Large scale functional MRI study on a production grid

Tristan Glatard[3,4], Remi S. Soleman[1], Dick J. Veltman[2],
Aart J. Nederveen[1], and Sílvia D. Olabarriaga[3,4]
[1] Radiology, [2] Psychiatry and [3] Clinical Epidemiology, Biostatistics and Bioinformatics
Academic Medical Center, University of Amsterdam, NL
[4] Institute of Informatics, University of Amsterdam, NL

Abstract

Functional magnetic resonance imaging (fMRI) analysis is usually carried out with standard software packages (e.g., FSL and SPM) implementing the General Linear Model (GLM) computation. Yet, the validity of an analysis may still largely depend on the parameterization of those tools, which has, however, received little attention from researchers. In this paper we study the influence of three of those parameters, namely (i) the size of the spatial smoothing kernel, (ii) the hemodynamic response function delay and (iii) the degrees of freedom of the fMRI-to-anatomical scan registration. In addition, two different values of acquisition parameters (echo times) are compared. The study is performed on a data set of 11 subjects, sweeping a significant range of parameters. It involves almost one CPU year and produces 1.4 Terabytes of data. Thanks to a grid deployment of the FSL FEAT application, this compute and data intensive problem can be handled and the execution time is reduced to less than a week. Results suggest that optimal parameter values for detecting activation in the amygdalae deviate from the default typically adopted in such studies. Moreover, robust results indicate no significant difference between brain activation maps obtained with the two echo times.

1. Introduction

Functional magnetic resonance imaging (fMRI [1]) is a noninvasive method for detecting brain activation that is now applied extensively in neuroscience, neurosurgical planning and drug research. fMRI detects changes in the oxyhaemoglobin/deoxyhaemoglobin ratio resulting from increased local perfusion in the brain following a rise in neural activity, the so-called blood oxygenation level dependent (BOLD) contrast. Functional MR data can be acquired rapidly and spatial resolution is high, so that large data sets are generated. fMRI data is usually analysed with software packages such as fMRIB Software Library (FSL) [2] and Statistical Parametric Mapping software (SPM) [3]. Although the user interface of these packages conceals much of the complexity of the image analysis process, the choice of parameters still plays an important role in fMRI analysis. For example, it has been shown that the delay of the hemodynamic response (HRF) may vary between subjects and brain regions [4], but also between groups of e.g. healthy control subjects and neuropsychiatric patients [5]. Given this variability, it would seem to be advisable to assess optimal modelling parameters for each fMRI paradigm and subject sample *a priori*. Most researchers, however, perform the analysis using standard (default) parameter settings without questioning the role of these parameters, mostly due to practical reasons. The comparison of results obtained for different parameter settings requires a large amount of computing resources for analysis and data storage, but such advanced IT infrastructures are normally not available in a typical neuroscience environment. On the other hand, open grids [6] hold the promise to provide such high capacity computational resources in a shared and distributed model, and could be used to perform such parameter studies. Although the feasibility of grids has been demonstrated for several applications in medical imaging, their adoption in practice is still challenging and reports illustrating successful stories that go beyond the demonstrator level are still scarce.

In this paper we present a practical example of a parameter study in fMRI in which we varied three different parameters in a standard data pre-processing procedure and General Linear Model (GLM) analysis. We used for this example an emotional-provoking task which is known to activate the amygdalae, a brain area mainly involved in processing of emotional reactions and memory. We also investigate the effect of one image acquisition parameter, namely the echo time. The analysis is performed on a production grid infrastructure, enabling this compute and data intensive problem to be tackled within a reasonable amount of time. This example illustrates the usefulness of such methodological experiments that employ massive computing resources to investigate the optimal parameters settings and the robustness of results obtained with fMRI.

The paper is organised as follows. Section 2 presents details of the fMRI problem and the designed parameter sweep experiment. The implementation of this experiment on a grid infrastructure is presented in Section 3. fMRI

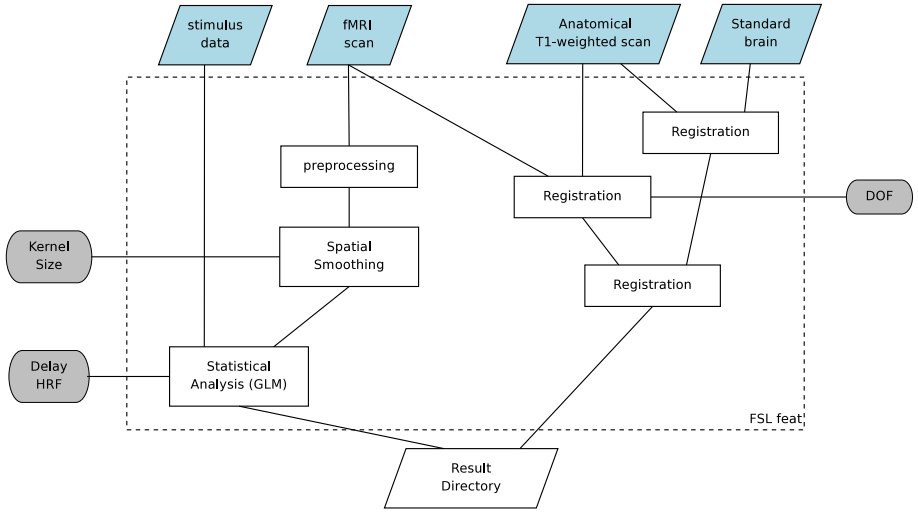

Figure 1: Simplified representation of FSL FEAT emphasising steps that are most relevant in this study. Rectangles are processing steps, diamonds are files and rounded boxes are parameters.

and application performance results are presented and discussed in Sections 4.1 and 4.2. Section 5 discusses related work and section 6 presents conclusions of this study.

2. fMRI Parameter Study

In this study the acquired data consists of a functional MRI scan containing time series of 3-D volumes, a high resolution MRI T1-weighted scan used for anatomical reference, and text files containing timing information about the adopted stimulus. The data is analysed with FSL, version 3.3, using the fMRI Expert Analysis Tool (FEAT), version 5.63. Each fMRI data set is first individually submitted to first-level analysis, which calculates brain activation maps as a result of a pipeline of image analysis steps. The most relevant steps for this study are illustrated in figure 1 and summarized below. After preprocessing (including motion correction, brain extraction and slice-timing), the fMRI scans are spatially smoothed using a kernel of a given size size. Then, the GLM analysis is performed, which creates a model to fit the fMRI scan data with respect to the timing of the stimulation paradigm employed during the scanning session. It is assumed that a good fit between the model and the data means that changes in BOLD-signal are causally related to the stimulation paradigm [1]. The output of an fMRI analysis is a brain activation map containing standardised activation probabilities (Z-scores). Spatial normalisation to the standard brain is performed in two steps: the first aligns the fMRI data to the anatomical scan, and the second aligns the anatomical scan to the reference brain. The image analysis steps are performed based on actual values of parameters, some of which represented in figure 1.

2.1. Parameters of Interest

The first part of the parameter sweep concerns the normalisation to a standard brain. To compare individual scans at a group level it is necessary to align them to a predefined reference brain. In this study we adopted the standard MNI152[1] brain distributed with FSL, and the FMRIB's Linear Image Registration Tool (FLIRT) [7]. The number of **degrees of freedom** (parameter D) in FLIRT indicates the type of transformations attempted to align two scans: 12 (affine), 9 (translation, rotation and anisotropic scaling), 7 (translation, rotation and isotropic scaling), 6 (rigid) and 3 (only translations). This parameter is typically adjusted to a lower level to prevent failures (e.g. complete flip of an image), or to reduce computing time. A wrong choice of D might result in a poor registration and different brain areas will be compared with each other. As shown in figure 1, the normalisation to the standard brain is performed through three different registrations. First, the high-resolution T1-weighted scan is mapped to the standard brain and to the fMRI scan using two independent registration procedures. Then, those two results are combined to produce the fMRI-to-standard-brain mapping. In this study we investigated the influence of the number of degrees of freedom to perform the first step only (fMRI-to-T1 registration), the fMRI-to-standard-brain registration being always performed using 12 degrees of freedom.

The second part of the parameter sweep concerns spatial smoothing. In many cases a smoothing kernel $S = 6$ mm (full width at half maximum) is used for a typical voxelsize of approximately 2-3 mm. Reducing the spatial resolution of fMRI data, i.e., smoothing with larger

[1]Montreal Neurological Institute and Hospital, *http://www.mni.mcgill.ca*

kernels, increases the signal to noise ratio (SNR), which improves the sensitivity. On the other hand, when using large **smoothing kernel sizes (S)** the spatial resolution may be lost without gain regarding to sensitivity. In this study we investigate this trade-off by varying the size of the spatial smoothing kernel.

The third part of the parameter sweep concerns the **delay of hemodynamic response (H)**. Functional MRI provides an indirect measure of brain activity, since fMRI detects changes in blood oxygenation level resulting from increased local perfusion following a rise in neural activity. Therefore, fMRI analysis needs to correct for this delayed hemodynamic response. This is performed by convolving the modeled activity pattern in the brain with a hemodynamic response function (HRF). In general a hemodynamic delay $H = 6$ s is used, although there is evidence that this value may vary within and between subjects [8, 9].

Apart from the effects of parameter manipulation, we are also interested in comparing MRI sequences. An important parameter in an MRI sequence is the **echo time (T)**, indicating the time window between the transmission of a radiofrequency pulse and the signal acquisition in fMRI. In general, the usage of a sequence with a shorter echo time will result in higher signal and smaller susceptibility artifacts in the images, whereas the contrast between high and low brain activity states will tend to decrease [10]. However, it is difficult to predict from theoretical considerations which acquisition scheme should be used to obtain optimal results. In addition we hypothesize that the differences observed in the brain maps obtained with these two protocols might also depend on the parameter values used for fMRI analysis.

2.2. Experiment Design

Data Acquisition. To assess those questions we used data of healthy volunteers acquired on a Philips 3.0 Tesla Intera scanner during an event-related task paradigm: viewing of emotional pictures, International Affective Picture System (IAPS) [11]. This task was chosen because it has been proven a highly robust paradigm, producing reliable activation of the bilateral anterior medial temporal lobe, i.e., the amygdala region [12]. The task was presented twice during a session with different echo times: 28 ms and 35 ms. The order of presentation was counterbalanced between subjects. Whole brain imaging was performed by scanning axial slices with an in-plane resolution of 1.9 mm and a slice thickness of 3.0 mm. For $T = 28$ ms and $T = 35$ ms we used a repetition time (time between successive volumes) of 2.7 s and 3.1 s respectively. Except for the echo times and the repetition times, all scanning parameters were kept identical. In total 22 fMRI scans $F_{i,T}$ were acquired for 11 subjects $P_i, i \in [1, 11]$ with echo times $T \in \{28, 35\}$.

	CPU time	Input size (MB)	Results size (MB)
Indiv. analysis	49 min	104.5	150
Group analysis	8.6 min	1650	30.5
Group difference	22.8 min	3300	55.5

Table 1: Characteristics of the individual and group analyses. CPU times have been measured on a Dual-Core AMD Opteron 2613.427 MHz with 3.5 GB of RAM. Groups summarise results of 11 individuals.

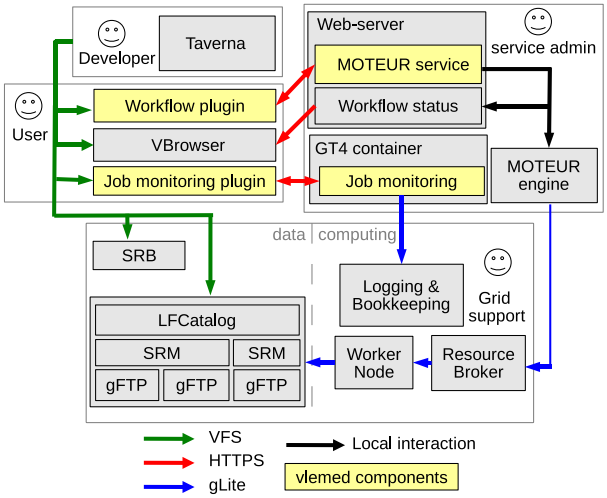

Figure 2: Overview of system used for the experiments. More details in [17, 18]

Data Analysis. Each scan $F_{i,T}$ was analysed individually using FSL FEAT first-level analysis with varying parameters: HRF delay $H \in \{2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5\}$, smoothing kernel size $S \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, and degrees of freedom for registration $D \in \{3, 6, 7, 9, 12\}$. Results of this analysis consist of activation maps $\alpha_{i,T}(H, S, D)$ obtained with each combination of parameter values (H, S, D). After all the individual activation maps were computed, an average activation map was calculated for groups of results using FSL FEAT high-level analysis. Results obtained for all P_i with identical (T, H, S, D) are averaged, generating group activation maps denoted as $\gamma_T(H, S, D)$. Finally, the results obtained with the two echo times were compared using FSL FEAT group difference analysis in a similar way. The generated difference maps are denoted $\Delta(H, S, D)$.

Comparison of Results. Results obtained with different parameters are compared based on a region of interest with robust activation for the adopted stimulus. IAPS is a picture set containing emotion-provoking pictures, both positive (e.g. erotic scenes) and negative (e.g., mutilations, attack scenes), which has often been used in emotion research. Previous studies have shown that viewing IAPS pictures vs. baseline activates regions within the

brain that are involved in emotional processing such as the amygdalae [13, 14]. The amygdalae are almond-shaped groups of nuclei within the left and right medial temporal lobe which are one of the key elements of emotion processing. Therefore, the mean Z-scores in the amygdalae were used as reference to compare activation maps obtained with different parameter settings. The amygdalae were located with a predefined anatomical atlas (AAL, [15, 16]) on the MNI152 template (see figure 7-right). The created mask was applied to the group activation maps $\gamma_{28}(H,S,D)$, $\gamma_{35}(H,S,D)$ and $\Delta(H,S,D)$, to extract the Z-scores within this region of interest. The mean Z-score indicating activation in the extracted region was calculated and respectively denoted $\mu_{28}(H,S,D)$, $\mu_{35}(H,S,D)$ and $\mu_D(H,S,D)$.

3. Grid implementation

The designed experiment involves large computation effort. Table 1 summarizes the data size and computing cost as measured on a machine representative of a user's desktop. This benchmark is used subsequently in this paper as a reference to quantify the gain provided by the grid execution, in particular CPU time and data size estimations. The individual analysis is the most compute-intensive one due to pre-processing, registration and linear model computation. It takes about 45 min. The group analysis is the most data intensive one, since it needs to access all the individual analysis results for averaging. Parameter sweeps are costly; in particular, when the parameter space has several dimensions (three in this study), widening the range of one of them swiftly increases the size of the computing problem. As a side effect, storing the produced data is often not manageable without special infrastructure. Based on the benchmark in table 1, the total estimated resources for this study add up to almost one CPU-year and 1.4 Terabytes of data. Obviously, this experiment could not have been performed without a high performance infrastructure.

Production grids have been designed to support such computing and data needs, but they are still seldom used by medical imaging researchers for such studies. Roadblocks need to be identified and carefully targeted by the grid community with new developments. To do that, the Dutch Virtual Laboratory for e-Science (VL-e[2]) set up a tight interaction between domain scientists, grid developers and grid service providers. The project gives access to a Proof-of-Concept (VL-e PoC) grid infrastructure that is part of EGEE[3]. In spite of its name, the resources and management of the VL-e PoC grid largely overlap with the EGEE production infrastructure, supporting daily computing for scientific research. Note that this production grid lacks advanced monitoring features, as available in experimental platforms used in computer science (e.g. DAS3 or Grid5000 [19]). In particular, information about the network traffic at the time of an experiment, the number of jobs in the system or the number of CPUs used by the application are difficult to obtain.

3.1. Setup description

Figure 2 shows the grid setup that we used for this experiment. The user front-end is provided by the Virtual Resource Browser (VBrowser) [20]. MOTEUR [21] is adopted to execute workflows on the grid. The workflow manager is configured to resubmit failed jobs until they successfully complete. Plugins have been developed to enable workflow enactment with MOTEUR and the Generic Application Service Wrapper (GASW) [22] from the VBrowser. Grid job submission and monitoring is transparently handled by MOTEUR using the gLite middleware. The various storage resources are accessed via the homogeneous Virtual File System interface provided by the VL-e software Toolkit[4].

The Scufl workflow language from the Taverna workbench [23] is used to describe workflows. Even if this application does not exhibit strong workflow requirements (like complex data dependencies between components), using such a language is useful to describe the parameter sweep in an elegant way. In short, Scufl supports lists as inputs, and the workflow is iterated for each list element. Operators specify how list elements should be combined as workflow inputs. The "cross product" $A \otimes B$ indicates that the workflow is iterated for all combinations of elements $a_i \in A$ and $b_j \in B$, and the "dot product" $A \oplus B$ indicates that the workflow is iterated for pairs (a_i, b_i). Using such a generic workflow framework instead of ad-hoc scripts enabled autonomous usage of the grid by end-users, which is an important aim of our research.

An extensive description of this grid setup is available in [17, 18].

3.2. Application Porting

As part of the VL-e software distribution[5], the FSL package is pre-installed in all computing nodes, therefore scripts can be submitted directly as grid jobs to perform FSL FEAT or other image analysis tasks. The image analysis tasks were wrapped as services, workflows were composed to describe the parameter sweep experiment, and the workflows were enacted on the grid. Two workflow descriptions were used to implement the parameter sweep experiment, one for the individual analyses and one for the group analyses.

The **individual analysis workflow** takes 10 inputs and produces a single output (see figure 3). The performed task consists of FSL FEAT first-level analysis and some data logistics (download/upload data). Six of the inputs correspond to the fMRI scan $F_{i,T}$ (100 MB) and

[2] http://www.vl-e.nl
[3] http://www.eu.eege.org
[4] http://www.vl-e.nl/vbrowser
[5] http://poc.vl-e.nl/

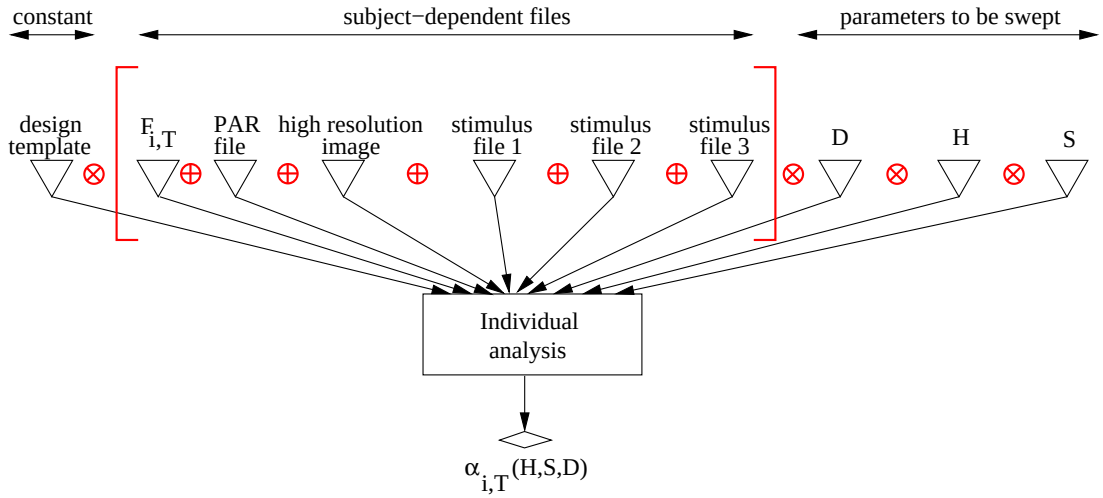

Figure 3: Individual analysis workflow described in Scufl. Triangles denote inputs, diamonds indicate outputs and boxes correspond to processors (here a script that executes FSL feat for individual analysis). Iteration strategies are drawn in red ($\otimes$ is the cross-product and $\oplus$ is the dot product). Scufl iteration strategies allows for easily distinguishing sweeping parameters from data-related inputs.

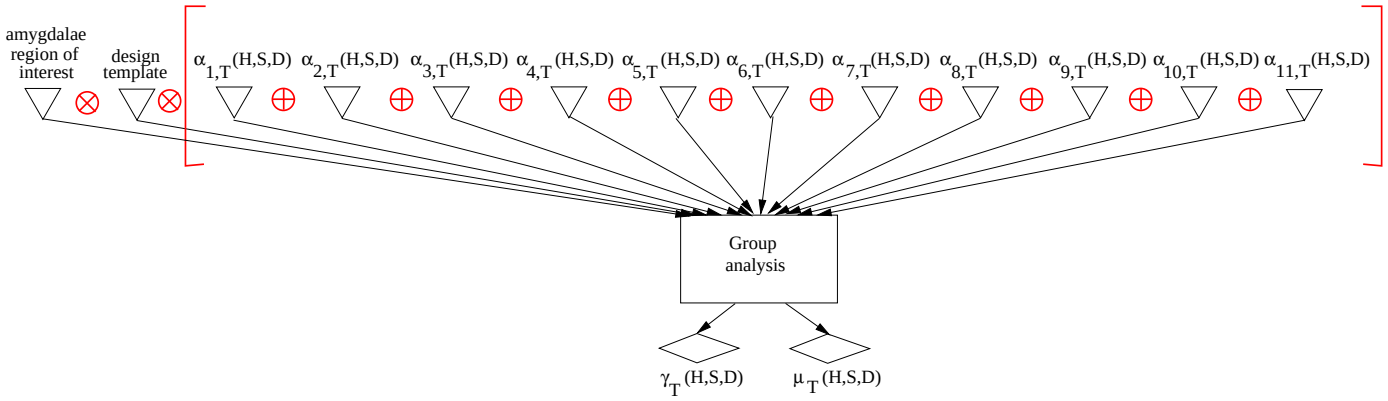

Figure 4: Group analysis workflow described in Scufl.

other subject-related data: MRI acquisition parameters (1.5 MB), the high resolution anatomical (T1-weighted) scan (1.5 MB) and 3 files with stimulus timing information (500 bytes each). Three other inputs correspond to the parameters (H, S, D) to sweep upon. The last input (design file) contains all the parameters for FSL FEAT that remain constant in all analyses. Each individual analysis produces a single directory with all the output files (150 MB when compressed), one of them being the brain activation map $\alpha_{i,T}(H, S, D)$ that is subsequently used in group analyses. In this experiment, inputs are given as lists of values associated with different scans, subjects, and parameters. The iteration strategies of Scufl allow to express the sweep on individual analysis tasks in a simple way: the 6 first data-related inputs are combined with the dot product operator, and the other parameters are swept with the cross product.

The **group analysis workflow** used to calculate average activation has 13 inputs and two outputs – see figure 4. Besides running the FSL FEAT high-level analysis, this workflow also calculates the mean value in the amygdalae using with the FSL AVWMATHS utility. Eleven inputs indicate directories containing pre-computed individual activation maps $\alpha_{i,T}(H, S, D), i \in [1, 11]$. These inputs are provided by lists, and the dot product is used to combine them, guaranteeing that the activation maps averaged during the workflow iteration were obtained with the same parameter settings (T, H, S, D). The remaining inputs are constant in this experiment, namely the design file with parameters for FSL FEAT, and the mask indicating the voxels in the amygdalae. The group analysis produces 2 outputs: a directory with results (30 MB when compressed), including the group activation map $\gamma_T(H, S, D)$, and a file containing the mean Z-score $\mu_T(H, S, D)$ in the

amygdalae region.

Similarly to the above, a workflow with 23 inputs and 2 outputs is used for **group difference analysis** to compare results obtained with different echo times T. Twenty two inputs indicate directories with pre-computed individual analyses corresponding to two groups fMRI data $F_{i,28}, F_{i,35}, i \in [1..11]$ acquired with different echo times. The outputs include the difference activation maps for these two groups $\Delta(H, S, D)$ and the mean Z-scores in the amygdalae $\mu_\Delta(H, S, D)$.

3.3. Application Deployment

The grid infrastructure of the Dutch VL-e project was used to run the experiment. Those resources are deployed, maintained and administrated with the goal to offer production support to the applications. Consequently, experiment conditions can hardly be controlled or monitored by the applications. At the time the experiment was performed our `vlemed` Virtual Organisation (VO) had access to 687 CPUs spread over 4 sites through 8 batch queues federated by the gLite middleware. This production grid is shared among several VOs, therefore there is no control on the number of available CPUs at a given instant. Four Storage Elements (SE) are accessible through a Storage Resource Manager (SRM[6]) interface and files can be organised in a single Logical File Catalog (LFC) independently from their physical location. In addition, a gridFTP and a Storage Resource Broker (SRB[7]) servers were also available at the time.

Porting the image analysis application to the grid was easily done using GASW, which integrates executables in Scufl workflows via a command-line descriptor and the Web Service Description Language (WDSL). The FEAT script was wrapped as shipped by the FSL distribution, with additional logistics steps to download and upload the data. Workflows are executed on this infrastructure using the software architecture depicted on figure 2.

4. Results and Discussion

4.1. fMRI results

Figure 5 illustrates the mean Z-score indicating activation within the amygdalae using different degrees of freedom D for the fMRI-to-anatomical scan registration. No significant activation differences were observed, a pattern that is also observed for other smoothing kernel sizes. In particular, note that a registration with $D = 3$ (translation only) produced similar results as obtained with $D = 9$ and $D = 12$. Given that both scans belong to the same subject and have been acquired during a single scanning session, only minor misalignments are expected, which could explain why only translations seem sufficient to capture

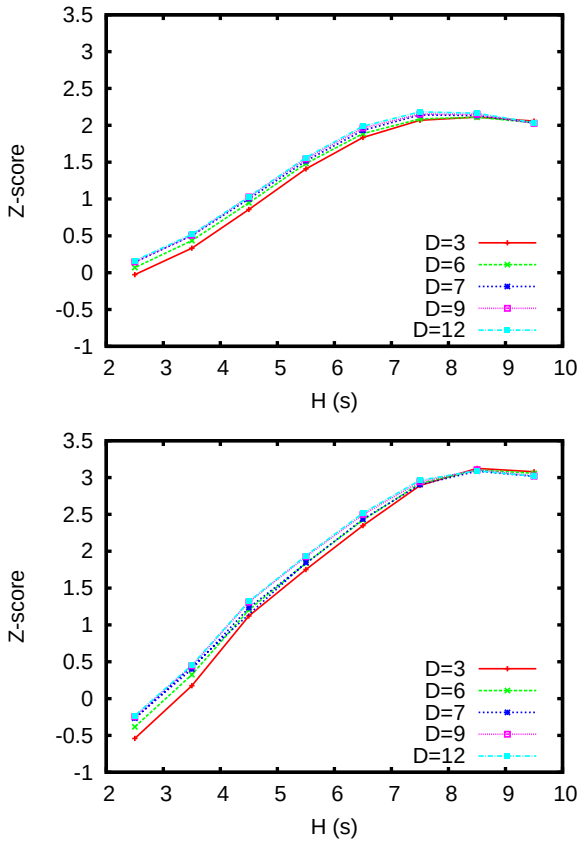

Figure 5: Influence of degrees of freedom D in the mean activation in the amygdalae as a function of delay for $T = 28$ ms using smoothing kernels $S = 5$ mm (top) and $S = 11$ mm (bottom).

them. The main contribution to the normalisation process is likely to be the anatomical-to-standard-brain registration, the second step in the registration process, which was fixed to 12 degrees of freedom in this study. Based on those results, a further analysis of the complete registration process will be performed in the future.

Figure 6-top shows the mean activation within the amygdalae obtained with different spatial smoothing kernel sizes and HRF delays. Note that a peak for the Z-score is observed for a delay of $H = 8.5$ s, which differs from the value adopted by default in FSL feat ($H = 6$ s). Furthermore, note that the mean Z-scores increase for larger spatial smoothing kernels. A visual inspection of the results (figure 7) suggests that with $S = 12$ mm spatial resolution is still acceptable and that even larger smoothing kernels might further increase sensitivity. However, whereas this increase in sensitivity is likely to result from blurring residual anatomical differences still present after warping individual brains to anatomical standard space, larger smoothing kernels will inevitably lead to loss of spatial resolution. Further investigation is needed to better understand the effect of spatial smoothing in data analysis, for example the role of anatomical variability in large subject groups or in clinical subjects.

[6] `http://sdm.lbl.gov/srm-wg/`
[7] `http://www.sdsc.edu/srb/`

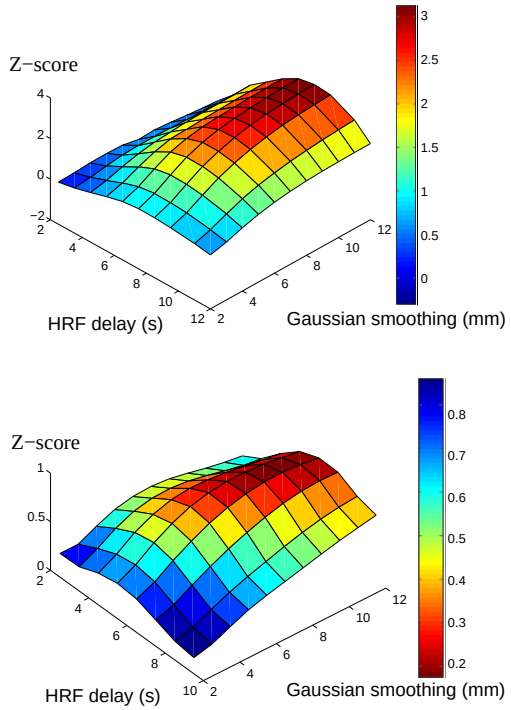

Figure 6: Results obtained for group analyses for varying smoothing kernel sizes S and HRF delays H. Top: Mean activation (Z-scores) in the amygdalae for $T = 28$ ms, $D = 12$. Bottom: Mean differences (Z-scores) of amygdalae activation between brain maps calculated with $T = 28$ ms and $T = 35$ ms, for all combinations of parameters S and H.

Our second research question concerns the effects of manipulating echo time (T). Pairwise T-tests were used to compare results obtained with both echo times. No significant differences in activation between the scan sequences could be detected by the comparison of these two echo times. As shown in figure 6-bottom, none of the parameter combinations cross the threshold of significance (we used a Z-value of 2.33, which corresponds to a p-value of 0.01, as threshold to distinguish noise from the signal [24]). These results are in agreement with the general understanding that BOLD sensitivity does not change a lot in this range of echo times, although to our knowledge there is no conclusive literature on this issue.

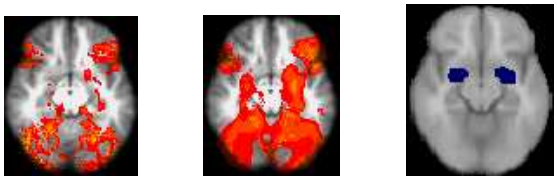

Figure 7: Impact of the smoothing kernel size for $H = 2.5$ s and $D = 12$. Left and Centre: slice of the brain activation map respectively for $S = 5$ mm and $S = 12$ mm smoothing. Right: slice of the amygdalae region of interest (in blue) overlaid on the template brain.

4.2. Grid performance

Table 2 presents a summary of performance results of the individual and the group analyses as measured during the execution of this parameter study on the VL-e grid. A total of 9680 individual analyses, 880 group analyses and 440 group differences was computed.

Individual analyses were submitted in 4 batches of about 2500 jobs each to avoid infrastructure flooding. Compared to local execution (as estimated from benchmark of table 1), the grid implementation provided a global speed-up of 115, which enabled us to compute the individual analyses in less than 3 days (68.7 hours). Note that this is a compute-intensive experiment, therefore the speed-up provides a rough indication of the number of CPUs that were used concurrently. Instead of reporting detailed efficient metrics, our main interest here is to quantify the gain provided by the grid compared to the common practice of using a single desktop computer. We see the grid as a single machine providing computing power and data storage. Note also the small job failure rate of 0.5% measured for this experiment.

Although pleasantly parallel, group analyses faced quite poor performance (global speed-up of 2.9). One of the reasons is that these are data intensive tasks, since all individual analyses are downloaded to the worker node before executing FSL FEAT. This is due to both the algorithm design of fMRI analysis (FSL FEAT in our case) and the setup of the experiment account for the slow pace of group analyses. Group analyses require results from several individual analyses to compute an average activation map. They are data-intensive tasks by nature, regardless the particular fMRI analysis implementation.Even if each analysis only transfers about 1 GB, concurrently initiating hundreds of such data transfers rapidly reduces the data servers throughput. Moreover, one of the Storage Elements limited the number of concurrent connections, causing jobs to fail and partially explaining the high 38.93% global failure ratio. This problem was even larger in the case of group difference analyses, where each of the 440 jobs concurrently attempted to download more than 3 GB. The global failure rate here was 83%. In spite of these problems, the experiment could be carried out in a reasonable time in the end, which could not have been possible without adopting a grid implementation. Still, the grid deployment could be further optimized. For instance, among the files produced by the individual analyses, only a subset is required for the group analyses, while the whole archive was transferred here. Or individual activation maps could be progressively downloaded as they become available, avoiding burst download requests to the SE.

All in all, the failure ratios reported here look good compared to other medical experiments conducted on EGEE. For instance, authors report in [25] a 44% success ratio on a metagenomics experiment conducted within the `biomed` EGEE VO. Two main reasons can explain such a differ-

	$\#P_i$	#T	#S	#D	#H	# Analyses	# Submit Jobs	Failure (%)	Elapsed (hours)	CPU (days)*	Data (TB)*	Speed -up*
Individual Analyses												
batch 1	11	1	5	5	8	2200	2200	0.00	14.9	74.9	0.31	120.5
batch 2	11	1	6	5	8	2640	2642	0.08	11.6	89.8	0.38	186.6
batch 3	11	1	6	5	8	2640	2687	1.75	32	89.8	0.38	67.38
batch 4	11	1	5	5	8	2200	2203	0.14	10.2	74.9	0.31	176.8
total	11	2	11	5	8	9680	9732	0.53	68.7	329.4	1.38	115
Group Analyses											(MB)	
batch 1		1	6	5	8	240	401	40.15	8.0	1.4	7.1	4.3
batch 2		1	6	5	8	240	240	0.00	9.5	1.4	7.1	3.6
batch 3		1	5	5	8	200	200	0.00	14.9	1.2	6	1.9
batch 4		1	5	5	8	200	600	66.67	11.3	1.2	6	2.5
total		2	11	5	8	880	1441	38.93	43.7	5.2	26.2	2.9
Group Difference Analyses											(MB)	
batch 1			11	5	8	440	2650	83.40	44.3	7	23.8	3.8

Table 2: Performance results for individual, group and difference analyses for each job batch: number of subjects, echo times, smoothing sizes S, degrees-of-freedom D, HRF delays H, and successful analyses; number of submitted jobs, job failure rate ((#submitted - #successful) / #submitted), total elapsed time from first submitted to last completed job, total CPU time, amount of produced data and speed-up of the experiment (CPU/elapsed). Values estimated from the benchmark of table 1 are denoted with *; other values correspond to actual measurements.

ence. First, our experiment is set up using a very limited number of sites (we used 4, whereas EGEE VOs like `biomed`, use hundreds) with strong support team on every grid site. Second, it has been carefully tuned in order to avoid some of the problems encountered in [25]. For instance, the total workload has been split into batches of reasonable size to avoid proxy expiration and software was pre-installed on the worker nodes to avoid runtime installation problems.

5. Related work

To our knowledge, this is the first study to investigate the effects of varying multiple parameters during spatial preprocessing and statistical modelling of functional MRI data. Previous studies have addressed single parameters only, for example adding temporal and dispersion derivatives to the canonical hemodynamic response function to account for minor time shifts of the BOLD response [26], or adjusting spatial smoothing size to correct for between-scanner differences in multicenter studies [27].

Using grid technology to enable large-scale fMRI experiments has been envisaged several times, with the FSL toolkit (e.g., in the LONI pipeline environment [28]) or with the SPM (e.g. in [29]). Yet, to our knowledge, previous attempts differ from the fMRI experiment presented here due to the scale (much smaller experiments have been performed) and the type of infrastructure adopted (usually a dedicated infrastructure, whereas we adopted an open production grid with all its pros-and-cons).

A couple of other initiatives have used high-performance computing to meet real-time clinical constraints in fMRI [30]. The scope of these works is slightly different though, since they mainly aim at optimizing the fMRI analysis algorithms for speed-up (latency), whereas we are exploiting data parallelism on massive data (throughput).

Workflow technology is now increasingly popular in neuroimaging [31]. Some systems focus on interoperability and data sharing [32] to automate fMRI analysis for clinical use (i.e. standardization and streamlining of image acquisition, processing, post-processing and results publication). Here workflows were adopted due to the parameter-sweep support in our experiment, similarly to [33].

EGEE is used by applications from several domains, including biomedical [34]. The experiment presented here is definitely smaller in terms of CPU time and number of jobs (table 3 of [34] reports 80 years of CPU time and about 70,000 submitted jobs). The output data size, however, is comparable, but the infrastructure we used was significantly smaller (58 sites are used in [34]). Moreover, we put much effort in using a grid setup that is now autonomously adopted by end-user.

6. Conclusion

In this paper, the benefit of performing parameter sweeps for methodological fMRI studies has been established by assessing amygdalae activation for different analysis parameter choices. Detecting activation in this brain area is in general cumbersome, emphasising the need of a good choice of parameters to obtain reliable results. Although

initial and with limited scope, the results of this experiment provide us means to perform more reliable data analysis for this study. For the number of degrees of freedom to register the fMRI data to the anatomical scan with FSL FLIRT, we found no significant influence in the average activation measured at the amigdala. For the kernel size for spatial smoothing, the obtained results are inconclusive, requiring further analysis. The findings for the HRF delay to measure activation at the amigdala indicate that the optimal value (8.5s) deviates from the default parameters that are typically used in this type of analysis in FSL FEAT. In the future we plan to perform this analysis for different fMRI paradigms and different brain areas and expect to find different optimal parameter values. Finally, the results show no significant difference in the measured average activation at the amigdala for data acquired with two fMRI sequences (TE=28ms and TE=35ms). This conclusion does not depend on the parameters used for the analysis. These findings correspond to expectations, however to our knowledge there is not conclusive literature on the subject.

Such an experiment could not have been done without a grid: the infrastructure enabled it within a reasonable time and highly simplified data storage management. Since the implementation adopts true grid technologies, i.e. using gsi authentication/access control, grid-wide scheduling among several sites and federation of distributed storage, the application is scalable, enabling the users to repeat such experiments easily. Moreover, the general set-up can be adapted for similar applications, requiring only the deployment of the services shown on figure 2 and the composition of new workflows.

Yet, this setup is still far from ideal. Data transfers, rather than computation, seem to be the limiting factor now. Strategies to reduce the data bottlenecks should be considered, such as improved data staging, data distribution or even replication among several Storage Elements. Moreover, failure management is still a challenging issue in production grids, with the consequence that either expert intervention or insane amounts of job resubmissions are needed. Anyways, users are now quite autonomous with our setup, which enables sustainable developments of this study, motivated by the results reported here.

7. Acknowledgements

We thank P.T. de Boer and K. Boulebiar for the development on the VBrowser front-end and plug-ins. We also warmly thank all the people operating behind mailing lists grid.support@{nikhef,sara}.nl. This work was carried out in the VL-e project, supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science (OC&W) and is part of the ICT innovation program of the Ministry of Economic Affairs (EZ).

References

[1] S. Smith, Overview of fMRI analysis, The British Journal of Radiology 77 (2004) S167–S175.

[2] S. Smith, et al., Advances in functional and structural MR image analysis and implementation as FSL, NeuroImage 23 (1) (2004) S208–S219.

[3] K. Friston, Statistical Parametric Mapping and Other Analysis of Functional Imaging Data, in: Brain Mapping: The Methods, Academic Press, 363–385, 1996.

[4] H. D.A., O. J.M., D. M., Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses, Neuroimage 21 (4) (2004) 1639–51.

[5] R. S.A., G. R., S. C.J., B. F., S. P., Delayed rather than decreased BOLD response as a marker for early Alzheimer's disease., Neuroimage 26 (4) (2005) 1078–85.

[6] I. Foster, C. Kesselman, S. Tuecke, The Anatomy of the Grid: Enabling Scalable Virtual Organizations, International Journal of Supercomputer Applications 15 (2001) 2001.

[7] M. Jenkinson, S. Smith, A global optimisation method for robust affine registration of brain images, Medical Image Analysis 2 (1) (2001) 143–156.

[8] M. Woolrich, et al., The variability of human BOLD hemodynamic responses, NeuroImage 21 (2004) 1748–1761.

[9] G. Aguirre, et al., The variability of human BOLD hemodynamic responses, NeuroImage 8 (4) (1998) 360–369.

[10] M. Gorno-Tempini, et al., Echo Time Dependence of BOLD Contrast and Susceptibility Artifacts, NeuroImage 15 (1) (2002) 136–142.

[11] P. Lang, et al., International affective picture system (IAPS): Technical manual and affective ratings, Tech. Rep., University of Florida, Center for Research in Psychophysiology, Gainesville, 1999.

[12] H. A.R., M. V.S., T. A., F. F. W. D.R., Neocortical modulation of the amygdala response to fearful stimuli., Biol Psychiatry 53 (6) (2003) 494.

[13] H. Breiter, et al., Response and Habituation of the Human Amygdala during Visual Processing of Facial Expression, Neuron 17 (5) (1996) 875–887.

[14] I. Liberzon, et al., Extended amygdala and emotional salience: a PET activation study of positive and negative affect, Neuropsychopharmacology 28 (4) (2003) 726–33.

[15] N. Tzourio-Mazoyer, et al., Automated anatomical labelling of activations in SMP using a macroscopic anatomical parcellation of the MNI MRI single subject brain, NeuroImage 15 (1) (2002) 273–289.

[16] C. Holmes, et al., Enhancement of MR images using registration for signal averaging, Journal of Computer Assisted Tomography 22 (2) (1998) 324–333.

[17] S. Olabarriaga, T. Glatard, K. Boulebiar, P. T. de Boer, From 'low-hanging' to 'user-ready': initial steps into a healthgrid, in: HealthGrid'08, Chicago, 70–79, 2008.

[18] T. Glatard, et al., Workflow integration in VL-e medical, in: 21st IEEE International Symposium on Computer-Based Medical Systems (CBMS'08), Jyväskylä, Finland, 144–146, 2008.

[19] F. Cappello, H. Bal, Towards an International Computer Science Grid, in: 7th IEEE Int. Symp. on Cluster Computing and the Grid (CCGrid'07), Rio de Janeiro, Brazil, 3–12, 2007.

[20] S. Olabarriaga, P. T. de Boer, K. Maheshwari, A. Belloum, J. Snel, A. Nederveen, M. Bouwhuis, Virtual Lab for fMRI: Bridging the Usability Gap, in: 2nd IEEE Conference on e-Science and Grid Computing (e-Science06), IEEE, Amsterdam, 2006.

[21] T. Glatard, J. Montagnat, D. Lingrand, X. Pennec, Flexible and efficient workflow deployment of data-intensive applications on grids with MOTEUR, International Journal of High Performance Computing and Applications (IJHPCA) 22 (3) (2008) 347–360.

[22] T. Glatard, et al., A Service-Oriented Architecture enabling dynamic services grouping for optimizing distributed workflows execution, Future Generation Computer Systems 24 (7) (2008) 720–730.

[23] T. Oinn, et al., A tool for the composition and enactment of bioinformatics workflows, Bioinformatics journal 17 (20) (2004) 3045–3054.
[24] K. Worsley, et al., A three-dimensional statistical analysis for CBF activation studies in human brain, Journal of Cerebral Blood Flow and Metabolism 12 (1992) 900–918.
[25] G. Aparicio, et al., A Highly Optimized Grid Deployment: the Metagenomic Analysis Example, in: Global Healthgrid: e-Science Meets Biomedical INformatics (Healthgrid'08), Healthgrid, IOS Press, Chicago, USA, 105 –115, 2008.
[26] R. Henson, C. Price, M. Rugg, R. Turner, K. Friston, Detecting Latency Differences in Event-Related BOLD Responses: Application to Words versus Nonwords and Initial versus Repeated Face Presentations, NeuroImage 15 (1) (2002) 83–97.
[27] L. Friedman, H. Stern, G. Brown, D. Mathalon, J. Turner, G. Glover, R. Gollub, J. Lauriello, K. Lim, T. Cannon, D. Greve, H. Bockholt, A. Belger, B. Mueller, M. Doty, J. He, W. Wells, P. Smyth, S. Pieper, S. Kim, M. Kubicki, M. Vangel, S. Potkin, Test-retest and between-site reliability in a multicenter fMRI study, Human Brain Mapping 8 (29) (2008) 958–72.
[28] D. Rex, J. Ma, A. W. Toga, The LONI Pipeline Processing Environment, NeuroImage 3 (19) (2003) 1033–1048.
[29] S. Bagnasco, F. Beltrame, B. Canesi, I. Castiglioni, P. Cerello, S. Cheran, M. Gilardi, E. Lopez Torres, E. Molinari, A. Schenone, L. Torterolo, Early Diagnosis of Alzheimer's Disease Using a Grid Implementation of Statistical Parametric Mapping Analysis, in: HealthGrid, Valencia, 2006.
[30] E. Bagarinao, T. Nakai, Y. Tanaka, Real-Time Functional MRI: Development and Emerging Applications, Magnetic Resonance in Medical Sciences 5 (3) (2006) 157–165.
[31] K. Fissell, Workflow-Based Approaches to Neuroimaging Analysis, Humana Press, 235–266, 2008.
[32] S. Erberich, M. Bhandekar, A. Chervenak, C. Kesselman, M. Nelson, funcLAB/G, NeuroImage 37 (1) (2007) S135–S143.
[33] W. Sudholt, K. K. Baldridge, D. Abramson, C. Enticott, S. Garic, C. Kondric, D. Nguyen, Application of grid computing to parameter sweeps and optimizations in molecular modeling, Future Generation Computer Systems 21 (1) (2005) 27–35.
[34] N. Jacq, J. Salzeman, F. Jacq, Y. Legré, E. Medernach, J. Montagnat, J. Maass, M. Reichstadt, H. Schwichtenberg, M. Sridhar, V. Kasam, M. Zimmermann, M. Hofmann, V. Breton, Grid-enabled Virtual Screening against malaria, Journal of Grid Computing (JGC) 6 (1) (2008) 29–43.