Completion and augmentation of connectomic datasets in dementia and Alzheimer's Disease using Virtual Patient Cohorts

Djouya Mohammad Arbabyazd[1,+], Kelly Shen[2], Zheng Wang[2], Martin Hofmann-Apitius[3], The Alzheimer's Disease Neuroimaging Initiative*, Anthony R. McIntosh[2], Demian Battaglia[1, §,@] & Viktor Jirsa[1, §,@]

[+]First author; [§] shared last authors; [@] corresponding author

*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

[1] Université Aix-Marseille, INSERM UMR 1106, Institut de Neurosciences des Systèmes, F-13005 Marseille, France;

[2] Rotman Research Institute, Baycrest Centre, Toronto, Ontario, M6A 2E1, Canada;

[3] Fraunhofer Institute for Algorithms and Scientific Computing, 53754 Sankt Augustin, Germany

Abstract

Large neuroimaging datasets, including information about structural (SC) and functional connectivity (FC), play an increasingly important role in clinical research, where they guide the design of algorithms for automated stratification, diagnosis or prediction. A major obstacle is, however, the problem of missing features (e.g., lack of concurrent DTI SC and resting-state fMRI FC measurements for many of the subjects).

We propose here to address the missing connectivity features problem by introducing strategies based on computational whole-brain network modeling. Using the ADNI dataset for proof-of-concept, we demonstrate the feasibility of virtual data completion (i.e., inferring "virtual FC" from empirical SC or "virtual SC" from empirical FC), by using self-consistent mean-field network simulations and analytic approaches. Furthermore, we use similar procedures to perform dataset augmentation, i.e., complementing the original dataset with a large number of realistic surrogate connectivity matrices. We thus show that algorithms trained on virtual SCs and/or FCs can achieve performance in the unsupervised classification of control subjects and patients comparable to when trained on actual empirical data. Furthermore, the combination of empirical with virtual data allows algorithms to learn better how to extract information relevant for discrimination, resulting ultimately in superior classification performance.

Introduction

Despite decades of massive investment in Alzheimer's disease (AD) research and the daunting literature on the topic, the partial and, sometimes contradictory nature of the reported results (World Alzheimer Report 2018) still prevents a complete understanding of the factors governing the progression of the disease (Braak & Braak, 1991; Braak et al., 2006; Komarova & Thalhauser, 2011; Henstridge et al., 2019) or of the diversity of cognitive deficits observed in different subjects (Iacono et al., 2009; Mungas et al., 2010; Allen et al., 2016). In this context, datasets that compile rich and diverse genetic, biomolecular, cognitive, and neuroimaging (structural and functional) features for a large number of patients are playing an increasingly important role (Rathore et al., 2017; Iddi et al., 2019). Example applications include: the early diagnosis and prognosis by using MRI images (Dennis & Thompson, 2014; Chiesa et al., 2017; De Vos et al., 2018); the use of machine learning for automated patient classification (Cuingnet et al., 2011; Zhang et al., 2012; Moore et al., 2019) or prediction of the conversion from early stages to fully developed AD (Rombouts et al., 2005; Moradi et al., 2015; Casanova et al., 2018); the extraction of decision networks based on the combination of semantic knowledge bases and data mining of the literature (Sanchez et al., 2011; Kodamullil et al., 2015; Iyappan et al., 2016).

Among the factors contributing to the performance of prediction and inference approaches in AD are not only the large size of datasets but also the multiplicity of features jointly available for each patient. Indeed, one can take advantage not only of the complementary information that different features could bring but also capitalize on possible synergies arising from their simultaneous knowledge (Wang et al., 2015; Zimmermann et al., 2016; Iddi et al., 2019). Unfortunately, even gold standard publicly available datasets in AD, such as the datasets released by the Alzheimer's Disease Neuroimaging Initiative (ADNI) consortium (Wyman et al., 2013; Beckett et al., 2015; Weiner et al., 2017), have severe limitations. Indeed, if they include neuroimaging features of different types –structural DTI and functional MRI– these features are simultaneously available for only a substantial minority of the subjects in the dataset (i.e., the feature coverage is not uniform over the dataset). Furthermore, if the number of subjects included is relatively large (hundreds of subjects), it still is too small to qualify as "big data" properly.

Here we propose a new solution aiming at relieving these problems of partially missing features and limited sample size. To do so, we build on the quickly maturating technology of mean-field whole-brain network modeling (see Deco et al., 2011 for review). Importantly, the use of whole-brain modeling for pragmatic applications is greatly facilitated by the development of dedicated neuroinformatic platforms –such as "The Virtual Brain" (TVB; Sanz-Leon et al., 2013, 2015; Woodman et al., 2014)– and personalized simulation pipelines (Schirner et al., 2015; Proix et al., 2016), with translational impact (Jirsa et al., 2017; Proix et al., 2017). Computational modeling provides a natural bridge between structural and functional connectivity, the latter emerging as the manifestation of underlying dynamical states, constrained but not entirely determined by the underlying anatomy (Ghosh et al., 2008; Kirst et al., 2016). Theoretical work has shown that average functional connectivity properties in the resting-state can be accounted for by the spontaneous collective activity of brain networks informed by empirical structural connectivity (SC) when the system is tuned to operate slightly below a critical point of instability (Deco et al., 2011, 2012). Based on this finding, simulations of a model constructed from empirical DTI connectomes and then tuned to a suitable slightly sub-critical dynamic working point are expected to provide a good rendering of resting-state functional connectivity (FC). It becomes thus possible to complete the missing information in a dataset about BOLD fMRI FC by running a TVB simulation in the right regime embedding the available empirical DTI SC (*SC-to-FC completion*). Analogously, mathematical formula (Galán, 2008; Saggio et al., 2016) or algorithmic procedures based on mean-field modeling steps (Gilson et al., 2016; 2018) can be used to address the inverse problem of inferring SC from FC (*FC-to-SC completion*). In this work, we provide initial proofs-of-concept of data completion using TVB-technologies, using them to "fill gaps" in the ADNI neuroimaging dataset, generating the missing connectivity SC (or FC) connectivity matrices from the available FC (or SC) ones.

Finally, both types of completion rely on algorithms, including stochastic elements. Therefore, running the FC-to-SC or SC-to-FC completion procedures multiple times yields different realistic surrogate connectomes. In this way, the completion procedure can be used to generate not just one, but as many as desired instances of artificially completed data (e.g. hundreds of virtual FCs associated with each given empirical SC). Schemes to artificially increase the size of a given sample by introducing surrogate copies of the original data, with a wide variety of "distortions" which preserve nevertheless the distinctive features of the initial objects are known in machine learning under the name of *data augmentation* (Yaeger et al., 1997; Taylor & Nitschke, 2018). In many situations, the generalization capabilities of machine learning algorithms for automated categorization can be

enhanced by training them not on the original "true" dataset but augmented copies of it. While data augmentation cannot create additional information besides the one already carried by the original data, yet it can make it simpler for machine learning algorithms to detect and extract desired information, a task that would have been more difficult based on the empirical samples alone. We here extend our TVB-based strategies for data completion to connectomic data augmentation. We show then that actual empirical connectomic data can be categorized –e.g., by separating mild cognitive impairment (MCI) or AD patients from control– by machine learning algorithms trained uniquely on virtual data. Remarkably, we find that the use of completed and augmented dataset lead to performances comparable to algorithms trained on the original data themselves.

To conclude, we provide systematic "recipes" for generating surrogate connectomic data via data-constrained mean-field models. We show that the information that we can extract from computationally inferred connectivity matrices is largely equivalent to the one carried by the original empirical data. This opens the way to the design and sharing of veritable "virtual cohorts" data, ready for easier multi-centric federation and machine-learning applications in clinics.

Results

ADNI connectomic data have gaps

In the framework of this study, we chose to focus on one of the first and most popular available datasets in AD research, including a substantial amount of structural and functional connectomic information, i.e. the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI is impressive for the variety of features it aimed at systematically gathering (Figure 1A). Importantly, based on the T1, DTI and resting-state (rs) BOLD fMRI images available through the ADNI data-sets, state-of-the-art processing pipelines can be used to extract subject-specific Structural and resting-state Functional Connectomes, compiled into connectivity matrices adapted to the brain parcellation of choice (Figure 1B, see *Materials and Methods* for details).

We had access to an empirical of 244 overall subjects (119 labeled as "MCI", 51 as "AD", in addition to 74 control subjects, see *Materials and Methods*) for which MRI data had been gathered. We could extract an FC matrix for 168 subjects (starting from rsfMRI) and an SC matrix (starting from DTI) for 88 subjects. However, only for a minority of 12 subjects rsBOLD and DTI information were both available. In a majority of cases, either DTI or rsBOLD were missing (Figure 1C). This reduced number of "complete" subjects constitutes a serious challenge to attempts of automatedly categorize them through machine learning or inference approaches capitalizing on both SC and FC features simultaneously. As a matter of fact, the total numbers of AD- and MCI-labeled subjects in this complete subset decreased respectively to just 2 and 4, against 6 controls. In these conditions, the development of effective data completion strategies would be an important asset toward the development of classifier schemes exploiting FC/SC synergies. Therefore, approaches to "fill gaps" (completion) and, possibly, even artificially boosting sample size (augmentation) are veritably needed.

Linking SC and resting-state FC via computational modeling

As previously mentioned, FC and SC are related only indirectly through the rich non-linear dynamics supported by brain networks (Ghosh et al., 2008; Deco et al., 2011; Kirst et al., 2016). Mean-field modeling of large-scale brain networks has emerged initially as the key tool to predict the emergent dynamic patterns of resting-state FC, from spontaneous dynamics constrained by SC (Ghosh et al., 2008). It is thus natural to propose the use of model-based solutions to perform data-completion, which, in both the SC-to-FC and FC-to-SC directions, requires to capture the inter-relation between the two as mediated by dynamics.

Large-scale mean-field brain network models are specified by: i) a parcellation of cortical and subcortical brain areas; ii) a co-registered input SC matrix in the same parcellation; iii) a forward solutions linking source and sensor space; iv) a neuronal mass model, describing the non-linear dynamics of the regions at each of the nodes of the SC matrix; v) a choice of a few global parameters (e.g. scale of strength of inter-regional connectivity or speed of signal propagation along fiber tracts); vi) an external input given to the different regions, that, in the simplest case, corresponds to simple white noise uncorrelated across each of the different sites and of homogeneous strength. The Virtual Brain enables the complete workflow from brain images to simulation (TVB; Sanz-Leon et al., 2013, 2015). Personalization is accomplished by the subject-specific structural skeleton –

ingredients (i) though (iv)–, which has been demonstrated to be predictive (Proix et al 2017; Melozzi et al 2019). Simulations of the model can be run to generate surrogate BOLD time-series of arbitrary length (see *Materials and Methods* for details) and the associated simulated resting-state FC, time-averaged (static FC) or even time-resolved (FC dynamics or FCD, Hansen et al., 2015). The thus obtained simulated FC will depend on the chosen global parameters, setting the *dynamic working point* of the model. The model dynamics will eventually switch between alternative dynamical regimes when its global control parameters cross specific critical points. Tuning global parameters will thus uniquely determine, in which regime the model operates. Mean-field large scale models constrained by empirical SC tend to generate simulated resting-state FC that best matches empirical observations when the dynamic working point of the model lies in the proximity of a model's critical point (Deco et al., 2011; Deco et al., 2013; Hansen et al., 2015).

We here chose one of the simplest possible whole-brain network model designs, which emphasizes activity-based network organization (as opposed to reorganization due to synchronization) and thus ignores inter-regional propagation delays. This approach is frequently used in the literature (e.g., Deco et al., 2013; Hansen et al., 2015; Aerts et al., 2018) and has the advantage of avoiding the need for complex delay differential equation integration schemes (see *Discussion* for more details). Activation-based approaches adopt particularly simple neural mass models such as the reduced Wong-Wang model (Deco et al., 2013), in which the dynamics of an isolated brain region is approximated by either one of two possible steady states, one "down state" at low firing rate and an "up state" at high firing rate, a feature initially meant to mimic bi-stability in working memory or decision making (Wong & Wang, 2006). By varying G the model will switch from a low-coupling regime, in which all regional activations are low to a high-coupling regime, in which all regional activations are high, passing through an intermediate range, in which both regimes can exist in a multistable manner and regions display spatially and temporally heterogeneous activations (a changing mix of high and low firing rates). The best fit between simulated and empirical FC occurs slightly before the critical rate instability, at which modes of activity with low firing rate disappear (Deco et al., 2013).

As alternatives to the just described *non-linear mean-field models* (MFMs) of resting-state brain dynamics, simpler *stochastic linear models* (SLMs) have also been considered (Goñi et al., 2014; Messé et al., 2014; Saggio et al., 2016). In these models, the activity of each region is modeled as a stochastic process (linear, in contrast to the non-linear neural mass dynamics of conventional MFMs), biased by the fluctuations of the other regions weighted by the SC connectome (see *Materials and Methods*). SLMs have also two different regimes. In the first regime, the activities of all regions converge to a fixed-point of constant mean fluctuating activities, while, in the second, regional activities diverge with exponential growth. Once again, the best fit between the simulated and the empirical resting-state FCs is observed when tuning the model parameters slightly below the critical point (Hansen et al., 2015; Saggio et al., 2016).

MFMs and SLMs provide thus two natural ways to generate simulated resting-state FCs, depending on the chosen dynamic regime, starting from a selected SC. Strategies have also been devised to approximately solve the inverse problem of determining which SC matrix should be used as input to a model in order to give rise to a simulated FC matching a specific, pre-determined target matrix. For the SLM, a simple analytic solution to the inverse problem exists (Saggio et al., 2016). For MFMs, inverse problems have not been studied with the same level of rigor, but algorithms have been introduced that iteratively adjust the weights of the SC matrix currently embedded in the model to improve the fit between simulated and target FCs (Gilson et al., 2016; 2018). We will show later that these "effective connectivity" algorithms have the potential to cope with the actual problem of MFM inversion.

Model-driven data completion

Figure 2 summarizes many of the modeling operations described in the previous section framing them in the specific context of connectomic data completion. MRI data can be used to generate empirical SC matrices SC_{emp} (from DTI) or FC_{emp} (from rs fMRI BOLD). By embedding the empirical matrix SC_{emp} into a non-linear MFM or a linear SLM, it is possible to compute surrogate FC matrices (Figure 2A, upward arrows), denoted, respectively, FC_{MFM} and FC_{SLM}. The MFM and SLM global parameters are suitably tuned (slightly subcritical) then FC_{MFM} and FC_{SLM} will be maximally similar to the empirical FC_{emp} (dynamic working point tuning, represented by dashed grey arrows in Figure 2A). Starting from the empirical matrix FC_{emp}, one can then infer surrogate SC matrices (Figure 2A, downward arrows), either by using a linear theory –developed by Saggio et al. (2016)– to compute a

surrogate SC_{SLM}; or by exploiting non-linear effective connectivity algorithm –generalized from Gilson et al. (2016; 2018)– to infer a surrogate SC_{MFM} starting from a random initial guess (see later section).
When connectomic data are incomplete (only SC_{emp} or only FC_{emp} are available, but not both simultaneously), computational simulation or inference procedures can be used to fill these gaps: by using FC_{MFM} or FC_{SLM} as virtual replacements for a missing FC_{emp} (Figure 2B); or by using SC_{MFM} or SC_{SLM} as virtual replacements for a missing SC_{emp} (Figure 2C). The quality of the model-generated virtual SCs and FCs can be assessed by comparing them with the actual empirical counterparts for the small subset of subjects for which both SC_{emp} and FC_{emp} are simultaneously available. Optimizing the quality of the virtually completed matrices on the subset of "SC_{emp}+FC_{emp}" subjects, also allows extrapolating target criteria for identifying when the model is operating a suitable dynamic working point, that can be evaluated solely based on simulated dynamics when a fitting target matrix is missing and thus fitting quality cannot be explicitly measured (cf. Figures 3 and 4). We can thus translate these criteria into precise algorithmic procedures that inform linear or non-linear SC-to-FC and FC-to-SC completion (see Tables 1-4).

We now, provide more details on implementation and performance for each of the four mentioned types of data completion.

Linear SC-to-FC completion

In linear SC-to-FC completion, a simple SLM (see *Materials and Methods*) is constructed based on the available SC_{emp} and its direct simulations or even, in a much faster manner, analytical formulas deriving from the model's theory are used to generate the associated virtual Pearson correlation matrix FC_{SLM} (Figure S1). In this stochastic linear modeling scheme, once the driving noise strength is arbitrarily chosen and fixed and the input connectome SC_{emp} is specified, there remains a single parameter to adjust, the global scale of long-range connectivity strength G. Figure S1A shows a systematic exploration, performed on subjects from the "SC_{emp}+FC_{emp}" subset, of how the completion quality depends on tuning this parameter G. As shown by the main plot in Figure S1A for a representative subject, increasing G the correlation between the empirical FC_{emp} and the simulated FC_{SLM}, derived here from direct SLM simulations, initially grows to peak in proximity of a critical value G^*. The correlation then drops dramatically when further increasing G beyond the critical point G^*.

The exact value of G^* depends on the specific personalized SC_{emp} connectome embedded into the SLM and is therefore different for each subject. The small boxplot inset in Figure S1A gives the distribution of the personalized G^* over all the subjects in the "SC_{emp}+FC_{emp}" subset. However, when performing linear FC completion because BOLD data and FC_{emp} are missing, the exact location of the fitting optimum cannot be determined. To perform linear SC-to-FC completion for the ADNI subjects with missing BOLD we chose to always use a common prescribed value $G^*_{ref} = 0.83$, set to be equal to the median of the personalized G^* over the "SC_{emp}+FC_{emp}" subset of ADNI subjects.

Once a G^*_{ref} value and a noise strength are set, the linear completion can be further sped-up by the fact that the covariance matrix FC_{SLM} for these frozen parameters can be analytically evaluated, as discussed in Saggio et al. (2016). Therefore, one can directly apply the SLM analytical formulas (see *Material and Methods*) on the available SC_{emp} as input, without the need for performing direct simulations to generate surrogate BOLD first. Figures S1B-C analyze the expected performance of this "simulation-less" procedure, as benchmarked by applying it on the "SC_{emp}+FC_{emp}" subset. The boxplot in S1B reports a median Pearson correlation between the linear virtual FC_{SLM} and the actual empirical FC_{emp} close to ~0.24. Panel S1C indicates then the percent loss in correlation that has been caused by using the common value G^*_{ref} and the analytical formula to evaluate the linear virtual FC_{SLM} rather than direct simulations at the actual personalized optimum G^* for each of the "SC_{emp}+FC_{emp}" subjects. The median quality loss is approximately 0.5%, indicating that the lack of personalized tuning of the SLM working point is only a minor issue and that is acceptable to speed-up completion by relying on analytical evaluations.

Table 1 provides a pseudo-code for the linear SC-to-FC completion procedure (see *Materials and Methods* for all details). Linear SC-to-FC completions for the DTI-only subjects in the considered ADNI dataset can be downloaded as Supplementary File S1.

Table 1. Pseudo-code for linear SC-to-FC completion

```
algorithm linear SC-to-FC completion is

external input:  empirical SC (SC_emp)
output: linear virtual FC (FC_SLM)
fixed parameters: noise level (σ), guess for optimal G (G*_ref)

begin
     1.  Evaluate the covariance matrix C from SC_emp based on SLM theory for G*_ref
     return FC_SLM = C
end
```

Non-linear SC-to-FC completion

In non-linear SC-to-FC completion, a more complex MFM (see *Materials and Methods*) is constructed based on the available SC_{emp} and is simulated to generate surrogate BOLD data and the associated Pearson correlation matrix FC_{MFM} (Figure 3). Non-linear mechanistic MFM models are more compliant with neurophysiology than the phenomenological SLMs. Furthermore, because of their non-linearities, they are potentially able to capture complex emergent collective dynamics resulting in non-trivial FCD (which SLMs cannot render, cf. Hansen et al., 2015). However, MFMs have also more parameters and are computationally costlier to simulate than SLMs.

We chose here to limit ourselves to MFMs based on a reduced Wong-Wang regional dynamics (see *Materials and Methods* for model equations), which has previously been used to successfully reproduce rsFC (Deco et al., 2013) and FCD (Hansen et al., 2015) starting from empirical SC, despite its relative simplicity with respect to other possible neural masses implemented in the TVB platform. In addition to the global scale of long-range connectivity strength G, the MFM model dynamics depend also on regional dynamics parameters. In Figure 3, we froze all local parameters but the NMDA decay time-constant τ, since they affected the dynamic behavior of the model less than the other control parameters and, in particular, did not alter qualitatively the repertoire of accessible dynamical regimes (compare Figure 3A with Figure S2). The simulated collective dynamics and the resulting non-linear virtual FC_{MFM} will depend on the choice of the free control parameters G and τ. In Figure 3A, we have explored the dependency of the correlation between FC_{MFM} and the actual empirical FC_{emp} as a function of G and τ achievable over the subjects in the "SC_{emp}+FC_{emp}" subset. As evident in figure 3A, this dependence is non-monotonic and the best-fitting qualities are concentrated in a narrow concave stripe across the G/τ plane. Panels 3B and 3C report zoom of Figure 3A into increasingly smaller regions, revealing an extended zone of high fitting quality which some absolute optimum parameters G^* and τ^* (here $G^* = \sim 1.5$ and $\tau^* = 25$).

Remarkably, this best-fitting quality zone on the G/τ plane is associated as well to other properties that can be evaluated just based on the simulated dynamics (and, therefore even when the actual target FC_{emp} is unknown and missing). We found that the best fit quality systematically occurs in a region where three criteria are jointly met (Figures 3D-F).

First, there is a mixture of "ignited" regions with large activation and of not yet ignited regions with a weaker firing rate (*spatial heterogeneity*, Figure 3D). Conversely, when moving out of the best-fitting zone, the activity becomes more spatially homogeneous, either with all regions stable at low (for $G <<< G^*$) or high (for $G >>> G^*$) firing rates.

Second, the time-averaged FC_{MFM} has a complex modular organization between order and disorder, associated to high average clustering coefficient, in contrast with the absence of clustering observed for $G <<< G^*$ or $G >>> G^*$ (*structured FC*, Figure 3E).

Third, the simulated collective dynamics give rise to meta-stability of FC along time, i.e. to a non-trivially structured FCD, which alternates between "knots" of transiently slowed-down FC network reconfiguration and "leaps" of accelerated reconfigurations. Such non-triviality of FCD can be detected by the inspection of the so-called FCD matrix (Hansen et al., 2015), representing the similarity between FC matrices computed at different time-windows (see *Materials and Methods*). In this FCD matrix analysis, FCD "knots" are visualized as blocks with high inter-FC correlations, while FCD "leaps" give rise to stripes of low inter-FC correlation. The prominence of the block structure of the FCD matrix can be measured by the FCD clustering coefficient (see *Material and Methods*), higher when the FCD matrix includes more evident knots. The FCD clustering coefficient is higher in the best fit zone, while it drops moving outside it toward $G <<< G^*$ or $G >>> G^*$ (*structured FCD*, Figure 3F).

By scanning the G/τ plane in search of a zone with simultaneous spatial heterogeneity of activations, structured FC and structured FCD, the MFM model parameters can be tuned to bring it in a zone invariantly resulting in relatively higher fitting quality. Figure 3G shows the analysis of the expected performance of this

procedure, as benchmarked by applying it on the "SC_{emp}+FC_{emp}" subset. We measured a median Pearson correlation between the non-linear virtual FC_{MFM} and the actual empirical FC_{emp} close to ~0.32.

Table 2 provides a compact pseudo-code for the non-linear SC-to-FC completion procedure (see *Materials and Methods* for all details). Non-linear SC-to-FC completions for the DTI-only subjects in the considered ADNI dataset can be downloaded as Supplementary File S2.

Table 2. Pseudo-code for non-linear SC-to-FC completion

```
algorithm non-linear SC-to-FC completion is

external input:  empirical SC (SC_emp)
output: non-linear virtual FC (FC_MFM)
fixed parameters: noise level (σ), simulation time (T), range to scan G_start ≤ G ≤ G_stop, range to scan
     τ_start ≤ τ ≤ τ_stop, other frozen Wong-Wang neural mass parameters

begin
      1.  Construct a MFM embedding SC_emp and the default frozen Wong-Wang neural mass parameters
      for G_start ≤ G ≤ G_stop
         for τ_start ≤ τ ≤ τ_stop
                  2.1 Simulate the MFM with current parameter values for a short time 0.2*T (discarding
                      an initial transient)
                  2.2 Compute surrogate BOLD from MFM time-series via Balloon-Windkessel model
                  2.3 Compute Corr(BOLD), i.e. the time-averaged FC matrix
                  2.4 Compute stream of time-resolved FC(t) and the associated FCD matrix
                  2.5 Compute and store Crit_1[G, τ] (Spatial heterogeneity of activations)
                  2.6 Compute and store Crit_2[G, τ] (Clustering Coefficient of time-averaged FC matrix)
                  2.7 Compute and store Crit_3[G, τ] (Clustering Coefficient of FCD matrix)
         end
      end
      3.  Identify G* and τ* for which Crit_1[G, τ], Crit_2[G, τ] and Crit_3[G, τ] are jointly optimum
      4.  Simulate the MFM with parameter values G* and τ* for a time T (discarding an initial transient)
      5.  Compute surrogate BOLD from MFM time-series via Balloon-Windkessel model
      6.  Compute C = Corr(BOLD), i.e. the time-averaged FC matrix at G* and τ*
      return FC_MFM = C
end
```

Linear FC-to-SC completion

In linear FC-to-SC completion, we use once again the analytic theory derived for the SLM (Saggio et al., 2016) to deterministically compute a surrogate SC_{SLM} as a function of the available FC_{emp} or, more precisely, of the resting-state $BOLD_{emp}$ time-series used to derive FC_{emp}. In this scheme, the linear virtual SC_{SLM} is indeed taken to be directly proportional to the *inverse covariance* of the BOLD time-series (see *Materials and Methods*). The proportionality constant would depend on the free parameters chosen for the SLM, serving as a link between FC and SC. Here we set arbitrarily this constant to the unit value.

Figure S3 shows the analysis of the expected performance of this procedure, as benchmarked by applying it on the "SC_{emp}+FC_{emp}" subset. We measured a median Pearson correlation between the linear virtual SC_{SLM} and the actual empirical SC_{emp} close to ~0.22.

Table 3 provides a pseudo-code for the linear FC-to-SC completion procedure (see *Materials and Methods* for all details). Linear FC-to-SC completions for the BOLD-only subjects in the considered ADNI dataset can be downloaded as Supplementary File S3.

Table 3. Pseudo-code for linear FC-to-SC completion

```
algorithm linear FC-to-SC completion is

external input:  empirical FC (FC_emp)
output: linear virtual SC (SC_SLM)
fixed parameters: noise level (σ), guess for optimal G (G*_ref)

begin
      1.  Evaluate the inverse matrix C^-1 from FC_emp
      2.  Build a matrix S proportional to C^-1 according to SLM theory and drop its diagonal
      return SC_SLM = S
end
```

Non-linear FC-to-SC completion

Non-linear FC-to-SC completion consists in the inference of an SC_{MFM} matrix that, used as input to an MFM, produces as output a simulated FC* matrix highly correlated with the available empirical FC_{emp} (Figure 4). This non-linear inverse problem is more sophisticated than linear FC-to-SC completion, because, for the MFM a theory providing an explicit formal link between input structural connectome (SC*) and output functional connectome (FC*) is not available, unlike for the SLM. Note indeed that MFMs, at the best-fitting dynamic working point, give rise not just to a single dynamical mode, but to a multiplicity of them (Deco & Jirsa 2012; Hansen et al., 2015; Golos et al., 2015) and that each of them may be associated, in general, to a different state-specific FC (Battaglia et al., 2012; Hansen et al., 2015; Kirst et al., 2016) so that the final static FC* results from averaging over a mixture of different states sampled in stochastic proportions. Therefore, to derive the FC* associated with a given input SC*, it is necessary to run explicit MFM simulations.

Gilson et al. (2016; 2018) have introduced iterative optimization procedures aiming at updating a current guess for the input SC* to a model in order to improve the match between the model output FC* and a target FC_{emp}. In this "effective connectivity" procedure –named as such by Gilson and coworkers, even if different from effective connectivity measures defined more typically in terms of statistical causality metrics (Valdes-Sosa et al., 2011)– connectome weights are iteratively and selectively adjusted as a function of the difference occurring between the current FC* and the target FC_{emp}. Such optimization leads to infer refined connectomes, that, with respect to empirical DTI SC matrix, may display non-symmetric connections (distinguishing thus between "feeder" and "receiver" regions as in Gilson et al., 2016) or enhanced inter-hemispheric connections, usually under-estimated by DTI (as in Gilson et al., 2018). Here we use a similar algorithm to learn a suitable non-linear virtual SC_{MFM}.

The initial $SC^*_{(0)}$ is taken to be a matrix with fully random entries. An MFM embedding such $SC^*_{(0)}$ is built and simulations are run to generate an output $FC^*_{(0)}$ which is compared to the target FC_{emp} of the subject for which FC-to-SC completion must be performed. The used $SC^*_{(0)}$ is then modified into a different $SC^*_{(1)} = SC^*_{(0)} + \lambda\Delta FC_{(0)}$ matrix, by performing a small update step in the direction of the gradient defined by the difference $\Delta FC_{(0)} = FC_{emp} - FC^*_{(0)}$. A new simulation is then run to produce a new $FC_{(1)}$. The produce is repeated generating new $SC_{(i)} = SC_{(i-1)} + \lambda\Delta FC_{(i-1)}$ until when the difference between $FC_{(i)}$ and the target FC_{emp} becomes smaller than a specified tolerance, i.e. $|\Delta FC_{(i)}| < \varepsilon$. The last generation $SC_{(i)}$ is then taken as non-linear virtual surrogate SC_{MFM} (see *Materials and Methods* for details).

Figure 4A provides an illustration of the nonlinear FC-to-SC completion when applied to subjects in the "SC_{emp}+FC_{emp}" subset. In the first step, the matrix $SC^*_{(0)}$ is random and there is no correlation between the output $FC^*_{(0)}$ and FC_{emp}. Advancing through the iterations, $SC^*_{(k)}$ develops gradually more complex internal structures and, correspondingly, the correlation between $FC^*_{(k)}$ and FC_{emp} increases until when it reaches the desired quality threshold, here set to $CC_{target} = 0.7$. This threshold quality is usually reached after ~1500 iterations. In the "SC_{emp}+FC_{emp}" subset we can even take advantage of the availability of the actual SC_{emp} to quantify as well the convergence of $SC^*_{(k)}$ toward SC_{emp}. Figure 4A shows that advancing through the iterations, the correlation between $SC^*_{(k)}$ and SC_{emp} improves as well. The expected quality of reconstruction, as estimated from results on the "SC_{emp}+FC_{emp}" subset is reported in Figure 4B and amounts to an expected correlation between SC_{MFM} and SC_{emp} of ~0.31.

We note that non-linear FC-to-SC completion, as for non-linear SC-to-FC completion, is a non- deterministic procedure, meaning that a different SC_{MFM} is generated depending on the starting initial condition $SC^*_{(0)}$. However, the different non-linear virtual surrogates lie at distances from the common actual ground truth SC_{emp} which are tightly concentrated around the median correlation. As revealed by Figure 4C, the reported correlations between SC_{MFM} and SC_{emp} were within a narrow interval of ±2.5% of the relative difference from the median distance for all the tested random initial conditions (30 per subject, see *Materials and Methods*), showing that the expected performance is poorly affected by the initial conditions. This stochastic aspect of the non-linear completion algorithm is going to allow us generating not just one but arbitrarily many completions, starting from each available empirical connectivity matrix (see later section).

Table 4 provides a compact pseudo-code for the non-linear FC-to-SC completion procedure (see *Materials and Methods* for all details). Non-linear FC-to-SC completions for the BOLD-only subjects in the considered ADNI dataset can be downloaded as Supplementary File S4.

Table 4. Pseudo-code for non-linear FC-to-SC completion

algorithm non-linear FC-to-SC completion **is**

external input: empirical FC (FC_{emp})
output: non-linear virtual SC (SC_{MFM})
fixed parameters: FC* fitting quality (CC_{target}), initial guess $SC^*_{(0)}$, learning rate λ, noise level (σ), simulation time (T), range to scan $G_{start} \leq G \leq G_{stop}$, range to scan $\tau_{start} \leq \tau \leq \tau_{stop}$, other frozen Wong-Wang neural mass parameters

begin
1. $FC^*_{(0)}$ = non-linear SC-to-FC completion starting from $SC^*_{(0)}$
2. Dist = corr($FC^*_{(0)}$, FC_{emp})
3. iteration = 0
while (Dist ≤ CC_{target})
iteration = iteration + 1
$SC^*_{(iteration)} = SC^*_{(iteration - 1)} + \lambda*(FC^*_{(iteration)} - FC^*_{(iteration)})$
$FC^*_{(iteration)}$ = non-linear SC-to-FC completion starting from $SC^*_{(iteration)}$
Dist = corr($FC^*_{(iteration)}$, FC_{emp})
end
return $SC_{MFM} = SC^*_{(iteration)}$
end

Data completion is self-consistent: bi-virtual connectivity matrices

SLM and MFM have thus the capacity to bridge from SC to FC or from FC to SC. When using these models for data completion, the input matrix is always an empirical matrix (SC_{emp} or FC_{emp}) and the output a surrogate virtual matrix (respectively, FC_{virt} or SC_{virt}, where the index "virt" refers generally to any completion algorithm, i.e. either using the SLM or the MFM models). However, the algorithms presented in Tables 1-4 can still be applied even when the input connectivity matrix is *already* a virtual matrix. In this case, the input could be surrogate matrices (SC_{virt} or FC_{virt}) from data completion and the output would be *bi-virtual* (respectively, FC_{bivirt} or SC_{bivirt}), i.e. twice virtual, since, to obtain them starting from an empirical input connectome, two different model-based procedures have to be chained. The final result of passing an original empirical connectome through two chained completion procedures is then a bi-virtual surrogate matrix of the same type (structural or functional) of the initially fed connectome. In other words, SC_{emp} is mapped to an SC_{bivirt} (passing through an intermediate FC_{virt} step) and FC_{emp} is mapped to an FC_{bivirt} (passing through an intermediate SC_{virt} step).

A first reason to evaluate bi-virtual connectivity matrices is to perform a self-consistency check of the data completion procedures. If the completion quality is good, then empirical connectomes and their bi-virtual counterparts should be highly related between them. A second reason would be to generate for each subject a fully virtual pair of connectomes, e.g. an SC_{virt} (or an FC_{virt}) paired with an FC_{bivirt} (or an SC_{bivirt}) that could be shared in a public dataset to avoid disclosing the actual private subject-specific empirical data FC_{emp} (or SC_{emp}).

Figure S4 shows the correspondence between empirical and bi-virtual SC and FC pairs, both when using chained linear (SLM-based) and nonlinear (MFM-based) completion procedures. We first evaluated the quality of SC_{bivirt} generation over the ADNI-subset of 88 subjects for which an SC_{emp} matrix was available (Figure S4A). Considering the linear bi-virtual completion chain SC_{emp} to FC_{SLM} to $SC_{bi\text{-}SLM}$ we obtained a median correlation between SC_{emp} and $SC_{bi\text{-}SLM}$ of ~0.63. Considering then the non-linear bi-virtual completion chain SC_{emp} to FC_{MFM} to $SC_{bi\text{-}MFM}$ we obtained a smaller median correlation between SC_{emp} and $SC_{bi\text{-}MFM}$ of ~0.58.

We then evaluated the quality of FC_{bivirt} generation over the ADNI-subset of 168 subjects for which an FC_{emp} matrix was available (Figure S4B). Considering the linear bi-virtual completion chain FC_{emp} to SC_{SLM} to $FC_{bi\text{-}SLM}$ we obtained a rather poor median correlation between FC_{emp} and $FC_{bi\text{-}SLM}$ of ~0.12. However, considering finally the non-linear bi-virtual completion chain FC_{emp} to SC_{MFM} to $FC_{bi\text{-}MFM}$ the median correlation between FC_{emp} and $FC_{bi\text{-}MFM}$ rose to ~0.59.

The empirical-to-bi-virtual correlations were always significant and, in all cases but $FC_{bi\text{-}SLM}$ well compatible with the empirically observed test-retest variability (Wang et al., 2012; Chen et al., 2015; Termenon et al., 2016). This non-trivial performance and, particularly, the fact that empirical-to-bi-virtual correlations in Figure S4 are even higher than empirical-to-virtual correlations in Figures 3, 4 or S1, establish the self-consistency of the model-based data completion procedures presented in Tables 1-4.

Note also that the lack of perfect identity between original seed empirical connectomes and their bi-virtual counterparts prevents the exact regeneration of the actual subject empirical data (see *Discussion* for the positive implications of this negative result).

Data completion is useful for unsupervised categorization

The compilation of large datasets, including connectivity data from structural and functional neuroimaging is considered essential for the development of algorithmic patient stratification and predictive approaches. Here, we have described four different types of connectomic data completion and studied their consistency. We now show that such completion procedures are useful for the algorithmic extraction of information.

For the sake of relative benchmarking, we study a proof-of-concept classification problem, separating the ADNI subjects into the three subgroups of *control*, *MCI*, and *AD*, based uniquely on empirical and/or virtual connectomic data. The target classification labels in these three groups are already provided within the ADNI dataset and we assume them to be exact (see *Materials and Methods* for a summary of the used stratification criteria). Possible input features for classification can be the following matrices (for the subject subsets for which they are available): empirical SC_{emp} or FC_{emp}; linear virtual SC_{SLM} or FC_{SLM}; nonlinear virtual SC_{MFM} or FC_{MFM}; and, finally, linear bi-virtual $SC_{bi\text{-}SLM}$ or $FC_{bi\text{-}SLM}$; and nonlinear bi-virtual $SC_{bi\text{-}MFM}$ or $FC_{bi\text{-}MFM}$. We chose to perform classification using a variant (Seiffert et al., 2010) of the random forest algorithm, which is particularly suitable when the number of input features is large with respect to the available data-points in the training set (Breiman, 2001), as it is in our case (every connectivity matrix has 4560 potentially independent entries, corresponding to the number of upper-diagonal matrix entries in our parcellation with $Q = 96$ regions, see *Materials and Methods*). For illustration, we present here results for the subset of ADNI subjects for which a DTI empirical connectome SC_{emp} is available. We thus discuss the relevance for classification of data completions, focusing thus on nonlinear SC-to-FC data virtual completion algorithms and their companion bi-virtual SC-to-FC-to-SC chains (Figure 5).

Here (and in the following), the dataset of 88 subjects with available SC_{emp} is randomly split into a training set and a testing set (with maintained relative proportions of subjects of the three Control, MCI and AD categories). The classifier is then trained on the training set and classification performance benchmarked on the testing set, to assess generalization capabilities. Classification performance can vary depending on the specific chosen split. Since we are here interested in quantifying lower bounds to the amount of information that the different types of features can potentially bear, more than assessing an expected performance, we repeated training multiple times to distillate a "purified subset" of high-performance classifiers, performing better than average classifiers (see *Materials and Methods* for details).

We considered first the baseline classification performance that we could tendentially achieve training random forest classifiers on the original SC_{emp} features themselves (Figure 5A, light blue; data completion not used at all). We then quantified tendential classification performance based on: the virtually completed FC_{MFM} (Figure 5A, dark green; empirical data not used at all); on the pairs made by the actual empirical SC_{emp} and the virtually completed FC_{MFM} (Figure 5A, magenta; exploits synergy between SC and FC, mixing empirical and virtual features); and, finally, on the pairs made by the virtually completed FC_{MFM} and the bi-virtual $SC_{bi\text{-}MFM}$ (Figure 5A, violet; exploits synergy between SC and FC features and empirical data are not used at all, replaced by their bi-virtual counterparts).

Figure 5B reports Receiver Operator Curve (ROC) analyses of the purified classifier performance for all Different ROC curves are derived for different choices of the target class to predict. In plain words, once the input features for a subject are fed into the classifier, the classifier returns a probability for this subject to belong to the target class. A hard threshold is arbitrarily set, such that a subject is labeled to be of the target type (or not) depending on the output probability being respectively larger (or smaller) than the adopted threshold. When a very high threshold is taken, only a few subjects will be labeled to belong to target class (there may be many "false negatives") but there will also be little "false positives". On the contrary, when the threshold is low, the number of "false negatives" will decrease but the number of "false positives" will increase. The ROC curve precisely describes how the fractions of "true" (TPR) and "false positives" (FPR) evolve by gradually lowering the probability threshold to classify a subject as belonging to target class, interpolating between the lower left corner (TPR = FPR = 0, maximum threshold, no subjects classified as belonging to target class) and the upper right corner (TPR = FPR = 1, minimum threshold, all subjects classified as belonging to target class). A ROC curve following the diagonal on the TPR/FPR plane would correspond to a classification performance level expectable from random guessing. Performance better than chance level is indicated by ROC curves lifting toward the upper left corner (TPR = 1, FPR = 0) or, equivalently, by increased "Area Under Curve" (AUC).

In Figure 5D, we also report distributions of the tendential classification *precision,* defined as the fraction of correctly classified subjects, at a fixed low *recall* level of 10%, defined as the fraction of subjects labeled by the classifier as belonging to the target class (see *Materials and Methods*). Once again in plain words, we expect that when the threshold is high, only a small fraction of subjects is going to be labeled as belonging to the target class

(low recall). However, we expect most of these subjects to be actual true positives since the classification threshold is conservative. Correspondingly the precision should be higher than for lower threshold when the recall will be higher but the number of false positives as well.

Figure 5B shows that for all the considered input features choices, above chance-level performances could be achieved by random forest classifiers after training and ensemble purification. Remarkably, the ROC curves for purified classifiers trained on empirical only features of a single type (SC_{emp}, light blue) or virtual only features of a single type (FC_{MFM}, dark green) were very similar, indicating that a comparable amount for classification-relevant information can be tendentially extracted from the original ADNI SC empirical data and the virtually completed FC connectomes. As anticipated, the use of SC_{emp} and FC_{MFM} as combined inputs (magenta) boosted tendential classification performance, leading to higher AUC (building on synergy redundancy or a mixture of both, see *Discussion*). Finally, this superior tendential performance of classifiers with combined SC and FC inputs was maintained when the original SC_{emp} was replaced by its bi-virtual counterparts $SC_{bi\text{-}MFM}$ (violet). Figure 5D shows that similar relations hold for the achieved tendential classifier precisions at 10% of recall. Thus, overall, we found that classification based on purely virtual features can be tendentially achieved after ADNI dataset completion and that tendential classification performance (in terms of both AUC and precision) is not expected to be much lower than when actual empirical data are used (see *Discussion* for the implications of these results).

Data augmentation

All data completion algorithms involve a stochastic component (apart from the purely deterministic linear FC-to-SC completion of Table 3). Therefore, running various times the algorithms will give rise to different virtual and bi-virtual connectomes, associated with the same initial empirical seed connectome. Such a feature allows generating out of a given empirical seed SC (or FC) connectome, not just one surrogate virtual FC (or SC) connectome but an arbitrarily large ensemble of surrogate virtual connectomes, forming the *virtual cohort* associated to a specific subject (see *Materials and Methods*). Every virtual cohort maintains a strict relation to its seed empirical subjects. In particular, distances between virtual connectomes sampled within two different virtual cohorts are always closely correlated to the distance between the respective seed connectomes of the two cohorts. Therefore, learning performed on virtual cohorts is expected to generate similar discrimination surfaces as learning on the original subjects, with the added benefits that training items in cohorts are way more numerous than the original empirical data. The use of wider ensemble of surrogate date with statistical distributions of multi-dimensional features equivalent to the original data is a common practice in machine learning, known as *data augmentation* (Yaeger et al., 1997; Taylor & Nitshcke, 2018), and very popular e.g. in object recognition (where surrogate training data are produced by clipping or variously transforming copies of the original training images). Data augmentation aims to expand the training dataset beyond the initially available data to boost the learning by a classifier of the target categories (e.g. object identities). Crucial for dataset augmentation applications is that the surrogate data generated are not just identical to the actual data with some added noise but are genuinely new and can serve as actual good guesses for alternative (unobserved) instances of data-points belonging to the same category. Obviously, new information cannot be created, but the extraction of the available information can still be facilitated, leading to a concrete improvement of the generalization performance of the classifiers after data augmentation. Given that inter-relations between virtual cohorts mirror inter-relations between empirical subjects, surrogate connectomes in virtual cohorts could be naturally used to improve connectome-based classification via data augmentation.

The close relation between the original data and the respective virtual cohorts is visually evident in Figure 5C, where a distance-respecting non-linear t-SNE projection (Van Der Maaten & Hinton, 2008) has been used to represent in two dimensions the virtual cohorts of surrogate virtual FC_{MFM}'s associated to the 88 subjects with available SC_{emp} (every dot corresponds here to the two-dimensional projection of a high-dimensional FC_{MFM}; 100 different virtual FC_{MFM}'s have been generated starting from each one of the 88 SC_{emp} connectomes). The connectomes composing the virtual cohorts represented in Figure 5C can be downloaded as Supporting File S5.

To prove that virtual cohorts can be used to perform data augmentation and boost connectome-based subject classification, we performed training of random forest classifiers expanding the training set to include not only one pair SC_{emp} / FC_{MFM} per subject but a larger group of pairs selected within the virtual cohort of Figure 5B. For training, SC was always set to SC_{emp} and the FC_{MFM} were chosen randomly among the 100 that were generated for each subject (excluding subjects in the testing set). The AUC curve for $SC_{bi\text{-}MFM}$ + FC_{MFM} classification augmented by the use of virtual cohorts of FC_{MFM} 's rather than just one FC_{MFM} instance in training is shown in Figure 5B as a dashed violet line. Comparison with the solid violet line indicates that classification performance was further improved by data augmentation. More specifically, we compare in Figure 5D the classification

precision (fraction of truly MCI subjects over the total number of subjects labeled as MCI by the classifier) at 10% recall (threshold selected to guarantee that at least 10% of the actual MCI subjects are classified as MCI subjects) or various classifier designs. The median precision at 10% recall that can be achieved classifying subjects based on actual SC_{emp} connectomes was of ~60.8%. The precision achievable by classifier trained on an equal number of surrogate FC_{MFM} connectomes (one surrogate FC_{MFM} connectome per subject) was slightly but significantly smaller, at ~59.5% (here, and in the following, significance assessed via Kruskal-Wallis group-level comparison, $p < 0.05$). However, by performing training on a 100-times larger cohort of virtual FC_{MFM} 's, the median performance level achievable with empirical data was not only restored but even significantly outperformed, reaching ~62.5%. This improvement is not a mere consequence of overfitting since we consider genuine generalization performance. In particular, performance is assessed via a cohort-adapted cross-validation approach, i.e. training data are selected from only an equilibrated subset of cohorts, and performance monitored on validation data from the complementary subset of cohorts. Last but not least, to a larger number of items used for training correspond also a larger number of validation items to classify (cohorts are used both in training and validation, see *Materials and Methods*). Therefore, the slight but significant precision improvement can be explained only by an enhancement of learning, due to the artificial boost of training dataset size provided by data augmentation.

The tendential precision of classification was further improved by the use of combined SC and FC features as input, even when fully virtual pairs (here, $SC_{bi\text{-}MFM}$ + FC_{MFM}) are used. In this case (and analogously, when mixed empirical/virtual pairs SC_{emp} + FC_{MFM} are used) the median precision rose indeed significantly to over ~64%. However, in this case, data augmentation by using cohorts rather than a single FC_{MFM} instance per subject did not yield further significant improvements of tendential performance.

Discussion

We have here demonstrated the feasibility of connectomic dataset completion (and even augmentation) using algorithms based on mean-field computational modeling. In particular, we have completed an ADNI gold standard connectomic dataset. We have then shown that the use of virtual connectomic data improves automated subject classification. Furthermore, classification based uniquely on surrogate data can approach the same performance levels as of empirical data. The capacity to extract more clinically-relevant information from empirical data is central for making progress in predictive and personalized neurology. Neurodegenerative diseases are a dramatic burden, linked to massive economic costs for healthcare growing in an aging society, outrun only by the quality of life decrease and intimate feeling of "mind fading away" experienced by patients and their friends and family. This is true in particular for Alzheimer's disease (AD), which is one of the most common types of dementia and one of the most widely studied progressive neurodegenerative disorders. Small and incomplete datasets for clinical research are certainly among the factors contributing to slow progress in the development of new diagnostic and therapeutic tools. Our methodological proposal aims precisely at relieving these two problems.

Data completion procedures allowed us to infer Functional Connectivity when only Structural Connectivity was available or Structural Connectivity (SC) when only Functional Connectivity (FC) were available. Such procedures for data completion could easily be implemented within popular neuroinformatic platforms as The Virtual Brain (TVB). TVB provides practical graphical interfaces or fully scriptable code-line environments for "plug-and-play" large-scale brain network modeling, signal emulation, and dataset management, including simulating SC and FC with adjustable complexity MFMs or SLMs (Sanz-Leon et al., 2013). In this way, capitalizing on the software built-in capabilities, even the more elaborated non-linear completion algorithms could become accessible to non-expert users with only a little training.

The possibility of having access to both types of connectomic information brought up by model-based data completion is vital because structural and functional connectivity convey complementary information. It has been shown for instance, that analyses of SC-to-FC inter-relations can yield better characterizations and group discriminations than analyses of SC or FC alone in a variety of pathologies or conditions (Zhang et al., 2011; Davis et al., 2012; Zimmermann et al., 2016; Straathof et al., 2019).

Indeed, FC networks in the resting-state do not merely mirror SC but are believed to be the by-product of complex dynamics of multi-scale brain circuits (Honey et al., 2007; Deco et al., 2011). As such, they are constrained but not entirely determined by the underlying anatomy (encoded in the SC matrix) and FC also carries valuable information about the dynamic regime giving rise to the observed resting-state activity fluctuations (Hansen et al., 2015). In particular, brain networks are thought to operate at a regime close to

criticality. For a fixed SC, the resulting FC would be different depending on how close to a critical working point dynamics is tuned (Deco et al., 2013; Hansen et al., 2015). This information that brain networks are supposed to operate close to a critical boundary is used to generate the surrogate virtual FC_{MFM}, when performing non-linear SC-to-FC completion. Thus, FC_{MFM} carries indirectly extra information about a (putative) dynamic regime that was not conveyed by the original empirical SC (nor by virtual completions with linear SLM-based pipelines). This may be one of the reasons why the classification performance using combined empirical SC and virtual FC is superior to the one based on empirical SC alone, even if the virtual FC_{MFM}'s have been derived from the empirical SC. The data completion procedure has "injected" useful information in FC_{MFM}, in particular, information about the dynamical regime, in which the brain is expected to be in. This generation of supplementary information is possible due to the mechanistic nature of the virtual brain models, taking advantage of the synergy of dynamic modeling constrained by subject-specific structural data. The performance improvement observed in classification suggests that the closeness-to-criticality hypothesis informing non-linear data completion is a reasonable dynamic network mechanism for the resting state. It also provides further support for the predictive capacity of connectome-based personalized brain network models.

When both empirical SC and FC were available, we could measure the quality of reconstruction achieved by our models. The correlation reached between empirical and reconstructed connectivity matrices is only moderate, however. There are multiple reasons for this limited performance. One evident reason is the simplicity of the neural mass model adopted in our proof-of-concept illustration. The Wong-Wang neural mass model is able only to express two states of lower or higher local activation (Wong & Wang, 2006). Instead, neuronal populations can display a much more extensive repertoire of possible dynamics, including e.g., coherent oscillations at multiple frequencies, bursting, or chaotic trajectories (Stefanescu & Jirsa, 2008; Spiegler et al., 2011). Synchronization in a network depends on various factors, including frequency, network topology, and time delays via signal propagation, all of which have been ignored here and in large parts of the literature (Deco et al., 2009; Petkoski & Jirsa, 2019). It is acknowledged that delay-less approaches serve as a useful approximation (Deco et al. 2015). Nevertheless, we are aware that our choice to restrict our analyses on the subset of activation-based mechanisms introduces critical limitations. Indeed, our models, ignoring delay-mediated synchronization, are incapable of capturing a range of dynamic oscillatory behaviors, such as multifrequency coupling or multiphase coupling. More sophisticated mean-field virtual brain models than the very simple one here explored could thus reach superior performance (see e.g. Stefanovski et al., 2019). Yet, even such a simple model, achieving such a limited reconstruction performance proved to be consistent and useful. First, when concatenating data completion pipelines to give rise to bi-virtual data, we found a robust self-consistency, i.e. remarkable matching between e.g. the original SC and the bi-virtual SCbi-MFM, generated via the intermediated FCMFM step. Second, classification performance can be improved by using our virtual data (or maintained, by using our virtual instead of empirical data). Another reason for the only moderate quality of reconstruction is our choice of second-order moments, which is covariance and FC, as data feature for performance evaluation. The associated generative model giving rise to a stationary Gaussian probability distribution is a linear network model (Jaynes, 1957; Haken, 1983). Only a linear model evaluated against a stationary time-series generated by a linear process would provide optimal performance values. As we here discussed several times, resting-state dynamics do not satisfy these properties, and thus, inferior performance is expected.

Together, these findings show that even if the reconstruction quality of our model-based completion procedures is modest, a meaningful relationship with the original seed data is still maintained, even after two steps of virtual completion. The use of simple models has the additional advantage of being less computationally expensive to simulate. This is particularly true for SLMs in which no attempt to emulate the non-linearity of neural population dynamics are made but for which the availability of analytical formulas for completion even removes the need for direct simulation. On the other hand, nonlinear MFM simulations can take very long times, which can become an obstacle, especially for nonlinear FC-to-SC completion, where multiple simulations must be run. However, SLM-based procedures cannot be used for data augmentation, because they cannot generate a multiplicity of different surrogate connectomes starting from a shared seed empirical one.

We have shown that data augmentation can boost classification performance. Indeed, the redundancy present in stochastic ensembles of virtual connectivity matrices, generated via swarms of MFM simulations with different random initial conditions and noise realizations, can help training. We stress once again that we should not interpret data augmentation as a way to increase the information contained in the dataset: we cannot create information! Such is a consequence of the information theoretic concept of data processing inequality (Cover & Thomas, 2006). But it is also known that redundant information can improve the performance of decoding and

classification (Guyon & Elisseeff, 2003), which is the key point that we exploit here in model-based connectomic data completion. Computational models such as MFM do not provide mappings between input and output connectomes, but rather between statistical ensembles of connectomes, with both mean and correlated dispersion realistically shaped by trustworthy non-linear dynamics. In other words, differences between alternative connectomes in a generated surrogate virtual cohort are not mere "noise", but reflect realistic data-compliant possibilities of variation. The different connectome realizations sample indeed the specific landscapes of possible FCs that may be compatible with a given SCs, degenerate because the allowed dynamics to unfold along with low-dimensional manifolds, rather than being frozen in strict vicinity of a trivial fixed point (Mehrkanoon et al., 2014; Pillai & Jirsa, 2017).

Data augmentation constitutes thus a possible way to generate more massive training sets for machine learning applications, as far as the seed empirical dataset is well representative of the expected population variability. However, by capitalizing exclusively on redundancy, augmentation cannot replace the gathering of more empirical data (Carrillo et al., 2012; Toga et al., 2016). Unfortunately, federation (or even mining) of data is often impeded by unavoidable juridical concerns linked to strict and diverse regulations (Dulong de Rosnay, 2017; Thorogood et al., 2018) The use of virtual cohorts may relieve this burden. Virtual cohorts maintain their statistical relation to the original data, in a way sufficiently good to be exploitable for classification, but do not precisely match the original data. They maintain an inherent variability, indeed, due to the nature of the SC-to-FC link mediated by complex dynamics. Virtual data carry information operationally equivalent to the one carried by empirical data but not the same information. As such, it is not possible to reconstruct the original subject data from virtualized connectomes, and privacy concerns are considerably reduced if not entirely removed. We thus anticipate a near future in which virtual cohorts, providing vast numbers of virtual and bi-virtual connectivity information, would play an increasing role in massive data-driven explorations of factors predictive of neurodegenerative disease progression.

Materials and Methods

Data Sample

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

Raw neuroimaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) GO/2 studies (Wyman et al., 2013; Beckett et al., 2015) were downloaded for 244 subjects. These included T1w images for all subjects, as well as DWI and rsfMRI images for separate cohorts of subjects. An additional 12 subjects for which both DWI and rsfMRI were acquired in the same session were identified and their data also downloaded.

A volumetric 96-ROI parcellation was defined on the MNI template and consisted of 82 cortical ROIs from the Regional Map parcellation (Kötter & Wanke, 2005) and an additional 14 subcortical ROIs spanning the thalamus and basal ganglia. Details on the construction of the 96-ROI parcellation can be found in Bezgin et al (2017).

Among the 244 subjects we downloaded, 74 were control subjects, while the others were patients at different stages of the pathology progression. In this study, we performed a rough coarse-graining of the original ADNI labels indicating the stage or type of pathology. We thus overall labeled 119 patients as "MCI" (grouping together the labels 4 patients as "MCI", 64 as "EMCI" and 41 as "LMCI") and 51 patients as "AD".

Overall, T1 and DTI were jointly available for 88 subjects (allowing to reconstruct structural connectivity (SC) matrix), and T1 and fMRI for 178 (allowing to reconstruct functional connectivity (FC)). However, among the 244 subjects we downloaded, only 12 subjects (referred to as the "SC_{emp}+FC_{emp}" subset) had a complete set of structural and functional images (T1, DTI, fMRI), hinting at how urgently needed are data completion and augmentation.

Data Preprocessing

Neuroimaging data preprocessing was done using a custom Nipype pipeline implementation (Gorgolewski et al., 2011). First, raw neuroimaging data were reconstructed into NIFTI format using the dcm2nii software package (https://www.nitrc.org/projects/dcm2nii/). Skull stripping was performed using the Brain Extraction Tool (BET) from the FMRIB Software Library package (FSL v5) for all image modalities prior to all other preprocessing steps. Brain extraction of T1w images using BET was generally suboptimal and was supplemented by optiBET (Lutkenhoff et al., 2014), an iterative routine that improved brain extractions substantially by applying transformations and back-projections between the native brain mask and MNI template space. Segmentation of the T1w images was performed using FSL's FAT tool with bias field correction to obtain into three distinct tissue classes.

To improve the registration of the ROI parcellation to native space, the parcellation was first nonlinearly registered to a publicly-available older adult template (aged 70-74 years, Fillmore et al., 2015) using the Advanced Normalization Tools (ANTS, Avants et al., 2011) software package before subsequent registrations.

Diffusion-weighted images were preprocessed using FSL's *eddy* and *bedpostx* tools. The ROI parcellation was first nonlinearly registered to each subject's T1w structural image and then linearly registered to the DWI image using ANTS.

rsfMRI data were preprocessed using FSL's FEAT toolbox. Preprocessing included motion correction, high-pass filtering, registration, normalization and spatial smoothing (FWHM: 5 mm). Subjects with excessive motion were excluded from our sample. Global white matter and cerebrospinal fluid signals (but not global mean signal) were linearly regressed from the rsfMRI data.

All images were visually inspected following brain extraction and registrations to ensure correctness.

SC Construction

Details of tractography methods for reconstructing each subject's structural connectome can be found in Shen et al (2019 a, b). Briefly, FSL's *probtrackx2* was used to perform tractography between all ROIs. The set of white matter voxels adjacent to a grey matter ROI was defined as the seed mask for that particular ROI. Grey matter voxels adjacent to each seed mask were used to define an exclusion mask. For intrahemispheric tracking, an additional exclusion mask of the opposite hemisphere was additionally defined. Tractography parameters were set to a curvature threshold of 0.2, 5000 seeds per voxel, a maximum of 2000 steps and a 0.5 mm step length. The connection weight between each pair of ROIs was computed as the number of streamlines detected between the ROIs, divided by the total number of streamlines sent from the seed mask. This connectivity information was compiled for every subject in a matrix of empirical structural connectivity SC_{emp}.

rsfMRI Timeseries and FC Construction

Empirical rsfMRI time-series for each ROI were computed using a weighted average approach that favored voxels nearer the center of each ROI (Shen et al., 2012). Each subject's matrix of empirical functional connectivity FC_{emp} was determined by Pearson correlation of these recorded rsfMRI time-series.

SLM models

The SLM model used in this study is a linear stochastic system of coupled Ornstein-Uhlenbeck processes which is deeply investigated in (Saggio et al., 2016). For each brain region, neural activity $x_i(t)$ is modeled as a linear stochastic model, coupled to the fluctuations of other regions:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \sigma\xi(t) \quad (1)$$

where $\boldsymbol{A}$ is the coupling matrix, ξ is a normal Gaussian white noise and σ the standard deviation of the local drive noise. The coupling matrix $\boldsymbol{A}$ can be written as:

$$\boldsymbol{A} = -\boldsymbol{I} + G.\boldsymbol{W} \quad (2)$$

where $\boldsymbol{I}$ is the identity matrix, G is the global coupling parameter and $\boldsymbol{W}$ is a weight matrix set to match SC_{emp}. The negative identity matrix guarantees that the nodes have a stable equilibrium point. If all the eigenvalues of $\boldsymbol{A}$ are negative, which happens for all positive values of $G < G_{critic} = 1/max(\lambda_i)$ (where λ_i are the eigenvalues of

W), the system will be in an equilibrium state. After some mathematical steps (Saggio et al., 2016), the covariance matrix between regional fluctuations can be analytically expressed at this critical point G_{critic} as:

$$\boldsymbol{C} = \frac{-\sigma^2}{2}\boldsymbol{A}^{-1} \quad (3)$$

whose normalized entries provide the strength of functional connectivity between different regions. The noise strength can be arbitrarily set at the critical point since it provides only a scaling constant to be reabsorbed into the Pearson correlation normalization. However, the only parameter that needs to be explored is G, whose range goes from $G_{min} = 0$, i.e. uncoupled nodes, to slightly before $G_{critic} = 1/max(\lambda_i)$, or $G_{max} = G_{critic} - \epsilon$. In Figure S1A, running explicit simulations of SLM models for different values of coupling G and evaluating on the "FC_{emp} + SC_{emp}" subset of subjects the match between the simulated and empirical activity correlation matrices, we confirm (cf. e.g. Hansen et al., 2015) that the best match (max of Pearson correlation between the upper-triangular parts of the empirical and virtual FCs) is obtained at a slightly subcritical point for $G^* = G_{critic} - \epsilon$.

Linear SC-to-FC and FC-to-SC completion

Every subject in the "FC_{emp} + SC_{emp}" subset has a different value of G^*. To infer FC_{SLM} from SC_{emp} for the subjects with missing functional data, we chose to always use a common value $G^*_{ref} = 0.83$, which is the median of G^* for all 12 "FC_{emp} + SC_{emp}" subjects (the error made in doing this approximation is estimated to be less than 1% in Fig. S1C). When the connectome FC_{emp} is not known, equations (2) and (3) can directly be used to evaluate the covariance matrix ***C*** (setting σ = 1 and $G = G^*_{ref}$). We then estimate the regional fluctuation covariance from these inferences and normalize it into a Pearson correlation matrix to infer FC_{SLM} (*See pseudo-code in Table 1, line 1*). Linear FC_{SLM} completions for our ADNI dataset can be downloaded as Supporting File S1.

To infer SC_{SLM} from FC_{emp}, we invert the analytical expressions of eqs. (2) and (3) and always set $\sigma = 1$ and $G = G^*_{ref}$ leading to:

$$\boldsymbol{W}^* = -\boldsymbol{C}^{-1}/G^*_{ref} \quad (4)$$

where ***C*** is the covariance matrix estimated from empirical BOLD time-series. The linearly completed SC_{SLM} is then set to be identical to ***W**** setting its diagonal to zero to avoid offsets, which would be meaningless given the conventional choice of noise σ which we have made (*see Table 3, line 3*). Note that all the free parameters of the SLM model appear uniquely as scaling factors and do not affect the (normalized) correlation of the inferred SC_{SLM} with the SC_{emp}. However, the absolute strengths of inferred structural connections remain arbitrary, with only the relative strengths between different connections being reliable (since unaffected by arbitrary choices of scaling parameters; *see pseudo-code in Table 3*). Linear SC_{SLM} completions for our ADNI dataset can be downloaded as Supporting File S3.

MFM models

For non-linear completion algorithms, we performed simulations of whole-brain mean-field models analogous to Deco et al. (2013) or Hansen et al. (2015). We used a modified version of the mean-field model designed by Wong and Wang (2006), to describe the mean neural activity for each brain region, following the reduction performed in (Deco et al., 2013). The resulting neural mass equations are given by:

$$\frac{dS_i}{dt} = \frac{-S_i}{\tau_S} + (1 - S_i)\gamma R_i + \sigma\eta_i(t) \quad (5)$$

$$R_i = \frac{ax_i - b}{1 - exp\,[-d(ax_i - b)]} \quad (6)$$

$$x_i = \omega J_N S_i + J_N G \sum_j C_{ij} S_j + I_0 \quad (7)$$

where S_i represents NMDA synaptic input currents and τ_S the NMDA decay time constant; R_i is collective firing rates; $\gamma = 0.641$ is a kinetic parameter; $a = 270(V.nC)^{-1}$, $b = 108Hz$, $d = 0.154s$ are parameters values for the input-output function; x_i are the total synaptic inputs to a regions; $J_N = 0.2609nA$ is an intensity scale for

synaptic currents; ω is the relative strength of recurrent connections within the region; C_{ij} are the entries of the SC_{emp} matrix reweighted by global scale of long-range connectivity strength *G* as a control parameter; σ is the noise amplitude, and η_i is a stochastic Gaussian variable with a zero mean and unit variance. Finally, I_0 represents the external input and sets the level of regional excitability. Different sets of parameters yield different neural network dynamics and, therefore, patterns of FC_{MFM} non-stationarity.

To emulate BOLD fMRI signals, we then transformed the raw model output activity x_i through a standard Balloon-Windkessel hemodynamic model. All details of the hemodynamic model are set according to Friston et al. (2003).

Non-linear SC-to-FC completion

In general, our simple MFM model has three free parameters at the level of the local neural mass dynamics (τ, ω, and I_0) and one free global parameter *G.* Since changing the values of ω and I_0 had lesser effects on the collective dynamics of the system (see Figure S2), we set their values to $\omega = 0.9$ and $I_0 = 0.32$ respectively and remain then just two free parameters which we allow to vary in the ranges $G \in [1\ 3]$ and $\tau \in [1\ 100]$ *ms* when seeking for an optimal working point of the model. As revealed by the analyses of Figure 3, the zone in this restricted parameter space associated with the best FC-rendering performance can be identified through the joint inspection of three scores, varying as a function of both *G* and τ. The first criterion is the spatial heterogeneity of activation (*see Table 2, line 2.5*) computed by taking the coefficient of variation of $BOLD_{MFM}$ time-series.

By computing the Pearson correlation coefficient of upper-triangular between FC_{MFM} and FC_{emp} for every subject from "SC_{emp} + FC_{emp}" subset (*see Table 2, line 2.3*), we obtained a best-fitting zone in a narrow concave stripe (see Figure 3A for one subject); (G^*, τ^*) parameter set, bring the system to this best-fitting zone and values lower than this is (G^-, τ^-) set and higher values are (G^+, τ^+). This non-monotonic behavior of yellow zone in G/τ plane occurs where three criteria are jointly met; the second criterion is the clustering coefficient of time-average FC_{MFM} matrices (*see Table 2, line 2.6*) and finally, the third criterion is the clustering coefficient of FCD_{MFM} matrices (*see Table 2, line 2.6*), where the FCD matrices were computed for an arbitrary window. By knowing the optimal working point of the system where all three criteria are jointly optimum (*see Table 2, line 2*), we freeze the algorithm and run the non-linear SC-to-FC data completion for 76 subjects (*see Table 2, lines 3 to 5*). Non-linear FC_{MFM} completions for our ADNI dataset can be downloaded as Supporting File S2.

Non-linear FC-to-SC completion

We implemented a heuristic approach to infer the most likely connectivity matrix (i.e. Effective Connectivity) that maximizes the similarity between empirical and simulated functional connectivity. As an initial point, we considered a random symmetric matrix and removed diagonal as $SC^*_{(0)}$ (*see Table 4, line 1*) and run the algorithm in Table 2 in order to simulate the $FC^*_{(0)}$. Then iteratively we adjusted the SC as a function of the difference between the current FC and empirical FC (*see Table 4, line 2*), in other words $SC^*_{(1)} = SC^*_{(0)} + \lambda \Delta FC_{(0)}$ where $\Delta FC_{(0)} = FC_{emp} - FC^*_{(0)}$ and λ is the learning rate (*see Table 4, line 3*). The iteration will stop when the correlation between FC_{emp} and $FC^*_{(k)}$ reaches to the threshold $CC_{target} = 0.7$, and giving the $SC^*_{(k)}$ as SC_{MFM}. All the parameter used in this section is identical to the non-linear SC-to-FC completion procedure. Nonlinear SC_{MFM} completions for our ADNI dataset can be downloaded as Supporting File S4.

Bi-virtual data completion

The pipelines for data completion described above can be concatenated, by performing e.g. FC-to-SC completion on a virtually FC or SC-to-FC completion on a virtual SC (rather than actual FC_{emp} or SC_{emp}, respectively). In this way, one can create bi-virtual counterparts $SC_{bi\text{-}MFM}$ ($FC_{bi\text{-}MFM}$) or $SC_{bi\text{-}SLM}$ ($FC_{bi\text{-}SLM}$) for any of the available empirical SC_{emp} (FC_{emp}) by applying in sequence non-linear MFM-based or linear SLM-based procedures for SC-to-FC and then FC-to-SC completion (or, conversely, FC-to SC followed by SC-to-FC completions). Linear bi-virtual completions for our ADNI dataset can be downloaded as Supporting File S6 (for $SC_{bi\text{-}SLM}$ and $FC_{bi\text{-}SLM}$) and non-linear completions as Supporting File S7 (for $SC_{bi\text{-}MFM}$ and $FC_{bi\text{-}MFM}$).

Supervised subject classification

As previously mentioned, we separated subjects in our ADNI-derived dataset in three subgroups: "controls", "MCI" and "AD". We focus here on the "one vs all" task of classifying "MCI" subjects against "controls" or "AD", given that early diagnosis is one of the aims for which our data completion and augmentation approaches could use in perspective. Subjects (the actual ones or their associated virtual counterparts) are thus labeled as "positive" when belonging to the MCI subgroup or "negative" otherwise. We construct classifiers based on different types of input features. For classification based on empirical data only, we used as input features a vector of the independent entries (upper-triangular part) of the SC_{emp} matrix (light blue color in Figure 5). For classification based on single-type virtual data only, we used as input features a vector of the independent entries of the non-linearly completed FC_{MFM} matrix (green color in Figure 5). We then performed classification based on combined (empirical) SC and (virtual) FC, by providing as input features the concatenated vectors of the upper triangular parts of SC_{emp} and FC_{MFM} (magenta color in Figure 5). Finally, we performed classification based on combined (bi-virtual) SC and (virtual) FC, by providing as input features the concatenated vectors of the upper triangular parts of $SC_{bi\text{-}MFM}$ and FC_{MFM} (violet color in Figure 5). The same training and performance assessment schemes were used for all types of input features (and could be generalized to other combinations of input features or target positive labels as well).

We chose as classifier a boosted ensemble of 50 shallow decision trees, training it using the RUSBoost algorithm (Seiffert et al., 2010), particularly adapted to data in which "positive" and "negative" labels are unbalanced. We adopted a 5-fold cross-validation approach and quantified fractions of true and false positives (numbers of true or false positives over total number of actual positives) as well as precision (number of true positives over numbers of items classified as positives, i.e. fraction of the positives correctly classified as such) during generalization (i.e. prediction performed on the folds of data not actually used for training). True positive fraction (or recall), false negative fraction and precision depend all on an arbitrary threshold to be applied to the classifier ensemble output to decide for positivity of not of the input data. In Figure 5B, receiver operator curves (ROC) generated by smoothly growing this threshold between a minimum and a maximum value and plotting how the True and False positive fractions vary for the different threshold values. In Figure 5D, precision was evaluated at the threshold providing the recall value closest to 10%. The entire cross-validated training and performance estimation value was repeated 5000 times. Since, in this context, we are interested in quantifying how far can be pushed lower bounds to the amounts of extractable information from the different type of features (rather than in building actual classifiers for specific applications), out of the 5000 classifier training experiments, we discarded the runs leading to weaker performance and retained only the 500 ones leading to cross-validated performances in the top decile. We thus plot in Figure 5B, the median ROCs over the 500 retained forest training experiments, for every chosen combination of inputs. Analogously, in Figure 5D, we plot distributions of achieved precision at 10%-recall over training runs in this top-decile. Our purified training procedure estimates thus how high is the performance that could be achieved tendentially by classifiers trained in different input feature combinations. We must keep in mind, however, to avoid confusion, that tendential performance estimates in Figures 5B and 5D are by construction superior to the expected medians, whenever pre-selection of top-performing classifiers was not applied. Once again, this is not a problem when being interested, as in the context of this study, in estimating lower bounds to tendentially achievable performances.

Virtual cohorts and data augmentation

In data augmentation, the size of the original dataset (limited by the total number of subjects with available empirical connectomic information of either SC or FC type) is artificially boosted by generating a much larger number of virtual subjects with multiple alternative (but all equally valuable) completions of the missing connectomic data. Concretely, to generate the virtual cohort dataset, we took the 88 subjects in the SC_{emp} only dataset (including 21 AD subjects, 35 MCI, and 32 Control subjects) and run for each of them the non-linear SC-to-FC completion algorithm 100 times, using each time a different random seed. The net result was a group of 100 alternative FC_{MFM} instances for each of the subjects, yielding in total a virtual cohort of 8800 FC_{MFM} matrices to be potentially used for classifier training. Such a cohort can be downloaded as Supporting File S5. To generate Figure 5C, showing a dimensionally reduced representation of the relative distances between these 8800 virtual matrices, we used an exact t-SNE projection (Van Der Maaten and Hinton, 2008) of the vectors of upper-triangular parts of the different FC_{MFM} 's toward a two-dimensional space, using a default perplexity value of 30 and no-exaggeration.

We then performed data augmentation during classifier training based on this virtual cohort. For classification based on single-type virtual features only (i.e. only FC_{MFM}), we used the entire cohorts of 8800 FC_{MFM} matrices to select from training and testing sub-folds during cross-validation. This means that both the size of training and validation sub-datasets were enlarged. Importantly, however, when sampling sub-folds for 5-fold cross-validated training, we still excluded from the training set all matrices deriving from a randomly selected fifth of the subjects. In this way the validation subset included not only different matrices than the one used for training but, beyond that, different matrices generated from subjects that did not contribute any matrix to the training set. In this way, we can be sure that improved cross-validated performance estimations do not reflect over-fitting but actually improved learning. To perform classification based on combined SC_{emp} and FC_{MFM} , we constructed an augmented dataset by including for each of the subjects 100 virtual pairs of connectomes, combining the same SC_{emp} with each of the possible 100 FC_{MFM} 's. Finally, when combining bi-virtual SC and virtual FCs, we analogously always combined the same $SC_{bi\text{-}MFM}$ instance with each of the 100 FC_{MFM} 's available for each subject. In principle, it would have been possible to generate as well a multiplicity of alternatives same $SC_{bi\text{-}MFM}$ counterparts, leading to an even more varied augmented dataset, but we were limited by the computational resources needed to build a multiplicity of alternative $SC_{bi\text{-}MFM}$'s for each of the subjects in our ADNI-derived dataset (nonlinear FC-to-SC completion is way harder computationally than SC-to-FC completion). Finally, tendential true and false positive fractions and precisions were evaluated in the same way as for the non-augmented datasets (multiple training runs, followed by purification), i.e. via retaining only the top decile of 5000 independent cross-validated training runs over the augmented datasets.

Acknowledgements:

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, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; 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. DB acknowledges support from the EU Innovative Training Network "i-CONN" (H2020 ITN 859937) and VJ acknowledges funding by the European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2) and H2020 Research and Innovation Action grants VirtualBrainCloud.

References

Aerts H, Schirner M, Jeurissen B, Van Roost D, Achten E, Ritter P, Marinazzo D. Modeling brain dynamics in brain tumor patients using The Virtual Brain. eNeuro. 2018; 5(3). https://dx.doi.org/10.1523/eneuro.0083-18.2018

Allen GI, Amoroso N, Anghel C, Balagurusamy V, Bare CJ, Beaton D, Bellotti R, Bennett DA, Boehme KL, Boutros PC, et al. Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease. Alzheimer's & Dementia. 2016; 12(6):645–653. https://doi.org/10.1016/j.jalz.2016.02.006

Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage. 2011; 54(3):2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025

Battaglia D, Witt A, Wolf F, Geisel T. Dynamic effective connectivity of inter-areal brain circuits. PLoS computational biology. 2012; 8(3):e1002438. https://doi.org/10.1371/journal.pcbi.1002438

Beckett LA, Donohue MC, Wang C, Aisen P, Harvey DJ, Saito N, Initiative ADN, et al. The Alzheimer's Disease Neuroimaging Initiative phase 2: Increasing the length, breadth, and depth of our understanding. Alzheimer's & Dementia. 2015; 11(7):823–831. https://doi.org/10.1016/j.jalz.2015.05.004

Bezgin G, Solodkin A, Bakker R, Ritter P, McIntosh AR. Mapping complementary features of cross-species structural connectivity to construct realistic "Virtual Brains". Human brain mapping. 2017; 38(4):2080–2093. https://doi.org/10.1002/hbm.23506

Braak H, Alafuzoff I, Arzberger T, Kretzschmar H, Del Tredici K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta neuropathologica. 2006; 112(4):389–404. https://doi.org/10.1007/s00401-006-0127-z

Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta neuropathologica. 1991; 82(4):239–259. https://doi.org/10.1007/BF00308809

Breiman L. Random forests. Machine learning. 2001; 45(1):5–32. https://doi.org/10.1017/CBO9781107415324.004

Carrillo MC, Bain LJ, Frisoni GB, Weiner MW. Worldwide Alzheimer's disease neuroimaging initiative. Alzheimer's & Dementia. 2012; 8(4):337–342. https://doi.org/10.1016/j.jalz.2012.04.007

Casanova R, Barnard RT, Gaussoin SA, Saldana S, Hayden KM, Manson JE, Wallace RB, Rapp SR, Resnick SM, Espeland MA, et al. Using high-dimensional machine learning methods to estimate an anatomical risk factor for Alzheimer's disease across imaging databases. NeuroImage. 2018; 183:401–411. https://doi.org/10.1016/j.neuroimage.2018.08.040

Chen B, Xu T, Zhou C, Wang L, Yang N, Wang Z, Dong HM, Yang Z, Zang YF, Zuo XN, et al. Individual variability and test-retest reliability revealed by ten repeated resting-state brain scans over one month. PLoS One. 2015; 10(12):e0144963. https://doi.org/10.1371/journal.pone.0144963

Chiesa PA, Cavedo E, Lista S, Thompson PM, Hampel H, Initiative APM, et al. Revolution of resting-state functional neuroimaging genetics in Alzheimer's disease. Trends in neurosciences. 2017; 40(8):469–480. https://doi.org/10.1016/j.tins.2017.06.002

Cover TM, Thomas JA. Elements of information theory. John Wiley & Sons; 2006. https://doi.org/10.1002/047174882X

Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert MO, Chupin M, Benali H, Colliot O, Initiative ADN, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. neuroimage. 2011; 56(2):766–781. https://doi.org/10.1016/j.neuroimage.2010.06.013

Davis SW, Kragel JE, Madden DJ, Cabeza R. The architecture of cross-hemispheric communication in the aging brain: linking behavior to functional and structural connectivity. Cerebral cortex. 2012; 22(1):232–242. https://doi.org/10.1093/cercor/bhr123

Deco G, Jirsa V, McIntosh AR, Sporns O, Kötter R. Key role of coupling, delay, and noise in resting brain fluctuations. Proceedings of the National Academy of Sciences. 2009; 106(25):10302–10307. https://doi.org/10.1073/pnas.0901831106

Deco G, Jirsa VK. Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. Journal of Neuroscience. 2012; 32(10):3366–3375. https://doi.org/10.1523/JNEUROSCI.2523-11.2012

Deco G, Jirsa VK, McIntosh AR. Emerging concepts for the dynamical organization of resting-state activity in the brain. Nature Reviews Neuroscience. 2011; 12(1):43. https://doi.org/10.1038/nrn2961

Deco G, Ponce-Alvarez A, Mantini D, Romani GL, Hagmann P, Corbetta M. Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. Journal of Neuroscience. 2013; 33(27):11239–11252. https://doi.org/10.1523/JNEUROSCI.1091-13.2013

Deco G, Tononi G, Boly M, Kringelbach ML. Rethinking segregation and integration: contributions of whole-brain modelling. Nature Reviews Neuroscience. 2015; 16(7):430. https://doi.org/10.1038/nrn3963

Dennis EL, Thompson PM. Functional brain connectivity using fMRI in aging and Alzheimer's disease. Neuropsychology review. 2014; 24(1):49–62. https://doi.org/10.1007/s11065-014-9249-6

Fillmore PT, Phillips-Meek MC, Richards JE. Age-specific MRI brain and head templates for healthy adults from 20 through 89 years of age. Frontiers in aging neuroscience. 2015; 7:44. https://doi.org/10.3389/fnagi.2015.00044

Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage. 2003; 19(4):1273–1302.

Galán RF. On how network architecture determines the dominant patterns of spontaneous neural activity. PLoS One. 2008; 3(5):e2148. https://doi.org/10.1371/journal.pone.0002148

Ghosh A, Rho Y, McIntosh AR, Kötter R, Jirsa VK. Noise during rest enables the exploration of the brain's dynamic repertoire. PLoS computational biology. 2008; 4(10):e1000196. https://doi.org/10.1371/journal.pcbi.1000196

Gilson M, Deco G, Friston KJ, Hagmann P, Mantini D, Betti V, Romani GL, Corbetta M. Effective connectivity inferred from fMRI transition dynamics during movie viewing points to a balanced reconfiguration of cortical interactions. Neuroimage. 2018; 180:534–546. https://doi.org/10.1016/j.neuroimage.2017.09.061

Gilson M, Moreno-Bote R, Ponce-Alvarez A, Ritter P, Deco G. Estimation of directed effective connectivity from fMRI functional connectivity hints at asymmetries of cortical connectome. PLoS computational biology. 2016; 12(3):e1004762. https://doi.org/10.1371/journal.pcbi.1004762

Golos M, Jirsa V, Daucé E. Multistability in large scale models of brain activity. PLoS computational biology. 2015; 11(12):e1004644. https://doi.org/10.1371/journal.pcbi.1004644

Goñi J, van den Heuvel MP, Avena-Koenigsberger A, de Mendizabal NV, Betzel RF, Griffa A, Hagmann P, Corominas-Murtra B, Thiran JP, Sporns O. Resting-brain functional connectivity predicted by analytic measures of network communication. Proceedings of the National Academy of Sciences. 2014; 111(2):833–838. https://doi.org/10.1073/pnas.1315529111

Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, Ghosh SS. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Frontiers in neuroinformatics. 2011; 5:13. https://doi.org/10.3389/fninf.2011.00013

Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003; 3(Mar):1157–1182.

Haken H. Synergetics. Nonequilibrium phase transitions and self-organization in physics, chemistry and biology. 3rd rev. enl. ed. New York: Springer-Verlag. 1983; http://dx.doi.org/10.1007/978-3-642-88338-5

Hansen EC, Battaglia D, Spiegler A, Deco G, Jirsa VK. Functional connectivity dynamics: modeling the switching behavior of the resting state. Neuroimage. 2015; 105:525–535. https://doi.org/10.1016/j.neuroimage.2014.11.001

Henstridge CM, Hyman BT, Spires-Jones TL. Beyond the neuron–cellular interactions early in Alzheimer disease pathogenesis. Nature Reviews Neuroscience. 2019; p. 1. https://doi.org/10.1038/s41583-018-0113-1

Honey CJ, Kötter R, Breakspear M, Sporns O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. Proceedings of the National Academy of Sciences. 2007; 104(24):10240–10245. https://doi.org/10.1073/pnas.0701519104

Iacono D, Markesbery W, Gross M, Pletnikova O, Rudow G, Zandi P, Troncoso JC. The Nun study: clinically silent AD, neuronal hypertrophy, and linguistic skills in early life. Neurology. 2009; 73(9):665–673. https://doi.org/10.1212/WNL.0b013e3181b01077

Iddi S, Li D, Aisen PS, Rafii MS, Thompson WK, Donohue MC, Initiative ADN, et al. Predicting the course of Alzheimer's progression. Brain informatics. 2019; 6(1):6. https://doi.org/10.1186/s40708-019-0099-0

Iyappan A, Kawalia SB, Raschka T, Hofmann-Apitius M, Senger P. NeuroRDF: semantic integration of highly curated data to prioritize biomarker candidates in Alzheimer's disease. Journal of biomedical semantics. 2016; 7(1):45. https://doi.org/10.1186/s13326-016-0079-8

Jaynes ET. Information theory and statistical mechanics. Physical review. 1957; 106(4):620. https://doi.org/10.1103/PhysRev.106.620

Jirsa VK, Proix T, Perdikis D, Woodman MM, Wang H, Gonzalez-Martinez J, Bernard C, Bénar C, Guye M, Chauvel P, et al. The virtual epileptic patient: individualized whole-brain models of epilepsy spread. Neuroimage. 2017; 145:377–388. https://doi.org/10.1016/j.neuroimage.2016.04.049

Kirst C, Timme M, Battaglia D. Dynamic information routing in complex networks. Nature communications. 2016; 7:11061. https://doi.org/10.1038/ncomms11061

Kodamullil AT, Younesi E, Naz M, Bagewadi S, Hofmann-Apitius M. Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis. Alzheimer's & Dementia. 2015; 11(11):1329–1339. https://doi.org/10.1016/j.jalz.2015.02.006

Komarova NL, Thalhauser CJ. High degree of heterogeneity in Alzheimer's disease progression patterns. PLoS computational biology. 2011; 7(11):e1002251. https://doi.org/10.1371/journal.pcbi.1002251

Kötter R, Wanke E. Mapping brains without coordinates. Philosophical Transactions of the Royal Society B: Biological Sciences. 2005; 360(1456):751–766. https://doi.org/10.1098/rstb.2005.1625

Lutkenhoff ES, Rosenberg M, Chiang J, Zhang K, Pickard JD, Owen AM, Monti MM. Optimized brain extraction for pathological brains (optiBET). PLoS One. 2014; 9(12):e115551. https://doi.org/10.1371/journal.pone.0115551

Maaten Lvd, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. 2008; 9(Nov):2579– 2605.

Mehrkanoon S, Breakspear M, Boonstra TW. Low-dimensional dynamics of resting-state cortical activity. Brain topography. 2014; 27(3):338–352. https://doi.org/10.1007/s10548-013-0319-5

Melozzi F, Bergmann E, Harris JA, Kahn I, Jirsa V, Bernard C. Individual structural features constrain the functional connectome. bioRxiv. 2019; p. 613307. https://doi.org/10.1101/613307

Messé A, Rudrauf D, Benali H, Marrelec G. Relating structure and function in the human brain: relative contributions of anatomy, stationary dynamics, and non-stationarities. PLoS computational biology. 2014; 10(3):e1003530. https://doi.org/10.1371/journal.pcbi.1003530

Moore P, Lyons T, Gallacher J, Initiative ADN, et al. Random forest prediction of Alzheimer's disease using pairwise selection from time series data. PLoS One. 2019; 14(2):e0211558. https://doi.org/10.1371/journal.pone.0211558

Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Initiative ADN, et al. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage. 2015; 104:398–412. https://doi.org/10.1016/j.neuroimage.2014.10.002

Mungas D, Beckett L, Harvey D, Tomaszewski Farias S, Reed B, Carmichael O, Olichney J, Miller J, DeCarli C. Heterogeneity of cognitive trajectories in diverse older persons. Psychology and aging. 2010; 25(3):606. https://doi.org/10.1037/a0019502

Patterson C. World Alzheimer Report 2018: the state of the art of dementia research: new frontiers. Alzheimer's Disease International (ADI): London, UK. 2018.

Petkoski S, Jirsa VK. Transmission time delays organize the brain network synchronization. Philosophical Transactions of the Royal Society A. 2019; 377(2153):20180132. https://doi.org/10.1098/rsta.2018.0132

Pillai AS, Jirsa VK. Symmetry breaking in space-time hierarchies shapes brain dynamics and behavior. Neuron. 2017; 94(5):1010–1026. https://doi.org/10.1016/j.neuron.2017.05.013

Proix T, Bartolomei F, Guye M, Jirsa VK. Individual brain structure and modelling predict seizure propagation. Brain. 2017; 140(3):641–654. https://doi.org/10.1093/brain/awx004

Proix T, Spiegler A, Schirner M, Rothmeier S, Ritter P, Jirsa VK. How do parcellation size and short-range connectivity affect dynamics in large-scale brain network models? NeuroImage. 2016; 142:135–149. https://doi.org/10.1016/j.neuroimage.2016.06.016

Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. NeuroImage. 2017; 155:530–548. https://doi.org/10.1016/j.neuroimage.2017.03.057

Rombouts SA, Barkhof F, Goekoop R, Stam CJ, Scheltens P. Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study. Human brain mapping. 2005; 26(4):231–239. https://doi.org/10.1002/hbm.20160

de Rosnay MD, The legal and policy framework for scientific data sharing, mining and reuse; 2017. https://doi.org/10.4000/books.editionsmsh.9082

Saggio ML, Ritter P, Jirsa VK. Analytical operations relate structural and functional connectivity in the brain. PLoS One. 2016; 11(8):e0157292. https://doi.org/10.1371/journal.pone.0157292

Sanchez E, Toro C, Carrasco E, Bonachela P, Parra C, Bueno G, Guijarro F. A knowledge-based clinical decision support system for the diagnosis of Alzheimer disease. In: *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services* IEEE; 2011. p. 351–357. https://doi.org/10.1109/HEALTH.2011.6026778

Sanz-Leon P, Knock SA, Spiegler A, Jirsa VK. Mathematical framework for large-scale brain network modeling in The Virtual Brain. Neuroimage. 2015; 111:385–430. https://doi.org/10.1016/j.neuroimage.2015.01.002

Sanz Leon P, Knock SA, Woodman MM, Domide L, Mersmann J, McIntosh AR, Jirsa V. The Virtual Brain: a simulator of primate brain network dynamics. Frontiers in neuroinformatics. 2013; 7:10. https://doi.org/10.3389/fninf.2013.00010

Schirner M, Rothmeier S, Jirsa VK, McIntosh AR, Ritter P. An automated pipeline for constructing personalized virtual brains from multimodal neuroimaging data. NeuroImage. 2015; 117:343–357. https://doi.org/10.1016/j.neuroimage.2015.03.055

Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano. RUSBoost. A Hybrid Approach to Alleviating Class Imbalance. IEEE Transactions on Systems. Man, And Cybernetics—Part A: Systems And Humans. 2010; 40(1). https://dx.doi.org/10.1109/tsmca.2009.2029559

Shen K, Bezgin G, Hutchison RM, Gati JS, Menon RS, Everling S, McIntosh AR. Information processing architecture of functionally defined clusters in the macaque cortex. Journal of Neuroscience. 2012; 32(48):17465–17476. https://doi.org/10.1523/JNEUROSCI.2709-12.2012

Shen K, Bezgin G, Schirner M, Ritter P, Everling S, McIntosh R. A macaque connectome for large-scale network simulations in TheVirtualBrain. bioRxiv. 2019a; p. 480905. https://doi.org/10.1038/s41597-019-0129-z

Shen K, Goulas A, Grayson DS, Eusebio J, Gati JS, Menon RS, McIntosh AR, Everling S. Exploring the limits of network topology estimation using diffusion-based tractography and tracer studies in the macaque cortex. NeuroImage. 2019b; 191:81–92. https://doi.org/10.1016/j.neuroimage.2019.02.018

Spiegler A, Knösche TR, Schwab K, Haueisen J, Atay FM. Modeling brain resonance phenomena using a neural mass model. PLoS computational biology. 2011; 7(12):e1002298. https://doi.org/10.1371/journal.pcbi.1002298

Stefanescu RA, Jirsa VK. A low dimensional description of globally coupled heterogeneous neural networks of excitatory and inhibitory neurons. PLoS computational biology. 2008; 4(11):e1000219. https://doi.org/10.1371/journal.pcbi.1000219

Stefanovski L, Triebkorn P, Spiegler A, Diaz-Cortes MA, Solodkin A, Jirsa V, McIntosh AR, Ritter P, Initiative ADN, et al. Linking molecular pathways and large-scale computational modeling to assess candidate disease mechanisms and pharmacodynamics in Alzheimer's disease. BioRxiv. 2019; p. 600205. https://doi.org/10.3389/fncom.2019.00054

Straathof M, Sinke MR, Dijkhuizen RM, Otte WM. A systematic review on the quantitative relationship between structural and functional network connectivity strength in mammalian brains. Journal of Cerebral Blood Flow & Metabolism. 2019; 39(2):189–209. https://doi.org/10.1177/0271678X18809547

Taylor L, Nitschke G. Improving deep learning with generic data augmentation. In: *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* IEEE; 2018. p. 1542–1547. https://doi.org/10.1109/SSCI.2018.8628742

Termenon M, Jaillard A, Delon-Martin C, Achard S. Reliability of graph analysis of resting state fMRI using test-retest dataset from the Human Connectome Project. Neuroimage. 2016; 142:172–187. https://doi.org/10.1016/j.neuroimage.2016.05.062

Thorogood A, Mäki-Petäjä-Leinonen A, Brodaty H, Dalpé G, Gastmans C, Gauthier S, Gove D, Harding R, Knoppers BM, Rossor M, et al. Consent recommendations for research and international data sharing involving persons with dementia. Alzheimer's & Dementia. 2018; 14(10):1334–1343. https://doi.org/10.1016/j.jalz.2018.05.011

Toga AW, Neu SC, Bhatt P, Crawford KL, Ashish N. The global Alzheimer's association interactive network. Alzheimer's & Dementia. 2016; 12(1):49–54. https://doi.org/10.1016/j.jalz.2015.06.1896

Valdes-Sosa PA, Roebroeck A, Daunizeau J, Friston K. Effective connectivity: influence, causality and biophysical modeling. Neuroimage. 2011; 58(2):339–361. https://doi.org/10.1016/j.neuroimage.2011.03.058

de Vos F, Koini M, Schouten TM, Seiler S, van der Grond J, Lechner A, Schmidt R, de Rooij M, Rombouts SA. A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease. Neuroimage. 2018; 167:62–72. https://doi.org/10.1016/j.neuroimage.2017.11.025

Wang JY, Abdi H, Bakhadirov K, Diaz-Arrastia R, Devous Sr MD. A comprehensive reliability assessment of quantitative diffusion tensor tractography. Neuroimage. 2012; 60(2):1127–1138. https://doi.org/10.1016/j.neuroimage.2011.12.062

Wang Z, Wang J, Zhang H, Mchugh R, Sun X, Li K, Yang QX. Interhemispheric functional and structural disconnection in Alzheimer's disease: a combined resting-state fMRI and DTI study. PLoS One. 2015; 10(5):e0126310. https://doi.org/10.1371/journal.pone.0126310

Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack Jr CR, Jagust W, Morris JC, et al. The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. Alzheimer's & Dementia. 2017; 13(5):561–571. https://doi.org/10.1016/j.jalz.2016.10.006

Wong KF, Wang XJ. A recurrent network mechanism of time integration in perceptual decisions. Journal of Neuroscience. 2006; 26(4):1314–1328. https://doi.org/10.1523/JNEUROSCI.3733-05.2006

Woodman MM, Pezard L, Domide L, Knock SA, Sanz-Leon P, Mersmann J, McIntosh AR, Jirsa V. Integrating neuroinformatics tools in TheVirtualBrain. Frontiers in neuroinformatics. 2014; 8:36. https://doi.org/10.3389/fninf.2014.00036

Wyman BT, Harvey DJ, Crawford K, Bernstein MA, Carmichael O, Cole PE, Crane PK, DeCarli C, Fox NC, Gunter JL, et al. Standardization of analysis sets for reporting results from ADNI MRI data. Alzheimer's & Dementia. 2013; 9(3):332–337. https://doi.org/10.1016/j.jalz.2012.06.004

Yaeger LS, Lyon RF, Webb BJ. Effective training of a neural network character classifier for word recognition. In: *Advances in neural information processing systems*; 1997. p. 807–816.

Zhang D, Shen D, Initiative ADN, et al. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage. 2012; 59(2):895–907. https://doi.org/10.1016/j.neuroimage.2011.09.069

Zhang Z, Liao W, Chen H, Mantini D, Ding JR, Xu Q, Wang Z, Yuan C, Chen G, Jiao Q, et al. Altered functional–structural coupling of large-scale brain networks in idiopathic generalized epilepsy. Brain. 2011; 134(10):2912–2928. https://doi.org/10.1093/brain/awr223

Zimmermann J, Ritter P, Shen K, Rothmeier S, Schirner M, McIntosh AR. Structural architecture supports functional organization in the human aging brain at a regionwise and network level. Human brain mapping. 2016; 37(7):2645–2661. https://doi.org/10.1002/hbm.23200

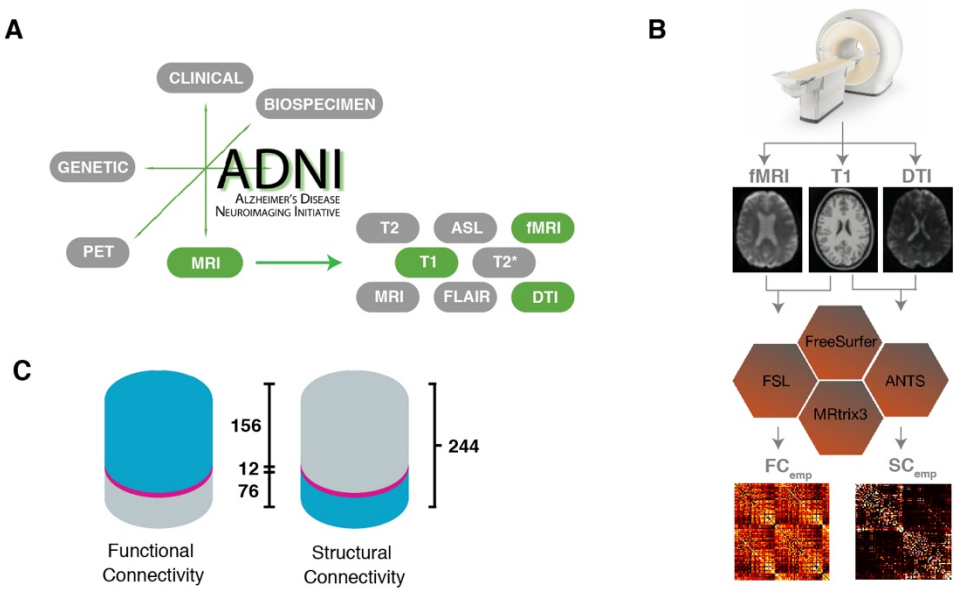

Figure 1. Connectomic information extracted from the ADNI dataset has gaps. A) The different dataset releases by the ADNI consortium include a variety of information relative to different biomarkers and imaging modalities. Here, we focus on structural and functional MRI features and, chiefly: T1, DTI (allowing to extract empirical structural connectomes); and resting-state fMRI BOLD time-series (allowing to extract empirical functional connectomes). B) Matrices SC_{emp} and FC_{emp} summarizing connectomic information about, respectively structural connectivity (SC) and functional connectivity (FC) are obtained via elaborated multi-step processing pipelines, using various software including FreeSurfer, FSL, ANTS, and MRtrix3. C) The total number of ADNI-derived subjects investigated in this study is 244, in which 74 subjects were control, while 119 subjects labeled as MCI, and 51 subjects as AD. Out of these 244, FC_{emp} could be extracted for 168 subjects, and SC_{emp} for 88. However, SC_{emp} and FC_{emp} were both simultaneously available for just a minority of 12 subjects (referred to as the "SC_{emp}+FC_{emp} subset"). The available data is shown in blue and the missing data in grey, the SC_{emp}+FC_{emp} subset is shown in pink.

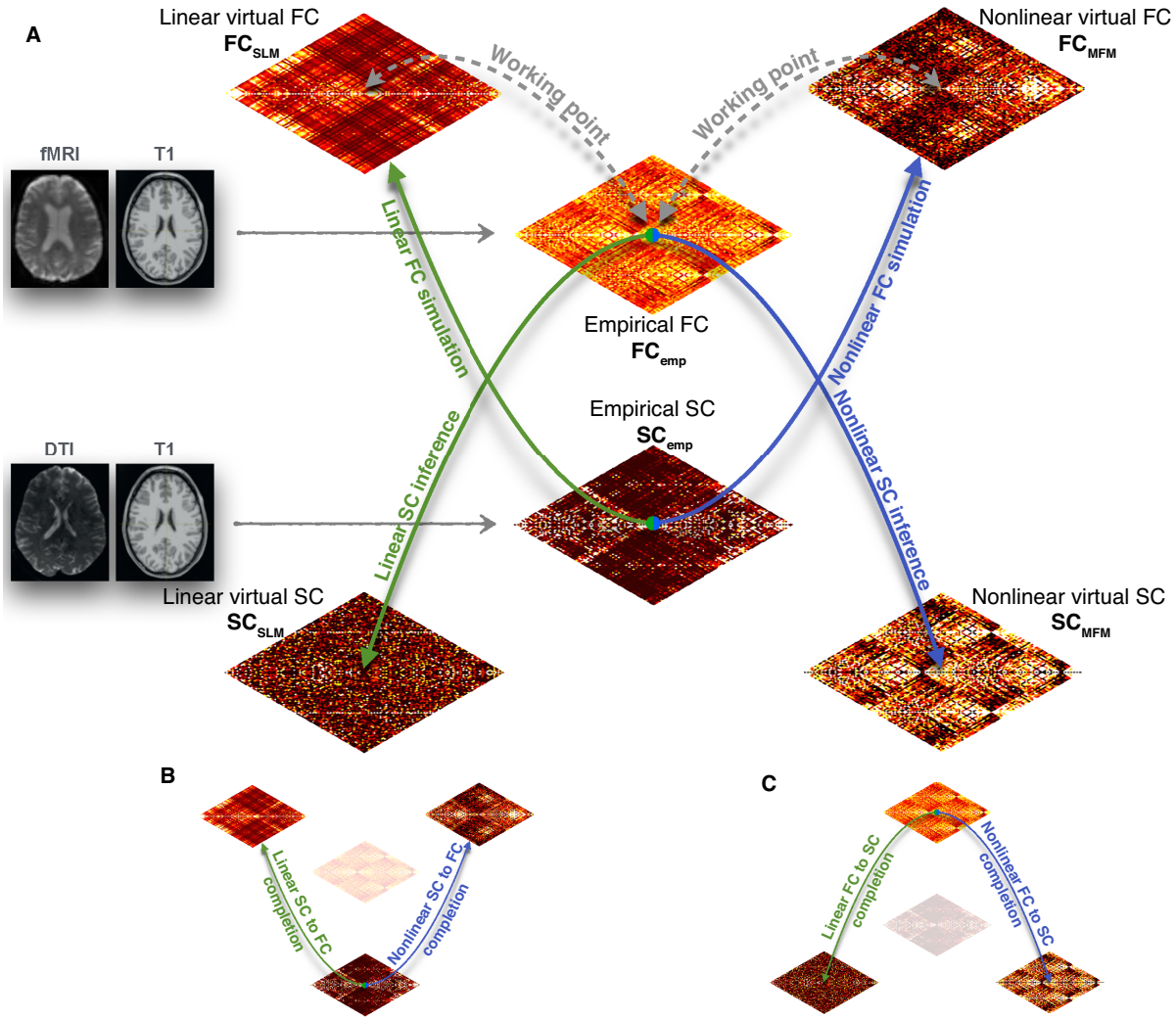

Figure 2. From mean-field modeling to connectomic data completion. A) We present here a graphical summary of the various computational simulation and inference strategies used in this study to bridge between different types of connectivity matrices. Mean-field simulation and the associated analytic theory can be used to generate virtual FC, through simulations of resting-state whole-brain models embedding a given input SC connectome (ascending arrows). Algorithmic procedures, that may still include computational simulation steps, can be used to perform the inverse inference of a virtual SC that is compatible with a given input FC (descending arrows). Both simulation and inference can be performed using simpler linear (green arrows) or non-linear (blue arrows) approaches. When the input SC (or FC) connectomes used as input for FC simulation (or SC inverse inference) correspond to empirical connectomes SC_{emp} (or FC_{emp}), derived from ADNI T1 and DTI (fMRI) images, then model simulation (inversion) can be used to complete gaps in the dataset, whenever FC_{emp} (or SC_{emp}) is missing. We refer then to these operations as: B) SC-to-FC completion; and, C) FC-to-SC completion. Both exist in linear and non-linear versions. D) The Virtual Brain neuroinformatic platform (whose logo is shown here) provides a simulation environment particularly suitable to perform operations of connectomic-dataset completion.

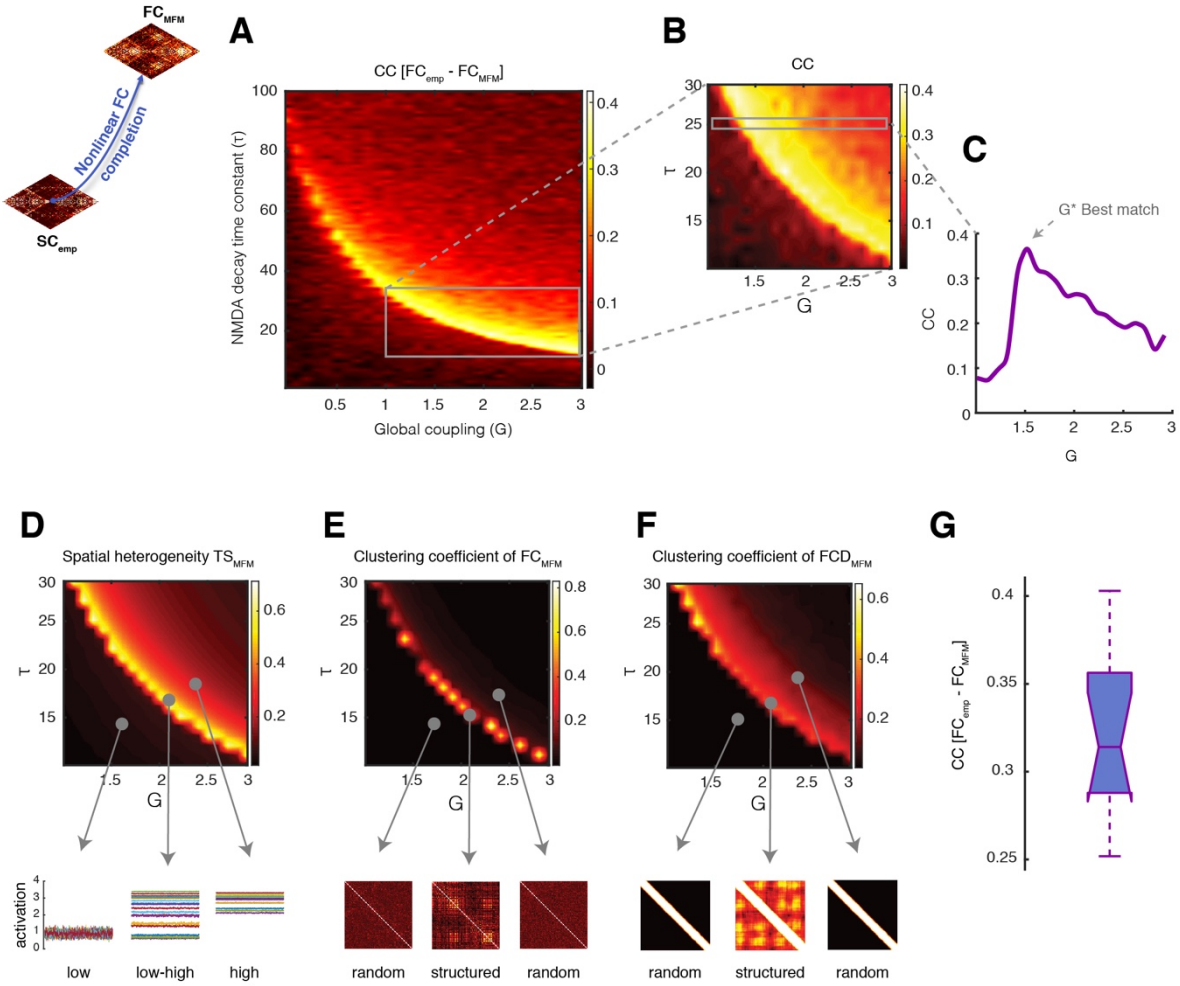

Figure 3. Non-linear SC-to-FC data completion. Simulations of a non-linear model embedding a given input SC_{emp} matrix can be used to generate surrogate FC_{MFM} matrices. A) Systematic exploration (here shown for a representative subject) of the dependency of the correlation between FC_{emp} and FC_{MFM} on the MFM parameters G (inter-regional coupling strength) and τ (synaptic time-constant of within-region excitation) indicates that the best fitting performances are obtained when parameters are concentrated in a narrow concave stripe across the G/τ plane. B) Enlarged zoom of panel A over the range $G \in [1\ 3]$ and $\tau \in [10\ 30]$. C) For a value of $\tau = 25$, representatively chosen here for illustration, we identify a value G^* for which the Pearson correlation between FC_{emp} and FC_{MFM} reaches a clear local maximum. Panels A-C thus indicate that it makes sense speaking of a best-fit zone and that reliable nonlinear SC-to-FC completion should be performed using MFM parameters within this zone. Three criteria help us identifying parameter combinations in this best fitting zone when the actual FC_{emp} is unknown. D) First criterion: we define the spatial coefficient of variation of the time-series of simulated BOLD activity TS_{MFM} as the ratio between the variance and the mean across regions of the time-averaged activation of different regions. The best fit zone is associated with a peaking of this spatial coefficient of variation, associated with a maximally heterogeneous mix or low and high activation levels for different regions (see time-series cartoons for three working points below the CV surface). E) Second criterion: in the best fitting zone, the resulting FC_{MFM} is neither randomly organized nor excessively regular (synchronized) but presents a complex clustering structure (see lower cartoons), which can be tracked by a peak in the clustering coefficient of the FC_{MFM}, seen as weighed adjacency matrix. F) Third criterion: in the best fitting zone, resting-state FC_{MFM} display a relatively richer dynamics than in other sectors of the parameter space. This gives rise to an "FCD matrix" (correlation between time-resolved FC observed at different times) which is neither random nor too regular but displays a certain degree of clustering (see lower cartoons). The emergence of complex dynamics of FC can be tracked by an increase in the clustering coefficient of the FCD matrix extracted from simulated resting-state dynamics. G) The boxplot shows the distribution of correlations between the actual FC_{emp} and FC_{MFM} estimated within the best fitting zone for all subjects from the "SC_{emp} + FC_{emp}" subset.

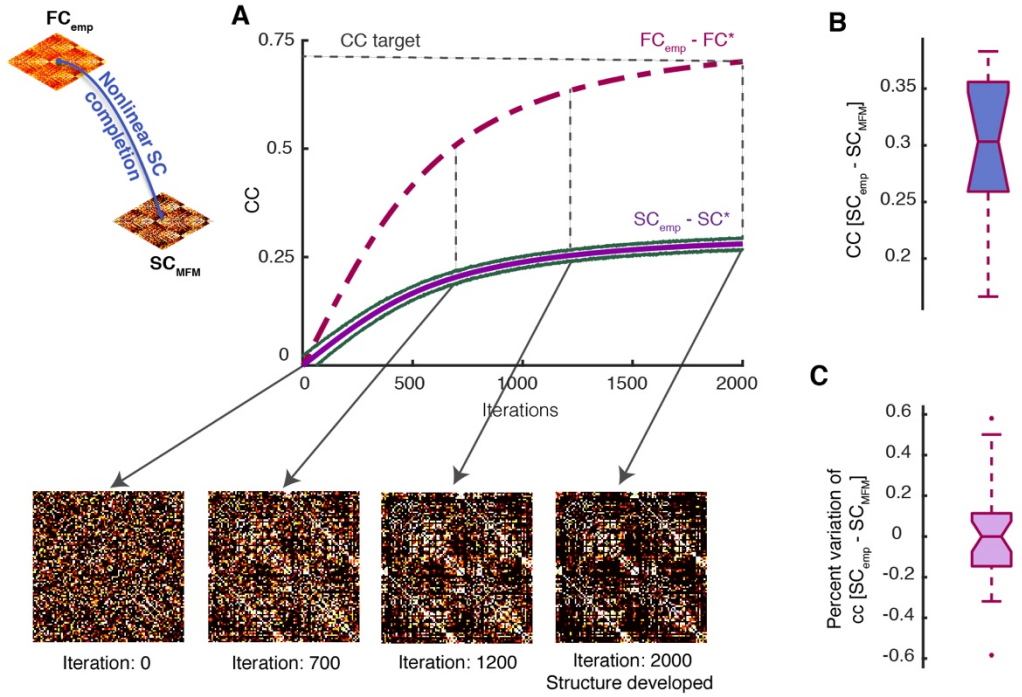

Figure 4. Non-linear FC-to-SC data completion. An iterative procedure can be used to perform resting-state simulations of an MFM model starting from a randomly guessed structural connectome SC* and progressively modify this SC* to make it compatible with a known target FC_{emp}. A) Starting from an initial random $SC^*_{(0)}$ matrix, there is no correlation between the target FC_{emp} and the generated $FC^*_{(0)}$ matrix. However, by adjusting the weights of the used SC* through the algorithm of Table 4, SC* gradually develops a richer organization, leading to an increase of the correlation between FC* and FC_{emp} (violet dashed line) and, in parallel, of the correlation between SC* and SC_{emp} (violet solid line), as shown here for a representative subject within the "SC_{emp}+FC_{emp}" subset. The algorithm stops when the correlation between FC* and the input target FC_{emp} reaches a desired quality threshold (here 0.7 after 2000 iterations) and the SC* at the last iteration is used as virtual surrogate SC_{MFM}. B) The boxplot shows the distribution of correlation between SC_{emp} and SC_{MFM} for all subjects in the "SC_{emp} + FC_{emp}" subset. C) The correlation between SC_{emp} and SC_{MFM} can vary using different random initial connectomes $SC^*_{(0)}$. Here we show a boxplot of the percent dispersions of the correlation values obtained for different initial conditions around the median correlation value. The fact that these dispersions lie within a narrow interval of ±2.5% indicates that the expected performance is robust against changes of the initial conditions.

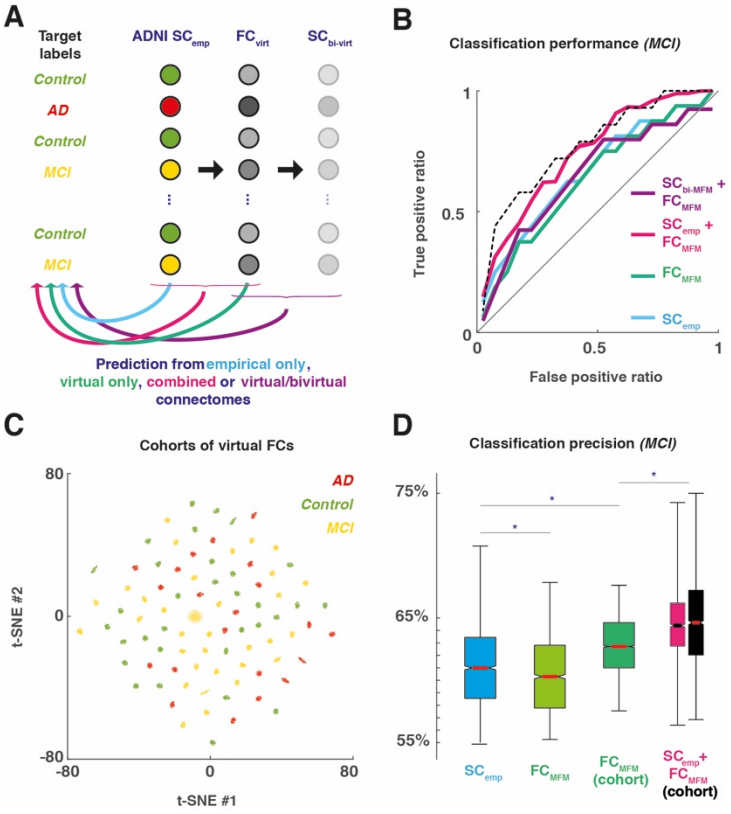

Figure 5. Machine-learning classification of MCI patients based on empirical and virtual connectomes. A) We quantified lower bounds to the best performance that machine-learning classifiers can reach in classifying subjects as being affected by early-stage dementia or not ("MCI vs all", i.e. neither labeled "AD" nor "Control", but "MCI"), based purely on connectome matrices of different types or type combinations: actual empirical connectomes (here, SC_{emp}, light blue color); their virtual counterparts (here, FC_{MFM}, non-linearly completed from SC_{emp}, green color); a combination of empirical and virtually-completed connectomes of different types (here, SC_{emp} combined with FC_{MFM}, magenta color); and, a combination of virtual and bi-virtual connectomes of different types (here $SC_{bi\text{-}MFM}$ combined with FC_{MFM}, violet color). B) We trained random forest classifiers based on these different combinations of input features, isolated a subset of the top-10% best random-ensemble classifiers we could train and assessed their average MCI detection performance (tendential classification performance, i.e. a lower bound to the best performance we could achieve). We show here median Receiver Operator Curves (ROCs) for MCI detection for different combinations of empirical and input features (generalization performance, via 5-fold cross-validation). High-performance classifiers trained on virtual and bi-virtual features can reach areas under curves indistinguishable from high-performance classifiers trained on empirical features. Non-linear data completion (here, complementing SC_{emp} combined with FC_{MFM}) leads to superior performance, especially when combined with data augmentation (dashed ROC curve). C) We generated virtual cohorts of surrogate FC data for data augmentation purpose (i.e. facilitating training, by increasing training dataset size with appropriate surrogates). We generated 100 different FC_{MFM} matrices for each of the 88 subjects with an available SC_{emp}. Shown here is a low-dimensional t-SNE projection of the resulting 8800 virtual FC_{MFM} 's. D) Classifier trained on augmented datasets (i.e. datasets where the training and validation datasets have an artificially enlarged size, by the use of the many alternative variants of virtual FC_{MFM} available for each of the SC_{emp} or $SC_{bi\text{-}MFM}$), achieve superior performance than classifiers trained without FM-based data augmentation. Shown here are boxplots of the precision (at 10% recall) achieved by high-performance classifiers, trained with different input datasets. Median precision for classifiers trained on FC_{MFM} matrices (light green) is smaller than when training on SC_{emp} matrices (light blue), however, the use of data augmentation restores and even improve precision (dark green). Combining SC_{emp} with FC_{MFM} matrices yields a further boost in performance (magenta), but, in this case, data augmentation is not helpful (black). A star denotes significantly different median precisions (Kruskal-Wallis, $p < 0.05$, Bonferroni corrected).

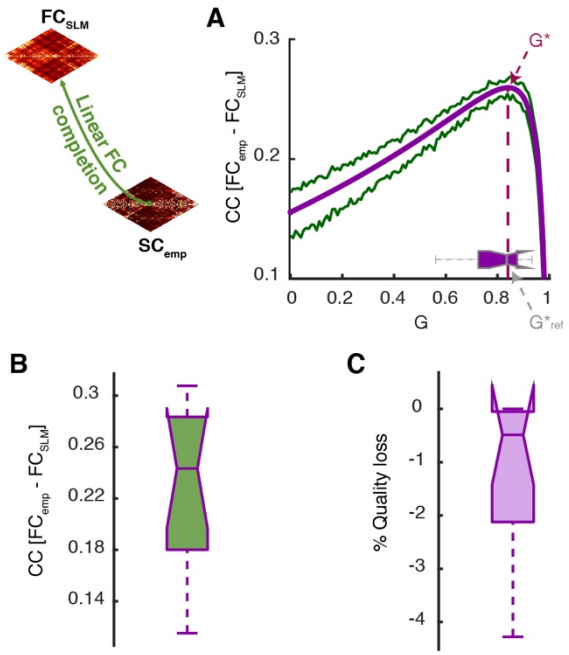

Figure S1. Linear SC-to-FC data completion. The functional data completion can also be done using the linear model starting from SC_{emp} matrices. A) the systematic exploration (for a representative subject) of the dependency of correlation between FC_{emp} and FC_{SLM} on the SLM parameter G (global scale of long-range connectivity strength) shown by the violet line indicates that the best fitting value *G** (dashed line) can be obtained slightly before the critical point of the system $G_{critic} = 1/max(\lambda_i)$ which since the SC_{emp} matrices are normalized to one $1/max(\lambda_i) = 1$ and $G_{critic} = 1$. The green lines display 5 and 95 percentiles of bootstrap resampling. The inset boxplot gives the distribution of *G** over all the subjects in the "SC_{emp} + FC_{emp}" subset; for the SLM SC-to-FC completion, we used a common value $G^*_{ref} = 0.83$, equal to the median of the boxplot. B) The boxplot reports the distribution of Pearson correlation between FC_{emp} and FC_{SLM} for all subjects from the "SC_{emp} + FC_{emp}" subset with a median equal to 0.243. C) In case of using the common value G^*_{ref} for all subjects instead of the actual personalized optimum *G** for each subject in the "SC_{emp} + FC_{emp}" subset, the value of quality loss for each subject is shown in the boxplot with median equal to 0.5%.

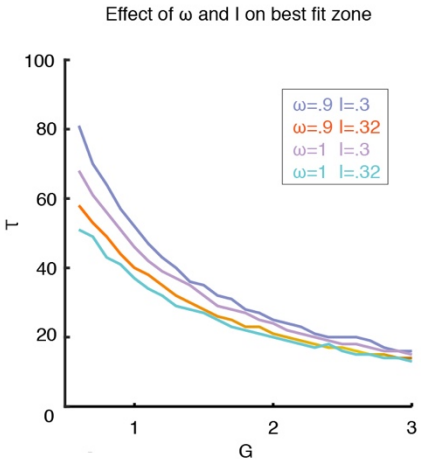

Figure S2. The dependency of best MFM fit zone on additional regional dynamics parameters. In the non-linear data completion, the global parameters of the MFM model are *G* (inter-regional coupling strength), τ (synaptic time-constant of within-region excitation), ω (relative strength of recurrent within-region connections) and *I* (external input) which parameters *G* and τ were investigated in this paper (see Figure 3). Here we showed for different values of ω and *I*, the narrow concave stripe of Figure 3.A as the representative of the best fitting zone is slightly shifted in the G/τ plane, suggesting G and τ are more sensitive parameters and need to be explored rather than ω and *I.*

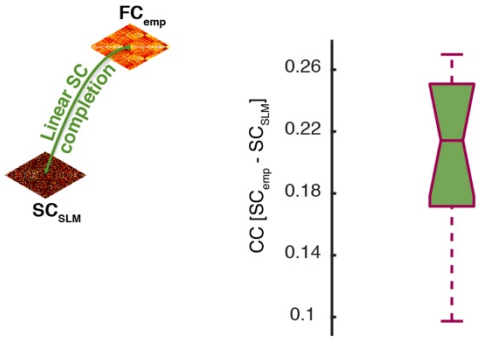

Figure S3. Linear FC-to-SC data completion. Using the linear model, it is equivalently possible to infer the structural SC_{SLM} matrices from FC_{emp}. Since in this approach the free parameters of SLM model appear as scaling factor, they don't affect the correlation of the inferred SC_{SLM} with the SC_{emp} so there is no need for parameter exploration here. The distribution of the correlation values for all the subjects from the "SC_{emp} + FC_{emp}" subset is shown in the boxplot with median equal to 0.21.

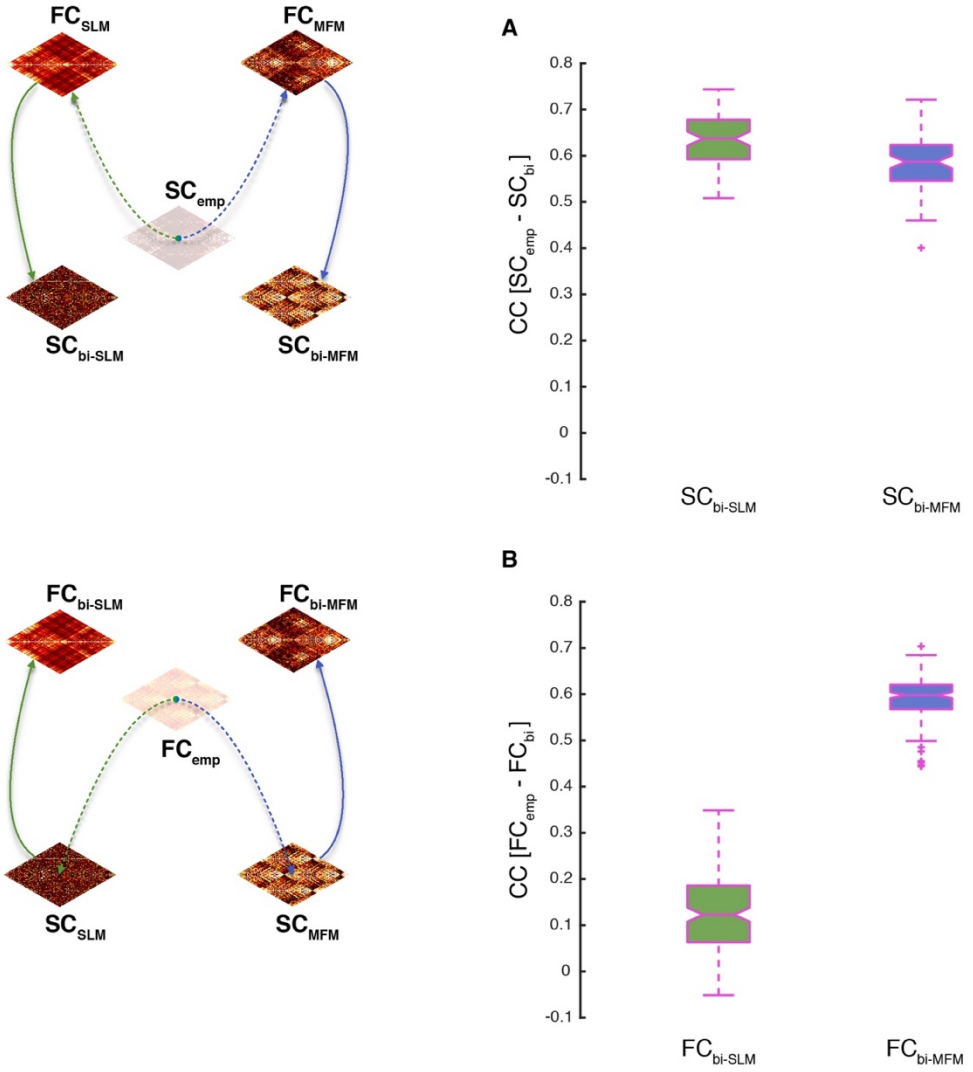

Figure S4. Bi-virtual connectomes. This figure shows the correspondence between empirical and bi-virtual SC and FC pairs, both when using chained linear (SLM-based) and nonlinear (MFM-based) completion procedures. A) For 88 subjects from the ADNI-subset with only SC_{emp} available, considering the linear bi-virtual completion chain SC_{emp} to FC_{SLM} to $SC_{bi\text{-}SLM}$, we obtained a median correlation between SC_{emp} and $SC_{bi\text{-}SLM}$ equal to 0.636 (green boxplot); simultaneously, considering the non-linear bi-virtual completion chain SC_{emp} to FC_{MFM} to $SC_{bi\text{-}MFM}$, we obtained a median correlation between SC_{emp} and $SC_{bi\text{-}MFM}$ equal to 0.583 (blue boxplot). B) For 168 subjects from the ADNI-subset with only FC_{emp} available, considering the linear bi-virtual completion chain FC_{emp} to SC_{SLM} to $FC_{bi\text{-}SLM}$, we obtained a median correlation between FC_{emp} and $FC_{bi\text{-}SLM}$ equal to 0.122 (green boxplot); simultaneously, considering the non-linear bi-virtual completion chain FC_{emp} to SC_{MFM} to $FC_{bi\text{-}MFM}$, we obtained a median correlation between FC_{emp} and $FC_{bi\text{-}MFM}$ equal to 0.597 (blue boxplot).