IMMERSE

Implementing Mobile MEntal health Recording Strategy for Europe

H2020 - 945263

D 4.2	**Algorithms and software environment for DTSM- based multi-modal big data integration**
Dissemination level[1]	PU
Contractual date of delivery	31.03.2024
Actual date of delivery	05.03.2024
Type[2]	OTHER: Software
Work package	WP4
Work package lead	Georgia Koppe

Author list

Organisation	Name	Contact information
P2 – CIMH	Georgia Koppe	georgia.koppe@zi-mannheim.de
P2 – CIMH	Manuel Brenner	Manuel.Brenner@zi-mannheim.de
P2 – CIMH	Daniel Durstewitz	daniel.durstewitz@zi-mannheim.de

[2] **Use one of the following codes:**
R: Document, report (excluding the periodic and final reports)
DEM: Demonstrator, pilot, prototype, plan designs
DEC: Websites, patents filing, press & media actions, videos, etc.
OTHER: Software, technical diagram, etc

Document history

Version	Date
(1)	05.03.2024

List of abbreviations

DSR: Dynamical systems reconstruction

EMA: Ecological Momentary Assessment

MVAE: Multimodal Variational Autoencoder

MTF: Multimodal Teacher Forcing

RNN: Recurrent Neural Network

WP: work package

Deliverable report

1 Summary

Deliverable 4.2 of WP4 focused on developing a completely novel algorithmic approach for multi-modal and big data integration in the context of deep time series models (DTSM) based on recurrent neural networks (RNNs). The developed algorithm is termed Multimodal Teacher Forcing (MTF), and is described in detail in (Brenner et al., 2023). Here, we therefore focus on describing the basic architecture and training algorithm of the proposed framework, the key ideas and insights underlying its development, as well as the extensive benchmarking that was performed for model validation.

2 Description of DTSM algorithm.

2.1 Basic architecture

Our data driven analysis approach rests on approximating a generative latent dynamics model from time series data in order to model the behavior of the latent dynamics underlying the observed data. Here we choose a mathematically tractable formulation of a piecewise linear RNN (PLRNN) model called the dendritic PLRNN (dendPLRNN, Brenner et al., 2022):

$$z_t = Az_{t-1} + W\phi(z_{t-1}) + h + Cs_t + \epsilon_t\,, \qquad \epsilon_t \sim N(0, \Sigma)$$

where z_t is the M-dimensional latent state vector at time t, $A \in \mathrm{R}^{\mathrm{MxM}}$ is a diagonal matrix of time constants, $W \in \mathrm{R}^{\mathrm{MxM}}$ is an off-diagonal weights matrix, $\phi(z) = \sum_{b=1}^{B} \max(z_{t-1} - h_b)$ is a nonlinearity composed of an expansion of piecewise linear functions (cf. Brenner et al., 2022), and h is a constant threshold vector. Finally, s_t is a (P×1)-dimensional input vector allowing for external (experimental) stimuli to directly affect the latent state via a regression coefficient matrix $A \in \mathrm{R}^{\mathrm{MxP}}$, and ϵ_t is a Gaussian noise vector with 0 mean and covariance Σ.

Through our novel developments, this model can now be coupled to different probabilistic decoder (observation) models, relating latent states of the RNN model z_t to observations x_t. In the case of Gaussian observations, this can take the simple form of a linear observation model:

$$x_t = Bz_t + \gamma_t , \qquad \gamma_t \sim N(0, \Lambda)$$

where $B \in \mathrm{R}^{\mathrm{MxN}}$, and γ_t is an observation noise vector with covariance $\Lambda \in \mathrm{R}^{\mathrm{NxN}}$.

Alternative, we can also couple the latent process to discrete observations, such as ordinal observations (as often assessed via EMA), using e.g. an ordered logit (proportional odds) observation model that couples the likelihoods of the occurrence of different discrete ratings to the underlying continuous process:

$$a_t \sim \mathrm{Ordinal}\ (\beta z_{t,}, \epsilon).$$

Note that this formulation is very general and allows the coupling of many different observation models for different continuous or discrete observations, such as Poisson or zero-inflated Poisson models for count data or categorical decoders for categorical data, and thus allows for reconstruction from discrete data, such as behavioral data, alone.

The framework also allows **to integrate over modalities**, that is, to integrate various observation models simultaneously by connecting different types of observations within a common latent dynamics space, enabling integration of various sources of information for optimal reconstruction.

The developed approach hence allows the joint extraction of complex dynamics from multiple modalities simultaneously, such as continuous GPS data with discrete ordinal Likert scale ratings and discrete step counts. It is one of the first models ever to perform multimodal data integration for dynamical systems reconstruction (DSR), and reaches state of the art performance, as shown in section 3.

2.2 Training framework.

Teacher forcing.

The developed approach is grounded on recent advances in training RNN-based time series models leveraging a technique called teacher forcing (TF) for effectively controlling gradient divergence during training while capturing important long-term dependencies present in time series data. Techniques like shallow TF (Mikhaeil et al., 2022) or Generalized TF (Hess et al., 2023) enable DSR even from challenging real-world data on which many previous methods failed. Hence, the key idea for making DSR from multiple, statistically distinct data sources work is the generation of a multimodal TF signal for efficient training of a reconstruction model.

Approaches based on TF leverage a combination of data-inferred states and forward-iterated latent states to balance the training process. This involves strategically replacing certain latent states with data-inferred states during training. These data-inferred states are derived by inverting the decoder model. This inversion is straightforward in the context of continuous data, e.g. when using a simple linear Gaussian observation model.

Inverting the decoder model is, however, not always possible, particularly not for discrete random variables. Moreover, in the case of multiple simultaneously observed data modalities, it is unclear how to combine the different data modalities to obtain an optimal estimate for the TF signal. To address the challenges posed by different data modalities and the infeasibility of inverting the decoder model in certain scenarios, we employ a Multimodal Variational Autoencoder (MVAE). This MVAE creates a joint latent representation from various data types, providing a versatile TF signal for training. The MVAE is linked to the observations through shared decoder models with the DSR model.

To further enforce consistency between the MVAE and DSR model latent codes, we assume that the MVAE prior is given through the DSR model. Training the MVAE thus involves minimizing the negative Evidence Lower Bound (ELBO), using the reparameterization trick for latent random variables. We experimented with different encoder models for the MVAE, such as those based on RNNs and Transformers, finding that a Convolutional Neural Network (CNN) yielded the best results.

Our total training loss combines the negative ELBO of the MVAE with the reconstruction loss of the DSR model and the consistency loss between encoder states and latent states of the DSR model. This integration ensures that our training process is both robust and effective, guiding the DSR model towards accurate dynamic system reconstruction. Fig. 1 illustrates the general MTF framework.

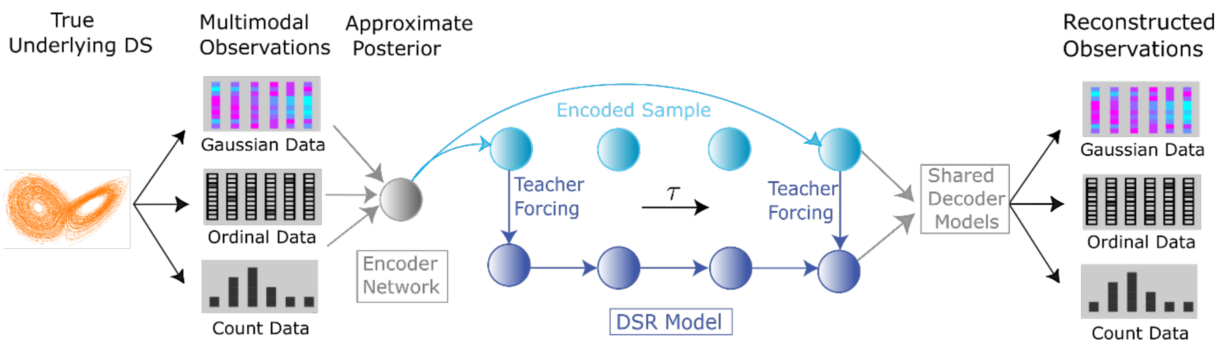

Fig. 1. General MTF framework. Multimodal time series observations are mapped to an approximate posterior distribution, using an encoder network. Samples from this distribution are used as an effective control signal to enable training of a DSR model for extracting a joint latent dynamics model of the multimodal observations. Both latent codes (that of the DSR model and that of the encoder network) are coupled to shared decoder models to compute likelihoods over the reconstructed observations. The framework can be used with any type of dynamics model (e.g., other types of RNNs), as well as with any set of encoder or decoder models (e.g., other types of observation modalities). Based on Brenner et al. (2023).

Parameter hierarchization framework.

Optionally, the MTF algorithm can be trained via a parameter hierarchization approach (Fig. 2). In this context, parameter hierachization entails varying levels of parameter inference, wherein the upper hierarchy parameters are inferred from time series data derived from multiple individuals, while the lower hierarchy is inferred from data of an individual. This allows the inference procedure to **integrate over data from different participants**, performing 'big data' integration. To achieve this, model parameters are partitioned into a low-dimensional trainable weight vector and projection matrices responsible for projecting this vector onto parameters of the RNN model. The projection matrices are jointly trained and shared across multiple individuals. The low-dimensional weight vectors are fine-tuned for each individual, thereby reducing the inter-individual differences to a low-dimensional, potentially interpretable parameter manifold.

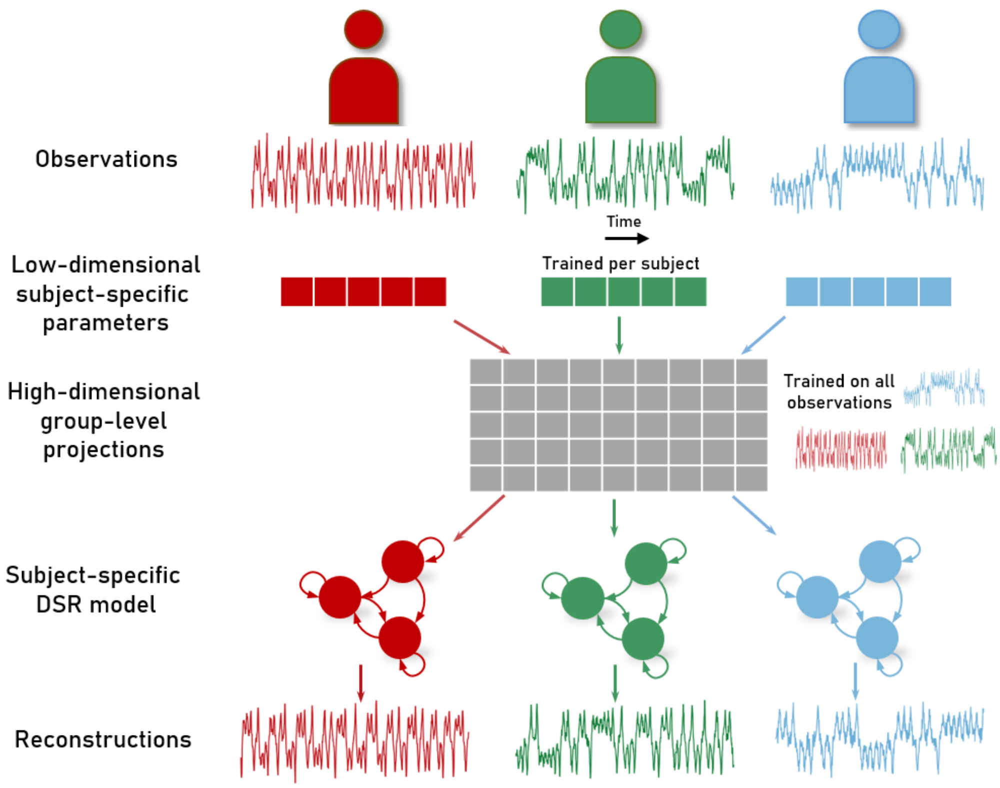

Fig. 2. Illustration of the hierarchization framework, combining low-dimensional subject specific parameter vectors with projections learned on the group level, mapping to the parameters of subject specific DSR models, such as the dendPLRNN, capturing individual differences while leveraging shared group level structure.

3 Model validation

We validated our implementation of the MTF framework on a series of benchmarks by testing its DSR performance, and comparing it to other DSR models. A range of tools and performance metrics were developed to assess DSR in the process.

We tested different scenarios relevant in the context of IMMERSE, showcasing different strengths of the MTF framework: multimodal data integration across continuous, ordinal and count observations (such as GPS data, Likert scale ratings and step counts), multimodal integration in a setting where continuous observations are highly distorted by noise but ordinal observations are available, and DSR exclusively from ordinal or categorical data.

3.1 DSR from multimodal benchmarks

To benchmark the approach in settings reflecting the experimental multimodal data collected in the IMMERSE consortium, we simulated time series from two nonlinear dynamical benchmark systems, the Lorenz attractor and a 6-dimensional chaotic network model, and filtered the simulated time series through Gaussian, ordinal, and count process decoder models.

We then compared the DSR performance of the MTF algorithm on these benchmarks to several alternative approaches. Firstly, we compared it with a sequential multimodal Variational Autoencoder (VAE), the only other general approach for DS reconstruction from multimodal data in the literature (Kramer et al., 2022). Secondly, we used classical RNN training with modality-specific decoder models, where observations are provided as inputs at every time step. Another strategy tested was the 'multiple shooting' approach, adapted by us to handle multimodal data.

Additionally, we explored approaches where multimodal data were transformed to approximate Gaussian distributions using Box-Cox transformations and Gaussian kernel smoothing. These transformed datasets were then used to train the RNN either via standard Backpropagation Through Time with Teacher Forcing (BPTT-TF) or a VAE-based Teacher Forcing (VAE-TF) without modality-specific decoder models, termed GVAE-TF. In all these comparisons, the same RNN architecture (dendPLRNN) was employed.

Our evaluations focused on capturing the geometrical structure in state space and the asymptotic temporal structure of the underlying DS (Fig. 3). We used measures like Kullback-Leibler divergence for state space structure and Hellinger distance or auto-covariance functions for temporal structure. We also computed mean-squared errors (MSE) for short-term ahead prediction. However, in chaotic systems, these prediction errors are not indicative of the system's longer-term behavior due to the exponential divergence of nearby trajectories. Here the MTF framework outperformed all other approaches by sometimes large margins, showing that it successfully learns a joint model over all three data modalities.

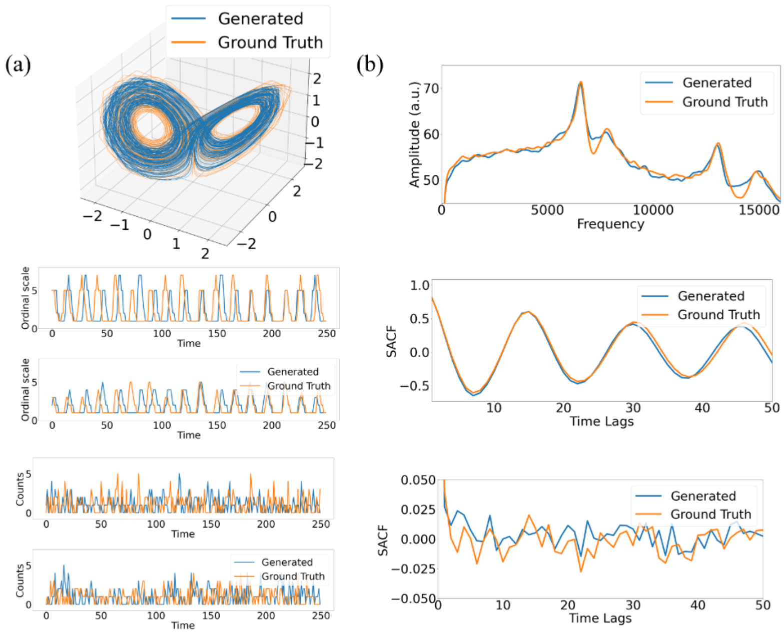

Fig. 3. a) Example reconstructions jointly reconstructed from continuous, ordinal and count data from the chaotic Lorenz-63 benchmark systems. b) Temporal agreement, based on the power spectrum (top) for continuous data and the Spearman autocorrelation function (bottom) for discrete observations. Taken from Brenner et al., 2023.

3.2 DSR for noise distorted continuous data and ordinal data

To mimic another clinically relevant situation, we then tested the algorithm in a setting where the underlying continuous data was highly distorted by noise (with 50% of the data variance), e.g. induced by noisy measurement devices or artifacts, but jointly measured ordinal data sampled from the same ground truth system was available. We compared DSR on this data with and without including additional data. We found that in this setting, including ordinal data allowed successful reconstructions of the chaotic Lorenz system, even when in the unimodal setting no such reconstructions were possible anymore (cumulative histograms over state space divergence and power spectrum agreement in Fig. 4 b,c).

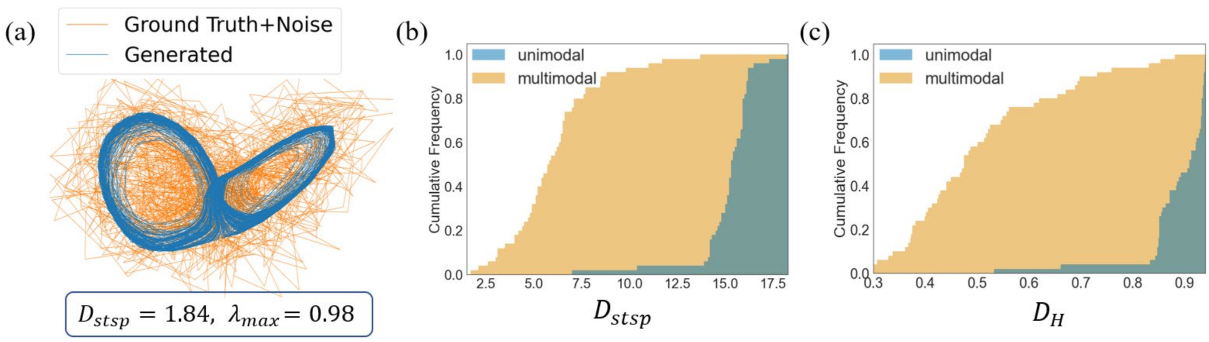

Fig. 4. DSR for a setting where continuous observations are heavily distorted by observation noise (here with 50% of the data variance), and simultaneously provided ordinal observations. Normalized cumulative histograms of geometrical attractor disagreement and power spectrum Hellinger distance show that reconstructions are still possible with ordinal observations on board, while largely failing when using solely the distorted continuous observations. Taken from Brenner et al., 2023.

3.3 DSR from discrete data

Motivated by these results, we explored the feasibility of DSR using solely ordinal data, a scenario relevant in psychiatric settings where often only ordinal or behavioral data are recorded. This approach represents a significant challenge compared to multimodal settings with Gaussian data, as ordinal data substantially coarse-grains the underlying continuous dynamical process, omitting intricate geometric and topological information.

To address this, we first tackled DS reconstruction using only ordinal data, randomly sampling 8 ordinal variables with 7 levels each, a scaling often employed in studies based on ecological momentary assessment (EMA) (see Fig. 5 top row).

Furthermore, we attempted DSR based on purely symbolic representations of the dynamics. This involved using symbols corresponding to sub-regions of a grid superimposed on the attractor (Fig. 5 bottom row). Our findings demonstrate that successful DSR is in principle possible from just a symbolic (categorical) coding of the underlying chaotic attractor. This is, to our knowledge, the first instance such a result has ever been shown.

These results have implications for experimental psychiatry, where data is often limited to discrete or ordinal forms, such as patient responses or observed behaviors, which do not capture the full complexity of the underlying processes. Our approach shows that even with this limited data, it is possible to reconstruct complex underlying dynamical systems.

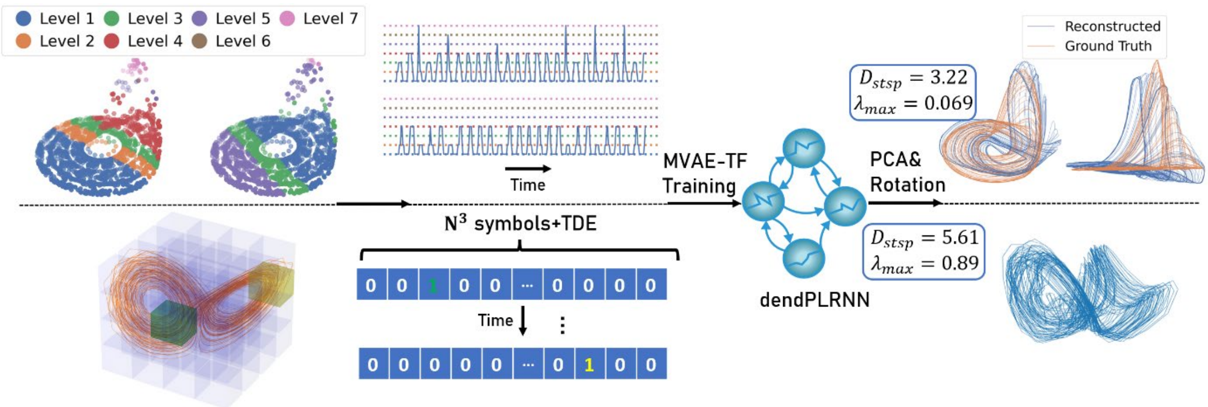

Fig. 5. DSR solely from discrete observations, using MTF. Top: Reconstruction of the chaotic Rössler attractor from only ordinal time series. Bottom: DSR from symbolic coding of Lorenz attractor.

3.4 Benchmark validation of hierarchization framework.

The hierarchisation framework was tested on very short time-series (with lengths common in the context of EMA time series), derived from the chaotic Lorenz system, where we tuned one of its parameters (commonly referred to as ρ) to induce a bifurcation between the cycle and chaotic regime of the Lorenz attractor, where different values of ρ are taken to represent different subjects.

Initial tests lead to positive results, where the subject specific models successfully captured individual differences, while at the same time leveraging group-level information for inferring dynamics from otherwise permissibly short time series.

Most notably, the low-dimensional parameter manifold could be related via linear regression to the ρ parameter of the ground truth system, allowing extrapolation of models to new dynamical regimes, and revealing clearly interpretable differences between subjects.

In experimental applications, the subject-specific low-dimensional parameter vector aims to capture all relevant subject-specific differences, and can be further related to psychological constructs and survey data after training. This could significantly aid in connecting individual differences between models to subject-specific differences and behavioral contingencies, which are otherwise much harder to discern from models solely trained on individual subjects.

Code base

A detailed codebase for training and testing the MTF-approach was implemented in Python. The MTF framework is implemented in Pytorch within a repository containing detailed documentation on training new models and examples for analyzing trained models. This repository is currently available to the IMMERSE consortium upon request.

References

Brenner, M., Hess, F., Mikhaeil, J., Bereska, L., Monfared, Z., Kuo, P. & Durstewitz, D. (2022), Tractable Dendritic RNNs for Reconstructing Nonlinear Dynamical Systems, Proceedings of the 39th International Conference on Machine Learning (ICML 2022)

Brenner, M., Koppe, G., & Durstewitz, D. (2023). Multimodal teacher forcing for reconstructing nonlinear dynamical systems. *In The 37th AAAI Conference on Artificial Intelligence (AAAI, Washington, MLmDS workshop, 2023)*.

Hess, F., Monfared, Z., Brenner, M. & Durstewitz, D. (2023), Generalized Teacher Forcing for Learning Chaotic Dynamics, Proceedings of the 40th International Conference on Machine Learning (ICML 2023)

Mikhaeil, J., Monfared, Z., & Durstewitz, D. (2023), On the difficulty of learning chaotic dynamics with RNNs, NeurIPS 2022