

Multiple Vergleiche & Effekte mehrerer Kontraste

Falk Eippert

20. September 2011

SPM 2011 – 14. Kurs zur funktionellen Bildgebung

FMRIB

University of Oxford



Überblick

Beurteilung der statistischen Karten

- Einzeltest vs. Multiple Tests
- Multiple Tests und Bonferroni-Korrektur
- Bonferroni-Korrektur und fMRI
- Korrektur in SPM: FWE (Bonferroni & Gaussian Random Fields) & FDR
- Small Volume Correction

Effekte mehrerer Kontraste

- Inklusive Maskierung
- Exklusive Maskierung
- Konjunktion

1. Teil: Beurteilung der statistischen Karten

Einzeltest vs. Multiple Tests

Einzeltest

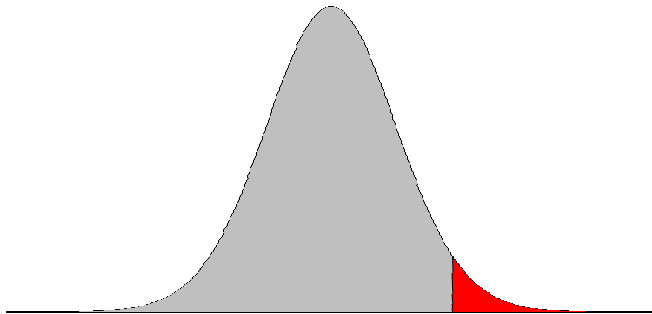
Ein statistischer Wert, der die Stärke des Effekts beschreibt (T, Z, F, χ^2)

Eine Nullhypothese, gegen die getestet wird

Es besteht **ein Mal** die Möglichkeit, die Nullhypothese zu verwerfen

$$T(40) = 1.69, p < 0.05$$

=> bei einem Test würde man ein falsch positives Testergebnis mit einer Wahrscheinlichkeit von 0.05 (5%) erwarten



Multiple Tests in fMRI

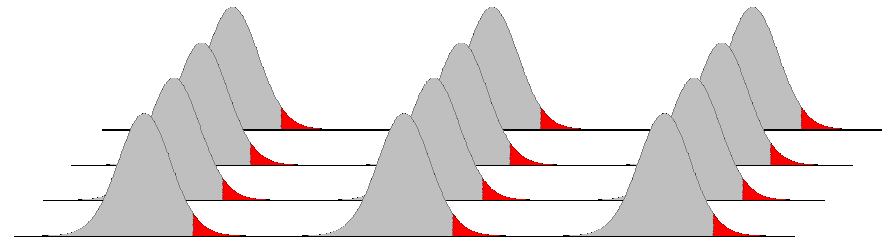
Tausende statistischer Werte, da Statistik in jedem Voxel einzeln berechnet wird

Eine Nullhypothese, gegen die getestet wird

Es besteht **mehrere tausend Mal** die Möglichkeit, die Nullhypothese zu verwerfen

$$T(40) = 1.69, p < 0.05 \text{ nur für dieses Voxel}$$

=> bei (64 x 64 x 40) 163840 Voxeln würde man 8192 falsch positive Testergebnisse erwarten



Ziel: Wahrscheinlichkeit, die Nullhypothese fälschlicherweise abzulehnen, soll 5% **bezogen auf das ganze Volumen** betragen

Multiple Tests und Bonferroni-Korrektur

Eine Irrtumswahrscheinlichkeit von $\alpha = 0.05$ pro Einzeltest ist nicht ausreichend, um für die Familie aller Tests eine Irrtumswahrscheinlichkeit von $\alpha = 0.05$ zu garantieren

Um die Irrtumswahrscheinlichkeit für die Familie aller Tests auf $\alpha = 0.05$ festzulegen und unabhängig von der Anzahl m durchgeführter Tests zu machen, wird folgende Korrektur benötigt:

$$\alpha' = \alpha / m$$

Unterschreitet mindestens ein Test diese Irrtumswahrscheinlichkeit α' , so wird die globale Nullhypothese auf einem α -Niveau von 0.05 verworfen

Annahme: Unabhängigkeit der einzelnen Tests

Bonferroni-Korrektur geeignet für fMRI?

Bonferroni-Korrektur und fMRI

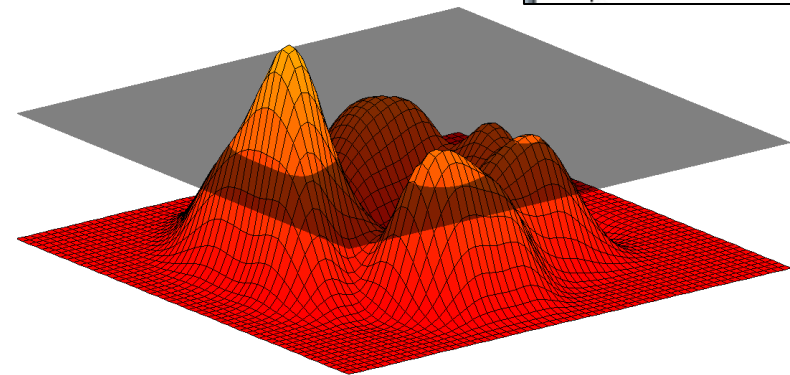
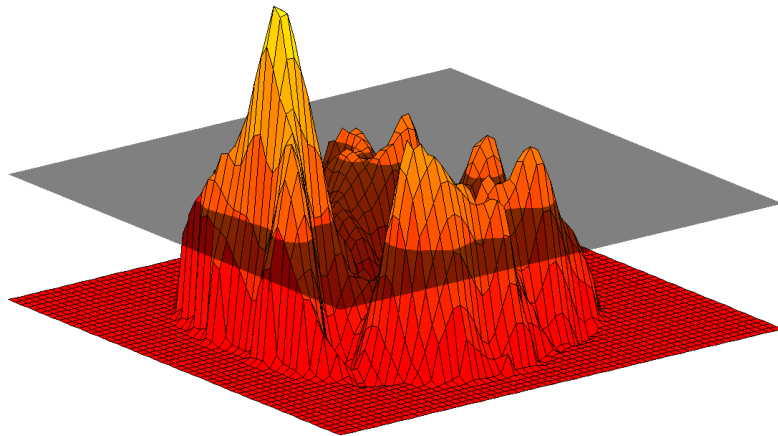
fMRI Daten sind räumlich korreliert => Abhängigkeitsstruktur

- Räumliche Ausbreitung der Aktivierung
- Nicht modellierte physiologische Einflüsse
- Vorverarbeitung: Interpolation und Glättung

=> Anzahl unabhängiger Datenpunkte < Anzahl der Voxel

=> **Bonferroni-Korrektur zu konservativ**

```
Plane 68/69 , block  
Plane 69/69 , block  
Spatial non-sphericity  
Spatial non-sphericity  
Saving results  
Completed
```



Glätte der statistischen Karte beeinflusst Größe und Anzahl der Aktivierungen

Korrektur für multiple Vergleiche in SPM

Zwei verschiedene Korrekturalgorithmen

Family Wise Error Rate (FWE)

Signifikanzkorrektur bezieht sich auf das Verhältnis von falsch abgelehnten Tests zu **allen durchgeführten** Tests

=> 5% Wahrscheinlichkeit, dass es ein falsch positives Ergebnis gibt

Entweder über Bonferroni oder Theorie der Gaussian Random Fields (GRF)

GRF schätzt die Glätte der Karte und passt Schwellen an, optimal bei großer Anzahl an Freiheitsgraden und hoher Smoothness der Karte

Bonferroni meistens zu konservativ, aber bei geringer Glätte der Karte manchmal liberaler als GRF

False Discovery Rate (FDR)

Signifikanzkorrektur bezieht sich auf das Verhältnis von falsch abgelehnten Tests zu **allen abgelehnten** Tests

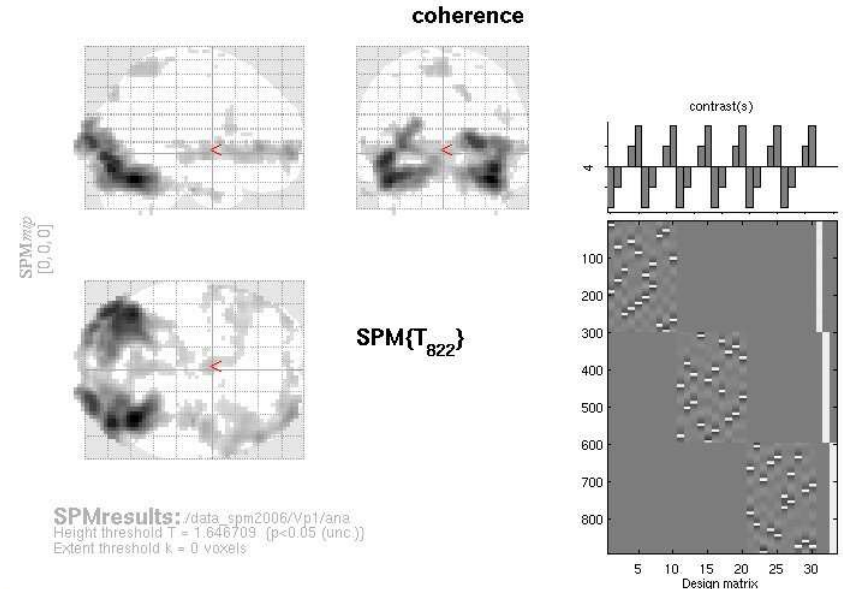
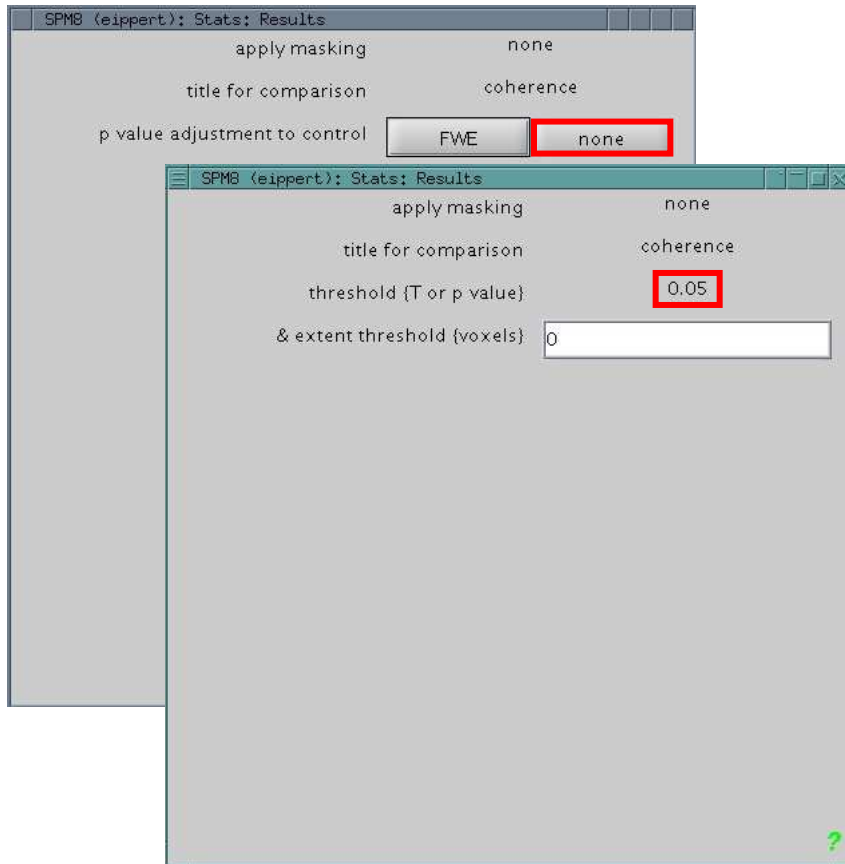
=> im Durchschnitt sind 5% der signifikanten Tests falsch Positive

Achtung: unterschiedliche Formen der FDR-Korrektur in SPM5 und SPM8

SPM5: voxel-wise FDR, **SPM8:** topological FDR (voxel-wise kann aktiviert werden über Änderung von `spm_defaults: defaults.stats.topo.fdr` von 1 auf 0 setzen)

FDR ist in vielen Fällen sensitiver als FWE (topological FDR konservativer als voxel-wise FDR)

Unkorrigierte p-Werte



Statistics: p-values adjusted for search volume

set-level		cluster-level				peak-level					
p	c	$p_{FWE-corr}$	$q_{FDR-corr}$	k_E	p_{uncorr}	$p_{FWE-corr}$	$q_{FDR-corr}$	T	$(Z_{\bar{e}})$	p_{uncorr}	mm mm mm
1.000	27	0.000	0.000	3448	0.000	0.007	0.013	5.01	4.97	0.000	36 -60 -18
		0.048	0.030			0.048	0.030	4.56	4.53	0.000	39 -78 -12
		0.052	0.030			0.052	0.030	4.54	4.51	0.000	-42 -63 -21
		0.392	0.135	809	0.010	1.000	1.000	2.82	2.82	0.002	33 63 0
		1.000	1.000	2.66	2.66	1.000	1.000	2.66	2.66	0.004	51 36 6
		1.000	1.000	2.39	2.38	1.000	1.000	2.39	2.38	0.009	45 -6 0
		1.000	0.948	40	0.538	1.000	1.000	2.48	2.47	0.007	39 -15 -27
		1.000	0.948	25	0.635	1.000	1.000	2.22	2.22	0.013	18 -78 45
		0.979	0.697	335	0.077	1.000	1.000	2.20	2.20	0.014	-3 -3 3
		1.000	1.000	2.18	2.18	1.000	1.000	2.18	2.18	0.015	-54 39 0
		1.000	1.000	2.16	2.16	1.000	1.000	2.16	2.16	0.016	-39 60 0
		1.000	0.948	147	0.228	1.000	1.000	2.19	2.19	0.014	0 -57 66
		1.000	1.000	2.17	2.16	1.000	1.000	2.17	2.16	0.015	-9 -48 66
		1.000	1.000	2.12	2.12	1.000	1.000	2.12	2.12	0.017	-9 -60 66
		1.000	0.948	16	0.714	1.000	1.000	2.09	2.09	0.018	51 3 -33
		1.000	0.948	20	0.677	1.000	1.000	2.06	2.07	0.019	-60 3 -30
		1.000	1.000	1.74	1.75	1.000	1.000	1.74	1.75	0.040	-51 -6 -30
		1.000	0.948	23	0.651	1.000	1.000	1.99	1.99	0.023	-39 15 21
		1.000	0.948	36	0.561	1.000	1.000	1.96	1.97	0.025	-9 -21 6
		1.000	1.000	1.79	1.80	1.000	1.000	1.79	1.80	0.036	-21 -33 6
		1.000	0.948	4	0.875	1.000	1.000	1.84	1.85	0.032	-39 -48 -42
		1.000	0.948	13	0.746	1.000	1.000	1.84	1.84	0.033	-3 -36 -18
		1.000	0.948	17	0.704	1.000	1.000	1.79	1.80	0.036	48 12 27
		1.000	1.000	1.72	1.73	1.000	1.000	1.72	1.73	0.042	45 15 36
		1.000	0.948	3	0.896	1.000	1.000	1.79	1.79	0.036	-15 -6 78
		1.000	0.948	1	0.948	1.000	1.000	1.77	1.78	0.038	27 -27 -24
		1.000	0.948	2	0.919	1.000	1.000	1.76	1.77	0.038	0 -21 -18

peak-level					mm mm mm
$p_{FWE-corr}$	$q_{FDR-corr}$	T	$(Z_{\bar{e}})$	p_{uncorr}	
0.007	0.013	5.01	4.97	0.000	36 -60 -18
0.048	0.030	4.56	4.53	0.000	39 -78 -12
0.052	0.030	4.54	4.51	0.000	-42 -63 -21
1.000	1.000	2.82	2.82	0.002	33 63 0
1.000	1.000	2.66	2.66	0.004	51 36 6
1.000	1.000	2.39	2.38	0.009	45 -6 0

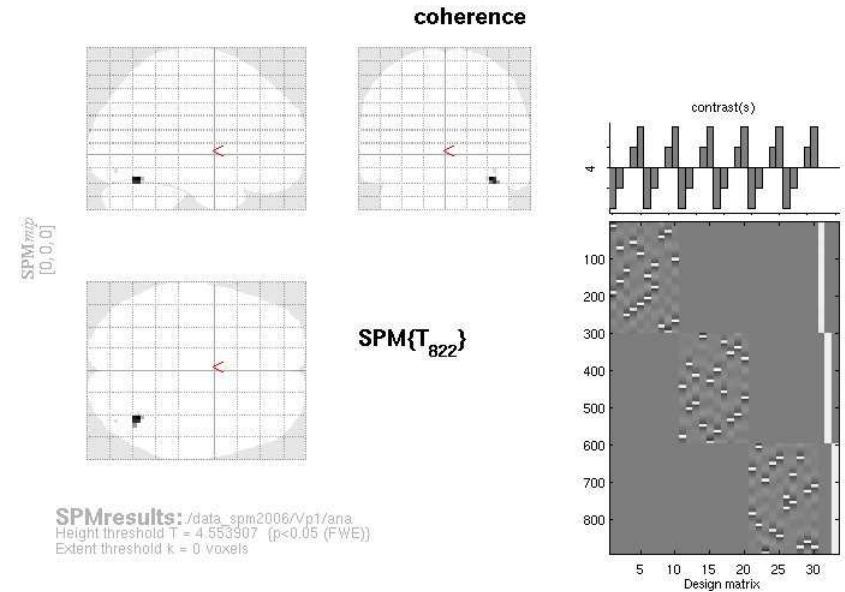
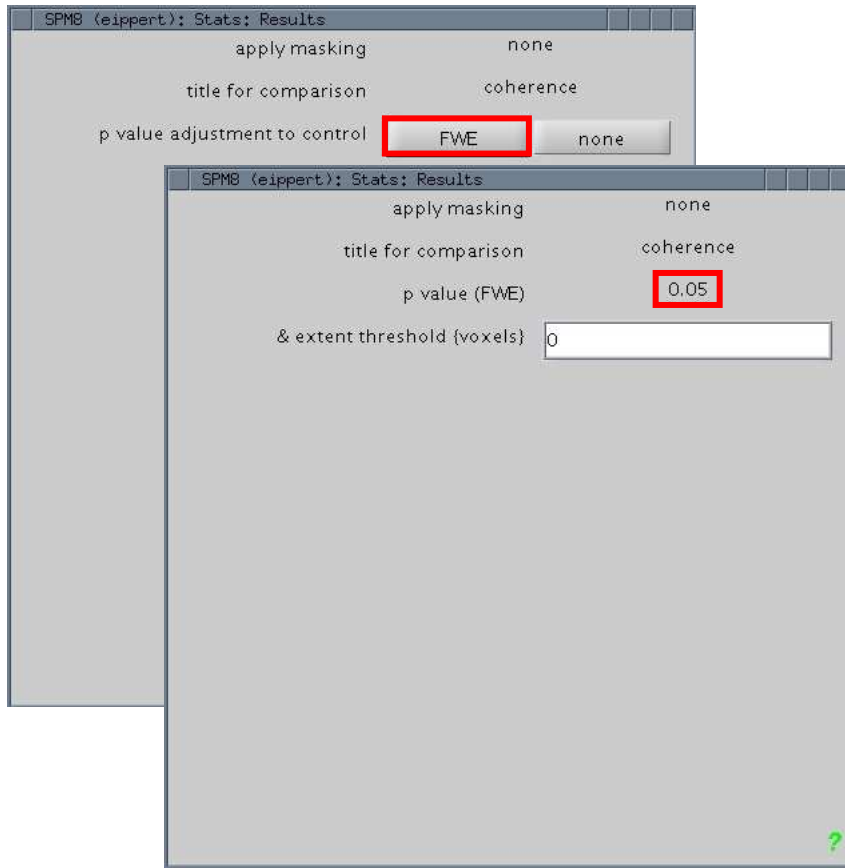
table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 1.65, p = 0.049 (1.000)
 Extent threshold: k = 0 voxels, p = 1.000 (1.000)
 Expected voxels per cluster, <k> = 108.822
 Expected number of clusters, <c> = 49.81
 WEp: 4.554, FDRp: 4.543, FWEc: 3448, FDRc: 3448

Degrees of freedom = (1.0, 822.0)
 FWHM = 16.4 16.0 11.6 mm mm mm; 5.5 5.3 3.9 (voxels)
 Volume: 1745253 = 64639 voxels = 517.3 resels
 Voxel size: 3.0 3.0 3.0 mm mm mm; (resel = 112.50 voxels)

Page 1

FWE Korrektur



Statistics: *p-values adjusted for search volume*

set-level		cluster-level				peak-level					mm mm mm		
p	c	$p_{FWE-corr}$	$q_{FDR-corr}$	k_E	p_{uncorr}	$p_{FWE-corr}$	$q_{FDR-corr}$	T	$(Z_{\bar{c}})$	p_{uncorr}			
0.001	2	0.004	0.168	12	0.084	0.007	0.287	5.01	4.97	0.000	36	-60	-18
		0.048	0.962	1	0.623	0.048	0.962	4.56	4.53	0.000	39	-78	-12

- Whole-brain FWE Korrektur ist ein sehr konservatives und nicht sehr sensibles Verfahren

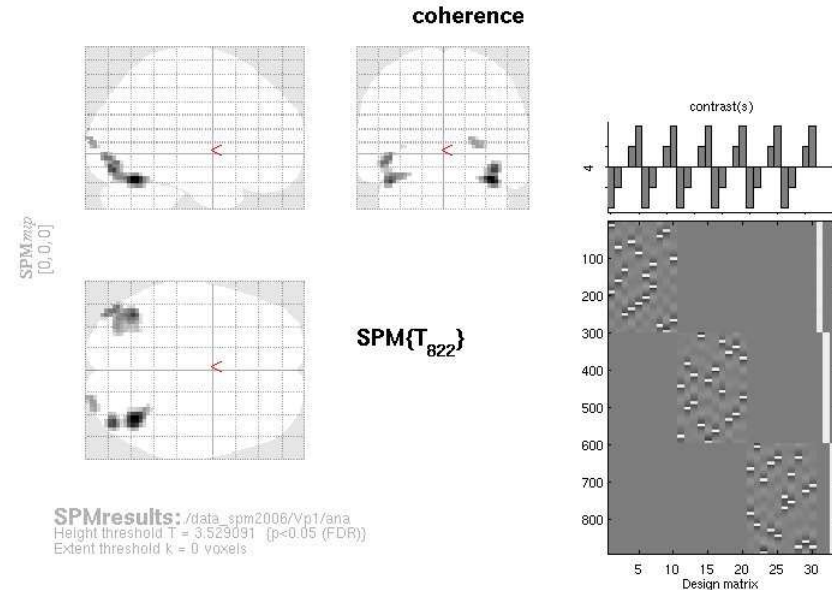
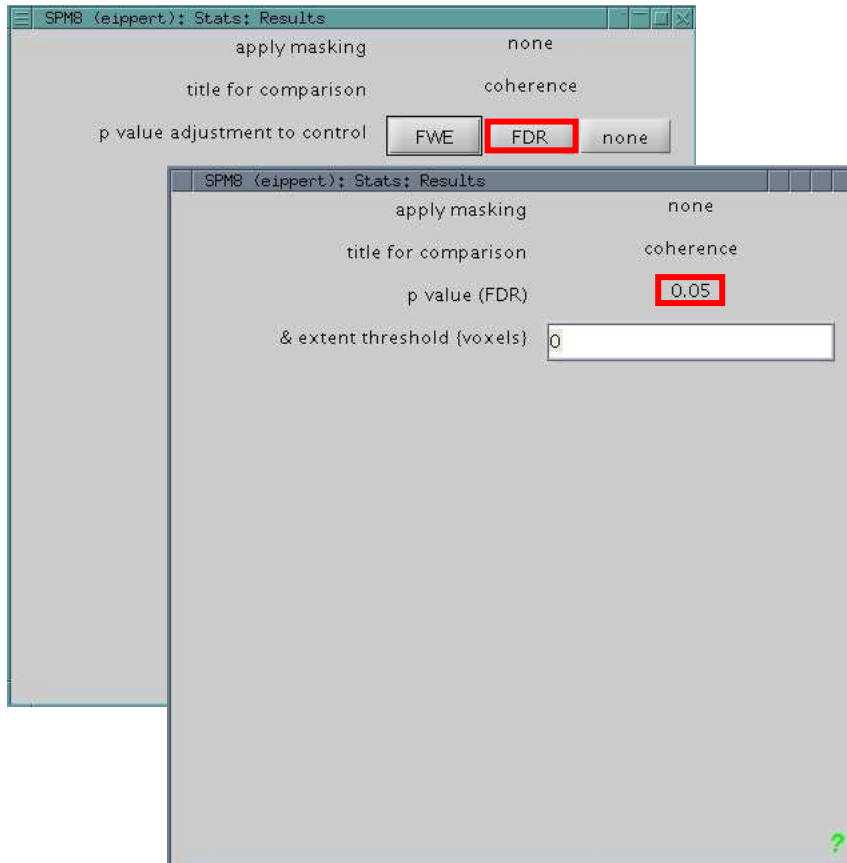
peak-level					mm mm mm		
$p_{FWE-corr}$	$q_{FDR-corr}$	T	$(Z_{\bar{c}})$	p_{uncorr}			
0.007	0.287	5.01	4.97	0.000	36	-60	-18
0.048	0.962	4.56	4.53	0.000	39	-78	-12

table shows 3 local maxima more than 8.0mm apart

Height threshold: $T = 4.55$, $p = 0.000$ (0.050)
 Extent threshold: $k = 0$ voxels, $p = 1.000$ (0.050)
 Expected voxels per cluster, $\langle k \rangle = 4.092$
 Expected number of clusters, $\langle c \rangle = 0.05$
 FWEp: 4.554, FDRp: Inf, FWEc: 1, FDRc: Inf

Degrees of freedom = [1, 0, 822, 0]
 FWHM = 16.4 16.0 11.6 mm mm mm; 5.5 5.3 3.9 (voxels)
 Volume: 1745253 = 64639 voxels = 517.3 resels
 Voxel size: 3.0 3.0 3.0 mm mm mm; (resel = 112.50 voxels)

FDR Korrektur (voxel-wise)



SPMresults: /data_spm2006/Vp1/ana
 Height threshold T = 3.529091 [p<0.05 (FDR)]
 Extent threshold k = 0 voxels.

Statistics: p-values adjusted for search volume

set-level		cluster-level				peak-level				mm mm mm	
p	c	$p_{FWE-corr}$	$q_{FDR-corr}$	k_E	p_{uncorr}	$p_{FWE-corr}$	$q_{FDR-corr}$	T	(Z_{\pm})	p_{uncorr}	
0.111	4	0.002		121	0.001	0.007	0.007	5.01	4.97	0.000	36 -60 -18
		0.001		139	0.001	0.048	0.011	4.56	4.53	0.000	39 -78 -12
						0.052	0.011	4.54	4.51	0.000	-42 -63 -21
						0.152	0.022	4.25	4.23	0.000	-45 -78 -6
		0.180		22	0.109	0.298	0.025	4.04	4.02	0.000	-33 -75 -15
		0.686		2	0.639	0.359	0.027	3.97	3.95	0.000	24 -96 9
						0.797	0.046	3.57	3.56	0.000	-33 -93 6

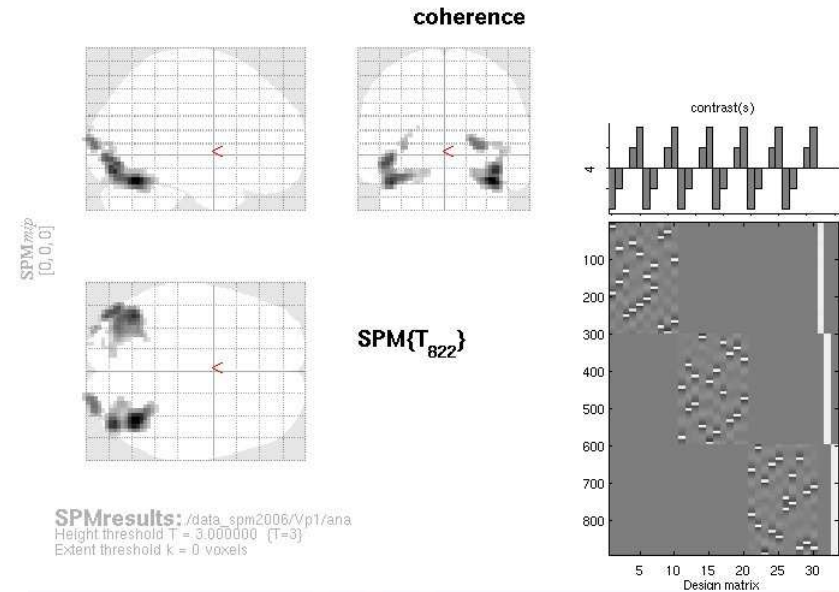
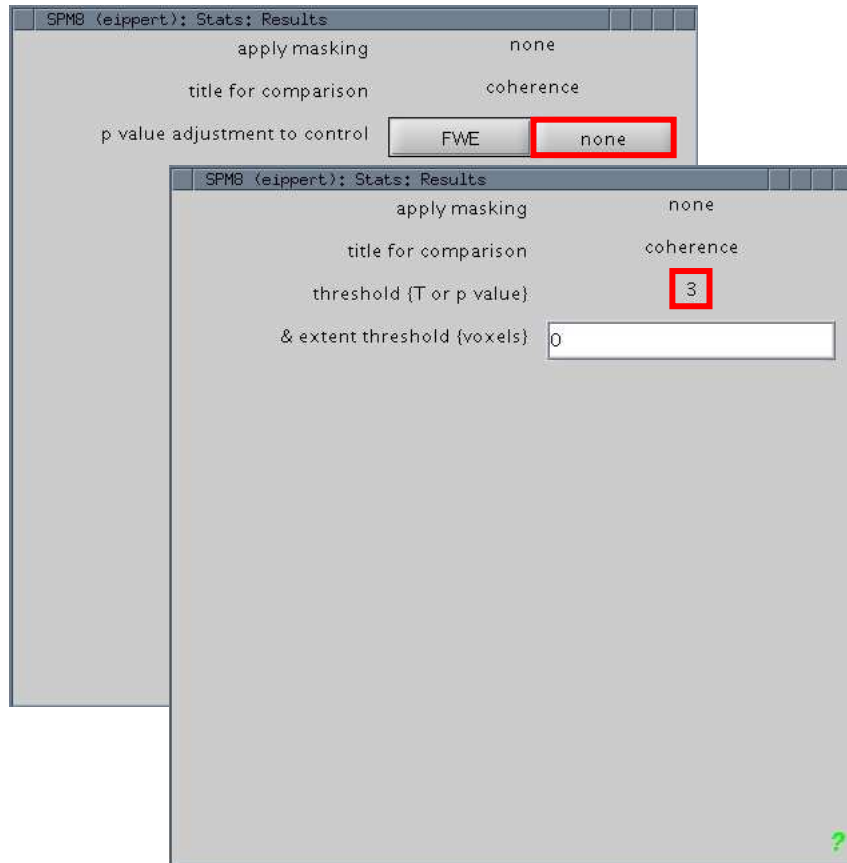
- nur erreichbar durch Veränderung von spm_defaults.m
- Standard ist in SPM8 die topologische FDR-Korrektur

peak-level					mm mm mm	
$p_{FWE-corr}$	$q_{FDR-corr}$	T	(Z_{\pm})	p_{uncorr}		
0.007	0.007	5.01	4.97	0.000	36	-60 -18
0.048	0.011	4.56	4.53	0.000	39	-78 -12
0.052	0.011	4.54	4.51	0.000	-42	-63 -21
0.152	0.022	4.25	4.23	0.000	-45	-78 -6
0.298	0.025	4.04	4.02	0.000	-33	-75 -15

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 3.53, p = 0.000 (0.837) Degrees of freedom = [1 0, 822, 0]
 Extent threshold: k = 0 voxels, p = 1.000 (0.837) FWHM = 16.4 16.0 11.6 mm mm mm; 5.5 5.3 3.9 (voxels)
 Expected voxels per cluster, <k> = 8.877 Volume: 1745253 = 64639 voxels = 517.3 resels
 Expected number of clusters, <c> = 1.81 Voxel size: 3.0 3.0 3.0 mm mm mm; (resel = 112.50 voxels)
 FWEp: 4.554, FDRp: 5.011, FWEc: 121, FDRc: 121

FDR Korrektur (topological)



peak-level					mm mm mm		
$p_{\text{FWE-corr}}$	$q_{\text{FDR-corr}}$	T	$(z_{\text{=}})$	p_{uncorr}			
0.007	0.015	5.01	4.97	0.000	36	-60	-18
0.048	0.037	4.56	4.53	0.000	39	-78	-12
0.359	0.120	3.97	3.95	0.000	24	-96	9
0.052	0.037	4.54	4.51	0.000	-42	-63	-21
0.152	0.076	4.25	4.23	0.000	-45	-78	-6
0.298	0.120	4.04	4.02	0.000	-33	-75	-15
0.797	0.365	3.57	3.56	0.000	-33	-93	6
0.998	0.851	3.07	3.06	0.001	-21	-99	21

- Abhängigkeit von „feature-inducing threshold“
- Wird vom Benutzer gesetzt (z.B. „none“ und $T = 3$)

Small Volume Correction

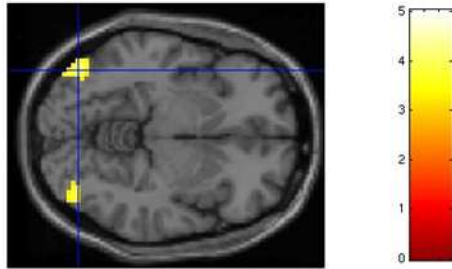
Was?

- A-priori Hypothese in Bezug auf eine Region
- Einschränkung des Suchvolumens auf einen bestimmten Bereich (VOI, volume of interest) des Bildes
- Anzahl der durchgeführten Tests, für die korrigiert werden muss, verringert sich => höhere Sensitivität

Wie?

- Form: Kugel (optimal), Box, Bild (z.B. selbst kreierte Maske des LOC)
- Größe: sollte das Volumen der Zielstruktur approximieren
- Algorithmus: Bonferroni, GRF, FDR

Small Volume Correction



Statistics: *p-values adjusted for search volume*

peak-level					mm mm mm
$p_{\text{FWE-corr}}$	$q_{\text{FDR-corr}}$	T	$(Z_{\text{=}})$	p_{uncorr}	
0.007	0.018	5.01	4.97	0.000	36 -60 -18
0.048	0.044	4.56	4.53	0.000	39 -78 -12
0.888	0.535	3.46	3.45	0.000	24 -72 -15
0.052	0.044	4.54	4.51	0.000	-42 -63 -21
0.152	0.090	4.25	4.23	0.000	-45 -78 -6

Statistics: *search volume: 50.0mm sphere at [-45,-78,-6]*

peak-level					mm mm mm
$p_{\text{FWE-corr}}$	$q_{\text{FDR-corr}}$	T	$(Z_{\text{=}})$	p_{uncorr}	
0.017	0.079	4.54	4.51	0.000	-42 -63 -21
0.052	0.079	4.25	4.23	0.000	-45 -78 -6

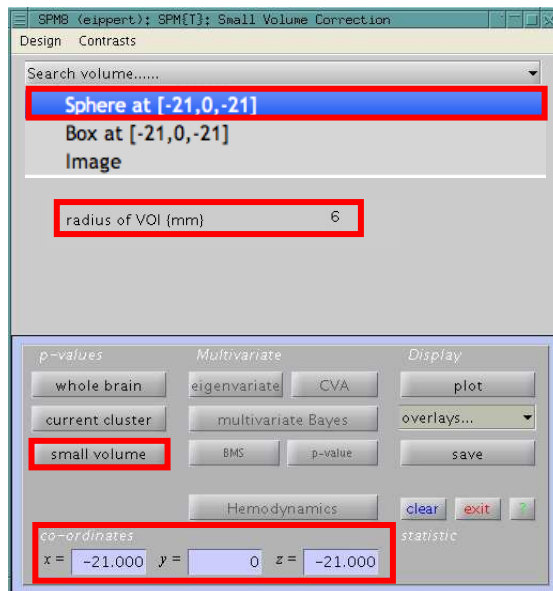
Statistics: *search volume: 25.0mm sphere at [-45,-78,-6]*

peak-level					mm mm mm
$p_{\text{FWE-corr}}$	$q_{\text{FDR-corr}}$	T	$(Z_{\text{=}})$	p_{uncorr}	
0.003	0.064	4.54	4.51	0.000	-42 -63 -21
0.008	0.064	4.25	4.23	0.000	-45 -78 -6

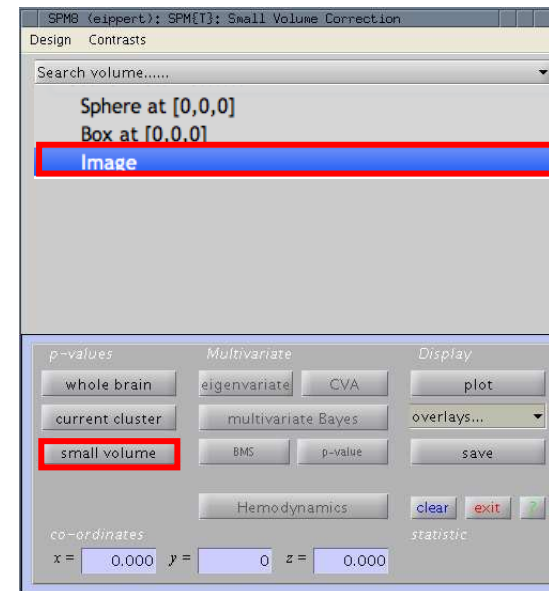
Small Volume Correction: Möglichkeiten

- Sphere / Box um Koordinaten, die von der eigenen Studie unabhängig sind (z.B. aus früheren Studien oder Meta-Analysen)
- Sphere / Box um Koordinaten, die zwar aus der eigenen Studie kommen, aber aus einem orthogonalen Kontrast (z.B. aus einem Haupteffekt, wenn die Interaktion getestet wird)
- Funktionelle Masken (z.B. aus einem vorgeschalteten „functional localizer“ Experiment)
- Anatomische Masken (z.B. selbstgezeichnet, Brodmann Areale, bestimmte Gyri oder subkortikale Strukturen, zytoarchitektonische Wahrscheinlichkeitskarten [Programme: z.B. AAL, Anatomy, WFU Pick Atlas])

Sphere um Koordinaten



Anatomische Maske



1. Teil: Zusammenfassung

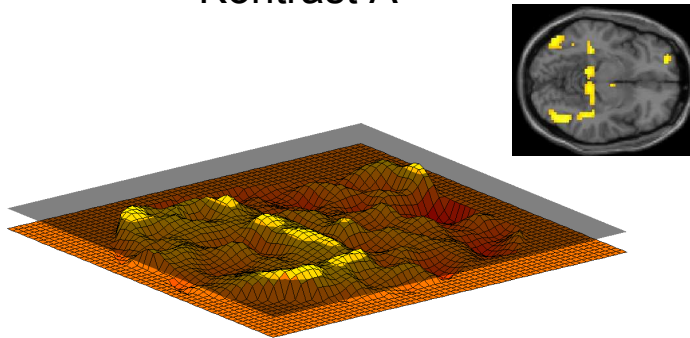
- Univariate fMRI Analysen führen tausende von Tests gegen eine globale Nullhypothese durch
- Multiple Tests führen zu einer drastischen Erhöhung der Irrtumswahrscheinlichkeit
- Bonferroni-Korrektur korrigiert die Irrtumswahrscheinlichkeit für die Familie aller Tests in Abhängigkeit von der Anzahl der Tests, so dass $p < 0.05$
- Bonferroni-Korrektur ist nur für unabhängige Tests angemessen und daher für fMRI Daten zu konservativ (Abhängigkeitsstruktur der Voxel)
- SPM beinhaltet zwei Korrekturmöglichkeiten: i) Kontrolle der *Family Wise Error Rate* über Gaussian Random Fields oder Bonferroni und ii) Kontrolle der *False Discovery Rate*
- Weitere Möglichkeit ist bei regionalen Hypothesen eine Einschränkung des Suchvolumens (Small Volume Correction) auf a-priori definierte Regionen

2. Teil: Effekte mehrerer Kontraste

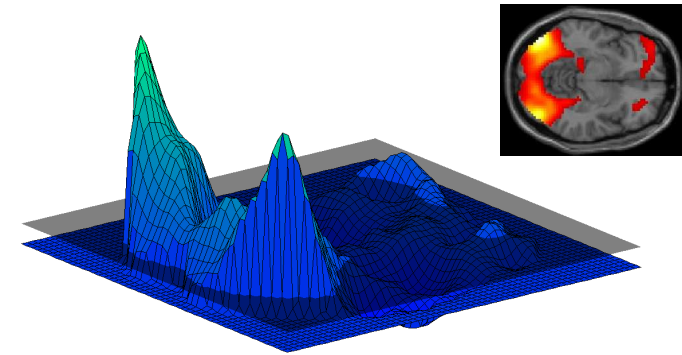
Inklusive Maskierung

A inklusiv maskiert mit B: Welche Aktivierungen überleben eine Schwelle von $p < \alpha_i$ in Kontrast A, die eine Schwelle von $p < \alpha_{ij}$ in Kontrast B überleben

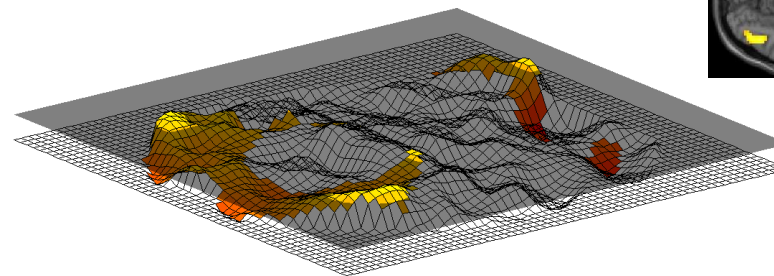
Kontrast A



Kontrast B



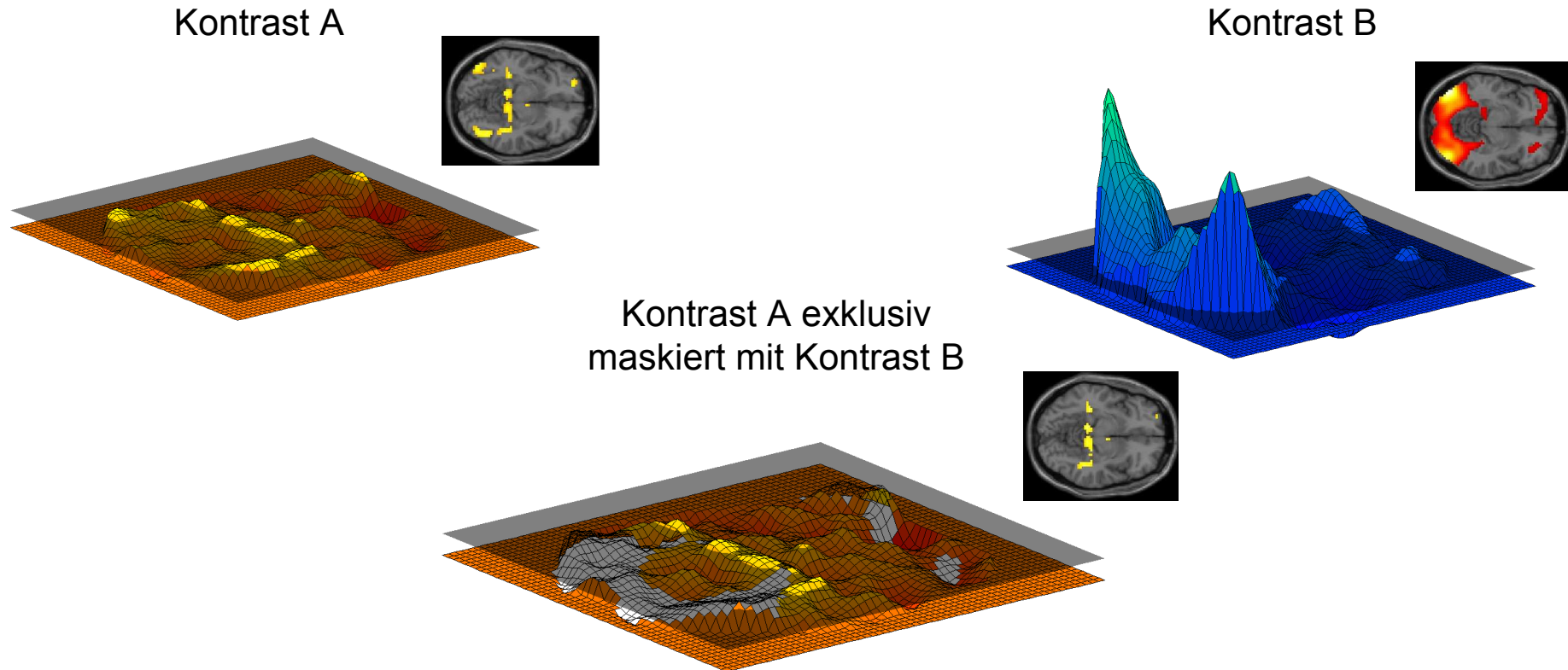
Kontrast A inklusiv
maskiert mit Kontrast B



Achtung: Inklusive Maskierung ist **kein statistischer Test** und ist nicht äquivalent zu Konjunktionstests

Exklusive Maskierung

A exklusiv maskiert mit B: Welche Aktivierungen überleben eine Schwelle von $p < \alpha_i$ in Kontrast A, die eine Schwelle von $p < \alpha_{ij}$ in Kontrast B nicht überleben



Achtung: Exklusive Maskierung ist **kein statistischer Test** und bedeutet nicht, dass in Voxel XYZ i) Kontrast A signifikant stärker aktiviert als Kontrast B oder ii) Kontrast B überhaupt nicht aktiviert

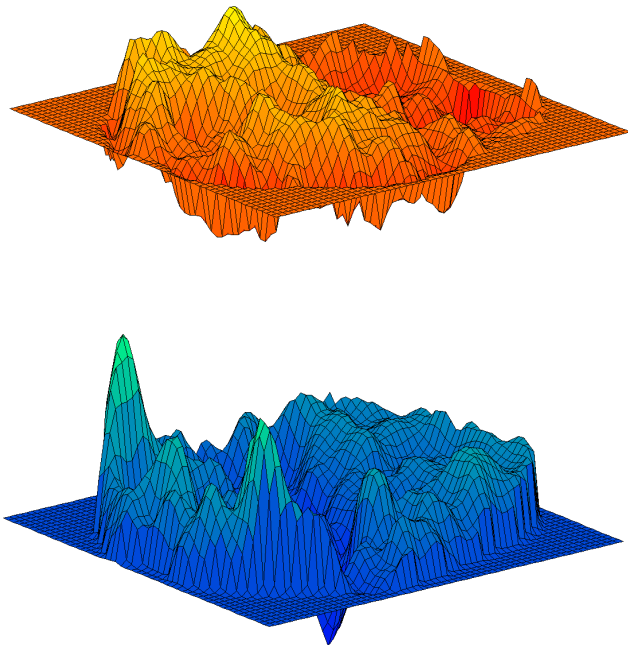
Konjunktion

Konjunktionsanalysen werden benutzt, um Gemeinsamkeiten verschiedener Kontraste zu ermitteln

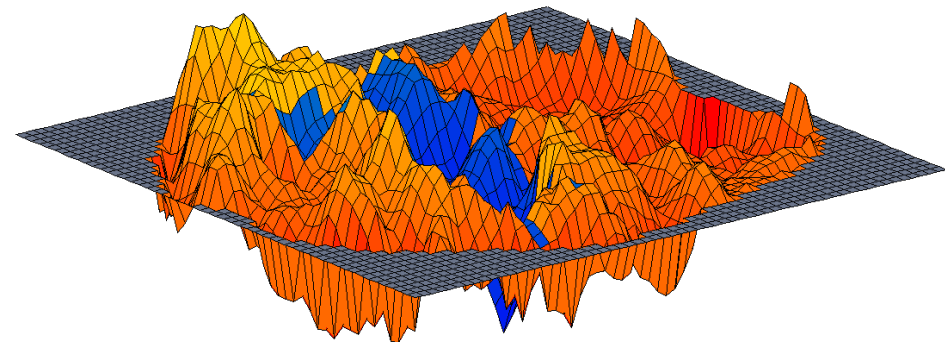
Konjunktionskarte besteht in jedem Voxel aus dem kleinsten t-Wert der ausgewählten Kontraste in diesem Voxel => **minimum t-statistic**

- großer Wert in allen Karten => großer Wert in Konjunktionskarte
- kleiner Wert in mindestens einer Karte => kleiner Wert in Konjunktionskarte

Ausgangskarten



Konjunktionskarte (Minimumkarte der Ausgangskarten)



Konjunktion

3 verschiedene Arten der Konjunktion in SPM

Global Null Conjunction

Nullhypothese: „0 effects under null“ - kein Effekt für alle n einzelnen Tests

Interpretation: Konjunktion von $k > 0$ Effekten

Testet gegen Minimum t-Verteilung

Tests müssen unabhängig sein

Intermediate Null Conjunction

Nullhypothese: „ u effects under null“ - kein Effekt für $n-u$ der n einzelnen Tests

Interpretation: Konjunktion von $k > u$ Effekten

Testet gegen Minimum t-Verteilung

Tests müssen unabhängig sein

Conjunction Null Conjunction

Nullhypothese: „ $n - 1$ effects under null“ - kein Effekt für einen der n einzelnen Tests

Interpretation: Konjunktion von $k > n - 1$ Effekten

Testet gegen t-Verteilung (conjunction of significances)

Unabhängigkeit nicht notwendig

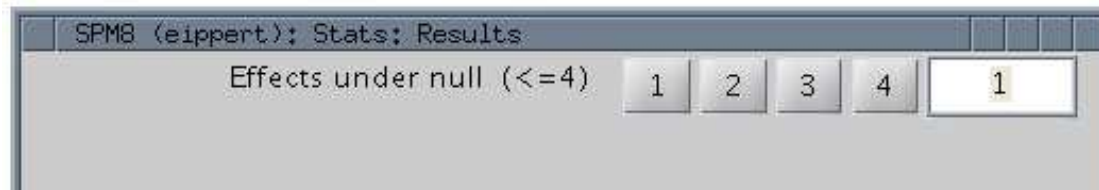
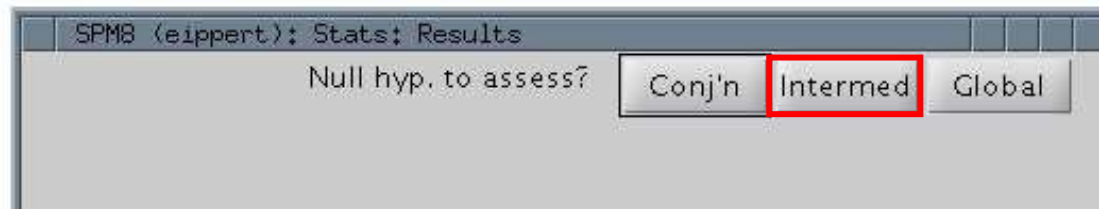
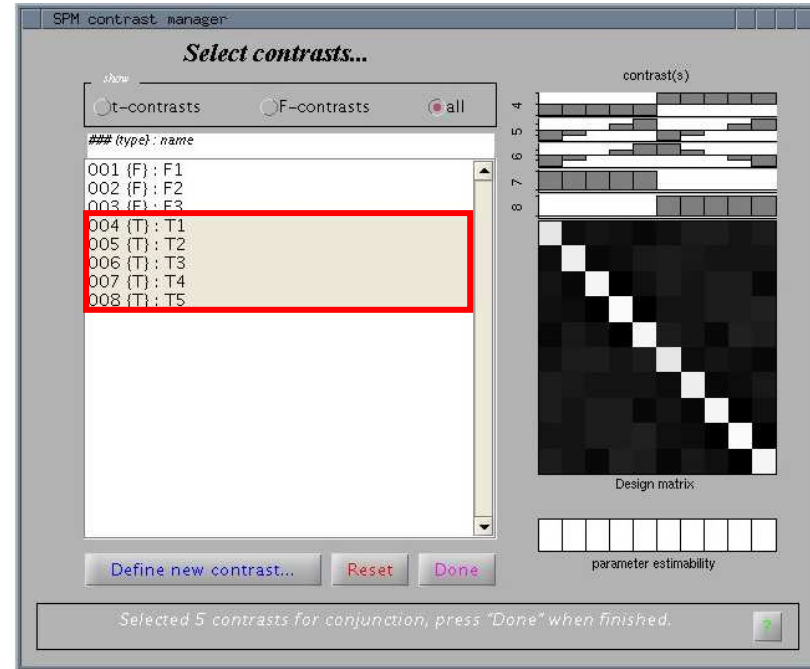
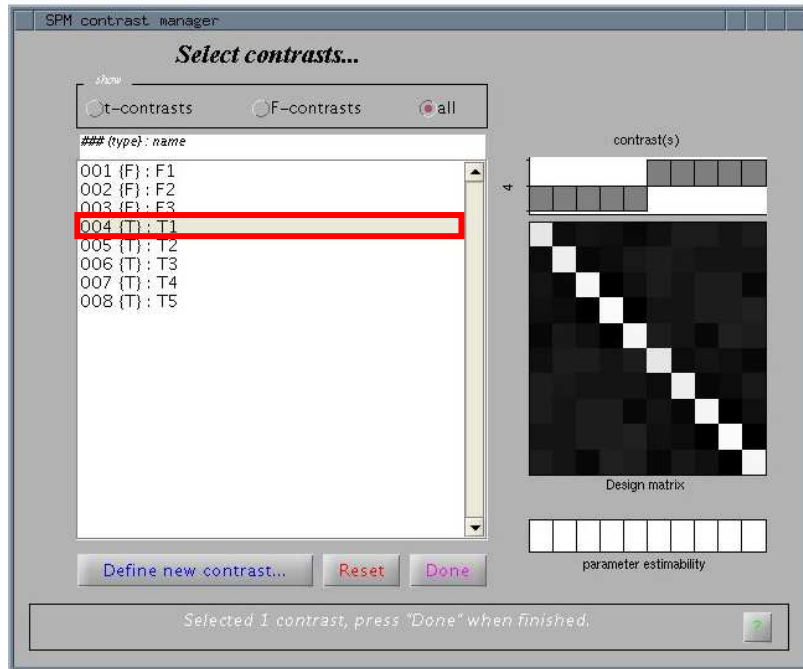
liberal



konservativ



Konjunktion



2. Teil: Zusammenfassung

- Maskierungs- und Konjunktionsanalysen betrachten Effekte, die mit einem einzelnen Kontrast nicht untersucht werden können
- Maskierungsverfahren (exklusiv, inklusiv) sind keine statistischen Tests, sondern nur Visualisierungshilfen
- Konjunktionsanalysen basieren auf einer Minimum t-Karte, die sich aus den in die Konjunktion eingehenden Kontrasten ergibt
- Konjunktionsanalysen in SPM (Global, Intermediate, Conjunction) haben unterschiedliche Nullhypothesen => Interpretation der Ergebnisse unterschiedlich
- Global Null Konjunktion ist am liberalsten, Conjunction Null Konjunktion ist am konservativsten

Literatur

Multiple Vergleiche

- Einführung: Brett et al., 2007 (SPM-Buch [siehe unten], Kapitel 17)
- Artikel zu multiplen Vergleichen und SVC in „Tools of the Trade“ Rubrik von SCAN
- Übersichtsartikel: Nichols & Hayasaka, 2003 (Statistical Methods in Medical Research)
- Fisch im Kernspin: Bennet et al., 2010 (Journal of Serendipitous and Unexpected Results)
- Artikel von Kriegeskorte et al. sowie Vul et al. zu „non-independence errors“ / double dipping
- FDR: Genovese et al., 2002; Chumbley & Friston, 2009; Chumbley et al., 2010, NeuroImage

Konjunktion

- Price & Friston, 1997, NeuroImage
- Friston et al., 1999a, 1999b, NeuroImage
- Nichols et al., 2005, NeuroImage, Friston et al., 2005 NeuroImage

=> www.jiscmail.ac.uk/cgi-bin/wa.exe?S1=spm (SPM mailing list)

=> **Buch: Statistical Parametric Mapping (Friston, 2007, Academic Press)**