

IRMf dans Memento

Rapport d'avancement - Avril 2024

Présentation des données.....	0
Premier contrôle qualité et passage en BIDS.....	1
Images.....	1
Métadonnées.....	1
Contrôle qualité.....	2
Dimensions permutées.....	2
Cerveaux à l'envers.....	5
Contrôle qualité automatique.....	6
Estimation du slice-timing et prétraitement.....	7
Analyse univariée.....	8
Étapes suivantes.....	10
Bibliographie.....	10

Ce rapport décrit d'abord le chemin parcouru depuis la découverte des données d'imagerie dans Memento jusqu'à la fin du prétraitement. Il s'agit d'un processus qui rend les images exploitables en corrigeant les défauts dûs au processus d'acquisition, en excluant les déchets inévitables dans une étude multicentrique de cette ampleur, et en rendant les séquences comparables entre elles grâce à une projection dans espace commun. Les résultats d'une première analyse statistique sont ensuite succinctement présentés afin de vérifier que les images contiennent suffisamment d'information pour discriminer les sujets déments des autres. Nous évoquons en fin de rapport les perspectives pour la suite de l'analyse.

Présentation des données

Les données qui nous ont été fournies sont divisées entre deux dossiers, en référence à la session d'acquisition : M000_M024 (mois 0 et mois 24 de l'étude) et M048 (mois 48). Il faut noter que M000_M024 contient également des fichiers du mois 48, mais pas tous. Le dossier est organisé par centre anonymisé, puis par sujet, puis par session d'acquisition (Fig 1).

Un sujet peut avoir été scanné à un, deux ou trois points temporels. À chaque session, on retrouve un nombre arbitraire de modalités : T1w, BOLD, T2w, diffusion, TEP, etc. Les noms de fichiers indiquent que certaines images ont été corrigées pour les

distorsions géométriques (*geo*), et/ou pour les niveaux de gris (*no_bias*), et parfois on ne sait pas (*unknown*).

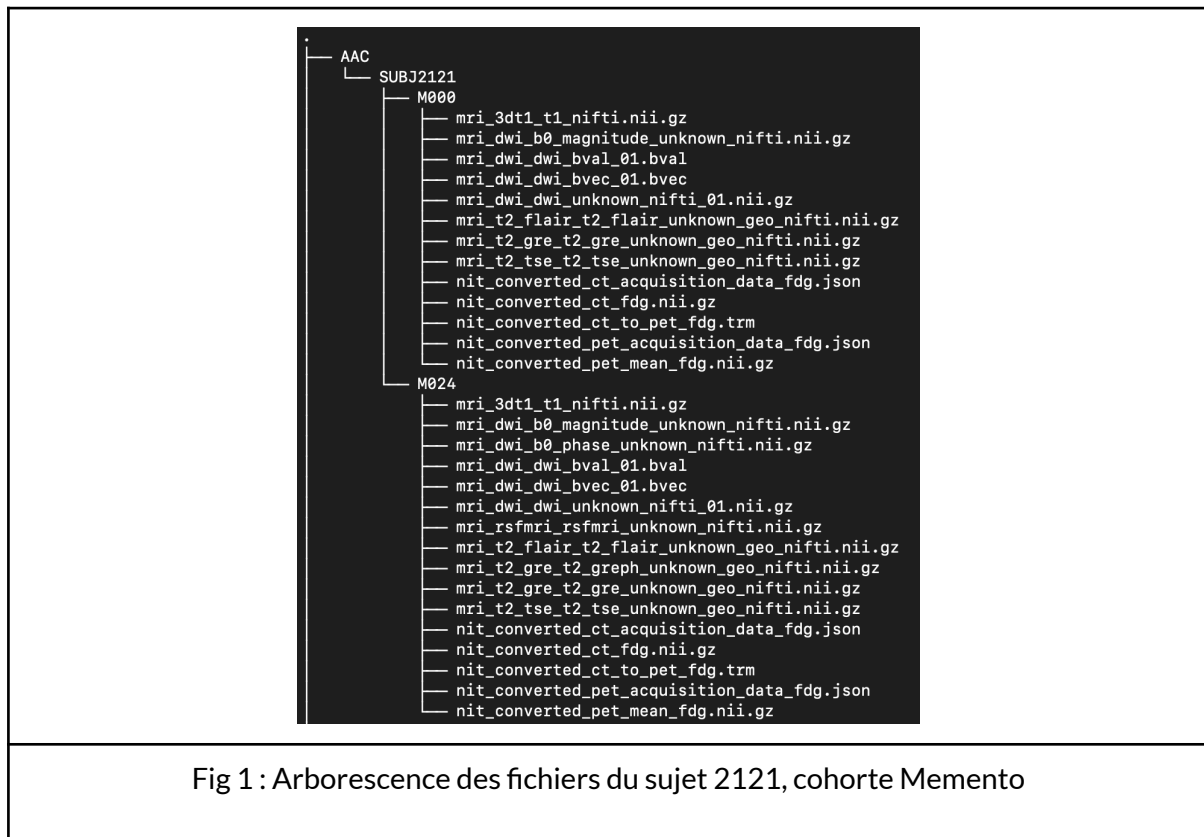


Fig 1 : Arborescence des fichiers du sujet 2121, cohorte Memento

De façon assez inhabituelle, les fichiers images ne sont pas associés à des fichiers de métadonnées (les *sidecars*), qui contiennent en général des informations essentielles sur les modalités d'acquisition : temps de répétition (TR), slice-timing, phase, etc.

L'objectif dans un premier temps est de réorganiser les données dans un format qui se prête au prétraitement, le format BIDS (Gorgolewsky et al., 2016).

Premier contrôle qualité et passage en BIDS

Images

Les images anatomiques (T1w) et fonctionnelles (BOLD) sont transférées en format BIDS avec un script python écrit pour l'occasion. À cette étape, on vérifie que le format des fichiers est correct : pas d'erreur de compression, bon nombre de dimensions, matrices qform correctes, etc. Lorsque plusieurs versions d'un même fichier existent, on garde systématiquement la plus corrigée.

Métadonnées

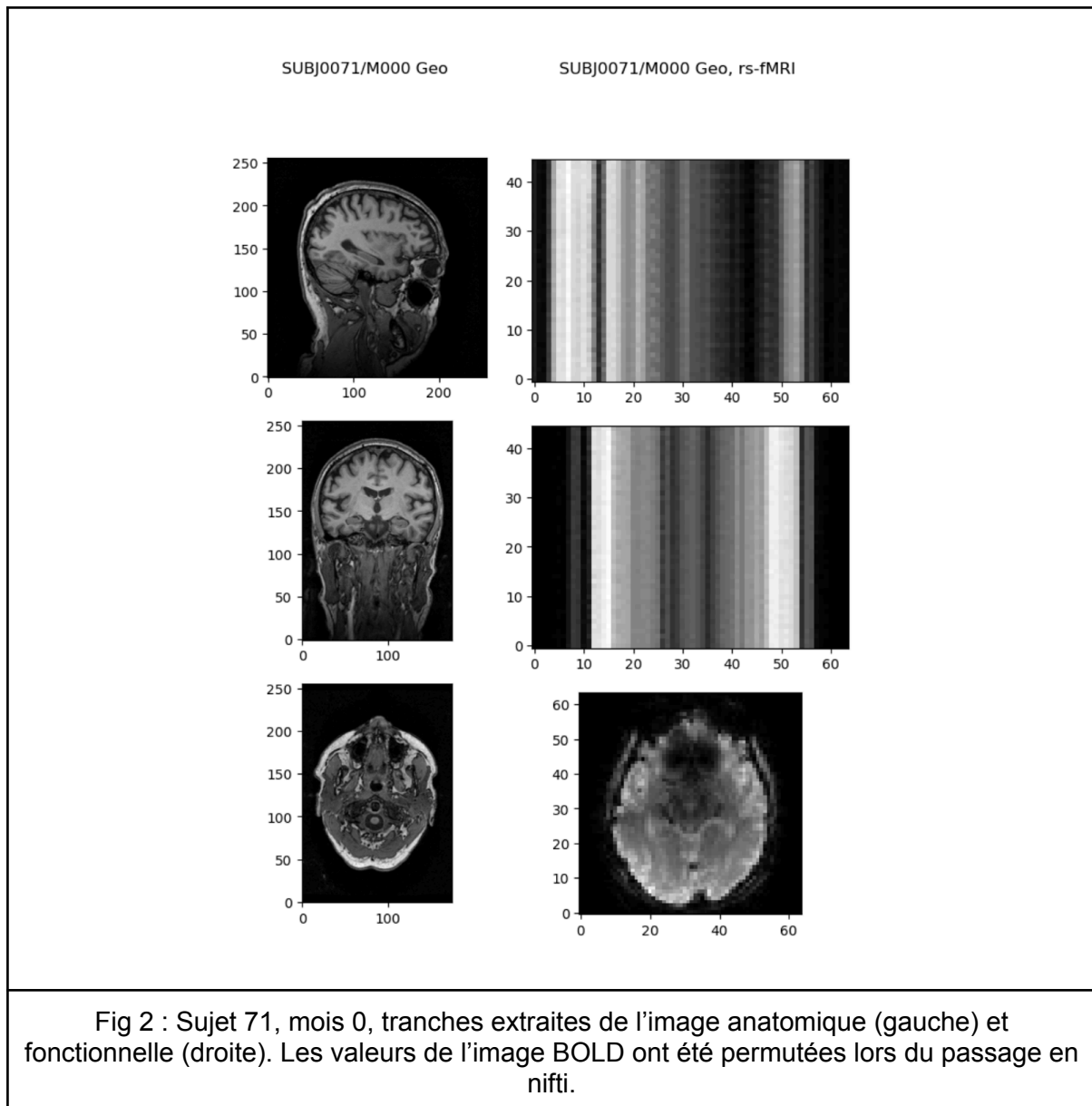
Afin de pallier l'absence des métadonnées, un fichier *sidecar* minimal est créé en récupérant le temps de répétition lorsque l'information est disponible dans l'entête des fichiers d'imagerie. Lorsque cette information est manquante, elle est inférée comme le TR médian des séquences provenant du même centre avec le même nombre de coupes. Un fichier des participants est également créé afin de se mettre en conformité avec le standard BIDS.

Contrôle qualité

Étant donné le volume de données de Memento, et les artefacts assez inhabituels que l'on peut y trouver, il est nécessaire de développer des outils pour repérer les images à réparer ou exclure. Pour chacun des problèmes ci-dessous, des scripts python ont été écrits et sont actuellement stockés sur les serveurs du GIN.

Dimensions permutées

Certains fichiers BOLD dans Memento présentent des rainures sur le plan axial et sagittal qui les rendent inexploitable. Il semble que les dimensions temporelles et spatiales aient été permutées lors de la transformation DICOM vers nifti.



Nous avons mis au point une procédure pour détecter automatiquement les fichiers corrompus. Il s'agit, pour chaque volume 4D, de calculer sa moyenne au cours du temps, d'extraire une image par plan de coupe anatomique, et de calculer leurs gradients dans toutes les directions. On obtient pour chaque plan une distribution des gradients des scans de Memento. On peut alors vérifier manuellement que les points aux valeurs de gradient anormalement basses sur le plan sagittal et coronal correspondent à des fichiers corrompus.

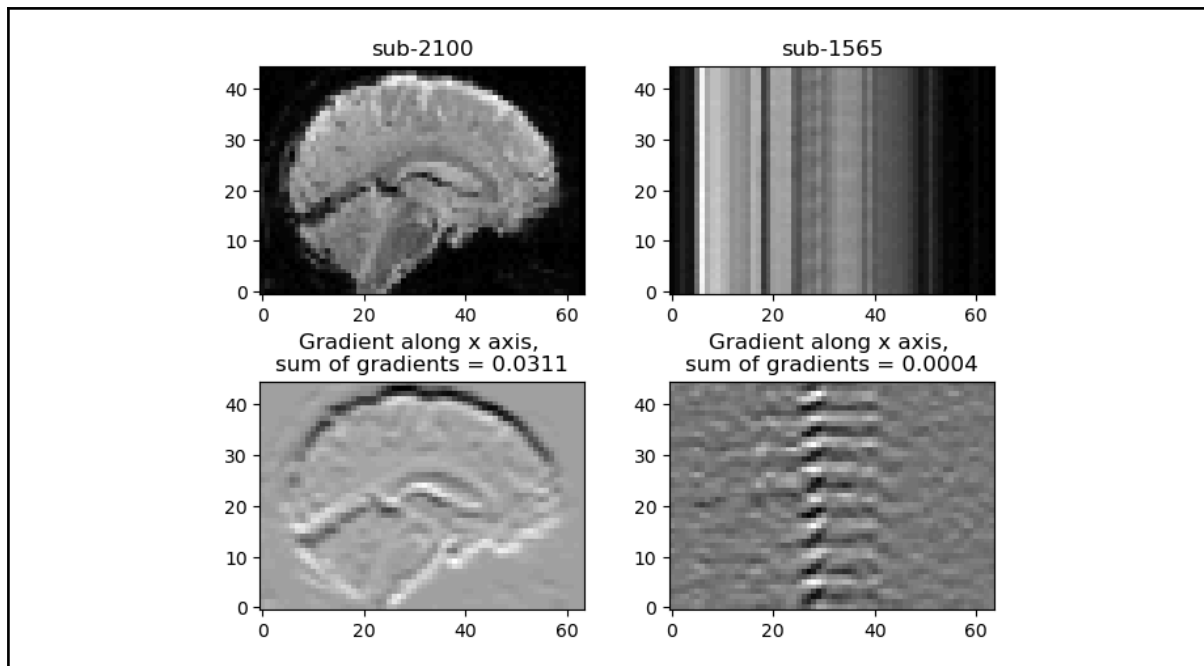


Fig 3 : Image normale et son gradient (gauche), comparés à une image corrompue (droite). On s'attend pour les images corrompues à des gradients de magnitude plus faible dans la direction des rainures, car l'image y présente peu de variation d'intensité.

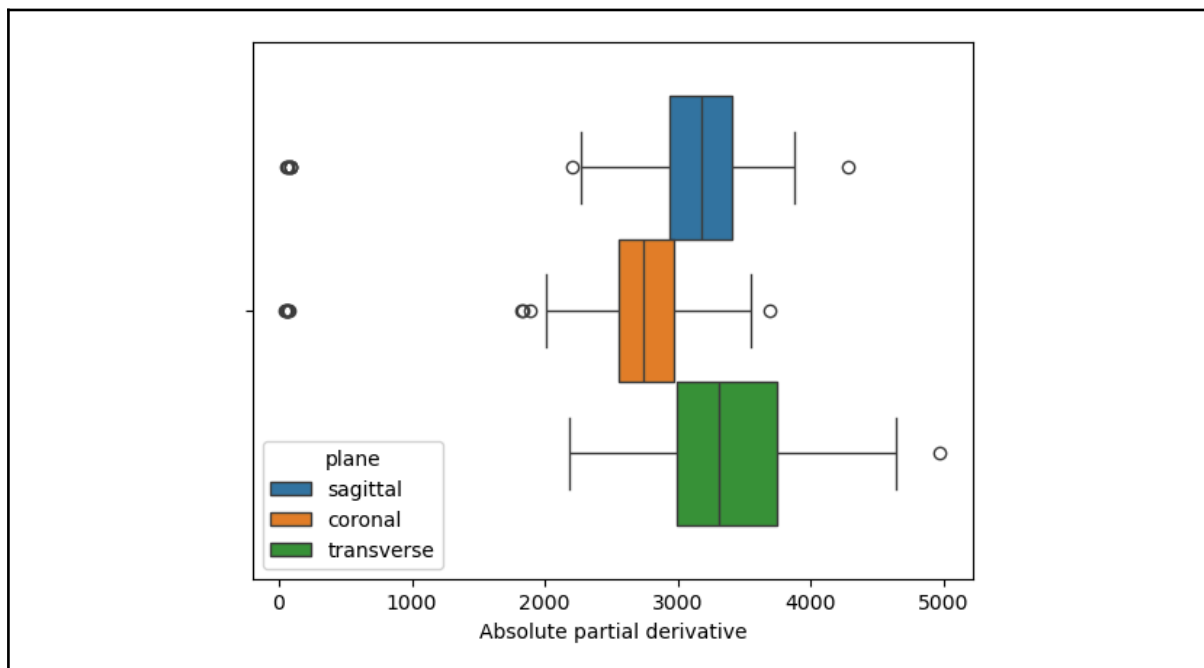


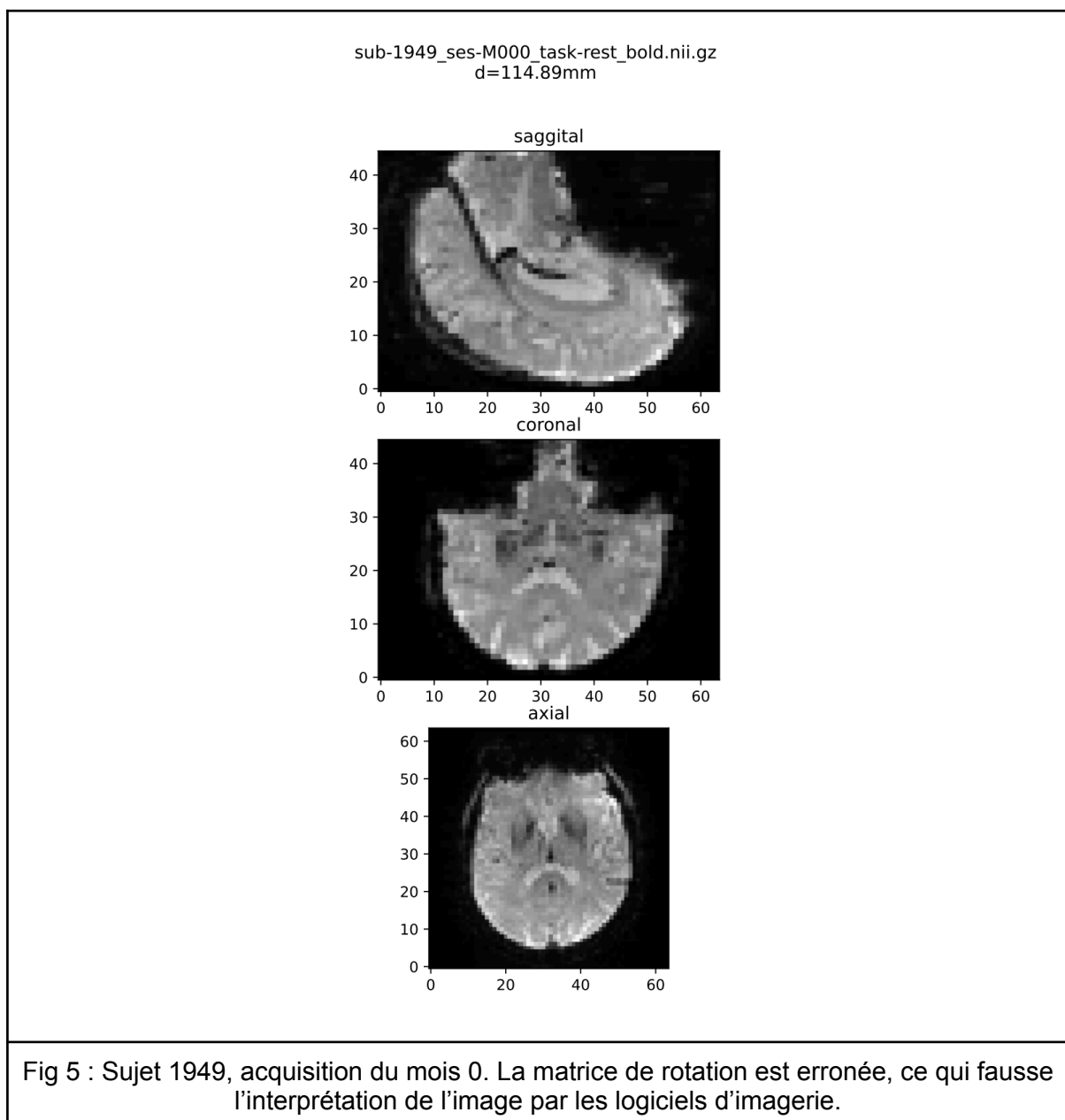
Fig 4 : Distribution des magnitudes des gradients. Les valeurs anormalement basses pour le plan sagittal et le plan transverse correspondent aux images corrompues.

Les fichiers corrompus contiennent encore toutes les informations nécessaires à l'analyse, ils ont simplement été reconstruits dans le mauvais sens. Nous avons mis au point un algorithme pour réarranger les données dans le bon ordre. À partir de chaque scan ainsi réparé, nous générons un rapport montrant les tranches du volume moyenné au cours du temps, de la dernière TR, ainsi que le signal global dans les domaines temporels et

fréquents. L'examen manuel de ces rapports nous permet de vérifier que les informations géométriques et temporelles ont été reconstruites de façon cohérente. 88 scans corrompus ont ainsi été détectés et réparés

Cerveaux à l'envers

Pour une raison inconnue, certaines séquences BOLD possèdent une matrice de rotation qui n'est pas cohérente avec les données numériques. Par conséquent, ces scans apparaissent à l'envers lorsqu'ils sont ouverts avec un logiciel d'imagerie, et il est difficile de les projeter dans un espace standard.

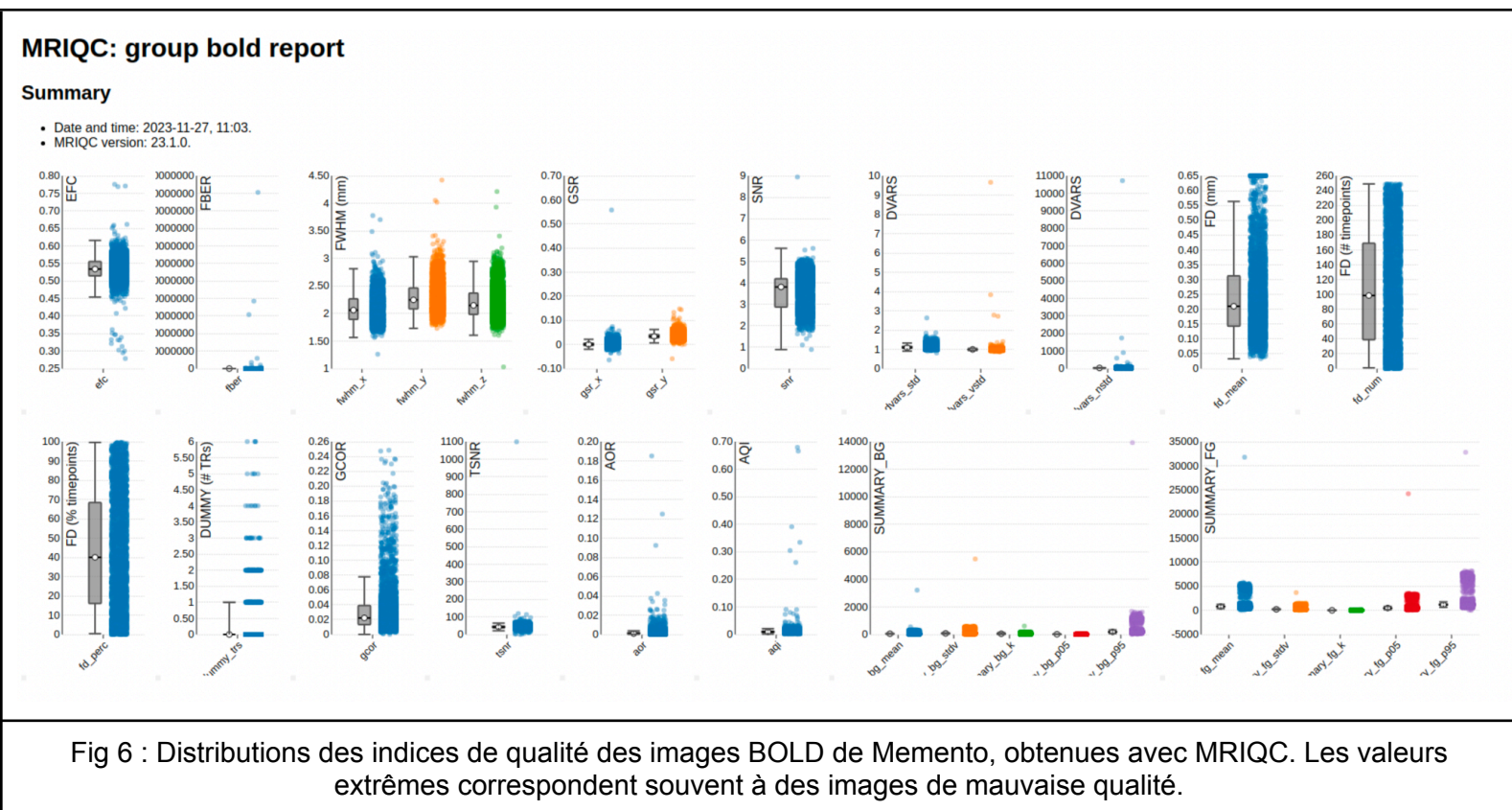


Il est possible de mesurer la distance euclidienne entre le centre de l'image BOLD et celle de l'image anatomique, en les projetant dans l'espace du scanner avec leurs matrices de transformation respectives. Lorsque la matrice de rotation de l'image BOLD est erronée, le centre de l'image BOLD est loin de celui de l'image anatomique ($d > 50\text{mm}$). En calculant cette distance pour tous les scans, il est possible de détecter les images "à l'envers". Nous en trouvons 44 dans Memento, qui sont exclues des analyses ultérieures.

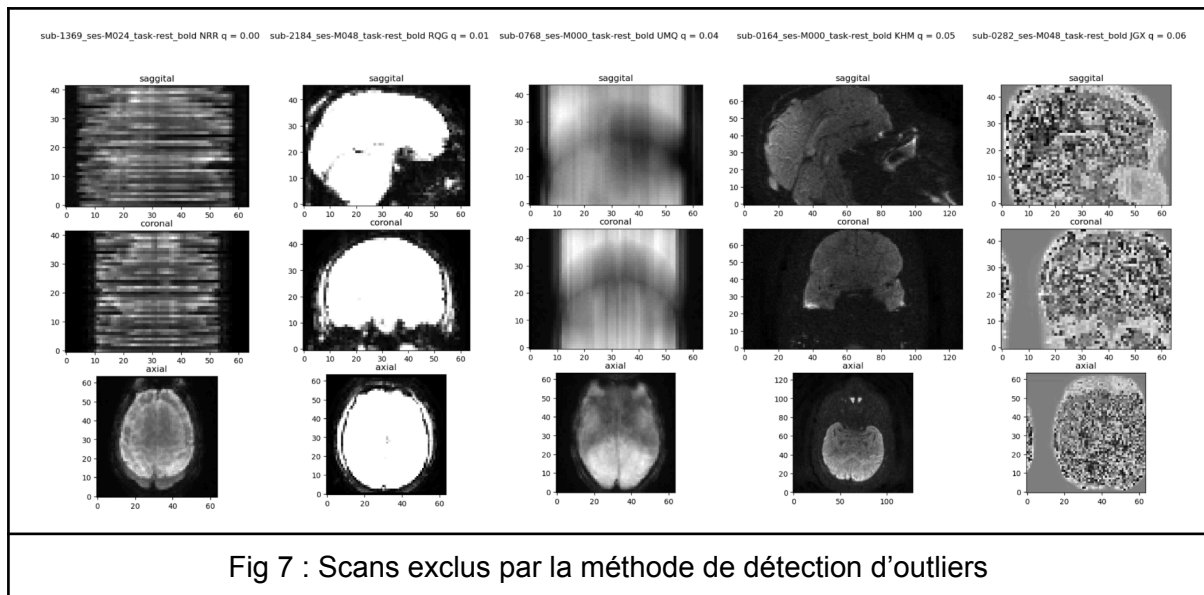
Contrôle qualité automatique

Les deux procédures ci-dessus permettent de repérer et traiter les problèmes les plus récurrents dans les scans Memento, mais il paraît difficile de mettre au point une procédure pour chaque type de problème. Il est préférable de tirer parti de la conversion au format BIDS pour mettre au point des méthodes plus générales.

Le logiciel MRIQC (Esteban et al., 2017) permet de calculer pour chaque scan un ensemble d'indices de qualité (*Image Quality Metrics*, IQM) dans les jeux de données au format BIDS. Les *outliers* dans la distribution d'un IQM correspondent généralement à des scans de mauvaise qualité.



Plutôt que de regarder les distributions une par une, nous avons calculé un indice global qui les prend en compte conjointement, grâce à un algorithme de détection d'*outliers* (Liu et al., 2012). En se basant sur cet indice, nous détectons et excluons cinq autres scans de la suite de l'étude.

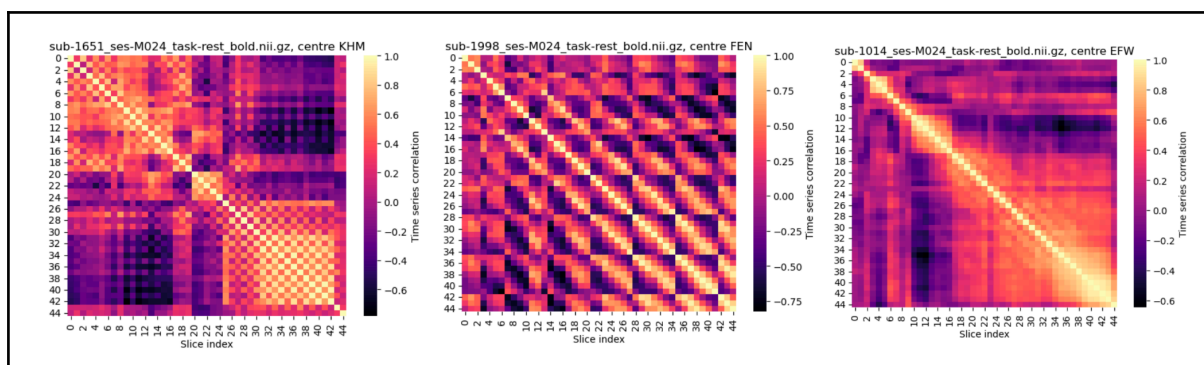


Estimation du slice-timing et prétraitement

L'information du slice timing est nécessaire pour effectuer des corrections sur les séquences BOLD qui permettent, selon le paradigme d'acquisition, d'augmenter la puissance des analyses statistiques ultérieures (Sladky et al., 2011).

Dans Memento, sans les métadonnées, nous ne savons pas comment les coupes ont été acquises, donc nous ne savons pas s'il faut corriger le slice timing, ni comment s'y prendre le cas échéant. Il faut donc mettre au point une méthode pour estimer cette information à partir des données.

En regardant la corrélation entre les signaux temporels moyens de chaque tranche au cours du temps, on obtient une partie de la réponse : certains scans semblent déjà corrigés, d'autres sont à corriger, et d'autres encore ont été acquis avec des méthodes d'acquisition dites "multibande" qui ne nécessitent pas de correction.



Il est inenvisageable de regarder les 3000 matrices pour déterminer scan par scan si la correction est nécessaire. Le slice-timing empirique est calculé automatiquement grâce à la corrélation croisée entre signaux des tranches paires et impaires, puis estimé pour chaque couple (centre, machine), et enfin ajouté au *sidecar* des scans nécessitant une correction.

Le prétraitement est réalisé avec fMRIPrep (Esteban et al., 2019) : coregistration, normalisation, estimation du bruit, segmentation, extraction du crâne, et correction du slice-timing lorsque c'est nécessaire. À l'issue du contrôle qualité des différentes étapes de prétraitement, il reste 1430 sujets, pour un total de 2924 scans BOLD.

Subjects

	All	Dementia converters	Non converters	Converter vs Non-converter, p-value
Sample size	1430	205	1225	-
Age, mean (SD)	71 (8.6)	74 (7.5)	71 (8.7)	<1e-3
Sex, %Female	63	56	64	0.024
baccalaureate or above, %	76	69	78	0.011

Scans

	All	Dementia converters	Non converters
M000 (% already converted)	1234 (30%)	173 (6.36%)	1061
M024 (% already converted)	1106 (24%)	143(40.56%)	791
M048 (% already converted)	899 (10%)	72(79.17%)	683
Total #scans	2924	389	2535

Fig 9 : Effectif de la cohorte Memento à la fin du prétraitement

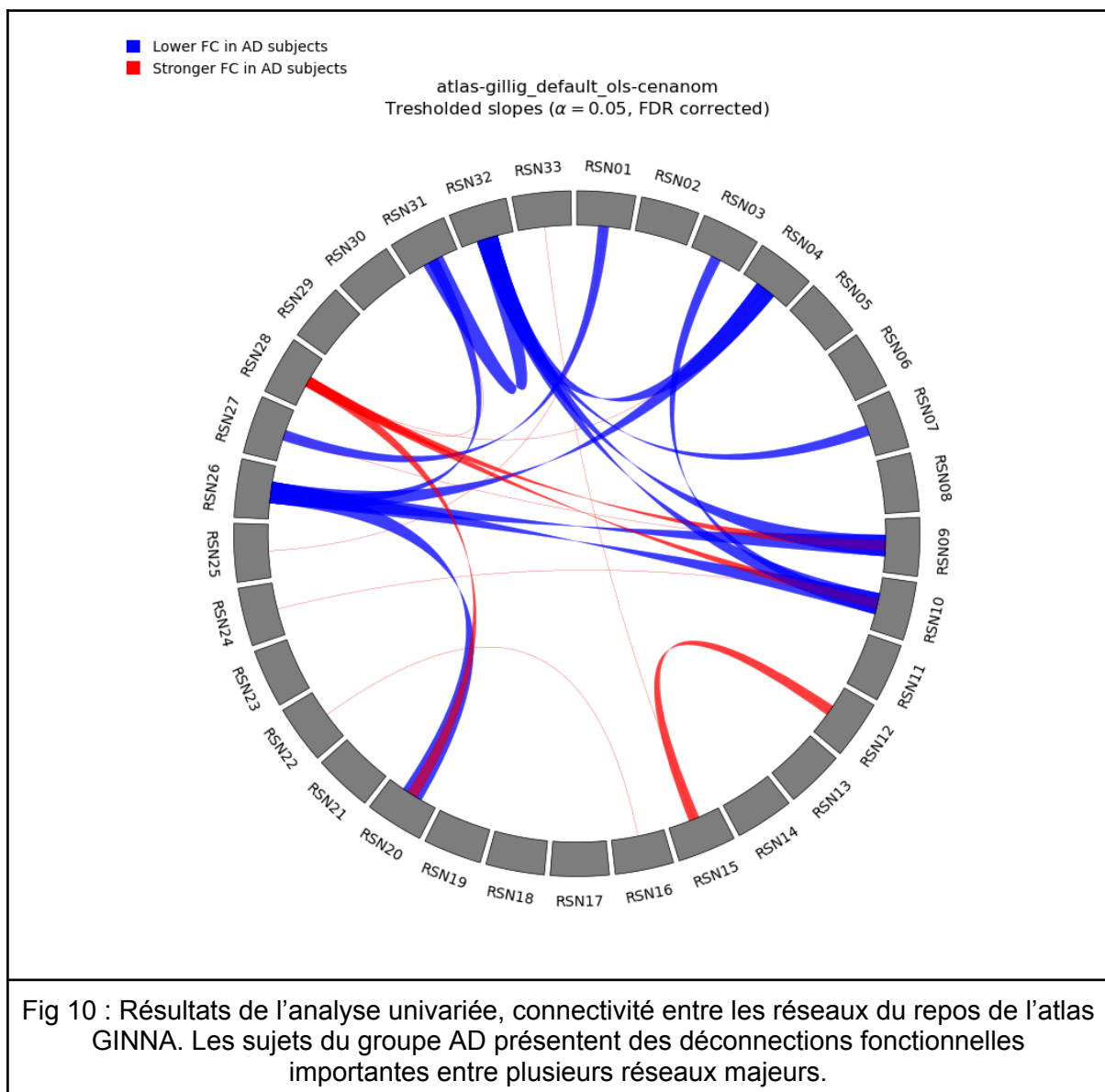
Analyse univariée

Afin de vérifier le succès du prétraitement, nous réalisons une première analyse de la connectivité fonctionnelle entre un ensemble de sujets qui ne se convertissent pas durant la durée de l'étude (*control*, N = 1221) et des sujets diagnostiqués avec la maladie d'Alzheimer avant leur scan (*AD*, N = 89). Il s'agit d'une analyse courante dans la littérature qui a déjà permis de révéler des différences significatives entre des sujets MCI et Alzheimer (Contreras et al., 2019).

L'analyse consiste, pour chaque scan, à :

1. moyenner le signal de l'IRM par région suivant un atlas, en l'occurrence l'atlas GINNA (Nozais and Naveau, en préparation);
2. nettoyer chaque signal en régressant les variables de bruit estimées durant le prétraitement et en conservant le signal résiduel;
3. calculer les corrélations des signaux deux à deux.

Cette procédure permet d'obtenir une matrice dite de connectivité pour chaque sujet. Un modèle linéaire, prenant également en compte l'âge, le sexe, le niveau d'étude et le volume intracranien est estimé pour chaque connexion. Même corrigés pour comparaisons multiples, les résultats indiquent des connexions significativement altérées dans le groupe AD, ce qui motive la poursuite de l'analyse sur des questions de recherche plus avancées.



Étapes suivantes

L'analyse précédente est intéressante comme preuve de concept mais elle présente certaines limites. Notamment, elle ne prend compte les connexions qu'une à une et ne permet pas de faire un lien direct avec la cognition. D'autre part, elle néglige l'aspect longitudinal de Memento. Une voie à explorer pour la suite consiste à récupérer les données issues des batteries de tests neuropsychologiques, de dégager différents profils d'évolution cognitifs puis d'explorer le potentiel de la connectivité fonctionnelle comme biomarqueur de ces évolutions. Nous cherchons également à renforcer la robustesse de l'analyse univariée afin d'en tirer une interprétation plus fine des interactions entre connectivité fonctionnelle et maladie d'Alzheimer.

Bibliographie

- Contreras, J. A., Avena-Koenigsberger, A., Risacher, S. L., West, J. D., Tallman, E., McDonald, B. C., Farlow, M. R., Apostolova, L. G., Goñi, J., Dziedzic, M., Wu, Y.-C., Kessler, D., Jeub, L., Fortunato, S., Saykin, A. J., & Sporns, O. (2019). Resting state network modularity along the prodromal late onset Alzheimer's disease continuum. *NeuroImage: Clinical*, 22, 101687. <https://doi.org/10.1016/j.nicl.2019.101687>
- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, 16(1), 111-116. <https://doi.org/10.1038/s41592-018-0235-4>
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1), 160044. <https://doi.org/10.1038/sdata.2016.44>
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2012). Isolation-Based Anomaly Detection. *ACM Transactions on Knowledge Discovery from Data*, 6(1), 3:1-3:39. <https://doi.org/10.1145/2133360.2133363>
- MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites | PLOS ONE. (s. d.). Consulté 26 avril 2024, à l'adresse <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184661>
- Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *NeuroImage*, 58(2), 588-594. <https://doi.org/10.1016/j.neuroimage.2011.06.078>