WP4: Data Analysis and AI Algorithms

Daniel Durstewitz, Georgia Koppe, Peter Kuppens, Manuel Brenner
Central Institute for Mental Health Mannheim

General Assembly meeting Edinburgh, 31/3/22

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 - Staff

Peter Kuppens

Daniel Durstewitz

Georgia Koppe

Manuel Brenner

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – Objectives&Deliverables

IMMERSE

Two key objectives:

- **Simple Statistics and Visualisations (first 9 months)**
- **RNN-based AI models for multimodal data integration and prediction (years 2 and 3)**

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – Deliverables & Milestones

1) Simple Statistics and Visualisations

Main results:

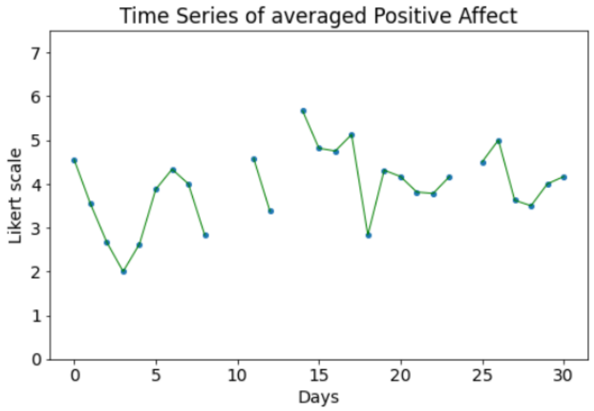
- Google document "RobustStatisticsAndVisualizations", which contains descriptions of statistics and visualisations and comments on minimal statistical requirements

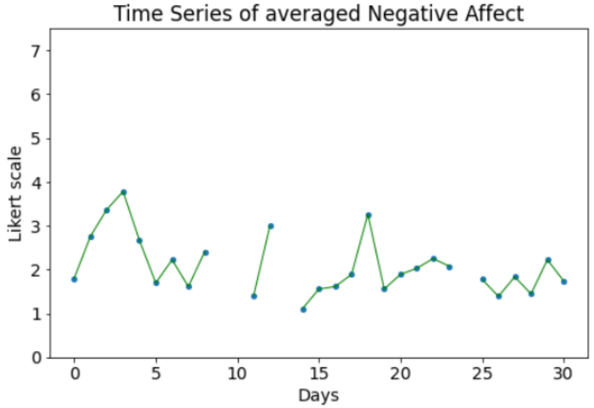
- Jupyter Notebook with example implementations in Python

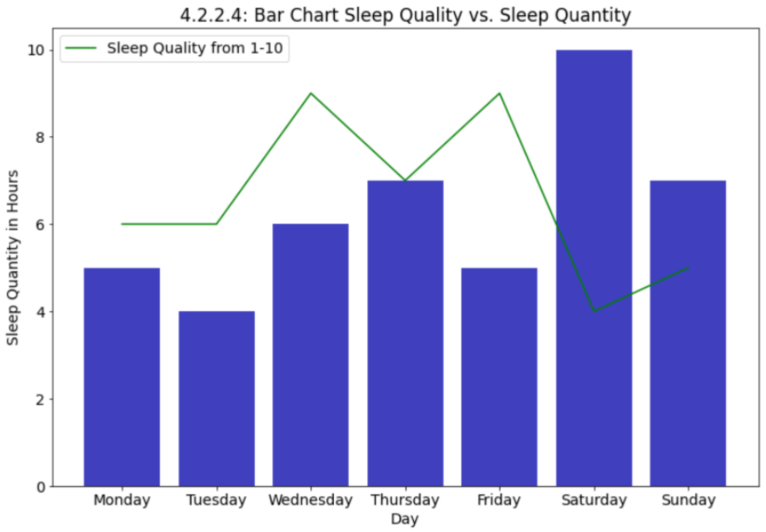
1. Preprocessing
2. Time series graphs (4.2.2.1.)
3. Positive and negative affect (for different time intervals) (3.4.2)
4. Means, medians (3.4.1)
5. Distribution plots (pointplot, boxplot, violinplot) (4.2.2.3)
6. Distribution plots with adjustable error bars and multiple items (4.2.2.3)
7. Pie charts (4.2.2.4)
8. Acitivity vs. affect distribution plots 4.2.2.4)
9. Bar chart for comparison between two items (4.2.2.4)
10. Ordinal correlation measures between items and significance tests (3.4.3)
11. Trends and Mann-Kendall test (3.4.4, 4.2.2.7)
12. Significant differences between time periods with Kruskal-Wallis tests (3.4.5.)
13. Starplot (4.2.2.8, Hexplot)

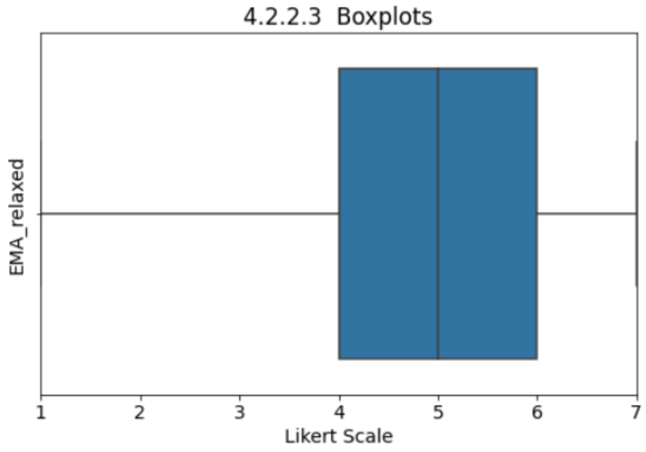
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

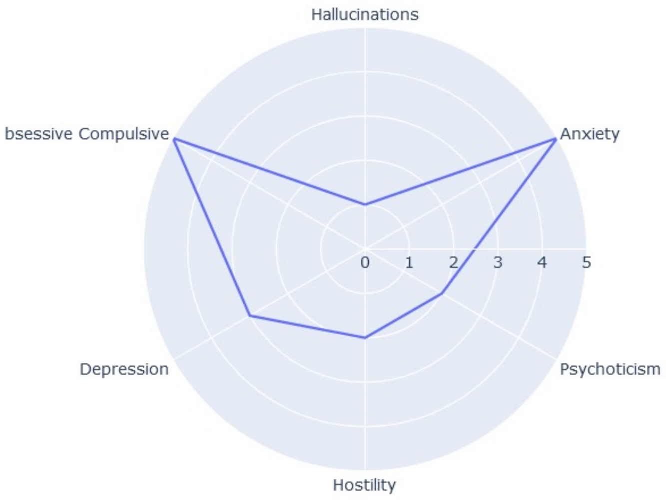
WP4 – Example Visualisations

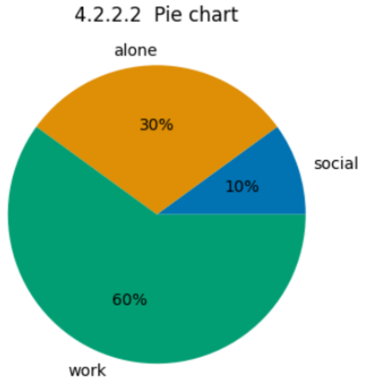








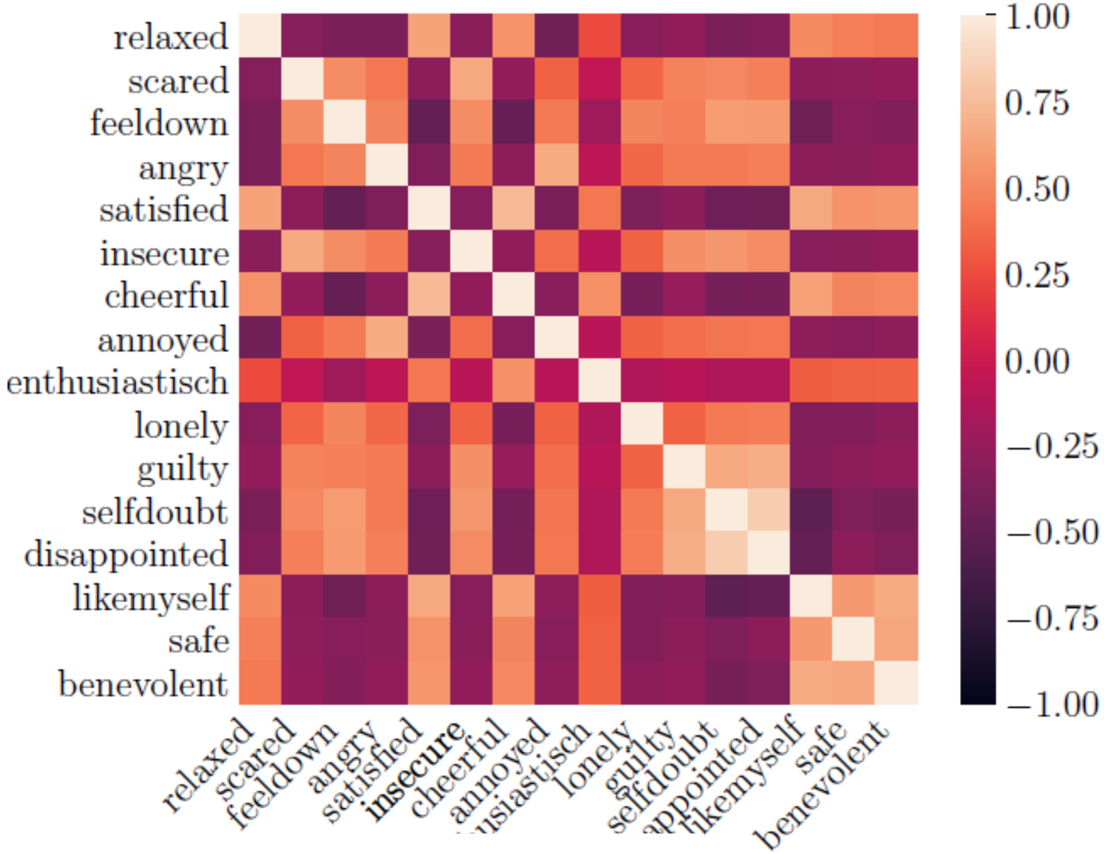




This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – Simple Statistics

- Spearman Rho significance test for correlation between different mood items, combined with text output



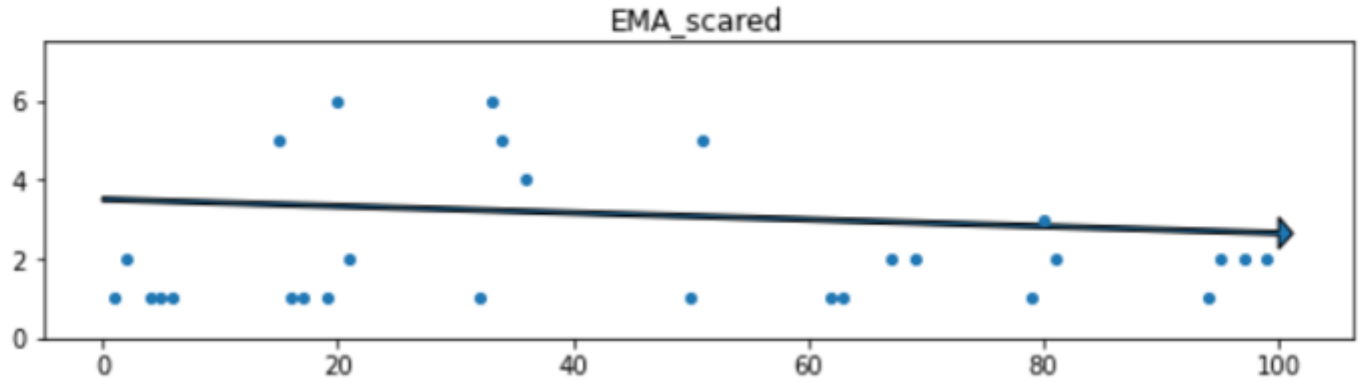

```
Positive Correlations:
Feature EMA_feeldown and feature EMA_scared are significantly postively correlated.
Feature EMA_satisfied and feature EMA_relaxed are significantly postively correlated.

Negative Correlations:
Feature EMA_satisfied and feature EMA_feeldown are significantly negatively correlated.
```

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – Simple Statistics

- **Kendall-Mann test for detecting trends in time series**

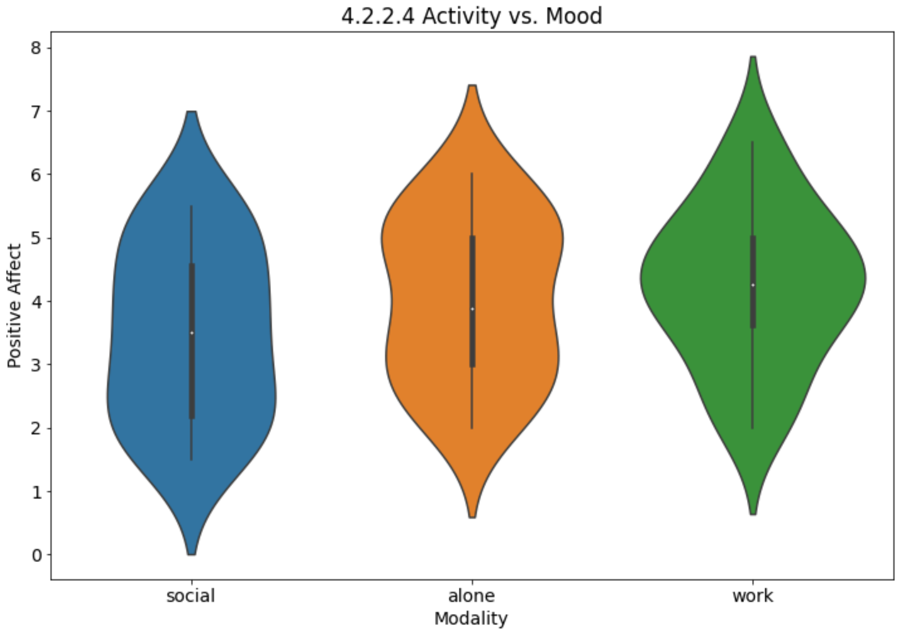


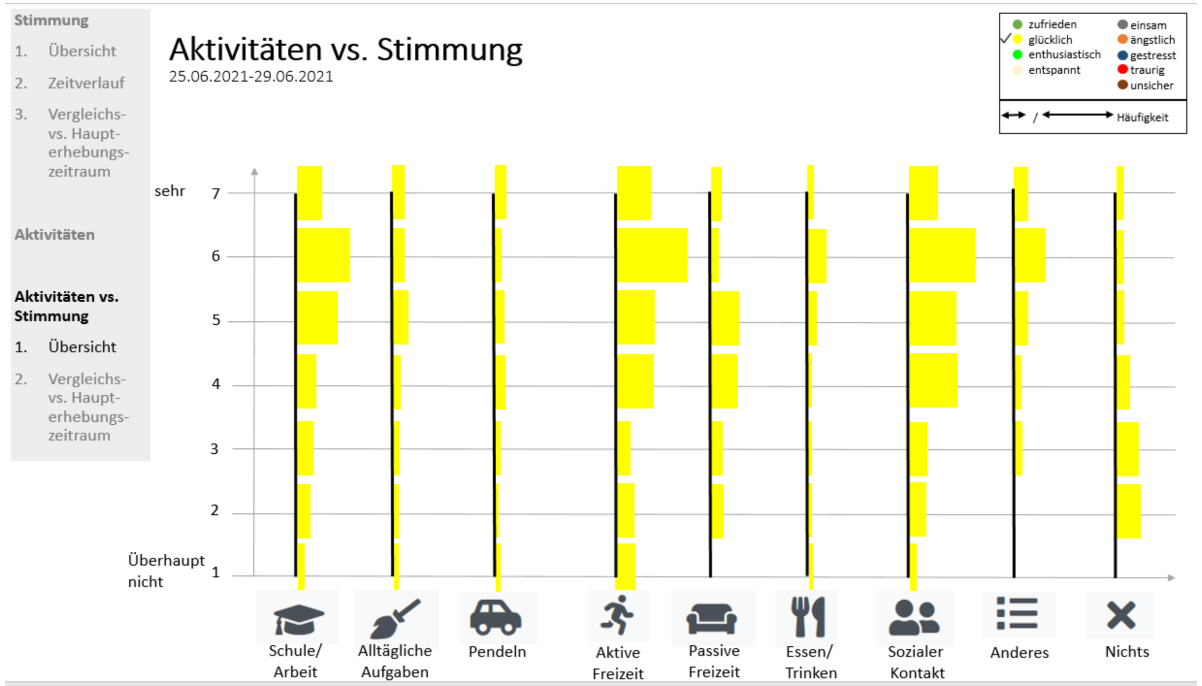
- **Example text output: “there is a significant negative trend in the “scared” mood item”**

This project has received funding from the European Union’s Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – Simple Statistics

- **Kruskal-Wallis test to assess whether distributions are significantly different between different activities (e.g. PA is higher while social compared to being alone)**

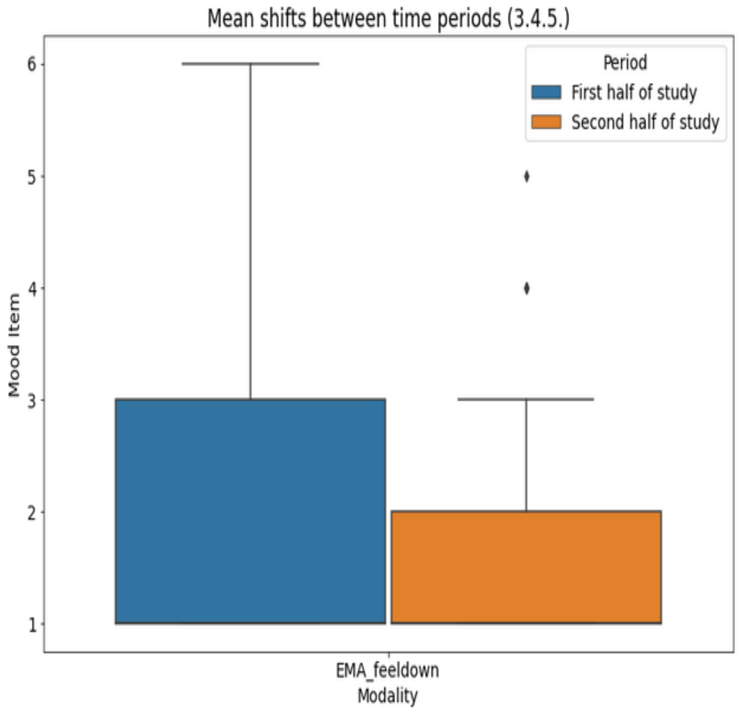




This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – Simple Statistics

- Kruskal-Wallis tests for shifts between between time periods
- → output if a mood item/PA/NA has significantly changed
 between different time intervals, e.g. beginning and end of study)



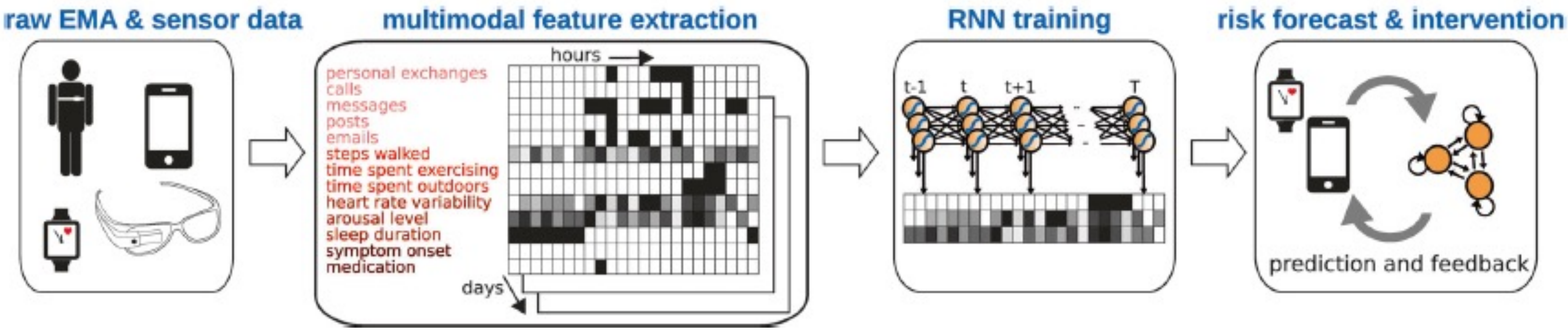
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

2) RNN-based AI models for multimodal data integration and prediction

Strengths of AI algorithms:

- 1) Extraction of complex features from data
- 2) Integration of multiple data modalities
- 3) Extraction of complex temporal dynamics
- 4) Ability to flexibly forecast future dynamics



Durstewitz*, D., Koppe*, G. & Meyer-Lindenberg, A. (2019).

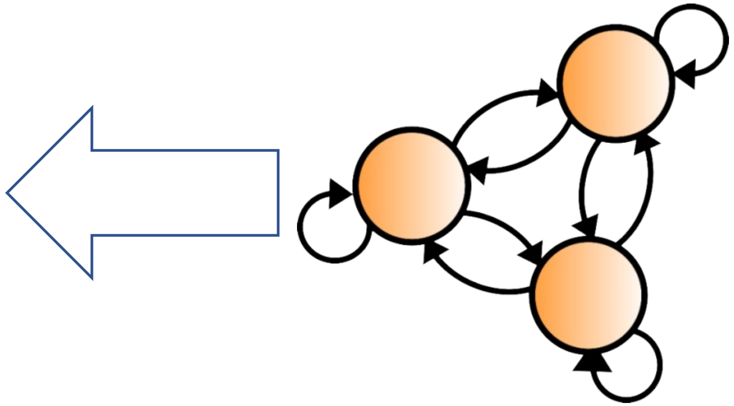
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

Recurrent neural networks approximate temporal dynamics

Lorenz Eqns.

$$\dot{x}_1 = s(x_2 - x_1)$$
$$\dot{x}_2 = rx_1 - x_2 - x_1x_3$$
$$\dot{x}_3 = x_1x_2 - bx_3$$

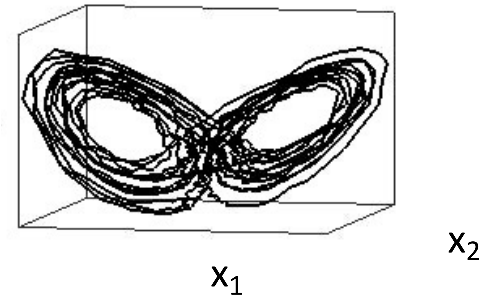


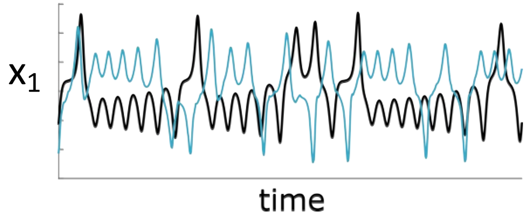
Generic RNN Eqns.

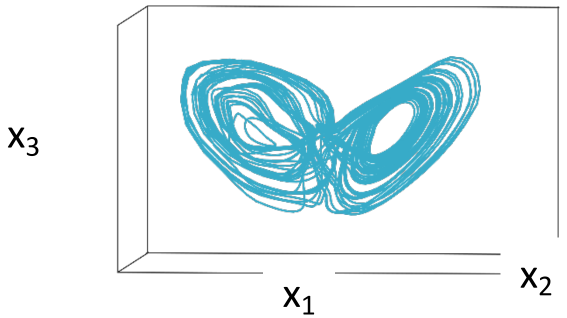
$$\mathbf{z}_t = f(\mathbf{z}_{t-1}, \mathbf{u}_t) + \boldsymbol{\varepsilon}_t$$
$$\mathbf{x}_t = g(\mathbf{z}_t) + \boldsymbol{\eta}_t$$

True trajectory

Simulated trajectory







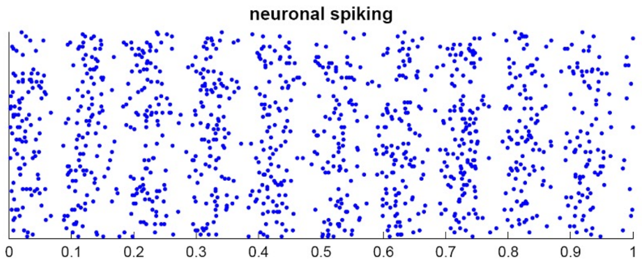
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

Observation models connect latent dynamics to measurements

Observation Models

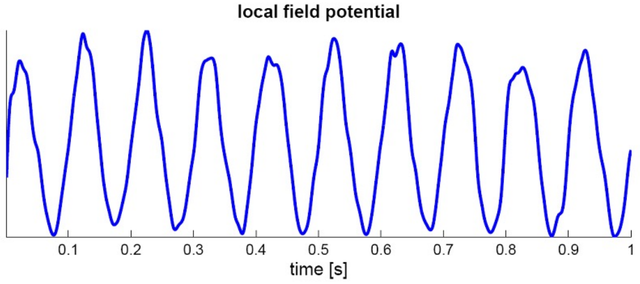
$$\mathbf{z}_t = f(\mathbf{z}_{t-1}, \mathbf{u}_t) + \boldsymbol{\varepsilon}_t$$
$$\mathbf{x}_t = g(\mathbf{z}_t) + \boldsymbol{\eta}_t$$



Zero-inflated Poisson model for spike data

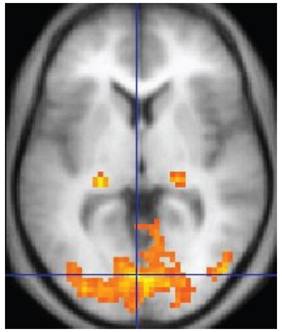
$$\mathbf{q}_t \mid \mathbf{z}_t \sim \mathrm{ZIP}(\boldsymbol{\pi}_t, \boldsymbol{\lambda}_t),$$

$$p(q_{st}|\mathbf{z}_t) = \begin{cases} \pi_{st} + (1-\pi_{st})e^{-\lambda_{st}} & \text{for } q_{st} = 0 \\ (1-\pi_{st})\frac{\lambda_{st}^{q_{st}}}{q_{st}!}e^{-\lambda_{st}} & \text{for } q_{st} > 0 \end{cases}$$



Hemodynamic response function for fMRI

$$\mathbf{x}_t = \mathbf{B}(hrf * \mathbf{z}_{\tau:t}) + \mathbf{J}\mathbf{r}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Gamma})$$

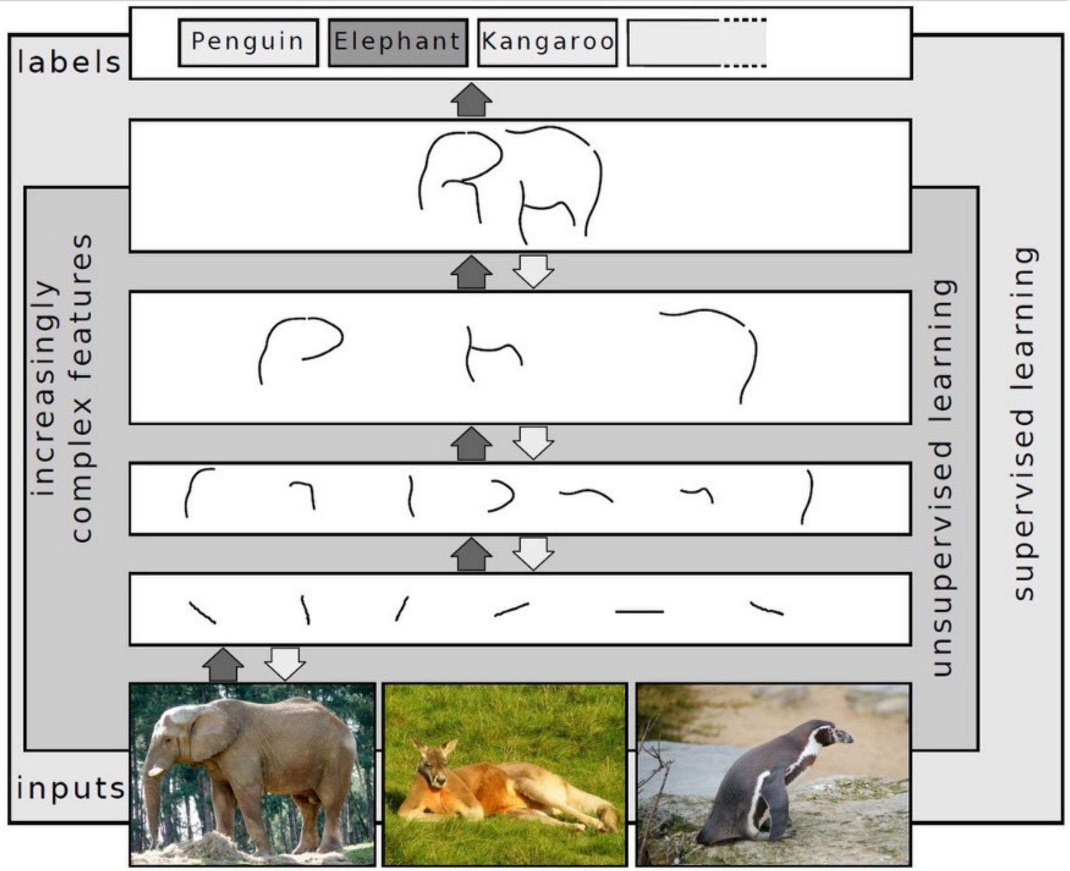


This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

Feature extraction

(Deep) neural networks learn to extract complex features from input data that is useful e.g. for classification (elephant vs. penguin) and bears some similarity e.g. to human visual cortex



From Sven Behnke [CC BY-SA 4.0]

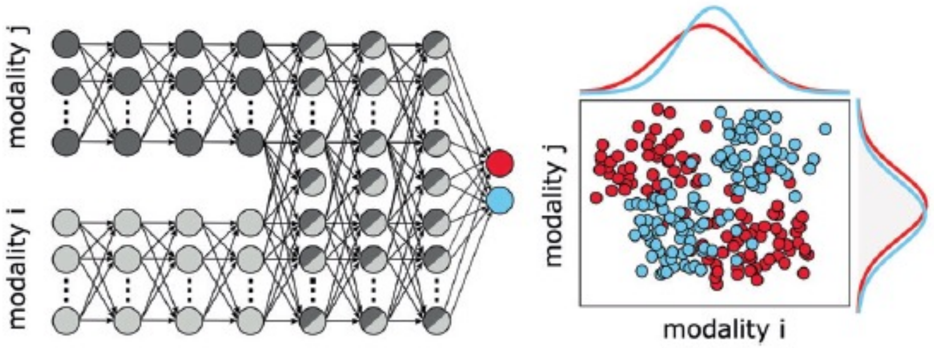
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

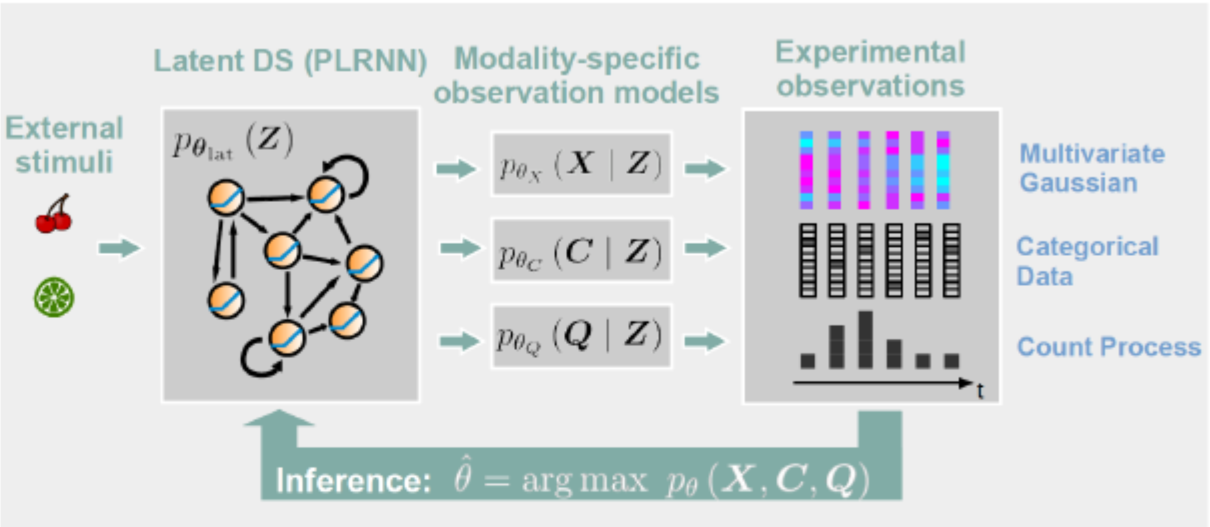
Multimodal Data Integration

Combine different data modalities
(e.g. GPS, step counts, EMA data)
to extract combined features beyond the capacity of preprocessing and manual feature extraction

Incorporates more robust data into training routine
→ stronger statistical grounding and improved models



From Durstewitz, Koppe, Meyer-Lindenberg 2019



Kramer et al., 2021

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

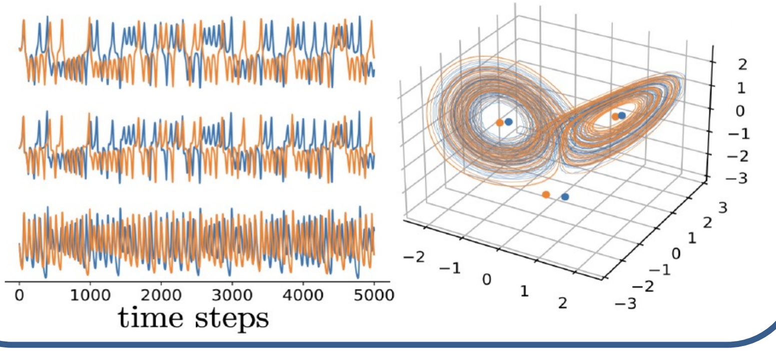
WP4 – AI Algorithms

Extraction and forecasting of complex (latent) dynamics

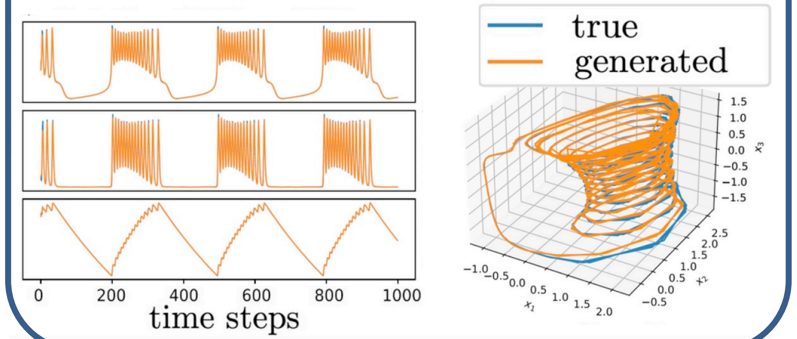
RNN-based algorithms allow us to extract complex temporal patterns in an unsupervised way from benchmark and empirical data

→ if we reconstruct the underlying dynamics, we can predict the future of the system and analyze its properties

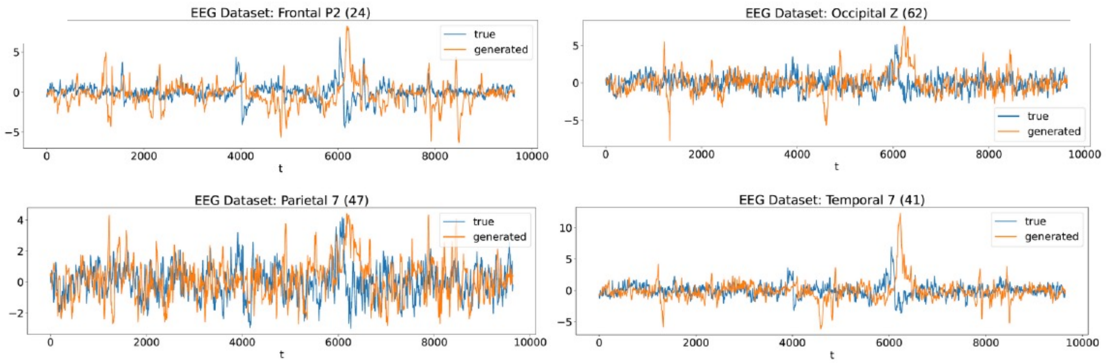
Lorenz-63 (atmospheric convection)



Biophysical bursting neuron model



Experimental EEG Data



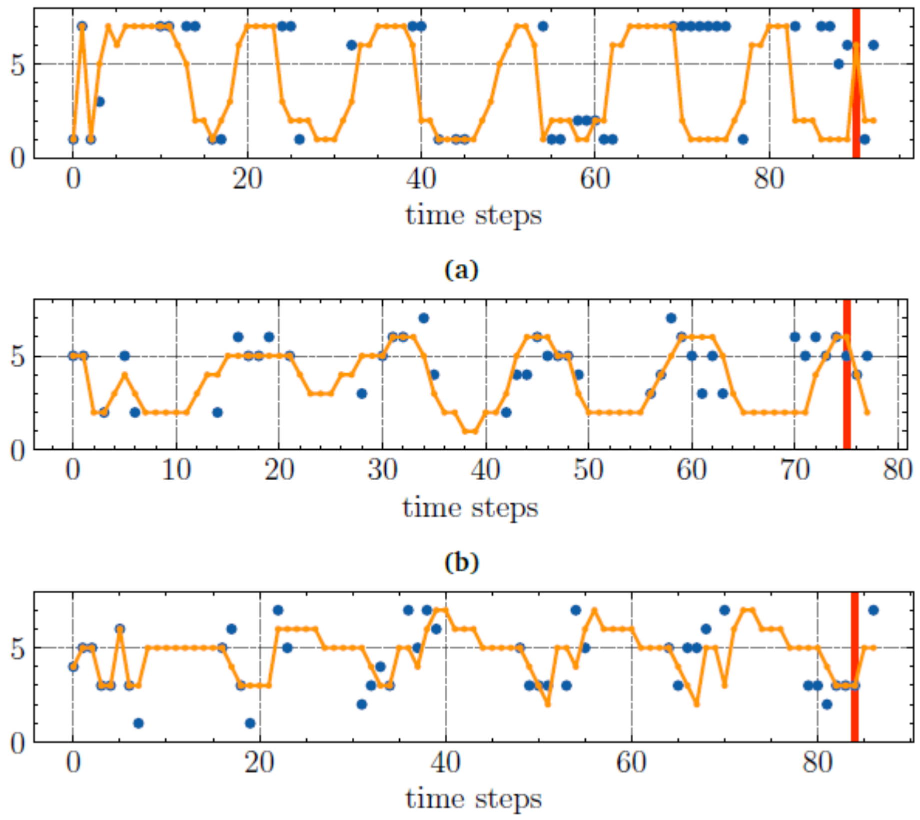
Brenner et al., 2022

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

IMMERSE

AI algorithms can automate extraction of complex nonlinear temporal patterns, And create forecasts for unseen data points

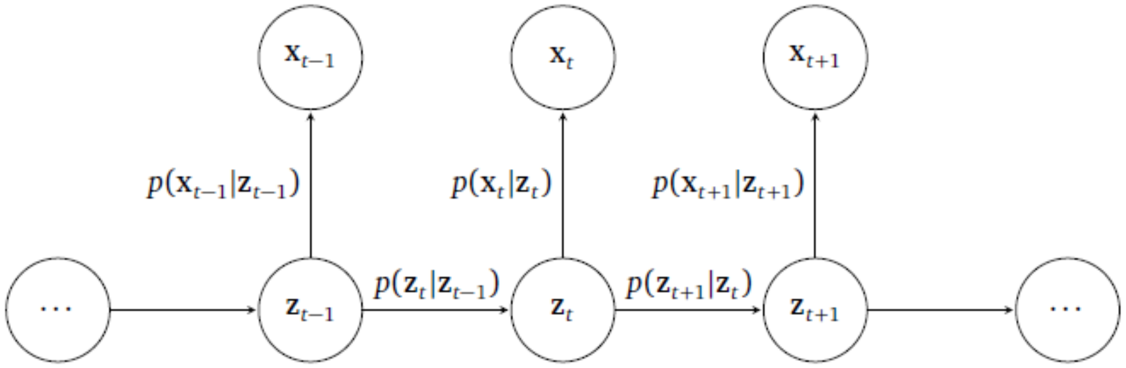


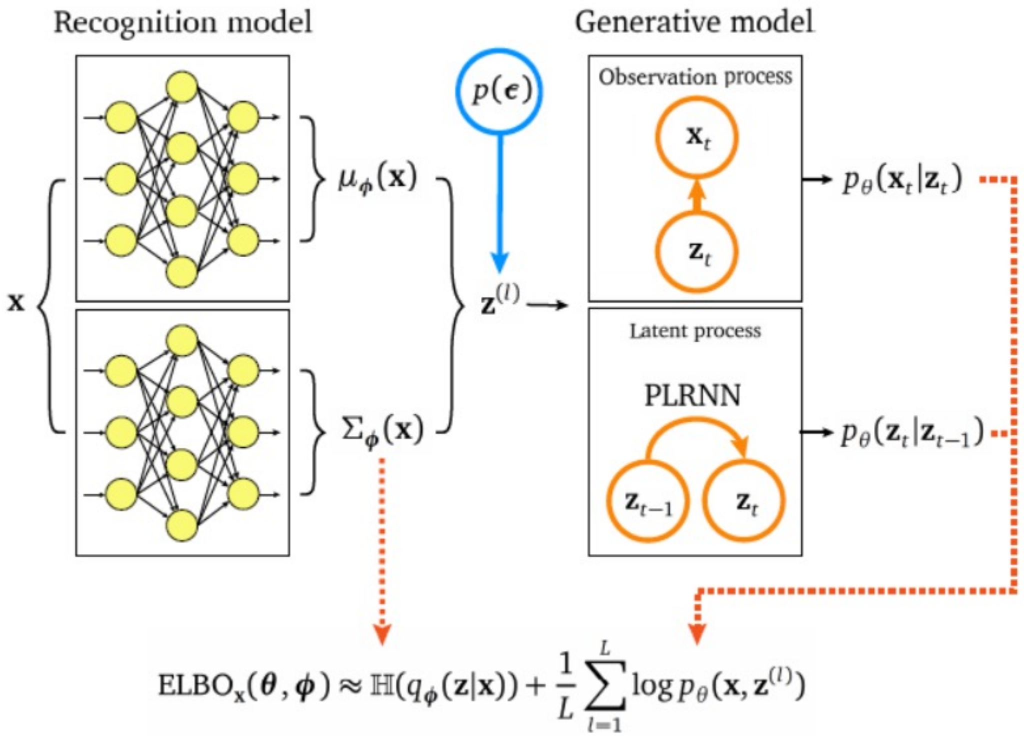
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

IMMERSE

- Training RNNs on nonlinear dynamical systens is challenging
  (e.g. exploding+vanishing gradients)
- Our approach combines popular and flexible machine learning frameworks such as variational autoencoders with sparse teacher forcing





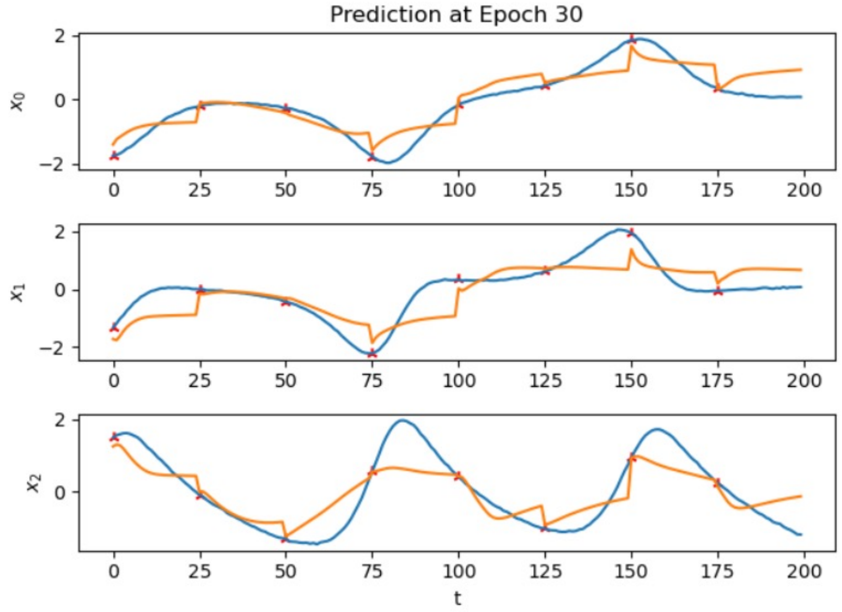
Provided by Unai Fischer

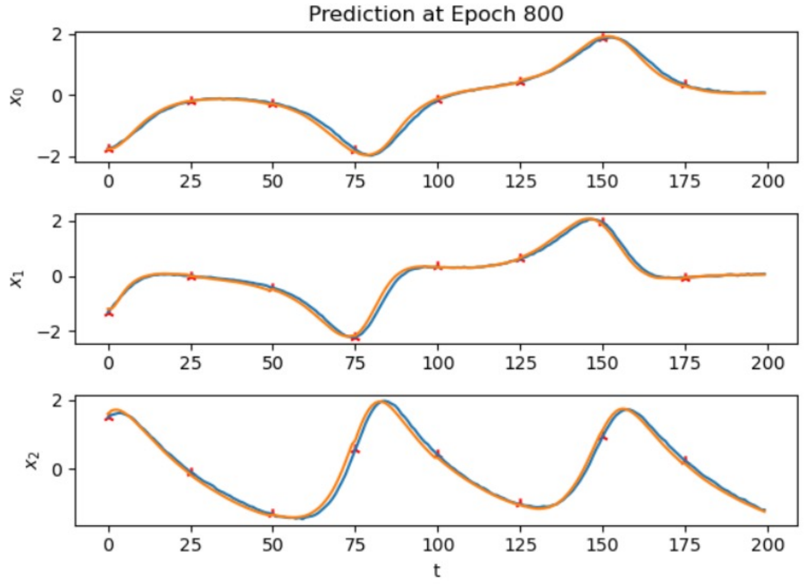
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

Sparse Teacher Forcing

- Grounded in Lyapunov theory of chaotic dynamical systems
- Tool to regularize gradients during training without loosing dynamical information
- Sparsely supplies a control signal in latent space during training
- Requires inversion of observation models
  → autoencoder-teacher-forcing hybrid approach

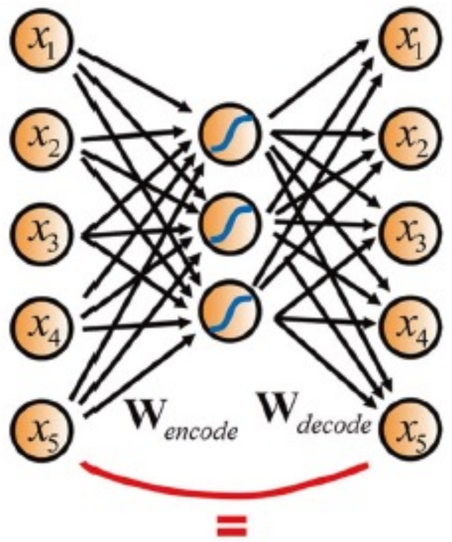


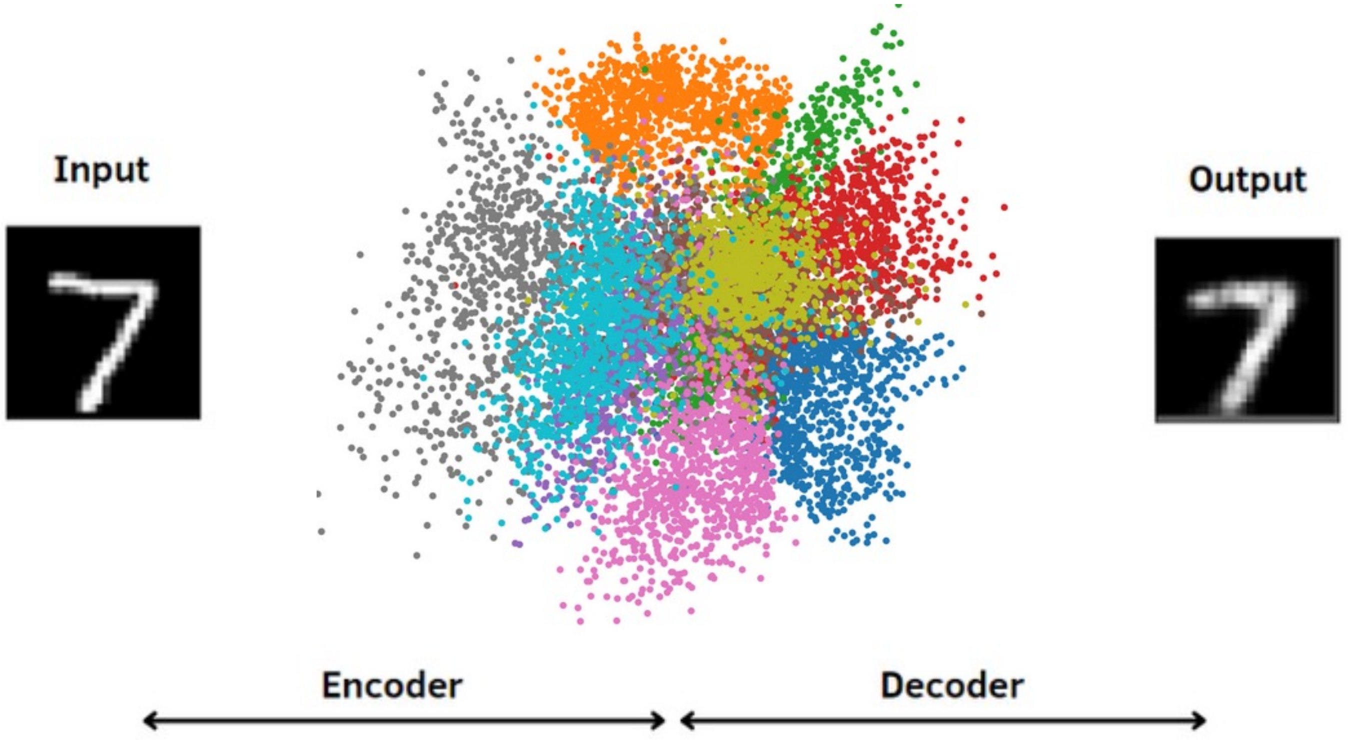


This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

Autoencoders allow for latent representation of complex patterns within (multimodal) data





Durstewitz*, D., Koppe*, G. & Meyer-Lindenberg, A. (2019).

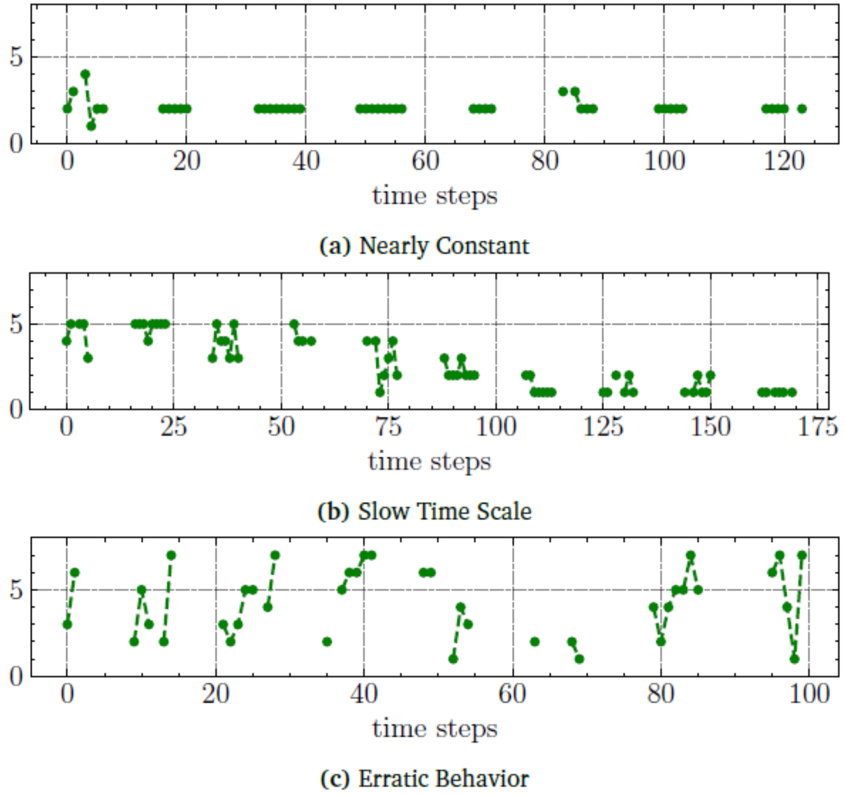
This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

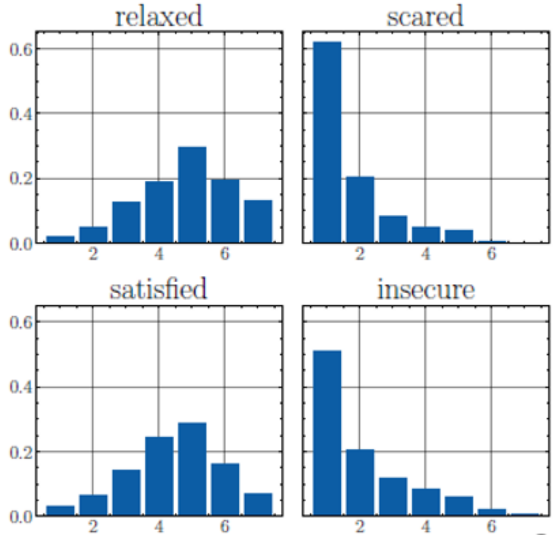
WP4 – AI Algorithms

Main challenges

EMA data is often:
- noisy
- non-stationary
- different time scales
- irregularly sampled (e.g. day vs. night)
- many missing values
- contingent on self-report
- small sample sizes
- Different distributional assumptions

→ Extremely difficult to train!





Provided by Unai Fischer

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

Challenge	Solution
Noisy data	Statistical models (e.g. variational autoencoders)
Multiple data distributions	Flexible observation models (e.g. cumulative link, zero-inflated Poisson)
Complex (chaotic) dynamics	Training via sparse teacher forcing
Small sample sizes	Hierarchisation of training algorithm, multimodal data
Different time scales	Line attractor regularization
Unreliable self-report	Integration of different data modalities (step counts, GPS)
Irregular sampling	Continuous time models, multimodal data
Missing values	Imputation methods, multimodal data
Non-stationary dynamics	Inclusion of non-stationary parameters

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – AI Algorithms

Model Hierarchisation

- augmenting predictive strength of models through group level information
- Training models on all subjects simultaneously
  →make use of all available data given experimental time series are often very short
- Introducing hierarchical set of parameters, some of which are trained on the group level and some of which are only trained on the individual level
  →retain individual expressiveness while leveraging group statistics

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 – Year 2022 Goals

- Multimodal data integration in sparse teacher forcing framework
- Development of benchmark data and preprocessing steps to evaluate algorithmic performance
- Building and testing a hierarchical inference framework
- Addition of further data modalities (e.g. step counts, activity data and GPS data with appropriate preprocessing)

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)

WP4 - Objectives

Collaboration / input other WPs

WP2: assistance with testing of statistics and visualisations, minimal requirements

WP3: agreeing on data modalities+preprocessing steps

This project has received funding from the European Union's Horizon 2020 research and innovation Programme under grant agreement 945263 (IMMERSE)