

Supplementary Materials

This appendix is divided into four sections. The first section is supplementary content for the verification strategy design. The second section is the verification of model stability based on Lyapunov's theorem. The third section is the verification of model stability based on Input-State Stability using Lyapunov's theorem. The fourth section is the analysis of the dynamic properties of the model.

Section S1: Supplementary Verification of Strategy Design

This work adopts a stratified retention validation strategy. The original dataset (N samples), subject-level division (M independent subjects). Among them, the training set $\mathcal{P}_{\text{train}}$ (80%), the validation set $\mathcal{P}_{\text{val}}$ (10%), and the test set $\mathcal{P}_{\text{test}}$ (10%). The representation of the data distribution is as follows, let f represent the set of all subjects, and the division process is

$$\begin{aligned}\mathcal{P}_{\text{train}}, \mathcal{P}_{\text{temp}} &\sim \text{StratifiedSplit}(\mathcal{P}, \text{test_size}=0.2, \text{random_state}=\text{SEED}) \\ \mathcal{P}_{\text{val}}, \mathcal{P}_{\text{test}} &\sim \text{StratifiedSplit}(\mathcal{P}_{\text{temp}}, \text{test_size}=0.5, \text{random_state}=\text{SEED})\end{aligned}$$

Among them, the time series gap constraint conditions need to be satisfied, $\forall i, j \in \{\mathcal{P}_{\text{train}}, \mathcal{P}_{\text{test}}, \mathcal{P}_{\text{val}}\}, |t_i - t_j| \geq g = 10$. The mutual exclusivity guarantee of the data set,

$$\mathcal{P}_{\text{train}} \cap \mathcal{P}_{\text{val}} = \emptyset, \quad \mathcal{P}_{\text{train}} \cap \mathcal{P}_{\text{test}} = \emptyset, \quad \mathcal{P}_{\text{val}} \cap \mathcal{P}_{\text{test}} = \emptyset.$$

Section S2: Proof of Stability

In this study, SleepXLSTM is considered as a discrete-time dynamical system. The state vector $\mathbf{s}_t = [\mathbf{c}_t; \mathbf{h}_t] \in \mathbb{R}^{2lh}$ is defined. The state concatenation of all layers is $\mathbf{c}_t = [\mathbf{c}_t^1, \dots, \mathbf{c}_t^l]$, and the hidden state concatenation of all layers is $\mathbf{h}_t = [\mathbf{h}_t^1, \dots, \mathbf{h}_t^l]$. where, l is the number of model layers and h is the hidden layer size. The system expression is

$$\mathbf{s}_t = F(\mathbf{s}_{t-1}, \mathbf{x}_t)$$

where F represents a composite function of SleepXLSTM update and attention mechanisms.

This study constructs Lyapunov functions as where the function $V(\mathbf{s}_t)$ is positive definite, i.e., $V(\mathbf{s}_t) \geq 0$ and $V(\mathbf{s}_t) = 0$ if and only if $\mathbf{s}_t = \mathbf{0}$ holds the difference for calculating Lyapunov function is

$$\Delta V(\mathbf{s}_t) = V(\mathbf{s}_t) - V(\mathbf{s}_{t-1}) = \sum_{l=1}^L \left(\|\mathbf{c}_t^l\|_2^2 - \|\mathbf{c}_{t-1}^l\|_2^2 + \|\mathbf{h}_t^l\|_2^2 - \|\mathbf{h}_{t-1}^l\|_2^2 \right)$$

For each layer, the cell state update expression is

$$\mathbf{c}_t^l = \mathbf{f}_t^l \odot \mathbf{c}_{t-1}^l + \mathbf{i}_t^l \odot \tilde{\mathbf{c}}_t^l$$

The parameters of the above formula conform to the definition of SleepXLSTM model. computing squared norm

$$\| \mathbf{c}_t^l \|_2^2 = \| \mathbf{f}_t^l \odot \mathbf{c}_{t-1}^l \|_2^2 + \| \mathbf{i}_t^l \odot \tilde{\mathbf{c}}_t^l \|_2^2 + 2 \langle \mathbf{f}_t^l \odot \mathbf{c}_{t-1}^l, \mathbf{i}_t^l \odot \tilde{\mathbf{c}}_t^l \rangle$$

Due to $\| \mathbf{f}_t^l \odot \mathbf{c}_{t-1}^l \|_2^2 \leq \| \mathbf{c}_{t-1}^l \|_2^2$ and $\| \mathbf{i}_t^l \odot \tilde{\mathbf{c}}_t^l \|_2^2 \leq h$, inferred from Cauchy-Schwarz inequality

$$\langle \mathbf{f}_t^l \odot \mathbf{c}_{t-1}^l, \mathbf{i}_t^l \odot \tilde{\mathbf{c}}_t^l \rangle \leq \| \mathbf{c}_{t-1}^l \|_2 \cdot \sqrt{h}$$

Therefore, it is possible to obtain

$$\| \mathbf{c}_t^l \|_2^2 \leq \| \mathbf{c}_{t-1}^l \|_2^2 + h + 2\sqrt{h} \| \mathbf{c}_{t-1}^l \|_2$$

Therefore

$$\| \mathbf{c}_t^l \|_2^2 - \| \mathbf{c}_{t-1}^l \|_2^2 \leq h + 2\sqrt{h} \| \mathbf{c}_{t-1}^l \|_2$$

Then the residual connection $\mathbf{h}_t^{(l)}$ is formed according to the hidden state $\tilde{\mathbf{v}}$ and the value

$\mathbf{h}_t^{(l)'} \in \mathbb{R}^{B \times T \times h}$ of the connection

$$\mathbf{h}_t^{(l)'} \leftarrow \underbrace{\mathbf{h}_t^{(l)}}_{B \times T \times h} + \underbrace{\tilde{\mathbf{v}}}_{B \times T \times h}$$

According to the attention mechanism available

$$\| \mathbf{c}_t^l \|_2 \leq M_w \sqrt{h}$$

Hence

$$\| \mathbf{h}_t^l \|_2 \leq \| \mathbf{h}_t^{l, Enhanced-XLSTM} \|_2 + \| \mathbf{c}_t^l \|_2 \leq \sqrt{h} + M_w \sqrt{h} = (1 + M_w) \sqrt{h}$$

Therefore

$$\| \mathbf{h}_t^l \|_2^2 \leq (1 + M_w)^2 h$$

Similarly, for the previous time step, we obtain

$$\| \mathbf{h}_t^l \|_2^2 - \| \mathbf{h}_{t-1}^l \|_2^2 \leq (1 + M_w)^2 h - \| \mathbf{h}_{t-1}^l \|_2^2$$

Substituting the total difference, we get

$$\Delta V(\mathbf{s}_t) \leq \sum_{l=1}^L \left(h + 2\sqrt{h} \| \mathbf{c}_{t-1}^l \|_2 + (1 + M_w)^2 h - \| \mathbf{h}_{t-1}^l \|_2^2 \right)$$

And then $v_{t-1} = \sqrt{V(\mathbf{s}_{t-1})}$, then $\sum_{l=1}^L \| \mathbf{c}_{t-1}^l \|_2 \leq \sqrt{L} v_{t-1}$, and $\sum_{l=1}^L \| \mathbf{h}_{t-1}^l \|_2^2 \geq 0$, so you can get

$$\Delta V(\mathbf{s}_t) \leq Lh + 2\sqrt{hL} v_{t-1} + L(1 + M_w)^2 h$$

Thus

$$V(\mathbf{s}_t) = V(\mathbf{s}_{t-1}) + \Delta V(\mathbf{s}_t) \leq v_{t-1}^2 + Lh[1 + (1 + M_w)^2] + 2\sqrt{hL} v_{t-1}$$

order $C = Lh[1 + (1 + M_w)^2]$, then

$$V(\mathbf{s}_t) \leq (v_{t-1} + \sqrt{hL})^2 + C$$

take the square root

$$v_t = \sqrt{V(\mathbf{s}_t)} \leq v_{t-1} + \sqrt{hL} + \sqrt{C}$$

by iteratively

$$v_t \leq v_0 + t(\sqrt{hL} + \sqrt{C})$$

of which, $v_0 = \sqrt{V(\mathbf{s}_0)} = 0$. Thus, for finite time steps $t \leq T$, the state norm is bounded.

Section S3: Stability proof based on input-output

This study assumes that the forgetting gate has an upper bound, i.e., there is $0 < \alpha < 1$, such that for all l and t there is $\|\mathbf{f}_t^l\|_\infty \leq \alpha$, which can be achieved by initializing the forgetting gate bias.

Cell status update expression is

$$\|\mathbf{c}_t^l\|_2 \leq \alpha \|\mathbf{c}_{t-1}^l\|_2 + \|\mathbf{i}_t^l \odot \tilde{\mathbf{c}}_t^l\|_2$$

Since $\|\mathbf{i}_t^l \odot \tilde{\mathbf{c}}_t^l\|_2 \leq \|\tilde{\mathbf{c}}_t^l\|_2$, and $\tilde{\mathbf{c}}_t^l = \tanh(\mathbf{W}_c^l[\mathbf{h}_{t-1}^l; \mathbf{x}_t^l] + \mathbf{b}_c^l)$, there is $\|\tilde{\mathbf{c}}_t^l\|_2 \leq K_1 + K_2 \|\mathbf{x}_t\|_2$, for some K_1 and $K_2 > 0$ (because the parameters are bounded). Hence

$$\|\mathbf{c}_t^l\|_2 \leq \alpha \|\mathbf{c}_{t-1}^l\|_2 + K_1 + K_2 \|\mathbf{x}_t\|_2$$

Recursive solution yields

$$\|\mathbf{c}_t^l\|_2 \leq \alpha^t \|\mathbf{c}_0^l\|_2 + (K_1 + K_2 \sup_{0 \leq \tau \leq t} \|\mathbf{x}_\tau\|_2) \sum_{k=0}^{t-1} \alpha^k \leq \alpha^t \|\mathbf{c}_0^l\|_2 + \frac{K_1 + K_2 \sup_{0 \leq \tau \leq t} \|\mathbf{x}_\tau\|_2}{1 - \alpha}$$

For hidden states, there are

$$\|\mathbf{h}_t^l\|_2 \leq (1 + M_w) \sqrt{h}$$

Therefore, for the whole state

$$\|\mathbf{s}_t\|_2 \leq C \alpha^t \|\mathbf{s}_0\|_2 + \frac{D}{1 - \alpha} (1 + \sup_{0 \leq \tau \leq t} \|\mathbf{x}_\tau\|_2)$$

where $C, D > 0$ is constant.

This satisfies the definition of Input-State Stability, i.e., there exists a $\mathcal{KL}$ function, $\beta(r, t) = C \alpha^t r$ and $\mathcal{K}$ functions, $\gamma(r) = \frac{D}{1 - \alpha} (1 + r)$, such that

$$\|\mathbf{s}_t\|_2 \leq \beta(\|\mathbf{s}_0\|_2, t) + \gamma(\sup_{0 \leq \tau \leq t} \|\mathbf{x}_\tau\|_2)$$

Section S4: Analysis of dynamic properties

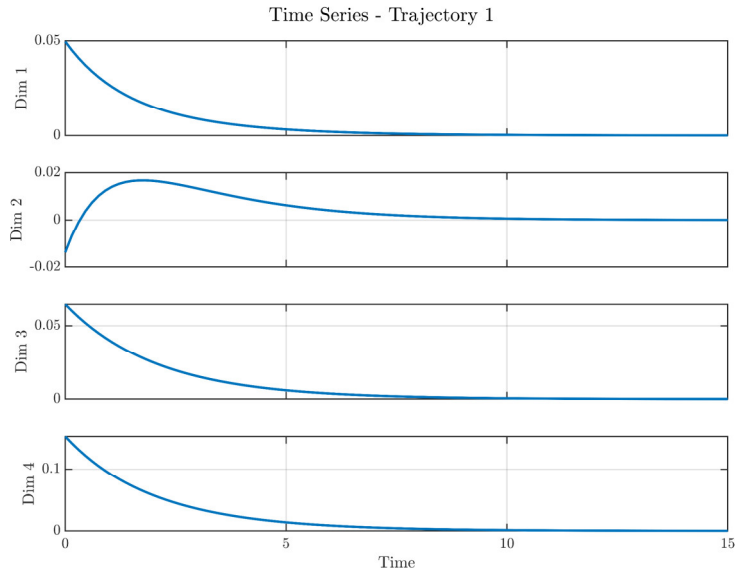


Figure S1. Time Series 1

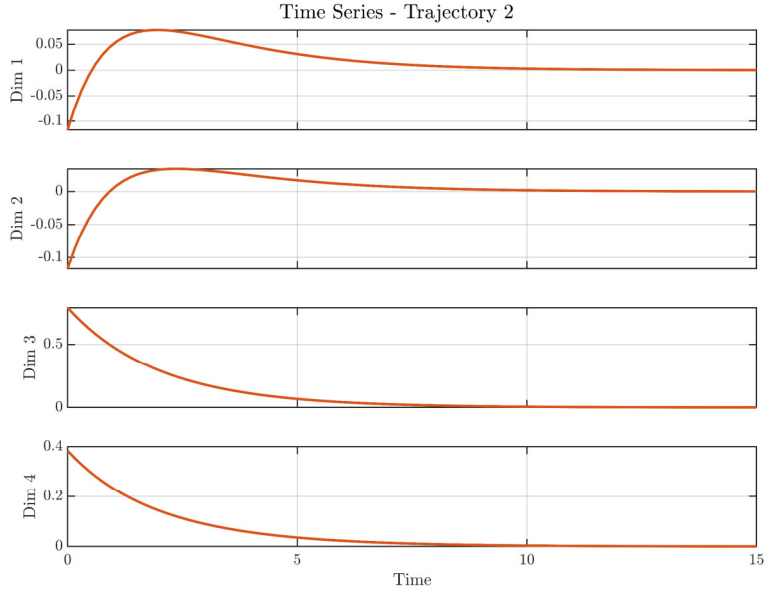


Figure S2. Time Series 2

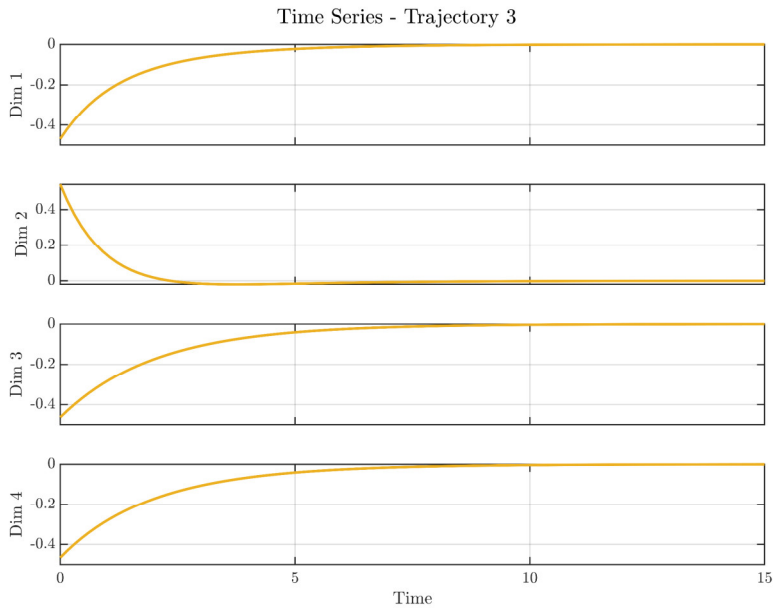


Figure S3. Time Series 3

Figures S1 - S3 represents the evolution of four state variables (State Dimension 1- State Dimension 4) about Time Series 1- Time Series 3. All state variables converge monotonically to zero at any initial value, indicating that the stability of the system is global and not limited to a subspace. It is worth noting that the sum of squares of each state $V(\mathbf{s})$, decreases to zero with time, indicating that $dV/dt < 0$ holds for all $\mathbf{s} \neq 0$. According to Lyapunov stability theory, it is proved that the system is asymptotically stable at the origin.

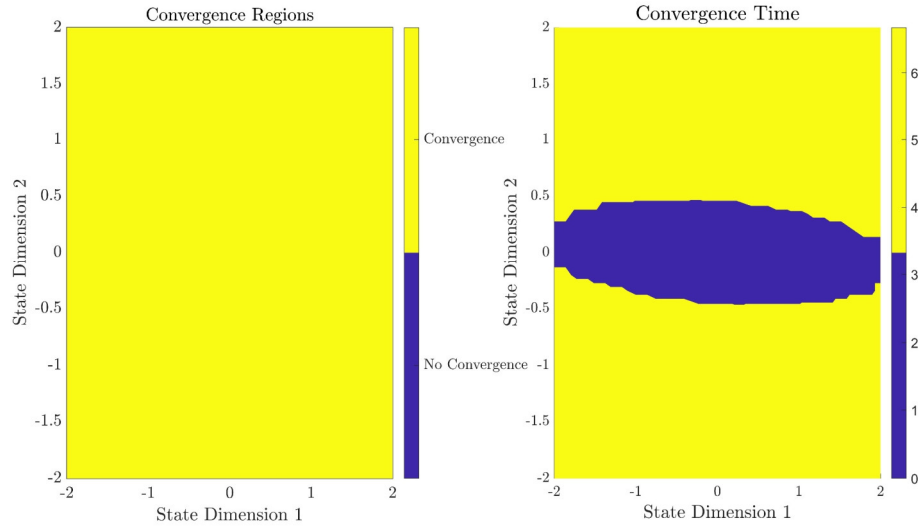


Figure S4. Basin of Attraction

Fig S4(left) forms a two-dimensional phase plane with State Dimension 1 and State Dimension 2. The green region represents that the trajectory of the system with initial conditions in this region will converge to the stable equilibrium point, and its range indicates that the equilibrium point has a large stable region, indicating that the system is robust to disturbances in the initial state.

Fig S4(right) further quantifies the convergence rate within the attraction basin. The color changes from dark blue to bright yellow, indicating that the convergence time changes from short too long. It means that the closer to the equilibrium point, the darker the color, which means that the convergence speed is very fast. The closer to the boundary of the attraction basin, the lighter the color, meaning that convergence takes significantly longer.

The left panel defines the stable boundary from geometric topology, and the right panel depicts the dynamic gradient within the stable region from convergence performance. The combination of the two shows that the model not only has equilibrium points, but also has a large range of attractive basins