*Advanced Science*

ADVANCED SCIENCE Open Access

www.advancedscience.com

RESEARCH ARTICLE OPEN ACCESS

# Linearizing and Forecasting: A Reservoir Computing Route to Digital Twins of the Brain

Gabriele Di Antonio[1,2] | Tommaso Gili[3] | Andrea Gabrielli[1,4] | Maurizio Mattia[2]

[1]"Enrico Fermi" Research Center - CREF, Rome, Italy | [2]Natl. Center for Radiation Protection and Computational Physics, Istituto Superiore di Sanità, Rome, Italy | [3]Networks Unit, IMT Scuola Alti Studi Lucca, Lucca, Italy | [4]Dip. di Ingegneria Civile, Informatica e delle Tecnologie Aeronautiche, Università degli Studi "Roma Tre", Rome, Italy

**Correspondence:** Maurizio Mattia (maurizio.mattia@iss.it)

**Received:** 3 September 2025 | **Revised:** 24 February 2026 | **Accepted:** 26 February 2026

**Keywords:** data-driven digital-twins | koopman operator | recurrent neural networks | reservoir computing | resting state fmri

## ABSTRACT

Exploring the dynamics of complex systems such as the human brain is challenging due to inherent uncertainties and the limited availability of high-quality data. Here, we develop a mathematical theory for noisy linear recurrent neural networks (lRNNs) within the reservoir computing framework and demonstrate their effectiveness in constructing autonomous in silico replicas – digital-twins – of brain activity. We show that the Laplace-transform poles of high-dimensional inferred lRNNs directly encode the spectral properties of observed systems and are linked to the kernels of auto-regressive models. Notably, our approach enables accurate recovery of the system's linear spectrum even when observations undergo conventional preprocessing, including band-pass filtering pipelines commonly used in neural recordings and resting-state fMRI. In these regimes, established techniques such as dynamic mode decomposition often produce spurious spectral estimates. Applying our framework to resting-state fMRI, we successfully predict and decompose BOLD activity into spatiotemporal modes in a low-dimensional latent state space confined around a single equilibrium point. The inferred lRNNs provide interpretable signatures that differentiate subjects and brain areas, supporting biologically meaningful clustering. This flexible digital-twin framework opens the door to virtual experiments and computationally efficient real-time adaptive learning, offering a promising avenue for personalized medicine and intervention strategies.

## 1 | Introduction

The study of complex dynamical systems has a central role in contemporary scientific research. Over recent decades, a plethora of data-driven methodologies has emerged, denoting a field in rapid evolution that is capable to addressing systems with increasing complexity. Among these methodologies, machine learning techniques have become increasingly popular due to their effectiveness in modeling complex datasets and delivering accurate predictions [1–3].

Within the landscape of machine learning methods, reservoir computing (RC) has emerged as a powerful framework for processing temporal data with remarkable computational efficiency. RC leverages high-dimensional recurrent neural networks (RNNs) with fixed internal couplings to transform input time series into rich, high-dimensional representations [4, 5]. This transformation enables the prediction of future samples by simply reading out the current state of the network, effectively capturing the system's dynamics [6–8]. The elegance of this approach lies in conceptualizing the readout as a projection onto a manifold learned through linear regression, echoing the efficient coding strategies observed in biological neural networks [9–11]. Consequently, RNNs have become invaluable tools in neuroscience, aiding in hypothesis generation and providing analytical frameworks for understanding neural computations [12–15].

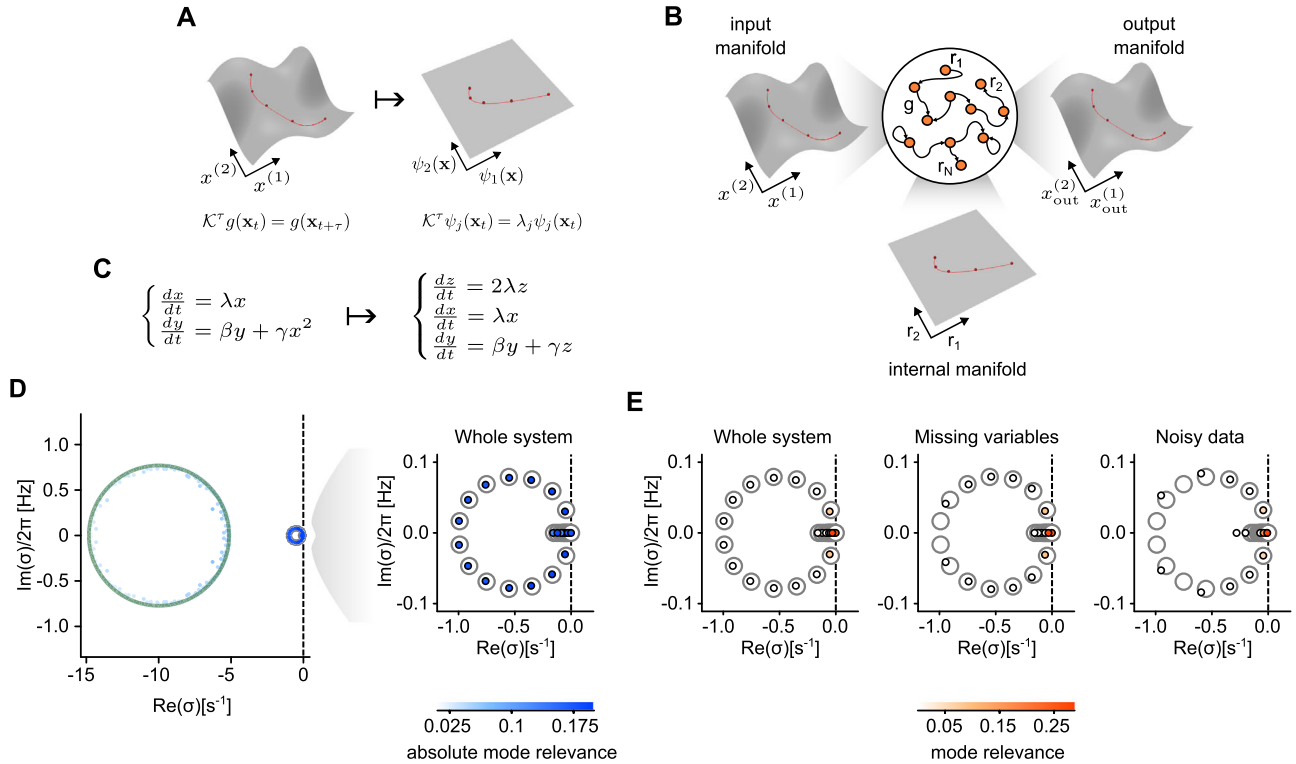


**FIGURE 1** | lRNN encodes a linear representation of the input system. (A) Nonlinear dynamics can be projected onto linear manifolds. Within the infinite functional space of the system observables, the eigenfunctions of the Koopman operator establish a basis for decomposing nonlinear dynamics into simpler linear processes. (B) Reservoir computing (RC) with recurrent neural networks (RNNs). After 'learning', the readout weights $\boldsymbol{W}^{\text{out}}$ are set to reproduce the system observables provided as input. The resulting network is composed of linear units (lRNN) and its post-learning autonomous dynamics encapsulates a linear representation of the original system. C) Example of a 2-D nonlinear system that can be decomposed into a set of three linearly evolving variables. (D) Spectrum of a lRNN with $N = 500$ units trained to mimic a 15-D version of the system in (C) composed of five linear $x$ variables and ten nonlinear $y$ variables. Green circles, spectrum of the lRNN before training. Gray circles, theoretically-derived eigenvalues of the 20-D linearized system to be replicated (see text for details). Small circles, Laplace transform poles from the post-learning autonomous lRNN replicating the example system. Filling colors code the absolute mode relevance $\langle|\Xi_{jk}|\rangle_j$. Right panel, zoom in of the left panel showing the 20 most significant modes. (E) Same spectrum as in (D) with filling colors coding the mode relevance $\langle|\Xi_{jk}v_k(0)|\rangle_j$ (left panel, see text for details). Central panel, spectrum of a lRNN trained on incomplete information (i.e., omitting all $x$ and 5 $y$ variables). Right panel, spectrum of a lRNN trained on the same observables but perturbed by noise.

However, as machine learning has become increasingly popular, some clarity of understanding may have been lost in the pursuit of better predictions. While the inferred models are often quite accurate, they can sometimes lack interpretability [16–18]. A valuable framework to address this limitation is provided by Koopman theory, which offers clear insights into the system dynamics by mapping nonlinear systems into simple yet higher-dimensional dynamics [19–22]. This approach involves the development of data-driven methods aimed at computing finite-dimensional linear representations of nonlinear systems within a suitable functional space. Once an equivalent linear system is inferred, interpretability becomes straightforward, allowing to gain understanding of the time scales and dynamical modes of the observed systems.

The capability to predict future states of a dynamical system based on past observations allows, in principle, to build a replica of the same system. Indeed, predicted observations can be fed back as input resulting in a generative model. Both Koopman-based methods (Figure 1A) and RC approaches (Figure 1B) have the potential to generate these digital copies by producing time series statistically equivalent to those generated by the original systems. The inferred generative models can then serve as 'digital twins' – a concept that has emerged as a transformative paradigm in the study of dynamical and natural systems [23, 24]. Such digital twins can be effective tools for analyzing and simulating their physical counterparts. Successful examples include digital copies of brain activity designed to assist neurosurgeons in dealing with drug-resistant epileptic patients [25–27]. Furthermore, these tools can provide a complementary workbench for designing stimulation approaches to be tested in silico, ultimately leading to control systems for neurorehabilitation [28].

In this study, we investigate the potential of RNNs with linear units to implement effective digital twins of single-subject brain dynamics, measured as the vascular response to neuronal activity (i.e., the BOLD signal) extracted from functional magnetic resonance imaging (fMRI). We demonstrate the autoregressive nature of this approach and its effectiveness in representing whole-brain activity in a low-dimensional latent state space. Such a reduced representational complexity mitigates the risk of overfitting often associated with one-to-one fitted physical models, especially in

realistic scenarios where only a limited number of observations are available and only a partial view of the system can be experimentally accessed. The integration of endogenous memoryless fluctuations is another essential component that enables the digital twin to self-sustain a noisy dynamics statistically equivalent to the one experimentally measured and, crucially, to recover the correct spectral content from filtered observations.

## 2 | Results

### 2.1 | Linear RNNs can Copy Nonlinear Systems

We start investigating the reservoir computing abilities of a RNN with $N$ linear units (lRNN). The state of the lRNN is the vector $\boldsymbol{r}(t)$ whose elements are the activities $r_i(t)$ of each unit, following the linear dynamics

$$\tau\dot{\boldsymbol{r}}(t) = \boldsymbol{W}\boldsymbol{r}(t) + \boldsymbol{W}^{\text{in}}\boldsymbol{\omega}(t) - \boldsymbol{r}(t) \tag{1}$$

The synaptic matrix $\boldsymbol{W} \in \mathbb{R}^{N\times N}$ determines the recurrent input $\boldsymbol{W}\boldsymbol{r}$, which, together with the contributions from the external environment $\boldsymbol{W}^{\text{in}}\boldsymbol{\omega}(t)$, composes the total input $\boldsymbol{h}$ that each unit receives. In general, these currents are nonlinearly transformed by an input–output gain function $\Phi(\boldsymbol{h})$ we assume here to be $\Phi(\boldsymbol{h}) = \boldsymbol{h}$ – a condition usually associated with unit activities perturbatively fluctuating around the quiescent state (see Methods Section). Network units receive a source of external input due to the $M$ observables $\omega_i(t)$ of the systems under investigation, mediated by the synaptic matrix $\boldsymbol{W}^{\text{in}} \in \mathbb{R}^{N\times M}$. From Equation (1) the network state at any time $t$ results then to be

$$\boldsymbol{r}(t) = \int_{-\infty}^{t} e^{(\boldsymbol{W}-\boldsymbol{I})(t-t')/\tau}\boldsymbol{W}^{\text{in}}\boldsymbol{\omega}(t')dt' \tag{2}$$

where $\boldsymbol{I} \in \mathbb{R}^{N\times N}$ is the identity matrix.

If the lRNN is capable to display the so-called 'generalized synchronization' property in response to the input $\boldsymbol{W}^{\text{in}}\boldsymbol{\omega}$ [29, 30], $\boldsymbol{r}(t)$ effectively embeds the dynamics of the 'whole' observed system into its state space. This because the lRNN incorporates the 'echo' [4, 31, 32] of past observations implementing a dimensional embedding *a la* Takens [33–35]. In this RC framework, for sufficiently high $N$ a simple linear transformation capable to approximately map the network state to a target output exists

$$\boldsymbol{\omega}(t) \approx \boldsymbol{W}^{\text{out}}\boldsymbol{r}(t) = \int_0^{\infty} \boldsymbol{G}(s)\boldsymbol{\omega}(t-s)ds \tag{3}$$

with synaptic weights $\boldsymbol{W}^{\text{out}} \in \mathbb{R}^{M\times N}$ computed via a simple linear regression [4, 5]. The linear kernels $\boldsymbol{G}(s) = \boldsymbol{W}^{\text{out}}E^{(\boldsymbol{W}-\boldsymbol{I})(s)/\tau}\boldsymbol{W}^{\text{in}} \in \mathbb{R}^{M\times M}$ (i.e., the Green functions) resulting from Equation (2) reveals the autoregressive nature of the RC approach in this lRNN case.

This concept can be better clarified by resorting to the eigendecomposition of the synaptic matrix $\boldsymbol{W} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}$. Here, $\boldsymbol{\Lambda}$ is the diagonal matrix of the eigenvalues ($\Lambda_{ii} = \lambda_i$) and the eigenvectors are the columns of $\boldsymbol{Q}$: $\boldsymbol{W}\,|Q_i\rangle = \lambda_i\,|Q_i\rangle$. By applying this decomposition to the kernel expression, a superposition of $N$ exponential modes results:

$$\boldsymbol{G}(s) = \sum_{n=1}^{N} \boldsymbol{J}_n e^{(\lambda_n-1)s/\tau} \tag{4}$$

where $\boldsymbol{J}_n = \boldsymbol{W}^{\text{out}}\,|Q_n\rangle\,\langle Q_n^{-1}|\,\boldsymbol{W}^{\text{in}}$ are $N$ rank-1 matrices with $\langle Q_n^{-1}|$ being the $n$-th row of the inverse matrix $\boldsymbol{Q}^{-1}$ (see Methods Section). The convolution in Equation (3) is integrable if the spectral radius $\rho = \max_{n\in[1,N]} \text{Re}\,\lambda_n$ of $\boldsymbol{W}$ is smaller than 1. Under this condition, unperturbed lRNNs ($\boldsymbol{\omega} = \boldsymbol{0}$) have a stable equilibrium point at $\boldsymbol{r} = \boldsymbol{0}$, and self-consistency equation Equation (3) describes a continuous-time autoregressive (AR) model with order $N$ (number of exponential modes in $\boldsymbol{G}(s)$). This relationship between RC with lRNNs and AR models generalizes previous results obtained in the discrete-time domain [36, 37]. As we will see later, the order of the equivalent AR model can be much smaller than $N$ as many $\boldsymbol{J}_n$ are usually close to $\boldsymbol{0}$.

By replacing the observables $\boldsymbol{\omega}(t)$ provided as input with the reconstructed once from Equation (3), the lRNN results to follow the autonomous dynamics

$$\tau\dot{\boldsymbol{r}} = \tilde{\boldsymbol{W}}\boldsymbol{r} - \boldsymbol{r} \tag{5}$$

where $\tilde{\boldsymbol{W}} = \boldsymbol{W} + \boldsymbol{W}^{\text{in}}\boldsymbol{W}^{\text{out}}$ is the updated (i.e., 'learned') synaptic matrix. If the network can correctly predict the future of the target output, Equation (5) provides a linear representation of the system under analysis. Relying as above on the eigenmode decomposition $\tilde{\boldsymbol{W}} = \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{Q}}^{-1}$, the projections $\boldsymbol{v}(t) \equiv \tilde{\boldsymbol{Q}}^{-1}\boldsymbol{r}(t)$ evolve in time as a set of decoupled variables:

$$v_k(t) = v_k(0)e^{(\tilde{\lambda}_k-1)t/\tau} \quad \forall k \in [1, N] \tag{6}$$

This last step makes it apparent the relationship with the Koopman-operator theory, as a suited linear combination of these modes eventually implements an effective decomposition of the time series $\boldsymbol{\omega}(t)$:

$$\boldsymbol{\omega}(t) = \boldsymbol{W}^{\text{out}}\boldsymbol{r}(t) \equiv \boldsymbol{\Xi}\boldsymbol{v}(t) \tag{7}$$

with $\boldsymbol{\Xi} = \boldsymbol{W}^{\text{out}}\tilde{\boldsymbol{Q}}$. The Laplace transform of the reconstructed observable will then have $N$ poles $\sigma_k = (\tilde{\lambda}_k - 1)/\tau$ with their related residues, allowing us to represent the dynamical properties of the 'digital twin' (i.e., the autonomous lRNN), and linking it to AR models and finite approximations of the Koopman operator (see Methods Section). Indeed, the $k$-th residue is proportional to $\Xi_{jk}v_k(0)$ and provides in absolute value the relevance of the $k$-th mode in describing the observable $\omega_j$, given the initial network state $\boldsymbol{r}(0)$.

As a testing ground for the theoretical framework derived above we consider the following nonlinear system:

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{C}_x\boldsymbol{x} \\ \dot{\boldsymbol{y}} = \boldsymbol{C}_y\boldsymbol{y} + \boldsymbol{C}_{yx}\boldsymbol{x} \odot \boldsymbol{x} \end{cases} \tag{8}$$

Choosing $\boldsymbol{C}_x$ as a diagonal matrix, this system can be mapped onto a linear one by introducing a set of new observables $\boldsymbol{z} = \boldsymbol{x} \odot \boldsymbol{x} = \textbf{diag}(\boldsymbol{x})\,\boldsymbol{x}$ (Figure 1C). In this case, the related infinitesimal

generator of the Koopman operators is equivalent to the finite matrix $\boldsymbol{K}$ with diagonal blocks $\boldsymbol{C}_x$, $2\boldsymbol{C}_x$, $\boldsymbol{C}_y$.

We trained a lRNN to replicate the relaxation dynamics of this system from the observables $\boldsymbol{\omega} = [\boldsymbol{x}; \boldsymbol{y}]$ with $\boldsymbol{x} \in \mathbb{R}^5$ and $\boldsymbol{y} \in \mathbb{R}^{10}$, starting from a random initial condition $\boldsymbol{\omega}(0)$. As initial synaptic matrix $\boldsymbol{W}$ we chosen the one associated to ring-like connectivity (i.e., $W_{i,i+1} = W_{N,1} = \rho$ are the only non-zero elements) known to have optimal performances in the case of linear units [38]. In Figure 1D, the poles $\sigma_k$ of the reconstructed observables from Equation (7) are shown together with the related mode relevance (color intensity). Despite the nonlinearity of the system to replicate, the overlap between the poles $\sigma_k$ and the theoretical spectrum of $\boldsymbol{K}$ is remarkable. Furthermore, only 20 out of the $N = 500$ modes have $\langle |\Xi_{jk}| \rangle_j$ significantly different from zero (blue-colored circles).

Decomposition modes of the trained and autonomous lRNN exhibit damped oscillations following Equation (6) for each pole $\sigma_k \in \mathbb{C}$. Each mode has an amplitude envelope $e^{\mathrm{Re}(\sigma_k)t}$ and oscillates at angular frequency $\mathrm{Im}(\sigma_k)$. Consequently, most modes decay rapidly and have little influence on the dynamics, whereas only a few slow modes, i.e. those with the largest (least negative) real parts, contribute appreciably (Figure 1E, red-colored small circles).

Such low-dimensionality is not altered even if during learning the lRNN receives noisy-contaminated observables or some of them are not provided as input. Indeed, the autonomous lRNN continues to predict accurately the system's future also under these conditions, although the less relevant poles move away from the imaginary axis (missing match between small and large circles).

## 2.2 | Stochastic lRNNs Successfully Forecast fMRI Time Series

To explore the applicability of the RC with lRNN, we examined time series data from a real-world physical system. We considered the brain activity of 20 healthy subjects at rest (i.e., in a state of quiet wakefulness), measured through fMRI. The experimental time series was derived from the BOLD signals recorded from a subset of voxels, which serve as proxies for neuronal activity across 11 cortical areas in the language network (Figure 2A, see Methods Section).

The BOLD time series were preprocessed using principal component analysis (PCA) to reduce redundancy and achieve dimensionality reduction, which was found to be relatively high (see Figure 2B, C). Specifically, we assessed the correlation between the information lost during dimensionality reduction and the actual signals. Our analysis revealed that the first principal components (PCs), which explain 99.5% of the variance (Figure 2D,E), excluded a residual activity that was indistinguishable from white noise. Based on this criterion, we determined that, on average, only four PCs per cortical area should be considered. Consequently, the average dimensionality of the dataset was $M \approx 44$ PCs per subject (refer to Methods Section for details). This relatively high-dimensional time series served as the input for the lRNN. The readout synaptic matrix $\boldsymbol{W}^{\mathrm{out}}$ was computed over the first 280 s (Figure 2F).

The decision to utilize a transformation of brain activity that results in information loss may appear counterproductive. However, projecting onto the first PCs was expected to be less sensitivity to the uncertainties inherent in the measured BOLD signals. By employing these first PCs as inputs, we provided the lRNN with denoized representations of brain activity. Moreover, thanks to the echo state property [4, 31, 32], the missing information regarding the state of the observed system can be effectively recovered through the dimensional embedding that the lRNN performs.

As previously described, with the learned $\boldsymbol{W}^{\mathrm{out}}$ – which varies from subject to subject – we established a closed loop between the input and output of the lRNN, effectively creating an autonomous digital twin of the brain network under investigation. This approach yielded a remarkable overlap between the reconstructed and measured time series (see Figure 2G for an example subject). At the population level, we evaluated the forecasting performance by calculating the correlations between replicated and experimental BOLD signals, as well as the root mean square errors (RMSEs) for the final unseen 37.5 s of the time series (Figure 2H). From this perspective, the inferred digital twin demonstrated high-quality reconstruction of brain activity, further validating its effectiveness.

The lRNN dynamics described thus far is dissipative, meaning that the state of the network ultimately relaxes to $\mathbf{0}$. However, brain activity associated to the BOLD signals fluctuates without rest. To reconcile this discrepancy in our digital twin, we incorporated an endogenous noise into each network unit. This modification introduces a generative mechanism that sustains activity over time. More specifically, the autonomous dynamics of the inferred (i.e., learned) lRNN dynamics is a multi-variate Ornstein-Uhlenbeck process

$$\tau d\boldsymbol{r} = (\tilde{\boldsymbol{W}} - \boldsymbol{I})\boldsymbol{r}dt + d\boldsymbol{\eta} \tag{9}$$

where $\boldsymbol{\eta}(t)$ is the the vector of Gaussian white noise each unit of the network independently receives ($\langle d\eta_j(t) d\eta_k(t') \rangle = \gamma^2 \delta_{jk} \delta(t - t')$) with noise intensity $\gamma$. Note that, since the system is linear, this formulation does not change the theoretical framework derived above; in this case, it applies to the expectation values $\mathbb{E}[\boldsymbol{r}]$ and $\mathbb{E}[\boldsymbol{\omega}] = \boldsymbol{W}^{\mathrm{out}}\mathbb{E}[\boldsymbol{r}]$.

## 2.3 | Stochastic lRNN Outperforms Hankel DMD on Filtered Datasets

A key advantage of this stochastic formulation emerges when fitting lRNNs to filtered data, a standard step in fMRI preprocessing pipelines [39, 40]. Figure 3 compares spectral recovery in stochastic lRNNs to Hankel dynamic mode decomposition (HDMD), a widely used autoregressive alternative [41, 42] (see Methods Section), in the task of learning a 10-D multivariate Ornstein-Uhlenbeck process. HDMD recovers the full distribution of poles when trained on unfiltered long datasets (green dots in Figure 3A, top). However, applying a low-pass filter induces an artificial collapse of the poles toward the imaginary axis (Figure 3A, middle

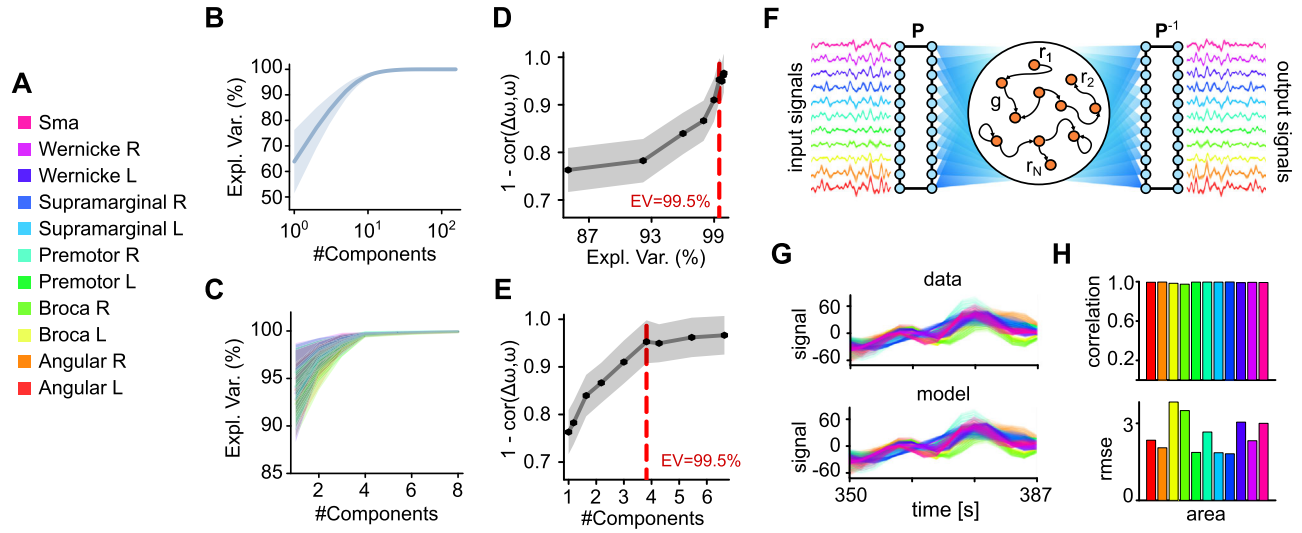


**FIGURE 2** | Compression and prediction of resting state fMRI activity. (A) Color legend associated with the 11 cortical areas taken from the language network. (B) Explained variance (EV) from PCA of the fMRI BOLD signals of the voxels in the language network. Solid line, average across the 20 considered subjects. Shaded area, standard deviation across subjects. (C) EV of the PCA for each cortical area. Colors as in (A). (D,E) 1 minus the correlation between the residuals from the PCA reconstruction ($\Delta\omega$) and the original data ($\omega$) per single area in the last (unseen) 37.5 s. A value equal to one means that the remaining information is just white noise. Vertical dashed lines, explained variance (D) and number of first PCs (E) required to reach the EV of 99.5%. (F) Schematic design of the used lRNN. The input from each area is compressed to have an EV of 99.5%. $\boldsymbol{P}$ is the matrix operating this linear transform leading to an average of 44 observables. The lRNN output reconstructing these observable is then uncompressed applying $\boldsymbol{P}^{-1}$. (G) Example of forecasted (bottom) and original (top) activities during the test window (last 37.5 s). Each curve is the BOLD signal of a voxel with color indicating the cortical area of belonging. (H) Forecast performances of the post-learning autonomous lRNN per area. Correlations between predicted and experimental activity (top) and root mean square error (bottom).

and bottom), thereby obscuring the true dissipative structure of the underlying dynamics. This structure is not recovered even when performing singular-value decomposition with a truncated rank that excludes noisier dimensions (Figure S1). By contrast, lRNNs with appropriately tuned noise (red dots in Figure 3A) preserve accurate spectral recovery even under filtering and in the small-dataset setting, closely matching the ground-truth eigenvalue distribution.

This robustness depends on selecting an optimal noise intensity. Figure 3B shows noise-level tuning using the maximum real part of the poles (top), prediction accuracy (middle), and long-run covariance matching (bottom). Increasing noise level strengthens the effective regularization during learning, which progressively reduces the real parts of the inferred poles by pushing them back toward the original random bulk. Focusing on the stable regime (white region in Figure 3B), where the maximum real part is compatible with having negative values (threshold: > 3 standard deviations), the optimal noise level maximizes short-horizon prediction accuracy while best matching the signal's statistical properties. In particular, (i) we computed the correlation between predictions and data over the final 20% of the dataset, and (ii) we ran the autonomous and stochastic lRNN for an additional duration equal to the training window (the first 80% of the dataset) and measured the discrepancy between the resulting covariance matrix and that of the training data.

An intermediate noise regime, near the stability boundary, tends to optimize all three metrics. Excessive noise intensities degrade prediction and prune too many modes, whereas insufficient noise fails to infer stable, realistic BOLD fluctuations eventually leading to divergence. This sweet spot enables consistent estimation across sample sizes, as reflected in the pole estimation errors in Figure 3C (red curves), whereas HDMD (green curves) continues to produce erroneous estimates.

Such results encapsulate the central distinction between the lRNN and standard AR/DMD approaches in their modeling assumptions. These methods estimate a linear map for the conditional mean evolution. Our lRNN instead fits a dissipative, noise-driven, higher-dimensional process. For heavily low-pass filtered signals such as BOLD, "best linear predictor" methods may converge to spectra that describe the smoothed average evolution (effectively a Fourier-like decomposition of the filtered signal), whereas the stochastic lRNN is constrained to produce a stable stochastic surrogate, which is precisely why spectral recovery remains robust under filtering and in the small-dataset regime.

## 2.4 | Stochastic lRNN as a Digital Twin of Brain Activity

To validate the effectiveness of this method in accurately replicating brain activity (low-pass filtered and with a limited number of samples), we numerically integrated the network dynamics for an additional 387.5 s (Figure 4A). We then compared the functional connectivity (FC) – defined as the correlation between BOLD signals of all possible voxel pairs – calculated from this simulation to that obtained from the experimental data (Figure 4B). Furthermore, the temporal evolution of the FC, referred to as functional connectivity dynamics (FCD), closely aligned with the experimental observations. This similarity was quantified by the

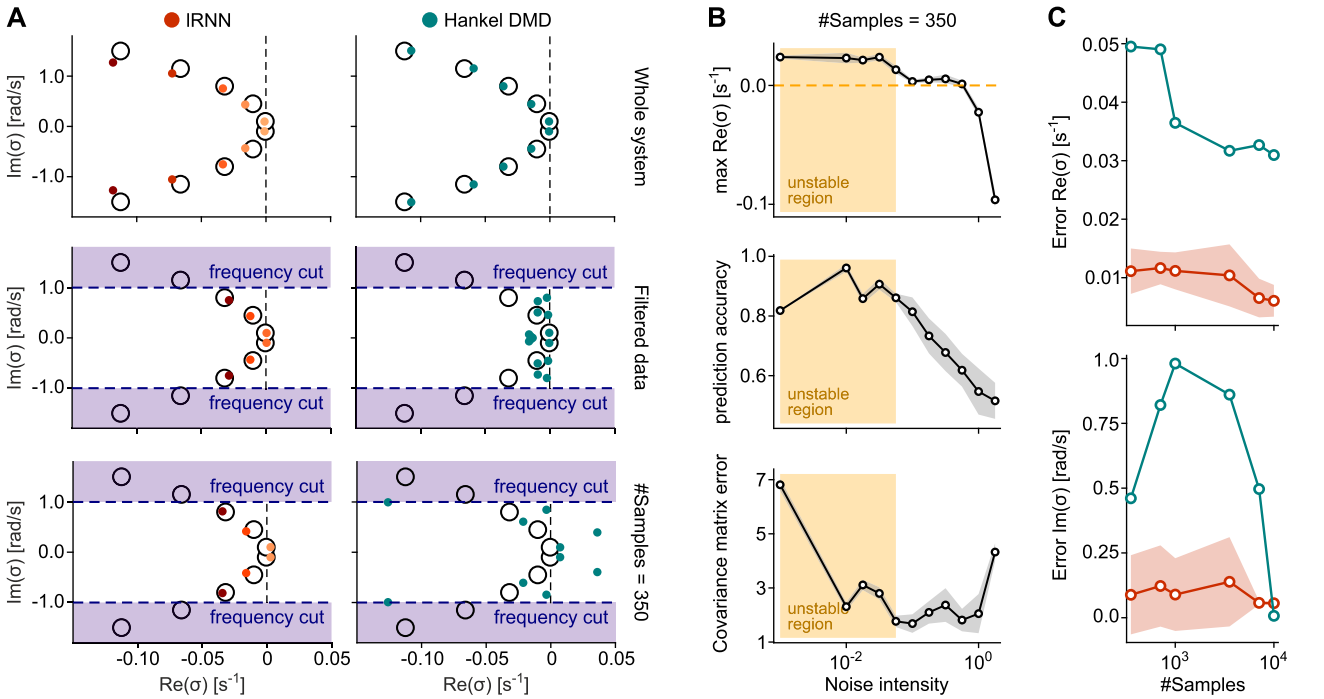


**FIGURE 3** | Optimal stochastic lRNNs recover filtered dynamics more accurately than Hankel DMD. (A) Comparison of estimated spectra for a 10-D multivariate Ornstein-Uhlenbeck process: lRNN with $N = 2000$ units and endogenous noise (red dots) versus Hankel DMD (green dots), with the ground-truth spectrum (black circles). From top to bottom: (i) training on the full unfiltered dataset; (ii) training on full low-pass filtered dataset with cut-off frequency at 1 rads$^{-1}$; (iii) training on a low-pass filtered dataset of 350 samples. (B) Tuning of lRNN noise intensity based on three metrics (mean over 20 noise realizations): the maximum real part of the inferred poles $\sigma_k$; correlation between predicted and true signals; and mismatch between the long-run reproduced covariance and the training-set covariance. Yellow region: unstable regime where max $\mathrm{Re}\,\sigma_k$ is significantly greater than 0 (threshold: > 3 standard deviations). (C) Estimation error of the real and imaginary parts of the first three eigenvalues (sorted by real part) for lRNN (red) and Hankel DMD (green) as a function of available number of samples in the filtered dataset. Shadings: standard deviation across the 20 noise realizations.

Jensen-Shannon distance (JSd) between the distribution values of the associated matrices (Figure 4C).

Finally, we conducted a comprehensive analysis of the optimal intensity of noise across all subjects (Figure 4), focusing on the accuracy of BOLD signal forecasting and the lRNN capability to reproduce both FC and FCD. To evaluate the quality of these aspects simultaneously, we introduced a success rate defined by applying fixed thresholds to the performance metrics: accuracy > 80%, FC correlation > 85% and JSd < 0.1. Additionally, we assessed the stability of the inferred lRNN by examining the maximum real part of the poles $\sigma_k$ (Figure 4D). According to what is shown in the previous Subsection, increasing the noise intensity during inference progressively shifts the maximum real part of the poles leftward, eventually making it negative. This spectral drift enforces dissipativity, thus preventing marginally stable or weakly damped solutions, and stabilizes trajectories, thereby improving out-of-sample predictions. This results further demonstrate that noise in our lRNN acts as a constructive element, essential for recovering the correct stochastic dissipative dynamics that best match the experimental time series. This analysis enabled us to identify a common optimal intensity $\gamma$ of endogenous noise across subjects with a standard deviation of $10^{-6}$ (Figure 4H). Interestingly, the existence of an optimal noise displays some similarities with the resonance phenomenon found in other machine-learning studies [43], which in our case robustly emerged in all the subjects.

## 2.5 | Dynamical Properties of the Digital Twin

We showed that RC with lRNNs can effectively reproduce resting state fMRI activity. Each simulated voxel can be described as a decomposition of linearly evolving modes. The dynamical properties of this decomposition can be effectively represented by the poles $\sigma_k$ of the Laplace transform of the inferred lRNN, which are the eigenvalues $\tilde{\lambda}_k$ of the learned synaptic matrix $\tilde{\boldsymbol{W}}$. This spectrum of eigenvalues is shown in Figure 5A for an example subject. For each of the modes, the relevance is averaged across all voxels and it is color coded, showing a pattern similar to what seen in Figure 1E where only a subset of poles are moved towards the imaginary axis. These poles are the most relevant and display a specific organization like a rotated parabola. The modes with the highest relevance (i.e., the darkest) appear to be distributed at specific frequencies falling into the range of infra-slow oscillations (< 0.1 Hz), a typical footprint of the resting state and of the unconscious brain activity [44–47]. As a representative example, in Figure 5A, the most relevant mode is one of the most persistent, characterized by a relatively small $\mathrm{Re}\,\sigma_k$, and exhibits a resonant frequency of approximately 0.02 Hz.

Subsequently, we investigated the stability of this representation exploring how the spectrum of $\tilde{\boldsymbol{W}}$ and the mode relevance change according to the number $N$ of units in the lRNN. As shown in Figure 5B, the density of eigenvalues in the complex plane ($\mathrm{Re}\,\sigma$, $\mathrm{Im}\,\sigma$) becomes less and less scattered with increasing

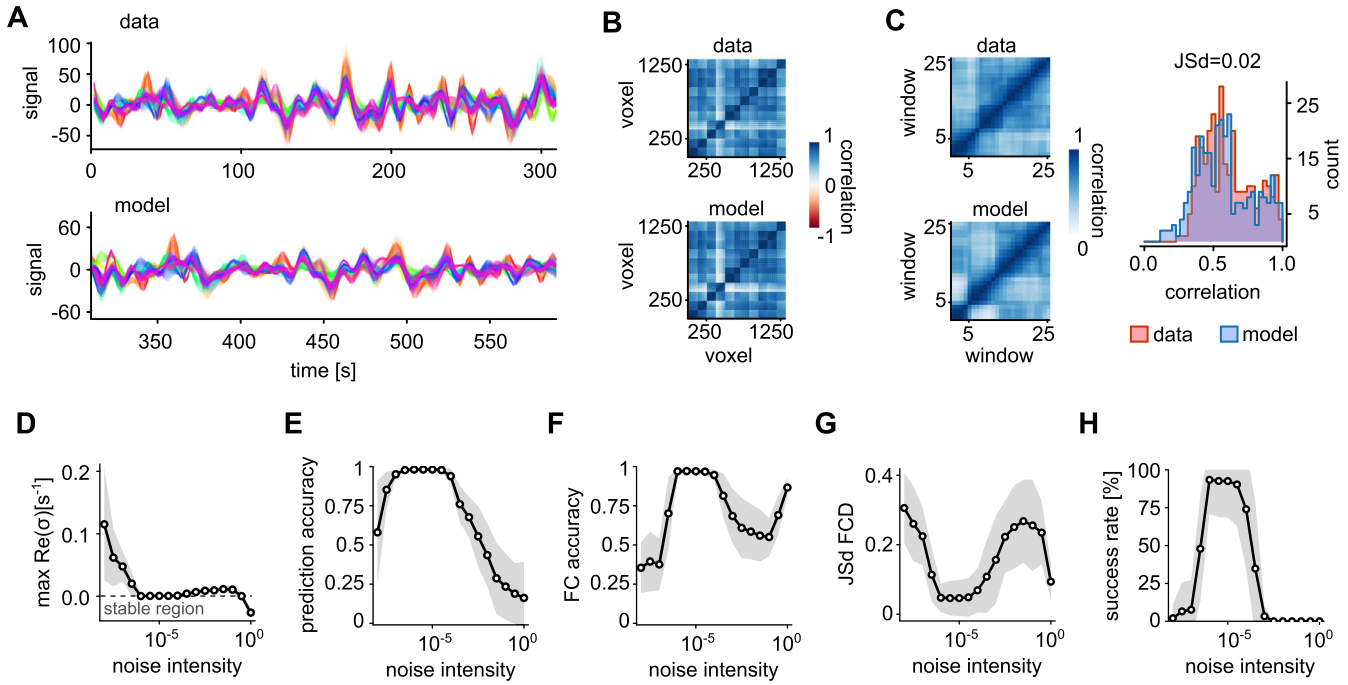


**FIGURE 4** | Endogenous noise in lRNN is needed to replicate brain activity. (A) Comparison of actual brain activity (top) with the corresponding output from a trained autonomous lRNN over the subsequent prediction interval (bottom). (B,C) Analysis of functional connectivity (FC) and functional connectivity dynamics (FCD) computed from the activity depicted in (A). Right, distribution of FCD values for lRNN-generated and exeprimental time series with the related Jensen-Shannon distance. (D–G) Features and performance statistics of inferred lRNNs across subjects varying the amplitude of endogenous noise: maximum pole real part (D), correlation between data and model prediction (E), FC correlation with actual data (F), and Jensen-Shannon distance between FCDs (G). (H) Success rate in reproducing above-threshold performances of the features in panels (E–G) (see main text for details).

network size. For clarity, and to save space, here we plotted only one half of the complex plane; the other half follows by complex-conjugate symmetry, which holds by having real-valued synaptic matrix. Besides, all the performance measures converged to fixed values corresponding to a high quality of the data replicate by the digital twin (Figure 5C-E), proving that a finite size lRNN can reach asymptotically high performances. Intriguingly, as convergence is approached the average number of relevant modes stands at about 60 independent components (Figure 5F). This number is significantly higher than the number of experimental PCs provided as input to the lRNN (44 on average). Thus, brain activity replicated by the inferred lRNN appear to live in a latent state space whose dimensionality is larger than the one determined by its observation.

Functional connectivity illustrated in Figure 6A-top for an example subject is then fully replicated by a limited number of independent modes of the inferred (and stochastic) lRNN. This effective copy occurs even though the dimensionality of the latent state space of the digital twin is significantly smaller than the number of voxels encompassing the examined language network. To gain a deeper understanding of this result, we 'opened the box' by examining the linear transformation $\boldsymbol{Z} \equiv \boldsymbol{P}^{-1}\boldsymbol{\Xi} \in \mathbb{R}^{L\times N}$. According to Equation (7), this matrix facilitates the mapping of the $N$ eigenmode projections $\boldsymbol{v}(t)$ onto the $L$-dimensional voxel-wise BOLD activity, represented as $\boldsymbol{x}(t) = \boldsymbol{Z}\boldsymbol{v}(t)$. The absolute values of the elements in this matrix are displayed in Figure 6B for the same subject. Unsurprisingly, the most significant contributions arise from the slowest (rightmost) modes, specifically those with the largest, albeit still negative, $\mathrm{Re}\,\sigma_k$. Additionally, by measuring the covariance between the eigenmode projections $v_k(t)$, it becomes evident from Figure 6C that they are largely uncorrelated. Only about 60 of the slowest modes exhibit significant variability, as indicated by the dark diagonal elements. This empirical evidence suggests that functional connectivity can be estimated directly using the cosine similarity $\mathrm{S_C}$ between the rows of the matrix $\boldsymbol{Z}$ (see Methods Section):

$$\mathrm{FC}_{jk} \approx \mathrm{S_C}\left(\langle Z_j|\,, \langle Z_k|\right) \tag{10}$$

In Figure 6A-bottom, the similarity matrix is presented, showcasing a remarkable overlap with the experimental functional connectivity **FC**. It is important to note that in computing $\boldsymbol{Z}$, we considered only the columns of $\boldsymbol{\Xi}$ associated with the relevant modes. This further confirms that the remaining $N - 60$ degrees of freedom are nearly irrelevant.

## 2.6 | lRNN as a Proxy to Characterize Subjects and Brain Areas

Given that inferred lRNNs reliably replicate observed brain activity and their modal decomposition defines their dynamical properties, a critical question emerges: Can lRNNs serve as proxies for understanding the similarities and differences among subjects and cortical areas? To address this, we characterized the spectrum of relevant eigenvalues by focusing on two key features: the linear relationship between their real and imaginary

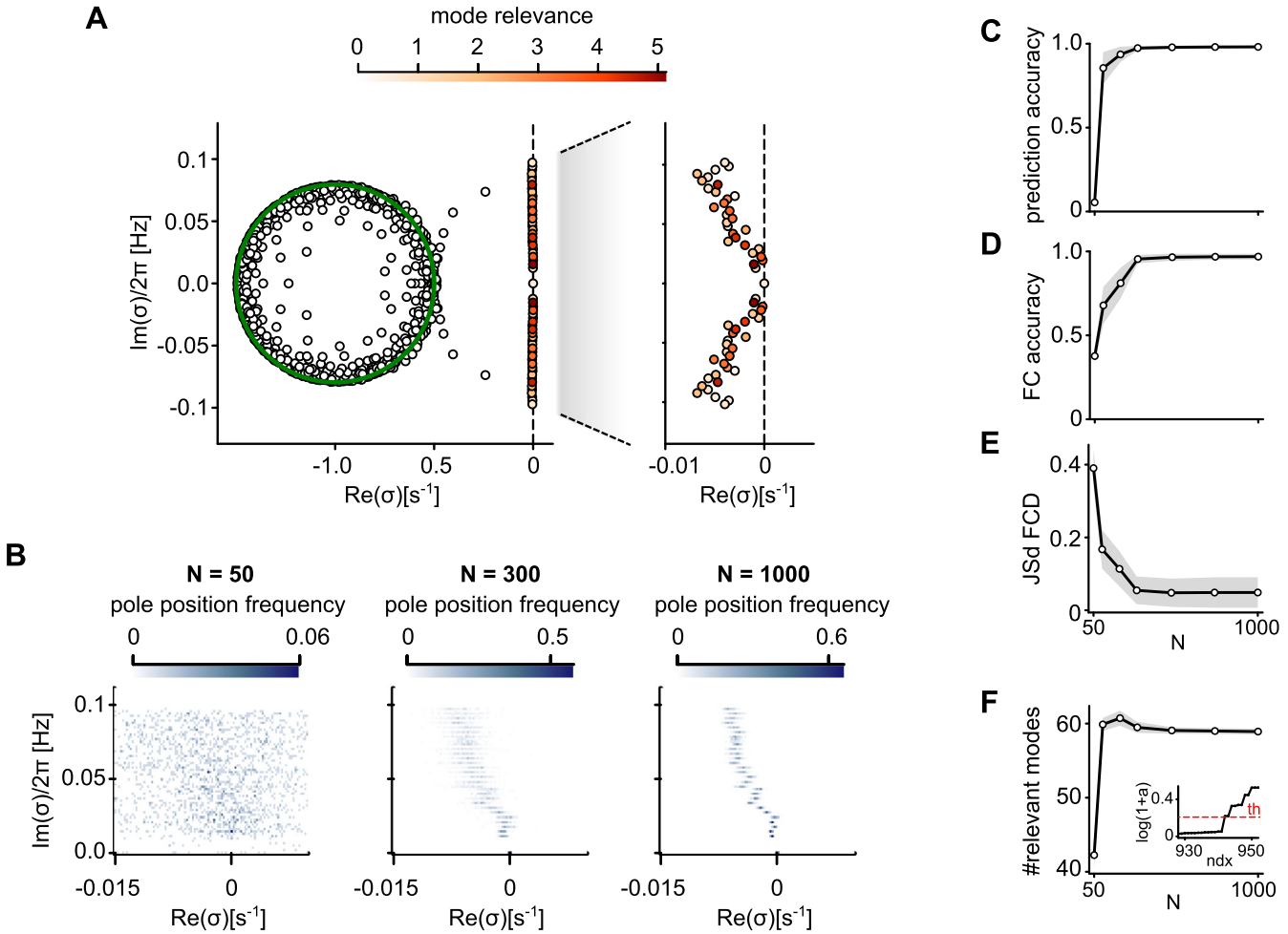


**FIGURE 5** | Convergence of the mode decomposition with increasing lRNN size. (A) Mode decomposition of a single cortical area for an example subject. Eigenvalues of $\tilde{W}$ are color-coded as in Figure 1E. (B) Distribution of eigenvalues for three distinct network sizes $N$, analyzed across 100 randomly sampled lRNNs. (C) Accuracy of forecasted activity. D, E) Reproduction fidelity of FC (D) and of FCD (E). (F) Number of relevant modes (i.e., with relevance greater than 5%). Inset, sorted mode relevance and related threshold (dashed line). Gray shadings and black lines, standard deviation and mean across subjects, respectively.

parts, and the overall significance of each oscillation frequency, as illustrated in Figure 7A. The slope of this linear relationship indicates the correlation between decay times and oscillation frequencies, while the coefficient of determination ($R^2$) assesses the goodness of fit for this linear model. In this 2D parameter space, subjects are systematically distributed among three distinct regions (Figure 7B). Those with high $R^2$ values display either steep or shallow slopes, whereas subjects with low $R^2$ values indicate a poor linear fit and suggest the presence of persistent isolated high-frequency oscillatory modes. Within this representational space, spectra characterized by steeper slopes (located at the bottom) are associated with poles that are closer to the imaginary axis. This suggests that those lRNNs (and consequently subjects) exhibit longer relaxation time scales, which may be linked to brain activity approaching a critical point where the resting state could become unstable and display more complex dynamics.

Shifting our focus to the frequency vectors derived from the inferred lRNN, we sought statistical regularities that reveal invariant features associated with individual brain areas and subjects. The frequency vector for a given brain area in a subject is represented by the histogram of $\mathrm{Im}\,\sigma_k$ from the corresponding lRNN, each counted with multiplicity proportional to its absolute mode relevance averaged across the voxels in that area: $\langle |\Xi_{jk}| \rangle_{j \in \mathrm{BA}}$. We then performed a linear discriminant analysis (LDA) in the frequency-vector space to identify patterns that effectively discriminate between different brain areas (Figure 7C). It is important to emphasize that the frequency vectors used in this analysis were obtained from 100 independently inferred lRNNs for each subject. This approach ensured a fair and robust discrimination analysis, as it explicitly incorporates the unavoidable variability introduced by the stochastic lRNN inference process, although (as shown in Figure 4B) the distribution of poles is already highly stable for the chosen network size. Despite originating from distinct subjects, the resulting data points clustered according to their respective brain areas. It is important to note that the compositions of the latent variables ($L_1$, $L_2$, and $L_3$) correspond to specific frequency patterns (Figure 7D). Among these, the pattern associated with $L_1$ appears most relevant, as data points on this linear manifold are distinctly clustered based on the brain area of belonging (Figure 7C).

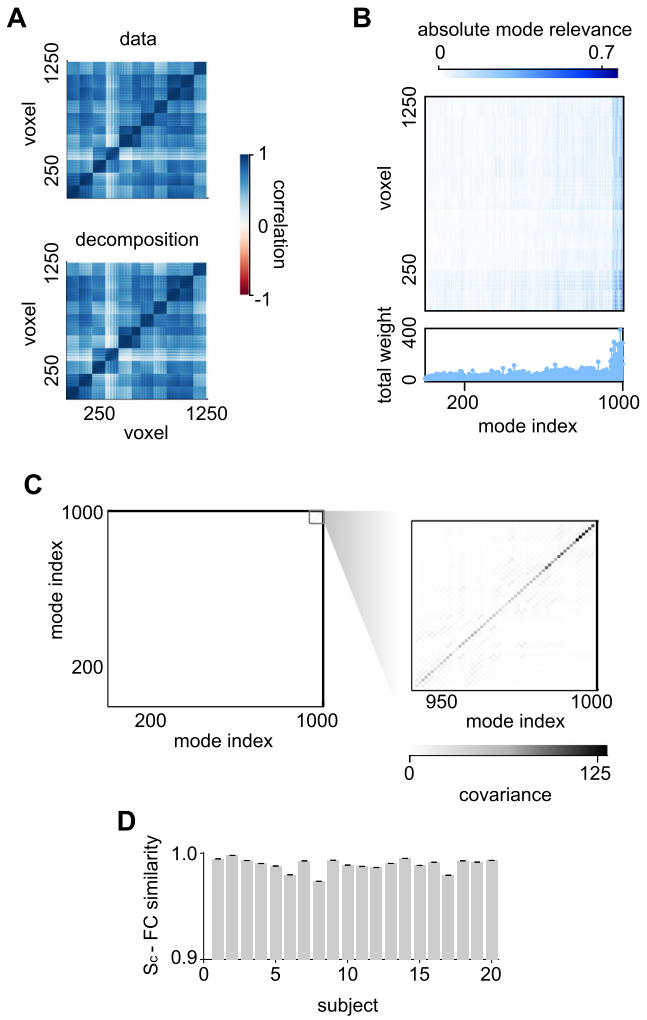


**FIGURE 6** | Low-dimensional mode decomposition explains functional connectivity. (A) Functional connectivity of an example subject (top) and cosine similarity (bottom) between rows of the matrix $\boldsymbol{Z}$ mapping eigenmode projections to voxel-wise BOLD activity from the inferred lRNN with $N = 1000$ units (see main text). (B) Absolute mode relevance $|Z_{jk}|$ for the same lRNN together with the total weight $\sum_j |Z_{jk}|$ of each of the $N$ modes (bottom). Modes are sorted according to the real part of the associated eigenvalue ($\mathrm{Re}\,\tilde{\lambda}_k$). (C) Covariance matrix of the eigenmode projections $v_k(t)$ (see Equation (7)) during the extended autonomous phase (387.5 s post learning in Figure 4A-bottom). Right, zoomed-in view on the slow and most relevant modes. (D) Correlation between the upper diagonal elements of the FC and the cosine similarity ($S_C$) of the rows of $\boldsymbol{Z}$ for each subject. Averaged over 100 noise realizations. Error bars, SEM.

To further isolate the frequencies at which each brain area exhibits unique dynamics, we computed the corresponding linear decoders (see Methods Section). The elements of the coefficient vectors characterizing each decoder are assigned high absolute values if the associated frequency plays a crucial role in distinguishing that area from others (Figure 7E). The resulting patterns are markedly distinct from the average frequency vector, which, in contrast, reveals a continuum of relevant frequencies (Figure 7F). In fact, the frequency vectors show significant similarities, making it challenging to differentiate between areas even if they are located in different hemispheres (Figure 7H). In contrast, the decoders exhibit a lower degree of similarity, except when comparing the corresponding areas of both hemispheres (Figure 7G). Interestingly, there is an exception to this trend: the decoder for the right Wernicke area appears to highlight frequencies that differ from those identified for the same area in the left hemisphere.

Further insights into brain lateralization are revealed by examining the accuracy of the decoders. Specifically, it is more difficult to linearly recognize left hemisphere patterns by this method for the Broca, premotor, and supramarginal areas (Figure 7I). This asymmetry likely reflects the well-established left-lateralization of the resting-state language network in right-handed individuals. In particular, the inferior frontal gyrus (including Broca's area), the supramarginal gyrus, and the premotor/pre-supplementary motor areas exhibit stronger intrinsic left-hemisphere connectivity at rest [48–50]. As a consequence, their BOLD time courses tend to covary more strongly within the left hemisphere, making these regions harder to dissociate and likely explaining the reduced decoder accuracy compared to their more independent right-hemisphere homologues. It is also worth noting that Wernicke's area appears as the most difficult region to recognize, thereby explaining the decoder dissimilarity observed in Figure 7G. Importantly, despite these heterogeneous coupling strengths across subnetworks, stochastic lRNNs can still be robustly inferred, demonstrating their ability to capture the effective dynamics of resting-state activity even in the presence of strong intra-network coherence.

In short, these findings demonstrate that inferred lRNNs can serve as powerful proxies to characterize the unique dynamics of subjects and areas of the brain, capturing their specific dynamical footprints.

## 3 | Discussion

The successful application of linear recurrent neural networks within the reservoir computing framework represents a significant methodological contribution to model resting-state fMRI data. Our results show that lRNNs effectively capture brain dynamics, suggesting that resting-state activity can be approximated as a linear system in a high-dimensional space, supporting recent studies advocating for linear models in this context [51]. Our approach provides valuable and easily accessible insights into the functional connectivity structure as similarity patterns between vectors associated with the eigenmode decomposition of the inferred lRNN. This low-dimensional description of the whole brain activity moves the focus to the mode's subspace offering a normative framework where brain areas and inter-subject comparisons are more straightforward.

Notably, the mode dynamics described in Equation (6) provides an equivalent linear representation for the observables of the system under investigation – namely, the single-subject voxel-wise BOLD activity. The resulting vector of all time derivatives (see Methods Section) serves as an alternative descriptor of the system's state [52], potentially encapsulating the trajectory evolution at any given time. This concept is reminiscent of the dynamic mode decomposition with Hankel matrix [41, 42], but with a significantly reduced dimensionality constraint.

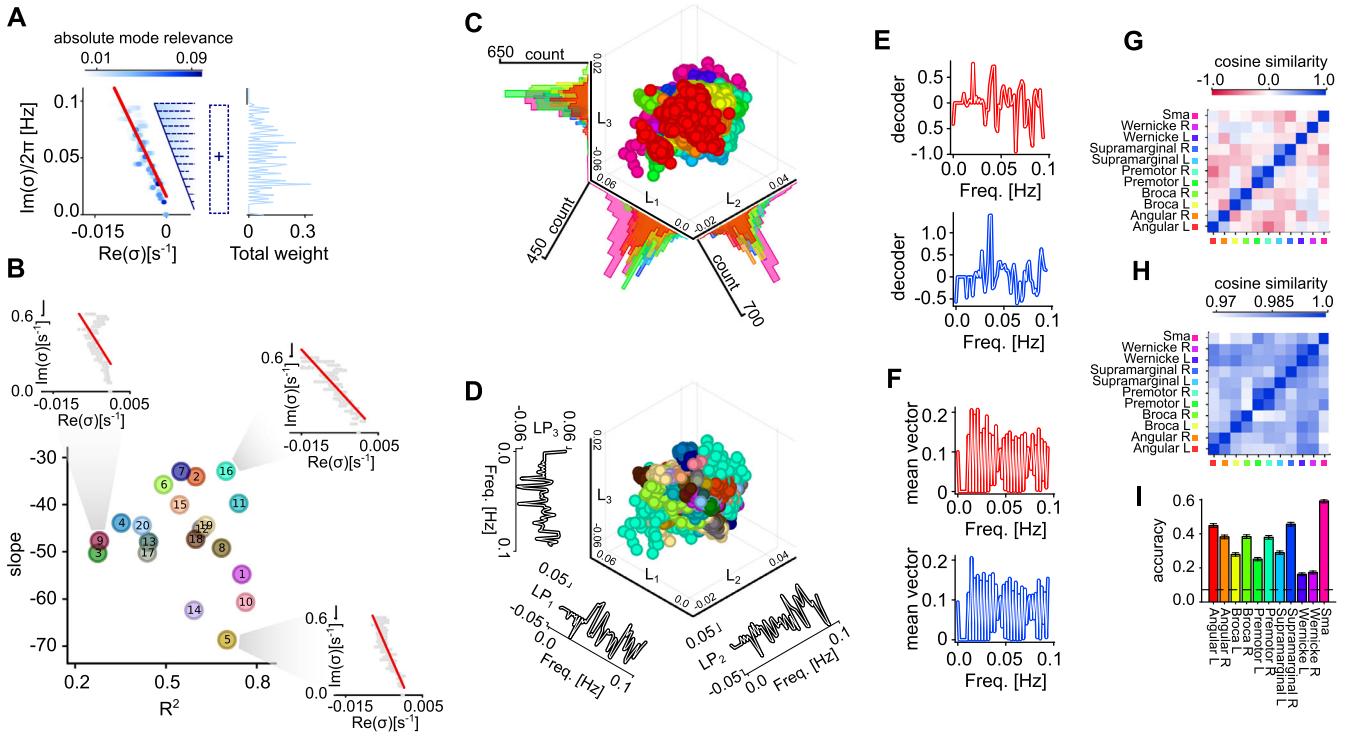


**FIGURE 7** | Characterization of subjects and brain areas. (A) Left, Spectrum of relevant eigenmodes for a sample subject. A linear regression (red line) is performed between the real and imaginary parts of the poles, $\sigma_k$, with the slope serving as the first feature that characterizes the subject. The coefficient of determination, $R^2$, from this regression represents the second feature. Right, Frequency vector, i.e., the histogram of Im $\sigma_k$ each counted with multiplicity given by the averaged absolute mode relevance across voxels in a single brain area: $\langle |\Xi_{jk}| \rangle_{j \in \mathrm{BA}}$. (B) Slope of linear regression and $R^2$ of the 20 subjects in the dataset. Insets: eigenmode spectra of three representative subjects with the related linear regressions (red lines, see main text for details). (C) First three latent variables derived from the linear discriminant analysis (LDA) applied to the frequency vectors of 11 brain areas across all 20 subjects. The densities of circles for each latent variable are also plotted. For each subject, 100 lRNNs are inferred by sampling different endogenous noise in their stochastic dynamics, resulting in a total of 22,000 circles. Colors code brain areas according to Figure 2A. (D) Same as (C) but with circles colored according with the subject color as in (B). Composition of the three latent variables $\mathrm{LP}_1$, $\mathrm{LP}_2$, $\mathrm{LP}_3$ are show on the side of their respective axes. (E,F) Linear decoders (E) discriminating the different areas in the frequency-vector space and average frequency vectors (F) for two sample brain areas (red and blue). (G,H) Cosine similarity between linear decoders (G) and between frequency vectors (H) per area averaged across subjects. (I) Accuracy of decoders along the training set. Dashed black line highlights the random chance level separation.

However, we showed that Hankel-DMD in practice tends to converge toward a Fourier-like representation when the data undergo standard frequency-band filtering (as commonly done in fMRI analysis pipelines), while our lRNN formulation can still recover the correct continuous-time spectrum, including nonzero decay rates. In filtered conditions, Hankel-DMD often yields poles whose real parts collapse toward 0, effectively favoring purely oscillatory modes and obscuring the dissipative structure of the underlying dynamics (Figure 3). This result also stems from considering an optimal amount of memory-less fluctuations integrated into the lRNN units' endogenous activity. This feature enables effective modeling of non-informative fast fluctuations in BOLD activity without requiring inference of additional high-frequency, fast-decaying modes, thereby reducing the dimensionality of the latent state space visited by the deterministic components of the inferred lRNNs. This characteristic makes the presented method particularly advantageous and flexible, allowing the inference of lRNN-based digital twins at the single-subject level using a limited amount of experimental data. In contrast, several recent approaches have applied deep reinforcement learning (DRL) to fMRI timeseries or brain-network analysis [53, 54]. These DRL-based methods typically require reward-driven optimization strategies and substantially larger training pipelines [55]. Differently from these approaches, we show that our stochastic lRNN framework achieves accurate, robust, and interpretable data-driven modeling by 'learning' only a limited number of readout weights through straightforward linear regression, similarly to classical echo state networks [4].

The reported results have the potential to open up several avenues for future research. Indeed, the possibility offered to represent the single-subject as a point into a low-dimensional map – i.e., a landscape (Figure 7B) – allows in principle to follow in time its drift in longitudinal studies like those focused on neurodegenerative diseases [56] or aging [57–59]. Here, the novelty is the fully data-driven nature of the approach to build such a landscape of subjects compared to those relying on model-driven inferences [60]. Characterizing the trajectory followed in this landscape by each subject could provide valuable insights into the alterations in functional connectivity and dynamics. For instance, the emergence of unstable modes may be linked to transitions into pathological states of a specific neurological disease. By perturbing the system with specific and personalized

rehabilitative or pharmacological treatments that aim to suppress these "pathology-related" modes, we may gain a means to intervene and reduce the occurrence of such transitions, ultimately helping to tailor optimal therapeutic strategies [28, 61].

In line with this, as in classic digital twins, the inferred lRNN can incorporate any additional available information about the system simply as new input. An external stimulation can potentially act as a selector for different dynamical regimes, allowing for two distinct representations with a single network [62]. The functional connectivity itself could be a proxy for state transition. Indeed, a significant change in functional connectivity might indicate that the decomposition has also changed, suggesting that the observed system is close to a different equilibrium point [63]. These changes can be tracked following the distribution of poles that characterize the autonomous dynamics of the inferred linear RNN. This is the case for transitions in global brain states – such as those governing the sleep-wake cycle – that arise from network destabilization, a hallmark of criticality [64–66]. Under these conditions, pole distributions correlate with longer relaxation time scales and steeper slopes (as illustrated in Figure 7B), possibly underlying previously observed state-dependent spectral signatures [67]. This metastable dynamics results in broader excursions within the latent state space of neural activity, making predictions of future BOLD activity more challenging for the inferred lRNN. The quality of lRNN predictions can then serve as a proxy for dynamical stability, akin to findings in intracortical local field potentials (LFPs) of macaque monkeys during anesthesia-induced transitions between wakefulness and unconsciousness [68].

While this study demonstrates the efficacy of lRNNs in modeling resting-state brain activity, several limitations must be acknowledged. The assumption of linearity, although supported by our results, may not fully capture the complexity of brain dynamics under all conditions. Nonlinear phenomena, transient states, and the influence of external stimuli could require more sophisticated modeling approaches [14, 69]. Alternatively, the apparent suitability of resting-state activity being well represented by a stationary multivariate Ornstein-Uhlenbeck process (i.e., Brownian motion) [70, 71] may reflect an intrinsic limitation of the BOLD signal in conveying detailed neural dynamics. Indeed, this hemodynamic-related signal acts as a lumped observable that inevitably linearizes the spiking activity of relatively large neuronal assemblies [51, 72]. The coexistence of endogenous noise with the relatively high-dimensionality of the BOLD time series can give rise to a multivariate Brownian motion exhibiting a rich repertoire of restless spatiotemporal modes [71]. These are captured in our lRNN by relatively wide distributions of poles. Within this linear stochastic system, noise spontaneously excites its normal modes, eliciting co-fluctuating patterns closely linked to the functional connectivity measured in rs-fMRI. Interestingly, this structured stochasticity may, at least in principle, exhibit similarities to turbulence [73, 74].

However, assuming that all voxel activities fluctuate around a single fixed point imposes a strong constraint on the brain's computational repertoire. Neural computation is thought to emerge from trajectories evolving across landscapes rich in saddles and metastable states [75, 76], whereas lRNNs cannot, by construction, represent the latent state space of a genuinely multistable system. Nonetheless, multistability could in principle be handled by partitioning the state space into quasi-linear regions, each represented by a local lRNN, as in piecewise-linear approaches [77], or by allowing unstable modes when external inputs drive transitions, similar to recent modeling of cortical trajectories [78].

Besides this modeling limitation, our results raise an important question: how can we reconcile the evidence that resting-state fMRI activity is well described by stochastic lRNNs with the necessity of nonlinear dynamics for brain computation? A tentative answer is that, during rest, endogenous fluctuations have a magnitude comparable to the deterministic excursions of neural activity associated with relaxation dynamics. Consequently, the nonlinear components of brain activity may only become apparent when the system is pushed far from equilibrium – such as when the brain engages in cognitive functions like motor or perceptual decision-making. A promising direction for future work would then be to test this hypothesis by assessing whether an effective lRNN can be inferred from BOLD time series recorded while subjects engage in such cognitive tasks.

All these considerations are particularly relevant because the metastable neural dynamics are tightly linked to both healthy and pathological brain function. In this framework, a failure to infer a stochastic lRNN that faithfully reproduces resting-state BOLD signals can be interpreted as a quantitative marker of changes in metastability. Such failures in inferring an effective lRNN may therefore serve as a sensitive biomarker of brain dysfunction, ultimately aiding the identification of disease-related alterations in dynamical coordination (e.g., schizophrenia, depression, Alzheimer's disease, epilepsy; see [79]).

## 4 | Methods Section

### 4.1 | Dataset Description

#### 4.1.1 | Subjects

The study included 20 healthy, right-handed subjects (mean age ± SD = 37 ± 12; 7 females, 13 males) with no history of neurological disorders. The study was approved by the Institutional Review Board at Memorial Sloan Kettering Cancer Center, and informed consent was obtained from each participant. During the resting-state condition, subjects were instructed to lie in the scanner, keep their eyes open, try to think of nothing in particular, and maintain fixation on a central cross on the screen.

#### 4.1.2 | MRI Methods

A GE 3T scanner (General Electric, Milwaukee, Wisconsin, USA) and a standard quadrature head coil was employed to acquire the MR images. Functional images covering the whole brain were acquired using a (T2*)-weighted imaging sequence sensitive to blood oxygen level-dependent (BOLD) signal (repetition time, TR/TE = 2500/40 ms; slice thickness = 4.5 mm; matrix = 128 × 128; FOV = 240 mm; volumes = 160). Functional matching axial T1-weighted images (TR/TE = 600/8 ms; slice thickness = 1 mm) were acquired for anatomical co-registration purposes.

#### 4.1.3 | Data Preprocessing

Functional MRI data were processed and analyzed using the software program Analysis of Functional NeuroImages (AFNI; Cox, 1996). Head motion correction was performed using 3D rigid-body registration. The first volume was selected to register all other volumes. The first volume was chosen because it was acquired before the anatomical scan. Both task and resting state fMRI scans were monitored using a real time post-processing software BrainWave (BrainWave RT, Medical Numerics, Germantown, MD) to monitor brain activity and the head motion. For subjects showing severe head motion over time, generally, the scan was repeated. For small head motion (less than 2 voxel size), a motion correction algorism (iterated linearized weighted least squares) considering three translation and three rotation parameters against a reference volume was applied. The obtained six parameter time courses were also integrated in the statistical analysis to regress out residual motion-correlated artifactual voxels. Spatial smoothing was applied to improve the signal-to-noise ratio using a Gaussian filter with a 6 mm full width of half maximum. Corrections for linear trend and high-frequency noise were also applied. Resting-state data requested some more preprocessing steps. They were corrected for head motion by regressing head motion data and the first five principal components of the white matter and CSF signals. They were also detrended, demeaned, and band-pass filtered (frequency range 0.01–0.1 Hz). All fMRI data were registered to the standard space (Montreal Neurological Institute MNI152 standard map).

### 4.2 | fMRI Dimensionality Analysis

Principal Component Analysis (PCA) is a widely used dimensionality reduction technique used to transform a dataset into a lower-dimensional space while preserving most of the information in the original data [80]. In the context of the given problem, PCA was applied to the set of BOLD signals to reduce redundancy and keep only the relevant information.

The PCA-based dimensional reduction of BOLD was carried out based on the initial 280 s (learning period) of each subject's data. Referring to Figure 2, the number of principal components taken into account were those capturing 99.5% of the variance in the time series of each area. The resulting transformation was represented by a block matrix $\boldsymbol{P}_0$, where each block's columns $\boldsymbol{p}_i$ correspond to the eigenvectors of the covariance matrix of the respective cortical area:

$$\boldsymbol{P}_0 = \begin{bmatrix} \boldsymbol{p}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{p}_{11} \end{bmatrix} \tag{11}$$

The output undergoes orthogonalization through an additional PCA step ($\boldsymbol{P}_1$) utilizing all components. Subsequently, normalization was performed by applying the diagonal matrix $\boldsymbol{N}$, which contains the reciprocals of the maximum values of the compressed signals observed during the learning period. This approach ensured that all currents enter the reservoir with the same maximum strength. The final transformation matrix $\boldsymbol{P}$ is computed as:

$$\boldsymbol{P} = \boldsymbol{N}\boldsymbol{P}_1\boldsymbol{P}_0 \tag{12}$$

While the antitransformation matrix $\boldsymbol{P}^{-1}$ is calculated as:

$$\boldsymbol{P}^{-1} = \boldsymbol{P}_0^T\boldsymbol{P}_1^T\boldsymbol{N}^{-1} \tag{13}$$

The data was effectively reduced and then restored, ensuring the preservation of essential information while reducing dimensionality for the analysis. The decomposition for voxel activity was derived by applying $\boldsymbol{P}^{-1}$ to the decomposition matrix for the actual network inputs.

### 4.3 | Koopman Operator

The discrete-time Koopman operator $\mathcal{K}^1 : \mathcal{F} \to \mathcal{F}$ is a linear operator defined in an infinite-dimensional space $\mathcal{F}$ of observables of the system state [19, 20, 81]. The operation of the Koopman operator on an observable $g$ is described by the equation:

$$\mathcal{K}^1 g(\boldsymbol{x}_t) = g(\boldsymbol{x}_{t+1})\,. \tag{14}$$

Here, $\boldsymbol{x}$ represents a point in the state space. The eigenfunctions of the Koopman operator were potential observables of the system themselves, and had the peculiar property of evolving linearly in time:

$$\mathcal{K}^1\psi_k(\boldsymbol{x}_t) = \psi_k(\boldsymbol{x}_{t+1}) = \lambda_k^1\psi_k(\boldsymbol{x}_t) \tag{15}$$

These eigenfunctions $\psi_k(\boldsymbol{x}_t)$ constitute a basis of the space $\mathcal{F}$, enabling the representation of all the possible observables of the system as a linear combination of linearly evolving functions.

In general there existed a family of Koopman operators, $\mathcal{K}^t$, that advances a function forward by a time $t$:

$$\mathcal{K}^t g(\boldsymbol{x}) = g(\mathcal{S}^t(\boldsymbol{x}))\,. \tag{16}$$

Here, $\mathcal{S}^t$ denotes the time flow operator, with $\boldsymbol{x}(t) = \mathcal{S}^t[\boldsymbol{x}(0)]$. The infinitesimal generator of the Koopman operator family $\mathcal{L}g = \lim_{t\to 0}\frac{\mathcal{K}^t g - g}{t}$ is the Lie operator which evaluates the temporal change of observables:

$$\mathcal{L}g(\boldsymbol{x}(t)) = \frac{d}{dt}g(\boldsymbol{x}(t)). \tag{17}$$

The family of Koopman operators can then be expressed in this term $\mathcal{K}^t = e^{\mathcal{L}t}$.

### 4.4 | Hankel Dynamic Mode Decomposition (Hankel DMD)

Hankel Dynamic Mode Decomposition was employed to compute a finite-dimensional approximation of the Koopman spectral properties from the measured observables by applying DMD to a delay-embedded representation of the time series [41, 42]. Given a sequence of snapshots $\boldsymbol{\omega}(t_k) \in \mathbb{R}^M$, sampled at constant interval

$\Delta t$, a block-Hankel (time-delay) data matrix of memory length $q$ is constructed as

$$\boldsymbol{H} = \begin{bmatrix} \omega(t_1) & \omega(t_2) & \cdots & \omega(t_K) \\ \omega(t_2) & \omega(t_3) & \cdots & \omega(t_{K+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \omega(t_q) & \omega(t_{q+1}) & \cdots & \omega(t_{K+q-1}) \end{bmatrix} \in \mathbb{R}^{qM \times K} \tag{18}$$

whose columns represent delay-embedded states, while the first block-row coincides with the original observations. Two embedded snapshot matrices were then defined as $\boldsymbol{H}_0 = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_{K-1}]$ and $\boldsymbol{H}_1 = [\boldsymbol{h}_2, \ldots, \boldsymbol{h}_K]$, where $\boldsymbol{h}_k$ denotes the $k$-th column of $\boldsymbol{H}$, and the best-fit linear map on the embedded space was sought in the form $\boldsymbol{H}_1 \simeq \boldsymbol{K}\boldsymbol{H}_0$.

As in standard DMD, a truncated singular value decomposition (SVD) of the embedded snapshot matrix was computed as $\boldsymbol{H}_0 \simeq \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^T$, where the truncation rank $r$ was selected according to the retained SVD energy. The reduced operator is then obtained as

$$\boldsymbol{K} = \boldsymbol{H}_1 \boldsymbol{V}_r \boldsymbol{\Sigma}_r^{-1} \boldsymbol{U}_r^T \tag{19}$$

and its eigendecomposition $\boldsymbol{K}\boldsymbol{w}_j = \lambda_j \boldsymbol{w}_j$ provides the DMD eigenvalues $\lambda_j$. The associated generator spectrum is obtained from $\sigma_j = \Delta t^{-1} \log(\lambda_j)$.

### 4.5 | Reservoir Computing

The reservoir computing approach involved a RNN with fixed and random internal couplings (namely the 'reservoir') whose state was fed forward to a second set of 'readout' units [4, 31]. In this framework the required 'echo-state property' spontaneously emerges from the RNN collective dynamics. The units composing the RNN are intended to model homogeneous and local neuronal assemblies of a cortical network [82–84]. Each of the $N$ units in the network has activity state $r_j(t)$ evolving in time as

$$\tau \dot{r}_j = \Phi(h_j) - r_j \tag{20}$$

with $j \in [1, N] \subset \mathbb{Z}$. The decay time constant $\tau$ and the sigmoidal activation function $\Phi(h_j)$ is the same for all units. Here, we consider the limiting case of weak recurrent coupling, where the activation function can be approximated to a linear function: $\Phi(h_j) = h_j$. The synaptic input $h_j$ is the weighted sum

$$h_j(t) = \sum_{k=1}^{N} W_{jk} r_k(t) + h_j^{\text{in}}(t)$$

where $W_{jk}$ are the elements of the synaptic matrix (i.e., the internal couplings) $\boldsymbol{W} \in \mathbb{R}^{N \times N}$, and $h_j^{\text{in}}(t)$ is the external input received by the unit $j$. In reservoir computing, the external input is driven by the measured observable $\omega$ of the inspected system, which is fed into the network using random synaptic couplings: $h_j^{\text{in}}(t) = W_j^{\text{in}}\, \omega(t)$. Due to the high dimensionality of the trajectories in the RNN state space, a simple linear transformation represented by $\boldsymbol{W}^{\text{out}}$ is often sufficient to accurately map the RNN state into a desired output time series. More precisely, given $\boldsymbol{R} \in \mathbb{R}^{N \times T}$ whose $k$-th row represents the time series of the $k$-th unit during a learning period lasting $T$ time steps, and $\boldsymbol{\Omega} \in \mathbb{R}^{M \times T}$ has rows given by the $M$ observables $\omega_k(t)$ to be replicated, the linear map is computed by a ridge regression:

$$\boldsymbol{W}^{\text{out}} = \boldsymbol{\Omega}\boldsymbol{R}^T(\boldsymbol{R}\boldsymbol{R}^T + \beta \boldsymbol{I})^{-1} \tag{21}$$

where $\beta$ is a regularization parameter and $\boldsymbol{I} \in \mathbb{R}^{N \times N}$ the identity matrix. The target output coincides with the external input. Then, the resulting linear map can be used to simulate the external stimulation and predict the future steps of the data. This is equivalent to update the synaptic matrix of the reservoir. In particular $\boldsymbol{W}^{\text{in}}\boldsymbol{\omega}(t) \approx \boldsymbol{W}^{\text{in}}\boldsymbol{W}^{\text{out}}\boldsymbol{r}(t)$ leading to the autonomous system:

$$\tau \dot{\boldsymbol{r}} = \boldsymbol{\Phi}[\tilde{\boldsymbol{W}}\boldsymbol{r}] - \boldsymbol{r} \tag{22}$$

where $\tilde{\boldsymbol{W}} = \boldsymbol{W} + \boldsymbol{W}^{\text{in}}\boldsymbol{W}^{\text{out}}$ is the updated synaptic matrix.

### 4.6 | Linear Recurrent Neural Networks (lRNN) and Autoregressive Models

In the established framework of reservoir computing with linear RNNs (lRNN, i.e., with a linear activation function) by setting as initial condition $r_k(-\infty) = 0$, the system Equation (20) has the following solution:

$$\boldsymbol{r}(t) = \sum_{k=1}^{M} \int_{-\infty}^{t} \omega_k(t') e^{(\boldsymbol{W}-\boldsymbol{I})(t-t')/\tau} \, |W_k^{\text{in}}\rangle \, dt' \tag{23}$$

As above $\omega_k(t)$ denotes the $k$-observable measured from the inspected system. Assuming a time $t^*$ exists after which a matrix $\boldsymbol{W}^{\text{out}}$ maps the network state to the input data $\boldsymbol{\omega}(t) = \boldsymbol{W}^{\text{out}}\boldsymbol{r}(t)$, we can rewrite the above equation as:

$$\boldsymbol{\omega}(t) = \sum_{k=1}^{M} \int_{-\infty}^{t} \omega_k(t') \boldsymbol{W}^{\text{out}} e^{(\boldsymbol{W}-\boldsymbol{I})(t-t')/\tau} \, |W_k^{\text{in}}\rangle \, dt' \tag{24}$$

By substituting $s = t - t'$, we obtain an autoregressive description of the signals:

$$\omega_j(t) = \sum_{k=1}^{M} \int_{0}^{\infty} G_{jk}(s) \omega_k(t-s) ds. \tag{25}$$

Here, $G_{jk}(s) = \boldsymbol{W}_j^{\text{out}} e^{(\boldsymbol{W}-\boldsymbol{I})s/\tau} \boldsymbol{W}_k^{\text{in}}$. These kernels can be expressed in terms of the synaptic matrix spectrum. The diagonal representation of the exponential matrix is given by:

$$e^{(\boldsymbol{W}-\boldsymbol{I})s/\tau} = \sum_{n=1}^{N} e^{(\lambda_n - 1)/\tau} |Q_n\rangle\langle Q_n^{-1}| \, . \tag{26}$$

In this representation, $\lambda_n$ denotes the eigenvalues and $|Q_n\rangle\langle Q_n^{-1}|$ the outer product between the right and left eigenvectors of $\boldsymbol{W}$. This representation allows us to rewrite the kernel as the weighted sum over exponentially decaying components:

$$G_{jk}(s) = \sum_{n=1}^{N} J_{jk}^{(n)} e^{(\lambda_n - 1)s/\tau} \tag{27}$$

where the tensor of weights is defined by the set of rank-1 matrices $\boldsymbol{J}_n = \boldsymbol{W}^{\text{out}} |Q_n\rangle \langle Q_n^{-1}| \boldsymbol{W}^{\text{in}}$.

### 4.7 | lRNN and the Laplace Transform of System State

Assuming the existence of a time $t^*$ such that a matrix $\boldsymbol{W}^{\text{out}}$ maps the lRNN state into the system observable given as input $\omega(t) = \boldsymbol{W}^{\text{out}}\boldsymbol{r}(t)$, the open- and closed-loop dynamics were in principle equivalent [30, 31]. This property allowed us to link the kernel of the autoregressive model (closed-loop dynamics) to the spectrum of the learned synaptic matrix $\tilde{\boldsymbol{W}}$, and consequently, to the resulting finite Koopman approximation, as we will see in the following.

#### 4.7.1 | Open-Loop Representation

The autoregressive formula expressed in Equation (3) can be reformulated as:

$$\omega_j(t) = \sum_{k=1}^{M} \int_0^t G_{jk}(t-t')\omega_k(t')dt' + \omega_j(0) \tag{28}$$

The convolution theorem for the Laplace transform leads to:

$$\text{LT}\{\omega_j\}(\sigma) = \sum_{k=1}^{M} \text{LT}\{G_{jk}\}(\sigma)\text{LT}\{\omega_k\}(\sigma) + \frac{\omega_j(0)}{\sigma} \tag{29}$$

here, $\text{LT}\{G_{jk}\} = \int_0^\infty G_{jk}(t)e^{-\sigma t}dt$ denotes the Laplace transform of the kernel that is

$$\text{LT}\{G_{jk}\}(\sigma) = \sum_{n=1}^{N} \frac{J_{jk}^{(n)}}{\sigma - \frac{\lambda_n - 1}{\tau}} \tag{30}$$

Thus, in matrix formalism the Laplace transforms of the system observables result to be

$$\text{LT}\{\boldsymbol{\omega}\}(\sigma) = \frac{\boldsymbol{L}^{-1}(\sigma)}{\sigma}\boldsymbol{\omega}(0) \tag{31}$$

where the matrix $\boldsymbol{L}(\sigma) \in \mathbb{R}^{M\times M}$ has elements $L_{ij}(\sigma) = \delta_{ij} - \text{LT}\{G_{ij}\}(\sigma)$.

#### 4.7.2 | Closed-Loop Representation

The closed-loop network evolved according to Equation (5). Assuming a diagonalizable synaptic matrix $\tilde{\boldsymbol{W}}$, it can be factorized as $\tilde{\boldsymbol{W}} = \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{Q}}^{-1}$, where $\tilde{\boldsymbol{Q}}$ is a matrix whose columns are the right eigenvectors of $\tilde{\boldsymbol{W}}$ and $\tilde{\boldsymbol{\Lambda}}$ is a diagonal matrix containing the associated eigenvalues. The spectral decomposition of the learned synaptic matrix leads to a decoupled set of linearly evolving projections $\boldsymbol{v}(t) = \tilde{\boldsymbol{Q}}^{-1}\boldsymbol{r}(t)$ whose $N$ elements evolves independently as

$$v_n(t) = v_n(0)e^{(\tilde{\lambda}_n - 1)t/\tau} \tag{32}$$

These variables serve as a basis to decompose the system observables $\omega_k(t)$ as the sum of linearly evolving modes, resembling the Koopman decomposition. Indeed, given $\boldsymbol{r} = \tilde{\boldsymbol{Q}}\boldsymbol{v}$ we can write

$$\omega_j(t) = \sum_{n=1}^{N} \Xi_{jn}v_n(t) = \sum_{n=1}^{N} \Xi_{jn}v_n(0)e^{(\tilde{\lambda}_n - 1)t/\tau} \tag{33}$$

with matrix $\Xi \in \mathbb{R}^{M\times N}$ defined as $\Xi = \boldsymbol{W}^{\text{out}}\boldsymbol{Q}$ such that $\boldsymbol{\omega}(t) = \Xi\boldsymbol{v}(t)$. The Laplace transform of the replicated observables is then

$$\text{LT}\{\omega_j\}(\sigma) = \sum_{n=1}^{N} \frac{\Xi_{jn}v_n(0)}{\sigma - \frac{(\tilde{\lambda}_n - 1)}{\tau}} \tag{34}$$

whose poles coincide with the eigenvalues of the estimated Koopman matrix.

Given the equivalence between the open and closed representations, $L_{ij}(\sigma)/\sigma$ shares the same poles, providing a link between the autoregressive modeling and the linear dynamic representation.

### 4.8 | lRNNs and the Spectrum of the Koopman Operator

The closed-loop representation provided a decomposition of the signal as linear evolving modes. We could explicitly express the functional form of the basis starting from open loop dynamics. By applying $\tilde{\boldsymbol{Q}}^{-1}$ to both hand sides of Equation (2) yields:

$$\boldsymbol{v}(t) = \sum_{k=1}^{M} \int_0^\infty \omega_k(t-s)\tilde{\boldsymbol{Q}}^{-1}e^{(\boldsymbol{W}-\boldsymbol{I})s/\tau}\,|W_k^{\text{in}}\rangle\, ds \tag{35}$$

By resorting to the spectral decomposition in Equation (26), the above eigenmode dynamics reduces to:

$$v_j(t) = \sum_{k=1}^{M}\sum_{n=1}^{N} \left[\tilde{\boldsymbol{Q}}^{-1}\,|Q_n\rangle\,\langle Q_n^{-1}|\,\boldsymbol{W}^{\text{in}}\right]_{jk} \int_0^\infty \omega_k(t-s)e^{(\lambda_n - 1)s/\tau}ds \tag{36}$$

We could then expand in Taylor series the observables: $\omega_k(t-s) = \sum_{r=0}^{\infty} \frac{d^r\omega_k(t)}{dt^r}(-1)^r s^r/r!$. This allowed to solve the above integral leading to a functional description of the estimated eigenmodes

$$v_j(t) = \psi_j[\boldsymbol{\omega}(t), \dot{\boldsymbol{\omega}}(t), ..] = \sum_{n=1}^{N}\sum_{k=1}^{M} A_{jk}^{(n)} \sum_{r=0}^{\infty} \frac{d^r\omega_k(t)}{dt^r}\left(\frac{\tau}{\lambda_n - 1}\right)^r \tag{37}$$

The vector of all the derivatives $d^r\omega_k(t)/dt^r$ fully determine the state of the observed system [33] and $\psi_j$ is the eigenfunction of the approximated Koopman operator depending on the full state of the system. The corresponding eigenvalue $\sigma_j = (\tilde{\lambda}_j - 1)/\tau$, and $A_{jk}^{(n)}$ is a tensor of complex weights defined by the rank-1 matrices $\boldsymbol{A}_n = \frac{\tau}{1-\lambda_n}\tilde{\boldsymbol{Q}}^{-1}\,|Q_n\rangle\,\langle Q_n^{-1}|\,\boldsymbol{W}^{\text{in}}$.

### 4.9 | Stochastic lRNN

The deterministic neural network described can be extended to the stochastic case, where a white noise input stimulates the reservoir, maintaining it out of equilibrium as a continuous source of new energy. Each unit receives a total input defined by the equation:

$$h_j(t) = \sum_{k=1}^{N} W_{jk} r_k(t) + h_j^{\text{in}}(t) + \eta_j(t) \tag{38}$$

where $\eta_j(t)$ represents an Ornstein-Uhlenbeck process (colored noise) with a relatively small correlation time such that for the purpose of this work it could be considered as memory-less: $\langle \eta_i(t)\eta_j(t')\rangle = \gamma^2 \delta_{ij}\delta(t-t')$. An optimal standard deviation $\gamma$ can be identified to maximize the accuracy of the linear map $\boldsymbol{W}^{\text{out}}$.

In the closed-loop formulation, the input can be expressed as:

$$h_j(t) = \sum_{k=1}^{N} \tilde{W}_{jk} r_k(t) + \eta_j(t) \tag{39}$$

Consequently, the associated Langevin equation for the reservoir is given by:

$$\tau d\boldsymbol{r} = -\boldsymbol{r}dt + \tilde{\boldsymbol{W}}\boldsymbol{r}dt + d\boldsymbol{\eta}. \tag{40}$$

In this context, the introduced autoregressive description and linear decomposition remain valid when considering the expected values of the system.

### 4.10 | Functional Connectivity and Functional Connectivity Dynamics

Functional Connectivity (FC) is a measure of the degree of co-activation over time of different brain regions. Although FC does not allow to infer the directional flow of information among the nodes of the brain network, it has proven to provide valuable insights into the functional organization of the central nervous system [85]. Mathematically, the FC between two brain regions labeled A and B, is usually measured as the correlation function

$$\text{FC}_{AB}(t;T) = \text{cor}[X_A(t;T), X_B(t;T)] \tag{41}$$

where cor is the Pearson correlation coefficient, and $X_A(t;T)$ and $X_B(t;T)$ are the BOLD signals from region A and B, respectively, in the time window $[t, t+T]$.

Temporal changes of functional connectivity were evaluated by computing the Functional Connectivity Dynamics (FCD) [86, 87]. FCD was defined as the correlation between vectorized FC matrices evaluated in time-shifted chunks of BOLD activity:

$$\text{FCD}_{A,B}(\delta t, t; T) = \text{ucor}[\text{FC}_A(t;T), \text{FC}_B(t+\delta t;T)] \tag{42}$$

where $\delta t$ is a constant time shift, and ucor is the correlation between the vectors composed of the upper-diagonal elements of the two FC matrices involved.

Throughout the whole paper FC was computed with $t = 0$ s and $T = 387.5$ s, while FCD uses $T = 75$ s time shift $\delta t$ multiples of 12.5 s.

### 4.11 | Functional Connectivity in lRNN

Once a lRNN capable to reproduce the systems observables $\omega_j(t)$ was available, the elements of the FC matrix could be inferred directly from the lRNN parameters. The computed Functional Connectivity of the simulated signals depends on the covariance matrix estimated from time $t$ over a time window of length $T$. Considering large time windows $T$ and assuming a zero time average $\langle \omega_j(t)\rangle_t \simeq 0$, it holds:

$$\begin{aligned}\text{cov}_{ij}(t;T) &= \int_t^{t+T} \omega_i^*(t')\omega_j(t')dt' \\ &= \sum_{r,s} \Xi_{ir}^* \Xi_{js} \int_t^{t+T} \upsilon_r^*(t')\upsilon_s(t')dt' \\ &= \sum_{r,s} \Xi_{ir}^* \Xi_{js} \Omega_{rs}(t;T)\end{aligned} \tag{43}$$

Here, we are making use of the spectral decomposition in Equation (7) and we define $\Omega_{rs}(t;T)$ as the estimated covariance between the OU modes $\upsilon_r$, $\upsilon_s$. From the simulations, it results that only a few slow uncorrelated modes had a significant weight $\Xi$ in the decomposition. In this scenario, the covariance matrix was well approximated by the dot product between the representational vectors in the functional space:

$$\text{cov}_{ij}(t;T) \approx \boldsymbol{\Xi}_i \cdot \boldsymbol{\Xi}_j \tag{44}$$

Consequently for the correlation matrix it holds

$$\text{FC}_{ij}(t;T) \approx \text{S}_\text{C}\left(\boldsymbol{\Xi}_i, \boldsymbol{\Xi}_j\right) \tag{45}$$

where $\text{S}_\text{C}$ indicates the cosine similarity function, i.e., the Pearson correlation between row and column vectors composing $\boldsymbol{\Xi}^*$ and $\boldsymbol{\Xi}$, respectively, under the assumption of a zero mean for the system observables.

### 4.12 | Simulation Parameters

Network parameters had been selected through grid search methods to optimize prediction accuracy. The reservoir state was initialized in a zero-activity state and was disrupted from equilibrium by the input for a transient period of 20 $s$ before learning. This period corresponds to the $t^*$ mentioned in the preceding paragraphs. The input signal was linearly interpolated, and the system's dynamics were resolved using the Euler–Maruyama method with a time step that was one-hundredth of the sampling interval. The spectral radius $\rho$ of the synaptic matrix, used for fMRI analysis, was set to 0.5, and the time constant $\tau$ was 1.0. The standard deviation $\text{g}^{\text{in}}$ of the input matrix $\boldsymbol{W}^{\text{in}}$ was set to 1. The thermal noise shown in Figure 4 had a regularizing effect. Consequently, we did not employ further regularization for the readout regression. For the example system depicted in Figure 1, the parameters were $\rho = 0.5$, $\tau = 0.1$ with no regularizers, and

$g^{in}$ equal to the inverse of the square root of the number of inputs. The stochastic scenario was characterized by thermal noise with a standard deviation of $10^{-5}$ and ridge regression with a regularization of $10^{-7}$. For cases with missing variables, the parameters were set to $\rho = 0.6$, $\tau = 0.25$.

Hankel DMD analyzes in Figure 3 were all performed choosing a truncation rank $r$ preserving 0.99999 of the energy in the computed SVD and memory length $q = 5$.

### 4.13 | Clustering Analysis

The clustering analysis was performed on the resulting spectrum of the data by discretizing the complex space into a grid with a real resolution of $0.00025s^{-1}$, and an imaginary resolution of $0.01s^{-1}$.

To further reduce the dimensionality of the data and get an interpretable 3D representation, Linear Discriminant Analysis (LDA) was employed. LDA was a widely used technique in machine learning [80]. It was a supervized learning method that aims to reduce the dimensionality of the data while preserving the discriminatory information between different classes. The core idea was to find a linear transformation that maximizes the ratio of the between-class scatter $\boldsymbol{S}_b$ to the within-class scatter $\boldsymbol{S}_w$. This optimization process equals to solve the following generalized eigenvalue problem

$$\boldsymbol{S}_b \boldsymbol{l} = \lambda \boldsymbol{S}_w \boldsymbol{l} \tag{46}$$

The optimal projection matrix $\boldsymbol{L}$ to a subspace of dimension $k$ was given by the eigenvectors $\boldsymbol{l}$ corresponding to the largest $k$ eigenvalues.

### 4.14 | Linear Decoders

To identify frequency patterns that discriminate brain areas in the discretized frequency-vector representation, we trained a multi-class linear decoder that maps each frequency vector to a one-hot target indicating the corresponding area. Let $M = 11$ denote the number of areas and $D$ the dimensionality of the discretized frequency representation. Pooling all subjects and all 100 reservoir realizations, we arranged the data into the feature matrix

$$\boldsymbol{F} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_S] \in \mathbb{R}^{D \times S}$$

where each column $\boldsymbol{f}_n \in \mathbb{R}^D$ is one sample and $S$ is the total number of samples (i.e., $20 \times 100 = 2,000$). We encoded area labels in the one-hot matrix

$$\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_S] \in \{0,1\}^{M \times S}$$

where $\boldsymbol{y}_n \in \{0,1\}^M$ has a single 1 at index corresponding the area of sample $n$, and zeros elsewhere. We estimated the decoding matrix $\boldsymbol{W} \in \mathbb{R}^{M \times D}$ by solving the matrix least-squares problem

$$\boldsymbol{W} = \arg\min_{\boldsymbol{W}_0} \|\boldsymbol{Y} - \boldsymbol{W}_0 \boldsymbol{F}\|_F^2$$

The minimum-norm solution is given by $\boldsymbol{W} = \boldsymbol{Y}\boldsymbol{F}^+$, where $\boldsymbol{F}^+$ denotes the Moore-Penrose pseudoinverse. The $k$-th row $\boldsymbol{W}_{k:}$ defines the linear decoder for area $k$. The output of the area-$k$ decoder for a single sample $\boldsymbol{f}_n$ is the scalar

$$\hat{y}_k(n) = \sum_{j=1}^{D} W_{kj} F_{jn}$$

Per-area accuracy for area $k$ was defined as the fraction of samples with true label $k$ for which $\hat{y}_k(n)$ attains the maximum across areas.

### 4.15 | Manuscript Language Optimization

To enhance the clarity, coherence and readability of the manuscript, we employed large language models (LLMs) for language editing and refinement. All edits assisted by the model were subsequently reviewed by us to ensure accuracy and to preserve the intended scientific meaning.

### 4.16 | Code Availability

All codes are available at https://github.com/gdianto/LinearRNNforFMRIdigitaltwins


#### Acknowledgements

The authors thank M. Allegra for the feedback on an earlier version of the manuscript and L. Bertini, I. Branchi, S. P. Caminiti, and G. V. Vinci for the stimulating discussions. Work partially funded by the NEXTGENERATIONEU and MUR (PNRR-M4C2I1.3) project MNESYS (PE0000006–DD 1553 11.10.2022) and project EBRAINS-Italy (IR0000011–DD 101 16.6.2022) to MM.


#### Conflicts Of Interest

The authors declare no conflicts of interest.

#### Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

#### References


1. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).

2. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016), http://www.deeplearningbook.org.

3. B. Mehlig, *Machine Learning With Neural Networks: An Introduction for Scientists and Engineers* (Cambridge University Press, 2021).

4. H. Jaeger and H. Haas, "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication," *Science* 304, no. 5667 (2004): 78–80.

5. M. Lukoševičius and H. Jaeger, "Reservoir Computing Approaches to Recurrent Neural Network Training," *Computer Science Review* 3, no. 3 (2009): 127–149.

6. J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-Free Prediction of Large Spatiotemporally Chaotic Systems From Data: A Reservoir Computing Approach," *Physical Review Letters* 120, no. 2 (2018): 024102.

7. P.-R. Vlachas, J. Pathak, B. R. Hunt, et al., "Backpropagation Algorithms and Reservoir Computing in Recurrent Neural Networks for the Forecasting of Complex Spatiotemporal Dynamics," *Neural Networks* 126 (2020): 191–217.

8. K. Srinivasan, N. Coble, J. Hamlin, T. Antonsen, E. Ott, and M. Girvan, "Parallel Machine Learning for Forecasting the Dynamics of Complex Networks," *Physical Review Letters* 128, no. 16 (2022): 164101, https://doi.org/10.1103/PhysRevLett.128.164101.

9. H. B. Barlow, "Possible Principles Underlying the Transformation of Sensory Messages," in *Sensory Communication*, W. A. Rosenblith, Ed. (MIT Press, 1961), 217–234.

10. E. P. Simoncelli and B. A. Olshausen, "Natural Image Statistics and Neural Representation," *Annual Review of Neuroscience* 24, no. 1 (2001): 1193–1216, https://doi.org/10.1146/annurev.neuro.24.1.1193.

11. S. Fusi, E. K. Miller, and M. Rigotti, "Why Neurons Mix: High Dimensionality for Higher Cognition," *Current Opinion in Neurobiology* 37 (2016): 66–74, https://doi.org/10.1016/j.conb.2016.01.010.

12. V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome, "Context-Dependent Computation by Recurrent Dynamics in Prefrontal Cortex," *Nature* 503, no. 7474 (2013): 78–84.

13. D. Sussillo and O. Barak, "Opening the Black Box: Low-Dimensional Dynamics in High-Dimensional Recurrent Neural Networks," *Neural Computation* 25 (2013): 626–649, https://doi.org/10.1162/NECO_a_00409.

14. D. Durstewitz, G. Koppe, and M. I. Thurm, "Reconstructing Computational System Dynamics From Neural Data With Recurrent Neural Networks," *Nature Reviews Neuroscience* 24, no. 11 (2023): 693–710.

15. G. Di Antonio, S. Raglio, and M. Mattia, "A Geometrical Solution Underlies General Neural Principle for Serial Ordering," *Nature Communications* 15, no. 1 (2024): 8238.

16. C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable Machine Learning—A Brief History, State-of-the-Art and Challenges," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Springer, 2020),417–431.

17. W. Gilpin, "Generative Learning for Nonlinear Dynamics," *Nature Reviews Physics* 6, no. 3 (2024): 194–206, https://doi.org/10.1038/s42254-024-00688-2.

18. H. Turbé, M. Bjelogrlic, C. Lovis, and G. Mengaldo, "Evaluation of Post-Hoc Interpretability Methods in Time-Series Classification," *Nature Machine Intelligence* 5, no. 3 (2023): 250–260, https://doi.org/10.1038/s42256-023-00620-w.

19. A. Mauroy, Y. Susuki, and I. Mezić, *Koopman Operator in Systems and Control* (Springer, 2020).

20. S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, "Modern Koopman Theory for Dynamical Systems," *SIAM Review* 64, no. 2 (2022): 229–340, https://doi.org/10.1137/21M1401243.

21. P. J. Schmid, "Dynamic Mode Decomposition and Its Variants," *Annual Review of Fluid Mechanics* 54, no. 1 (2022): 225–254, https://doi.org/10.1146/annurev-uid-030121-015835.

22. G. Di Antonio and G. V. Vinci, "Nonlinear Correlations Underlie Linear Response and Causality," *Physical Review Research* 7 (2025): L042029, https://doi.org/10.1103/vjvt-gdn1.

23. F. Tao, B. Xiao, Q. Qi, J. Cheng, and P. Ji, "Digital Twin Modeling," *Journal of Manufacturing Systems* 64 (2022): 372–389, https://doi.org/10.1016/j.jmsy.2022.06.015.

24. S. A. Niederer, M. S. Sacks, M. Girolami, and K. Willcox, "Scaling Digital Twins From the Artisanal to the Industrial," *Nature Computational Science* 1, no. 5 (2021): 313–320, https://doi.org/10.1038/s43588-021-00072-5.

25. M. Schirner, S. Rothmeier, V. K. Jirsa, A. R. McIntosh, and P. Ritter, "An Automated Pipeline for Constructing Personalized Virtual Brains From Multimodal Neuroimaging Data," *NeuroImage* 117 (2015): 343–357.

26. V. K. Jirsa, T. Proix, D. Perdikis, et al., "The Virtual Epileptic Patient: Individualized Whole-Brain Models of Epilepsy Spread," *NeuroImage* 145 (2017): 377–388, https://doi.org/10.1016/j.neuroimage.2016.04.049.

27. V. Jirsa, H. Wang, P. Triebkorn, et al., "Personalised Virtual Brain Models in Epilepsy," *The Lancet Neurology* 22, no. 5 (2023): 443–454, https://doi.org/10.1016/S1474-4422(23)00008-X.

28. L. S. Fekonja, R. Schenk, E. Schröder, R. Tomasello, S. Tomšič, and T. Picht, "The Digital Twin in Neuroscience: From Theory to Tailored Therapy," *Frontiers in Neuroscience* 18 (2024): 1–8, https://doi.org/10.3389/fnins.2024.1454856.

29. L. Kocarev and U. Parlitz, "Generalized Synchronization, Predictability, and Equivalence of Unidirectionally Coupled Dynamical Systems," *Physical Review Letters* 76, no. 11 (1996): 1816–1819, https://doi.org/10.1103/PhysRevLett.76.1816.

30. Z. Lu, B. R. Hunt, and E. Ott, "Attractor Reconstruction by Machine Learning," *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28, no. 6 (2018), https://doi.org/10.1063/1.5039508.

31. H. Jaeger, "The "Echo State" Approach to Analysing and Training Recurrent Neural Networks," German National Research Center for Information Technology, Technical Report, 2001.

32. I. B. Yildiz, H. Jaeger, and S. J. Kiebel, "Revisiting the Echo State Property," *Neural Networks* 35 (2012): 1–9, https://doi.org/10.1016/j.neunet.2012.07.005.

33. F. Takens, "Detecting Strange Attractors in Turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L.-S. Young, Eds. (Springer Berlin Heidelberg, 1981), 366–381.

34. T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *Journal of Statistical Physics* 65, no. 3–4 (1991): 579–616, https://doi.org/10.1007/BF01053745.

35. J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke, "Delay Embeddings for Forced Systems. II. Stochastic Forcing," *Journal of Nonlinear Science* 13 (2003): 519–577, https://doi.org/10.1007/s00332-003-0534-4.

36. E. M. Bollt, "On Explaining the Surprising Success of Reservoir Computing Forecaster of Chaos: The Universal Machine Learning Dynamical System With Contrast to VAR and DMD," *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31, no. 1 (2021): 013108.

37. D. J. Gauthier, E. M. Bollt, A. Griffith, and W. A. Barbosa, "Next Generation Reservoir Computing," *Nature Communications* 12, no. 1 (2021): 1–8.

38. P. Verzelli, C. Alippi, L. Livi, and P. Tiňo, "Input-to-State Representation in Linear Reservoir Dynamics," *IEEE Transactions on Neural Networks and Learning Systems* 33, no. 9 (2021): 4598–4609.

39. G. P. Krishnan, O. C. González, and M. Bazhenov, "Origin of Slow Spontaneous Resting-State Neuronal Fluctuations in Brain Networks," *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 26 (2018): 6858–6863, https://doi.org/10.1073/pnas.1715841115.

40. P. J. Drew, C. Mateo, K. L. Turner, X. Yu, and D. Kleinfeld, "Ultra-Slow Oscillations in fMRI and Resting-State Connectivity: Neuronal and Vascular Contributions and Technical Confounds," *Neuron* 107, no. 5 (2020): 782–804, https://doi.org/10.1016/j.neuron.2020.07.020.

41. H. Arbabi and I. Mezic, "Ergodic Theory, Dynamic Mode Decomposition, and Computation of Spectral Properties of the Koopman Operator," *SIAM Journal on Applied Dynamical Systems* 16, no. 4 (2017): 2096–2126.

42. S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, "Chaos as an Intermittently Forced Linear System," *Nature Communications* 8, no. 1 (2017): 19.

43. Z.-M. Zhai, L.-W. Kong, and Y.-C. Lai, "Emergence of a Resonance in Machine Learning," *Physical Review Research* 5, no. 3 (2023): 033127.

44. E. J. W. Van Someren, Y. D. Van Der Werf, P. R. Roelfsema, H. D. Mansvelder, and F. H. Lopes da Silva, "Slow Brain Oscillations of Sleep, Resting State, and Vigilance," *Progress in Brain Research* 193 (2011): 3–15, https://doi.org/10.1016/B978-0-444-53839-0.00001-6.

45. S. Lecci, L. M. J. Fernandez, F. D. Weber, et al., "Coordinated Infraslow Neural and Cardiac Oscillations Mark Fragility and Offline Periods in Mammalian Sleep," *Science Advances* 3, no. 2 (2017), https://doi.org/10.1126/sciadv.1602026.

46. D. Gutierrez-Barragan, M. A. Basson, S. Panzeri, and A. Gozzi, "Infraslow State Fluctuations Govern Spontaneous fMRI Network Dynamics," *Current Biology* 29, no. 14 (2019): 2295–2306.e5, https://doi.org/10.1016/j.cub.2019.06.017.

47. N. Tort-Colet, C. Capone, M. V. Sanchez-Vives, and M. Mattia, "Attractor Competition Enriches Cortical Dynamics During Awakening From Anesthesia," *Cell Reports* 35, no. 12 (2021): 109270, https://doi.org/10.1016/j.celrep.2021.109270.

48. K. A. Smitha, K. M. Arun, P. G. Rajesh, B. Thomas, and C. Kesavadas, "Resting-State Seed-Based Analysis: An Alternative to Task-Based Language fMRI and Its Laterality Index," *American Journal of Neuroradiology* 38, no. 6 (2017): 1187–1192, https://doi.org/10.3174/ajnr.A5169.

49. P. Branco, D. Seixas, and S. L. Castro, "Mapping Language With Resting-State Functional Magnetic Resonance Imaging: A Study on the Functional Profile of the Language Network," *Human Brain Mapping* 41, no. 2 (2020): 545–560, https://doi.org/10.1002/hbm.24821.

50. L. Labache, T. Ge, B. T. T. Yeo, and A. J. Holmes, "Language Network Lateralization Is Reflected Throughout the Macroscale Functional Organization of Cortex," *Nature Communications* 14, no. 1 (2023): 3405, https://doi.org/10.1038/s41467-023-39131-y.

51. E. Nozari, M. A. Bertolero, J. Stiso, et al., "Macroscopic Resting-State Brain Dynamics Are Best Described by Linear Models," *Nature Biomedical Engineering* 8, no. 1 (2024): 68–84, https://doi.org/10.1038/s41551-023-01117-y.

52. H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, 2003).

53. Y. Lu, J. Liu, J. Ji, H. Lv, and M. Huai, "Brain Effective Connectivity Learning With Deep Reinforcement Learning," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2022): 1664–1667.

54. Z. Zhang, J. Ji, and J. Liu, "Metarlec: Meta-Reinforcement Learning for Discovery of Brain Effective Connectivity," *Proceedings of the AAAI Conference on Artificial Intelligence* 38, no. 9 (2024): 10 261–10 269.

55. Q. Yan, H. Ouyang, Z. Tao, et al., "Multi-Wavelength Optical Information Processing With Deep Reinforcement Learning," *Light: Science & Applications* 14, no. 1 (2025): 160, https://doi.org/10.1038/s41377-025-01846-6.

56. R. Zhang, L. Aksman, D. Wijesinghe, J. M. Ringman, D. J. J. Wang, and K. Jann, "A Longitudinal Study of Functional Brain Complexity in Progressive Alzheimer's Disease," *Alzheimer's & Dementia* 17, no. 1 (2025): 1–11, https://doi.org/10.1002/dad2.70059.

57. D. C. Van Essen, S. M. Smith, D. M. Barch, et al., "The WU-Minn Human Connectome Project: An Overview," *NeuroImage* 80 (2013): 62–79, https://doi.org/10.1016/j.neuroimage.2013.05.041.

58. M. P. Harms, L. H. Somerville, B. M. Ances, et al., "Extending the Human Connectome Project Across Ages: Imaging Protocols for the Lifespan Development and Aging Projects," *NeuroImage* 183 (2018): 972–984, https://doi.org/10.1016/j.neuroimage.2018.09.060.

59. S. Y. Bookheimer, D. H. Salat, M. Terpstra, et al., "The Lifespan Human Connectome Project in Aging: An Overview," *NeuroImage* 185 (2019): 335–348, https://doi.org/10.1016/j.neuroimage.2018.10.009.

60. Y. Sanz Perl, S. Fittipaldi, C. Gonzalez Campo, et al., "Model-Based Whole-Brain Perturbational Landscape of Neurodegenerative Diseases," *eLife* 12 (2023): 1–25, https://doi.org/10.7554/eLife.83970.

61. L. L. Oganesian and M. M. Shanechi, "Brain–Computer Interfaces for Neuropsychiatric Disorders," *Nature Reviews Bioengineering* 2, no. 8 (2024): 653–670.

62. L.-W. Kong, G. A. Brewer, and Y.-C. Lai, "Reservoir-Computing-Based Associative Memory and Itinerancy for Complex Dynamical Attractors," *Nature Communications* 15, no. 1 (2024): 4840, https://doi.org/10.1038/s41467-024-49190-4.

63. H. Sanhedrai, J. Gao, A. Bashan, M. Schwartz, S. Havlin, and B. Barzel, "Reviving a Failed Network Through Microscopic Interventions," *Nature Physics* 18, no. 3 (2022): 338–349, https://doi.org/10.1038/s41567-021-01474-y.

64. G. Deco and V. K. Jirsa, "Ongoing Cortical Activity at Rest: Criticality, Multistability, and Ghost Attractors," *Journal of Neuroscience* 32, no. 10 (2012): 3366–3375, https://doi.org/10.1523/JNEUROSCI.2523-11.2012.

65. M. V. Sanchez-Vives, M. Massimini, and M. Mattia, "Shaping the Default Activity Pattern of the Cortical Network," *Neuron* 94, no. 5 (2017): 993–1001, https://doi.org/10.1016/j.neuron.2017.05.015.

66. S. di Santo, P. Villegas, R. Burioni, and M. A. Muñoz, "Landau-Ginzburg Theory of Cortex Dynamics: Scale-Free Avalanches Emerge at the Edge of Synchronization," *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 7 (2018): E1356–E1365, https://doi.org/10.1073/pnas.1712989115.

67. C. Song, M. Boly, E. Tagliazucchi, H. Laufs, and G. Tononi, "fMRI Spectral Signatures of Sleep," *Proceedings of the National Academy of Sciences of the United States of America* 119, no. 30 (2022): 1–12, https://doi.org/10.1073/pnas.2016732119.

68. A. J. Eisen, L. Kozachkov, A. M. Bastos, et al., "Propofol Anesthesia Destabilizes Neural Dynamics Across Cortex," *Neuron* 112, no. 16 (2024): 2799–2813.e9, https://doi.org/10.1016/j.neuron.2024.06.011.

69. G. Koppe, H. Toutounji, P. Kirsch, S. Lis, and D. Durstewitz, "Identifying Nonlinear Dynamical Systems via Generative Recurrent Neural Networks With Applications to fMRI," *PLOS Computational Biology* 15, no. 8 (2019): e1007263, https://doi.org/10.1371/journal.pcbi.1007263.

70. M. Gilson, E. Tagliazucchi, and R. Cofré, "Entropy Production of Multivariate Ornstein-Uhlenbeck Processes Correlates With Consciousness Levels in the Human Brain," *Physical Review E* 107, no. 2 (2023): 024121, https://doi.org/10.1103/PhysRevE.107.024121.

71. V. Sip, M. Hashemi, T. Dickscheid, K. Amunts, S. Petkoski, and V. Jirsa, "Characterization of Regional Differences in Resting-State fMRI With a Data-Driven Network Model of Brain Dynamics," *Science Advances* 9, no. 11 (2023): eabq7547, https://doi.org/10.1126/sciadv.abq7547.

72. J. D. Touboul and A. Destexhe, "Can Power-Law Scaling and Neuronal Avalanches Arise From Stochastic Dynamics?" *PLOS ONE* 5, no. 2 (2010): e8982, https://doi.org/10.1371/journal.pone.0008982.

73. A. Santoro, F. Battiston, G. Petri, and E. Amico, "Higher-Order Organization of Multivariate Time Series," *Nature Physics* 19, no. 2 (2023): 221–229, https://doi.org/10.1038/s41567-022-01852-0.

74. G. Deco and M. L. Kringelbach, "Turbulent-Like Dynamics in the Human Brain," *Cell Reports* 33, no. 10 (2020): 108471, https://doi.org/10.1016/j.celrep.2020.108471.

75. M. Khona and I. R. Fiete, "Attractor and Integrator Networks in the Brain," *Nature Reviews Neuroscience* 23 (2022): 744–766, https://doi.org/10.1038/s41583-022-00642-0.

76. B. A. W. Brinkman, H. Yan, A. Maffei, et al., "Metastable Dynamics of Neural Circuits and Networks," *Applied Physics Reviews* 9, no. 1 (2022): 011313, https://doi.org/10.1063/5.0062603.

77. M. Brenner, C. J. Hemmer, Z. Monfared, and D. Durstewitz, "Almost-Linear RNNs Yield Highly Interpretable Symbolic Codes in Dynamical Systems Reconstruction,"*Advances in Neural Information Processing Systems 37 (2024): 36829-36868*.

78. J. Soldado-Magraner, V. Mante, and M. Sahani, "Inferring Context-Dependent Computations Through Linear Approximations of Prefrontal

Cortex Dynamics," *Science Advances* 10, no. 51 (2024): 1–22, https://doi.org/10.1126/sciadv.adl4743.

79. F. Hancock, F. E. Rosas, A. I. Luppi, et al., "Metastability Demystified–The Foundational Past, the Pragmatic Present and the Promising Future," *Nature Reviews Neuroscience* 26, no. 2 (2025): 82–100, https://doi.org/10.1038/s41583-024-00883-1.

80. R. O. Duda, P. E. Hart, and D. D. Stork, *Pattern Classification*, 2nd ed. (John Wiley & Sons, 2001).

81. B. O. Koopman, "Hamiltonian Systems and Transformation in Hilbert Space," *Proceedings of the National Academy of Sciences of the United States of America* 17, no. 5 (1931): 315–318.

82. A. Treves, "Mean-Field Analysis of Neuronal Spike Dynamics," *Network: Computation in Neural Systems* 4, no. 3 (1993): 259–284.

83. B. W. Knight, "Dynamics of Encoding in Neuron Populations: Some General Mathematical Features," *Neural Computation* 12, no. 3 (2000): 473–518.

84. M. Mattia and P. Del Giudice, "Population Dynamics of Interacting Spiking Neurons," *Physical Review E* 66, no. 5 Pt 1 (2002): 051917.

85. E. T. Bullmore and O. Sporns, "The Economy of Brain Network Organization," *Nature Reviews Neuroscience* 13, no. 5 (2012): 336–349.

86. E. C. A. Hansen, D. Battaglia, A. Spiegler, G. Deco, and V. K. Jirsa, "Functional Connectivity Dynamics: Modeling the Switching Behavior of the Resting State," *NeuroImage* 105 (2015): 525–535.

87. G. Deco, M. L. Kringelbach, V. K. Jirsa, and P. Ritter, "The Dynamics of Resting Fluctuations in the Brain: Metastability and Its Dynamical Cortical Core," *Scientific Reports* 7, no. 1 (2017): 1–14.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.

**Supporting File:** advs74692-sup-0001-SuppMat.pdf.