Simulating Alzheimer's disease progression with personalised digital brain models

Igor Koval[1,2,5,**], Alexandre Bône[1,2,**], Maxime Louis[1,2], Simona Bottani[2,1], Arnaud Marcoux[2,1], Jorge Samper-González[1,2], Ninon Burgos[1,2], Benjamin Charlier[6,1,2], Anne Bertrand[1,2,3,†], Stéphane Epelbaum[1,2,3], Olivier Colliot[1,2,3], Stéphanie Allassonnière[4,5] & Stanley Durrleman[2,1,*] for the Alzheimer's Disease Neuroimaging Initiative[‡]

[1] *Institut du Cerveau et de la Moelle épinière (ICM) & Inserm, U 1127 & CNRS, UMR 7225 & Sorbonne Université, F-75013 Paris, France*

[2] *Inria, Aramis project-team, Paris, France*

[3] *AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France*

[4] *Centre de Recherche des Cordeliers, Université Paris Descartes, Paris, France*

[5] *Centre de Mathématiques Appliquées, Ecole Polytechnique, Palaiseau, France*

[6] *Laboratoire Alexandre Grotendieck, Université de Montpellier, Montpellier, France*

$*$ corresponding author

$**$ contributed equally to this work

† deceased on March 2, 2018.

‡ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within

the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Simulating the effects of Alzheimer's disease on the brain is essential to better understand, predict and control how the disease progresses in patients. Our limited understanding of how disease mechanisms lead to changes seen in brain images and clinical examination hampers the development of biophysical simulations.

We propose here a statistical learning approach, where the repeated observations of several patients over time are used to synthesise personalized digital brain models. They provide spatiotemporal views of structural and functional brain alterations and associated scenarios of cognitive decline at the individual level.

We show that the personalization of the models to unseen subjects reconstructs their progression with errors of the same order as the uncertainty of the measurements. Simulation of synthetic patients generalise the distributions of the data in the training cohort. The analysis of factors modulating disease progression evidences a prominent sexual dimorphism and probable compensatory mechanisms in APEO-ε4 carriers.

This first simulator of its kind offers an unparalleled way to explore the heterogeneity of the disease's manifestation on the brain, and to predict its progression in each patient.

Numerical simulation has long been a central approach to understand complex systems, identify their determinants, and predict their behaviour. Recently, simulation has also proved to be key

in artificial intelligence, for instance for its ability to simulate a large number of go games that has made it possible to build a computer program that can learn to play better than a human[1]. Simulating a go game is easy because the rules are perfectly known and easy to implement. Simulating a brain developing Alzheimer's disease is more challenging because the biological mechanisms leading to the effects that are visible in brain images and clinical examinations are too imperfectly known[2], like the reason why these mechanisms lead to so heterogeneous effects across individuals. However, as with any complex system, simulating the disease is certainly a very promising way to better understand how it develops, identify the factors that modulate its manifestation in different individuals, and predict its progression in each patient.

We address here this simulation problem with a statistical learning approach. We design a computer program that automatically learns how Alzheimer's disease affects brain structure and function from the repeated observations of several patients in time. It estimates a typical long-term scenario of change by normalizing, re-aligning in time and combining several individual short-term data sequences. During training, the model learns how this typical scenario should be varied to reproduce the heterogeneity of progression profiles seen in the data by allowing adjustments in terms of age at onset, pace of disease progression and appearance of the model. Once trained, the model can be personalized to new subject's data or used to simulate entirely synthetic disease trajectories.

Statistical approaches to model disease progression have mostly remained descriptive so far, and do not generate long-term disease trajectories that are shown to accurately reproduce the het-

erogeneity of the progression at the individual level[3–9]. In the absence of reliable markers of disease progression, a central difficulty is to distinguish in data the differences due to disease progression from those due to the inter-individual variability that is independent of it. For instance, it is not clear whether differences usually found in clinical studies are not confounded by the fact that one compares subject's data who may be at different disease stages. It has recently been understood that seeing trajectories of data changes in the mathematical framework of the Riemmanian geometry allows one to ensure a unique decomposition between the variability in dynamics of disease progression (i.e. differences in age at onset or in pace of progression) and the inter-individual variability at any given disease stage[10,11]. The former is encoded by the temporal parameterisation of the followed trajectory, the latter by the position of the trajectory in space.

We use data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). In order to reproduce the natural history of the disease from the pre-clinical to the clinical stage, we selected the 322 subjects in this database who were included as cognitively normal (as defined in the ADNI protocol) or with mild cognitive impairments, and who had a confirmed clinical diagnosis of Alzheimer's disease at a later time-point in the study.

Whenever available, we use at each visit:

- regional measurements of standard uptake value ratio (SUVR) of Fluoro-DeoxyGlucose (FDG)-Positron Emission Tomography (PET) to build models of hypometabolism across brain regions,

- maps of cortical thickness defined on a mesh of the cortex and extracted from T1-weighted Magnetic Resonance Images (MRI) to build models of cortical thinning,

- surface meshes of the hippocampus of both hemispheres segmented also from T1-weighted MRI to build models of hippocampal atrophy, and

- scores of the Mini-Mental State Examination[12] (MMSE) and Alzheimer's Disease Assessment Scale - Cognitive Subscale with 13 items[13,14] (ADAS-Cog), the latter being divided into four sub-scores assessing memory, language, concentration and praxis, to build models of cognitive decline,

which amounts to 687 visits with PET images, 1,993 visits with MRI data and 1,235 visits with neuro-psychological assessments (See Methods and Extended Data Table 1 for summary statistics).

We represent the data as points on a multi-dimensional Riemannian manifold (see Fig. 1). For each modality, we choose a manifold that is adapted to the structure of the data: normalized measurements, image, or shape with a specific topology. Repeated observations of the same subject are thus seen as noisy samples along a curve on the manifold. Furthermore, we assume that such individual curves result from random spatiotemporal transformations of a geodesic curve that is common to the population. This hierarchical structure forms a mixed-effects statistical model[10,11].

The population geodesic is parameterized by an initial point on the manifold p_0 of the same type as the data, a velocity v_0 and a time-point t_0. By an appropriate choice of the Riemannian metric, we prescribe a certain form for this curve. For neuro-psychological assessments, each

score follows a logistic curve[10,11]. Cortical thickness decreases at a linear rate at each vertex of the surface, while ensuring that slopes (v_0) and intercepts (p_0) vary smoothly over the surface[15]. Regional SUVR maps also decrease at a linear rate with smoothly varying parameters across neighbor regions. The shape of the hippocampus meshes is changed by a smooth and invertible 3D deformation called diffeomorphism[16–18]. The use of the Riemmannian setting allows us to deal with all these data types with the same method and very similar algorithms (see Methods).

Subject-specific curves derive from the population average by random spatiotemporal transformations. Each transformation is composed of a parallel shift of the geodesic curve on the manifold called Exp-parallelization[11,19] combined with a linear time-reparameterization of the trajectory. The former is defined by a direction on the tangent-space of the manifold at some reference point, called "individual space-shift". The latter is defined by an acceleration factor and a time-shift encoding differences in pace of progression and delay at onset. Space-shifts encode variability in the magnitude of the effects, ordering of events, or change in the spatial pattern of alterations. More precisely, for SUVR regional measurements and cortical thickness maps, the space-shift encodes inter-subject variations in the ordering and relative timing of the alterations across the regions. For hippocampus meshes, it encodes variations in the shape of the structure for different individuals. For neuro-psychological assessments, it encodes the variations in the ordering and timing among different scores. We ensure that the effects of the space-shifts are not confounded by the changes due to disease progression along the population average trajectory by imposing an orthogonality condition between space shifts and the velocity of the geodesic at all time-points[10,17]. It makes the statistical model identifiable.

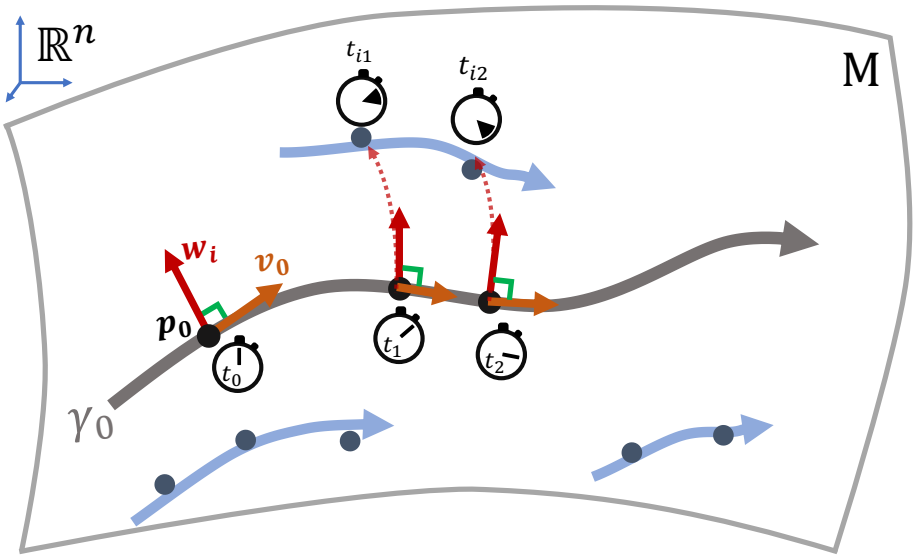


Figure 1: Hierarchical statistical model. Individual data acquired at age t_{ij} are seen as noisy samples along a subject-specific curves (blue) lying on a Riemannian manifold M. These curves derive from an Exp-parallelization (in the direction w_i in red) of a common population geodesic curve (in grey, parameterized by a point p_0, a velocity v_0 and a time t_0) and a time-reparametrization. The maximization of the model likelihood given longitudinal data estimates a typical long-term scenario of change, which is informed by a series of individual short term data sequences that are normalized and temporally aligned. Orthogonality condition ensures unique decomposition between changes due to inter-individual variability at the same disease stage and the ones due to disease progression. Once trained, the model can be used to fit new data, or generate entirely synthetic trajectories.

All in one, we define a mixed-effects statistical model, which may be written as $y_{ij} = f(\theta, z_i, t_{ij}) + \varepsilon_{ij}$, where y_{ij} is the j-th observation of the i-th subject observed at age t_{ij}, f is a non-linear function that is specific to each data type, and ε_{ij} is a residual noise. The vector θ contains the fixed-effects p_0, v_0, t_0, the variance of the random-effects and the variance of the noise, and the vector z_i corresponds to the random-effects: acceleration factors, time-shifts and space-shifts, which are specific to each individual. We add priors on the coordinates of the vector θ in a Bayesian setting. When t is varied, the curve $f(\theta, z_i, t)$ represents the subject-specific trajectory at any time t.

We now consider three successive statistical tasks:

- **calibration:** given the longitudinal data set $\{y_{ij}, t_{ij}\}_{i=1,\dots,N, j=1,\dots,N_i}$ for a certain type of data, we find the value of parameters θ that maximizes the joint likelihood $p(\{y_{ij}\}_{ij}, \theta) = p(\{y_{ij}\}_{ij}|\theta)p(\theta)$. The optimal value $\hat{\theta}$ fully specifies the model of disease progression;

- **personalization:** for the optimal value of the parameter $\hat{\theta}$, we personalize the model to the repeated data of a subject i (either a training subject, or a test subject in a cross-validation setting) $\{y_{ij}, t_{ij}\}_{j=1,\dots,N_i}$ by finding the optimal value of the random-effect $\hat{z}$ that maximizes the conditional likelihood $p(\{y_{ij}\}_j, z|\hat{\theta})$;

- **simulation:** for the optimal value of the parameter $\hat{\theta}$, we can simulate random-effects z and generate synthetic data y at any user-defined time-point t by computing $y = f(\hat{\theta}, z, t)$ and adding noise.

We use a stochastic approximation of the Expectation-Minimization algorithm[20,21] for calibration, gradient-descent based method or Powell's method for personalization, and kernel density estimation together with dimension reduction for simulation (see Methods).

Multimodal disease progression models

For each data type, we calibrate the model parameters using all available visits of the selected subjects. The resulting normative models of progression are estimated relatively to a different temporal reference frame for each data type. For visualization and interpretation purposes, we

synchronize the different models using affine time-reparametrization with the cognition model as a reference. Furthermore, we used the age at which each subject has been diagnosed with the disease to find the corresponding stage of progression at diagnosis on the reference time-line (see Methods).

Fig. 2 shows the synchronized models of hypometabolism, cortical thinning, hippocampal atrophy and cognitive decline at four representative time-points encompassing 16 years before diagnosis and 8 years after. These models may be visualized at a fine temporal resolution in the form of an interactive visualization at the website: `www.digital-brain.org`.

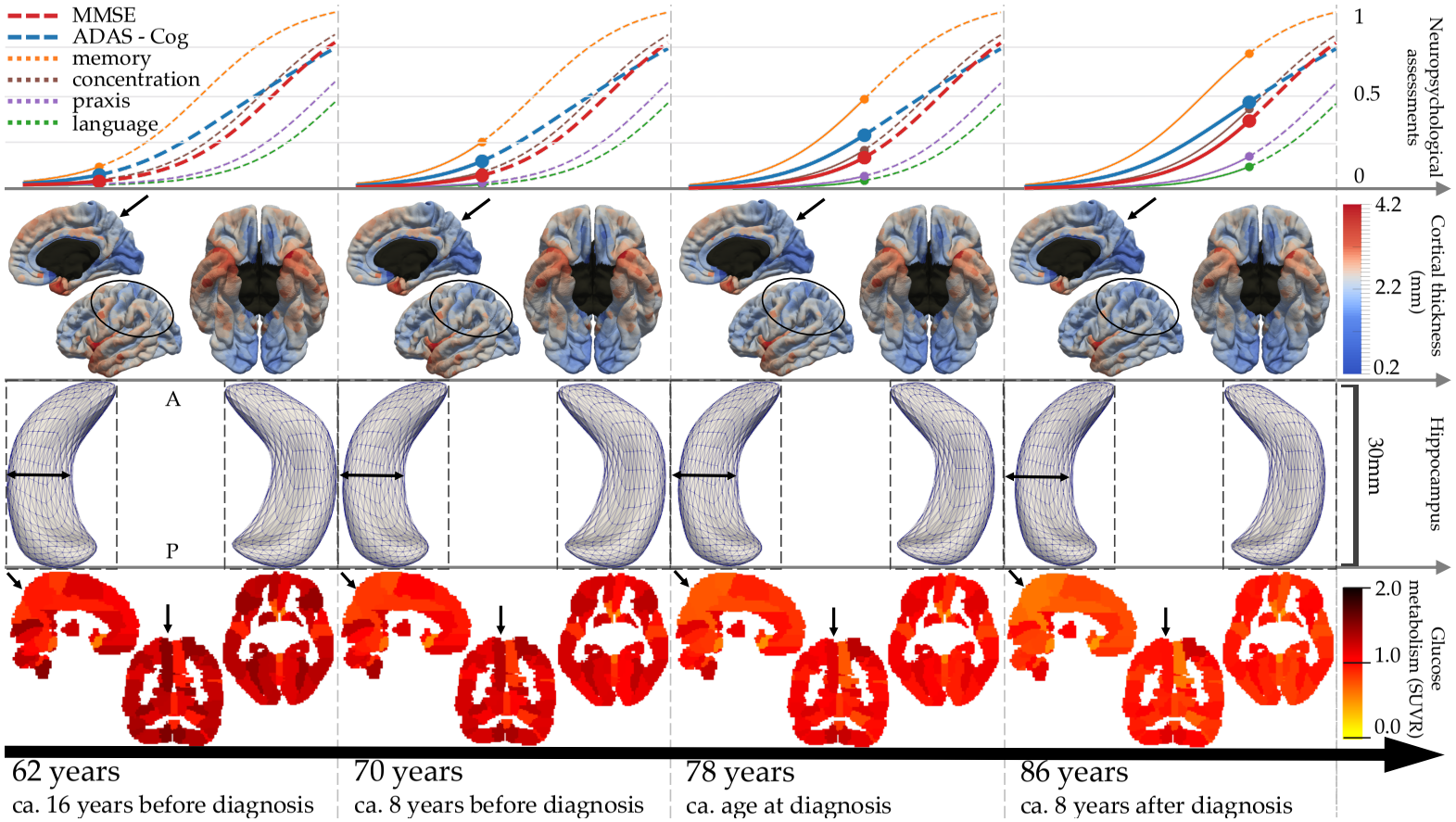


Figure 2: Normative model of Alzheimer's disease progression shown at 4 time-points with estimated time until/from diagnosis. Bottom to top rows show alteration of brain glucose metabolism, hippocampal atrophy, cortical thinning and onset of cognitive decline. Black arrows and ellipses indicate some areas of great changes.

The greatest alterations of glucose hypometabolism are found in the following regions that are consistent with previous studies: precuneus[22–24], prefrontal areas[25] and the parahippocampal region[26] (see Fig. 2 and Extended Data Fig. 1 showing annual SUVR decrease rate at age of diagnosis).

The greatest cortical atrophy (Fig. 2 and Extended Data Fig. 2 showing annual atrophy rate at age of diagnosis) also occurs in regions that are usually associated with it in the literature: the enthorinal cortex, the hippocampal gyrus, the temporal pole and the fusiform gyrus[27,28], cortical association areas (inferior parietal lobe[29] and temporal lobe[30]) and the precuneus[31]. As expected, very little atrophy is shown to occur in the occipital lobe and the cingulate gyrus. More suprisingly, the model shows atrophy in the precentral gyrus and the paracentral lobule. Whether these regions are affected by cortical thinning due to Alzheimer's disease is still a debated question[32]. It is worth noting that the noise of the measurement is by far the greatest in these areas, as measured by the residual noise of a linear regression performed for each subject independently, which present coefficients of determination R^2 lower than 0.25. Therefore, the high level of uncertainty in cortical thickness measurement must be taken into account when interpreting results in this region, and is probably a reason for disagreements across studies.

Deformation of the hippocampus due to neuronal loss during disease progression exhibits a complex pattern with deformation occurring more in the lateral parts of the hippocampus than in the antero-posterior direction. This complex pattern of shape changes is likely to be the consequence of tissue remodeling occurring within the temporal lobe due to neuronal loss. It suggests

that shape descriptors may be a more sensitive marker of disease progression than just the volume that is usually used in clinical studies[33,34].

The model of cognitive decline shows a typical sequence of cognitive impairments, as measured by combinations of the ADAS-Cog with 13 items, during the course of the disease starting with memory, followed by concentration 9.6 years after, praxis 9.8 years after, and finally language 3.3 years after. It has been shown that Alzheimer's disease diagnosis occurs when the ADAS-Cog is comprised between 18.6 and 28.9 (i.e. between 0.21 and 0.34 in the normalized scale)[35], which is reached between 74 and 80 years old in our normative time-line. Similarly, the diagnosis usually occurs for a MMSE score comprised between 27 and 23 (i.e. 0.1 and 0.23 in the normalized scale)[36], which occurs between 74 and 81 years old in our normative time-line. The age at diagnosis in the normative time-line has been estimated at 78 (± 5.6) years old. The consistency of these estimates shows that the algorithm was able to correctly align the individual short term data sequences around the diagnosis time, by using solely the analysis of the spatiotemporal patterns of data changes and not the age at which the subjects were diagnosed.

Interestingly, we notice that the estimated scenario of hypometabolism and cortical atrophy, which encompasses 16 years before diagnosis, shows the greatest alterations in the associative areas of the parietal lobe and in the medial frontal lobe, and to a lesser extend in the entorinal and para-hippocampus regions. It suggests that the most prominent changes in the hippocampus regions, a well known effect of Alzheimer's disease, must have occurred at least 15 years before the onset of memory impairment and the clinical diagnosis. This multi-modal disease model confirms

the large time gap between development of brain lesions seen in images and the onset of cognitive decline seen in clinical observations.

Reconstruction errors and generalization to unseen data

We use a five-fold cross-validation procedure to replicate model calibration five times on 80% of the training data set. The consistency of the fixed-effects estimates in this cross-validation setting shows the robustness of the estimation algorithm in different runs against resampling in the training set (see Extended Data Table 2). Furthermore, the delay between impairment of cognitive functions, as they are defined by our division of the ADAS-Cog with 13 items, is, relatively to memory, of $9.4 \pm 1.6(std)$ yrs for concentration (9.6 yrs using all data), $19.9 \pm 2.0(std)$ yrs for praxis (19.4 yrs using all data), $23.3 \pm 2.6(std)$ yrs for language (22.7 yrs using all data).

We personalize now the estimated models to the repeated observations of any subject. On the one hand, we personalize the model to the training subjects using the whole data set. It yields a set of individual parameters for each subject. On the other hand, we estimate the model using 80% of the subject and then personalize it to the remaining 20% subjects, yielding a set of individual parameters for test subjects only. After five splits, we recover a full set of individual parameters estimated in a cross-validation setting. We show that the discrepancy between individual effects estimated as training or test sample is small with r^2 comprised between 0.93 and 0.99 (see Extended Data Fig. 3).

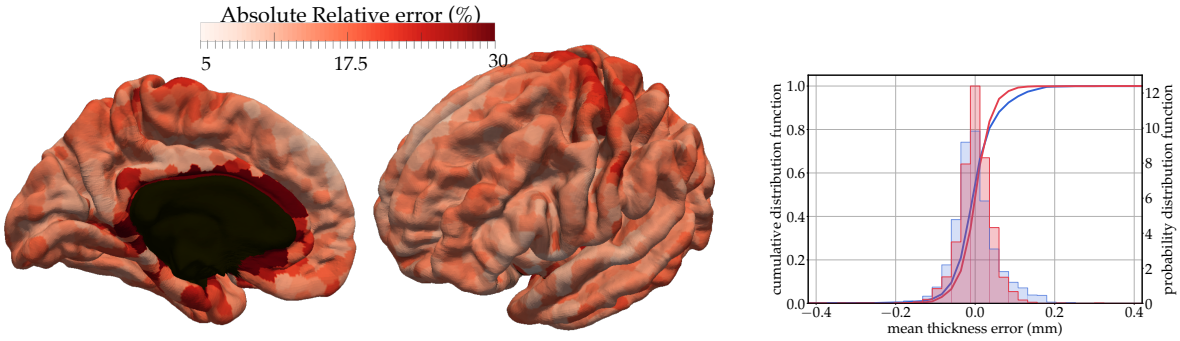
After showing the robustness of both the fixed and random effects estimates, we assess now

the goodness-of-fit of the model by measuring the reconstruction errors between the fitted model and the observed data. We do not expect a perfect match between prediction and observations as we imposed smoothness constraints on the spatial and temporal variations of the data and estimated a level of noise during model training with the aim to avoid over-fitting and allow better generalization. Assessing the accuracy of goodness-of-fit is a difficult task, as one does not know the true level of noise of the measurements. We estimated this measurement uncertainty using test / re-test MRI sessions, PET data at baseline and follow-up for amyloid negative cognitively normals subjects and a literature review of reproducibility of neuro-psychological assessments (see Methods).

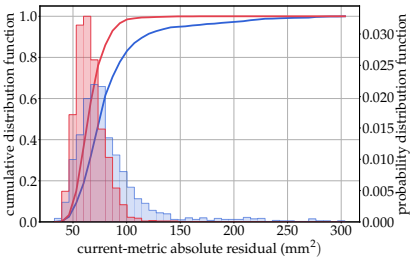
Fig. 3 shows the superimposition of the empirical distribution of reconstruction errors with the empirical distribution of the noise for all data types. Overall, the two distributions largely overlap, and the standard error is of the same order than the measurement noise (see Extended Data Table 3). We notice that the reconstruction errors in brain regions are not evenly distributed. For PET data, the largest errors are found mostly in smaller regions. For cortical thickness, larger errors are found at the boundary of the mesh with the corpus callosum, mostly due to interpolation errors. These errors are much smaller than the best possible image resolution of 1 mm isotropic, thus making our reconstructions at sub-voxel precision.

We measure distance between hippocampus meshes using the currents distance, which allows one to compare shapes with different samplings while being robust to small protrusion or topology changes[37]. As a consequence, the personalization of the model tends to ignore the many spikes

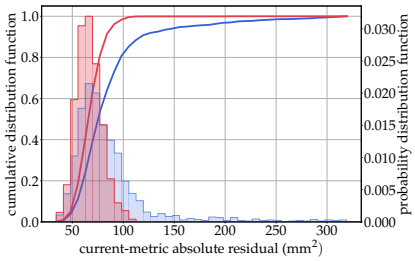
(a) FDG-PET images



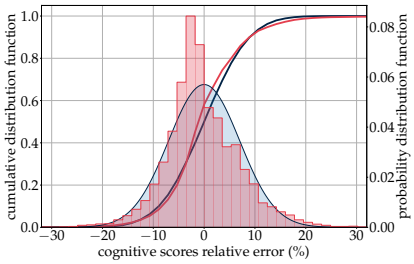
(b) Cortical Thickness maps



(c) Left hippocampus mesh



(d) Right hippocampus mesh



(e) Neuro-psychological assessments

Figure 3: Error of reconstruction of the model. The empirical distribution of errors (red) is superimposed with the estimated distribution of test / re-test differences (in blue). For FDG-PET images and cortical thickness maps the absolute relative error is shown in every brain region. Mean and standard errors are given in Extended Data Table 3.

pointing outward that are often seen in the segmentations. Reconstructed meshes are smoother than observations, resulting in an under-estimation of the volume of the observation (see Extended Data Fig. 4). It is more desirable to accurately reconstruct the shape rather than the volume, which is very sensitive to small segmentation errors. For instance, 83% of the subjects shows sequences of segmentation volume that are not monotonously decreasing, compared to only one subject for the volume of reconstructed meshes. Nevertheless, one should keep in mind that our reconstructions present a systematic bias in volume compared to the volume of the original segmentations.

Eventually, we measured the same reconstruction errors but when personalizing the model to data that were not seen during model calibration in the five-fold cross-validation setting. Distributions of these reconstruction errors are essentially identical with the previous ones obtained by calibrating and personalizing the model on the whole data set (see Extended Data Fig. 5). Only hippocampus shows a slightly higher generalization errors but still below the noise level estimated with test / re-test data. The reconstruction of unseen data is as good as the reconstruction of the training data, thus showing that the personalization of the model generalizes well to new individual data sequences.

Simulation of virtual cohorts

We now take advantage of the generative aspect of the statistical model to simulate entirely synthetic patients developing Alzheimer's disease. Calibration yields a series of individual parameters, from which we estimate the empirical posterior distribution. We sample random parameters from

this distribution, and use them to generate synthetic trajectories and then synthetic data as noisy samples along the simulated trajectory.

More precisely, for each gender status, we compute a kernel density estimation for the joint probability distribution of the temporal parameters: age at baseline, time-shift and speed factors. We then compute the multivariate Gaussian distribution of all other parameters conditionally to the temporal parameters. We draw samples from these two empirical distributions, and generate synthetic data at any given age of these virtual patients (see Methods).

To validate such simulations, we replicate the original ADNI data set by randomly picking a baseline age and simulating men and women subjects with the same sex ratio, the same number and same frequency of observations as in the original cohort. We then compute the distributions of simulated regional SUVR, cortical thickness, hippocampus volume and neuro-psychological assessments, and superimposed them with the distributions of the original data, and the data we reconstructed previously by model personalization.

The superimposition of the distributions shows that the simulated data closely replicate the reconstructed data for all modalities (see Fig. 4). For the hippocampus volume, the simulated data have the same bias than the reconstructed data in comparison to the real data. This fact is expected as the simulation reproduces the variability learned by the model. This experiment shows that the model accurately reproduces the diversity of disease progression patterns observed in the training cohort. It can be used therefore as a simulator of subjects developing Alzheimer's disease, which replicates the heterogeneity of the disease progression.

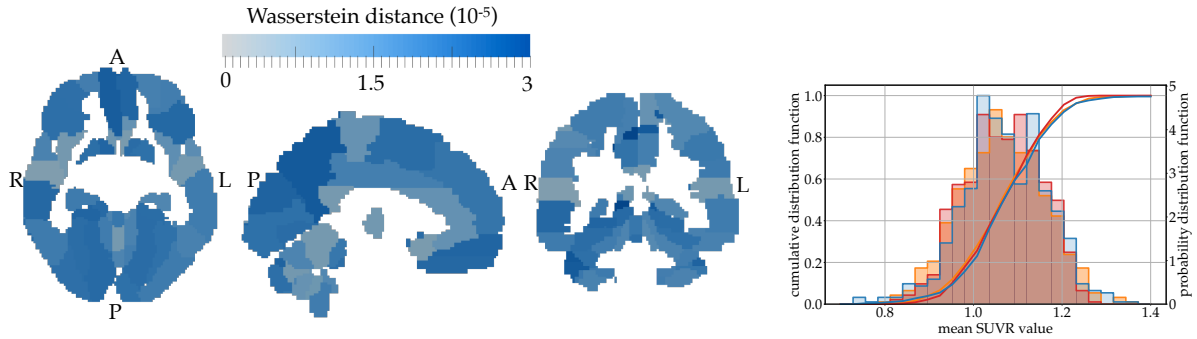
This simulator can be used to arbitrarily increase the number of subjects, number of visits and visits frequency in the training cohort. These data augmentation and resampling techniques are essential to improve the performance of machine learning algorithms. It can also be used to create large validation sets to better evaluate their generalization errors.

These models can be seen as an anonymous replicate of the training cohort, which can therefore be transferred and shared without regulatory constraints in lieu of the data itself. In this way, they allow the comparison and combination of multiple cohorts that would otherwise be very difficult to inter-operate. They can thus be used to detect the respective biases of these cohorts, and possibly correct them by simulating patients with a re-balanced disease stage distribution, sex ratio, or ratio of APOE-ε4 carriers for instance. The identification of such biases is essential because they are then found in the predictive systems trained on these cohorts. In this regard, it would be relevant to compare our results obtained on a research cohort with other types of cohorts, particularly epidemiological ones, if there are any with as many imaging modalities.

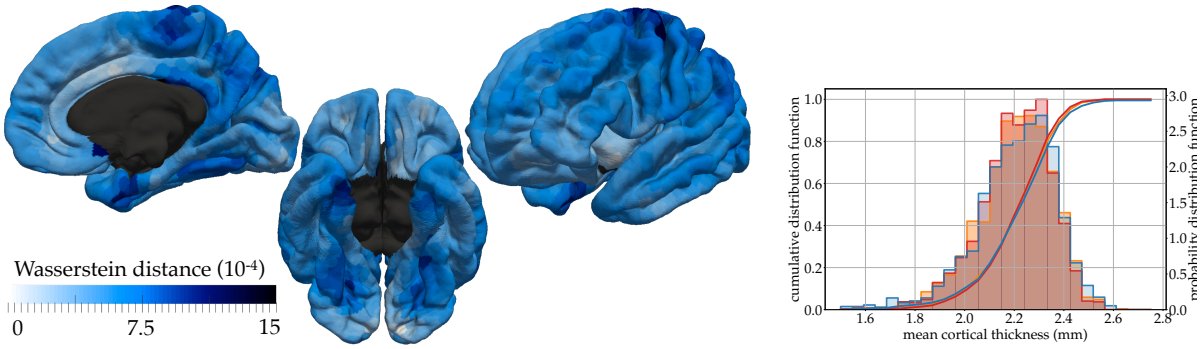
Analysis of factors modulating disease progression

We have just shown that the empirical distribution of the model parameters allows us to precisely reproduce the heterogeneity of the disease progression profiles. We are now in a position to examine how certain factors can explain this heterogeneity, or in other words, whether these factors determine particular progression patterns.

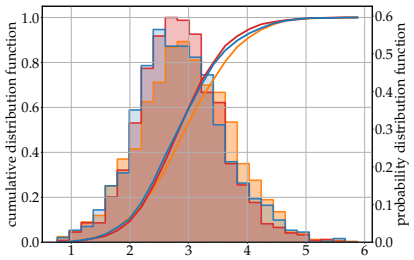
We recall that three parameters control the progression profile of the disease at the individual



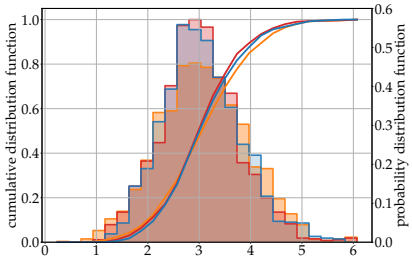
(a) FDG-PET images



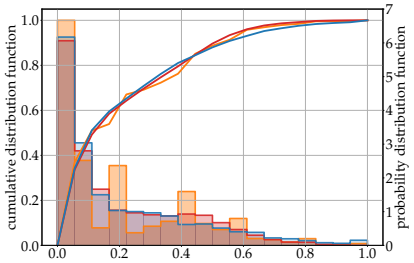
(b) Cortical Thickness maps



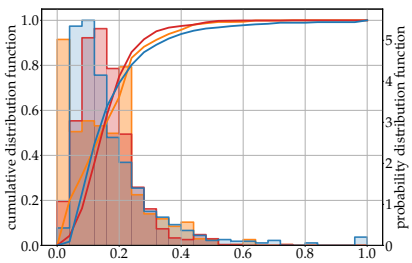
(c) Left hippocampus mesh



(d) Right hippocampus mesh



(e) ADAS-Cog



(f) MMSE

Figure 4: Statistics of simulated data. Superimposition of empirical distributions for simulated data (blue), reconstructed errors (red, as in Fig. 3) and real data (orange).

level: a time-shift that accounts for delay at onset, an acceleration factor that accounts for different pace of progression, and a multivariate space-shift that accounts for differences in the spatial pattern of the alterations or the delay between the decline of cognitive functions.

For each modality, we perform a multivariate linear regression between each of these parameters and a series of genetic, biological and environmental factors: sex, APOE-ε4 genotype, presence of amyloidosis, marital status and education level. We identify statistically significant association using a two tailed t-test at 5% significance level corrected for multiple comparisons with the false discovery rate method (see Methods). Note that in this study, we discard subjects without assessments of amyloidosis (see Extended Data Table 1 for corresponding number of samples).

We find that (see Table 1 for adjusted p-values and confidence intervals):

- no factor is associated with progression of brain glucose hypometabolism,
- atrophy of the hippocampus
  - progresses faster in women than in men by a factor 1.23 and 1.21 in left and right hemispheres respectively; starts earlier in women by 12.4 and 8.7 months for left and right hemispheres respectively; and exhibits a different pattern of deformation for men and women in both hemispheres (Extended Data Fig. 6);
  - progresses 1.22 times faster in the left hemisphere of the APOE-ε4 carriers, and arises earlier by 35.8 and 32.5 months for left and right hemispheres respectively;
  - progresses faster in amyloid-positive subjects by a factor 1.52 and 1.67 for left and

287 right hemispheres respectively;

288 – progresses 1.14 times faster in the left hemisphere of the married subjects; and starts
289 earlier by 42.5 and 36.3 months in the left and right hemispheres respectively; as com-
290 pared to non-married subjects;

291 – starts earlier in educated subjects by 3.73 and 6.97 months per year of education for
292 left and right hemispheres respectively;

293 • decrease in cortical thickness

294 – exhibits a different spatial pattern for men and women (Extended Data Fig. 7);

295 – occurs 1.42 times faster in APOE-ε4 carriers;

296 – exhibits a specific spatial pattern of thinning for amyloid positive subjects (Extended
297 Data Fig. 8);

298 • cognitive decline

299 – progresses 1.46 times faster in women and starts 36.8 months earlier than in men;

300 – progresses 1.25 times faster in APOE-ε4 carriers than in non-carriers;

301 – starts 21.9 months earlier for amyloid positive subjects than in amyloid negative sub-
302 jects;

303 – starts 32.6 months earlier for married subjects than non-married subjects.

			hypometabolism (FDG-PET)	hippocampus atrophy (MRI) left hemisphere	hippocampus atrophy (MRI) right hemisphere	cortical thinning (MRI)	cognitive decline (ADAS+MMSE)
genetic	gender female vs. male	speed factor		$\times 1.27$ CI=[1.11, 1.45] p=2.26e-3**	$\times 1.26$ CI=[1.08, 1.45] p=6.15-3**		$\times 1.46$ CI=[1.10, 1.92] p=8.42e-3 **
		time-shift		-33.6 CI=[-55.8, -11.6] p=3.71e-3**	-29.0 CI=[-53.0, -4.91] p=2.31e-2*		-36.8 CI=[-62.0, -11.6] p=4.48e-3 **
		space-shift		± 0.55 CI=[0.28, 0.82] p=4.00e-4***	± 0.60 CI=[0.34, 0.86] p=3.89e-5****	± 0.48 CI=[0.22, 0.75] p=2.24e-3 **	
	APOE-ε4 carrier vs. non-carrier	speed factor		$\times 1.17$ CI=[1.02, 1.33] p=2.77e-2*		$\times 1.42$ CI=[1.12, 1.82] p=2.17e-2*	$\times 1.25$ CI=[1.03, 1.51] 2.17e-2 *
		time-shift		-45.0 CI=[-66.9, -23.2] p=1.57e-4***	-36.8 CI=[-60.5, -13.0] p=4.27e-3**		
		space-shift					
biological	amyloid positive vs. negative	speed factor		$\times 1.18$ CI=[1.06, 1.32] p=8.20e-3**	$\times 1.23$ CI=[1.09, 1.39] p=4.03e-3**		
		time-shift					-21.9 CI=[-41.2,-2.5] p=2.70e-2 *
		space-shift				± 0.28 CI=[0.05, 0.50] p=2.24e-3 **	
environmental	marital married vs. non-married	speed factor		$\times 1.25$ CI=[1.07, 1.48] p=1.08e-2*			
		time-shift		-59.5 CI=[-86.6, -32.5] p=1.06e-4***	-52.7 CI=[-82.2, -23.2] p=1.28e-3**		-32.6 CI=[1.8, 63.3] p=3.78e-2*
		space-shift					
	education nb. of years of education	speed factor					
		time-shift		-6.04 CI=[-9.67, -2.42] p=1.95e-3**	-7.60 CI=[-11.55, -3.64] p=9.53e-4***		
		space-shift					

Table 1: Significant associations of individual parameters with genetic, biological and environmental factors: effect sizes, confidence intervals at 95%, and adjusted p-values. Only adjusted p-values below 5% significance level are shown. Time-shifts are in months, other quantities have no units. Directions of space-shift are not signed.

These results show that the alteration of brain metabolism progresses in an undifferentiated manner unlike atrophy that exhibits different spatiotemporal patterns according to the characteristics of the subjects. If both atrophy and hypometabolism are believed to reflect the accumulation of neuro-fibrillary tangles in the brain, then the relationship between tangles and hypometabolism must be quite different in nature and less sensitive to genetic and environmental factors than that between tangles and atrophy. Similarly, the hypothesis that atrophy may be the late consequence of hypometabolism cannot be reduced to a simple mechanical effect resulting from the progressive loss of neurons.

The absence of associations between cofactors and profiles of hypometabolism may be explained also by the fact that focal effects on specific brain areas may be diluted in non-specific regions of interest[38]. Previous findings showing associations are also likely to be due to the comparison of subjects at different ages or disease stages[38,39]. In this regard, it is interesting to notice that, except in four occasions, we found associations with parameters that modulate the dynamics of disease progression, not its trajectory. This fact suggests that previous findings showing association of these usual factors with the severity of atrophy, hypometabolism or cognitive decline are likely to be due to a non-proper temporal alignment of individual data.

There is a bilateral asymmetry in the hippocampus atrophy, with slightly more associations found in the left hemisphere. This fact is in line with previous findings suggesting that subjects with language impairment are more easily detected by clinical examination and neuro-psychological assessments, thus yielding to a higher prevalence of subjects with more pronounced atrophy in the

left hemisphere in clinical studies[40].

Our results also show the predominant role of genetic factors to explain the heterogeneity of the manifestation of the disease. In particular, disease progression presents a strong sexual dimorphism for hippocampus atrophy and cognitive decline. This question raises more and more attention in the scientific community, although its consequences for clinical trials and care have not yet been drawn[41–44]. The accelerated and earlier atrophy in women translates into an accelerated and even earlier cognitive decline. This dimorphism does not seem to be alleviated by compensatory mechanisms. By contrast, APOE-ε4 carriers also exhibit earlier and more pronounced alterations of their hippocampus, but this effect is, to some extend, alleviated in the onset of cognitive decline, which does not occur earlier than non-carriers, but still at a greater pace. It is as if brain plasticity is able to compensate for the advance of almost 3 years in hippocampal atrophy, but that once the compensation is made, cognitive decline still manifests itself at a faster rate than in subjects without the mutation.

Independently of disease progression, we found a sex dimorphism in the shape of hippocampus in both hemispheres (significant space shifts in Table 1). The position of the hippocampus presents a greater angle of rotation with respect to the brain stem in women, which makes it more forward-facing than in men (Extended Data Fig. 6). Sex differences are also found in the spatial pattern of cortical thickness with a more pronounced bilateral asymmetry in women than in men. These differences are the consequences of the well-known dimorphism in brain development, and are independent of disease progression. No such differences in brain structure are found for other

co-factors.

The presence or absence of amyloid plaques in the development of the disease tends to change the spatial patterns of cortical atrophy, which may be due to accumulation of plaques in specific brain areas (Extended Data Fig. 8). However, the pattern of cognitive decline is similar. The current definition of the disease makes the diagnostic of Alzheimer's disease inappropriate in subjects without amyloidosis[45]. The similar cognitive presentation of these subjects may explain the difficulty for clinicians to distinguish between cases. Subjects without amyloidosis and diagnosed with Alzheimer's disease show a later onset of cognitive decline, which occurs nearly 2 years after (21.9 months, CI=[2.5,41.2]) the general case of amyloid positive subjects. Older subjects may be more difficult to diagnose with more overlapping symptoms and co-morbidities[46].

Marital status seems to be the environmental factor having the strongest effect on disease progression. Married subjects tend to experience an alteration of their hippocampus more than 4 years earlier than divorced, widowed or never married subjects. This delay at onset is reduced to 2.6 years at the cognitive level. Compensation effects are not surprising as the marital status is likely to be linked with social habits that may be associated with disease progression[47]. Further interpretation is difficult since this status covers very heterogeneous individual situations.

Finally, we compute the co-variations among the individual parameters: time-shifts, acceleration factors, space-shifts and age at baseline (see Extended Data Fig 9). These co-variations present a consistent pattern for all modalities. Age at baseline strongly correlates with time-shift, showing that subjects were included in the study at similar disease stages. Time-shift and speed

364 factors anti-correlate, showing that early onset individuals tend to progress faster, and conversely
365 that late onset individuals tend to progress slower[48,49]. Space-shift correlates with age at baseline,
366 notably for the hippocampus atrophy, suggesting that early onset individuals also present a specific
367 pattern of atrophy than late onset individuals.

368 **Conclusion**

369 The digital brain models we have built provide, for the first time, a comprehensive view of how
370 structural and metabolic alterations propagate in the brain, both in space and time, and how they
371 relate to specific sequences of decline in cognitive functions. These models not only show the
372 typical trajectory of disease progression, but also allow the heterogeneity of this progression to be
373 accurately reproduced. In this way, they offer us an incomparable way to identify the factors that
374 influence this progression at the individual level, and to show how they modulate it. We were able
375 to highlight the strong sexual dimorphism in the rate and precocity at which the disease progresses,
376 as well as probable compensation mechanisms in carriers of some genetic risk factors.

377 We have shown that these models form simulators for multi-modal images and neuro-psychological
378 assessments of virtual patients whose characteristics reproduce those of the patients observed in
379 the training cohort. From now on, it will no longer be necessary to make voluminous medical data
380 repositories available, which is always difficult from a technical and regulatory point of view. It
381 will be sufficient to share a simple computer code that will be able to regenerate the cohort in any
382 laboratory, and even increase the number of samples, homogenize the frequency of observations

of subjects, or even correct some bias in the composition of the cohort.

If calibrated on a few data from a new patient, these simulations reconstruct the patient's trajectory with the same precision as the uncertainty of the observations themselves. These personalized simulations may thus serve to predict the future state of the subject's brain and cognitive functions, measure related risks, and tomorrow measure and predict the effect of a treatment. They represent therefore a decisive step towards the advent of precision medicine in neurology.

References

1. Silver, D. *et al.* Mastering the game of go without human knowledge. *Nature* **550**, 354–359 (2017).

2. Khanal, B., Lorenzi, M., Ayache, N. & Pennec, X. A biophysical model of brain deformation to simulate and analyze longitudinal mris of patients with alzheimer's disease. *NeuroImage* **134**, 35 – 52 (2016).

3. Fonteijn, H. M. *et al.* An event-based model for disease progression and its application in familial, alzheimer's disease and huntington's disease. *NeuroImage* **60**, 1880–1889 (2012).

4. Jedynak, B. M. *et al.* A computational neurodegenerative disease progression score: method and results with the alzheimer's disease neuroimaging initiative cohort. *Neuroimage* **63**, 1478–1486 (2012).

5. Villemagne, V. L. *et al.* Amyloid deposition, neurodegeneration, and cognitive decline in

sporadic alzheimer's disease: a prospective cohort study. *The Lancet Neurology* **12**, 357 – 367 (2013). URL http://www.sciencedirect.com/science/article/pii/S1474442213700449.

6. Donohue, M. C. *et al.* Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **10**, S400–410 (2014).

7. Zhang, X. *et al.* Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in alzheimer's disease. *Proceedings of the National Academy of Sciences* **113**, E6535–E6544 (2016). URL http://www.pnas.org/content/113/42/E6535. http://www.pnas.org/content/113/42/E6535.full.pdf.

8. Guerrero, R. *et al.* Instantiated mixed effects modeling of alzheimer's disease markers. *NeuroImage* **142**, 113 – 125 (2016). URL http://www.sciencedirect.com/science/article/pii/S1053811916302981.

9. Khanna, S. *et al.* Using multi-scale genetic, neuroimaging and clinical data for predicting alzheimers disease and reconstruction of relevant biological mechanisms. *Scientific Reports* **8** (2018).

10. Schiratti, J.-B., Allassonnière, S., Colliot, O. & Durrleman, S. Learning spatiotemporal trajectories from manifold-valued longitudinal data. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R. (eds.) *NIPS 28*, 2404–2412 (Curran Associates, Inc., 2015).

11. Schiratti, J.-B., Allassonnière, S., Colliot, O. & Durrleman, S. A bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. *The Journal of Machine Learning Research* **18**, 4840–4872 (2017).

12. Folstein, M. F., Folstein, S. E. & McHugh, P. R. mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research* **12**, 189–198 (1975).

13. Rosen, W. G., Mohs, R. C. & Davis, K. L. A new rating scale for alzheimer's disease. *The American journal of psychiatry* (1984).

14. Mohs, R. C. *et al.* Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the alzheimer's disease assessment scale that broaden its scope. *Alzheimer disease and associated disorders* (1997).

15. Koval, I. *et al.* Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 451–459 (Springer, 2017).

16. Durrleman, S. *et al.* Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage* (2014).

17. Durrleman, S. Geometrical approaches in statistical learning for the construction of digital models of the human brain. Habilitation à diriger des recherches, Pierre and Marie Curie University, Paris (2018). URL `https://who.rocq.inria.fr/Stanley.Durrleman/papers/Durrleman_hdr_lr.pdf`.

18. Bône, A., Colliot, O. & Durrleman, S. Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 9271–9280 (2018).

19. Louis, M., Charlier, B., Jusselin, P., Pal, S. & Durrleman, S. A fanning scheme for the parallel transport along geodesics on riemannian manifolds. *SIAM Journal on Numerical Analysis* **56**, 2563–2584 (2018).

20. Allassonnière, S., Durrleman, S. & Kuhn, E. Bayesian mixed effect atlas estimation with a diffeomorphic deformation model. *SIAM Journal on Imaging Science* **8**, 13671395 (2015).

21. Kuhn, E. & Lavielle, M. Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM: Probability and Statistics* **8**, 115–131 (2004).

22. Mosconi, L. Brain glucose metabolism in the early and specific diagnosis of alzheimers disease. *European journal of nuclear medicine and molecular imaging* **32**, 486–510 (2005).

23. Chen, K. *et al.* Twelve-month metabolic declines in probable alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the alzheimer's disease neuroimaging initiative. *Neuroimage* **51**, 654–664 (2010).

24. Pagani, M. *et al.* Early identification of mci converting to ad: a fdg pet study. *European Journal of Nuclear Medicine and Molecular Imaging* **44**, 2042–2052 (2017).

458 25. Drzezga, A. *et al.* Cerebral metabolic changes accompanying conversion of mild cognitive
459 impairment into alzheimer's disease: a pet follow-up study. *European journal of nuclear*
460 *medicine and molecular imaging* **30**, 1104–1113 (2003).

461 26. Mosconi, L. *et al.* Hippocampal hypometabolism predicts cognitive decline from normal
462 aging. *Neurobiology of aging* **29**, 676–692 (2008).

463 27. Hyman, B. T., Van Hoesen, G. W., Damasio, A. R. & Barnes, C. L. Alzheimer's disease:
464 cell-specific pathology isolates the hippocampal formation. *Science* **225**, 1168–1170 (1984).

465 28. Gómez-Isla, T. *et al.* Profound loss of layer ii entorhinal cortex neurons occurs in very mild
466 alzheimers disease. *Journal of Neuroscience* **16**, 4491–4500 (1996).

467 29. Greene, S. J., Killiany, R. J., Initiative, A. D. N. *et al.* Subregions of the inferior parietal lobule
468 are affected in the progression to alzheimer's disease. *Neurobiology of aging* **31**, 1304–1311
469 (2010).

470 30. Chan, D. *et al.* Patterns of temporal lobe atrophy in semantic dementia and alzheimer's disease.
471 *Annals of neurology* **49**, 433–442 (2001).

472 31. Jacobs, H. I., Van Boxtel, M. P., Jolles, J., Verhey, F. R. & Uylings, H. B. Parietal cortex
473 matters in alzheimer's disease: an overview of structural, functional and metabolic findings.
474 *Neuroscience & Biobehavioral Reviews* **36**, 297–309 (2012).

475 32. Suva, D. *et al.* Primary motor cortex involvement in alzheimer disease. *Journal of neu-*
476 *ropathology and experimental neurology* **58**, 1125–1134 (1999).

33. Apostolova, L. G. *et al.* 3d comparison of hippocampal atrophy in amnestic mild cognitive impairment and alzheimer's disease. *Brain* **129**, 2867–2873 (2006). URL http://dx.doi.org/10.1093/brain/awl274.

34. Frisoni, G. B. *et al.* Mapping local hippocampal changes in alzheimer's disease and normal ageing with mri at 3 tesla. *Brain* **131**, 3266–3276 (2008). URL http://dx.doi.org/10.1093/brain/awn280. /oup/backfile/content_public/journal/brain/131/12/10.1093/brain/awn280/2/awn280.pdf.

35. Skinner, J. *et al.* The alzheimers disease assessment scale-cognitive-plus (adas-cog-plus): an expansion of the adas-cog to improve responsiveness in mci. *Brain imaging and behavior* **6**, 489–501 (2012).

36. Raghavan, N. *et al.* The adas-cog revisited: novel composite scales based on adas-cog to improve efficiency in mci and early ad trials. *Alzheimer's & Dementia* **9**, S21–S31 (2013).

37. Vaillant, M. & Glaunès, J. Surface matching via currents. In *Biennial International Conference on Information Processing in Medical Imaging*, 381–392 (Springer, 2005).

38. Knopman, D. S. *et al.* 18f-fluorodeoxyglucose positron emission tomography, aging, and apolipoprotein e genotype in cognitively normal persons. *Neurobiology of aging* **35**, 2096–2106 (2014).

39. Jack, C. R. *et al.* Age, sex, and apoe $\varepsilon 4$ effects on memory, brain structure, and β-amyloid across the adult life span. *JAMA neurology* **72**, 511–519 (2015).

40. Wolf, H. *et al.* Hippocampal volume discriminates between normal cognition; questionable and mild dementia in the elderly. *Neurobiology of Aging* **22**, 177 – 186 (2001). URL http://www.sciencedirect.com/science/article/pii/S0197458000002384.

41. Mielke, M. M., Vemuri, P. & Rocca, W. A. Clinical epidemiology of alzheimers disease: assessing sex and gender differences. *Clinical Epidemiology* **6**, 1179–1349 (2014).

42. Fisher, D. W., Bennett, D. A. & Dong, H. Sexual dimorphism in predisposition to alzheimer's disease. *Neurobiology of Aging* **70**, 308 – 324 (2018). URL http://www.sciencedirect.com/science/article/pii/S019745801830126X.

43. Mielke, M. M., Ferretti, M. T., Iulita, M. F., Hayden, K. & Khachaturian, A. S. Sex and gender in alzheimer's disease; does it matter? *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **14**, 1101–1103 (2018).

44. Cavedo, E. *et al.* Sex differences in functional and molecular neuroimaging biomarkers of alzheimer's disease in cognitively normal older adults with subjective memory complaints. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **14**, 1204 – 1215 (2018).

45. Dubois, B. *et al.* Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *The Lancet Neurology* **6**, 734–746 (2007).

46. Nelson, P. T. *et al.* new old pathologies: Ad, part, and cerebral age-related tdp-43 with sclerosis (carts). *Journal of Neuropathology & Experimental Neurology* **75**, 482–498 (2016).

514 47. Wajman, J. R., Mansur, L. L. & Yassuda, M. S. Lifestyle patterns as a modifiable risk factor for
515 late-life cognitive decline: A narrative review regarding dementia prevention. *Current Aging*
516 *Science* **11** (2018).

517 48. C, B., J, C., R, R., X, S. & P, A. Age and rate of cognitive decline in alzheimer's disease:
518 Implications for clinical trials. *Archives of Neurology* **69**, 901–905 (2012). URL +http:
519 //dx.doi.org/10.1001/archneurol.2011.3758.

520 49. Holland, D., Desikan, R. S., Dale, A. M., McEvoy, L. K. & for the Alzheimer's Disease
521 Neuroimaging Initiative. Rates of decline in alzheimer disease decrease with age. *PLOS ONE*
522 **7**, 1–12 (2012). URL https://doi.org/10.1371/journal.pone.0042325.

523 50. Clark, C. M. *et al.* Cerebral pet with florbetapir compared with neuropathology at autopsy for
524 detection of neuritic amyloid-β plaques: a prospective cohort study. *The Lancet Neurology*
525 **11**, 669–678 (2012).

526 51. Landau, S. M. *et al.* Amyloid-β imaging with pittsburgh compound b and florbetapir: compar-
527 ing radiotracers and quantification methods. *Journal of nuclear medicine: official publication,*
528 *Society of Nuclear Medicine* **54**, 70 (2013).

529 52. Schindler, S. E. *et al.* Cerebrospinal fluid biomarkers measured by elecsys assays compared
530 to amyloid imaging. *Alzheimer's & Dementia* (2018).

531 53. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in spm using a
532 macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* **15**,
533 273–289 (2002).

534 54. Rolls, E. T., Joliot, M. & Tzourio-Mazoyer, N. Implementation of a new parcellation of the
535 orbitofrontal cortex in the automated anatomical labeling atlas. *Neuroimage* **122**, 1–5 (2015).

536 55. Routier, A. *et al.* Clinica: an open source software platform for reproducible clinical neuro-
537 science studies. In *Annual meeting of the Organization for Human Brain Mapping-OHBM*
538 *2018* (2018).

539 56. Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J. & Nichols, T. E. *Statistical para-*
540 *metric mapping: the analysis of functional brain images* (Elsevier, 2011).

541 57. Samper-González, J. *et al.* Reproducible evaluation of classification methods in alzheimer's
542 disease: framework and application to mri and pet data. *bioRxiv* 274324 (2018).

543 58. Fischl, B. & Dale, A. M. Measuring the thickness of the human cerebral cortex from magnetic
544 resonance images. *Proceedings of the National Academy of Sciences* **97**, 11050–11055 (2000).

545 59. Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures
546 in the human brain. *Neuron* **33**, 341–355 (2002).

547 60. Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation
548 for unbiased longitudinal image analysis. *Neuroimage* **61**, 1402–1418 (2012).

549 61. Woolrich, M. W. *et al.* Bayesian analysis of neuroimaging data in fsl. *Neuroimage* **45**, S173–
550 S186 (2009).

551 62. Ahrens, J., Geveci, B. & Law, C. Paraview: An end-user tool for large data visualization. *The*
552 *visualization handbook* **717** (2005).

553 63. Jian, B. & Vemuri, B. C. Robust point set registration using gaussian mixture models. *IEEE*
554 *transactions on pattern analysis and machine intelligence* **33**, 1633–1645 (2011).

555 64. Gori, P. *et al.* A Bayesian Framework for Joint Morphometry of Surface and Curve meshes
556 in Multi-Object Complexes. *Medical Image Analysis* **35**, 458–474 (2017). URL `https:`
557 `//hal.inria.fr/hal-01359423`.

558 65. Durrleman, S., Pennec, X., Trouvé, A., Thompson, P. & Ayache, N. Inferring brain variability
559 from diffeomorphic deformations of currents: an integrative approach. *Medical image analysis*
560 **12**, 626–637 (2008).

561 66. Allassonnière, S., Kuhn, E. & Trouvé, A. Construction of bayesian deformable models via a
562 stochastic approximation algorithm: a convergence study. *Bernoulli* **16**, 641–678 (2010).

563 67. Fishbaugh, J., Prastawa, M., Gerig, G. & Durrleman, S. Geodesic regression of image and
564 shape data for improved modeling of 4D trajectories. In *ISBI 2014 - 11th International Sym-*
565 *posium on Biomedical Imaging*, 385 – 388 (2014).

566 68. Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. A limited memory algorithm for bound constrained
567 optimization. *SIAM Journal on Scientific Computing* **16**, 1190–1208 (1995).

568 69. Jack Jr, C. R. *et al.* The alzheimer's disease neuroimaging initiative (adni): Mri methods.
569 *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for*
570 *Magnetic Resonance in Medicine* **27**, 685–691 (2008).

571 70. Jack Jr, C. R. *et al.* Update on the magnetic resonance imaging core of the alzheimer's disease
572 neuroimaging initiative. *Alzheimer's & Dementia* **6**, 212–220 (2010).

71. Clark, C. M. *et al.* Variability in annual mini-mental state examination score in patients with probable alzheimer disease: a clinical perspective of data from the consortium to establish a registry for alzheimer's disease. *Archives of neurology* **56**, 857–862 (1999).

72. Hensel, A., Angermeyer, M. C. & Riedel-Heller, S. G. Measuring cognitive change in older adults: reliable change indices for the mmse. *Journal of Neurology, Neurosurgery & Psychiatry* (2007).

73. Standish, T. I. *et al.* Improved reliability of the standardized alzheimer's disease assessment scale (sadas) compared with the alzheimer's disease assessment scale (adas). *Journal of the American Geriatrics Society* **44**, 712–716 (1996).

74. Abdi, H. Partial least square regression (pls regression). *Encyclopedia for research methods for the social sciences* **6**, 792–795 (2003).

Alexandre Bône and Igor Koval These authors contributed equally to the work.

Acknowledgements This work has been partly funded by the European Research Council (ERC) under grant agreement No 678304, European Unions Horizon 2020 research and innovation program under grant agreement No 666992, and the program Investissements davenir ANR-10-IAIHU-06. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimers Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Bio-

gen; Bristol-Myers Squibb Company; CereSpir, Inc.;Cogstate;Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies;Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Contributions I.K, A.B., M.L. performed the research, S.B., A.M., J.S. managed and pre-processed data, N.B., B.C., A.B.,S.E., O.C., S.A., S.D. designed the research, I.K., A.B., N.B., S.E., O.C., S.A, S.D wrote the paper.

Competing Interests A patent has been filed by INSERM Transfer under the reference PCT/IB2016/052699 and is currently under investigation (inventors: J.-B. Schiratti, S. Allassonnière, O. Colliot, S. Durrleman). It aims to cover the uses of the presented work for predicting subject's progression, assessing individual risks, predicting diagnosis and symptom onset, identifying therapeutic target and biomarkers, and screening a compound. Authors declare that they have no other competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to S.D. (email: stan-

ley.durrleman@inria.fr).

Methods

Data Set Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[a]. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

We used all available visits from ADNI, ADNI-GO and ADNI-2 data sets for all subjects who:

- have been diagnosed with Alzheimer's Disease (AD) at least at one visit;
- have been diagnosed as Mild Cognitive Impaired (MCI) subjects at least at one visit;
- did not revert to Cognitively Normal (CN) stage after being diagnosed as MCI or AD, nor revert to MCI or CN stage after being diagnosed with AD.

350 subjects satisfied the first two criteria. The third criterion excludes subjects with doubtful diagnoses: 28 subjects were then excluded, leading to a subset of 322 subjects representing a total

[a]http://adni.loni.usc.edu/

of 2136 visits. We define 3 overlapping sub-sets by selecting different data types: ADAS-Cog & MMSE, FDG-PET images and MRI images. Table 1 provides summary statistics of these data sets.

For each subject, we used the following additional data: age at each visit, sex, marital status, educational level, Apolipoprotein E (ApoE) polymorphism, and presence of amyloidosis. More precisely, we define:

- marital status as: married versus non-married meaning widowed, divorced, or never married;
- educational level as the number of years of education;
- ApoE-ε4 carriership as the presence of at least one allele ε4 of the ApoE gene;
- Amyloid status as positive if one of these conditions was met at one visit at least:
  - a Standard Updake Value ratio (SUVR), normalized by the entire cerebellum, greater than 1.1 in a PET image acquired with Florbetapir (AV-45) compound[50,51];
  - an average SUVR, normalized by the cerebellum, greater than 1.47 in a PET image with a Pittsburgh compound B (PiB)[51];
  - a level of beta amyloid 1-42 (Aβ42) (measured with the Roche Elecsys assays[b]) in the cerebrospinal fluid (CSF) lower than 1098 pg/mL[52];

  unknown if no values of CSF biomarkers and no AV45 or PiB PET images were available at any visit in the ADNI-merge file; and negative otherwise.

[b] http://adni.loni.usc.edu/new-csf-a%CE%B21-42-t-tau-and-p-tau181-biomarkers-results-from-adni-biomarker-core-using-elecsys/

Not counting 7% of the population with an unknown amyloid status, 83% of the remaining held a stable positive status status across all their visits, while 9% have their visits consistently negative – the last 8% present an evolution of its status through time. The stable positive and negative individuals allows to distinguish the subjects who have developed Alzheimer's Disease in presence of amyloidosis, from those who developed the clinical signs of the disease without the significant development of amyloid plaques.

Pre-processing and feature extraction We used the global MMSE score and aggregated scores from the 13 items of the ADAS-Cog. Furthermore, we pooled the 13 items into four sub-categories: memory by adding items 1, 4, 7, 8 and 9, language by adding items 2, 5, 10, 11 and 12,praxis by adding items 3 and 6, and concentration with item 13. Each value is normalized by the maximum possible value for the global score or for each category.

Regional FDG-PET SUVR were extracted using the second version of the Automated Anatomical Atlas[c] (AAL2)[53,54] with 120 regions covering the cortex and the main subcortical structures, using the open-source community software Clinica[d] [55]. The software performs intra-subject registration of the FDG-PET image into the space of the subject's T1-weighted MRI image using Statistical Parametric Mapping[e] (SPM) software (version 12)[56]. The PET image is then spatially normalized into MNI space using DARTEL deformation model of SPM, and its intensities normalized using the average uptake value in the pons as reference region. The SUVR map is obtained by averaging resulting intensities in each region of the atlas [57].

[c] http://www.gin.cnrs.fr/fr/outils/aal-aal2/
[d] http://clinica.run/doc/Pipelines/PET_Volume
[e] www.fil.ion.ucl.ac.uk/spm/

The MRI images were first processed independently with the cross-sectional pipeline of the FreeSurfer[f] software (version 5.3.0) [58,59]. The longitudinal FressSurfer pipeline is then used to create subject-specific templates from the successive data of each subject and refine image segmentations[60]. These segmented images are used then to extract a cortical thickness map, and a mesh of the left and right hippocampus.

We used the cortical surface mesh projected onto the average space called FSaverage with 163,842 vertices. For dimensionality reduction purposes, we then

- inflate the FSAverage mesh to a sphere using FreeSurfer, on which 3,658 vertices (called patch-nodes) are selected to map the whole sphere uniformly,
- associate each vertex to its closest patch-node, resulting in a parcellation of the cortical mesh into 3,658 patches that are uniformly distributed over the surface, where a patch contains on average 44 vertices,
- compute the average value of the cortical thickness in each patch.

We also align the skull-stripped images with an affine 12-degrees-of-freedom transformation onto the Colin27 template brain[g], using the FSL 5.0 software[h61]. Mesh representations of the geometry of the left and right hippocampus result from the following steps:

- the volumetric segmentations of the hippocampi obtained by FreeSurfer are transformed into

[f] https://surfer.nmr.mgh.harvard.edu
[g] http://www.bic.mni.mcgill.ca/ServicesAtlases/Colin27
[h] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/

meshes using the aseg2srf software[i],

- the resulting meshes are decimated by a 88% factor using Paraview, 5.4.1[j62],
- then aligned using the previously-computed global affine transformation estimated with the FSL software,
- residual pose differences among subjects are then removed by rigidly aligning the meshes from the baseline image of each subject to the corresponding hippocampus mesh in the Colin27 atlas image, this transformation with 6 degrees of freedom being computed with the GMMReg software[k63],
- the same transformation is eventually used to align the meshes from the follow-up images of the same subject.

Data representation and choice of Riemannian metrics The statistical model may be written as:

$$y_{ij} = \eta^{w_i}(\gamma_0)(\psi_i(t_{ij})) + \varepsilon_{ij} \tag{1}$$

where

- $\gamma_0 : t \to \mathrm{Exp}_{p_0}((t - t_0)v_0)$ is the population average trajectory in the form of a the geodesic passing at point p_0 with velocity v_0 at time t_0 (Exp denotes the Riemannian exponential as a concise way to write geodesics),

[i] https://brainder.org (version of July 2009)
[j] www.paraview.org
[k] https://github.com/bing-jian/gmmreg (version of July 2008)

- $\eta^{w_i}(\gamma_0) : t \rightarrow \mathrm{Exp}_{\gamma_0(t)}\left(P_{\gamma_0}^{t_0,t}(w_i)\right)$ is the exp-parallelisation of the geodesic γ_0 in the subject-specific direction w_i, called space-shift, as depicted in Fig. 1 ($P_{\gamma_0}^{t_0,t}(w_i)$ denotes the parallel transport of the vector w_i along the curve γ_0 from $\gamma_0(t_0)$ to $\gamma_0(t)$),

- $\psi_i : t \rightarrow \alpha_i(t - t_0 - \tau_i) + t_0$ is a time-reparameterizing function, where α_i is a subject-specific acceleration factor and τ_i a subject-specific time-shift.

For identifiability purposes, we impose the vectors w_i to be orthogonal to the velocity v_0 in the tangent-space at point p_0. Parallel transport being isometric, this property then holds at any time point. The random effects of the model are:

- an acceleration factor α_i, which accounts for the variations in pace of disease progression, and therefore distinguishes the fast from the slow progressing individuals,

- a time-shift τ_i, which accounts for the variations in age at onset, and therefore distinguishes the early from the late onset individuals,

- a space-shift w_i (a vector pointing a direction on the manifold), which accounts for the variations in the position of the individual trajectory, and therefore captures differences in patterns of disease progression (magnitude of the effects, re-ordering of events, change in the spatial pattern of alterations for instance, as detailled below).

Their prior distributions are a log-normal distribution for the acceleration factors, zero-mean Gaussian distribution for the time-shift. Space-shifts are decomposed into a series of independent components: $w_i = As_i$ where the columns of A contains a pre-defined number of vectors in the

orthogonal space of v_0, called components, and s_i are random weights, called sources and distributed according to a normal distribution for non-Euclidean metrics and a Laplace distribution if the manifold is Euclidean, for identifiability purposes.

We concatenated the aggregated MMSE score and the four sub-categories of the ADAS-Cog to build a 5-dimensional feature vector, which is seen as a point in a 5-dimensional hyper-cube $[0,1]^5$. We provide this manifold with a diagonal metric tensor which ensures that a geodesic in this hyper-cube is formed by 5 logistic curves, that are further assumed to be parallel to each others: $\gamma_{0,k}(t) = \gamma_{\text{logit}}(t+\delta_k)$ with $\gamma_{\text{logit}}(t) = \left(1+\frac{1-p_0}{p_0}\exp\left(\frac{-v_0(t-t_0)}{p_0(1-p_0)}\right)\right)^{-1}$. A parallel shift of the population geodesics in this hyper-cube translates into a change in the temporal delay between the logistics curves of each coordinate[10,11]: $\eta_k^{w_i}(\gamma_0)(t) = \gamma_{\text{logit}}\left(t+\delta_k+\frac{w_{i,k}}{\dot{\gamma}_{\text{logit}}(t_0+\delta_k)}\right)$.

Maps of cortical thickness take the form of a vector of 3,658 coordinates corresponding to the measurements values at every patch node, seen as a point in the Euclidean space $\mathbb{R}^{3,658}$. Geodesics are straight-lines in this space, where each coordinate $k \in \{1,\ldots,3,658\}$ is a one-dimensional straight-line of the form: $\gamma_k = p_k + v_k(t-t_0)$. The exp-parallelisation in the Euclidean space corresponds simply to a translation, so that each coordinate is transformed into[15]: $\eta_k^{w_i}(\gamma_0) = p_k + w_{i,k} + v_k(t-t_0)$. The fixed-effects p_0 and v_0 are vectors of size 3,658 whose k-th coordinate p_k and v_k are the reference intercept and slope at the k-th patch respectively. We select a sub-set of 911 control nodes $(c_i)_{1\leq i\leq 911}$ among the patch nodes, and create a mapping which generates 3,658 values from the 911 values using a manifold-kernel smoothing interpolation. Let the k-th path node be $x_k \in \mathbb{R}^3$, corresponding to the Euclidean coordinate of the center of the path. The value $p_k = p(x_k) = \sum_{i=1}^{911}\exp\left(-\frac{d(x_k,c_i)^2}{\sigma^2}\right)\beta_i$ corresponds to the value of the parameter at the k-th node.

The β_i are the 911 values at the control nodes c_i, the distance $d(x_k, c_i)$ is the geodesic distance on the cortical surface mesh between patch node x_k and control nodes c_i, and σ is a scalar parameter taken equal to $20\,mm$, which is approximately 2.5 times the average distance between neighbors control nodes (namely the three closest control nodes to a given control node). The same kernel mapping is used to generate the values $(v_k)_{1\leq k\leq 3{,}658}$. By construction, the maps generated by this operation are varying smoothly over the surface mesh and are controlled by a smaller number of parameters.

Each PET measurement is characterized by a vector in $\mathbb{R}^{120}$ whose k-th coordinate corresponds to the the average SUVR value on the k-th region of interest (ROI) of the AAL2 atlas. We take the same approach as for the cortical thickness maps. The centroids of the regions in the AAL2 anatomical atlas is considered as a fully connected graph (so that the geodesic distance on the graph is the Euclidean distance between centroids), and all centroids are taken as control nodes. Spatial smoothing parameter is taken here of $\sigma = 15$ voxels $= 22.5\,mm$.

For hippocampus meshes, we consider a finite-dimensional manifold of diffeomorphisms of the ambient 3D space that contains the hippocampus[16,17]. This manifold is parameterized by a set of momentum vectors $(m_k)_k$ attached to a set of control points $(c_k)_k$. This set of control points is seen as a dynamic system of particles which follows geodesics derived from the Hamiltonian: $H(c,m) = \sum_{k,l} \exp\left(-\frac{||c_k - c_l||^2}{\sigma^2}\right) m_k^T m_l$, where T denotes the transpose of a vector. The exponential function is a positive definite kernel defining the co-metric on this manifold as the matrix $K(c) = \left[\exp\left(-\frac{||c_i - c_j||^2}{\sigma^2}\right)\right]_{i,j}$. The deformation scale σ is an hyperparameter of this metric, and is set to $10\,mm$ in this application. For each configuration of control point $c(t)$ and momentum vec-

tor $m(t)$ at time-point t, we derive a continuous vector field $v_t(x) = \sum_k \exp\left(-\frac{||c_k(t)-x||^2}{\sigma^2}\right) m_k(t)$ for any point x. The trajectory of a set of control points and attached momenta therefore translate into a time-dependent family of vector fields. These vector fields are integrated in time from the identity map into a flow of diffeomorphisms. Diffeomorphisms along these geodesics are applied to a template shape $\mathcal{O}$ to give a smooth trajectory of shape deformation: $t \rightarrow \phi^{c,m}(t)\,(\mathcal{O})$, where we denote by $\phi^{c,m}(t)$ the diffeomorphism arising from control points c, momentum vectors m at time-point t. The set of control points and the template shape play the role of the point p_0, and momentum vectors the role of the cotangent-space vector $K(c)^{-1}v_0$.

This construction allows the exp-parallelisation of the trajectory of control points in the manifold, which translates into another trajectory of shape $\eta^{w_i}(\phi^{c,m})(t)(\mathcal{O})$. This parallel trajectory transports the deformation patterns of the baseline geodesics into a new geometry[18].

In this construction, the template shape $\mathcal{O}$ becomes a new fixed-effect of the statistical model. We use the metric on currents[37] to measure a distance between the deformed template and the observations, which are meshes with different topology and number of vertices. This distance appears when maximizing the likelihood of the residual noise ε_{ij}[17,64]. It is homogeneous to an area, and its units is therefore in mm^2. One of its main advantage is that it smooths out small protrusions and is insensitive to small holes or topology changes in the meshes, making it robust to segmentation errors and avoiding intensive mesh pre-processing. The scale at which the metric is insensitive to these artifacts is an hyperparameter of this attachment metric[64,65], and is set to $5\,mm$ in this work.

Calibration We use the Monte-Carlo Markov Chain Stochastic Approximation Expectation Maximization (MCMC-SAEM) algorithm [20,21,66] to calibrate the model. It is an iterative algorithm that solves the following approximate optimization problem at each iteration:

$$\theta_{k+1} = \text{argmax}_\theta \sum_{i=1}^{N} \int \log \left[p(\{y_{ij}\}_j, z_i; \theta)\right] p(z_i|\{y_{ij}\}_j; \theta_k) dz_i \tag{2}$$

At each iteration, it loops over the three following steps.

- simulation of candidate value of the random-effects z_k by running several steps of a Metropolis-Hasting method within a block Gibbs sampler with $p(z|\{y_{ij}\}_j, \theta_k)$ as ergodic distribution. This step essentially draws a candidate from a random walk sampler and accept or reject this candidate depending notably on the value of the complete likelihood $p(\{y_{ij}\}_j, z_k, \theta_k)$, which measures how well the data generated with the candidate z_k, i.e. $f(\theta_k, z_k, \{t_{ij}\}_j)$, resembles the actual observations $\{y_{ij}\}_j$.

- stochastic approximation using a Robbins-Monro method which keeps adding the terms within the integral with decreasing gains. For distributions belonging to the curved-exponential family (which is ensured in all cases but hippocampus by assuming parameters to be drawn from a prior distribution), it amounts to keep track of a set of sufficient statistics.

- maximization over the parameters, which is done by updating the parameters with a fixed number of gradient descent steps for hippocampus meshes, or in closed form in other cases.

The following procedures are preceded for the initialization of the algorithm. For the hip-

pocampus meshes, an average model was first computed by estimating an atlas [64] to initialize the template shape and the matrix A, individual geodesic regressions [67] were then estimated to initialize the velocity vector v_0. For the cortical thickness and SUVR maps, the coordinates p_k of the initial position p_0 corresponds to the mean value over all the data on the corresponding region. As for the initial velocity v_0, each coordinate v_k corresponds to the average slope of linear regressions performed on each subject independently. In the case of the cognitive scores, a random initialization was used.

The implementation of this algorithm is available in the software Deformetrica[l] for the longitudinal shape model, and in the Leasp software[m] for the other cases.

Model synchronization. The time-warp functions $\psi_i^{[m]}(t_{ij})$ maps the age of the i-th subject at the j-th visit, t_{ij} to a disease stage on the normative time-line for the data type m. Taking the model of cognitive decline as a reference ($m = \text{cog}$), we look for a temporal mapping $\Phi^{[m]}(t) = \lambda^{[m]} \cdot t + \mu^{[m]}$ between the normative time-line for data type m and the one of the cognitive decline so that $\Phi^{[m]} \circ \psi_i^{[m]}(t_{ij})$ is as close as possible to $\psi_i^{[\text{cog}]}(t_{ij})$ by minimizing $\sum_{i=1}^{N} \sum_{j=1}^{N_i} \left| \lambda^{[m]} \cdot \psi_i^{[m]}(t_{ij}) + \mu^{[m]} - \psi_i^{[\text{cog}]}(t_{ij}) \right|^2$, which admits a closed form solution. This steps allows the synchronization of different models of disease progression.

Estimation of age of diagnosis. The time-point $\psi_i^{[\text{cog}]}(t_i^{\text{diag}})$ maps the age at which the i-th subject was diagnosed with the disease, i.e. t_i^{diag}, to a disease stage that ideally would be the same for all subject. In practice, we used the average stage $t^{\text{diag}} = \frac{1}{N} \sum_{i=1}^{N} \psi_i^{[\text{cog}]}(t_i^{\text{diag}})$ as an estimate of the

[l] www.deformetrica.org

[m] https://gitlab.icm-institute.org/aramislab/longitudina

diagnosis time on the normative time-line of the model of cognitive decline. Note that this estimate is the best predictor of the age at diagnosis, as it minimizes $\sum_{i=1}^{N} \left| \{\psi_i^{[m]}\}^{-1}(t^{\text{diag}}) - t_i^{\text{diag}} \right|^2$.

Personalization Once the model is calibrated on a longitudinal data set, we personalize it to the temporal sequence $\{y_{ij}, t_{ij}\}_j$ of any target subject i by finding the values of the random-effects z_i that maximizes the posterior log-likelihood:

$$\log p(z_i|\{y_{ij}\}_j, \hat{\theta}) = \log p(\{y_{ij}\}_j|z_i, \hat{\theta}) + \log p(z_i|\hat{\theta}) + \text{Constant}. \tag{3}$$

The first term $\log p(\{y_{ij}\}_j|z_i, \hat{\theta}) \propto -\sum_{j=1}^{N_i} \left\| y_{ij} - f(z_i, \hat{\theta}, t_{ij}) \right\|^2$ measures the distance between the observations and the current fit of the model to this data. The norm considered is the one appearing in the noise likelihood: sum of squared differences for neuro-psychological assessments, PET images and cortical thickness maps, and the currents distance between meshes for hippocampus meshes[37]. The second term is a prior on the likelihood of the random-effects. This minimization problem is solved using Powell's method for the hippocampus meshes, and the L-BFGS algorithm [68] for all other modalities. Both algorithms were taken from the SciPy 1.1.0 library[n].

At convergence, the residual $\epsilon_{i,j} = y_{ij} - f(\hat{z}_i, \hat{\theta}, t_{ij})$ for the optimal value of the random-effect $\hat{z}_i$ is called the **reconstruction error** of the j-th observation of the i-th subject. Note that in the case of the hippocampus meshes, only the absolute reconstruction error $|\epsilon_{ij}|$ can be computed, because the currents representation is a multivariate vector, of which we take the norm[37].

[n] https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html

827 We compare the distribution of the reconstruction errors with the uncertainty in the measure-
828 ments, which is estimated as follows. In the ADNI protocol[69,70], most MRI sessions consist of a
829 pair of test and re-test MRI, namely two scans performed on the same day one immediately after
830 the other one. For 1841 out of 1993 MRI sessions, we measure therefore the differences between
831 the MRI derived data (hippocampus meshes and cortical thickness maps) when using the test or
832 the re-test image. These differences give an empirical distribution of the noise due to variations in
833 image acquisition and processing.

834 For PET derived data, we use the baseline and follow-up scans of stable cognitively normal
835 and amyloid negative subjects in ADNI, as a proxy to test / re-test data (125 subjects, 244 visits
836 with a follow-up time of 18 months). For those subjects, the changes in glucose metabolism over
837 a 18 months period is supposed to be negligible compared to all the other factors affecting the
838 measurements such as variations in reaction to radiotracers, and methods for PET reconstruction,
839 image correction and extraction of regional measurements.

840 Test / re-test studies have shown a that the MMSE, which scales from 0 to 30, is subject
841 to a difference between two sessions, whose standard deviation ranges from 1.3 for a one-month
842 interval[71] up to 1.82 for a 1.5 year long interval[72], thus representing a standard deviation of 4.3% to
843 6%. Another study[73] measured the former ADAS-Cog that scales between 0 and 70 three times at
844 a 2-week interval, with an agreement between raters. The inter-ratter standard deviation is of 9.64
845 between the first and second test, and of 6.79 between the second and third test. The intra-rater
846 standard deviation is of 8.16 between the first and third visit. This corresponds to a standard devia-
847 tion ranging from 9.7% to 13.8%. On average, we consider such neuro-psychological assessments

to have a zero-mean Gaussian distribution of noise with standard deviation of order 7%.

Simulation The calibration of the model and its personalization to the training subjects produce a series of values of the random-effects z_i. As sex has been found to be one of the most discriminative co-variates, we separate the random-effects in two groups: men and women. For each group, we estimate two empirical multivariate distributions of these random-effects:

- a kernel density estimation is performed to finely capture the empirical distribution of the time-related parameters, i.e. the speed factor, time-shift, and baseline age;
- a multivariate Gaussian distribution is fitted on all the individual parameters, i.e. the time-related ones augmented with the space-shift-encoding sources.

A new individual is simulated by drawing new random-effects z_i^{sim} according to the following procedure:

- its speed factor, time-shift and baseline age are drawn from the previously-estimated kernel density;
- its sources are drawn from the multivariate Gaussian conditional distribution with respect to its already-drawn time parameters.

We then generate the corresponding data by computing $f(\hat{\theta}, z_i^{\text{sim}}, t)$ at any arbitrary age t.

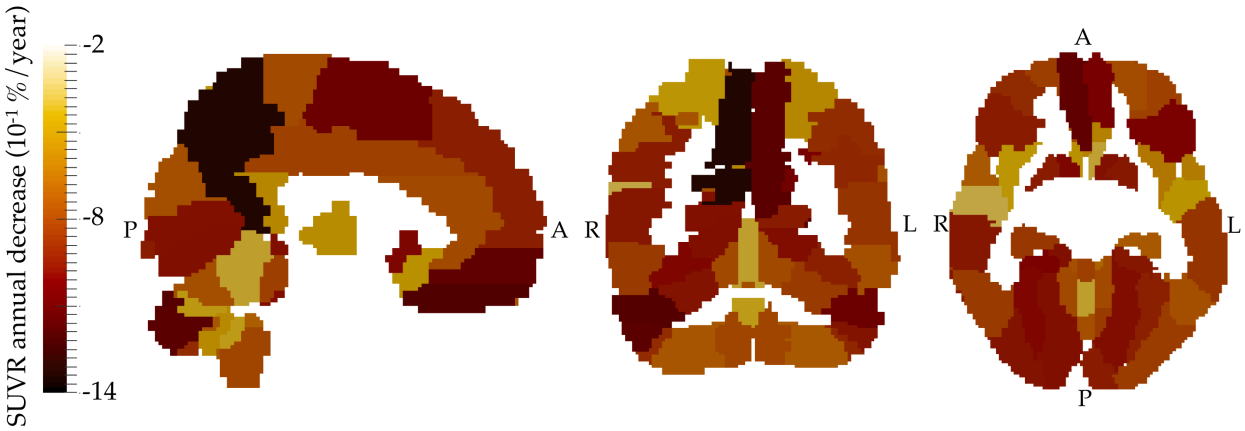
To validate our simulation method, we simulate a number of subjects equal to the number of training subjects for each modality, with the same sex ratio as in the training set, and then

compare how well the simulated cohort replicates the original cohort by comparing statistics of the simulated data with the corresponding statistics of the original data set.

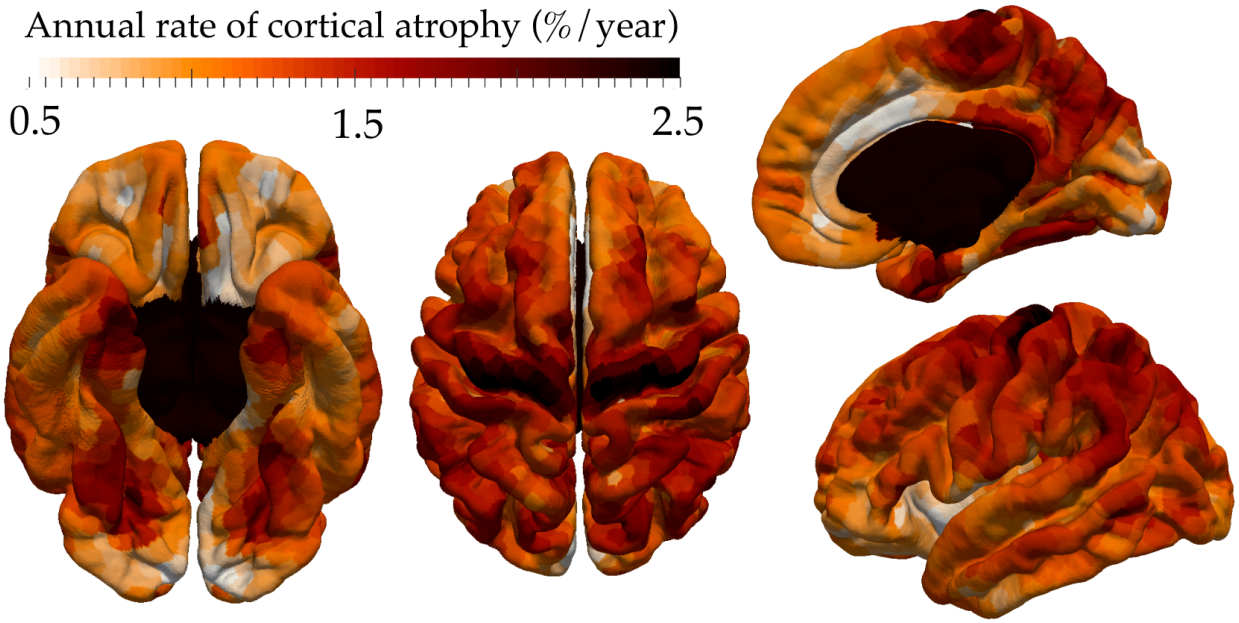
Cofactor analysis We take the series of random-effect estimates after model calibration and personalization on a given training data set. For each data type, we look for correlations between the values of these random-effects and a series of co-factors: gender, APOE status, marital status, level of education and amyloid status. On the one hand, the series of co-factor is regressed against the uni-dimensional temporal random effects (time-shift τ_i and acceleration factor α_i); the statistical significance of the slope coefficients is assessed by a two-sided t-test. On the other hand, for the multivariate vector of sources (s_i), we perform a 2-blocks partial least square [74] method to identify correlations between a linear combination of sources and co-factors. The resulting series of p-values are corrected for multiple comparisons using the False Discovery Rate (FDR) method.

When a significant association between a linear combination of sources (i.e. a vector d in the multivariate space of sources) and a categorical co-factor has been found, we project the individual source estimates on this direction (i.e. $b_i = d^T s_i$) and compute the distance between the empirical means of each class ($\delta_{12} = \overline{b}_2 - \overline{b}_1$). We select two points in the source space at $u = \pm a\delta_{12}/2$ to represent the typical configuration of each class, where $a = 1$ (for the cortical thinning) or 3 (for the hippocampus shape) is a factor to amplify differences for better visualisation. We then reconstruct the corresponding typical data by computing the exp-parallel curve in the direction u at a given time-point t: $\eta^{Au}(\gamma_0)(t)$.

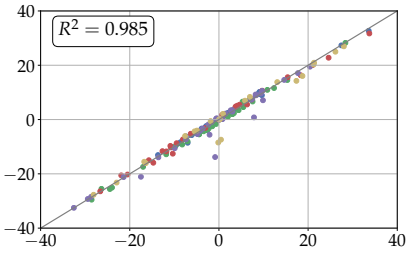
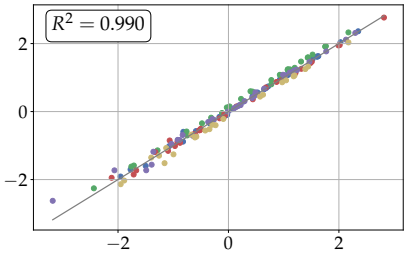
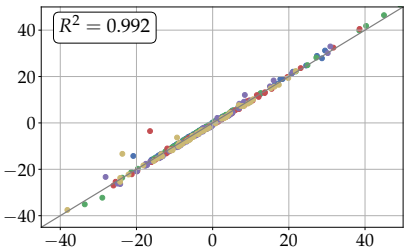
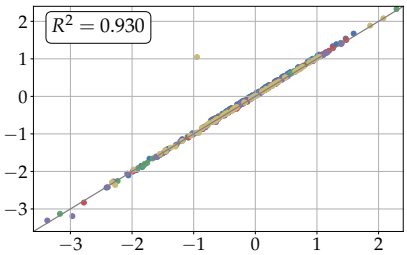
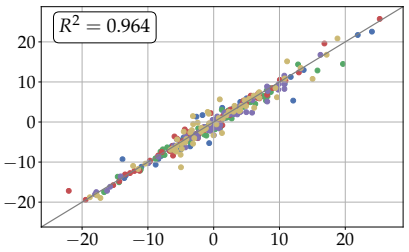
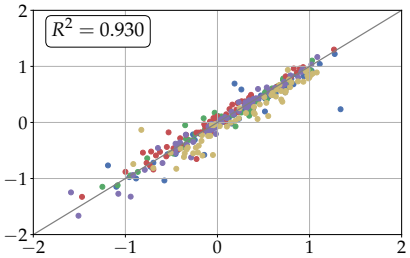
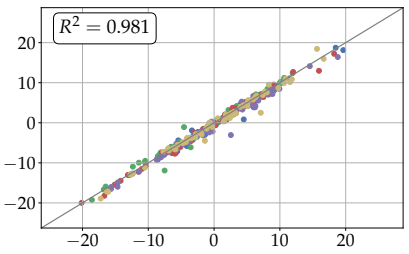
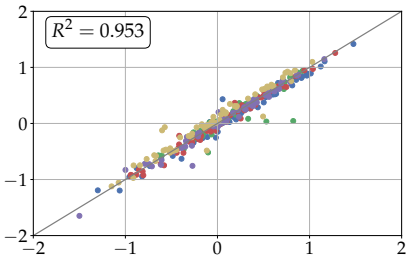
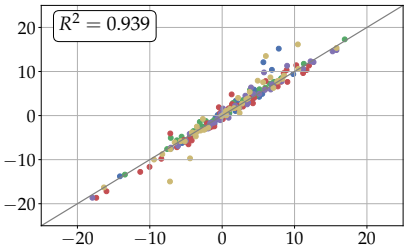
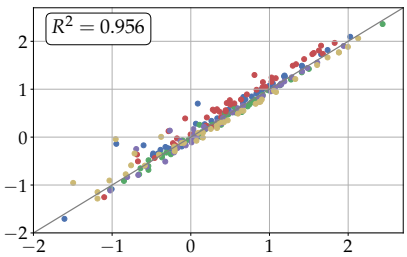
A Extended Data



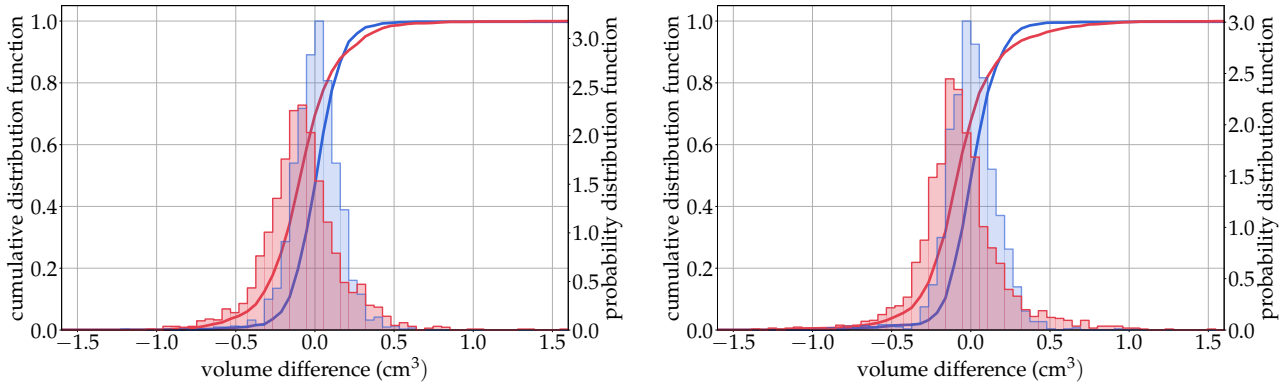
Extended Data Figure 1: Map of the annual rate of SUVR decrease at age of diagnosis computed from the model of hypometabolism shown in Fig. 2.



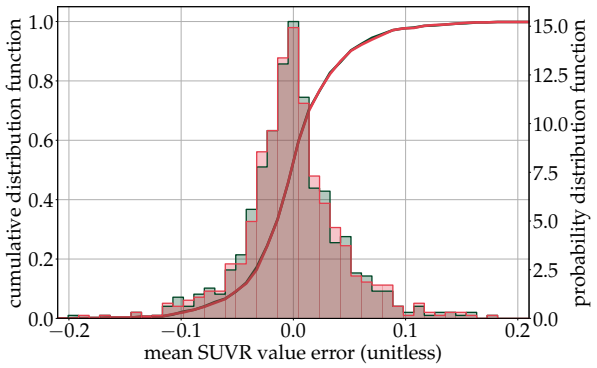
Extended Data Figure 2: Map of the annual rate of cortical atrophy at age of diagnosis computed from the model of cortical thinning shown in Fig. 2.

Modality	Time-shift	Speed factor
FDG-PET SUVR		
Cortical thickness		
Right hippocampus		
Left hippocampus		
Neuro-psychological scores		

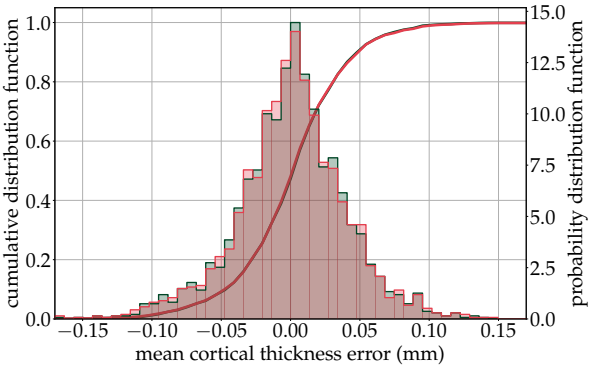
Extended Data Figure 3: Robustness of model calibration and personalization. Estimated time-shifts and speed factors when the individual belongs to the training data (x-axis) or to the test-set (y-axis). The five colors correspond to the folds the individuals belong to.



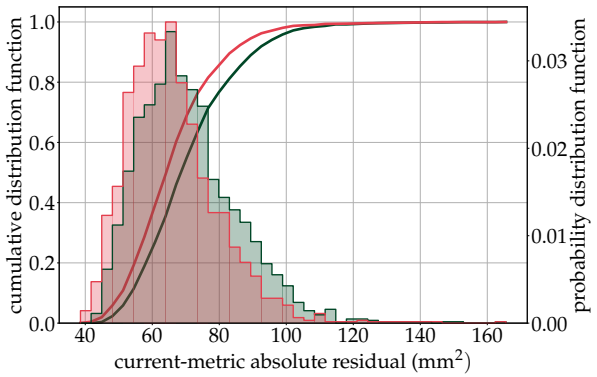
Extended Data Figure 4: Reconstruction errors in hippocampus volume. Superimposition of the distribution of the reconstruction errors (in red) and test / re-test differences (in blue) measured as volumes for the left and right hippocampus (left and right panel respectively). Whereas the distribution of the test / re-test differences is centered (empirical mean of 0.5 mm^3 for the left hippocampus and -1.2 mm^3 for the right hippocampus), the distribution of the reconstruction errors has an empirical mean of -84.5 mm^3 for the left hippocampus and -67.3 mm^3 for the right hippocampus. The standard deviations of the distributions are: 208.6 mm^3 and 210.2 mm^3 for the test / re-test differences for left and right hippocampus respectively, to be compared to 243.2 mm^3 and 267.2 mm^3 for the reconstruction errors.



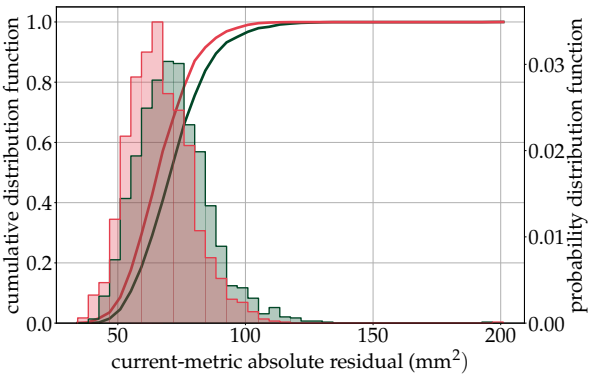
(a) FDG-PET SUVR values. The mean error is of $1.0 \times 10^{-4} \pm 0.044$ (red), and $-1.3 \times 10^{-4} \pm 0.044$ (green).



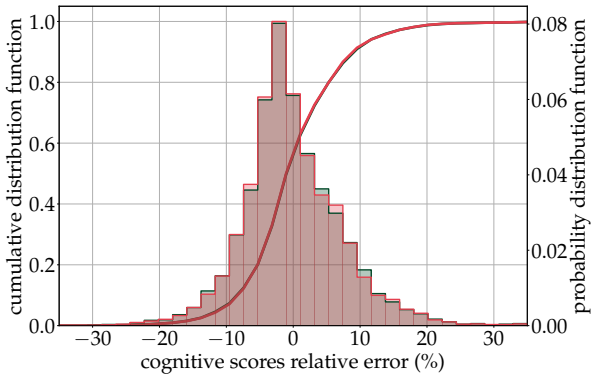
(b) Mean cortical thickness. The mean error is of $5.8 \times 10^{-4} \pm 0.040mm$ (red) and $6.1 \times 10^{-4} \pm 0.040mm$ (green).



(c) Left hippocampus. The mean error is $66.0 \pm 13.6\,mm^2$ (red), and $70.7 \pm 14.9\,mm^2$ (green).

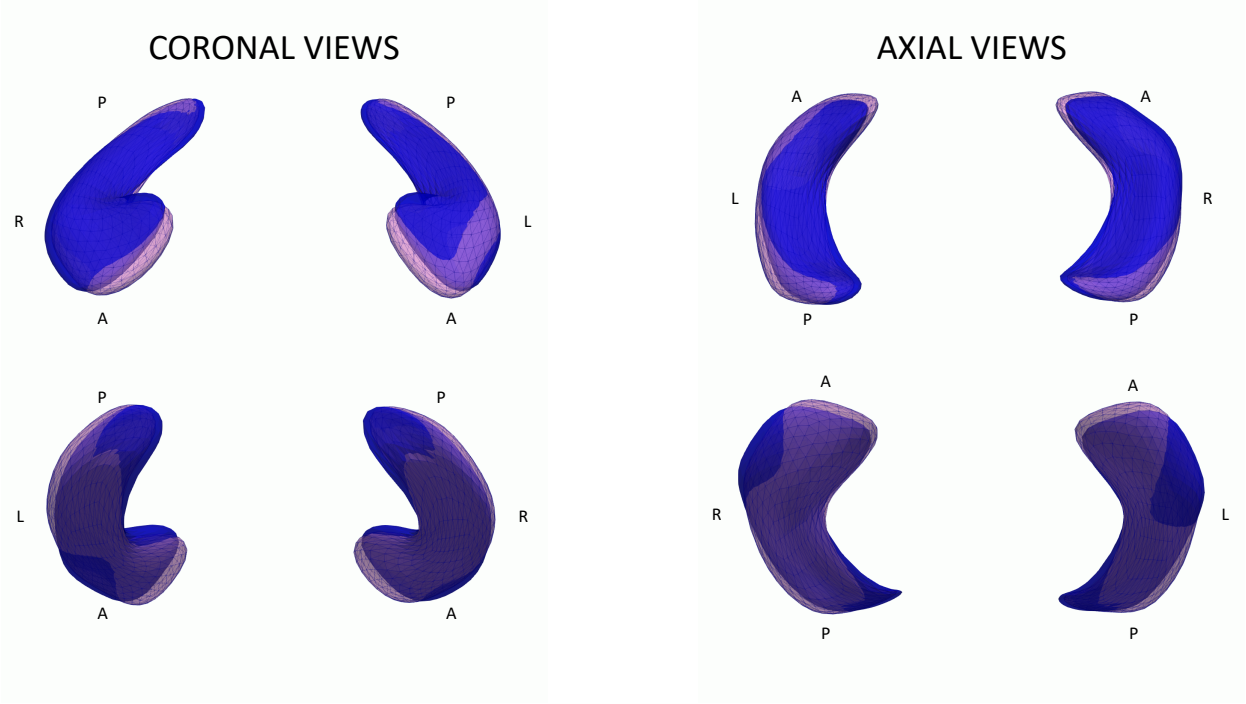


(d) Right hippocampus. The mean error is $66.6 \pm 12.8\,mm^2$ (red), and $71.7 \pm 14.0\,mm^2$ (green).

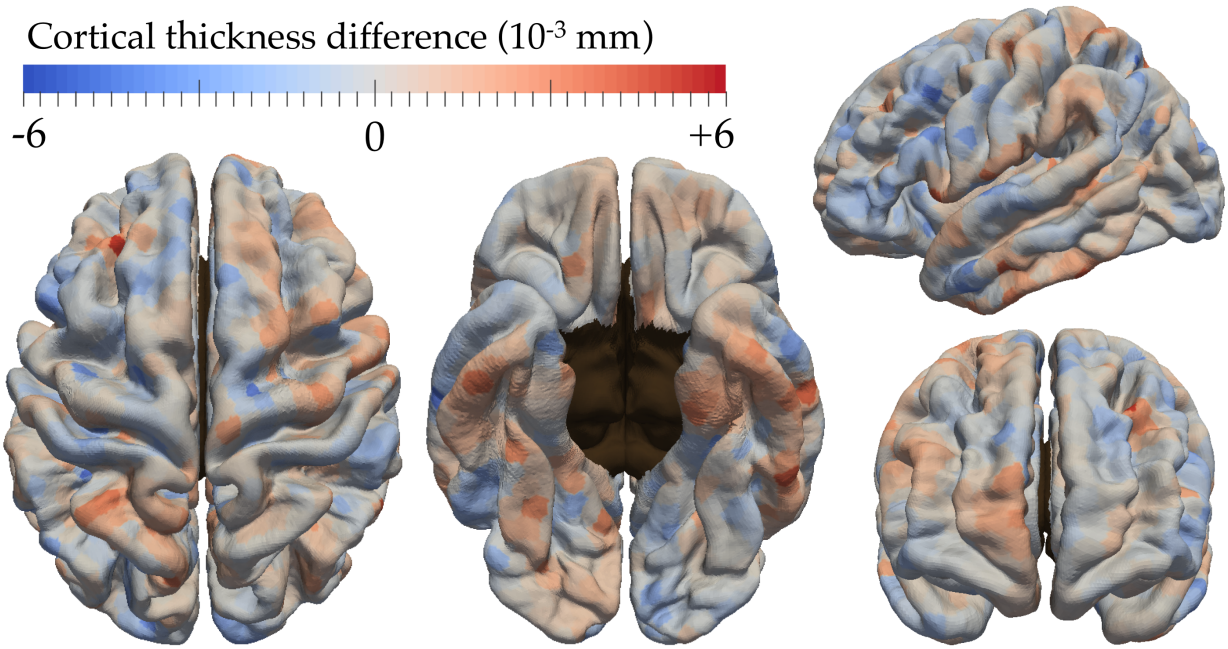


(e) Neuro-psychological assessments. The mean error is $-0.19 \pm 7.5\,\%$ (red), and $-0.14 \pm 7.5\,\%$ (green).

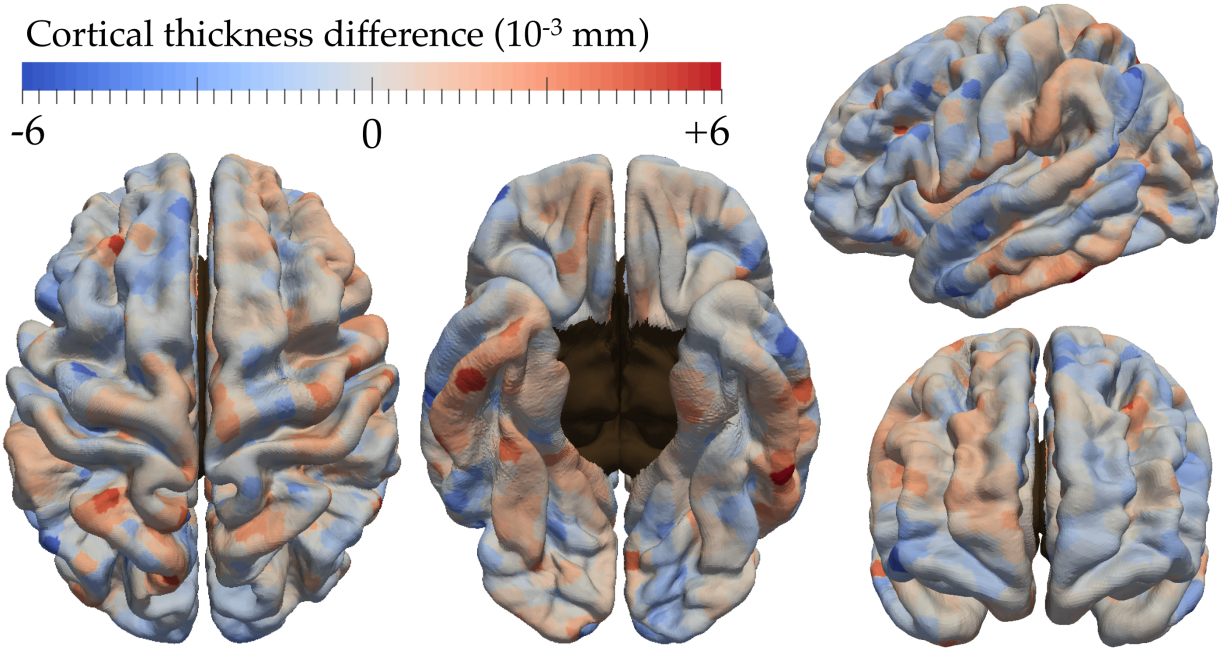
Extended Data Figure 5: Generalization error to unseen data. The distribution of reconstruction errors when calibration and personalization are done on the whole data set (in red, as in Fig. 3) is superimposed with the one estimated in the cross-validation procedure (in green).



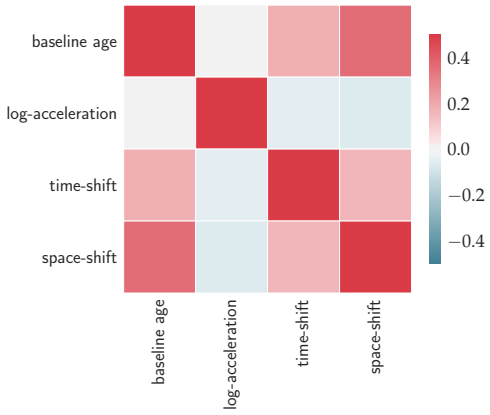
Extended Data Figure 6: Sex differences in hippocampus shape at age of diagnosis. Deformations of left and right hippocampi are shown in the direction of the significant space-shifts viewed from front (top-left), back (bottom-left), top (top-right) and bottom (bottom-right). Blue shapes are deformed in the direction of men. Red shapes are deformed in the direction of women. Amount of deformation has been magnified by a factor 3 in each direction for visualization purposes.



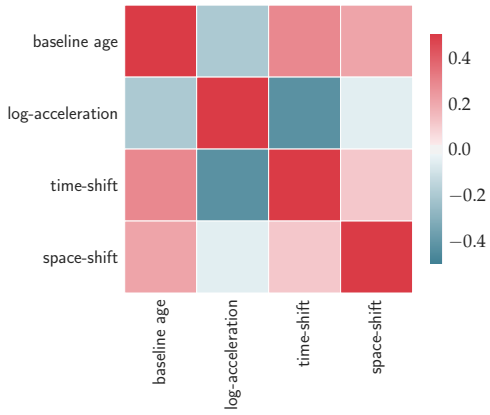
Extended Data Figure 7: Sex differences in cortical thickness at age of diagnosis. Color encodes in each brain region the estimated difference in cortical thickness between women and men at the stage of diagnosis.



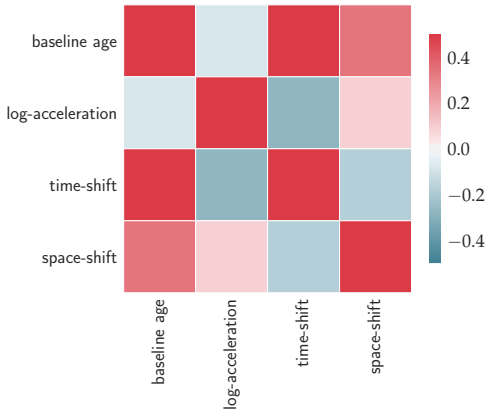
Extended Data Figure 8: Differences in cortical thickness at age of diagnosis between amyloid positive and amyloid negative subjects. Color encodes in each brain region the estimated difference in cortical thickness between amyloid positive subjects and amyloid negative subjects at the stage of diagnosis.



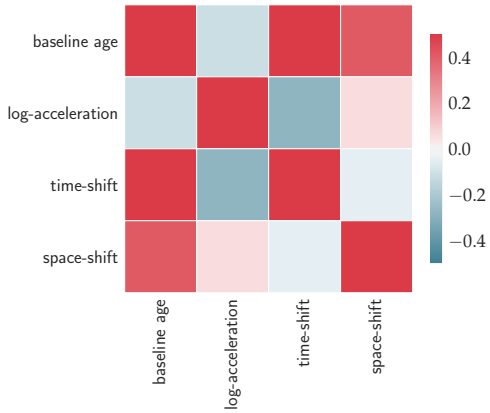
FDG-PET SUVR



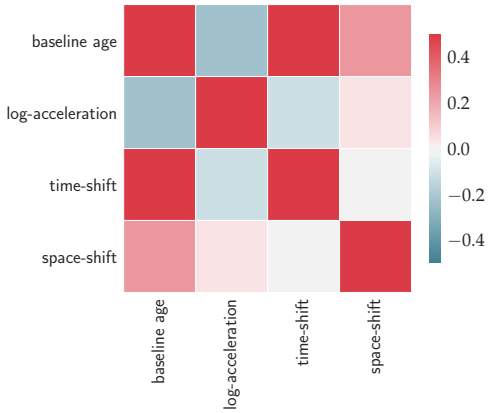
Mean cortical thickness



Right hippocampus



Left hippocampus



Neuro-psychological assessments

Extended Data Figure 9: Empirical variance-covariance matrices of the temporal random effects. We use baseline age, speed factor (aka log-acceleration factor), time-shift and the unidimensional projection of the spatial random effects learned by a Partial Least Square regression of the sources.

Extended Data Table 1: Summary statistics of the subject subsets for each data type

	ADAS & MMSE	PET	MRI
Number of subjects	223	157	322
Number of visits	1235	690	1993
Average number of visits per subject (± std)	5.5 (±1.1)	4.4 (± 2.1)	5.8 (± 2.4)
Average age (± std)	76.2 (± 6.9)	74.0 (± 7.2)	74.0 (± 6.7)
Sex ratio (F/M in %)	39 / 61	44 / 56	41 / 59
Amyloid status (+/-/unknown in %)	65.5 / 7.2 / 27.3	77.4 / 7.3 / 15.3	73.2 / 7.1 / 19.7
APOE carriership (%)	62.8	64.2	65.2
Education (mean ± std, in years)	15.8 (± 2.8)	15.8 (± 2.7)	15.9 (± 2.8)

Modality	Parameters	All data	Cross-validation
FDG-PET images	σ (no units)	0.101	0.101 ($\pm$ 0.001)
	t_0 (years)	75.5	74.9 ($\pm$ 0.9)
	σ_τ (years)	11.9	11.5 ($\pm$ 0.3)
	σ_ξ (no units)	1.30	1.28 ($\pm$ 0.03)
Cortical thickness	σ (mm)	0.442	0.442 ($\pm$ 0.001)
	t_0 (years)	82.0	82.7 ($\pm$ 0.7)
	σ_τ (years)	16.9	18.2 ($\pm$ 0.7)
	σ_ξ (no units)	0.99	1.03 ($\pm$ 0.02)
Right hippocampus	σ (mm^2)	2.49	2.60 ($\pm$ 0.03)
	t_0 (years)	76.2	75.7 ($\pm$ 0.3)
	σ_τ (years)	9.15	10.04 ($\pm$ 0.66)
	σ_ξ (no units)	0.71	0.78 ($\pm$ 0.03)
Left hippocampus	σ (mm^2)	2.67	2.74 ($\pm$ 0.04)
	t_0 (years)	76.3	76.3 ($\pm$ 0.3)
	σ_τ (years)	8.53	9.09 ($\pm$ 0.50)
	σ_ξ (no units)	0.66	0.68 ($\pm$ 0.03)
Cognitive scores	σ (no units)	0.081	0.081 ($\pm$ 0.001)
	t_0 (years)	71.5	72.4 ($\pm$ 0.8)
	σ_τ (years)	7.29	7.36 ($\pm$ 0.25)
	σ_ξ (no units)	1.07	1.11 ($\pm$ 0.11)

Extended Data Table 2: Fixed-effects estimates using calibration on the whole data set (first column) and in a five fold cross-validation setting (second column) where mean and standard deviations of the five estimates are shown.

	Mean Error ($\pm$ std)		Mean Absolute Error ($\pm$ std)	
Modality (unit)	Reconstruction	Measurement noise	Reconstruction	Measurement noise
FDG-PET images)	$1.1 \times -10^{4}(\pm 0.10)$	$-3.0 \times 10^{-3}(\pm 0.095)$	$7.6(\pm 6.5) \times 10^{-2}$	$6.8(\pm 9.4) \times 10^{-2}$
Cortical thickness (mm)	$5.8 \times 10^{-4}(\pm 0.44)$	$-1.1 \times 10^{-3}(\pm 0.28)$	$0.35(\pm 0.28)$	$0.19(\pm 0.20)$
Right hippocampus (mm^2)	N/A	N/A	$69.8(\pm 15.0)$	$85.2(\pm 40.1)$
Left hippocampus (mm^2)	N/A	N/A	$68.5(\pm 15.9)$	$83.2(\pm 36.0)$
Cognitive scores	$-2.2 \times 10^{-3}(\pm 0.075)$	$0(\pm 0.070)$	$5.5(\pm 5.0) \times 10^{-2}$	N/A

Extended Data Table 3: Comparison between the statistics of the reconstruction errors and the ones of the distribution of the measurement noise.