual cortex embraces this sulcus, and thus distortion in this region appears to be minor even after deformation procedure. This point of view is confirmed by the fact that areas in temporal and frontal lobes from this representation of the Brodmann map are very inconsistent with their homologues in other maps, since the nearest landmarks are located quite remotely from those regions.

These problems had to be addressed algorithmically while implementing the SORT procedure. An intuitive solution was to weigh the nodes in the surface template in such a way that captures the probability of a node belonging to a particular area, but not another one. More specifically: since nodes lying “deep within” the area are more likely to belong to that area, they should have higher weights; by contrast, nodes lying close to areal boundaries, should have lowest weights due to the uncertainty of their localization. The approach is a simple one: nodes lying in the borders are assigned a value of 1; immediate neighbours

within this area are assigned value of 2 – and so on, until the "centre" of the area. Here follows a more formal description:

Let *M* be the set of all nodes on the surface template. Then

$$\forall\, n \in M\ \exists\ N(n) \subset M \qquad (3.3.2),$$

where *n* is the node on that surface, and *N*(*n*) is the neighbourhood of the node *n*, an enumeration of immediate neighbors such that:

$$n_1 \in N(n_2) \Leftrightarrow n_2 \in N(n_1) \qquad (3.3.3)$$

Let now *A* be the set of all nodes belonging to the area *A*, then $A \subset M$; an element from this subset is *a*, such that $a \in A$. The function $f_A: A \to \mathbb{N}$ gives a natural value to each node obtained by the following transformation:

$$f_A(a) = \begin{cases} 1, \text{ if } n \in N(a)\ \exists : n \notin A \\ \min\{f_a(n) + 1 :\ n \in N(a)\},\ if\ N(a) \subset A. \end{cases} \qquad (3.3.4)$$

This procedure weighs the nodes in such a way as we were aiming at: peripheral nodes are the "lightest", and nodes in the centre are the "heaviest" (see Fig. 3.3.3). Given the distortion of the borders and imprecise information about their positions (as on Fig. 3.3.2), this method serves efficiently for this purpose.

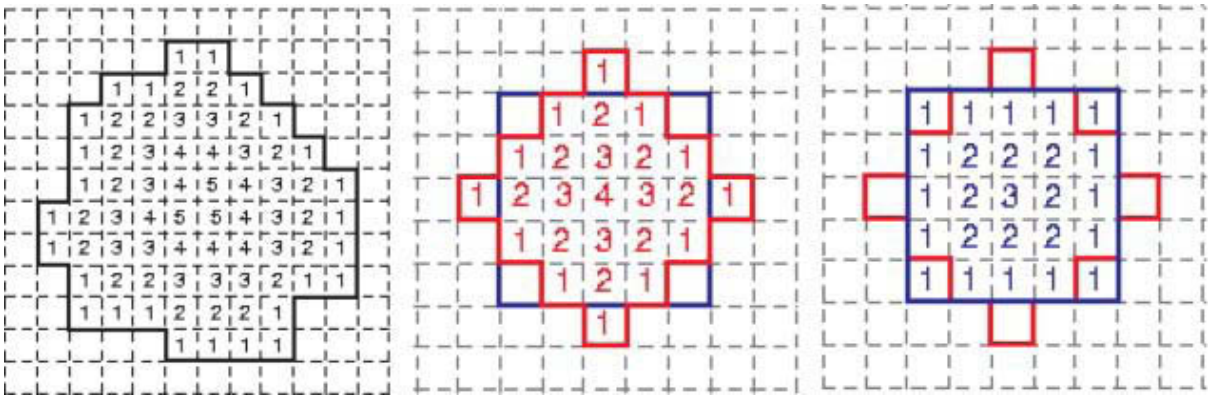


Figure 3.3.3. Illustration of the weighing procedure. A: an example weighed 71-node surface; B: a demonstration showing two abstract areas (colour-coded red (A) and blue (B)) which have a slight variability in border positions, but have the same portion of vertices in each of them, included in their intersection. The resulting values are as follows: $R_A = R_B = 0.84$, whereas $R'_A \approx 0.91, R'_B \approx 0.88$.

Speaking about values R_A and R_B obtained earlier (formula 3.3.1), one can now propose R'_A and R'_B on the basis of weight functions $f_A(a)$ and $f_B(b)$ from (3.3.4), such that:

$$R'_A = \sum_{a \in A \cap B} f_A(a) / \sum_{a \in A} f_A(a)$$
$$R'_B = \sum_{b \in A \cap B} f_B(b) / \sum_{b \in B} f_B(b) \qquad (3.3.5)$$

Figure 3.3.4 illustrates the efficiency of this approach: the point cloud containing logically identical regions is much less scattered, and R'_A and R'_B values become higher at that instance.

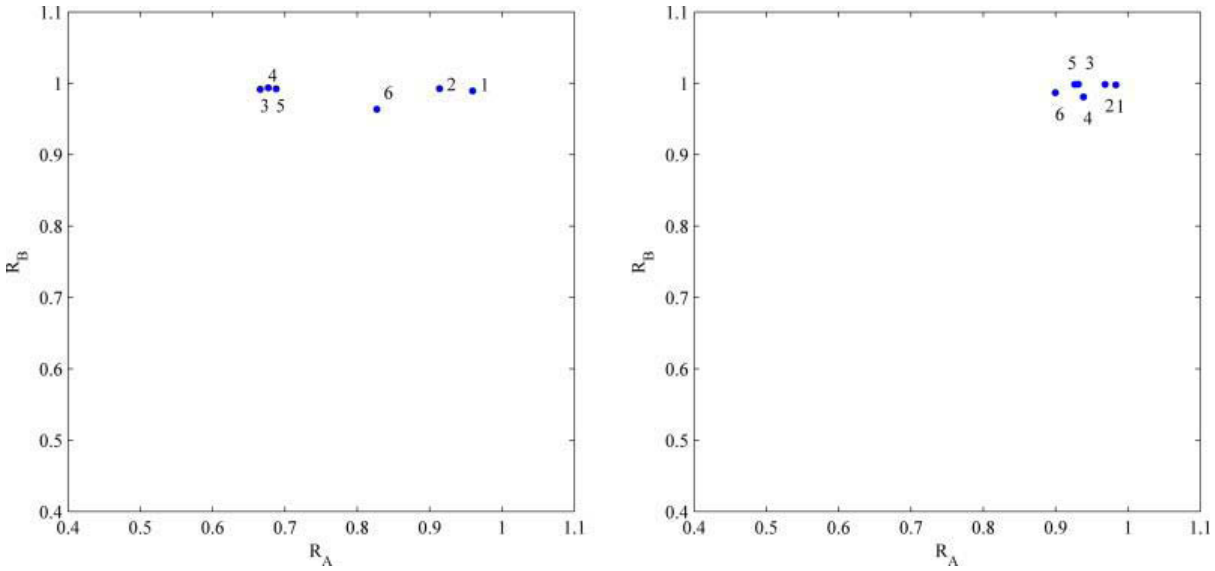


Figure 3.3.4. Comparison between several RC=I measurements taking into account homogeneous, or unweighted, nodes (left) versus weighted nodes (right).

3. Obtained values were further evaluated against CoCoMac data (those retrieved from the literature and generated by ORT). The resulting training set was used to establish appropriate constraints linking spatial relationships with logical statements. All the pairs of areas that had relationships with each other were checked for existence of statements in CoCoMac. Plotting R'_A against R'_B helped us to evaluate presumable regions that points with particular relationships must occupy (Fig. 3.3.5); it was an exploratory step, and these regions obtained by computational procedures described in the next subsection.

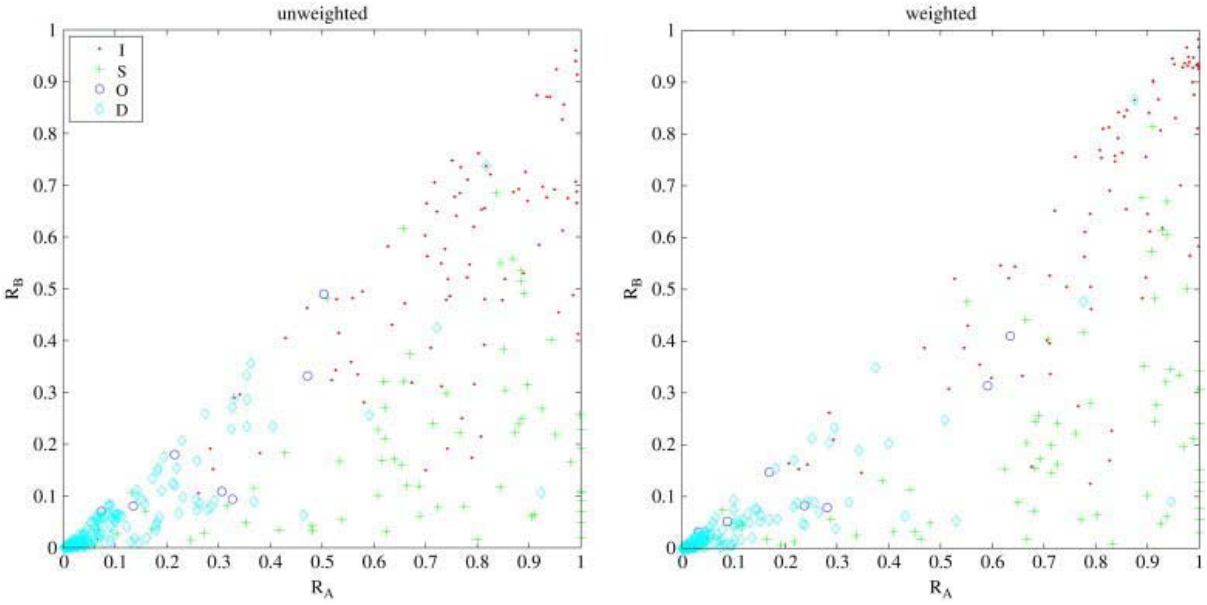


Figure 3.3.5. Training dataset based on statements obtained from surface data and also existing in CoCoMac. Newly introduced Regional Map is not included due to uncertainty of associated statements that were existing at that time.

The plot shown on Fig. 3.3.5 was standardized in such a way that $R'_A < R'_B$, so that all the points above the diagonal are mirrored to their appropriate positions below the diagonal. This space can be partitioned in such a way that data points standing for a particular relationship statement will be lying regularly within a particular region, and to a lesser extent in any other region. Namely, points with I-statement (identity) are lying in the vicinity of (1, 1). In the case case where A is included in B (A is a subarea of B), R'_A is approaching 1, while R'_B should be less than 1. If A is expanding over B (A is a larger area than B), R'_A is less than 1 and R'_B roughly equals 1; in this case, these points are not represented in the plot because of its standardization. We simply made use of a convenient property of L-statement (expansion) which is fully symmetric to S-statement (inclusion). O-statement (overlap) should be normally distinguished by low R'_A and R'_B, and disjoint areas have to be represented by data points lying around (0, 0).

4. As one can see in the output of the step 3, values related to particular logical relationship statements are considerably dispersed; nevertheless, particular patterns are obvious. Initially, we will disregard those data points whose positions are highly different from the main point cloud (for example, the red dots located close to R_A=0.3 and R_B=0.2, or those green points found in the vicinity of (0, 0) are inconsistent). Such occurrences might result from an insufficient specificity of logical relationship statements that were accessed, or surface to-

pology errors associated with these areas are increasingly high – or both. After such cases being eliminated, the training set is ready.

3.4 Training the model

To ensure that the ultimate output of the procedure provides a greater reliability, we devised three different approaches, each used for training the model on datasets with weighted and non-weighted nodes. The first approach is based on the assumption about data distribution, and therefore referred to as "hypothesis-driven". For the second model, we developed a theory about the locations that appropriate data points must presumably occupy. It is hence called a "theory-driven" approach. Finally, the third procedure involves partitioning the space fully on the basis of data, with no assumptions made. It is called therefore a "data-driven" approach. Each of these techniques and related issues are described in the following subsections.

3.4.1 Hypothesis-driven approach based on Bayesian statistics

When a researcher makes a hypothetical assumption about the domain of his or her interest, it is frequently convenient to make use of Bayes' theorem. In this framework, the sought-for value is represented by the posterior probability of the hypothesis given the evidence acquired from the observed domain. This value is obtained using its inverse, i.e. the likelihood of the evidence given the hypothesis - a parameter which can always be derived from any non-random dataset in a meaningful way. This approach can be especially useful for multidimensional spaces, where deterministic approaches may cause difficulties – hence its utility for the two-dimensional R_AR_B space might be desirable. The general form in which such a construction can be represented is as follows:

$$P(RC|(R_A, R_B), R_A \neq 0,\ R_B \neq 0) = \frac{P\left((R_A,R_B)|RC\right) P(RC)}{P(R_A,R_B)} \qquad (3.4.1.1),$$

where RC stands for "relationship code", i.e., the statement of relationship between two areas from CoCoMac (I, S, O or D). The term in the denominator, a normalizing constant, can be represented in the following way (Papoulis, 1984):

$$P(R_A, R_B) = \sum_{f:\{I,S,O,D\}} P\left((R_A, R_B)\middle|f\right) P(f) \qquad (3.4.1.2),$$

i.e., the sum of all likelihood values of positioning the data point on the R_AR_B space given each of these relationship statements.

At this point, the assumption on data distribution has to be made to supply an appropriate form for the first term in the numerator of (3.4.1.1). This term is the likelihood of the data point position given the relationship code (i.e., the evidence given the hypothesis, as mentioned above), and it should therefore be represented as a probability density function (PDF) in R_AR_B two-dimensional space, where each point in this space is assigned a probability of belonging to a particular class of a corresponding relationship statement. After visual inspection, it was concluded that the data distribution across both axes can be meaningfully approximated to normal one. Therefore we employed bivariate Gaussian distribution function with the mean having an appropriate value in two-dimensional space, and two variance values whose ratio determines the eccentricity of the ellipse representing the profile of the distribution (i.e., the extent to which the ellipse is stretched). For the O-statement (overlap), the eccentricity of the ellipse is 0, which means that it becomes a circle (i.e., both variances are equal: $\sigma_{R_A} = \sigma_{R_B}$). Such a property is chosen due to the fact that the number of O-statements in CoCoMac is much lower than that of any other statements, and the bivariate normal distribution of the points related to this statement can be rendered only circular ensuring an appropriate degree of accuracy. See Fig. 3.4.1.1 for an illustration of this concept.

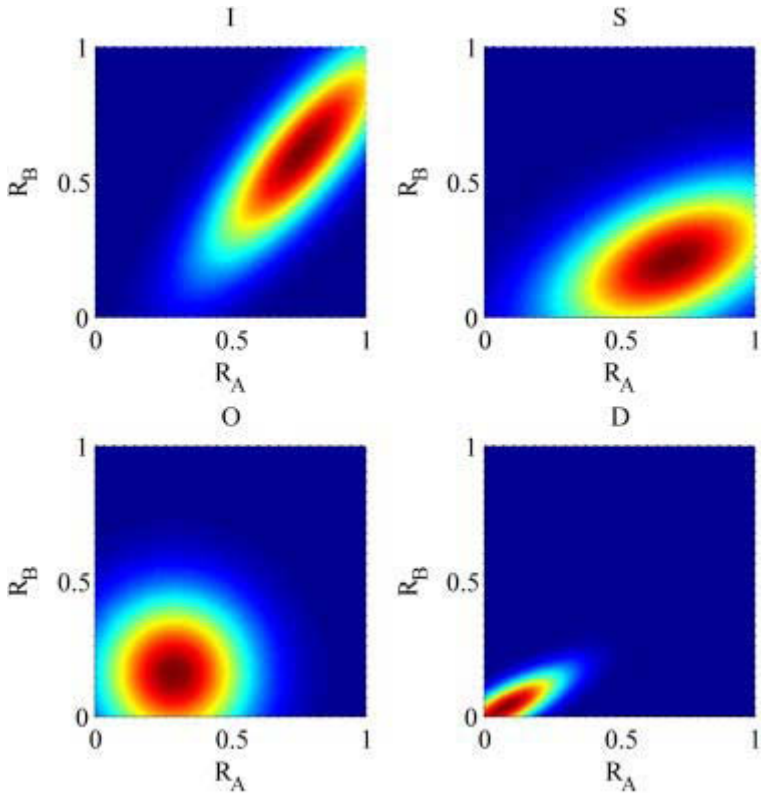


Figure 3.4.1.1. Likelihood of R_AR_B values given *RC*: {I,S,O,D}, as computed by (3.4.1.1); here, only those data generated with weighted nodes are taken into account.

The likelihood parameter in our model is therefore obtained in a following way:

$$P\big((R_A,R_B)\big|RC\big)=\frac{\exp\left(-\frac{\left(R_A-\mu_{R_A}\right)^2}{2\sigma_{R_A}^2}-\frac{\left(R_B-\mu_{R_B}\right)^2}{2\sigma_{R_B}^2}\right)}{2\pi\sigma_{R_A}\sigma_{R_B}} \qquad (3.4.1.3)$$

By introducing the flat prior ($P(RC)$ = 0.25), replacing the first term in the numerator of (3.4.1.1) by the right part of (3.4.1.3) and denominator by the right part of (3.4.1.2), we get the solution (see also Fig. 3.4.1.2):

$$P(RC|(R_A,R_B),R_A\neq 0,\ R_B\neq 0)=\frac{\exp\left(-\frac{\left(R_A-\mu_{R_A}\right)^2}{2\sigma_{R_A}^2}-\frac{\left(R_B-\mu_{R_B}\right)^2}{2\sigma_{R_B}^2}\right)}{8\pi\sigma_{R_A}\sigma_{R_B}\sum_{f:\{I,S,O,D\}}P\big((R_A,R_B)|f\big)\,P(f)} \qquad (3.4.1.4)$$

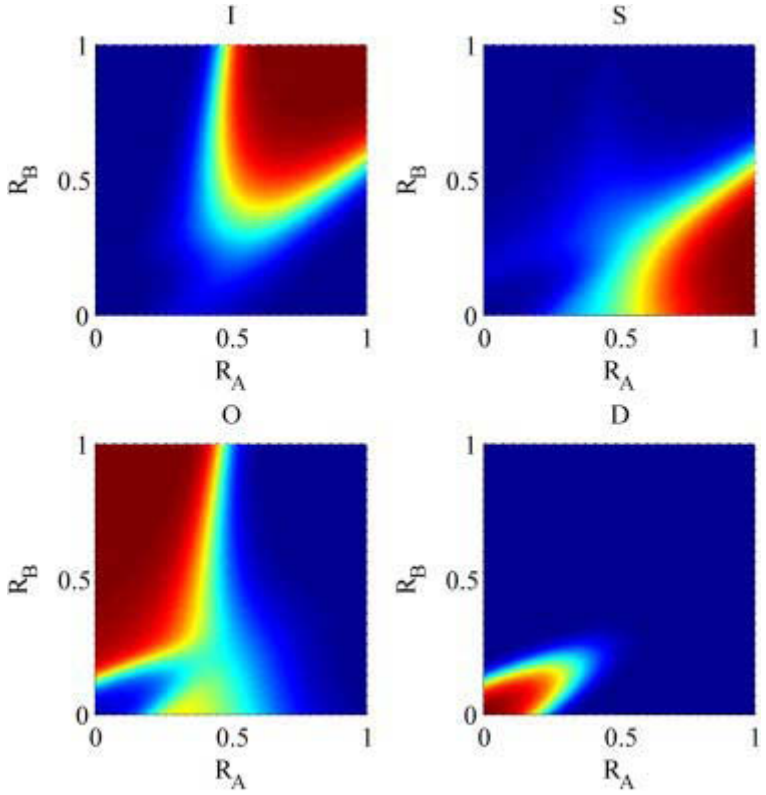


Figure 3.4.1.2. Posterior probability of RC given the location in R_AR_B space, as computed by (3.4.1.4); here, only those data generated with weighted nodes are taken into account.

It is natural to assume that the data we are dealing with has a normal distribution. Choosing a flat prior is also justified due to the fact that, in terms of spatial relationships, there can be any of existing statements appearing with equal probability, even though the statements in CoCoMac are distributed unevenly (e.g., there are much less O-statements than any other). The approach allows us to parcellate the R_AR_B space in such a way: for each data point, one

has to find the posterior probability assignments for each of the relationship statements, derived by (3.4.1.4), and then take maximum of these assignments. As a result, each of the relationship statements is assigned to a contiguous region on R_AR_B space (Fig. 3.4.1.3), and each point in this space possesses a value of probability of being assigned to a particular relationship code.

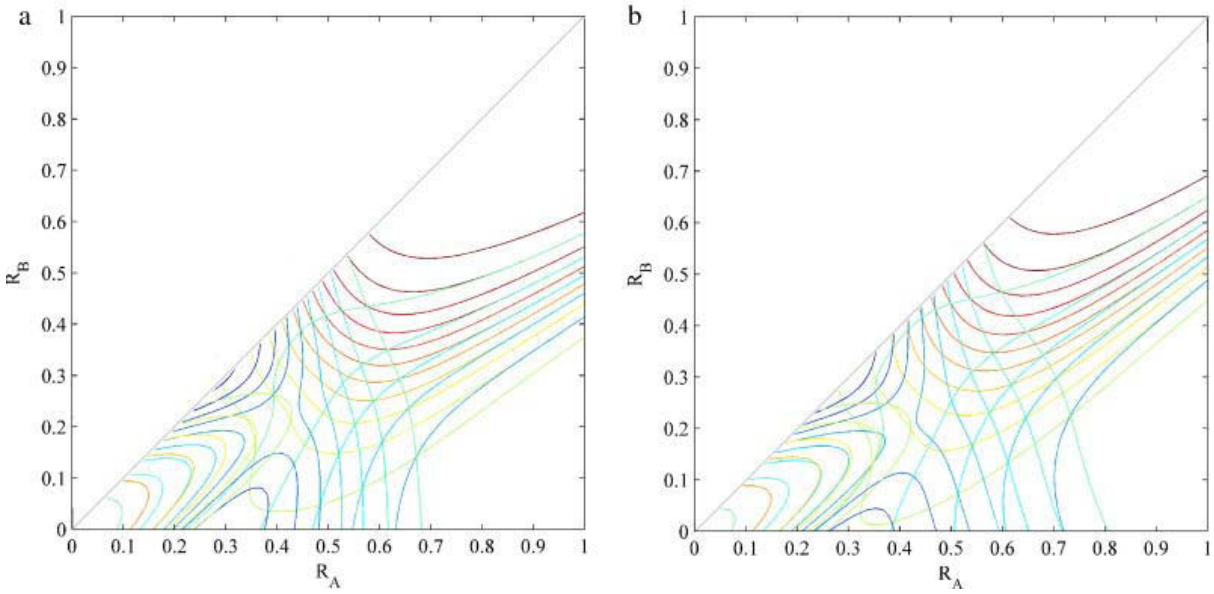


Figure 3.4.1.3. Partitioning the R_AR_B space using hypothesis-driven approach.

3.4.2 Theory-driven approach

As was mentioned previously, assuming that the subset of data related to a particular relationship statement has Gaussian distribution is always an approximation which is quite different from the reality. Hence it might be of potential utility to devise a complementary approach based on the theory about presumable regions that data points assigned to particular relationship statements must occupy. The notion is rather intuitive as one takes a look at Figure 3.3.5 and may conclude that most of the I-points are located in the vicinity of (1, 1) – i.e., infinitely small neighbourhood of (1, 1) would refer to absolutely identical areas, whereas in reality this neighbourhood is rather large. Similarly, D-points are normally occupying the neighbourhood of (0, 0) in a comparable way. S-points are then lying within the region bordered on the left by $R_A = 1 - \varepsilon, \varepsilon > 0$. Infinitely small ε would be the ideal case. O-points occupy the rest of R_AR_B space. Therefore the problem was reduced to optimization of three parameters: radii of I- and D-regions and width of S-region (Fig. 3.4.2.1).

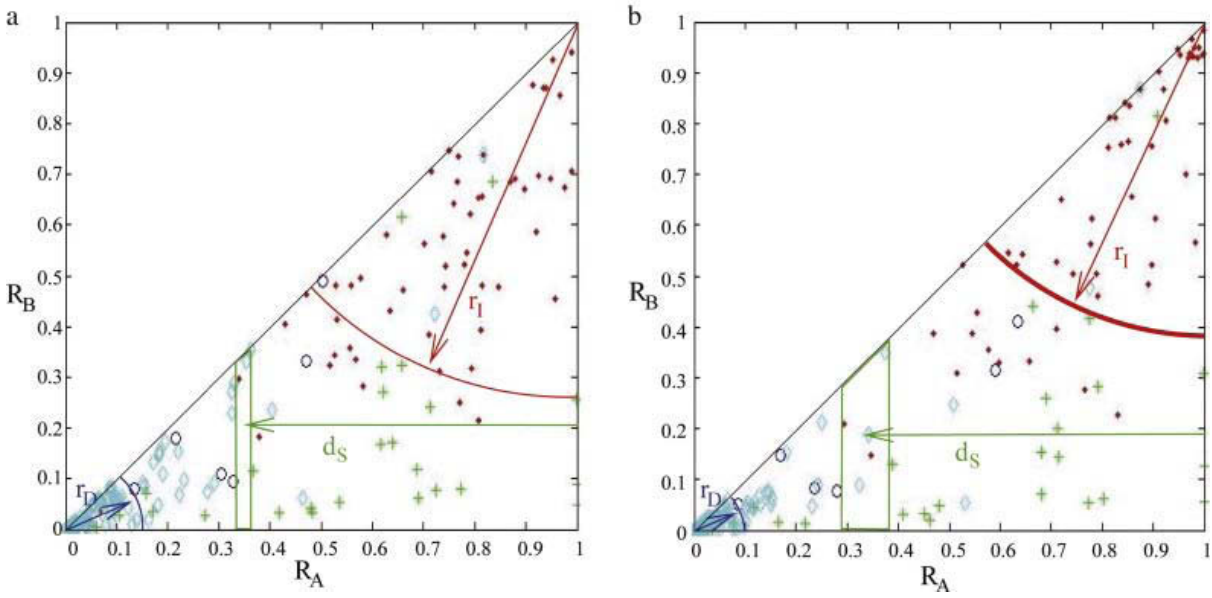


Figure 3.4.2.1. Partitioning the R_AR_B space using the theory-driven approach. Here, r_I, d_S and r_D are values to be optimized.

Ratio values in this figure are fractions of appropriate points within a region:

$$RC_{k\_ratio} = \frac{\#[X_k(R_{A_i},R_{B_j} | X_k \xrightarrow{yields} RC_k)]}{\#[X_k(R_{A_i},R_{B_j})]}; \qquad RC_k: \{I, S, O, D\} \qquad (3.4.2.1)$$

Ratios are the following: i_ratio = 0.75; s_ratio = 0.667; o_ratio = 0.667; d_ratio = 0.747 — for unweighted, and i_ratio = 0.732; s_ratio = 0.692; o_ratio = 0.667; d_ratio = 0.736 — for weighted nodes regions.

This approach is proposed by Antje Krumnack and described in more detail in her thesis (Krumnack, 2008).

3.4.3 Data-driven approach based on Voronoi Diagrams

It is always necessary to have a "backup" approach which treats the data as they are, without any presupposition about their distribution. Separation of space into regions using Voronoi diagrams has a wide application because of its simplicity and omnipresent occurrence in nature; methods for their efficient computation are reported elsewhere (e.g., Bowyer, 1981). The approach we employed can be described in a following way:

(1) explore the space within and in its neighborhood of R_AR_B, the space $R_A^*R_B^*$, such that $\{R_A^* \in [-1, 2] \Leftarrow R_A \in [0, 1]\} \wedge \{R_B^* \in [-1, 2] \Leftarrow R_B \in [0, 1]\}$;

(2) within $R_A^* R_B^*$, find a quadruple of points {I,S,O,D} to which most of the points of a particular RC-class have the closest proximity (defined globally for all RC-classes altogether)

The quadruple of points defined in the second step specifies cells where data points related to a particular statement should be located (Fig. 3.4.3.1):

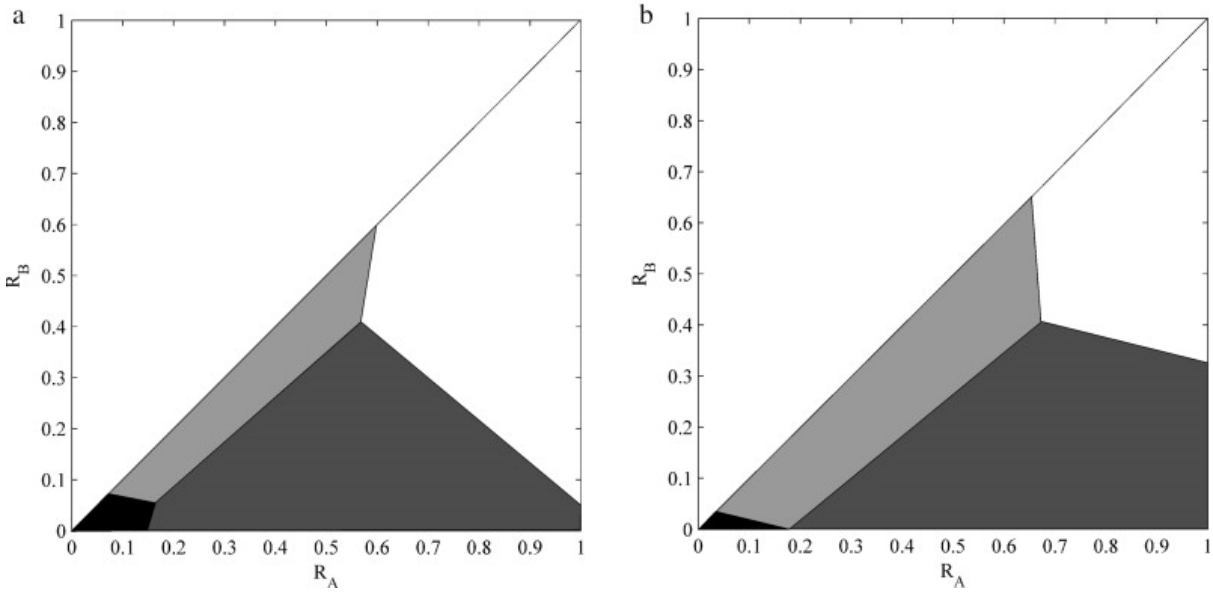


Figure 3.4.3.1. Partitioning the $R_A R_B$ space using the data-driven approach. Locations of points in the $R_A^* R_B^*$ space and their ratios are the following: I (1.8, 0.4), ratio = 0.732; S (0.8, −0.8), ratio = 0.667; O (−0.6, 0.8), ratio = 0,667; D (−0.8, −0.4), ratio = 0.714 for unweighted nodes data, and four points I (1.8, 1.4), ratio = 0.679; S (1.2, −1), ratio = 0.654; O (−0.6, 1.2), ratio = 0.667; D (1, −0.8), r atio = 0.678 for weighted nodes data (B).

The ratio values are defined by (3.4.2.1).

3.5 Results and their utility

The training dataset we used for our procedure contained statements in both CoCoMac and SuMS databases – namely, it included 8 parcellation schemes described in section 3.3. Evaluation of the procedure was made against two independent datasets: (1) all statements taken from the procedure output, given that they were not present in CoCoMac; (2) statements related to areas from the Regional Map, a recently proposed primate species-independent parcellation scheme (Kötter & Wanke, 2005). Although several statements from this map were already entered into CoCoMac before introduction of SORT, their certainty was highly questionable.

Taking the first evaluation dataset into account (statements generated by the procedure and not present in CoCoMac), one can note that the matching of data across methods described in sections 3.4.1-3.4.3 was consistent in 68% of cases for non-weighted nodes data and in 74% for weighted nodes data. These numbers are consistent with ratio values signifying fractions of appropriate data points devised for methods 3.4.2 and 3.4.3. This circumstance also confirms the fact that each of the three approaches has different requirements and applications – see next section for more detail on this issue.

It was also found within the same dataset that assignments of probability derived from the multivariate normal distribution were, on average, 84% for statements consistent across methods, and 64% for inconsistent ones. Given that the minimum probability assignment to any of the relationship statements is 33%, this difference is sufficiently significant for constraining the output to ensure statements have high fidelity.

Remarkably, comparison with CoCoMac produced even more striking evidence for the efficiency of employment of three different techniques and weighing the nodes. Data on weighted nodes exhibited higher consistency between CoCoMac RCs and those RCs generated by hypothesis-, theory- and data-driven approaches. Likewise, probability values assigned to relationship codes obtained by the hypothesis-driven approach were increased for data with weighted nodes as well (see Table 3.5.1).

Rates of matching to CoCoMac relationships

	unweighted	**weighted**
Data-driven	0.632	0.704
Hypothesis-driven	0.776	0.796
Theory-driven	0.684	0.717

Probability values obtained from the posterior PDF

false	**true**	**false**	**true**
0.713	0.894	0.721	0.92

Table 3.5.1. Comparison of the SORT output with CoCoMac statements. Top: the rates of matching between CoCoMac RCs and those produced by three approaches specified in the first column. The rates are generally higher for data from surface with weighted nodes, compared to those obtained from the homogeneous surface. Bottom: PDF values obtained from the hypothesis-driven approach; for each weighted and unweighted nodes data, there are average probabilities of non-matching (false) and those of matching (true) with CoCoMac relationship codes displayed.

Despite the fact that the data produced by SORT are generally consistent with the data in CoCoMac, we realized that knowing the sources of occasional inconsistencies would be re-

levant for necessary adjustments to the routine and, on the long run, for forthcoming investigations in the field. Figure 3.5.1 highlights these cases.

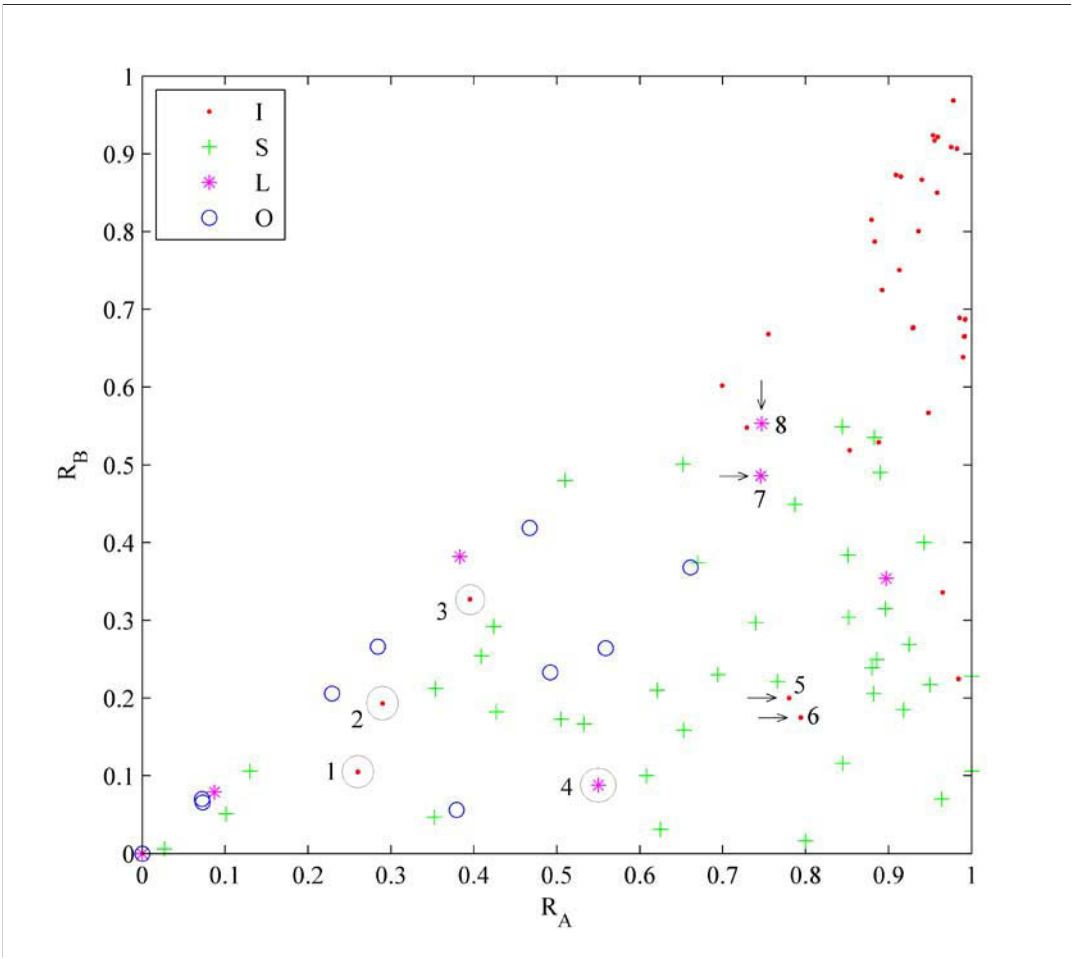


Figure 3.5.1. Several highly inconsistent cases. See main text for their analysis.

Explanations of the most persistent problems are as follows.

1. BB47-TB↔RM-A2=I,

2. BB47-IB↔RM-Ip=I,

(The area TB (IB) in Bonin and Bailey (1947) is identical to the area A2 (Ip) in Regional Map, Kötter and Wanke (2005)),

as declared in CoCoMac, while in both cases our procedure perceived them as non-identical. A possible explanation is that the BB47 parcellation scheme is delineated imprecisely. It exhibits gradually fading borders, which could explain the diversity of statements in the literature in regard to this issue. The authors indicate precisely (p. 38): "…(areas – G.B.) TC and TB

in the macaque cannot be clearly demarcated from each other". Contrarily, area IB has well defined boundaries, although even then this representation remains very imprecise, because the laminar pattern of that area is unclear. In addition, the concept of the Regional Map (RM) is defined by converging topographical, microstructural and functional criteria commonly consistent across individuals. As was mentioned in 3.2.2, CoCoMac associates Precision of Description Code (PDC) to each entered statement; in the absence of original information from the study, the RC only be assumed given its name or by comparing the maps using anatomical landmarks, due to unavailability of sufficiently detailed information, which was initially the case for the Regional Map before SORT statements were entered into the database.

3. RM-CCs↔PHT00-25=I,

4. B05-3↔RM-S1=L.

(Paxinos et al. (2000) is the stereotaxic atlas of macaque brain; Brodmann (1905) is the map of macaque cortex created by Brodmann).

Both problems arose primarily from shape variability between different subjects. The PHT00 map is a delineation study representing brain structures in a stereotaxic space (see Chapter 4 for more details). For the purpose of spatial comparison between various parcellation schemes, this map was deformed from its original stereotaxic space, to the standard template exploited by our procedure. Concerning the second case, Brodmann's map is a "fundamental" parcellation of the cortex of many species, given that cercopithecus is close to macaque in terms of brain morphology. This statement is nevertheless most likely an error made by the data collator and should be merely deleted.

Notably, our procedure revealed several gaps in CoCoMac; this is especially the case for cingulate, prefrontal, temporal and parietal cortex (Fig. 3.5.2). For instance, CoCoMac introduced 14 target areas for centrotemporal cortex in the Regional Map (RM_TCc), whereas the SORT procedure outputted 25 relationship statements associated with that area, which is almost twice as many. Given that the number of target maps in CoCoMac is obviously much higher than the number of parcellations already delineated on the surface template, increasing the number of parcellation schemes drawn on surface would increase the number of statements drastically. The data generated by SORT were subsequently uploaded to the CoCoMac database to ensure those gaps mentioned above are eliminated to a large extent.

As a quantitative analysis of the efficiency of SORT, we evaluated the ORT output for both the case where the database included statements generated by SORT, and the case where these data were not entered yet. Specifically, the output represents the result of the Floyd-Warshall algorithm for finding all possible mapping paths between various brain regions in different maps stored in the database. The statistics on this output are shown in the Table 3.5.2. One can note that along with the number of mapping statements, their uncertainty also increased in several aspects, but in general SORT statements had a very significant impact. Here we provide a full justification of this notion. First, the number of trusted resulting relations has increased by 45.4%; the term "resulting relations" stands for those statements that are produced by the Floyd-Warshall algorithm (see e.g. Belousov and Tkachev, 2004) for calculating all the shortest paths between all areas. This is different to the term "primary relations", which stands for those statements derived directly from literature (or algorithmically, in case of SORT). With such an increase of trusted statements, the number of those with conflicting relation codes also increased, but merely by 44.2%, which is slightly less than for trusted relations. "Conflicting relation codes" are related to identical pairs of areas with different relationship statements – for example, for areas A and B it might be stated that they are identical, but also that they are related in some other way (e.g., partial overlap). Such inconsistencies, including the last two rows of the table, are resolved by ORT at its last step.

ORT output statistics	without SORT	with SORT
Elapsed time	61 seconds	119 seconds
Number of updates	2285064	4453638
Number of brain sites	5594	5629
Number of primary relations	15984	16744
Number of primary conflicts	56	92
Total number of resulting relations	1000018	1847396
Number of trusted resulting relations	545880	793516
Of these, ... have conflicting relation codes	52518 (9.6%)	75746 (9.5%)
Number of paths with invalid chain of relation codes	0	0

Table 3.5.2. The ORT output evaluation, taking into account database entries with and without SORT statements.

After filling the gaps in mapping data, the connectivity output set is also changed in a meaningful way. Since the number of mapping statements is increased, one can obtain related connectivity from a greater number of maps (Fig. 3.5.2). Moreover, some new regions were

introduced, namely G (gustatory cortex), VAC subdivisions (VACd and VACv) and PFCorb subdivisions (PFCoi, PFCol and PFCom).

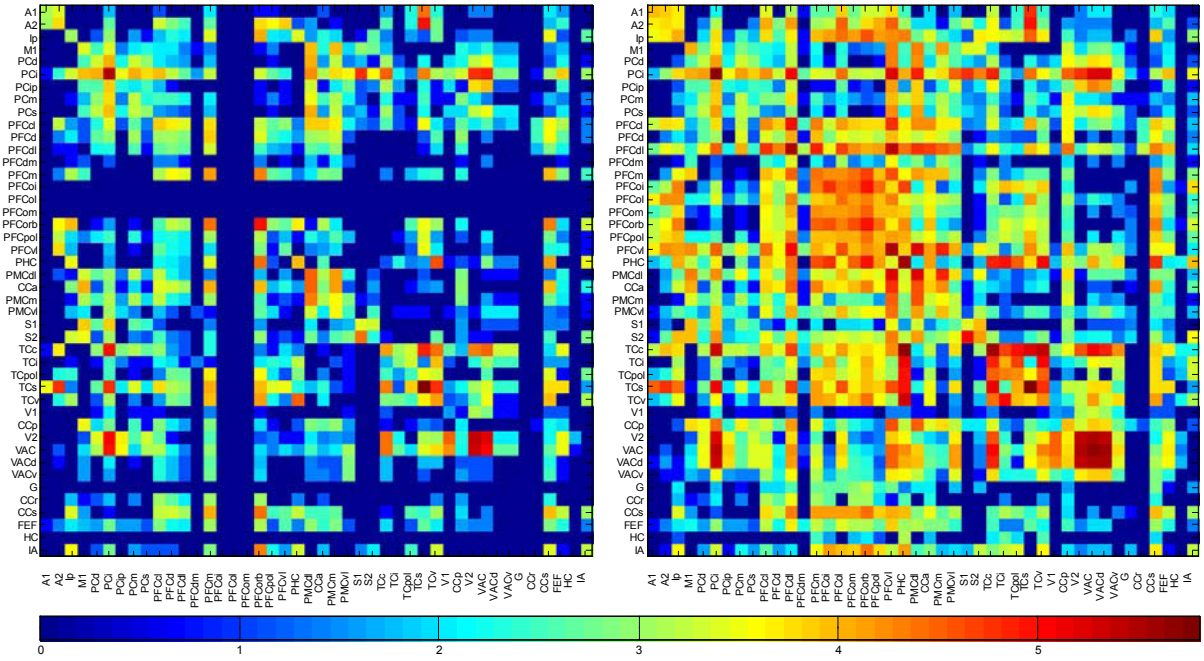


Figure 3.5.2. Impact of mapping statements generated by SORT on the number of labelled sites used for each of the entries in the RM connectivity matrix. In this figure, the log values of the number of labelled sites used for each entry are displayed for the case before (left) and after (right) entering SORT statements. Sources are displayed on Y axis, targets are on X axis, as typically done for adjacency matrices on brain connectivity. Note the total absence of information on particular regions (G, PFCoi, PFCol and PFCom) before entering SORT statements which introduced these regions.

An important characteristic of the impact of mapping statements on connectivity would be the number of labelled sites (sites being labelled following a tracer injection) used for each matrix entry. On average, each of the RM matrix entries utilized data on 11 labelled sites (5814 distinct ones in total) without SORT data, and 28 labelled sites (6532 distinct ones in total) with SORT data, following the output displayed in Fig. 3.5.2.

3.6 Discussion

The approach described here might best be applied to neuroimaging data. At present, numerous labs around the world generate and subsequently analyze voxel data with high spatial resolution. Anatomical substrates as structural references for such voxel data were not in use until recent times, when several attempts to create voxel-based atlases were made

(see for reference Kruggel et al., 2003; Schleicher et al., 1998; Mazziotta et al., 1995 and Tzourio-Mazoyer et al., 2002). One of the SORT follow-up approaches, the one described in (Bohland et al., 2009), is developed for neuroimaging data registered to particular anatomical templates and deals exclusively with quantitative characteristics derived from spatial data, without taking textual data into account. All these approaches allow a researcher to reference the spatial location of a functional pattern to valid landmarks (typically sulci or gyri). The main anatomical reference for functional data is still the ubiquitous Brodmann parcellation scheme, along with its modern successor, atlas of Talairach and Tournoux (1988), assigned to the space of individual's brain.

One important issue is the fact that different functional attributes can be associated with different parcellation schemes, whose criteria are more appropriate for these particular characteristics. For example, studies on visual modality could be meaningfully attributed to the connectivity map by (Felleman and Van Essen, 1991), whereas the nature of prefrontal activations can be investigated in regard to anatomical wiring obtained from the map of Walker (Walker, 1940). Applied to these issues, the SORT approach is very generic, because it can be trained on different datasets. Such a feature is also due to the fact that several techniques were applied for partitioning of the $R_A R_B$ space; each of these techniques has different application possibilities. In case of application of this approach to fMRI data or another modality with a similar spatial precision, it is presumably better to use the partitioning based on the bivariate Gaussian distribution, since the spatial resolution of this imaging modality is relatively high, and the data assigned to neighbouring voxels should fit the model to a large extent. EEG data, which is spatially much more crude, can be fit to the model representing the partitioning of data space using Voronoi diagrams; it is purely data-driven and low spatial resolution might speak to avoidance of any assumptions on distribution of data across space. The theory-driven approach can be applied to any of these data with roughly equal degree of success.

Data produced by SORT are evenly distributed across the nomenclature of extracted relationship statements (hence the choice of the flat prior, as stated in 3.4.1); if the entire cortex is parcellated, statements reveal no gaps. Neither can be there any contradictory statements, which are commonly the case for logical databases, where most entries are highly user-dependent and do not describe all cortical regions to the same precision and extent.

For instance, the data on connectivity of the macaque prefrontal cortex are covering extensively the entire region; by contrast, the temporo-parieto-occipital region has much less information on connectivity. Also, many parcellation schemes define particular areas of interest while ignoring the rest, which also causes gaps in such datasets (Fig. 3.6.1).

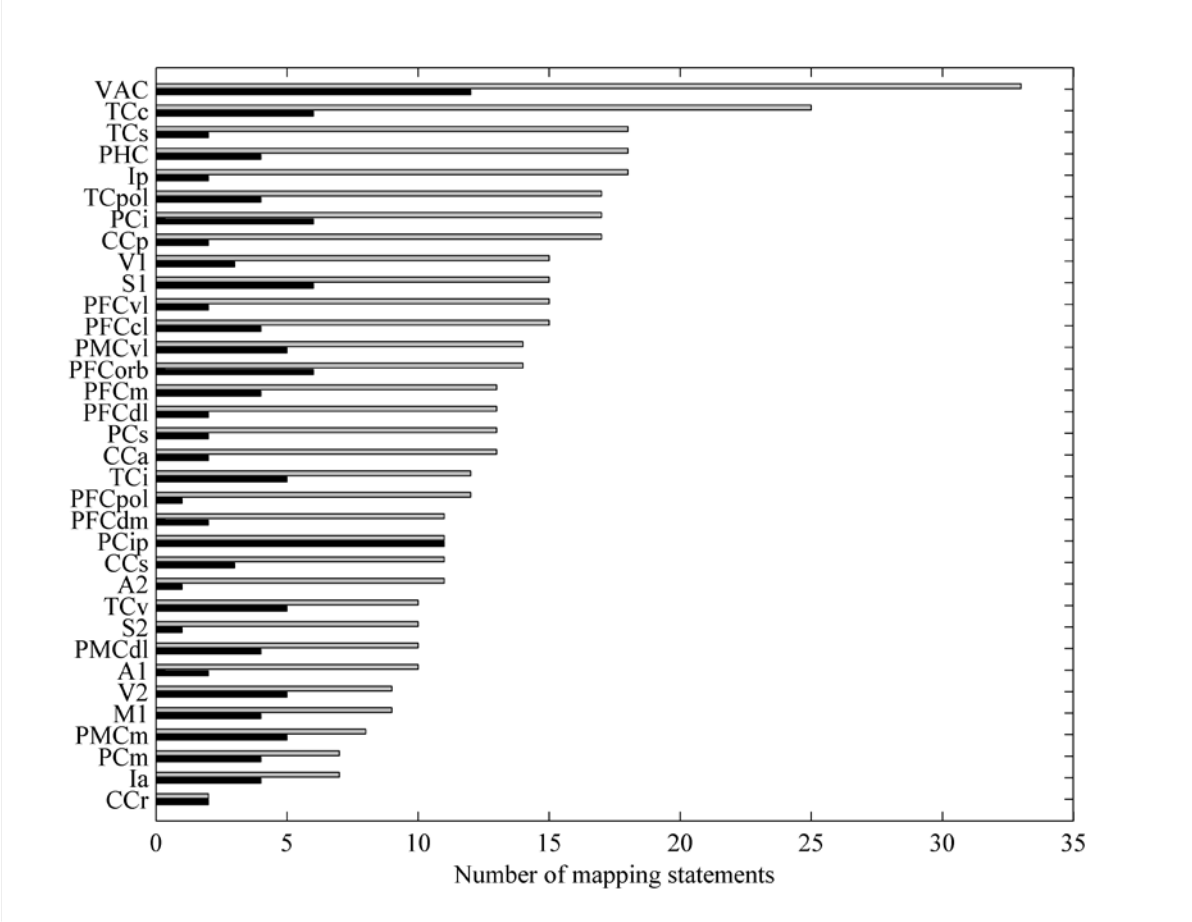


Figure 3.6.1. The number of mapping statements for Regional Map areas derived using the SORT procedure (grey), and those obtained from CoCoMac (black), taking into account the same set of target maps.

For the literature database, it can be very difficult to obtain an estimate of which areas have a lacking number of relationship statements and where in the brain they should be localized. SORT is observer-independent, which solves this problem to a large extent; the only decisive factor we need to rely on is the accuracy of spatial localization of a particular brain area – a criterion that highly affects its relationships with other brain areas. Obviously, when different authors represent their maps on the same spatial template, these maps may appear very different in spatial positions, which might cause a problem to the procedure; therefore, a training stage of SORT should be done deliberately. On the other hand, such caveats are also advantageous: differences between maps could indicate differences between individuals, as well as the difference in used methods; in this case, only spatial methods and relationship statements derived with their aid would serve as a meaningful support for warping, registration to surface or volume, or any other related procedure in which landmark-based approach failed and more appropriate registration criteria are needed.

To summarize, the SORT approach has the following main purpose: to evaluate spatial (coordinate-based) against logical (coordinate-independent) arguments. This will allow the enhancement of the wealth of statements in textual databases and their utility in neuroimaging approaches. Demonstrated results have pointed out that SORT may serve as a powerful tool for detection of errors arising from either incorrect space localization or inappropriate literature statements, or both – the latter issue is especially the case when results vary considerably between various experimental studies. Because of "topographic" (to a large extent) nature of the RM parcellation scheme, most of the discrepancies appeared either from failures in deformation algorithms, crude spatial representation, or insufficient or contradictory data obtained from studies describing coordinate-free delineations.

Chapter 4

CoCoMac-Paxinos3D tool: a user interface linking a stereotaxic atlas with a connectivity database

Atlases of various species' brains are extensively exploited in experimental neuroscience as useful substrates for identifying locations of particular targets in the brain. One of the difficulties in this endeavour is the provision of interoperability between such delineations and hierarchically strict ontologies (vocabularies with rules) obtained from animal databases. The aim of the project presented in this chapter is the development of an interface between a stereotaxic atlas and a connectivity database. More specifically, such an interface implies a link between volumetric structures and appropriate ontological entities. This concept was implemented in Java tools called *CoCoMac-Paxinos3D*, or briefly *CP3D*, where the name is signifying the link between CoCoMac and the atlas by (Paxinos et al., 2000). Having such a link, we also extended the tool for retrieval, visualization and analysis of connectivity data. Atlas structures rendered by the tool were drawn manually; the principles of linking these structures to appropriate ontological entities are described in detail ensuring the possibility to extend this tool for dealing with different species and in general with different concepts. Some teaching applications, including the experiment performed this year are introduced.

4.1 Why neuroscientists need such a link

Textual databases on invasive animal studies are usually distinguished by their relational complexity, non-intuitiveness when data are represented as text, and typically confusing nomenclature. Indeed, the variability of entities handled within such datasets along with intricate relationships between them becomes obvious when one compares different investigations with one another (Lewis and Van Essen, 2000; Gamberini et al., 2009). Moreover, such intricacy in data representation may frequently appear unfamiliar and confusing not only to non-experts in the field. Researcher involved in such particular field might spend a long time trying to find out what all this explosion of acronyms, relationship statements and other database entries might mean. Providing an intuitive graphical user interface (GUI) can serve as a meaningful solution for this problem.

Another difficulty related to the issue, partially addressed in the previous chapter, is the mapping between geometric space and animal databases, usually represented in a textual form. In this chapter, another aspect of such mapping is addressed, considering that animal databases are hierarchically organized, as contrary to brain atlases, whose hierarchies are

typically plain. As opposed to the previous chapter, instead of providing a backward compatibility allowing the linkage of voxel data to textual databases, this approach is in a way the opposite: taking a number of tracing studies into account, how can one represent their data on the finer parcellation scheme, registered in the three-dimensional (stereotaxic) space. Given that this is implemented, one can display connectivity data from any related study on the single map represented in the atlas. Further analysis on these connectivity data would be of an additional benefit.

In general, 3D-rendering is more intuitive than 2D representation in a sense that it gives an observer a realistic shape which can be manipulated, viewed from different angles, compared between selected structure and the rest of the brain mass. Stereotaxic frame can be also displayed alongside to ensure the location of structures can be evaluated against the neighbouring structures or other areas of interest (e.g., the ones obtained by the connectivity query). At the same time, these areas of interest, in case of having a hierarchical organization between each other, can be displayed in a viewer appropriately, so that the user will see immediately which area is a subregion of a given structure – the feature which is not possible to maintain in two-dimensional atlases (see subsection 4.2.1).

Further, atlas data originally lacks any information concerning interactions between regions. On the other hand, textual databases typically do not contain accurate spatial coordinates. By means of coordinate-independent mapping (Kötter and Wanke, 2005), one can retrieve all the necessary statements of mapping to other parcellation schemes for attributing resulting connectivity data to atlas volumes, organized hierarchically.

4.2 Arising problems

Although the representation of 2D-structures in 3D-space and provision of additional features is an efficient solution for the abovementioned problems, it takes several important and non-trivial steps during this development for these problems to be eliminated. In the following subchapters, these issues are described together with solutions for them.

4.2.1 Single-level representation of atlas structures versus strictly hierarchical organization of regions in the database

The Paxinos atlas used within this project (Paxinos et al., 2000) has a flat list of terms where each single area is a particular entity in Euclidean space. All the terms assigned to these ent-

ities were allocated in accordance with the limitations and requirements of the microscopically distinct features, attempting to make these terms compatible with other species' terminologies. By contrast, the CoCoMac database is devised in such a way that all its entities are organized in a strictly hierarchical fashion, reflecting the relationships between areas taken from various literature reports. An example (Fig. 4.2.1) can be a better illustration for this problem: area 10, located in the frontal pole, can be divided cytoarchitectonically into three parts: dorsal, medial, and ventral. The naming of this partitioning is defined according to the corresponding cortical locations, each distinguished microstructurally from one another; the area 10 is partitioned completely, in such a way as shown in Fig. 4.2.1. Given that the partitioning in the atlas is done section-wise, there is a high possibility that in a given section one cannot discern two or more subdivisions of the same area due to an absence of visible criteria. This is exactly what happened to the representation of the area 10 in the rostralmost section: it was merely impossible to specify 10D, 10M and 10V subdivisions in this particular section – and therefore it appears as area 10 in the atlas – the name which has to be assigned to the entire structure, including all its subdivisions, not just this tiny subvolume.

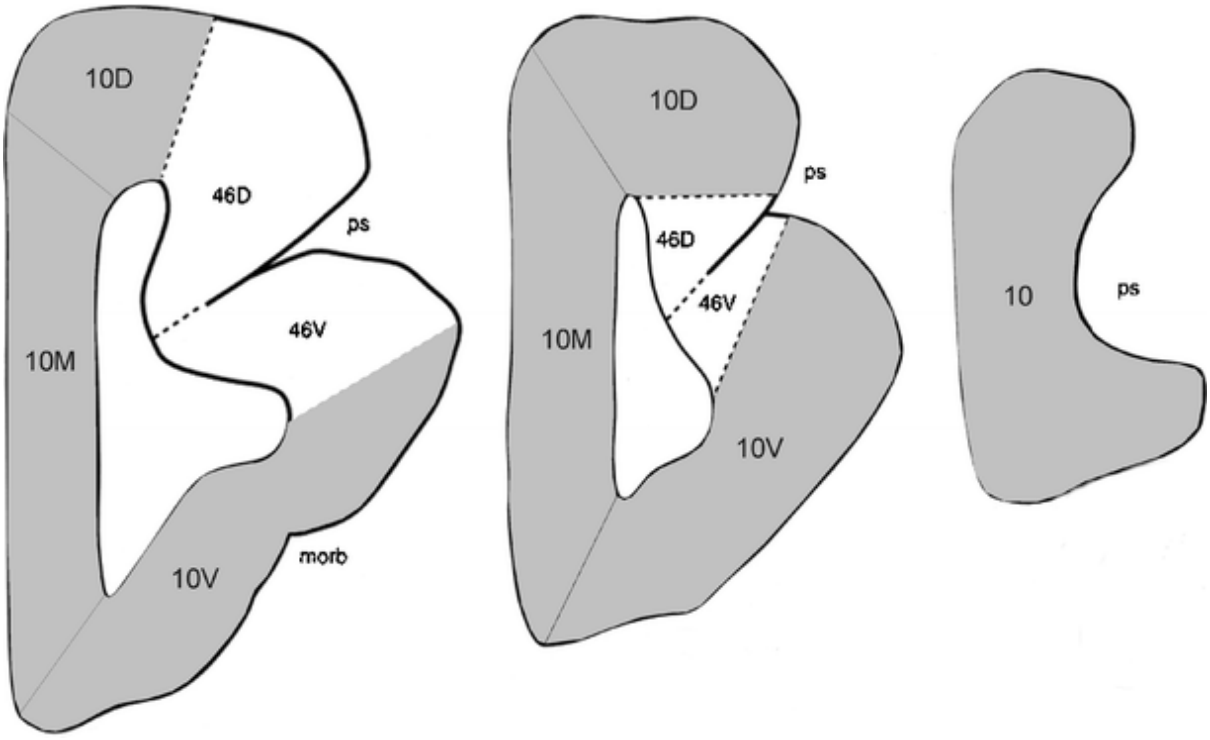


Figure 4.2.1. Discrepancy between geometrical atlas representation and the hierarchical ontological tree. Three subsequent sections drawn after the original paper atlas (plates 3, 2 and 1—from left to right) contain some regions, area 10 is among them (shaded in gray). On the left and middle sections, area 10 is subdivided into dorsal (10D), medial (10M) and ventral (10V) parts; the section on the right located most rostrally lacks this subdivision. These kind of unassigned substructures appear commonly at the beginning or end of a stack where the brain structure is included.

By contrast to this representation, CoCoMac defines areas rigorously in respect to their relationship with each other. This relationship may span multiple levels: some area may have subdivisions, some of which have subdivisions as well, etc. It is a non-trivial problem to handle in spatial representation. In anatomical labelling of MR images, it is commonly ignored, albeit several attempts for annotating such images with ontologically related entities were recently proposed (Golbreich et al., 2005). The rationale for such representation is the possibility to view brain structures in different scales, and to retrieve relevant database data related to each of these scales.

4.2.2 Absence of connectivity data in the stereotaxic atlas and hence a need for utility of database resources on mapping

Speaking about connectivity data retrieval using a stereotaxic atlas which originally contains no such information, one has to consider the issue of mapping between different parcellation schemes. Connectivity statements in the literature are describing connectivity information taken from one particular study, occasionally utilizing statements from previous studies in their original investigations. By contrast, relation of one parcellation scheme or connectivity map to appropriate statements from other literature resources can be obtained only subsequently. This is what ORT and SORT are designed for (Stephan et al., 2001, see 3.2.2; Bezgin et al., 2008, see Chapter 3). With such approaches one can obtain connectivity data from any map or group of them and relate these data to a particular parcellation scheme, given that these maps contain identical or anyhow else related regions. This feature can be useful in case if some detailed investigations on a particular subsystem are superimposed on higher scale model (for example, data on auditory or visual system applied to a coarser map of the entire cortex). In such a case, the detailed knowledge about a particular subdomain is transferred to a larger region, whose explorations spanned less detailed features in an original investigation. Indeed, combination of knowledge from several domains tells more information than each of the investigations separately.

Concerning the atlas data, the situation is even more prominent: being a highly detailed parcellation scheme registered in a stereotaxic space, it was not intended to be a connectivity map, and therefore no such data are available there at all. Luckily, the authors of this atlas were performing other mapping and also tracing experiments years before creating this atlas and used, in most of the cases, the same terminology, hence that knowledge base these previous studies provided could be partially imposed to the atlas. The rest of the gaps

in connectivity data were filled with statements generated by ORT and SORT. Section 4.4 provides more detail on this issue.

4.3 Tool development

For initiating the development of the CP3D tool, we had three core resources: (1) CoCoMac (Kötter, 2004) with interfaces for the output in XML format and automated queries; (2) the macaque brain atlas by (Paxinos et al. 2000) with a full cortical parcellation, whose entities were coexistent with CoCoMac; (3) a visualization Java-toolset previously developed for the rat brain (Howell et al. 2005). These components were subsequently assembled by the ultimate CP3D tool (see Fig. 4.3.1).

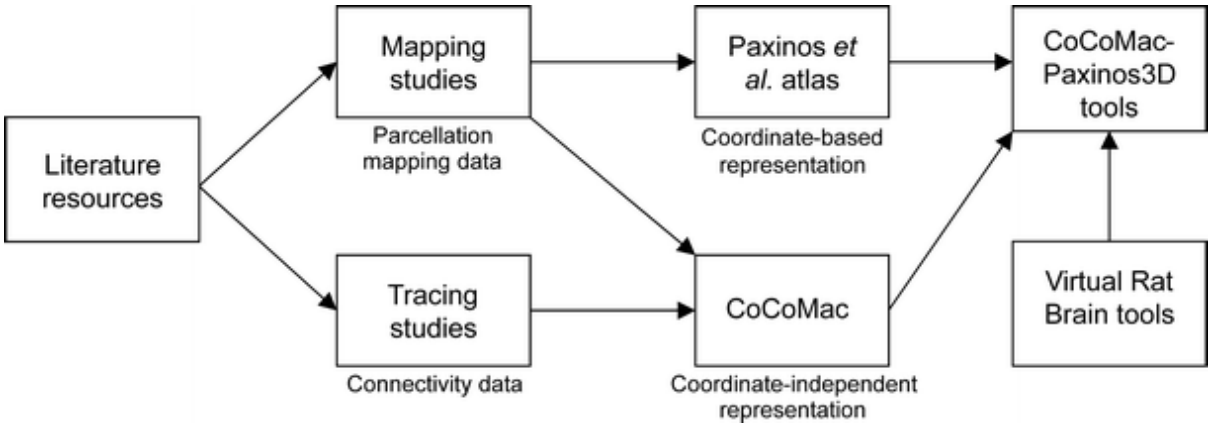


Figure 4.3.1. Components comprising the CP3D tools. All the data are obtained from literature reports; CoCoMac contains both connectivity (physical connections between regions) and mapping (logical relationships between areas from different maps). Paxinos et al. atlas serves as a geometrical and partially terminological basis for CP3D; the software code is based upon formerly developed *Virtual Rat Brain* tools.

For the description of the CoCoMac database, see subsection 3.2.2. Here we provide a portrayal of the *Virtual Rat Brain* tools, putting the accent on those features most useful for the development stages of CP3D.

The *Virtual Rat Brain* (VRB) application programming interface (API) was developed by several neuroinformatics researchers at the University of Edinburgh in the Institute for Adaptive and Neural Computation (http://www.inf.ed.ac.uk/research/ianc/) in 2004, in collaboration with the lab of Laszlo Zaborszky, engaged primarily in basal forebrain research (http://zlab.rutgers.edu/). The primary aim of VRB was visualization and analysis of 3D anatomical data from the brain of the rat (Howell et al. 2005). It featured also the analysis tool,

which allowed researchers to identify topological association patterns (or spatial overlap) of particular cell populations and morphometry of various anatomical units. Apart from that, the tool supported several XML-based formats, including MorphML (Crook et al., 2005) and NeuroML (Goddard et al., 2001), making the portability of neuronal morphology data among researchers possible and flexible. Several additional clients were included in the API, namely the one providing connection to the database, and some modules for browsing and visualization of data. In general, the software was designed in a flexible way, with provision of modularity, which made it possible to adapt this tool for rendering way more complex primate brain structures.

Indeed, the heart of CP3D, together with CoCoMac, is the Paxinos et al. atlas, which has to be described here as well. The atlas was developed with involvement of multiple experts in cytoarchitecture, who performed an advanced partitioning and compiled the list of terms (see *Terminology* paragraph below).

(1) *Contents of the atlas*. As it was mentioned in the section 2.3, this atlas is the first one up to now featuring a fine-grained macaque cortical parcellation registered in Euclidean space. The process of surgical and histological techniques involved the dissection of the left hemisphere of a single brain into more than 1,500 coronal sections. Subsequent delineation of anatomical entities was performed on every tenth slice, so that the atlas contained 151 sections.

(2) *Coordinate system*. Spatial coordinates in relation to bregma (top of the skull) and the interaural line are provided for each section, making up the resulting stereotaxic grid. The anterior commissure is to be found 18.3 mm anterior and 14.8 mm dorsal to the interaural axis, and the posterior commissure is 3.9 mm anterior and 12.2 mm dorsal to the interaural axis. The section-to-section distance is commonly 0.45 mm, slightly varying in several sections and increasing at the frontal and occipital poles. Some misalignment between a few pairs of adjacent sections is due to distortions and problems in locating the fiducial horizontal mark in some of the sections which contain ventricles or because of systematic drift if there had been no track visible for several millimeters. Moreover, indeed, some other problems occurred related to distortions of sections, related to the fact that slices were obtained from fresh tissue and without usage of the block-face imaging.

(3) *Terminology*. The complete taxonomy with all full names and acronyms, including descriptions of mapping techniques used in the study, is presented at the beginning of the atlas' first chapter. The atlas mainly uses generally accepted acronyms, consistent with terminology used in preceding atlases on a range of species (see e.g., Paxinos and Watson 1998),

which is important for reduction of inconsistencies in terminology. An example of such inconsistency is the basal ganglia structure called “accumbens nucleus”, which had been abbreviated in more than 20 various ways, as becomes evident from CoCoMac. The acronym used in the current atlas is Acb, consistent with the majority of previous atlases on different species. Several ambiguous cases were identified and handled in a way as represented in the table 4.3.1.

Area full name name as in the atlas	Commonly used acronym	Acronym as in the atlas	Acronym as in CoCoMac
parietal area PF	PF	PFCx	PF#1
parafascicular thalamic nucleus	Pf	PF	Pf#2
nucleus basalis Meynert B	many	B	BM#4
central amygdaloid nucleus, medial division	CEm	CeM	CEm#1
parieto-occipital area	PO	PO	PO#1
posterior thalamic nucleus	Po	Po	Po#5

Table 4.3.1. Several instances of ambiguity between atlas and database acronyms.In most of the studies parietal area PF is being referred to by acronym “PF”, in the atlas this abbreviation refers to a subcortical nucleus; acronym pairs such as CeM and CEm, PO and Po, would be ambiguous in case-insensitive queries. Translation between acronyms is handled by the querying system provided by the CP3D tools.

Altogether, these various features spoke to the usefulness of a provision of 3D-rendering interface for linking all these components, which motivated this development.

4.3.1 General interface description: Java3D features

CP3D is implemented using the publicly available, open source Java3D API from Sun Microsystems. It can be downloaded from http://java.sun.com/products/java-media/3D/. Variety of extensive application programming interfaces (APIs), portability and security were the most attractive features of this programming language, fitting perfectly to our purposes. Windows XP was used as a test platform; however, most of the features work also on UNIX systems.

All the iso- and allo-cortical structures, a few blood vessels together with several subcortical nuclei of thalamus, amygdala and basal ganglia were manually re-drawn after the original representations in the atlas. Resulting polygons were converted to Scalable Vector Graphics file format (SVG, http://www.w3.org/Graphics/SVG/). All the area names and acronyms were listed in a Microsoft Excel® file. An additional sheet of the same file contains the de-

scription of hierarchical relationship between these brain regions. Unique RGB specifications enable the mapping between geometric structures and their specifications in the Excel file. The tool's GUI (graphical user interface) consists of a viewer displaying three-dimensional renderings of atlas structures, a tree panel displaying these structures in a hierarchical way; important information about a selected entity appears on a panel below (Fig. 4.3.1.1).

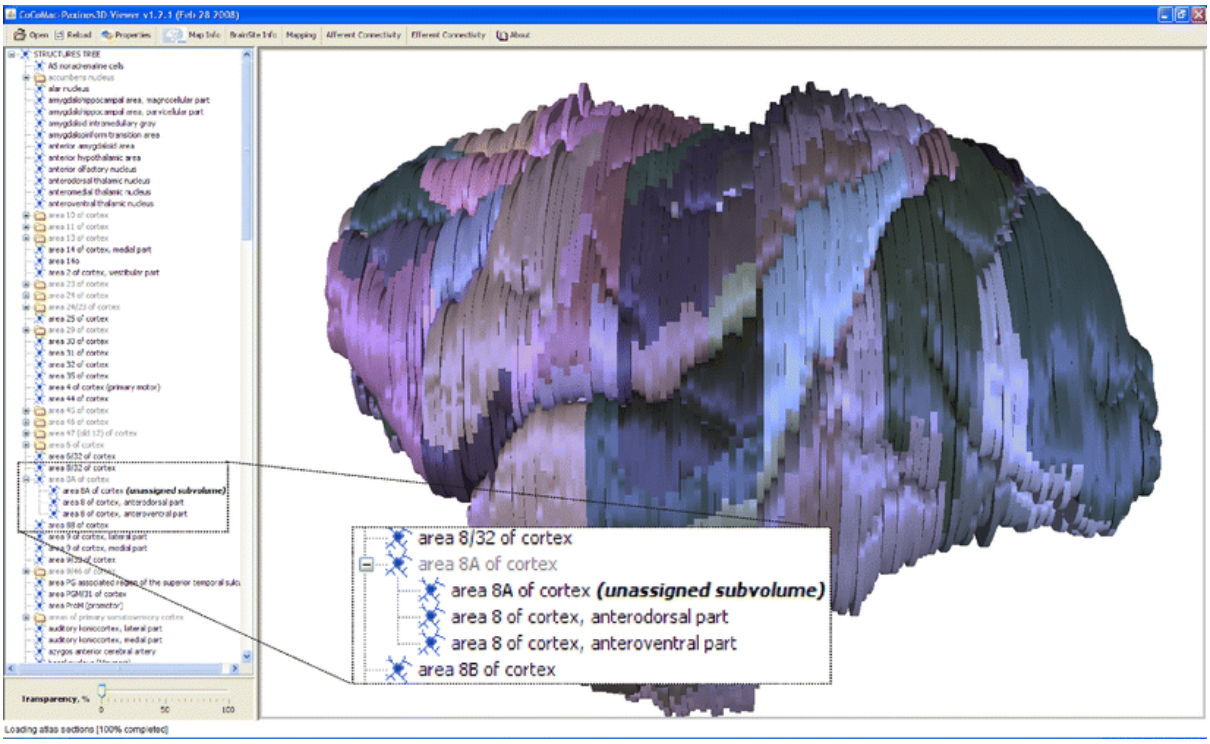


Figure 4.3.1.1. The general interface view of CP3D after loading the brain structures.

Loading all SVG files and Excel file allows the entire left hemisphere of the Rhesus monkey brain to be displayed. If only one or several brain sections are of interest for the user, they can be loaded separately from the rest of slices. Transparency can be adjusted using transparency slider located on the bottom left. Areas selected on the 3D window or at the tree always have 100% opacity, which makes them distinguishable from the rest of the brain (Fig. 4.3.1.2a,b). If the transparency slider is set to the 100% transparency, one can see the structure of interest without being distracted by other loaded regions (Fig. 4.3.1.2c,d).

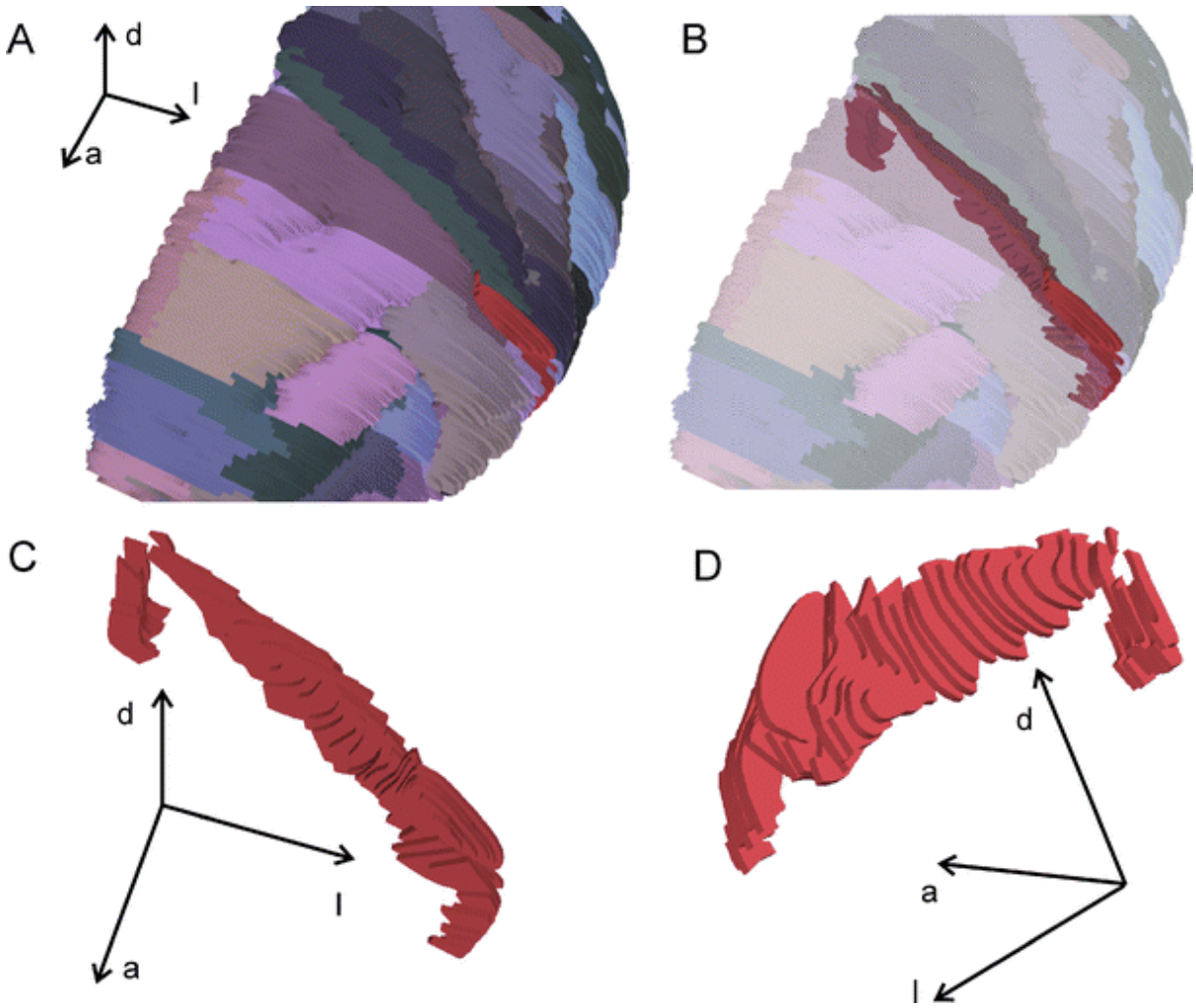


Figure 4.3.1.2. Transparency features of CP3D. A: when the brain mass is completely opaque, one can see only the outer surface of the selected structure, at this instance – area 3b of primary somatosensory cortex (red); B: semi-transparent mode allows one to achieve the "glass brain" effect; C,D: 100% transparency makes the main brain mass invisible; this is helpful for viewing a selected structure from different angles (a=anterior; d=dorsal; l=lateral).

Additionally, the user can manipulate a selected region separately from other structures. Rotating and moving operations can be used with the entire brain mass, and also with the structure of interest only. The usefulness of such a possibility is explained by the option of viewing the structure from different angles and exploring its location in relation to the areas located in the immediate neighbourhood (Figs. 4.3.1.2c,d).

In the *Properties* dialog, one can specify background colour, set the orientation of selected structure to the user (which is useful in 100% opacity mode for viewing superficial cortical regions) and display a stereotaxic grid in relation to both bregma and interaural line.

Some details on implementations of several features that could be useful for new related developments are provided in Appendix A.4.3.1.

4.3.2 Multi-level brain areas representation

As was mentioned in the subsection 4.2.1, brain atlas is conceptionally different from database in a sense that it treats areas as non-overlapping volumes, whereas databases represent structures with relation to each other, and Paxinos et al. atlas versus CoCoMac is not an exceptional case in this respect.

Multilevel arrangement of atlas areas permits one to achieve several goals. First, one can choose whether a selection of one particular subregion is profitable in a particular situation, or selection of all subdivisions in their entirety would be more profitable. Next, such multilevel representation delivers additional essential information: it helps to see whether a particular ontologically organized brain region is represented in the atlas as one entity or as a group of volumetric structures each possessing a different name. The hierarchical tree contains as its leaves all the full names of atlas regions, which are themselves non-overlapping volumes. By clicking such a leaf, one gets the corresponding atlas area highlighted in the three-dimensional panel (Figs. 4.3.2.1a–c). If the user clicks on higher level units of the tree, combining a number of brain regions, the ontological structure, which is not depicted in the atlas, is selected (Fig.4.3.2.1d). Notably, there are a number of occasions in which the branch holds the same name as the leaf. Although such an occurrence is somewhat counterintuitive, it is typical for brain atlases. The origin of this problem becomes understandable when considering the viewpoint described in the section 4.2.1 for the area 10 located in the frontal pole, which is more formally depicted in Fig. 4.3.2.1 for the visual area 4 (V4): due to limitations of the methods and material in the process of preparing the brain sections it has been not always possible to distinguish various subdivisions of the same area on a particular given slice.

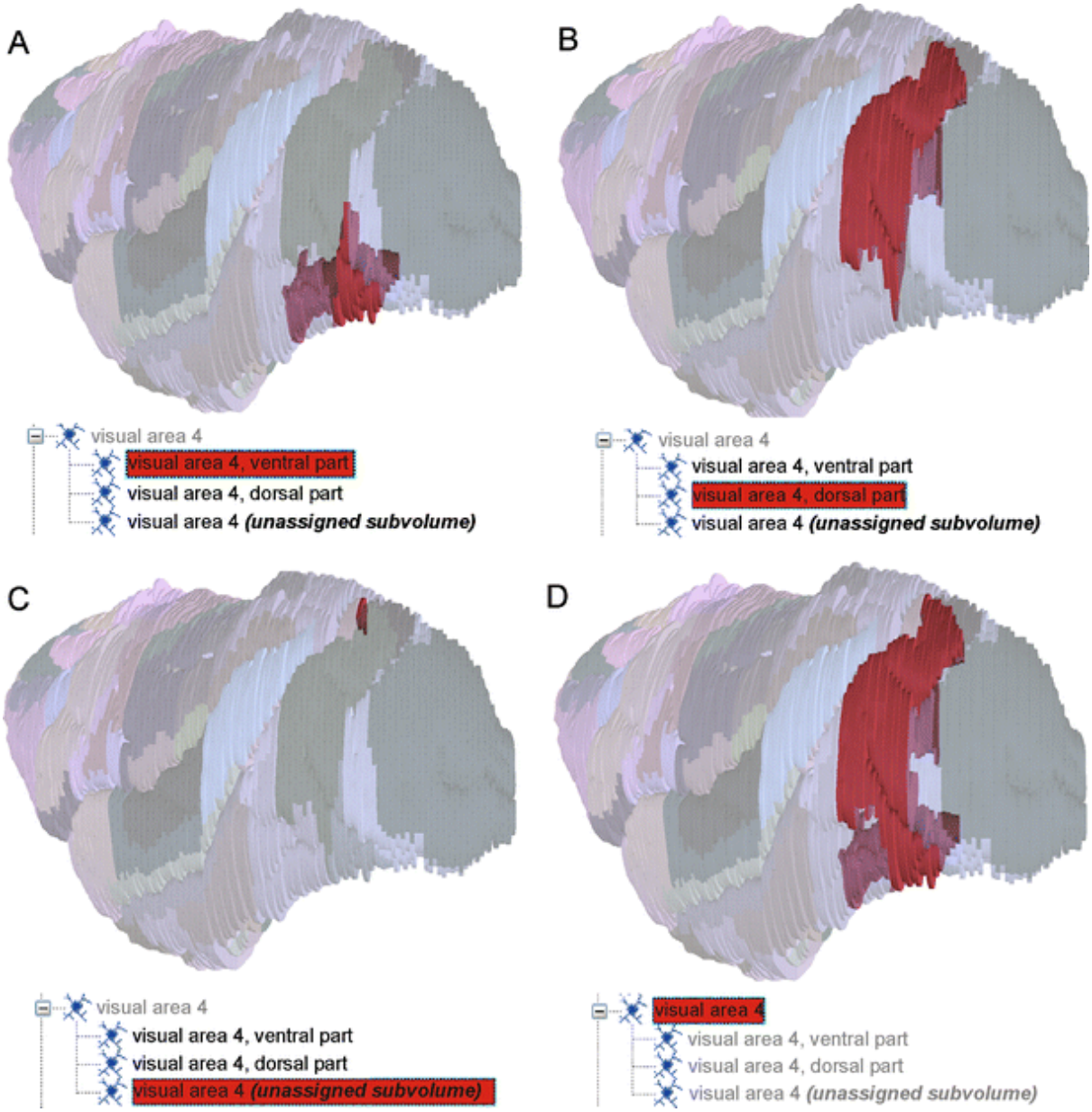


Figure 4.3.2.1. Managing contradictions between atlas volumetric structures and CoCoMac ontology. A: visual area 4, ventral part (V4V); B: visual area 4, dorsal part (V4D); C: visual area 4 as specified in the atlas (V4); D: visual area 4 according to the CoCoMac nomenclature (incorporating all the listed subregions).

The third prominent example of this issue is cortical area 29, which is logically subdivided into subregions a, b, c and d.

Despite this fact, there is an atlas entity referred to as area 29a-c and another one referred to as area 29d; a separate volume is assigned with the label “area 29”. Areas 29a-c and 29d cannot be considered subvolumes of the volume area 29 conceptually due to the same reasons as in two previous examples. Altogether, anyway, they comprise the total volume of

area 29, which is also the ontological unit as represented in the database. Any finer rational subdivision is not compatible with parcellation supplied by this atlas; this is explained by the fact that area 29d, for instance, should possess the volumetric structure 29d as represented in the atlas, plus an unknown part of the unassigned subvolume called 29. In the described framework, this is a truly minor problem since unassigned subvolumes are represented by typically small regions usually allocated within one slice, as compared to areal subdivisions occupying substantially greater amount of space (Fig. 4.3.2.1).

4.4 Connectivity database interface

For proper interoperability between the atlas and CoCoMac enabling the display of connectivity data on atlas structures (including those transformed to ontologically defined entities), there has been a need to establish a link to the database from the CP3D software. This link was maintained by the JDBC-ODBC bridge, described in the following subsection. Moreover, the textual data display was enabled as well by means of CoCoMac URL specifications allowing the user to see the data in the browser, in addition to the graphical display described in the subsection 4.4.2. Further connectivity analysis was a desirable feature, and therefore implemented as well; it is described in the subsection 4.4.3.

4.4.1 JDBC-ODBC bridge and the utility of SQL statements

Conveniently for our development purposes, the Sun Microsystems has developed a suitable API for accessing any database from Java. Following this framework, the Java Database Connectivity (JDBC) driver establishes a connection to a target database via Open Database Connectivity (ODBC) driver, regardless of what this target database is. The CP3D software is client-side, and therefore this bridge will work without a need for installation of the ODBC driver, installed on Windows machines by default; this allows the user to be secure about this functionality without caring about any additional installation necessities.

Having the JDBC-ODBC bridge established, one can perform queries to the target database. These queries are mediated by statements written in the Structural Query Language (SQL), proposed in early seventies of the last century (Chamberlin and Boyce, 1974). It constitutes a declarative approach for making queries to relational databases and remains to a large extent compatible between platforms and applications.

For querying CoCoMac in particular, one has to consider the following issues. First, one has to distinguish relationship statements obtained from original studies from those generated by ORT or SORT. These classes of relationship statements possess different PDC-values (precision of description code), and the output, especially in case of contradictory or redundant statement, should be handled appropriately in this regard. Second, mapping statements are used for obtaining connectivity data from tracing studies related to the Paxinos et al. atlas. To ensure more complete output, the connectivity query can be modified in a way that incorporates not only areas of a direct interest used for a query, but also those having particular relationship to them. To outline an example, let's first introduce the general rule for URL specifications in CoCoMac (http://cocomac.org/WWW/cocomac_urlsearch.html):

```
http://cocomac.org/URLSearch.asp?user=User&password=Password&Search=
        SearchCategory&Parameters &SearchString=Search string
```

The simplest search string for performing a connectivity query can be as follows:

```
SearchString=(BrainSiteAcronym)[TARGETSITE]
```

The corresponding SQL statement will look like this:

```
SELECT ID, ID_SourceSite, SourceSite_A, PDC_Site_A, Hemisphere_A    ,
Density, PDC_Density, Course, ID_TargetSite, TargetSite_B           ,
PDC_Site_B, Hemisphere_B FROM ORT_IntegratedPrimaryProjections WHERE

      (ID_TargetSite IN  (SELECT  ID FROM  BrainMa ps_BrainSites WHE RE
      ID_BrainMaps_BrainSiteAcronyms="+acronym_id+")+" OR
      ID_TargetSite IN (SELECT ID_BrainSite_A FROM InterMapRelation s
      WHERE ID BrainSite B IN

            (SELECT ID FROM BrainMaps_BrainSites WHER                E
            ID_BrainMaps_BrainSiteAcronyms="+acronym_id+")

      )

)
```

Here, the important data for further connectivity (at this instance, data on afferent connectivity – i.e., the data on incoming connections) data obtaining are retrieved from the `IntermapRelations` table, containing the information about the mapping between brain regions from different studies. Now, we can modify the query in such a way that it would deliver a greater amount if relevant information:

```
SearchString=(BrainSiteAcronym EXT 'I,L')[TARGETSITE],
```

and the corresponding SQL-code:

```
SELECT ID, ID_SourceSite, SourceSite_A, PDC_Site_A, Hemisphere_A   ,
Density, PDC_Density, Course, ID_TargetSite, TargetSite_B          ,
PDC_Site_B, Hemisphere_B FROM ORT_IntegratedPrimaryProjections WHERE

     (ID_TargetSite IN (SELECT ID FROM BrainMaps_BrainSites WHE RE
     ID_BrainMaps_BrainSiteAcronyms="+acronym_id+")+" OR
     ID_TargetSite IN (SELECT ID_BrainSite_A FROM InterMapRelation s
     WHERE RelationCode IN ('I','L') AND ID_BrainSite_B IN

          (SELECT ID FROM BrainMaps_BrainSites WHER              E
          ID_BrainMaps_BrainSiteAcronyms="+acronym_id+")

     )

)
```

The modification made means the following. If we know that area A connects to area B, we can also say that all the areas larger than A connect to all the areas that are larger than B. Thus, for the introduced example it means that if we want to retrieve all the afferents targeting to area B, we have to find all the afferents that are targeting areas smaller than B – i.e., those area for which area B has a relationship “larger than”.

Such connectivity retrieval is acronym-based and, although ambiguous cases are handled (see table 4.3.1), another approach involving retrieval of mapping paths information would be of additional benefit. Mapping path in this context means the sequence of relationships between the number of different areas from various maps.

For this purpose, the SQL query is modified in such a way:

```
SELECT ID, ID_SourceSite, SourceSite_A, PDC_Site_A, Hemisphere_A   ,
Density, PDC_Density, Course, ID_TargetSite, TargetSite_B          ,
PDC_Site_B, Hemisphere_B FROM ORT_IntegratedPrimaryProjections WHERE

     (ID_TargetSite IN (SELECT ID FROM BrainMa ps_BrainSites WHE RE
     ID_BrainMaps=map_id) AND

          (ID_TargetSite IN (SELECT ID FROM BrainMaps_BrainSite   s
          WHERE ID_BrainMaps_BrainSiteAcronyms=     acr_id) O     R
          ID_TargetSite IN

               (SELECT ID_BrainSite_B FROM InterMapRelations WHERE
               RelationCode IN ('I','L') AND ID_BrainSite_A IN (SE-
               LECT ID FROM BrainMaps_BrainSites WHERE
               ID_BrainMaps_BrainSiteAcronyms=acr_id)

          )

     )
```

```
) OR (enumeration of all possible map_id/acr_id combinations)
```

This enumeration of all possible maps and acronyms related to a given area is obtained by the following Java code with SQL strings embedded:

```
private static Set<String> getRelatedAreasSet(String acronym) {

	Set<String> acrs_set = new HashSet<String>();
	String id = getIdsFromBrainSite("PHT00-" + acronym);
	int acronym_id = -1;
	if (id==null || id.length()==0) return acrs_set;
	try {
		acronym_id = Integer.valueOf(id.substring(id.indexOf("_")
		+ 1));//getAreaID(acronym);
	} catch (NumberFormatException se) {
		se.printStackTrace();
		return acrs_set;
	}
	//performCoCoMacQuery(String query, String column_label)
	List<String> areas_a = performCoCoMacQuery("SELECT BrainSite_B
	FROM ORT_ResultingRelations WHERE " +
	"(ID_BrainSite_A IN (SELECT ID FROM BrainMaps_BrainSites WHERE
	ID_BrainMaps=239) AND ID_BrainSite_A " +
	"IN (SELECT ID FROM BrainMaps_BrainSites WHERE
	ID_BrainMaps_BrainSiteAcronyms="+acronym_id+") AND (Category=1
	OR Category=3))", "BrainSite_B");
	acrs_set.addAll(areas_a);
	List<String> areas_b = performCoCoMacQuery("SELECT BrainSite_A
	FROM ORT_ResultingRelations WHERE (ID_BrainSite_B IN (SELECT
	ID FROM BrainMaps_BrainSites WHERE ID_BrainMaps=239) AND
	ID_BrainSite_B IN (SELECT ID FROM BrainMaps_BrainSites WHERE
	ID_BrainMaps_BrainSiteAcronyms="+acronym_id+") AND (Category=1
	OR Category=2))", "BrainSite_A");
	acrs_set.addAll(areas_b);

	return acrs_set;
}
```

In this code, “Path Category = 1” refers to mapping path where all the areas have an identity relationship to each other, “Path Category = 2” means that the areas can be either identical to or smaller than each other; finally, “Path category = 3” stands for paths including either “identity” or “larger” relationship.

Overall, the utility of SQL statements is a convenient tool since this language has a fluent interoperability with Java, and is increasingly consistent across platforms and various relational databases.

4.4.2 Connectivity data visualization

Retrieved connectivity data can be visualized using CP3D. Since no information on the fiber trajectory is available either in the atlas or in the database, the tool visualizes projections using straight arrows of thickness and colour patterns associated with connection densities. For illustrating this approach, let us consider the following example. Parietal area PG, which is in fact the caudal part of the inferior parietal lobule, has an intricate pattern of connectivity with other cortical areas, as has been confirmed in a number of studies (Rizzolatti et al. 1998; Ding et al. 2000; Clower et al. 2001; Zhong and Rockland 2003; Andersen et al. 2004). Having the CP3D tool with its connectivity visualization features at hand, the process of obtaining further relevant and detailed information on this issue becomes possible – let us consider the following example.

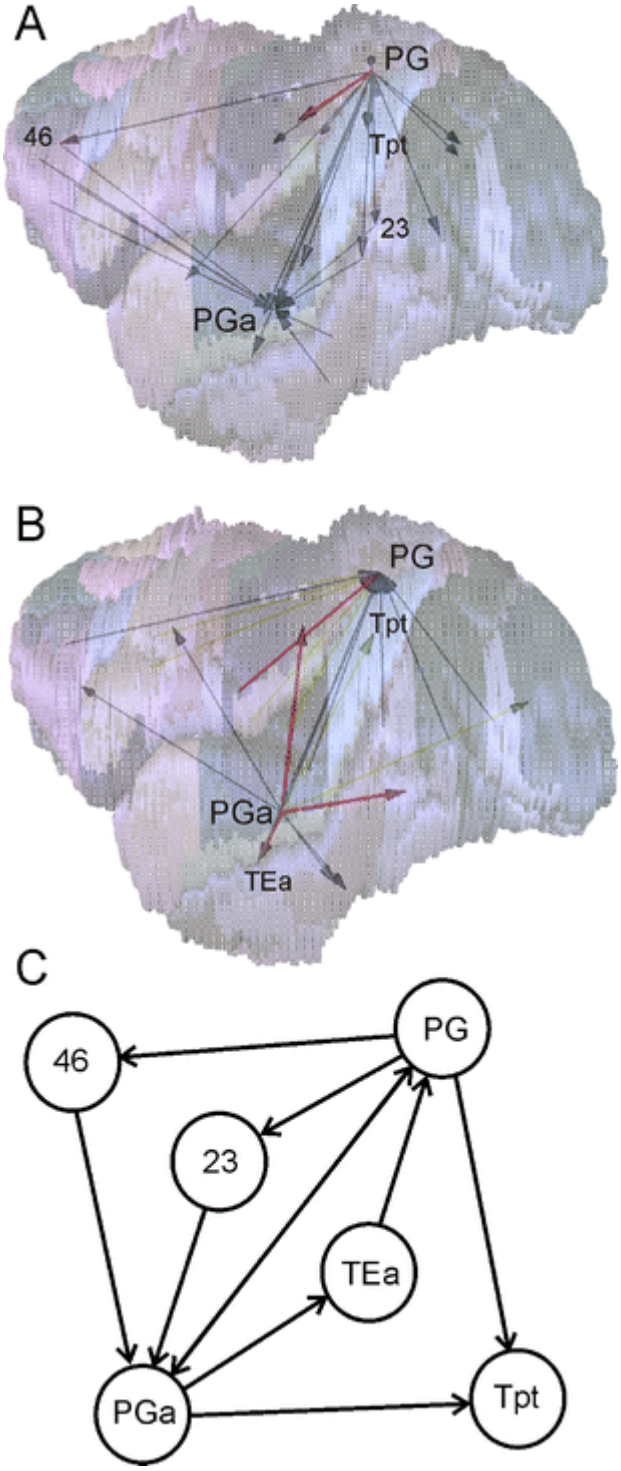


Figure 4.4.2.1. Visualization of connectivity data using CP3D.

For instance, we start with obtaining the connectivity information on all the efferents (outgoing projections) from area PG. It can be seen that the largest bundle of connections stretches to a region located in the close proximity of the superior temporal sulcus (Fig. 4.4.2.1).

In the set of all efferents resulting from the query, there is a region called “PG associated area of the superior temporal sulcus (PGa)”, which, according to the previous studies investigating the inferior parietal lobule (Golbreich et al. 2005) and associated areas (Seltzer and Pandya 1978), is reciprocally (mutually) connected with area

PG. In addition, these two regions cooperate in some indirect ways, namely they not only establish connections directly between each other, but use other regions in frontal, temporal and occipital lobes as mediators. For obtaining the full set of information about other regions projecting to PGa, we perform an additional query (afferents to PGa). Eventually, a number of afferents are received from regions in the parietal lobe, but also from regions in other parts of the brain, particularly the frontal lobe (Fig. 4.4.2.1a). Among them is prefrontal area 46. Notably, this region receives also one of the projections from PG, the following path exists: PG → 46→ PGa. The path is unidirectional, which is evidenced by two other queries, namely for PGa efferents and PG afferents (Fig. 4.4.2.1b). From these queries, it becomes obvious that the path PGa→ 46 → PG is non-existent. Instead, there is a prominent projection from PGa to the area TEa located in the temporal lobe; area TEa serves as a transit point, projecting further to the sought-for PG area. These projection paths (Fig. 4.4.2.1c) are in accordance with previous findings (Seltzer and Pandya 1978; Zhong and Rockland 2003).

4.4.3 Connectivity data analysis

Obviously, the connection path data obtained by means of performing database queries are relevant in itself and can provide an important information for a researcher investigating interactions between corresponding areas. However, in case of large connection circuits resulting from multiple queries, this data representation may appear confusing because of a large overlap between connection arrows. In such a case, the connectivity pattern must be unveiled in an automatic way – moreover, some selectable output might be also of a benefit; this is why the feature of connectivity path data analysis is implemented.

Following that CP3D is primarily a database query tool, and query objects are therefore of main interest for the user, the ideal choice for path analysis technique was Dijkstra's single source shortest path algorithm (Dijkstra, 1959), which considers all areas used in queries as sources, and all the variety of brain areas in the atlas as targets. The most attractive feature of this algorithm, in our framework, is its quadratic time complexity – so that even if the entire set of atlas structures is loaded, computation will be performed without a sensible delay. The algorithm halts as soon as the desired target is reached in the least costly way. The optimal path is then merely the one (or more than one) with the minimal sum of costs associated with all successive arcs in the path from node *a* to node *b* via one or more transitional arcs d_{ij}:

$$cp_{ab} = \min \sum_{i \in A, j \in B} d_{ij} \qquad (4.4.3.1)$$

where cp_{ab} is a connection path between node *a* and node *b*, *A* (*B*) is the set of all the starting (final) points on the path connected by the set of arcs with corresponding costs d_{ij}. These costs can be defined in a number of ways, each reflecting different properties of connectivity paths. The simplest approach is to weigh paths with the number of intermediate stations; in this case, the physical reachability is not necessarily optimal in a sense of Euclidean distance (although it is usually the case due to the "optimal" wiring topology of the primate brain). Another potentially helpful cost measure is the minimal connection strength found among other connections on the path (in the case if CoCoMac is used for this purpose: rank order values with 1<2<3; label 'X' means that the projection exists, but the density is unknown – it is treated as 2). The minimal connection strength is among the most intuitive density measures: the "bottleneck" property tells an observer about the capacity of the path. Aiming to relate lower connection strength value to a higher cost, we employed the following measure:

$$cd_{ab} = \min_{i \in A, j \in B}\left(\frac{d_{max} - d_{ij}}{d_{max}}\right) \qquad (4.4.3.2)$$

Finally, the third measure for path cost is the Euclidean distance. As was mentioned in the previous subsection, we had to consider that each pair of areas (their mass centres) is connected with a straight line, since no information on fibre trajectory is available.

As an example, let's consider the following task: to identify the most optimal paths between areas V1 (primary visual cortex) and 8A (area 8A of cortex, generally referred to as "frontal eye field"), according to the criteria mentioned above. The results are provided in the Table 4.4.3.1; schematic drawing of paths is depicted in Fig. 4.4.3.1. As one can see in Table 2, in this case there is only one pair of forward (V1→8A) and one pair of backward (8A→V1) paths are optimal according to all of the three criteria, i.e. V1→V2→8A and 8A→TPO→PG→V3A→V1, respectively.

<table>
<tr><th colspan="2">Feedforward (V1→8A)</th><th colspan="2">Feedback (8A→V1)</th></tr>
<tr><th>Path</th><th>Value</th><th>Path</th><th>Value</th></tr>
<tr><td colspan="4">Cost measure #1: the number of steps</td></tr>
<tr><td rowspan="2">[V1→V2→8A]</td><td rowspan="2">2</td><td>[8A→23→PG→V3A→V1]</td><td rowspan="2">4</td></tr>
<tr><td>[8A→TPO→PG→V3A→V1]</td></tr>
<tr><td colspan="4">Cost measure #2: the lowest projection density along entire path, see 4.4.3.2 (0<dc<1)</td></tr>
<tr><td>[V1→V2→8A]</td><td rowspan="4">0.3</td><td>[8A→TPO→PG→V3A→V1]</td><td rowspan="4">0.3</td></tr>
<tr><td>[V1→V2→V3A→8A]</td><td>[8A→TPO→PG→V3A→V2→V1]</td></tr>
<tr><td>[V1→V2→PG→V3A→8A]</td><td>[8A→23→PG→V3A→V1]</td></tr>
<tr><td>[V1→V2→23→PG→V3A→8A]</td><td>[8A→23→PG→V3A→V2→V1]</td></tr>
<tr><td colspan="4">Cost measure #3: Euclidean distance (mm)</td></tr>
<tr><td>[V1→V2→8A]</td><td>101.8</td><td>[8A→TPO→PG→V3A→V1]</td><td>155.1</td></tr>
</table>

Table 4.4.3.1. Path analysis using CP3D.

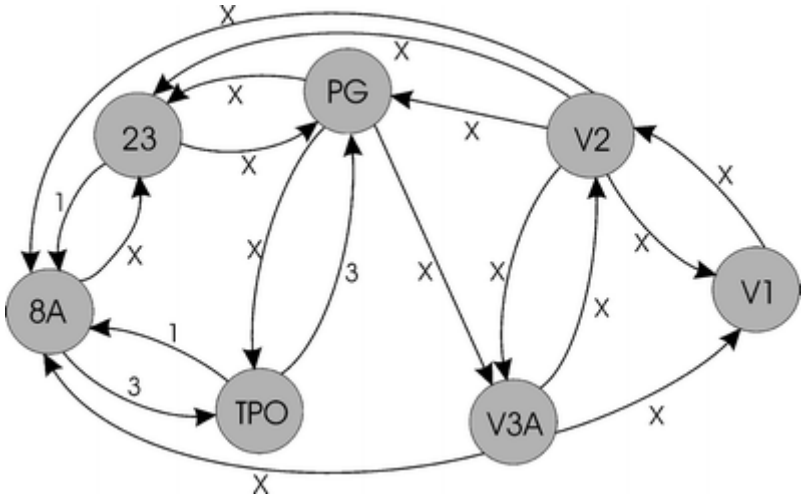


Figure 4.4.3.1. Schematic drawing of connections mediating interactions between V1 and 8A.

4.5 Code organization

The CP3D code is organized as depicted in Fig. 4.5.1. Gray rectangles in the middle are input file handlers (structure names Excel file and SVG graphics). Structure name file interacts with SVG graphics by means of RGB patterns encoding.

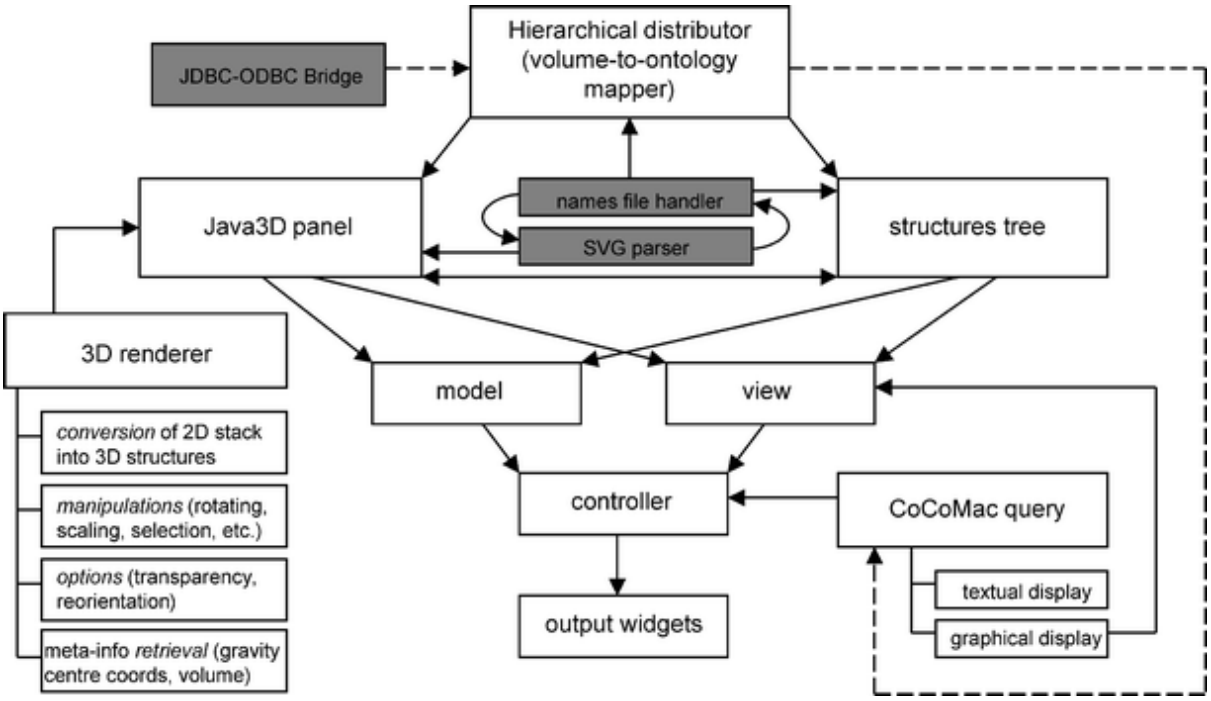


Figure 4.5.1. Schematic outline of the CP3D source code. Filled rectangles constitute the input/output system of the tool.

The module called *volume-to-ontology mapper* is dealing with compound brain structures containing multiple atlas areas. It contains the list of all compound structures; for each such structure, it specifies the list of the volumetric atlas regions (considered as "children"), based on the database nomenclature. Finally, it updates the multilevel tree with its selection properties on the basis of the list of structures and lists of substructures.

Both modules *Java3D panel* and the *structures tree* are synchronized. Java3D panel processes input from the geometrical renderer of three-dimensional structures, which represents a generic API for manipulating objects in Euclidean space. Such capabilities allow selection and handling of separate brain structures, as well as of the entire brain.

The *Controller* module joins together *view* and *model*; the first module handles a GUI (graphical user interface) of the tool; the second module triggers all the handlers that operate the software. The CoCoMac query module operates as described in the section 4.4.

The information bar notifies the user about selected structures, their volumes and mass centres, connectivity outputs, etc. For calculation of the structure volumes, the following form, suitable for non-convex polygons, was used:

$$A = \frac{1}{2}\sum_{i=1}^{n-1}\begin{vmatrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{vmatrix} + \begin{vmatrix} x_n & x_1 \\ y_n & y_1 \end{vmatrix} \quad (4.5.1),$$

where *n* is the number of points in the polygon. Subsequently, the volume is obtained by multiplication of (4.5.1) with the value designating a sum between semi-distance to previous and next section, respectively. Moreover, this formula was useful for another purpose - namely, for defining the directionality of the polygons. If a polygon is represented by the sequence of points accumulating successively counterclockwise, the result of (4.5.1) will be positive; if this orientation is clockwise, the result is negative. This feature is important for our issue due to the fact that the orientation of polygons in Java3D does matter for lighting issues; hence, in case of opposite orientation, two adjacent sections can be lighted differently (e.g., one highly shadowed and another very bright in a particular position), which does not look nice from the esthetic point of view. Using (4.5.1), we oriented all the polygons counterclockwise, fixing the lighting problem and ensuring volume values are always positive.

4.6 Teaching applications

Indeed, such an intuitive representation of macroscopic brain data and convenient ways to retrieve information from a complicated database speak to a possibility of utilization of the CP3D tool for teaching purposes. Such an experiment has been successfully applied within a group of medical master students without extensive knowledge in informatics/databasing. The following paragraphs provide a description of this experiment.

The computer practicum aimed at learning how to attribute localizations of functional activations obtained from magnetic resonance images, to anatomical areas and structural interactions between them. In the Fig. 4.5.1, the thresholded activity (dominant in two particular regions) of the brain of a monkey performing a particular visual task (similar to the one described in (Smith et al. 2007)) is shown in three different perspectives: posterior/anterior (P/A), lateral/medial (L/M) and dorsal/ventral (D/V). Coordinates are provided for all limits of three axes, and for centres of activations. The top row shows the first area of activation, and the bottom row shows the second area.

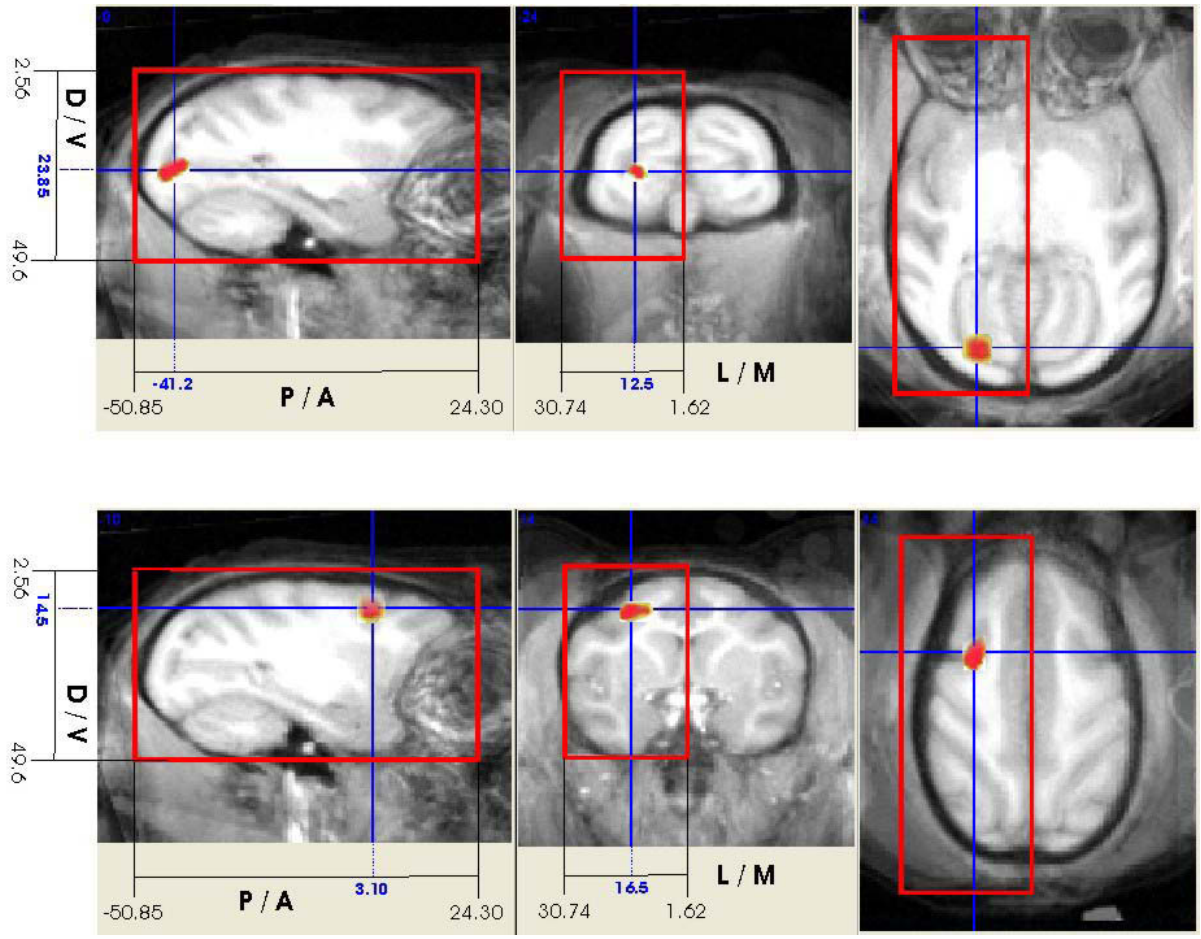


Figure 4.6.1. MR images of the macaque brain performing a particular visual task. Two areas are prominently activated, one shown at the top and another at the bottom. Reference to the three-dimensional coordinate system is provided.

For attributing these activations to anatomical regions and connections between them, the CP3D tool is utilized. Namely, it is used for identifying anatomical structures capturing the nearest neighborhood of functionally active regions revealed by fMRI. Further, it is investigated whether a synchronous activation in different regions has its anatomical substrate, i.e., whether the abovementioned activation areas are connected structurally. For this, CP3D is used for a retrieval of relevant data from the CoCoMac database on macaque anatomical connectivity (Kötter 2004).

After such introduction to the problem that needs to be solved, there are instructions for using CP3D given; namely, students have to identify atlas regions lying in the closest proximity to activations specified in Fig. 4.6.1 (given that the stereotaxic space is the same). Furthermore, they have to identify physical connections between these regions, including those mediated by other regions. These connectivity data from database, displayed with CP3D,

provides an insight concerning the means of functional interactions between sought-for regions shown in the Fig. 4.6.1.

The final evaluation questions can be as follows:

1. What areas are most likely the ones that got activated in the fMRI scan image?
2. What are the basic functions connected with these areas?
3. Make a diagram of the possible pathways that may connect these areas (including the intermediate processing stations)
4. Which pathway(s) is the most likely one for mediating synchronous activity of the two areas (argue how you come to this conclusion)?

Conceptually, such a practicum allows students to get a grasp on fMRI data analysis with their precise localization specifications – and further, to learn about the interpretation of time series data on the basis of structural connections. Such structure-function relationships are far from being straightforward (Honey et al., 2009), and the practicum delivers this notion by the fact that activation regions are frequently not directly connected in an anatomical sense, but these connections are mediated by intermediate regions.

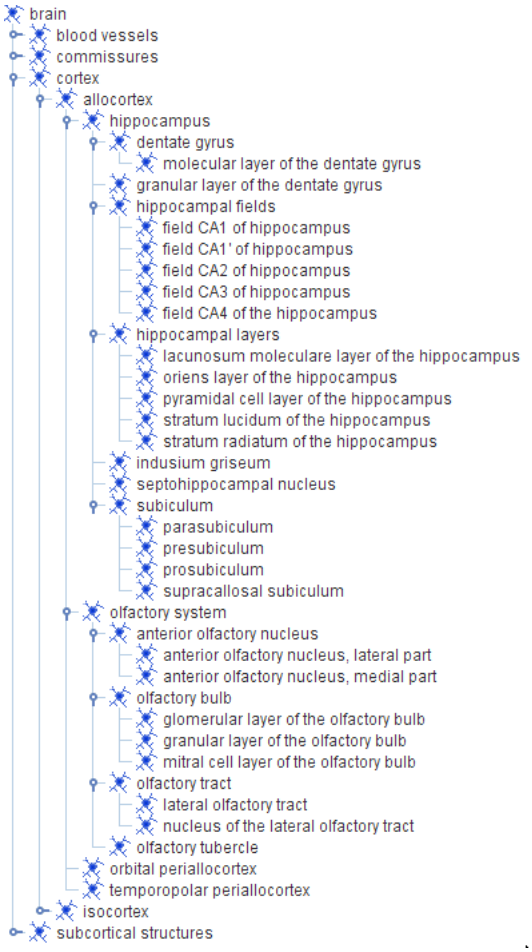


Figure 4.7.1. An example of multilevel organization of atlas structures, expanding only those related to allocortex.

4.7 Perspectives

One of the short-term updates planned to be made is creating a tree of structures with an arbitrary number of levels. This will allow the representation of the entire macaque brain as a hierarchy spanning multiple levels of organization. Fig. 4.7.1 depicts such a tree, with only branches representing allocortical structures expanded and other branches collapsed. The profit of such a representation is twofold: first, it allows for more extensive database queries, involving also those structures on the lower precision scale; the second advantage is compatibility of atlas structures with brain vocabulary systems such as NeuroLex (www.neurolex.org), employing a unified scheme of brain regions used by a various resources. The feature is at the moment

partially implemented and will appear on the forthcoming 3.2.1 version of the tool, together with possibility of multiple structure selection and correspondingly adjusted connectivity queries.

Surface smoothing has been an issue as well, although a certain degree of freedom in aligning adjacent sections may lead to a significant loss of important spatial information. Therefore, section alignment would make the brain look esthetically nice, although many individual features would be lost. Nevertheless, an algorithm which is able to recognize methodological misalignments in a proper way would be a meaningful tool to employ in this paradigm.

Speaking about database applications, the features can be further extended from mere retrieval and analysis of connectivity data. For example, displaying additional data modalities like DSI (Tuch et al., 2005) in combination with displayed tracing data would also provide information on fiber trajectory; data on minicolumns and single neurons available for macroscopic regions from the atlas can complement the tool in a way as described in a human connectome paper (Sporns et al., 2005). Finally, the possibility to generalize the tool for other species' templates awaits appropriate datasets, which can involve different brains, different species, or even different geometric data modalities, not necessarily related to neuroinformatics.

Selection of multiple areas can enable an adjustment of queries for obtaining fully specified connectivity matrices attributable to a specific set of regions. Several conceptual issues should be taken into account in this regard. First, the mapping data search should be done extensively throughout the entire database, so as to avoid gaps in the resulting matrix, or contaminating it with wrong arguments. Secondly, redundant and contradictory statements should be also treated appropriately and possibly taken into account while doing subsequent connectivity analysis. For instance, two statements referring to the same kind of connection but with different densities are specified; the connectivity path analysis should take all of those into account and possibly provide several alternative outputs. Since such a feature would complicate the analysis procedure, at the moment matrix is displaying the most reliable output. Finally, such extensive queries usually take rather long time to process – hence the first issue (mapping data search) has to be adapted in order to reduce overall query processing time.

Moreover, connectivity data can be displayed in a different way by taking into account injection sites and labeled regions. That would provide a possibility to visualize raw data from the database, and also tell how redundant and contradictory statements might be influenced by

the fact that different tracer injections were used – i.e., one at soma and another at axonal terminals. Ideal option would be to provide a possibility for several modes of connectivity display (raw versus processed data, density versus reliability, etc.) so that user would choose a preferred one.

4.8 Discussion

Conceptually, the tool presented in this chapter enables linking between macroscopic volumetric structures and their ontological representations. The final tool complements both of these representations: on the one hand, paper atlases partition the brain into spatially segregated anatomical entities, whereas textual databases represent these entities with related set of rules and without accurate spatial references. CP3D takes both of these benefits and combines them together. This link is provided by two aspects. First, all areas are organized in a hierarchical way, compatible with a database and extended for representing the brain as a multi-level neuroanatomical structure. Second, these hierarchically represented areas are further interfaced with a database on connectivity. The CP3D tool is filling the gap in available software resources as stated in (Van Essen and Dierker 2007, Table 1, 12th row).

The tool GUI provides such functionalities as connectivity visualization and analysis, stereotaxic grid with volume and centre-of-mass information, along with various three-dimensional manipulation possibilities. Transparency can be adjusted using the slider and is particularly useful for viewing subcortical nuclei and cortical structures located within deep sulci, and also for a graphical display of connectivity data with billboard labels indicating corresponding names of sources and targets resulting from a database query. All these standalone capabilities are useful for such applications as stereotaxic surgery planning. In this respect, additional probabilistic information on macaque neuroanatomy would be of a benefit. Such information is planned to be included as soon as it becomes available, which will help researchers and surgeons to reduce uncertainties caused by intersubject variability.

The easiest and fastest way to obtain necessary connectivity data from the database is an acronym mapping. This concept is sufficiently robust given that the cases of ambiguous acronyms are handled (Table 4.3.1); nevertheless, queries based on mapping paths, as described in the subsection 4.4.1, are providing more reliable output, since all statements processed by ORT are derived from original tracing and delineation studies. With the forthcoming update of CoCoMac and its migration from MS Access to MySQL, this process will become of increased efficiency, both in terms of time and the output reliability. The former aspect is explained by adjusted ORT routines with a decreased time complexity, and the latter is explained by the fact that more modern studies will be included in the database, gradually becoming a community project. Adjustment of software components to these new features is expected to be straightforward since SQL statements, which are the heart of all queries, are in general consistent between different relational databases. Inclusion of algorithmically generated statements obtained using SORT (see chapter 3; Bezgin et al., 2008) will

further increase the number of statements and fill the gaps in mapping and connectivity data.

In comparison to the surface-based tools (e.g., Caret, introduced in the Chapter 3), there are several aspects to be mentioned. The major difference is that such tools represent cortex as a folded sheet which can be unfolded or flattened for visualization convenience. Obviously, such data are in general easier to navigate due to the visualization intuitiveness and lower dimensionality (technically, cortex is a two-dimensional sheet in such a representation). Of relatively higher ease are also deformation procedures: intersubject or interspecies warping is generally easier to implement in two than in three dimensions. Therefore a number of overlays of structural and functional cortical data can be displayed on various surface templates, either individual or population-average, taking into account various species. The Paxinos et al. atlas overlay also exists in Caret, making a comparison with our tools possible. Clearly, no subcortical structures can be specified on Caret surfaces – nevertheless, Caret was one of the first tools to retrieve connectivity data from CoCoMac, although this interface was provided in a highly simplified way.

CP3D, as compared to Caret and other surface-based analysis tools, provides such functionalities as volumetric measurements and inclusion of numerous subcortical nuclei and non-neuronal structures such as blood vessels and ventricles. Moreover, volume-based structural data, registered in stereotaxic space, can be aligned with structural MRI templates, meaning that it can be also further overlaid with fMRI time series. Such an approach can be carried out with preserving of cortical thickness and other volumetric properties, but requires linear or non-linear volume deformation. The subject of combining sectional anatomy with neuroimaging data and subsequent 3D reconstruction has been recently addressed in the context of rodent brain research (Hjornevik et al. 2007).

An associated tool called Scalable Brain Atlas, supported by INCF, provides additional functionalities like viewing sections slice-by-slice and (most importantly) a link to NeuroLex, a dynamic lexicon of neuroscientific concepts – the initiative supported by the Neuroscience Information Framework (NIF). The tool is currently being under development headed by Rembrandt Bakker and will be a major part of the querying interface for the new version of the CoCoMac database (Fig 4.8.1).

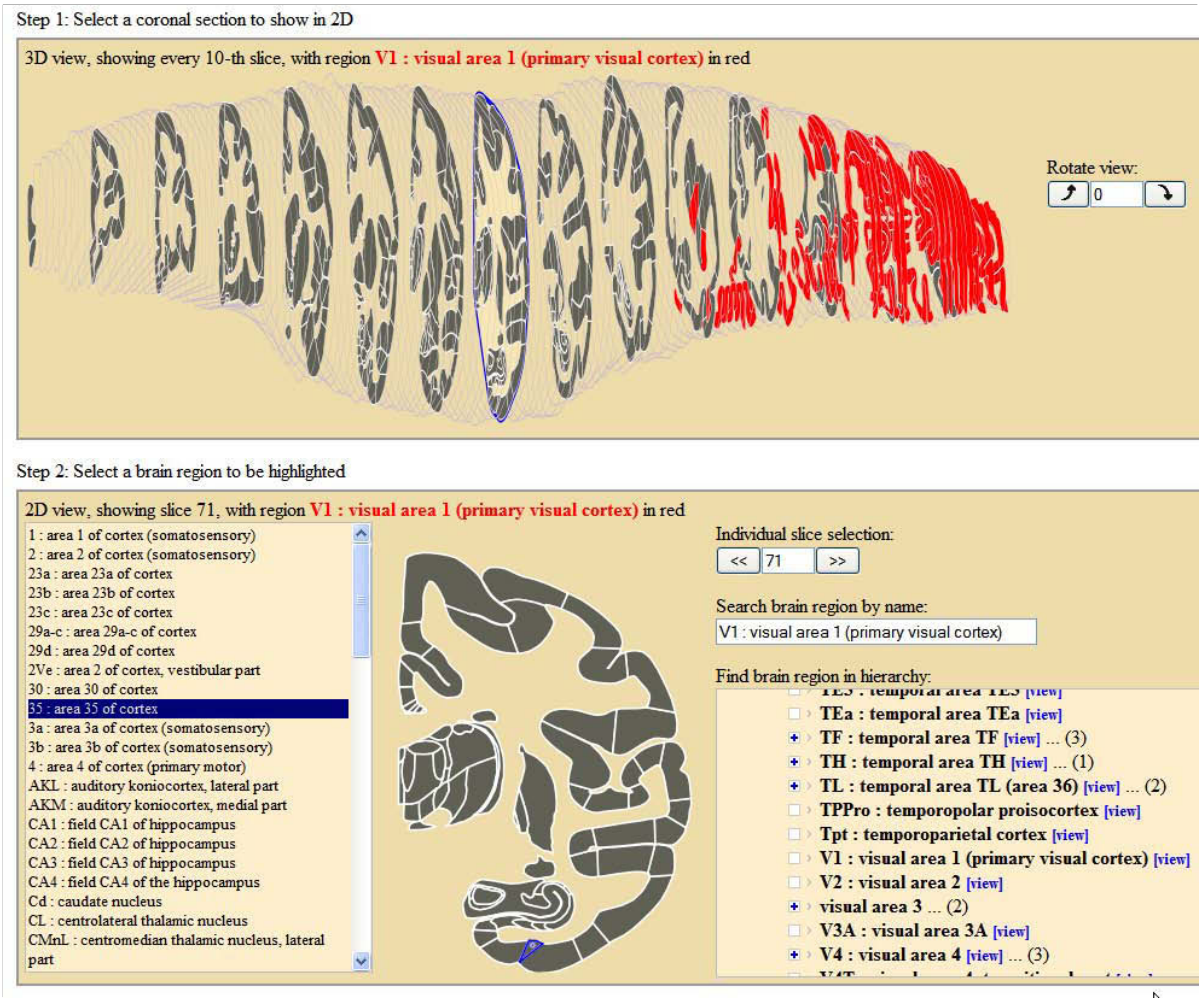


Figure 4.8.1. A snapshot of the Scalable Brain Atlas (www.scalablebrainatlas.incf.org).

Teaching applications of the CP3D tool can involve not only neuroimaging data analysis, but also other ways of use like planning of stereotaxic surgery, or superimposition of a connectivity query output on tractography data from DTI. The latter concept would link tracing data having information about directionality, with data on probabilistic tracking, which is missing any information about direction but describing presumable trajectories of fiber bundles. The approach for linking these two data modalities can be empowered by the concept similar to the one described in the subsection 3.4.1, on the basis of Bayes' theorem. Possibly the best way to proceed in this paradigm is to set tract tracing data as a prior, given its higher reliability and uncertainty about DTI tracts.

Chapter 5

General discussion

With currently growing wealth of health and disease studies on the primate brain, one of the most important paradigms in neuroscience is handling and interpreting data they produce. Indeed, the experimental research protocols are constantly changing, a greater precision increases the complexity of data, combination of various modalities also adds to the overall intricacy of the issue. Many of these aspects can be foreseen – at least, data management tools can be provided with increased flexibility, and the policy of sharing data and exploiting the tools for their mining can be adapted for extrapolated requests from the coming years of brain research.

Combining various measurement modalities on the one hand and utility of the former invasive animal studies on the other hand in particular is already discussed by the broad research community and its necessity is beyond doubt. Indeed, up to now most of the relevant insights about brain structure, function and chemistry were provided by animal studies, therefore their utility for the human brain research has been already used in previous investigations. The novelty of approaches described in this thesis is in fact represented by the issue of automatization of these efforts, which were previously performed mostly on the manual level and were therefore highly researcher-dependent.

Development of algorithms is also researcher-dependent in a sense that in the heart of the model that a particular neuroscientist devises usually lies his heuristic assumption about how he sees the model's output. However, non-parametric models search for objective criteria of a particular sort of output, based on data and not on assumptions derived from these data. This fact doesn't imply that one has to ignore parametric models: "…you cannot do inference without making assumptions…" (Bishop, 2006, Ch. 3, p. 51); sometimes the model has to be parametrized to ensure it gives a reliable output. The question is which model to choose – and the best way to find this out is to make an objective comparison between various methods (sections 3.4.1-3.4.3). In this paradigm, SORT approach can be further generalized on the basis of provided data, given that forthcoming neuroimaging studies will offer more accurate spatial precision, and animal experiments will include further fine details about structures with all relations and connections between these structures.

New generation of brain atlases is distinguished by a tendency to involve a number of measurement modalities, provision of a growing wealth of metadata, and also by increased number of accompanying software tools. The latter aspect is actually requested by the first

two: various resources scattered across libraries and web-pages are merely impossible to combine for gaining a mutual profit unless the necessary software is provided. CP3D tool is a good example in this regard: a stereotaxic atlas with a plain list of brain structures on one hand, and connectivity database with a knowledge about the domain of areal relationships on the other are initially two very distinctive resources, and to find correspondence between them on the manual level could take a lifetime. CP3D combines these features in a systematic way, gaining benefits from both of these domains.

To join research endeavours of various research groups and to ensure none of these groups are reinventing the wheel, it is necessary to organize collaborative efforts worldwide. Frequently, different groups are tackling similar issues and therefore can be combined as subsequent steps constituting a pipeline. In this paradigm, SORT and CP3D can be systematized as one large pipeline, in which the former approach extracts appropriate data from spatial and logical domains, whereas the latter tool uses these generated data to further process them in an analytic way. Specifically, data generated by SORT comprises statements of geometric relationship between different areas in different maps. It represents a highly efficient way to fill the gaps between existing statements in animal databases, since experimental data are generally provided by researchers who are interested in a very specific region, commonly not taking into account interaction between different subsystems. SORT generates such statements on the basis of geometrical relationships, which are not biased to any particular cortical subsystem. Further, these generated data are included in the database and, with an aid of approaches like ORT, new connectivity statements are produced. These statements are further retrieved, visualized and analyzed by CP3D. This pipeline is summarized in Fig. 5.1.

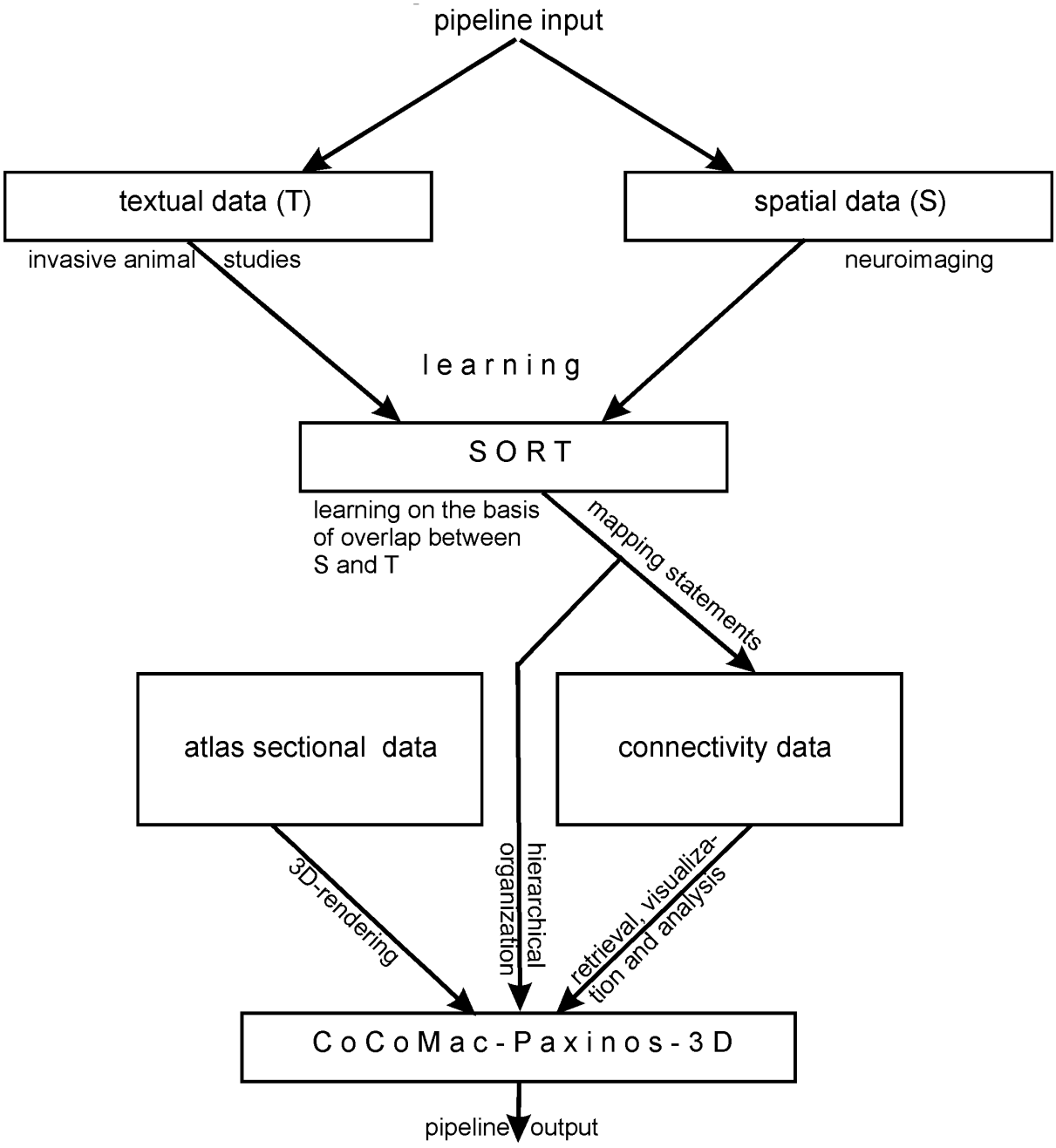


Figure 5.1. The SORT-CP3D pipeline summarizing concepts described in this thesis. Pipeline takes as an input both spatial and logical data, learns the correspondence between the two modalities by means of SORT, which further generates mapping statements for the database. These statements are used to generate new connectivity data which is visualized by the CP3D tool representing an electronic version of a sectional atlas with structures organized hierarchically to ensure compatibility with the database.

The example on the utility of SORT for filling the gaps in the database is demonstrated in Fig. 3.5.2. For evaluating gaps in connectivity within other maps, we checked the same measures

for the Brodmann's macaque parcellation scheme (Brodmann, 1905), representing a very coarse map (29 cortical regions) versus a fine partitioning by Paxinos et al. (207 cortical regions) – see Fig. 5.2.

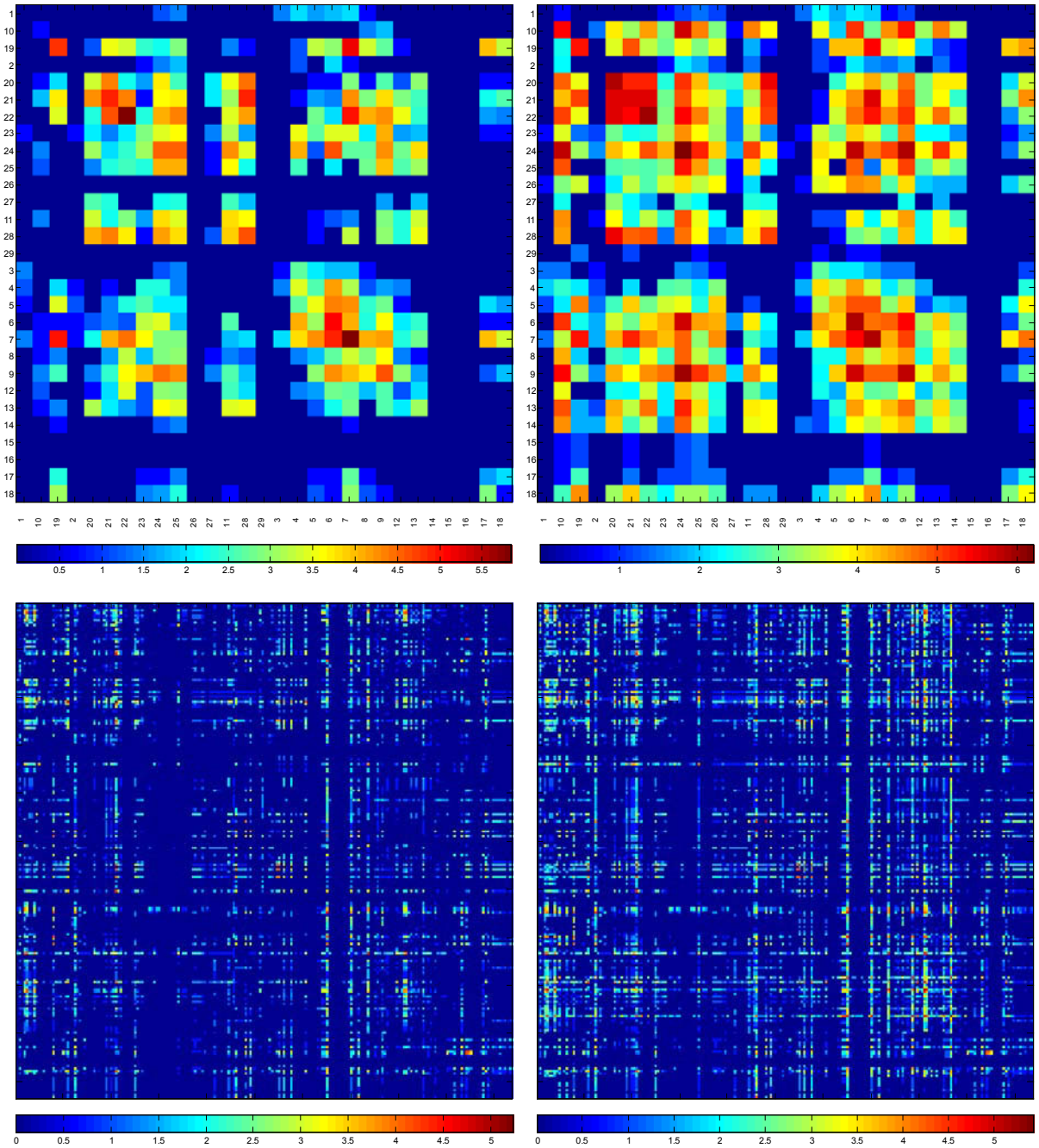


Figure 5.2. Impact of SORT statements on connectivity data in CoCoMac associated with the Brodmann's (top) and Paxinos' parcellations. See Fig. 3.5.2 for an explanation of displayed values.

Notably, SORT statements fixed 45% gaps (matrix entries with no associated labelled sites) in Brodmann's map and 11% in the Paxinos et al. atlas. Such a difference seems to appear

from the notion that connectivity patterns on finer scales are still rather uncertain; even a significant increase in the number of necessary mapping statements does not greatly influence information about connections, since animal studies still do not provide much of extensive quantitative data on connectivity within such detailed parcellations. This issue is getting resolved to a certain extent as contemporary animal brain studies become more accurate and comprehensive. At the moment, connectivity data retrieval, visualization and analysis (in particular, by the CP3D tool) are empowered by mapping statements and usability of larger structures in the hierarchical tree.

The higher reliability of SORT is provided by objective criteria for learning on existing spatial and logical mapping statements. From a neuroimaging-analytical viewpoint, such statements might be considered less accurate than spatial references in Euclidean space; nevertheless, such statements are serving as a necessary aid for spatial relationships given that the major wealth of structural mapping and connectivity data are attributed to a non-human primate brain, and the level of detail and comprehensiveness of its vocabulary would provide a major benefit for voxel-based or any other spatial data on human brain, initially missing abovementioned metadata. An example on how CP3D delivers such metadata to an accurate spatial reference could also serve as a workbench for future developments on corresponding human brain software tools, as such data becomes more available. In this paradigm, the following way might be considered: deformation from monkey to human and import of corresponding logical statements from monkey brain areas to their corresponding human brain homologues. It might also be that advances in neuroimaging will allow direct assignment of logical statements to appropriate regions – in this case, the CP3D would serve as a template framework for development of similar software on human brain, or usage of CP3D itself with newly defined regions and relationships between them.

Summary

In this thesis, the concept of mapping between macroscopic brain regions is presented, taken into account spatial (registered in the 3D space) versus non-spatial (represented by text/figures) datasets. The issue is timely and is addressed by the broad research community, particularly by researchers involved in neuroimaging, where one has to attribute functional activations located in space to anatomical entities. The latter ones are frequently hard to discern without taking into account extensive anatomical findings from past years, most of which are obtained from invasive animal studies. Such studies are collated in large databases with precise hierarchical vocabularies yet without spatial reference frames, making these anatomical datasets hard to compare with neuroimaging data. To cope with these problems, we devised two techniques, serving together as a pipeline for linking spatial with ontological brain entities.

The first tool is called SORT which stands for "Spatial Objective Relational Transformation" and represents a machine learning approach for generating animal database-compatible statements of relationship between various brain regions, from spatial (including human) data. The approach makes use of set theory, Bayes' theorem and Voronoi diagrams among other techniques employed for learning these statements.

Another tool is called CoCoMac-Paxinos-3D (CP3D) and is linking sectional atlas data with a connectivity database for further visualization and analysis. Connectivity dataset is empowered by the SORT output, providing missing statements. Apart from the database interface, the CP3D tool organizes brain regions as a hierarchy, providing taxonomy compatible with widely used neurological thesauri (e.g. NeuroLex, www.neurolex.org). The tool is exploiting the Paxinos et al. Rhesus Macaque stereotaxic atlas and the CoCoMac tract tracing database, but is flexibly generalizable to other atlases and databases as well. The examples of utility of both tools are discussed in various chapters of this thesis, and their free availability on the internet allows users to explore the tools and give a feedback necessary for further developments.

Zusammenfassung

In dieser Arbeit wird das Konzept vorgestellt, welches die Kartierung von makroskopischen Hirnregionen ermöglicht, wobei räumliche gegen nicht-räumliche Datensätze einbezogen werden. Das Thema ist von großer Aktualität und beschäftigt eine große Anzahl von Wissenschaftlern, insbesondere im Bereich der neurowissenschaftlichen Bildgebung (Neuroimaging), wo es von Bedeutung ist, Aktivierungen im Raum anatomischen Objekte zuzuordnen. Letzteres kann häufig nicht ohne Berücksichtigung umfangreicher anatomischer Erkenntnisse der vergangenen Jahre erreicht werden, die hauptsächlich durch invasive Tier-Studien gewonnen wurden. Solche Studien werden üblicherweise in großen Datenbanken mit präzisem hierarchischem Vokabular aber ohne räumlichen Bezugsrahmen zusammengestellt, so dass diese Informationen nur schwer mit den Daten aus bildgebenden Verfahren verglichen werden können. Um diese Probleme zu lösen, haben wir zwei Techniken entwickelt, mit denen räumliche und ontologische Daten des Gehirns verbunden werden können.

Die erste Technik nennt sich „SORT“, was für "Spatial Objective Relational Transformation" steht. Sie basiert auf einem Machine- Learning-Ansatz, mit dem Aussagen aus räumlichen Daten (einschließlich solcher aus Studien mit Menschen) generiert werden können, die mit jenen Datenbanken kompatibel sind, die auf Tierstudien basieren. Diese Aussagen werden durch die Mengenlehre, Bayes-Theoreme, Voronoi-Diagramme und andere Techniken erlernt.

Die zweite Technik nennt sich „CoCoMac-Paxinos-3D (CP3D)“. Dieses Tool ermöglicht als ein Interface zwischen anatomischen Schnitt-Atlas-Daten und der Konnektivitätsdatenbank die weitere Visualisierung und Analyse. Fehlende Informationen in der Konnektivitätsdatenbank werden durch die Ergebnisse aus der SORT-Analyse ergänzt. Zusätzlich zu dem Datenbank-Interface, organisiert das CP3D-Tool Hirnregionen als eine Hierarchie, die in ihrer Taxonomie mit weit verbreiteten anderen neurologischen Thesauri kompatibel sind (z. B. NeuroLex, www.neurolex.org). Das Tool nutzt den stereotaktischen Rhesusaffen-Atlas von Paxinos et al. sowie die Tracing-Datenbank CoCoMac, kann aber flexibel auf andere Atlanten und Datenbanken angewendet werden. Die freie Verfügbarkeit der Tools im Internet ermöglicht den Benutzern, die Instrumente auszuprobieren und ein Feedback zu geben, welches für die weitere Entwicklung hilfreich ist. Diese und weitere Beispiele für den Nutzen der entwickelten Techniken werden an verschiedenen Stellen in der vorliegenden Arbeit diskutiert.

Samenvatting

Dit proefschrift beschrijft de concepten van het mappen van macroscopische hersengebieden, en vergelijkt spatiële- en niet-spatiële datasets. Aan dit actuele onderwerp wordt onderzoek gedaan door vele wetenschappelijke richtingen, voornamelijk neuroimaging, waar het geacht wordt causale verbanden bloot te leggen tussen activaties in de ruimte en hersen-anatomische entiteiten. De laatste zijn vaak moeilijk te beschrijven zonder gebruik te maken van anatomische ontdekkingen, vaak verkregen uit dierexperimenten, van de afgelopen jaren. Deze studies zijn bijeengebracht in grote databases en zijn voorzien van een duidelijke, hiërarchische naamgeving, maar ontberen ruimtelijke referentiekaders waardoor ze moeilijk te vergelijken zijn met data verkregen door neuroimaging. Om deze problemen het hoofd te bieden hebben we twee softwareprogramma's ontwikkeld die samen de spatiele- en de ontologische hersengebieden met elkaar verbinden.

Het eerste programma, SORT (Spatial Objective Relational Transformation) maakt gebruikt van machinaal leren voor het maken van dierlijke database-gekoppelde voorspellingen over spatiele (inclusief menselijke) data. Deze benadering maakt naast technieken die gebruikt worden om deze voorspellingen te leren gebruik van de verzamelingenleer, theorema van Bayes en Voronoi diagrammen.

Het tweede programma, CoCoMac-Paxinos-3D (CP3D), verbindt sectionele atlas data met een connectiviteits-database omwille van verdere visualisatie en analyse. Deze connectiviteits-database wordt gevoed bij de output van SORT, en zorgt voor de opvulling van ontbrekende stellingen. Naast de interface van de database rangschikt CP3D de hersengebieden hiërarchisch, wat zorgt voor een taxonomy die verenigbaar is met veelgebruikte neurologische thesaurus (bijvoorbeeld NeuroLex, www.neurolex.org). Het programma gebruikt van zowel de "Rhesus Maquaque stereotactic atlas" als de "CoCoMac tract tracing database" van Paxinos et al. maar is gemakkelijk te gebruiken met andere atlassen en databases. Voorbeelden van het gebruik van deze programma's worden besproken in verschillende hoofdstukken van dit proefschrift. Doordat deze programma's op het internet vrij toegankelijk zijn hebben gebruikers de mogelijkheid deze te gebruiken en feedback te geven die kan helpen deze programma's verder te ontwikkelen.

Краткий обзор

В данной докторской диссертации представлена концепция картирования между различными макроскопическими регионами мозга, учитывая, что эти регионы представлены, с одной стороны, в виде пространственных координат, и с другой - в виде текстовых данных. Данная тема актуальна и адресована многими исследователями - в особенности теми, которые занимаются нейроимаджингом, где необходимо картировать активации различных регионов к анатомическим структурам. Последние зачастую вызывают сложность в спецификации, если не брать в оборот данных исследований на животных прошлых лет. Подобные данные зачастую собраны в больших базах данных с точными иерарчическими таксономиями, но без пространственных координат, что вызывает трудность в сравнении этих данных с данными нейроимаджинга. Для разрешения этих проблем мы предлагаем две техники, представляющие вместе некий "конвейер" для связи пространственных и текстовых данных нейронауки.

Первый инструмент называется SORT и представляет собой технику из области машинного обучения для генерирования новых текстовых данных на основе пространственных координат (в том числе данных человеческого мозга). В этой парадигме используются теория множеств, теорема Байеса и диаграммы Вороного, помимо прочих техник, необходимых для генерирования этих текстовых данных.

Другой инструмент называется CP3D и служит для связи между атласом мозга в пространственных координатах и базой данных связности, пополненной данными, генерированными также с помощью SORT. Помимо связи с базой данных, CP3D организует все зоны мозга иерархически, обеспечивая таксономию, совместимую с другими нейрологическими тезаурусами (например, NeuroLex, www.neurolex.org). CP3D использует атлас мозга, выполненный Паксиносом и др. в 2000, а также базу данных CoCoMac - однако, гибкий интерфейс позволяет адаптировать этот инструмент для других атласов и баз данных. Примеры использования обоих инструментов обсуждены в данной диссертации, а открытый доступ к ним в интернете предполагает широкий круг пользователей, чья обратная связь необходима для последующих разработок.

References

1. Amari S., Beltrame F., Bjaalie J.G., Dalkara T., De Schutter E., Egan G.F., et al. (2002) Neuroinformatics: the integration of shared databases and tools towards integrative neuroscience. *J Integr Neurosci.* **1(2)**: 117-28.
2. Amunts K., Schleicher A., Bürgel U., Mohlberg H., Uylings H.B.M., Zilles K. (1999) Broca's region revisited: cytoarchitecture and intersubject variability. *CompNeurol* **412**, 319-341.
3. Andersen R.A., Asanuma C., Essick G., Siegel R.M. (2004) Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. *J. Comp. Neurol*. **296**, 65-113..
4. Bakker R., Schubert D., Levels K., **Bezgin G.**, Bojak I., Kötter R. (2009) Classification of cortical microcircuits based on micro-electrode-array data from slices of rat barrel cortex. *Neural Networks* **22**, 1159-1168.
5. Belousov A., Tkachev S. (2004) Discrete Mathematics. *Moscow Technical University Press*.
6. **Bezgin G.**, Wanke E., Krumnack A., Kötter R. (2008) Deducing logical relationships between spatially registered cortical parcellations under conditions of uncertainty. *Neural Networks* **21(8)**: 1132-1145.
7. **Bezgin G.**, Reid A.T., Schubert D., Kötter R. (2009) Matching spatial with ontological brain regions using Java tools for visualization, database access, and integrated data analysis. *Neuroinformatics* **7(1)**: 7-22.
8. Bishop C.M. (2006) Pattern recognition and machine learning. *Springer*. ISBN: 978-0-387-31073-2.
9. Bjaalie J.G. (2002) Localizations in the brain: new solutions emerging. *Nature reviews* **3**, 322-325.
10. Black K.J., Koller J.M., Snyder A.Z., Perlmutter J.S. (2004) Atlas template images for nonhuman primate neuroimaging: baboon and macaque, in Imaging.(*Methods in Enzymology 385), Conn, P.M., ed., Elsevier, New York*, pp. 91-102.
11. Bohland J.W., Bokil H., Allen C.B., Mitra P.P. (2009) The brain atlas concordance problem: quantitative comparison of anatomical parcellations. *PLoS ONE* **4(9)**: e7200. doi:10.1371/journal.pone.0007200.
12. Bonin G.V., von, Bailey P. (1947) The neocortex of Macaca mulatta. *The University of Illinois Press, Urbana, Illinois.*
13. Bota M. and Swanson L.W. (2007) Online workbenches for neural network connections*. J. Comp. Neurol.* **500**, 801-814.
14. Bowden D.M. and Martin R.F. (1995) NeuroNames brain hierarchy. *NeuroImage* **2**, 63-83.
15. Bowyer A. (1981). Computing Dirichlet tessellations. *The Computer Journal* **24(2)**:162-166.
16. Brodmann K. (1905) Beitrage zur histologischen Localisation der Grosshirnrinde. Dritte Mitteilung. Die Rindenfelder der niederen Affen. *Journal für Psychologie und Neurologie* **4**, 177-226.
17. Brodmann K. (1909) Vergleichende Lokalisationslehre der Grosshirnrinde. *The Journal of Nervous and Mental Disease* **37**, 783-784.
18. Cachia A., Mangin J.-F., Riviere D, Kherif F., Boddaert N., Andrade A., Papadoupolos-Orfanos D., Poline J.-B., Bloch I., Zilbovicius M., Sonigo P., Brunelle F., Regis J. (2003) A primal sketch of the cortex mean

curvature: a morphogenesis based approach to study the variability of the folding patterns. *IEEE transactions on medical imaging,* **vol. 22**, 754-765.

19. Chamberlin D.D., Boyce R.F. (1974) SEQUEL: a Structured English Query Language. *Proceedings of the 1974 ACM SIGFIDET Workshop on Data Description, Access and Control (Association for Computing Machinery)*: 249–264.
20. Chicurel M. (2000) Databasing the brain. *Nature* **406**, 822-825.
21. Clower D.M., West R.A., Lynch J.C., Strick P.L. (2001) The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. *J. Neurosci*. **21**, 6283-6291.
22. Crook S., Beeman D., Gleeson P., Howell F. (2005) XML for model specification in neuroscience. *Brain, minds & media* 1, 1-24.
23. Dickson J., Drury H., Van Essen D. (2001) 'The surface management system' (SuMS) database: a surface-based database to aid cortical surface reconstruction, visualization and analysis. *Phil. Trans. R. Soc. Lond.* B **356***,* 1277-1292.
24. Dijkstra E.W. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik* **1**: 269-271.
25. Ding S.-L., van Hoesen G., Rockland K. (2000) Inferior parietal lobule projections to the presubiculum and neighboring ventromedial temporal cortical areas. *J. Comp. Neurol*. **425**, 510-530.
26. Dubois D., Foulloy L., Mauris G., Prade H. (2004) Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable computing* **10**, 273-297.
27. Ebbesson S.O.E. (1980) The parcellation theory and its relation to interspecific variability in brain organization, evolutionary and ontogenetic development, and neuronal plasticity. *Cell and Tissue Research* **213**, 179-212.
28. Economo C., von, Koskinas, G.N. (1925) Die Cytoarchitektonik der Hirnrinde des Erwachsenen Menschen. *Springer, Berlin*.
29. Felleman D.J., Van Essen D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* **1**, 1-47.
30. Ferry A.T., Öngür D., An X., Price J.L. (2000) Prefrontal cortical projections to the striatum in macaque monkeys: evidence for an organization related to prefrontal networks. *The Journal of Comparative Neurology* **425**, 447-470.
31. Fischl B., Sereno M.I., Tootel R.B.H., Dale A.M. (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping* **8**, 272-284.
32. Gamberini M., Passarelli L., Fattori P., Zucchelli M., Bakola S., Luppino G., Galletti C. (2009) Cortical connections of the visuomotor parietooccipital area V6Ad of the macaque monkey. *J. Comp. Neurol.* **513(6)**: 622-642.
33. Gardner D., Akil H., Ascoli G.A., Bowden D.M., Bug W., Donohue D.E., et al. (2008) The Neuroscience Information Framework: a data and knowledge environment for neuroscience. *Neuroinformatics* **6**, 149-160.

34. Gering D. T., Nabavi A., Kikinis R., Hata N., O'Donnell L. J., Grimson W. E. L., et al. (2001) An integrated visualization system for surgical planning and guidance using image fusion and an open MR. *J. Magn. Res. Im.* **13**, 967–975.
35. Geyer S., Luppino G., Ekamp H., Zilles K. (2005) The macaque inferior parietal lobule: cytoarchitecture and distribution pattern of serotonin 5-HT1A binding sites. *Anat. Embryol*. **210**, 353-362.
36. Goddard N.H., Hucka M., Howell F., Cornelis H., Shankar K., Beeman D. (2001) Towards NeuroML: model description methods for collaborative modeling in neuroscience. *Phil. Trans. R. Soc. Lond*. **356**, 1209-1228.
37. Golbreich C., Bierlaire O., Dameron O., Gibaud B. (2005) Use case: Ontology with rules for identifying brain anatomical structures. In: *W3C Workshop on Rule Languages for Interoperability*.
38. He B.J., Snyder A.Z., Zempel J.M., Smyth M.D., Raichle M.E. (2008) Electrophysiological correlates of the brain's intrinsic large-scale functional architecture. *PNAS* **105(41)**: 16039-44.
39. Hilgetag C.C., Barbas H. (2005) Developmental mechanics of the primate cerebral cortex. *Anat Embryol* **210**, 411-417.
40. Hilgetag C.C., Barbas H. (2006) Role of mechanical factors in the morphology of the primate cerebral cortex. *PLoS Computational Biology* **3**, 146-159.
41. Hinshaw K.P., Poliakov A.V., Moore E.B., Martin R.F., Shapiro L.G., Brinkley J.F. (2001) Shape-based cortical surface segmentation for visualization brain mapping. *NeuroImage* **16**, 295-316.
42. Hjornevik T., Leergaard T. B., Darine D., Moldestad O., Dale A. M., Willoch F., et al. (2007) Three-dimensional atlas system for mouse and rat brain imaging data. *Frontiers in Neuroinformatics*, **1**, 1–11.
43. Howell F., McDonnell N., Zaborszky L., Varsanyi P. (2005) Tools for integrating 3D single cell and population neuroanatomy data. In: *UK e-science All Hands Meeting 2005 Nottingham UK*.
44. Koslow S.H., Huerta M.F. (1997) Neuroinformatics: an overview of the Human Brain Project. *Lawrence Erlbaum Associates, Washington*.
45. Kötter R. (2001) Neuroscience databases: tools for exploring brain structure-function relationships. *Phil. Trans. R. Soc. Lond. B* **356**, 1111-1120.
46. Kötter R. (2004) Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics* **2**, 127-144.
47. Kötter R. (2007) Anatomical concepts of brain connectivity. Jirsa V., McIntosh A.R. (eds.) Handbook of Brain Connectivity, Series: Understanding Complex Systems. *Springer, Berlin*, pp. 149-166.
48. Kötter R., Reid A.T., **Bezgin G.** (2008) An Introduction to the CoCoMac-Paxinos-3D Viewer. Chapter 4 in: Paxinos G., Huang X.-F., Petrides M., Toga A.W. (eds.) The Rhesus Monkey Brain in Stereotaxic Coordinates, 2nd edition. *Elsevier, San Diego.*
49. Kötter R., Wanke E. (2005) Mapping brains without coordinates. *Phil. Trans. R. Soc. B* **360**, 751-766.
50. Kruggel F., Brückner M.K., Arendt Th., Wiggins C.J., von Cramon D.Y. (2003) Analyzing the neocortical fine-structure. *Medical Image Analysis* **7**, 251-264.

51. Krumnack A. (2008) Algorithmische Analyse makroskopischer Verbindungsstrukturen im Primatengehirn. *PhD thesis*, Heinrich Heine University Düsseldorf.
52. Kwon H., Ow A.W., Pedatella K.E., Lotspeich L.J., Reiss A.L. (2004) Voxel-based morphometry elucidates structural neuroanatomy of high-functioning autism and Asperger syndrome. *Developmental Medicine & Child Neurology* **46**, 760-764.
53. Le Bihan D. (2003) Looking into the functional architecture of the brain with diffusion MRI. *Nature Reviews Neuroscience* **4**, 469-480.
54. Lewis J.W. and Van Essen D.C. (2000a) Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol*. **428**, 112-137.
55. Lewis J.W. and Van Essen D.C. (2000b) Mapping of architectonic subdivisions in the macaque monkey, with emphasis on parieto-occipital cortex. *J. Comp. Neurol*. **428**, 79-111.
56. Mangin J.-F., Riviere D, Cachia A., Duchesnay E., Cointepas Y., Papadoupolos-Orfanos D., Collins D. L., Evans A. C., Regis J. (2004) Object-based morphometry of the cerebral cortex. *IEEE transactions on medical imaging,* **vol. 23**, 968-982.
57. Mangin J.-F., Riviere D, Cachia A., Duchesnay E., Cointepas Y., Papadoupolos-Orfanos D., Scifo P., Ochiai T., Brunelle F., Regis J. (2004) A framework to study the cortical folding patterns. *NeuroImage* **23**, 129-138.
58. Martin R.F. and Bowden D.M. (1996) A stereotaxic template atlas of the macaque brain for digital imaging and quantitative neuroanatomy. *NeuroImage* **4**, 119-150.
59. Mazziotta J.C., Toga A.W., Evans A., Fox P., Lancaster J. (1995) A probabilistic atlas of the human brain: theory and rationale for its development. *NeuroImage* **2**, 89-101.
60. Mechouche A., Golbreich C., Gibaud B. (2006) Semantic description of brain MRI images. In: *First International Workshop on Semantic Web Annotations for Multimedia.*
61. Morel A., Liu J., Wannier T., Jeanmonod D., Rouiller E.M. (2005) Divergence and convergence of thalamocortical projections to premotor and supplementary motor cortex: a multiple tracing study in the macaque monkey. *Eur. J. Neurosci*. **21**, 1007-1029.
62. Mussen A.T. (1967) A cyto-architectural atlas of the brain stem of Macacus rhesus. In A.T. Mussen, The Cerebellum and Red Nucleus. Anatomy and Physiology. *Hafner, New York*.
63. Ohser J., Mücklich F. (2001) Statistical analysis of microstructures in materials science. West Sussex, UK: *John Wiley and Sons Ltd*.
64. Oommen B., Nicholson D., Conway T., Schönning A., Ionescu I. Development of finite element mesh for traumatic brain injury using patient specific CT and MRI data of the cranio-cerebral complex: *Neurocomputing, in press*.
65. Papoulis, A. (1984) "Bayes' Theorem in Statistics" and "Bayes' Theorem in Statistics (Reexamined)." §3-5 and 4-4 in Probability, Random Variables, and Stochastic Processes, 2nd ed. New York: *McGraw-Hill*, pp. 38-39, 78-81, and 112-114.
66. Paxinos G. and Watson C. (1998) The Rat Brain in Stereotaxic Coordinates, 4th ed., *Academic, San Diego.*

67. Paxinos G., Huang X-F., Toga A.W. (2000) The Rhesus Monkey Brain in Stereotaxic Coordinates. *New York, Academic Press.*
68. Petrides M., Pandya D.N. (2007) Efferent Association Pathways from the Rostral Prefrontal Cortex in the Macaque Monkey. *The Journal of Neuroscience* **27**, 11573-11586.
69. Preuss T.M., Goldman-Rakic P.S. (1991) Myelo- and cytoarchitecture of the granular frontal cortex and surrounding regions in the strepsirhine primate Galago and the anthropoid primate Macaca. *J. Comp. Neurol.* **310**, 429-474.
70. Ramnani N., Owen A.M. (2004) Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Reviews* **5**, 184-194.
71. Reid A.T., Krumnack A., Wanke E., Kötter R. (2009) Optimization of cortical hierarchies with continuous scales and ranges. *NeuroImage* **47(2)**:611-617.
72. Rizzolatti G., Luppino G., Matelli M. (1998) The organization of the cortical motor system: new concepts. *Electroencephal. Clin. Neurophysiol*. **106**, 283-296.
73. Roland P. E., Geyer S., Amunts K., Schormann T., Schleicher A., Malikovic A., Zilles K. (1997) Cytoarchitectural maps of the human brain in standard anatomical space. *Human Brain Mapping* **5**, 222-227.
74. Rolls E.T., Baylis L.L. (1994) Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex. *The Journal of Neuroscience* **14**, 5437-5452.
75. Rubens N.O. (2006) The application of fuzzy logic to the construction of the ranking function of information retrieval systems. *Computer modelling and new technologies* **10/1**, 20-27.
76. Saleem K.S., Logothetis N.K. (2006) A Combined MRI and Histology Atlas of the Rhesus Monkey Brain. *Elsevier, London*.
77. Schleicher A., Amunts K., Geyer S., Morosan P., Zilles K. (1998) Observer-independent method for microstructural parcellation of cerebral cortex: a quantitative approach to cytoarchitectonics. *NeuroImage* **9**, 165-177.
78. Schmahmann J.D., Pandya D.N., Wang R., Dai G., D'Arceuil H.E., de Crespigny A.J., Wedeen V.J. (2007) Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* **130(3)**: 630-653.
79. Seltzer B. and Pandya D.N. (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Research* **149**, 1-24.
80. Smith O.A., Kastella K.G., Randall D.C. (1972) A stereotaxic atlas of the brain stem of Macaca mulatta in the sitting position. *J. Comp. Neurol*. **145**, 1-23.
81. Smith M.A., Kelly R.C., Lee T.S. (2007) Dynamics of response to perceptual pop-out stimuli in macaque V1. *J. Neurophysiol*. **98(6)**:3436-3449.
82. Smits M., Peeters R.R., van Hecke P., Sunaert S. (2007) A 3 T event-related functional magnetic resonance imaging (fMRI) study of primary and secondary gustatory cortex localization using natural tastants. *Neuroradiology* **49**, 61-71.
83. Sporns O., Tononi G., Kötter R. (2005) The human connectome: a structural description of the human brain. *PloS Comput. Biol*. **4**, 245-251.

84. Stephan K.E., Kamper L., Bozkurt A., Burns G.A.P.C., Young M.P., Kötter R. (2001) Advanced database methodology for the collation of connectivity data on macaque brain (CoCoMac) *Phil. Trans. R. Soc. Lond. B* **356**, 1159-1186.
85. Stephan K.E., Zilles K., Kötter R. (2000) Coordinate-independent mapping of structural and functional data by objective relational transformation (ORT) *Phil. Trans. R. Soc. Lond.* **B 355***,* 37-54.
86. Szabo J. and Cowan W.M. (1984) A stereotaxic atlas of the brain of the Cynomolgus monkey (Macaca fascicularis) *J. Comp. Neurol*. **222**, 265-300.
87. Takada M., Nambu A., Hatanaka N., Tachibana Y., Miyachi S., Taira M., Inase M. (2004) Organization of prefrontal outflow toward frontal motor-related areas in macaque monkeys. *Eur. J. Neurosci*. **19**, 3328-3342.
88. Talairach J. and Tournoux P. (1988) Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - an Approach to Cerebral Imaging. *Thieme, New York.*
89. Toga A.W. and Thompson P.M. (2001) Maps of the brain. *The Anatomical Record (The New Anatomist)* **265**, 37-53.
90. Toga A.W., Thompson P.M., Mori S., Amunts K., Zilles K. (2006) Towards multimodal atlases of the human brain. *Nature Reviews* **7**, 952-966.
91. Tuch D. S., Wisco J. J., Khachaturian M. H., Ekstrom L. B., Kötter R., & Vanduffel W. (2005) Q-ball imaging of macaque white matter architecture. *Phil. Trans. R. Soc. London. B* **360**, 869–879.
92. Tzourio-Mazoyer N., Landeau B., Papathanassiou D., Crivello F., Etard O., Delcroix N., Mazoyer B., Joliot M. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain*. NeuroImage* **15**, 273-89.
93. Ungerleider L.G., Desimone R. (1986) Cortical connections of visual area MT in the macaque. *J. Comp. Neurol.* **248**, 190-222.
94. Van Essen D. C., Dierker D. L. (2007) Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron* **56**, 209–225.
95. Van Essen D.C. (2004) Surface-based approaches to spatial localization and registration in primate cerebral cortex. *NeuroImage* **23**, S97-S107.
96. Van Essen D.C., Harwell J., Hanlon D., Dickson J. (2003) Surface-based atlases and a database of cortical structure and function. *Databasing the Brain: From Data to Knowledge* (pp. 369-388) NJ: *John Wiley & Sons, Inc.*
97. Van Essen D.C., Harwell J., Hanlon D., Dickson J. (2005) Surface-based atlases and a database of cortical structure and function, in Databasing the Brain: From Data to Knowledge (*Neuroinformatics*), Koslow, S.H. and Subramaniam, S., *Wiley, New York*, ch. 22.
98. Vogt C., Vogt O. (1919) Ergebnisse unserer Hirnforschung. 1.-4. Mitteilung. *J Psych. Neurol.* **25**, 279–461.
99. Walker A.E. (1940) A cytoarchitectural study of the prefrontal area of the macaque monkey. *J Comp Neurol* **73**: 59–86.

100. Wang A.Y., Leow A.D., Protas H.D., Toga A.W., Thompson P.M. Brain warping via landmark points and curves with a level set representation: *in press*.

101. Worsley K.J., Charil A., Lerch J., Evans A.C. (2005) Connectivity of anatomical and functional MRI data. *Neural Networks* **3**, 1534-1541.

102. Yoshida W., Ishii S. (2006) Resolution of uncertainty in prefrontal cortex. *Neuron* **50**, 781-789.

103. Zadeh L. (2001) From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. *New York Academy of Sciences* **929**, 221-252.

104. Zhong Y. and Rockland K.S. (2003) Inferior parietal lobule projections to anterior inferotemporal cortex (area TE) in macaque monkey. *Cereb. Cortex* **13**, 527-540.

105. Zhou Y., Thompson P.M., Toga A.W. (1999) Extracting and representing the cortical sulci. *IEEE Computer Graphics and Applications* Volume **19**, Issue **3**, 49-55.

106. Zilles K., Schleicher A., Langemann C., Amunts K., Morosan P., Palomero-Gallagher N., Schormann T., Mohlberg H., Bürgel U., Steinmetz H., Schlaug G., Roland P. E. (1997) Quantitative analysis of sulci in the human cerebral cortex: development, regional heterogenity, gender difference, asymmetry, intersubject variability and cortical architecture. *Human Brain Mapping* **5**, 218-221.

Appendix

A1. CP3D code organization

Aiming at the audience interested in collaboration/development, here we provide the conceptual description of the CP3D code.

Java3d Handler module introduces widgets for the rendering of three-dimensional structures, in addition to those presented in Java3d API itself. It consists of the following classes:

– Java3DBehavior.java handles various structure manipulation options, such as rotating, scaling, shifting, and mouse behaviour.

– Java3DColours.java consists of static methods representing various colors options; all those methods return objects of type Color3f except for one, which generates a Material instance (see Java3D API javadoc,http://java.sun.com/products/java-media/3D/forDevelopers/J3D_1_2_API/j3dapi/javax/media/j3d/Material.html)

– Java3DDebug.java—debugger widgets.

– Java3DHandler.java, the main module class which implements basic rendering, parameter setting, and event handling.

– Java3DModel.java is a supplementary class for the previous one, providing additional necessary capabilities and setting default options.

– Java3DPanel.java integrates the features into the graphical user interface (GUI).

– Java3DPopupMenu.java does the obvious.

– Java3DSurface.java presents several options for surface rendering.

– Java3DToolbar.java is a supplementary class for Java3DPanel.java.

The MorphML handler package contains classes necessary for rendering particular brain structures. It also contains parsers for Neurolucida

(http://www.mbfbioscience.com/neurolucida/) and Accustage (http://www.accustage.com/) file formats, not used by our tools, which do not deal with microscopic structures and therefore for practical reasons exploit SVGs for graphical representations instead. Therefore, we would make an outlook at those classes being used for our particular tool. Here are the descriptions of these classes:

– Feature.java and ComplexFeature.java. They introduce the entities representing the brain structures. The former class displays entities as they are depicted in the original atlas, whereas the latter one describes those brain structures consisting of sub-features, according to the CoCoMac nomenclature.

– MorphMLHandler.java class instances are called by reference when the features have to be loaded into the atlas. The class implements the creation of new entities, which represent brain structures.

– MorphMLJava3DPanel.java extends Java3DPanel.java class (see description above), adapting it to deal with concrete brain structures. In addition to capabilities for rendering geometrical primitives (e.g., cubes, spheres, cylinders), provided in the parent class, here those for rendering specific brain features are presented.

– MorphMLRenderer.java introduces various options for brain structures rendering—representing them as a wireframe mesh, or as solid structures of particular thickness. For our tool, we used only the latter representation.

– MorphMLTree.java organizes the structures list hierarchically. In the Fig. 9, it is shown as a separate module (Structures tree) for clarity.

– MorphMLUtil.java provides additional rendering widgets, and necessary geometrical algorithms.

– Morphology.java class represents the entity containing all Feature.java and ComplexFeature.java instances, and methods for rendering the entire set of brain structures loaded into the electronic atlas.

The remaining modules were created specifically for the present tool.

The file parsers shown as SVG parser and Structure names file handler modules contain two following classes:

– SVGHandler.java. This class provides an appropriate handling for instructions given in SVG scripts for displaying the atlas graphics files. Among other instructions, there are also cubic Bezier curves present, to allow a smoother polygon representation. The following parametric form represents the mathematical basis of this instruction:

$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t) P_2 + t^3 P_3,\ t \in [0,1] \quad \text{(A1.1)}$$

using starting point P 0, end point P 3 and two control points, P 1 and P 2.

– PaxinosHandler.java. This class uses JXL API (http://docjar.com/docs/api/jxl/) developed for working with Excel files in Java. It reads the data file containing area names, acronyms, RGB patterns and descriptions of hierarchical organization for complex features.

– CoCoMac query module is represented by the classes CoCoMacHandler.java and CoCoMacQueryMaker.java. The former class enables the display of information related to connectivity, mapping, literature and brain sites in a system default internet browser, whereas the latter class connects to the database directly via a JDBC-ODBC interface bridge, and handles appropriate SQL statements necessary for displaying connectivity data graphically.

The rest of the modules are parts of uk.ac.ed.paxinos3d package, which contains appropriate widgets necessary exclusively for the current project. Here are the descriptions of the classes included:

– DataVisualisation.java and Query.java handle the graphical display of connectivity data specific for the current atlas.

– View.java creates a GUI specific for the current project. It integrates the features introduced by two classes in MorphML handler API, namely MorphMLJava3DPanel.java and MorphMLTree.java, providing a native look and feel.

– Model.java handles all the events occurring in the above class.

– Controller.java class initializes all the event handlers for the previous class.

The last three classes are integrated by the class Main.java which launches the software.

Finally, there are some supplementary classes—e.g., the one allowing to display a URL in a browser, regardless of the operating system used (Previewer.java), and a class providing native look and feel (WindowUtilities.java); constants related to this particular atlas (Paxinos et al. *2000*) are handled in a separate class Constants.java.

There are some more classes that are currently under development – most of them are concerned with the new multilevel tree structure.

A2. CP3D practicum instructions

Here we provide a complete set of instructions from the manual of the practicum described in the section 4.6.

1. Open the CP3D tool.
2. Click the *Open* button; in the dialog that appeared, choose in the first field the *Excel* file located in the data folder, which will appear in the dialog by default.
3. The second field in the same dialog will display the *svg* files containing atlas sections; select all of them by selecting one and subsequent clicking *Ctrl+A*, as shown below.

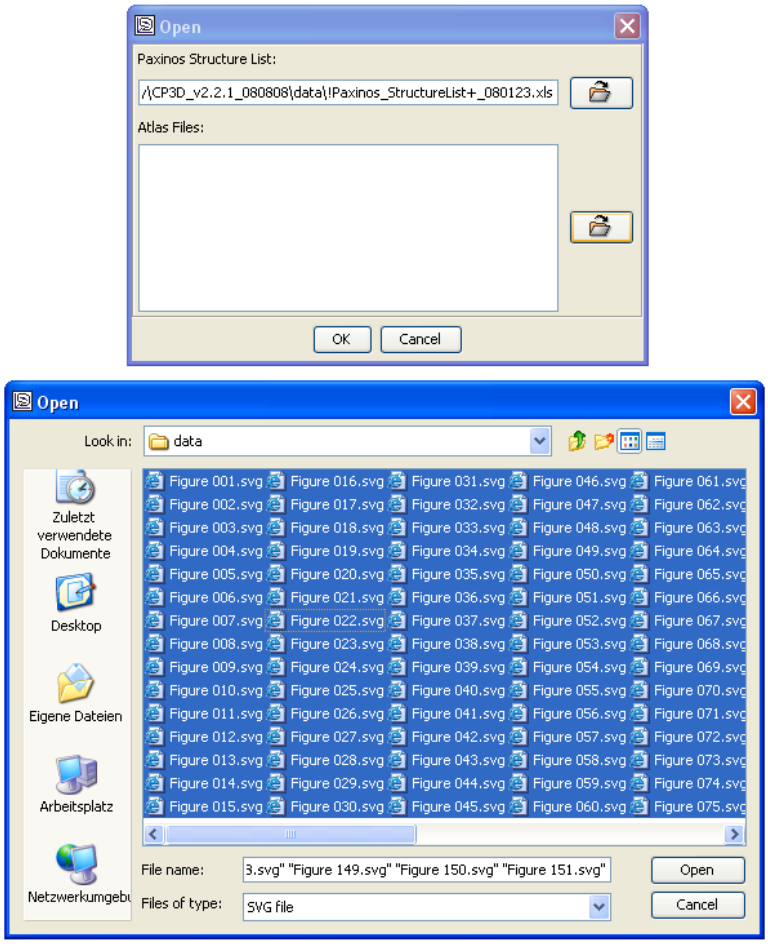


Figure A2.1. *Load atlas* dialog of the CP3D tool

4. When the atlas files are loaded, you are prepared to identify regions revealed by fMRI study shown in Figure 1. Note that stereotaxic space is the same (i.e. the coordinates shown in Figure one do correspond to those used by the tool).

5. Click the *Properties* button, and check the tick-box *Show stereotaxic grid*, then click *OK*. In the displayed grid, we are interested only in those axes displaying coordinates relative to Bregma (point on the top of the skull), coloured green. Note: to view the axis it is more convenient to zoom out the brain a bit; for this, right-click on the screen displaying the brain, choose *Zoom* mode and drag the mouse down.
6. Find the areas located **as close as possible** to two areas of activation shown in the Figure 1. WHENEVER YOU SELECT AN AREA, THESE COORDINATES ARE DISPLAYED AT THE BOTTOM (*Center relative to bregma [mm]:…*). For the top row, that is most likely the lower cortical visual area such as V1 or V2. In the bottom row, the activation area is located in a very close vicinity of so-called frontal eye fields, thus one of the areas 6, 8A, 8B or 9/46 can be the best candidate for this (**for a correspondence between full area names and their acronyms, look up the *!Paxinos_StructureList+_080123.xls* file in the data folder**).
7. When you have chosen the most appropriate **pair** of areas (let's call them the **query areas**), the next step would be to identify whether there are anatomical connections between them, presumably enabling simultaneous functional activity observed with fMRI. For this, the buttons *Afferent connectivity* and *Efferent connectivity* (for incoming and outgoing projections, respectively) are used. In the current practicum, we use **only** the *Efferent connectivity* button.
8. Select one of the two query areas (IT DOES NOT MATTER WHICH, BUT, FOR THE SAKE OF CONSISTENCY, LET IT BE THE ONE FROM THE VICINITY OF FRONTAL EYE FIELDS), and then click the *Efferent connectivity* button. Uncheck "*…as text in the browser*" box, click *OK*. Do not change anything in the window appeared afterwards, just click *OK*. Outgoing anatomical projections will be displayed. Look up the labels – if none of them are standing for another area in the query pair, there is no **direct** connection between them.

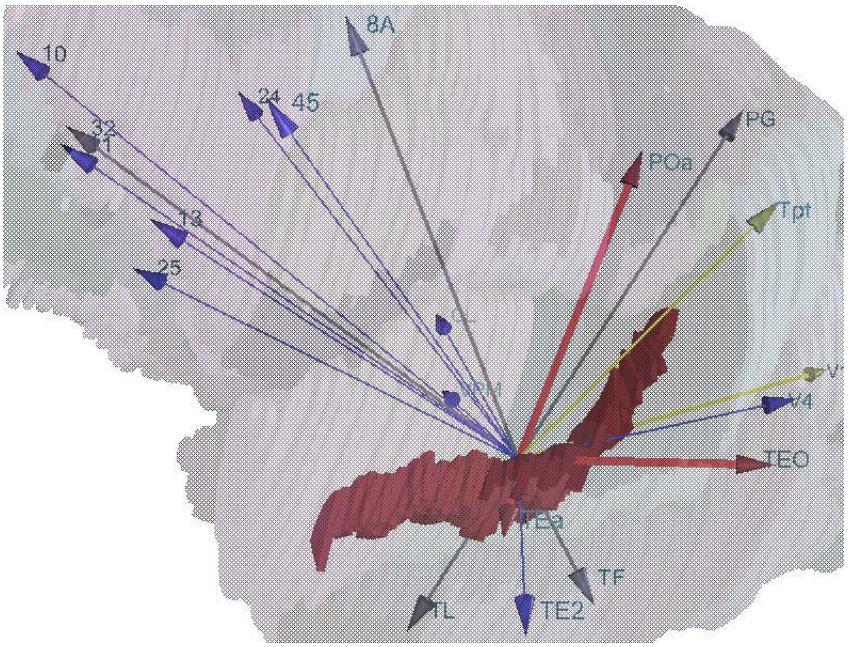


Figure A2.2. Example graphical output for CoCoMac connectivity data. Efferents from a single area.

9. Check those projections stretching in the direction to another area from the query pair, especially those with a high density. Usually it is either TPO or PG, denoting "transfer points" for visual streams in temporal and parietal lobes, respectively. Select one of these areas, repeat step 8 for it again, and then perhaps again for appeared areas, until the "path" reaches another area from the query pair.
10. As soon as the second area in the query pair is reached, apply step 8 to it. After that, click Analyse button. In the appeared table, find two paths between query areas (both ways). Is there a direct path connecting them in any direction? Is there a path at all? If yes, is the number of steps for reaching frontal eye fields (6, 8A, 8B or 9/46) from lower areas (V1 or V2) is the same as in the opposite direction? Note that the table displays **only** the shortest paths between area pairs. Can you tell whether there are alternative (longer) paths by mere visual inspection of the graphical output? For this, it would be helpful to uncheck *0 (absent)* box in the *Density values for displayed projections* panel – that will make actually absent projections invisible.

A3. The CoCoMac-Paxinos-3D manual

A3.1. System requirements

The software is fully tested on MS Windows XP platform. Several components are compatible with Linux and Macintosh OS. To use the Paxinos3D visualization tool (version 2.2.1), the following must be installed:

- **Java Runtime Environment, JRE** (1.6 or later) – available at http://java.sun.com/javase/downloads/index.jsp
- **Java3D API** (1.3 or later) – available at http://java.sun.com/products/java-media/3D/download.html

NOTE: if you have updated your JRE to a new version, make sure Java3D is reinstalled after that. This is necessary because reinstallation of JRE removes previously installed APIs, including Java3D.

An internet connection is also required to enable access to the online connectivity database CoCoMac, although viewing of data files is possible offline. It is also possible to display connectivity data from CoCoMac graphically without internet connection, since for these purposes the *.mdb file located in the `data` folder is used.

This version was designed and tested for Java SE 1.6 from Sun Microsystems, which is freely available at the above URL. Small problems may arise as a result of using earlier versions, in particular, with viewing the "About" screen.

A3.2. Software installation

Unzip the archive which contains the software (`CP3D_v2.2.1_080808.zip`) to an appropriate location on your hard disk (e.g., `Program Files` for Windows users). This should result in the creation of a folder called `CP3D_v2.2.1_080808`, and two subfolders: `bin` and `data`. The `bin` folder contains Java archive folders which constitute the executable code for CoCoMac-Paxinos3D tool. The `data` folder contains atlas data files (Paxinos *et al*., 2000), consisting of:

- An Excel file encoding RGB patterns into structure acronyms: `!Paxinos_StructureList+_080123.xls`
- 151 Scalar Vector Graphic (`svg`) files containing digitized atlas data
- the CoCoMac database file (`CoCoMac_online.mdb`) used by the software to display connectivity data graphically

To launch the Paxinos3D tool in Windows XP, run the batch file `CP3D.bat` from the main directory. To run on other platforms, it is necessary to run the main file `uk.ac.ed.paxinos3d.Main` class from the file `paxinos3d.jar`, including all `jar` files in the classpath. See the contents of the `CP3D.bat` file for an example.

A3.3. Features

The software interface contains four main features: **buttons**, the **structures tree**, the **Java3D panel**, and the **information bar**, as is illustrated below.

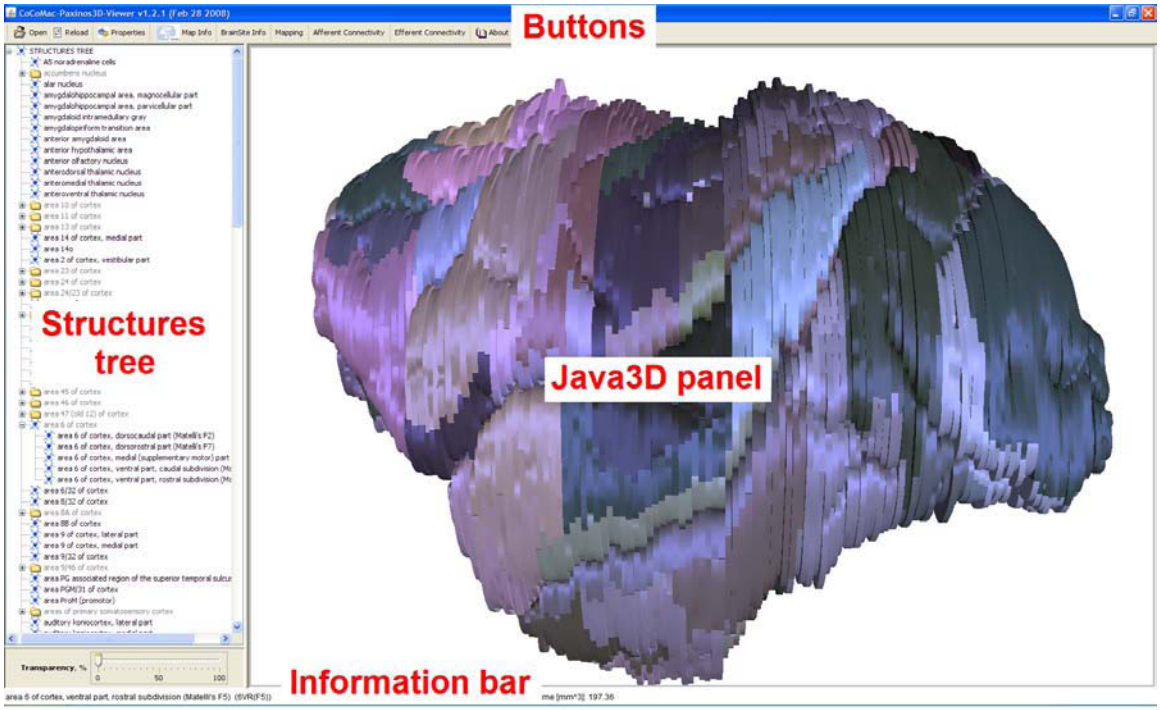


Figure A3.3.1. Paxinos3D interface features

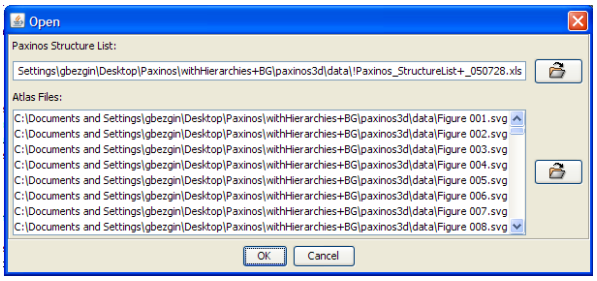
Just below the structures tree, there is a transparency slider allowing the user to adjust transparency parameters for clearer data display.

The following sections describe the features of this interface.

A3.3.1. Buttons

Open file

This button allows the user to load the atlas data. Clicking the button will bring up the following dialog:



Use this dialog to load data from the data directory. Both the structure file (`!Paxinos_StructureList+_080123.xls`), and some or all of the individual sections (`svg` files) can be selected. Note that the `svg` files, numbered 1 through 151, correspond to individual sections in the Paxinos *et al*. atlas.

Reload data

Use this button to reload the data from file. All the meta-info retrieved from the database (e.g., graphical representation of the connectivity data) will be removed by clicking this button.

Set properties for the viewer

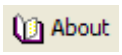
This button displays an options dialog which allows the user to specify settings.

- **Re-orient atlas on feature selection**. When selected, will alter the camera position in the Java3D display window to best display the structure selected from the structures list (see **Structures List**, below).
- **Show stereotaxic grid**. When selected, displays the stereotaxic grid relative to bregma, as represented in the original paper atlas.
- **Set background colour**. A combo box, where the items represent three different options for background colour: white, grey or black.

Various CoCoMac buttons

The buttons labelled **Map Info**, **BrainSite Info**, **Mapping**, **Afferent Connectivity**, and **Efferent Connectivity**. They operate in a following way. Each of the buttons **Map Info**, **BrainSite Info**, **Mapping**, sends a query to the CoCoMac database (http://cocomac.org) in order to display the relevant information about the selected structure. The information is displayed in the system default web browser, using the online CoCoMac interface. It includes the

mapping scheme in which the selected structure is represented and the connectivity patterns for this structure, as exist in the literature and are collated in the CoCoMac database. By contrast, **Afferent Connectivity**, and **Efferent Connectivity** buttons provide additional capabilities of displaying the connectivity data retrieved from the CoCoMac database as graphics. Each of these two buttons brings a dialog window with checkboxes enabling the user to choose whether the connectivity information should be displayed in the same way as it is done by the previous three buttons, shown graphically, or both.



About CoCoMac-Paxinos3D viewer v2.2.1

Displays a dialog containing the version number of the current product, a license agreement, a list of credits, and buttons linking to both http://cocomac.org and this document.

A3.3.2 Structures Tree

This tree, located at the left of the screen, contains all currently loaded structures. Any structure may be selected by clicking on its text. A selected structure will appear in the Java3D panel coloured red and opaque (regardless of the transparency setting). The structure name will also appear highlighted in the list, as illustrated below. Some structures include substructures – those are designated on the tree by a folder icon. Clicking on such structure will highlight all the substructures which serve as children for this superstructure. Clicking on "+" sign will browse all those children, which can be selected separately from their parent structure. Those children who have the same name as their parent structure are indicated by "*unassigned subvolume*" mark, which means that they do not have an appropriate representation in CoCoMac, due to an inconsistency with its ontology, which implies strictly hierarchical representation of structures. This concept, along with the current software, will be described in detail in forthcoming publication (Bezgin *et al*., submitted).

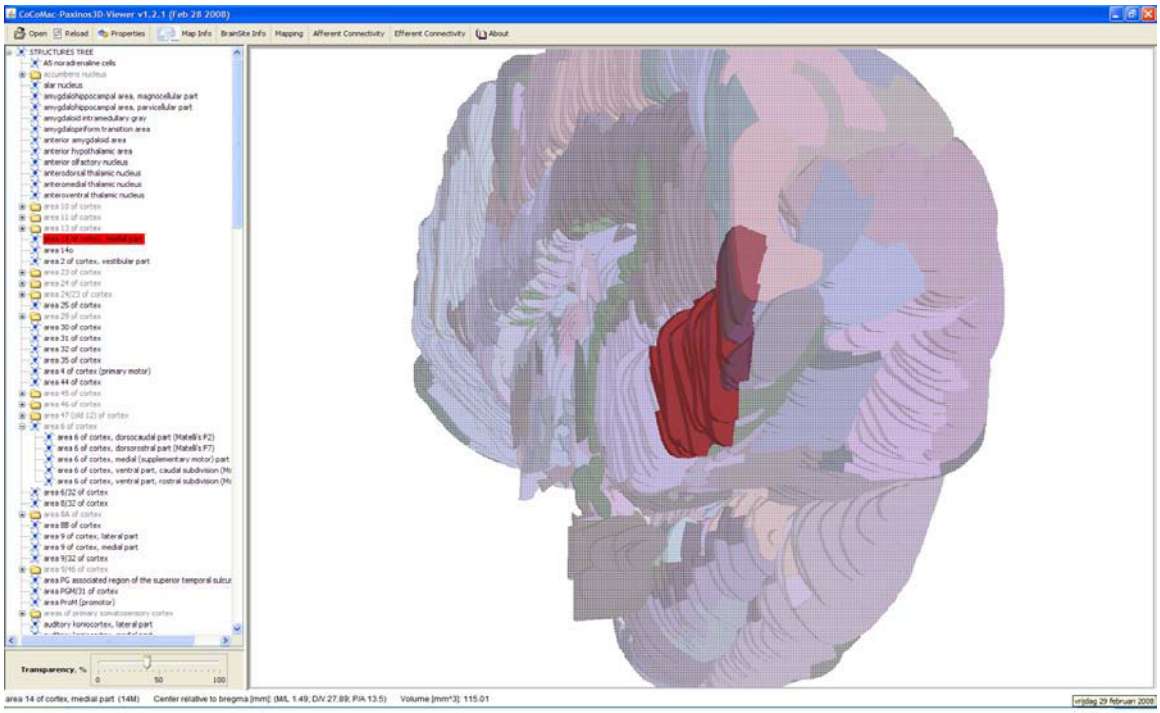


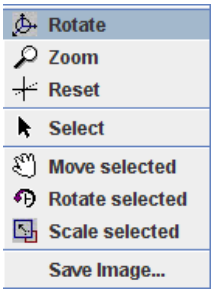
Figure A3.3.2.1. Selecting a structure from the `Structures Tree`

Right-clicking on a structure in the list also gives the user an option to show or hide the particular structure. A structure which is hidden will not be rendered in the Java3D window. This is useful for removing large overlying structures, in particular when transparency equals zero.

The user can also choose to hide all structures and show only a few select ones. To do this, right-click on the root node at the top of the tree, labelled “Morphology”, and select **hide**. The Java3D panel will appear empty. Now it is possible to display only individual structures, by right-clicking their names on the tree and selecting **show**.

A3.3.3 Java3D Panel

The Java3D panel is where the rendering of atlas structures occurs. Users can interact with the 3D model in a number of ways. The interaction mode can be specified by right-clicking on the panel, which brings up the following menu.

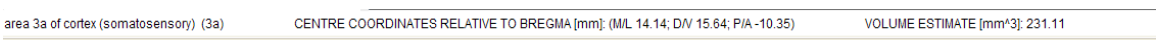


The modes are described below.

- **Select mode.** This is the default mode, which allows users to click on particular structures and select them. This is equivalent to clicking a structure name in the structures list, except that no change in camera position occurs. It is useful for identifying structures graphically.
- **Rotate.** This mode allows the user to rotate the camera around the model by dragging the mouse, and thus examine it from various orientations.
- **Zoom.** This allows the camera distance to be altered, which occurs by dragging the mouse in a particular direction. It is useful for adjusting the scope of the viewing panel.
- **Reset.** This resets the camera to its default zoom and orientation, which is a lateral view displaying the entire model.
- **Move/Rotate/Scale selected.** These options enable the user to manipulate single structures apart from the entire model. This allows a structure to be removed from the model, and inspected on its own. While there is currently no intuitive means of returning the structure to its original size and orientation, this can be done by hiding the structure using the structures list and then showing it again (see **Structures Tree**, above).
- **Save image.** This allows the current model view to be saved as an image file in `png` format. Note: in order to save the file properly, make sure that the extension `png` is specified: `myfile.png`.

A3.3.4 Information Bar

The information bar located at the bottom of the screen provides information about selected structures and (when displayed) connectivity data. This information includes the structure name and acronym, its center of mass expressed in stereotaxic coordinates relative to bregma (mm), and its estimated volume (mm^3). It is shown below. When connectivity data are displayed graphically, the information bar provides the hash table in a format {target (source) site *X*=density of connection to (from) target (source) site *X*} (the "=" sign here does not mean "equals", it indicates a mapping between hash table components).

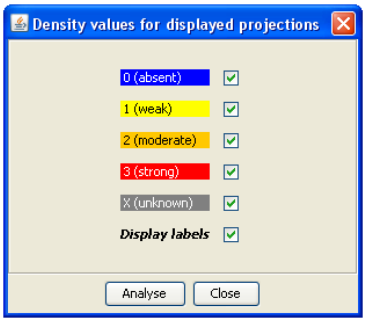


A3.3.5. CoCoMac interface features

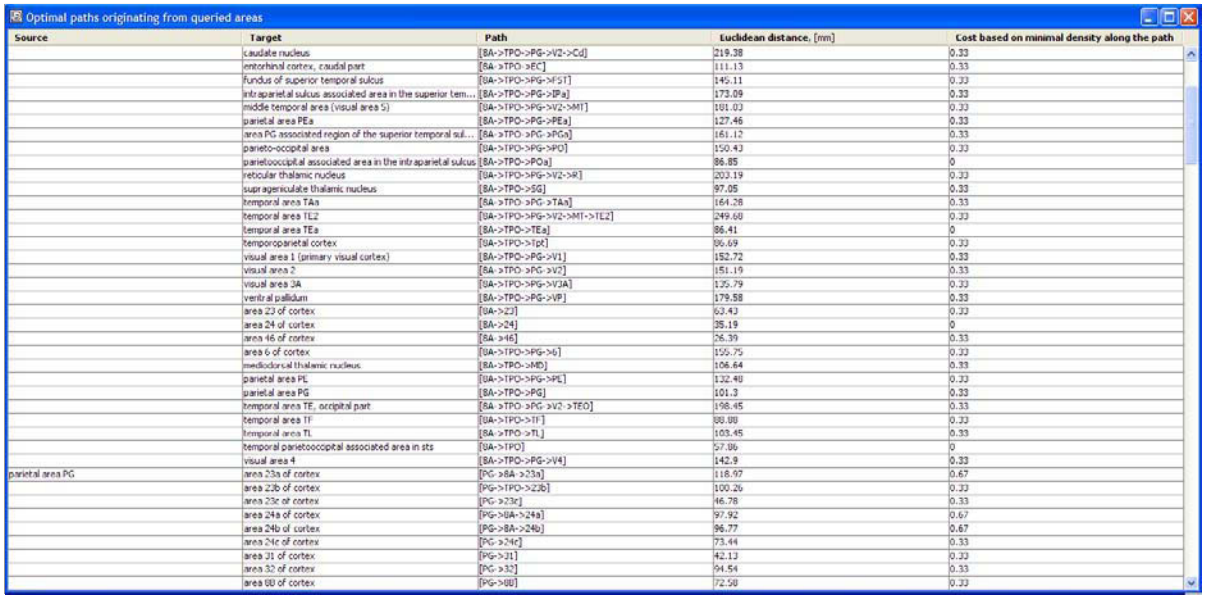
As shown above, the buttons **Afferent Connectivity** and **Efferent Connectivity** provide two options each for displaying the connectivity data from the CoCoMac database, related to the brain structures currently loaded into the atlas: to display these data either as text in a default system browser, or graphically. If the latter option is chosen and confirmed by the **OK** button, the user will be asked to choose one of three types of queries, namely:

1. Query based on *primary projections*, i.e. those connectivity statements extracted from the literature, based on explicitly specified acronyms from the Paxinos *et al*. atlas.
2. *Integrated primary projections* are primary projections with redundant and contradictory data being filtered out.
3. Connectivity data retrieved from the database by obtaining so-called *mapping paths*, i.e. sequences of mapping statements derived from the literature. They allow, in particular, to map the data from the Paxinos *et al*. atlas, to which no connectivity data are attributed, to other database resources containing those connectivity data. Therefore this type of query provides greater reliability than the first two. Note that it might take up to several minutes to process this query, depending on the amount of mapping data available in the database for a particular brain region.

When this choice is made, it will be indicated in the information bar that the query is being processed. When the data are displayed, and the information bar displays query results, the following dialog will also appear:



For each connection datum entered into the CoCoMac, the density (connection strength) information is specified. The dialog allows making the connections of particular density invisible, if they are of no interest for the user. Usually, this is the case for absent (zero-density, colour coded blue) connections, which are not essential for most users to be seen. Note that if the atlas is reloaded, these graphical data are no longer displayed.
Clicking the **Analyse** button brings a table containing all the optimal connection paths for areas being queried, based on Dijkstra's single source shortest path algorithm, like shown below:



Optimal paths originating from queried areas

Source	Target	Path	Euclidean distance, [mm]	Cost based on minimal density along the path
	caudate nucleus	[8A->TPO->PG->V2->Cd]	219.38	0.33
	entorhinal cortex, caudal part	[8A->TPO->EC]	111.13	0.33
	fundus of superior temporal sulcus	[8A->TPO->PG->FST]	145.11	0.33
	intraparietal sulcus associated area in the superior tem...	[8A->TPO->PG->IPa]	173.09	0.33
	middle temporal area (visual area 5)	[8A->TPO->PG->V2->MT]	101.03	0.33
	parietal area PEa	[8A->TPO->PG->PEa]	127.46	0.33
	area PG associated region of the superior temporal sul...	[8A->TPO->PG->PGa]	161.12	0.33
	parieto-occipital area	[8A->TPO->PG->PO]	150.43	0.33
	parietooccipital associated area in the intraparietal sulcus	[8A->TPO->POa]	86.85	0
	reticular thalamic nucleus	[8A->TPO->PG->V2->R]	203.19	0.33
	suprageniculate thalamic nucleus	[8A->TPO->SG]	97.05	0.33
	temporal area TAa	[8A->TPO->PG->TAa]	164.28	0.33
	temporal area TE2	[8A->TPO->PG->V2->MT->TE2]	249.60	0.33
	temporal area TEa	[8A->TPO->TEa]	86.41	0
	temporoparietal cortex	[8A->TPO->Tpt]	96.69	0.33
	visual area 1 (primary visual cortex)	[8A->TPO->PG->V1]	152.72	0.33
	visual area 2	[8A->TPO->PG->V2]	151.19	0.33
	visual area 3A	[8A->TPO->PG->V3A]	125.79	0.33
	ventral pallidum	[8A->TPO->PG->VP]	179.58	0.33
	area 23 of cortex	[8A->23]	63.43	0.33
	area 24 of cortex	[8A->24]	35.19	0
	area 46 of cortex	[8A->46]	26.39	0.33
	area 6 of cortex	[8A->TPO->PG->6]	155.75	0.33
	mediodorsal thalamic nucleus	[8A->TPO->MD]	106.64	0.33
	parietal area PL	[8A->TPO->PG->PL]	132.48	0.33
	parietal area PG	[8A->TPO->PG]	101.3	0.33
	temporal area TE, occipital part	[8A->TPO->PG->V2->TEO]	198.45	0.33
	temporal area TF	[8A->TPO->TF]	88.88	0.33
	temporal area TL	[8A->TPO->TL]	103.45	0.33
	temporal parietooccipital associated area in sts	[8A->TPO]	57.86	0
	visual area 4	[8A->TPO->PG->V4]	142.9	0.33
parietal area PG	area 23a of cortex	[PG->8A->23a]	118.97	0.67
	area 23b of cortex	[PG->TPO->23b]	100.26	0.33
	area 23c of cortex	[PG->23c]	46.78	0.33
	area 24a of cortex	[PG->8A->24a]	97.92	0.67
	area 24b of cortex	[PG->8A->24b]	96.77	0.67
	area 24c of cortex	[PG->24c]	73.44	0.33
	area 31 of cortex	[PG->31]	42.13	0.33
	area 32 of cortex	[PG->32]	94.54	0.33
	area 8B of cortex	[PG->8B]	72.58	0.33

The first two columns of this table display full names of source and target, respectively; the third column provides the optimal path between source and target (displayed in an abbreviated form), the corresponding Euclidean distance is shown in the fourth column. The fifth column introduces a measure associated with minimal projection strength (density) along the specified path; the values are distributed in a range [0, 1) and are higher for lower density projections and vice versa. Zero cost indicates that all the projections in the path are of the highest density.

A4. Data used for training the SORT model

Unweighted					Weighted				
A	**B**	**R_A**	**R_B**	**CoCoMac RC**	**A**	**B**	**R'_A**	**R'_B**	**CoCoMac RC**
B05-14	BB47-IB	0.836	0.685	S	B05-14	BB47-IB	0.91	0.815	S
B05-17	BB47-OC	0.94	0.871	I	B05-17	BB47-OC	0.979	0.949	I
B05-17	FV91-V1	0.953	0.924	I	B05-17	FV91-V1	0.976	0.967	I
B05-17	FV91-V2	0.018	0.015	D	B05-17	LV00-V1	0.995	0.933	I
B05-17	LV00-V1	0.975	0.675	I	B05-17	PHT00-V1	0.911	0.901	I
B05-17	PHT00-V1	0.793	0.62	I	B05-17	UD86-V1	0.989	0.948	I
B05-17	UD86-V1	0.949	0.692	I	B05-18	BB47-OC	0.018	0.015	D
B05-18	FV91-V2	0.503	0.49	O	B05-18	LV00-V1	0.033	0.023	D
B05-18	LV00-V1	0.069	0.056	D	B05-18	PHT00-V1	0.055	0.042	D
B05-18	PHT00-V1	0.086	0.079	D	B05-18	UD86-V1	0.007	0.005	D
B05-18	UD86-V1	0.028	0.024	D	B05-19	BB47-OA	0.293	0.21	I
B05-19	BB47-OB	0.194	0.175	D	B05-20	BB47-TE	0.69	0.257	S
B05-19	FV91-V2	0.219	0.154	D	B05-21	BB47-TE	0.664	0.442	S
B05-20	BB47-TE	0.621	0.271	S	B05-22	BB47-TA	0.752	0.726	L
B05-21	BB47-TE	0.619	0.321	S	B05-24	BB47-FL	0.123	0.073	L
B05-24	BB47-FL	0.166	0.124	L	B05-24	BB47-LA	0.26	0.027	L
B05-24	BB47-LA	0.33	0.072	L	B05-27	BB47-TF	0.088	0.052	O
B05-27	BB47-TF	0.135	0.081	O	B05-3	FV91-4	0.065	0.06	D
B05-4	BB47-FA	0.731	0.312	I	B05-4	BB47-FA	0.779	0.611	I
B05-4	FV91-4	0.717	0.706	I	B05-4	FV91-4	0.814	0.81	I
BB47-FB	B05-6	0.532	0.415	I	BB47-FB	B05-6	0.553	0.43	I
BB47-IA	B05-15	0.767	0.185	L	BB47-IA	B05-15	0.797	0.257	L
BB47-IA	B05-16	0.233	0.116	L	BB47-IA	B05-16	0.203	0.12	L
BB47-IB	B05-13	0.089	0.067	L	BB47-IB	B05-13	0.052	0.034	L
BB47-OA	B05-19	0.341	0.296	I	BB47-OB	B05-17	0.037	0.02	D
BB47-OB	B05-17	0.064	0.06	D	BB47-OB	B05-18	0.744	0.505	I
BB47-OB	B05-18	0.703	0.563	I	BB47-OB	B05-19	0.182	0.154	D
BB47-OB	FV91-V1	0.091	0.082	D	BB47-OB	FV91-V1	0.044	0.031	D

BB47-OB	FV91-V2	0.738	0.577	I	BB47-OB	LV00-V1	0.051	0.023	D
BB47-OB	LV00-V1	0.098	0.064	D	BB47-OB	PHT00-V1	0.044	0.023	D
BB47-OB	PHT00-V1	0.08	0.058	D	BB47-OB	UD86-V1	0.017	0.007	D
BB47-OB	UD86-V1	0.042	0.029	D	BB47-OC	LV00-V1	0.952	0.935	I
BB47-OC	B05-18	0.064	0.059	D	BB47-TF	B05-20	0.281	0.079	O
BB47-OC	FV91-V2	0.053	0.047	D	BB47-TF	FV91-TF	0.859	0.655	I
BB47-OC	LV00-V1	0.897	0.67	I	BB47-TF	LV00-TF	0.576	0.355	I
BB47-OC	PHT00-V1	0.759	0.641	I	BB47-TG	B05-20	0.169	0.147	O
BB47-OC	UD86-V1	0.88	0.693	I	BB47-TG	B05-21	0.591	0.314	O
BB47-TA	B05-22	0.658	0.616	S	BB47-TG	B05-22	0.237	0.083	O
BB47-TF	B05-20	0.327	0.094	O	BB47-TH	B05-20	0.164	0.017	S
BB47-TF	FV91-TF	0.785	0.547	I	FV91-1	LV00-1	0.527	0.521	I
BB47-TF	LV00-TF	0.556	0.359	I	FV91-3b	LV00-3a	0.008	0.003	D
BB47-TG	B05-20	0.215	0.18	O	FV91-46	B05-9	0.714	0.199	S
BB47-TG	B05-21	0.472	0.332	O	FV91-7a	B05-7	0.388	0.13	S
BB47-TG	B05-22	0.306	0.109	O	FV91-7a	BB47-PG	0.79	0.281	S
BB47-TH	B05-20	0.171	0.036	S	FV91-7a	LV00-7a	0.875	0.866	D
FV91-3b	LV00-3a	0.01	0.008	D	FV91-7a	LV00-7a	0.875	0.866	I
FV91-4	B05-3	0.104	0.059	D	FV91-7b	LV00-7a	0.026	0.022	D
FV91-46	B05-9	0.64	0.171	S	FV91-FST	BB47-TA	0.026	0.001	S
FV91-7a	B05-7	0.369	0.116	S	FV91-LIP	B05-7	0.41	0.032	S
FV91-7a	BB47-PG	0.714	0.24	S	FV91-LIP	BB47-PG	0.68	0.072	S
FV91-7a	LV00-7a	0.817	0.737	D	FV91-MDP	B05-7	0.803	0.064	S
FV91-7a	LV00-7a	0.817	0.737	I	FV91-MIP	B05-7	0.451	0.032	S
FV91-7b	LV00-7a	0.053	0.049	D	FV91-MSTd	B05-7	0.076	0.014	S
FV91-FST	BB47-TA	0.061	0.005	S	FV91-MSTd	BB47-PG	0.004	0.001	S
FV91-LIP	B05-7	0.417	0.035	S	FV91-MT	LV00-MT	0.897	0.523	I
FV91-LIP	BB47-PG	0.689	0.062	S	FV91-STPa	BB47-TA	0.48	0.049	S
FV91-MDP	B05-7	0.724	0.078	S	FV91-STPp	BB47-TA	0.999	0.307	S
FV91-MIP	B05-7	0.482	0.033	S	FV91-TH	BB47-TH	0.516	0.308	I
FV91-MSTd	B05-7	0.104	0.027	S	FV91-TH	LV00-TF	0.242	0.09	D
FV91-MSTd	BB47-PG	0.013	0.003	S	FV91-V1	B05-18	0.029	0.026	D
FV91-MT	LV00-MT	0.848	0.478	I	FV91-V1	BB47-OC	0.968	0.929	I

FV91-STPa	BB47-TA	0.537	0.054	S	FV91-V1	LV00-V1	0.998	0.936	I
FV91-STPp	BB47-TA	0.998	0.257	S	FV91-V1	UD86-V1	0.978	0.932	I
FV91-TF	LV00-TF	0.627	0.582	I	FV91-V2	B05-17	0.005	0.003	D
FV91-TH	BB47-TH	0.526	0.343	I	FV91-V2	B05-18	0.635	0.41	O
FV91-TH	LV00-TF	0.292	0.091	D	FV91-V2	B05-19	0.251	0.212	D
FV91-V1	B05-18	0.069	0.061	D	FV91-V2	BB47-OB	0.844	0.842	I
FV91-V1	BB47-OC	0.915	0.874	I	FV91-V2	BB47-OC	0.024	0.022	D
FV91-V1	LV00-V1	0.99	0.707	I	FV91-V2	LV00-V1	0.1	0.042	D
FV91-V1	PHT00-V1	0.808	0.653	I	FV91-V2	PHT00-V1	0.128	0.067	D
FV91-V1	UD86-V1	0.926	0.697	I	FV91-V2	UD86-V1	0.101	0.047	D
FV91-V2	LV00-V1	0.327	0.271	D	FV91-V3	BB47-OA	0.218	0.013	S
FV91-V2	PHT00-V1	0.354	0.333	D	FV91-V3A	BB47-OA	1	0.128	S
FV91-V2	UD86-V1	0.33	0.288	D	FV91-V3A	LV00-V4	0.007	0.001	D
FV91-V3	BB47-OA	0.272	0.029	S	FV91-V3A	PHT00-V2	0.012	0	D
FV91-V3A	BB47-OA	1	0.09	S	FV91-V3A	UD86-V2	0.082	0.005	D
FV91-V3A	LV00-V4	0.016	0.004	D	FV91-V3A	UD86-V3A	0.616	0.546	I
FV91-V3A	PHT00-V2	0.008	0.001	D	FV91-V3A	UD86-V4	0.019	0.002	D
FV91-V3A	UD86-V2	0.13	0.011	D	FV91-V4	B05-19	0.145	0.076	D
FV91-V3A	UD86-V3A	0.528	0.48	I	FV91-V4	BB47-OA	0.776	0.417	S
FV91-V3A	UD86-V4	0.041	0.006	D	FV91-V4	UD86-V4	0.897	0.756	I
FV91-V4	B05-19	0.231	0.098	D	FV91-VIP	B05-7	0.715	0.146	S
FV91-V4	BB47-OA	0.658	0.322	S	FV91-VP	BB47-OA	0.772	0.055	S
FV91-V4	UD86-V4	0.814	0.656	I	LV00-1	B05-2	0.029	0.004	D
FV91-VIP	B05-7	0.687	0.119	S	LV00-1	FV91-2	0.011	0.005	D
FV91-VP	BB47-OA	0.774	0.08	S	LV00-1	FV91-4	0.024	0.002	D
LV00-1	B05-2	0.046	0.01	D	LV00-1	PHT00-3a	0.045	0.032	D
LV00-1	FV91-1	0.471	0.463	I	LV00-1	PHT00-3b	0.035	0.013	D
LV00-1	FV91-2	0.017	0.01	D	LV00-1	PHT00-4(F1)	0.019	0.003	D
LV00-1	FV91-4	0.038	0.007	D	LV00-31	PHT00-31	0.711	0.527	I
LV00-1	PHT00-3a	0.059	0.044	D	LV00-3a	FV91-3a	0.632	0.522	I
LV00-1	PHT00-3b	0.038	0.025	D	LV00-3a	PHT00-3b	0.375	0.349	D
LV00-1	PHT00-4(F1)	0.029	0.011	D	LV00-3b	PHT00-4(F1)	0.057	0.028	D
LV00-31	PHT00-31	0.635	0.431	I	LV00-4	B05-3	0.105	0.053	D

LV00-3a	FV91-3a	0.578	0.495	I	LV00-4	B05-4	0.89	0.483	I
LV00-3b	PHT00-4(F1)	0.101	0.051	D	LV00-4	FV91-4	0.982	0.565	I
LV00-4	B05-3	0.181	0.049	D	LV00-4	PHT00-4(F1)	0.855	0.834	I
LV00-4	B05-4	0.813	0.392	I	LV00-LIPd	FV91-VIP	0.005	0.002	D
LV00-4	FV91-4	0.957	0.455	I	LV00-LIPv	FV91-VIP	0.776	0.477	D
LV00-LIPd	FV91-VIP	0.008	0.005	D	LV00-MDP	PHT00-MT(V5)	0.003	0	D
LV00-LIPv	FV91-VIP	0.722	0.425	D	LV00-MT	FV91-V4	0.015	0.002	D
LV00-MDP	PHT00-MT(V5)	0.006	0.001	D	LV00-MT	PHT00-MT(V5)	0.79	0.505	I
LV00-MT	FV91-V4	0.027	0.006	D	LV00-PIP	FV91-PIP	0.778	0.563	I
LV00-MT	PHT00-MT(V5)	0.794	0.316	I	LV00-TF	B05-19	0.052	0.005	D
LV00-PIP	FV91-PIP	0.74	0.479	I	LV00-TF	FV91-TF	0.721	0.652	I
LV00-TF	B05-19	0.091	0.013	D	LV00-TF	FV91-V4	0.007	0.002	D
LV00-TF	FV91-V4	0.016	0.005	D	LV00-TF	PHT00-V4V	0.342	0.189	D
LV00-TF	PHT00-V4V	0.325	0.229	D	LV00-V2d	B05-17	0.009	0.001	D
LV00-V2d	B05-17	0.024	0.008	D	LV00-V2d	BB47-OC	0.001	0	D
LV00-V2d	BB47-OC	0.004	0.001	D	LV00-V2d	FV91-V1	0.003	0.001	D
LV00-V2d	FV91-V1	0.011	0.003	D	LV00-V2d	PHT00-V1	0.004	0	D
LV00-V2d	PHT00-V1	0.009	0.002	D	LV00-V2v	B05-17	0.001	0	D
LV00-V2v	B05-17	0.006	0.003	D	LV00-V2v	BB47-OC	0.012	0.004	D
LV00-V2v	BB47-OC	0.045	0.022	D	LV00-V2v	BB47-TE	0	0	D
LV00-V2v	BB47-TE	0.001	0.001	D	LV00-V2v	FV91-V1	0.001	0.001	D
LV00-V2v	FV91-V1	0.006	0.003	D	LV00-V2v	PHT00-V1	0.032	0.012	D
LV00-V2v	PHT00-V1	0.073	0.03	D	LV00-V2v	UD86-V1	0.003	0.001	D
LV00-V2v	UD86-V1	0.014	0.005	D	LV00-V3	BB47-OB	0.099	0.007	D
LV00-V3	BB47-OB	0.161	0.019	D	LV00-V3	FV91-V2	0.007	0	D
LV00-V3	FV91-V2	0.017	0.002	D	LV00-V3	FV91-V4	0.233	0.048	D
LV00-V3	FV91-V4	0.223	0.067	D	LV00-V3	PHT00-V2	0.088	0.008	D
LV00-V3	PHT00-V2	0.139	0.018	D	LV00-V3	UD86-V2	0.53	0.053	D
LV00-V3	UD86-V2	0.464	0.063	D	LV00-V3	UD86-V4	0.02	0.003	D
LV00-V3	UD86-V4	0.04	0.01	D	LV00-V3A	FV91-V3A	0.905	0.612	I
LV00-V3A	FV91-V3A	0.814	0.48	I	LV00-V3A	PHT00-MT(V5)	0.005	0	D

LV00-V3A	PHT00-MT(V5)	0.007	0.001	D	LV00-V3A	UD86-V4	0.02	0.001	D
LV00-V3A	UD86-V4	0.041	0.004	D	LV00-V4	B05-19	0.102	0.016	D
LV00-V4	B05-19	0.122	0.041	D	LV00-V4	FV91-V4	0.925	0.807	I
LV00-V4	FV91-V4	0.87	0.687	I	LV00-V4	PHT00-V2	0.01	0.003	D
LV00-V4	PHT00-V2	0.025	0.008	D	LV00-V4	UD86-V2	0.002	0.001	D
LV00-V4	UD86-V2	0.008	0.003	D	LV00-V4	UD86-V4	0.964	0.701	I
LV00-V4	UD86-V4	0.919	0.585	I	PG91-46v	LV00-46v	0.598	0.329	I
PG91-46v	LV00-46v	0.581	0.281	I	PG91-6Ds	LV00-6Ds	0.546	0.387	I
PG91-6Ds	LV00-6Ds	0.518	0.324	I	PG91-6Vb	LV00-6Vb	0.025	0.003	I
PG91-6Vb	LV00-6Vb	0.036	0.008	I	PG91-8Ac	LV00-8Ac	0.049	0.019	I
PG91-8Ac	LV00-8Ac	0.069	0.036	I	PHT00-1	LV00-2	0.508	0.247	D
PHT00-1	LV00-2	0.405	0.234	D	PHT00-3a	LV00-2	0.002	0.001	D
PHT00-3a	LV00-2	0.003	0.002	D	PHT00-3a	LV00-3a	0.644	0.544	I
PHT00-3a	LV00-3a	0.559	0.482	I	PHT00-3a	LV00-4	0.076	0.016	D
PHT00-3a	LV00-4	0.089	0.043	D	PHT00-3b	LV00-2	0.001	0.001	D
PHT00-3b	LV00-2	0.003	0.002	D	PHT00-4(F1)	B05-3	0.323	0.089	D
PHT00-3b	LV00-3a	0.362	0.356	D	PHT00-MST	UD86-MST	0.658	0.333	I
PHT00-4(F1)	B05-3	0.279	0.075	D	PHT00-MT(V5)	FV91-V4	0.118	0.055	D
PHT00-4(F1)	LV00-4	0.751	0.748	I	PHT00-MT(V5)	LV00-V1	0.008	0.003	D
PHT00-MST	UD86-MST	0.569	0.335	I	PHT00-MT(V5)	LV00-V4	0	0	D
PHT00-MT(V5)	FV91-V4	0.156	0.085	D	PHT00-MT(V5)	UD86-V4	0.058	0.027	D
PHT00-MT(V5)	LV00-V1	0.023	0.003	D	PHT00-TFL	BB47-TF	0.036	0.003	S
PHT00-MT(V5)	LV00-V4	0.001	0.001	D	PHT00-TFM	BB47-TF	0.462	0.018	S
PHT00-MT(V5)	UD86-V4	0.092	0.04	D	PHT00-TFO	BB47-TF	0.682	0.154	S
PHT00-TFL	BB47-TF	0.047	0.007	S	PHT00-THO	BB47-TH	0.116	0.044	S
PHT00-TFM	BB47-TF	0.48	0.042	S	PHT00-V1	BB47-OC	0.921	0.867	I
PHT00-TFO	BB47-TF	0.615	0.169	S	PHT00-V1	FV91-V1	0.911	0.903	I
PHT00-THO	BB47-TH	0.158	0.071	S	PHT00-V1	LV00-V1	0.988	0.928	I
PHT00-V1	LV00-V1	0.967	0.856	I	PHT00-V1	UD86-V1	0.972	0.937	I

PHT00-V1	UD86-V1	0.934	0.871	I	PHT00-V2	B05-17	0.153	0.063	D
PHT00-V2	B05-17	0.183	0.153	D	PHT00-V2	B05-18	0.711	0.396	I
PHT00-V2	B05-18	0.66	0.472	I	PHT00-V2	BB47-OB	0.837	0.758	I
PHT00-V2	BB47-OB	0.766	0.685	I	PHT00-V2	BB47-OC	0.131	0.05	D
PHT00-V2	BB47-OC	0.172	0.133	D	PHT00-V2	FV91-V1	0.143	0.075	D
PHT00-V2	FV91-V1	0.182	0.148	D	PHT00-V2	FV91-V2	0.811	0.754	I
PHT00-V2	FV91-V2	0.743	0.519	I	PHT00-V2	LV00-V1	0.151	0.051	D
PHT00-V2	LV00-V1	0.19	0.11	D	PHT00-V2	UD86-V1	0.115	0.042	D
PHT00-V2	UD86-V1	0.152	0.093	D	PHT00-V2	UD86-V2	0.826	0.813	I
PHT00-V3V	FV91-V4	0.029	0.013	D	PHT00-V3V	FV91-V4	0.024	0.004	D
PHT00-V3V	LV00-V4	0.003	0.002	D	PHT00-V3V	LV00-V4	0.001	0.001	D
UD86-FST	PHT00-FST	0.38	0.183	I	UD86-FST	PHT00-FST	0.347	0.146	I
UD86-MT	FV91-V4	0.007	0.001	D	UD86-MT	FV91-V4	0.003	0	D
UD86-MT	LV00-MT	0.429	0.405	I	UD86-MT	LV00-MT	0.469	0.387	I
UD86-TEO	FV91-V4	0.108	0.064	D	UD86-TEO	FV91-V4	0.079	0.036	D
UD86-TEO	LV00-V4	0.082	0.062	D	UD86-TEO	LV00-V4	0.043	0.027	D
UD86-TF	BB47-OB	0.052	0.006	D	UD86-TF	BB47-OB	0.044	0.002	D
UD86-TF	BB47-OC	0.039	0.004	D	UD86-TF	BB47-OC	0.024	0.001	D
UD86-TF	BB47-TE	0.005	0.001	D	UD86-TF	BB47-TE	0.002	0	D
UD86-TF	PHT00-V2	0.079	0.01	D	UD86-TF	PHT00-V2	0.074	0.004	D
UD86-V1	LV00-V1	0.99	0.94	I	UD86-V1	BB47-OC	0.948	0.946	I
UD86-V2	B05-17	0.046	0.037	D	UD86-V1	LV00-V1	0.998	0.983	I
UD86-V2	B05-19	0.088	0.083	D	UD86-V2	B05-17	0.018	0.008	D
UD86-V2	BB47-OC	0.072	0.054	D	UD86-V2	B05-19	0.061	0.049	D
UD86-V2	FV91-V1	0.085	0.066	D	UD86-V2	BB47-OC	0.034	0.014	D
UD86-V2	FV91-V2	0.78	0.522	I	UD86-V2	FV91-V1	0.035	0.019	D
UD86-V2	LV00-V1	0.097	0.054	D	UD86-V2	FV91-V2	0.851	0.764	I
UD86-V2	PHT00-V1	0.102	0.064	D	UD86-V2	LV00-V1	0.045	0.015	D
UD86-V2	PHT00-V2	0.768	0.735	I	UD86-V2	PHT00-V1	0.069	0.027	D
UD86-V3d	PHT00-V3D	0.77	0.25	I	UD86-V3d	PHT00-V3D	0.766	0.275	I
UD86-V3v	PHT00-V3V	0.71	0.386	I	UD86-V3v	PHT00-V3V	0.791	0.462	I
UD86-V4	FV91-V2	0.034	0.013	D	UD86-V4	FV91-V2	0.011	0.005	D
UD86-V4	LV00-V2v	0.064	0.054	D	UD86-V4	LV00-V2v	0.038	0.024	D

UD86-V4	PHT00-V2	0.039	0.021	D	UD86-V4	PHT00-V2	0.016	0.008	D
UD86-VIP	BB47-PG	1	0.049	S	UD86-VIP	BB47-PG	1	0.056	S
UD86-VIP	FV91-VIP	0.807	0.215	I	UD86-VIP	FV91-VIP	0.831	0.227	I