Daniel Durstewitz

Dept. Theoretical Neuroscience (\RGs)

Academic profile

Our department follows three major lines of research:

1) Development of novel machine learning (ML)/ artificial intelligence (AI) methods for data analysis

Our group is driving forward ML/AI methods both at the mathematical and algorithmic level, with a focus on time series data like neuroimaging, EEG/MEG, or sequential behavioral data as obtained through mobile devices (e.g., Ecological Momentary Assessments, EMA). We follow a strongly theory-driven approach in methods development, deeply rooted in statistical and dynamical systems theory (DST). Given this strong backbone in theory-driven methodological development, our methods and algorithms are at the absolute forefront in the field of deep learning, raising the opportunity for completely novel insights into neural and behavioral processes which are beyond the scope of more ‘traditional’ machine/ deep learning methods. In particular, our DST perspective on time series analysis leads to ML/AI methods which yield *mechanistically interpretable models* of the nonlinear dynamics underlying observed time series. This is because, at some level, any known physical, biological, or societal system can be mathematically formalized through systems of coupled differential equations. As a result, our DST-driven approaches bear two major advantages compared to more traditional machine learning and statistical methods for time series forecasting: 1) they allow for *theoretically optimal* predictions of future time series events at the single subject level; 2) they yield, in addition, a *computational model of the underlying dynamics* that can be analyzed in depth, simulated, and systematically manipulated to gain mechanistic insights into the (neural or behavioral) processes that generate the observed time series. In essence, these methods return computational models of an individual’s behavior and/or brain dynamics that can be used for forecasting personal disease trajectories or probing possible interventions. Our framework for identifying the dynamical system underlying time series observations is based on deep Recurrent Neural Networks (RNNs) which are known to be universal approximators of dynamical systems (i.e., can emulate any other dynamical system).

In this context we also develop methods for *multi-modal data integration*: For instance, in a typical neuroscience experiment we may have simultaneous neurophysiological recordings and behavioral choice time series, or we may have EMAs on top of various sensor readings and other activity markers (e.g., number of in- & outgoing messages) sampled through mobile devices. We may also have non-sequential information, like genetic or metabolic profiles, that we would like to include in the form of (Bayesian) priors for informing computational model design.

While RNN/DST-based approaches are currently our major focus, we also develop other statistical time series methods, e.g. for unsupervised detection of spatio-temporal patterns in high-dimensional multivariate time series, or for detection of change (‘tipping’) points in time series, including paramtetric or bootstrap based significance tests.

2) Data analysis applications in biomedical research and practice

We apply our innovative methodology for diagnostic and prognostic (predictive) purposes mainly in psychiatry and neurology (see also *RG Computational Psychiatry*), but also in some other medical areas like the analysis of ICU patient time series for early sepsis detection (new project within the Heidelberg-Mannheim AI for Health Alliance). However, as alluded to above, our methods do not only serve prediction (diagnosis and prognosis), but are also aimed to gain mechanistic insight into underlying disease mechanisms and simulating the effect of potential interventions. For instance, based on generative RNNs we can infer dynamical models of an individual's brain from functional magnetic resonance imaging (fMRI), electroencephalographic (EEG), or magnetoencephalographic (MEG) measurements, or dynamical models of an individual's behavior from mobile sensors and Smartphone data. Such RNN-based dynamical systems models, trained on individual subjects, can then be used to forecast future behavioral or neuronal trajectories to enable early intervention. They can also be simulated to study the effect of potential pharmacological or behavioral treatments. In fact, we have developed a whole theory of how dynamical signatures derived using RNN-based algorithms (see above) from behavioral and/or neural data may relate to psychiatric phenotypes (Durstewitz et al., 2020).

In addition, we use formal reinforcement learning models of behavior to study learning processes in psychiatric conditions (see *RG Computational Psychiatry* for details).

3) Computational Neuroscience and computational dynamics of neural systems

At a more basic level, we are interested in how the brain implements computations, and how these computational processes may go astray in psychiatric conditions. Again, we approach this topic from the direction of DST (see #1), assuming that computations in the brain are implemented in terms of the underlying neural system dynamics, the most popular theoretical line in the area of computational neuroscience. Here, we rely mainly on the analysis of multiple single-unit (MSU) recordings and optogenetic data from rodents, with a focus on brain areas supporting higher cognitive functions, like the prefrontal cortex and hippocampus. While we use various ML/AI tools for analyzing spatio-temporal patterns, change points, or coding processes in MSU data, our major methodological vehicle for gaining insight into neural computations are again deep RNNs for reconstructing dynamical systems, as developed in our research focus #1 above. We also develop mathematical models of brain function at a more biophysical level, as well as statistical approaches for inferring such models directly from experimental observations like multi-cell recordings or neuroimaging data.

All these three research areas are deeply embedded in, and supported by, various currently running research clusters and consortia: We pursue our research focus #1, the more mathematical and methodological developments, mainly within the STRUCTURES excellence cluster at Heidelberg University (which has a focus on AI/ML methods in the natural sciences) and colleagues in there, further supported by currently two individual DFG grants with a focus on methods development. In area #2 we bring in our methods and algorithms into a number of national and international research consortia, including the DFG-funded TRR-265 inter-regional research consortium on regaining and loosing control in drug addiction, the state-funded local research alliance on AI for Health (ICU data), the EU-funded cluster IMMERSE (EMA data), and the state-funded 'AI real-world lab' on forecasting individual behavioral trajectories in mental health in younger adults.

In area #3 finally, we receive support by individual DFG grants, but mainly currently apply our methods within the DFG-funded FOR-5159 research group on neural mechanisms of prefrontal flexibility.

Key outputs 2020 – now

Especially with regards to our innovative ML/AI methodology (area #1), we made a number of major breakthroughs in the past two years, as collected in the set of publications below. To put this list into context for researchers not from the ML/AI field, I would like to remark that computer science and related theoretical disciplines follow a publication model very different from medicine and the natural sciences: Theoretical/ methodological papers in this area are usually published as open-access conference proceedings, for which there are numerous (on the order of hundreds) different potential outlets. These are indeed thoroughly peer-reviewed (usually by at least four referees), in a multi-stage process, full-length papers, with a typical length comparable to that of traditional peer-reviewed journals. The three most prestigious and selective deep learning conferences, perhaps comparable in reputation and impact with the Nature series journals in the natural sciences, are the International Conference on Machine Learning (ICML), Neural Information Processing Systems (NeurIPS), and International Conference of Learning Representations (ICLR). Our most important ML methods publications from the past two years are in one of these three (and all of them come with *software packages* available open-source at https://github.com/DurstewitzLab):

1) Monfared Z, **Durstewitz D** (2020) Transformation of ReLU-based recurrent neural networks from discrete-time to continuous-time. *Proceedings of the 37th International Conference on Machine Learning (ICML)*, PMLR 119:6999-7009.

→ Introduces a theoretical framework for translating between continuous and discrete time representations in RNNs, allowing for a tighter connection between RNN training in discrete time and real-world processes formulated in continuous time.

2) Schmidt D∗, Koppe G∗, Monfared Z, Beutelspacher M, **Durstewitz D** (2021) Identifying nonlinear dynamical systems with multiple time scales and long-range dependencies. *International Conference on Learning Representations (ICLR).* pp. 1-29.

→ Introduces a novel RNN training framework that eases training of RNNs on dynamical processes that evolve on many widely differing time scales (e.g., heart modulations within seconds vs. motivational states within days).

3) Kramer D*, Bommer P*, Tombolini C, Koppe G, **Durstewitz D** (2022) Reconstructing nonlinear dynamical systems from multi-modal time series data. *Proceedings of the 39th International Conference on Machine Learning (ICML)*, PMLR 11613–11633.

→ Developed within the TRR-265, this framework for the integration of many different types of data sources into the same RNN model was conceived with the multi-modal mobile data generated within the TRR-265 in mind.

4) Brenner M*, Hess F*, Mikhaeil JM, Bereska L, Monfared Z, Kuo P-C, **Durstewitz D** (2022) Tractable Dendritic RNNs for Reconstructing Nonlinear Dynamical Systems. *Proceedings of the 39th International Conference on Machine Learning (ICML)*, PMLR 2292–2320.

→ To improve interpretability of trained RNN models, this novel approach harvests ideas from computational neuroscience to profoundly reduce the dimensionality of trained models without compromising their expressivity.

5) Mikhaeil JM*, Monfared Z*, **Durstewitz D** (2022) On the difficulty of learning chaotic dynamics with RNNs. *Neural Information Processing Systems (NeurIPS)*.

→ Most complex biological and physical systems express chaotic temporal dynamics, but this makes it notoriously hard for RNNs to learn to represent these systems. In this very theoretical work we analyze why, and based on our mathematical insights derive a novel RNN training protocol.

Two other key publications were:

6) **Durstewitz D**, Huys, QJM, Koppe G (2021). Psychiatric Illnesses as Disorders of Network Dynamics. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(9), 865-876.

→ Here we flesh out our ideas on the dynamical systems perspective on psychiatric disorders and what it can offer in terms of diagnosis and treatment, along many specific examples from neuroscience and psychiatry.

7) Koppe G, Meyer-Lindenberg A, **Durstewitz D** (2020) Deep learning for small and big data in psychiatry. *Neuropsychopharmacology.* https://doi.org/10.1038/s41386-020-0767-z

→ Provides a comprehensive overview over deep learning methods and what they offer for the comparatively small datasets often encountered in psychiatry.

The three most important grant proposals for which received funding within this period for us were:

8) Du 354/15-1: Theoretical framework and bifurcation analysis for deep recurrent neural networks inferred from neural measurements (~427,000 €); individual DFG grant for further support of ML/AI methods development in neuroscience, 2022-2025.

9) Du 354/14-1: Reconstructing neuro-dynamical principles of prefrontal cortical computations across cognitive tasks and species; TP8 within DFG-funded FOR 5159 (627,200 €); major grant within a national neuroscience consortium, 2021-2025.

10) AI Living Lab for digital personalized mental health prevention in young adults (AI4Y) – together with Drs. Koppe, Reininghaus, & Krumm (our share: ~250,000 €).