RESEARCH ARTICLE

Deep learning and differential equations for modeling changes in individual-level latent dynamics between observation periods

Göran Köber1,2 | Raffael Kalisch3 | Lara M.C. Puhlmann3 | Andrea Chmitorz4 | Anita Schick5 | Harald Binder1,2

1Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
2Freiburg Center for Data Analysis and Modelling, University of Freiburg, Freiburg, Germany
3Leibniz Institute for Resilience Research (LIR), Mainz, Germany
4Faculty of Social Work, Education and Nursing Sciences, University Esslingen, Esslingen am Neckar, Germany
5Department of Public Mental Health, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

Abstract

When modeling longitudinal biomedical data, often dimensionality reduction as well as dynamic modeling in the resulting latent representation is needed. This can be achieved by artificial neural networks for dimension reduction and differential equations for dynamic modeling of individual-level trajectories. However, such approaches so far assume that parameters of individual-level dynamics are constant throughout the observation period. Motivated by an application from psychological resilience research, we propose an extension where different sets of differential equation parameters are allowed for observation subperiods. Still, estimation for intra-individual subperiods is coupled for being able to fit the model also with a relatively small dataset. We subsequently derive prediction targets from individual dynamic models of resilience in the application. These serve as outcomes for predicting resilience from characteristics of individuals, measured at baseline and a follow-up time point, and selecting a small set of important predictors. Our approach is seen to successfully identify individual-level parameters of dynamic models that allow to stably select predictors, that is, resilience factors. Furthermore, we can identify those characteristics of individuals that are the most promising for updates at follow-up, which might inform future study design. This underlines the usefulness of our proposed deep dynamic modeling approach with changes in parameters between observation subperiods.
1 INTRODUCTION

Deep learning techniques, that is, artificial neural networks with several layers, typically are associated with impressive performance on image data, such as in biomedicine (Esteva et al., 2017). However, there now also is a surge of proposals for combining deep learning with dynamic modeling, specifically using differential equations in a dimension-reduced latent representation obtained by neural networks (Chen, Rubanova, et al., 2018; Kidger et al., 2020; Rubanova et al., 2019). In our own work with an application from psychological resilience, we could use such techniques for identifying individual-level temporal trajectories of mental health in relation to external stressor load (Köber et al., 2022). We could thus quantify resilience of individuals and use regularized regression techniques for identifying resilience factors, that is, baseline characteristics that predict resilience. Yet, this application also led to the need to allow for changes in resilience in observation subperiods, that is, changes in the individual-level differential equation parameters.

We consider modeling with ordinary differential equations (ODEs) in a dimension-reduced latent space obtained by artificial neural networks, specifically variational autoencoders (VAE; Kingma & Welling, 2013). We focus on identifying the latent space dynamics due to their prominent role in resilience research—with subject-matter experts being primarily interested in identifying the individual dynamics and interplays of two latent concepts (mental health and stressor load)—but also due to their broad relevance across disciplines and research areas, as opposed to the often incomprehensible number of dynamics and interplays in the observed data space. Modeling such latent space dynamics with ODEs is appealing because of their natural way of dealing with irregularly spaced or entirely missing observations. Parameterizing ODEs with neural networks (Chen, Rubanova, et al., 2018; Rackauckas et al., 2019) allows us to combine tailored loss functions with the comprehensive modeling flexibility of differential equations. Most importantly, we introduce subperiod-specific ODE parameters and allow for discontinuous jumps. In more detail, we build upon Köber et al. (2022) and propose an approach for estimating separate sets of differential equation parameters for the two intra-individual subperiods. We acknowledge the potential similarity between the subperiods by tying each ODE parameter between the two subperiods together with a penalty term, where estimation is enabled by differentiable programming (Hackenberg et al., 2021; Innes et al., 2019). Moreover, we provide Julia code and showcase an analysis of simulated data in a Jupyter Notebook on GitHub.

Our method is applicable in real-world data situations with small datasets. In our case, small real-world data means 145 individual time series with, on average, 13 observations spanning several years, a low signal-to-noise ratio, missing observations, and unequal intervals between the observations. We utilize individual time series of a minimum of only four observations for training; only two observations are necessary to estimate the dynamics in a subperiod. Each time series is allowed to have entirely missing observations or unequal periods between the observations. To stabilize parameter estimation in this setting, we utilize structural information about the study (e.g., regarding the subperiod window size) to support the training mechanism. Specifically, our choice of two subperiods is motivated by the design of the MARP (Mainzer Resilience Project) study, where potential resilience factors (predictors) are repeatedly measured approximately every 1.75 years in laboratory battery time points (B0, B1, ...). In contrast, repeated online measures of stressor load and mental health problems—serving to determine the individual stressor reactivity in continuous time with ODEs—are regularly conducted at higher frequency (every 3 months, time points T0, T1, ...) at and between the battery time points (Kalisch et al., 2021). At the current stage of ongoing data collection, we use B0 to predict stressor reactivity between B0 and B1 (subperiod 1: T0–T6) and B1 to predict reactivity afterward (subperiod 2: T6 and later). We subsequently demonstrate how to use L1-regularized regression, that is, the lasso, to identify predictors of resilience in the subperiods and thus characteristics of individuals where follow-up measurement might be valuable.

While there are other differential equation approaches for modeling dynamics as fixed effects and normally distributed individual deviations from these fixed effects in psychology (Driver & Voelkle, 2018; Montpetit et al., 2010), these do not allow for intrapersonal changes in these dynamics. When not requiring modeling in a latent representation, that is, when fitting dynamic models at the observed level, there are many potentially useful regression modeling frameworks (Putter & van Houwelingen, 2017; Rizopoulos, 2012), and correspondingly various regression modeling approaches could be considered for updates of parameters. For example, L1-regularized regression could be considered (Inan & Wang, 2017;
Schelldorfer et al., 2011, 2014; Wang et al., 2012) or coupling the likelihood of multiple points in time (Schmidtmann et al., 2014; Zölle et al., 2016), as proposed in our own work. Moreover, there are many spline-based approaches that allow for modeling of changes in the dynamics (Bringmann et al., 2017; Hong & Lian, 2012; Meng et al., 2021; Refisch et al., 2022; Wang et al., 2007). Kamalabad and Grzegorczyk (2021) suggested to learn which temporal segments should be coupled, allowing the parameters to be informed by the previous segment, or uncoupled where the parameters of the segment can be estimated without previous information. While this perspective opens directions for future development, we here aim for deriving prediction targets from predefined temporal segments. Kim and Nelson (1999) suggested to learn categorical transitions between qualitatively different dynamics regimes, which is implemented in R with the package dynr (Ou et al., 2019). While we could extend our model to individual (changing) setpoints (Chen, Chow, et al., 2018), we assume that the structure of our system of differential equations remains stable across the study period. Rather than learning transitions between differing ODE systems, we allow the parameters to vary between individuals and intra-individual subperiods. To our knowledge, there is no approach that allows to simultaneously identify mappings to a latent space for dimension reduction and changes in individual-level dynamics with differential equations and artificial neural networks without making an assumption about the distributions of the inter-individual differences.

In the following, we briefly describe the psychological resilience application that motivates our methods development in Section 2 before describing the proposed approach in Section 3 and illustrating it with results from the application in Section 4. We subsequently discuss the potential usefulness of our approach in other application settings and potential further extensions.

2 A PSYCHOLOGICAL RESILIENCE APPLICATION OF LATENT DYNAMIC MODELING

Psychological resilience is the maintenance or rapid recovery of a healthy mental state during and after times of adversity (Kalisch et al., 2017). One element of the definition is that both mental health problems (mh) and stressor load (sl) can change over time and may even do so permanently. Influential resilience studies (Bonanno et al., 2011) investigated how mental health changes in response to one single potentially traumatic life event and found groups of similar individual mental health trajectories such as resilient (showing stably good or improving mental health in the months or years after the event) or vulnerable (showing stably poor or worsening mental health). These studies implicitly assume that the observed temporal changes in mental health are due to only a single stressor event and that individual differences in the mental health trajectory can be explained by some baseline individual characteristic. However, most individuals are continuously exposed to more or less severe stressors. These may include macrostressors (severe life events) but also more “mundane” microstressors, or daily hassles (Hahn & Smith, 1999), which also have an impact on mental health (Kalisch et al., 2021; Serido et al., 2004). In short, investigating resilience should ideally involve repeated longitudinal measurements of stressors and mental health (Bonanno et al., 2011; Kalisch et al., 2015, 2017, 2019); such longitudinal studies pose manifold problems for data collectors and analysts.

The Mainz Resilience Project (MARP) is an ongoing study that started in 2016 with a planned study duration per participant of 7 years. MARP is conducted by the University Medical Center Mainz and the Leibniz Institute for Resilience Research (Kalisch et al., 2021). Participants were recruited in a critical life phase aged 18 to 20 at study inclusion with a prehistory of critical life events.

The minimum requirement to be included in the longitudinal modeling is to have at least four observations. We efficiently use all individual time series with at least four observations for training the ODEnet. Suppose subjects have less than two observations in a subperiod but fulfill the just-mentioned minimum longitudinal requirement. In that case, we use them for training the ODEnet without penalizing the subperiod differences (using an $1f$-condition the loss function utilizing differentiable programming). This yields $N = 145$ subjects with on average 13.3 observations ($\text{min} = 4$, $\text{max} = 23$, $\text{sd} = 4.8$). Two or more observations in both subperiods are available for $N_{\text{sp}2} = 110$ subjects, here our intra-individual subperiod penalization is applied. Incomplete predictor data leads to a further decline in sample size for the lasso analysis, leaving us with $N_{\text{sp1lasso}} = 107$ and $N_{\text{sp2lasso}} = 77$. The total number of observations which are used to train the VAEs is $N_d = 2,052$ and $N_{\text{mh}} = 2,087$; the VAEs training data are not affected by the minimum longitudinal requirement. The majority of the longitudinal measurements for dynamic modeling of resilience are gathered via an online assessment roughly every 3 months. Each observation comprises $p = 28$ mental health items (GHQ-28; Goldberg et al., 1997) and $e = 58$ daily hassle items (MIMIS battery; Chmitorz et al., 2020). One item is a question to either rate 28 different mental health problems on a scale from 0 to 3 or indicate the days last week (0–7) where certain unpleasant circumstances

---

KÖBER ET AL.
This diagram shows the intra-individual learning of potentially changing dynamics schematically. We parameterize our system of ODEs with a feed-forward neural network (ODEnet), which receives summary statistics of the respective subperiod (sp1 or sp2) as inputs. The critical advancement of our algorithm is the coupled fitting of dynamic models for two observation subperiods for each respondent. Subperiod 1 (sp1) covers all measurements from the baseline battery measurement (B0) to the repeated resilience factor testing batteries (B1). Subperiod 2 (sp2) comprises all observations of mental health and stressor load after that. We indicate the solution of sp2 with a dotted line.

(stressors) were experienced. Respondents in a first subperiod of 1.75 years. In addition to longitudinal measurements, there is an extensive baseline battery of potential resilience factors for characterizing individuals, where we currently consider questionnaire data. Assessment of the baseline battery is repeated after 1.75 years, that is, at the beginning of the second subperiod (sp2).

3 | METHODS

We first provide a brief overview of the proposed approach, before subsequently describing its components in detail. In the first step, the VAEs are used for separate dimensionality reduction of mental health and stressor items. Longitudinal summary statistics of the resulting time series in the latent space are fed into a second neural network—the ODEnet—which provides individual-level ODE parameters as output. The ODE parameters are obtained for each individual, separately for the two subperiods; this is the key advancement compared to Köber et al. (2022) and depicted in the schematic illustration of Figure 1. We introduce individual-level dependency by penalizing the sum of each ODE parameter between the subperiods (but not the initial conditions). Thereby, the ODEnet is trained to find ODE parameters that best describe the dynamics of the respective subperiod and, at the same time, enforce some correlation between parameter sets from the same respondent, reflecting that the data is provided from the same respondent.

3.1 | Dimensionality reduction per time point with VAEs

We choose VAEs (Kingma & Welling, 2019) as a popular deep generative model to find lower dimensional latent distributions \( q_\phi(z|x) \) making use of variational inference (Blei et al., 2017). Intuitively, VAEs learn distributions in the latent space with an encoder net; simultaneously, the decoder net learns to reconstruct samples from the latent space distributions similar to the inputs. This comparatively easy structure is very extendable, for example, to hierarchical data structures (Vahdat & Kautz, 2020) or dynamical systems (Girin et al., 2021); we prefer an only slightly adapted variant of the vanilla VAE in combination with a dynamic model due to our limited sample size. At the same time, VAEs offer many potential
downstream applications, for example, multiple imputation (Nazábal et al., 2020), sample size calculation (Treppner et al., 2021), and anomaly detection (Li, Yan, et al., 2020), making them an appealing deep generative model.

Let $x_i^{mh} = (x_{i,1}^{mh}, \ldots, x_{i,p}^{mh})$ be a vector of $p$ items of $mh$ and $x_i^{sl} = (x_{i,1}^{sl}, \ldots, x_{i,e}^{sl})$ be a vector of items of $sl$ and $i \in \{1, \ldots, N_{sl}\}$ or $i \in \{1, \ldots, N_{mh}\}$, respectively (see Section 2). The latent variables of sl or mh are denoted as $z_i^{sl}$ and $z_i^{mh}$ for each realized observation $i$. Then, the VAEs are trained by maximizing the evidence lower bound of the marginal log-likelihood

$$
\log p(x_1, \ldots, x_N) = \sum_{i=1}^{N} \log p(x_i)
\geq \sum_{i=1}^{N} (E_{q(z|x)} \log p(x_i|z) - KL(q(z|x)_i||p(z)))
$$

where the first term on the right-hand side is the expectation of the log-likelihood of $x_i$ given $z$ with respect to $q(z|x)$. The Kullback–Leibler divergence ($D_{KL}$) penalizes deviations of the posterior from the prior. Temporal dependence between the observations is established by the system of ODEs.

VAEs comprise an encoder and a decoder neural network. The purpose of training the encoder (aka inference or recognition model) is to find good variational parameters $\phi$ that parameterize the distributions $q_{\phi}(z|x)$ in the latent space, in a way the decoder can reconstruct samples of that latent distribution that resemble the inputs $x_i$. In our case, the former is achieved by two separate EncoderNets

$$
(\mu_i^{mh}, \log \sigma_i^{mh}) = \text{EncoderNet}_{\phi_{mh}}(x_i^{mh})
$$

$$
(\mu_i^{sl}, \log \sigma_i^{sl}) = \text{EncoderNet}_{\phi_{sl}}(x_i^{sl}),
$$

which yield the latent distributions

$$
q_{z}(z_i^{mh}|x_i^{mh}) = \mathcal{N}(\mu_i^{mh}, \log \sigma_i^{mh})
$$

$$
q_{z}(z_i^{sl}|x_i^{sl}) = \mathcal{N}(\mu_i^{sl}, \log \sigma_i^{sl})
$$

for each complete observation $i$. The two separate DecoderNets($\theta$) then take samples of $q_{\phi}(z|x)$ and are trained to increase the log-likelihood of the inputs given samples of the posterior distributions. Please note, while the expected value of the posterior distribution $\mu$ captures the position of the observation in the latent space, the variance $\sigma$ captures the uncertainty that is related to this mapping (see also Kingma & Welling, 2019).

For computation, we plug in the Poisson log-likelihood and the closed form of $D_{KL}$ for a Gaussian prior and posterior. Thereby, training objective of our VAEs become

$$
\text{Loss}(\theta, \phi; x_i) = -\frac{1}{2} \sum_{i=1}^{N} (1 + \log((\sigma_i)^2) - (\mu_i)^2 - (\sigma_i)^2)
\underbrace{D_{KL}(\phi)}_{\text{KL divergence}}
+ \sum_{i=1}^{N} (\lambda_i - x_i \times \log(\lambda_i))
\underbrace{\text{Poisson reconstruction error (}\theta\text{)}}_{\text{Poisson reconstruction error (}\theta\text{)}}
+ \lambda_{VAE} \times (\sum_{i}^{\phi} + \sum_{i}^{\theta})
\underbrace{\text{Weight regularization (}}_{\text{Weight regularization (}}\text{VAE}}
$$

where $\mu$ and $\sigma$ are the mean and standard deviation of the latent distribution depending on the observed values $x_i$. The expected value of the Poisson distribution of each item is denoted as $\lambda_i$ depending on the observation. To prevent overfitting, we regularized the encoder ($\phi$) and decoder ($\theta$) weights of both VAEs with $\lambda_{VAE}$. 
We quantify the reconstruction error assuming a Poisson distribution because of the observed data’s value ranges and empirical distributions. Both value ranges are a subset of the natural numbers, including zero $\mathbb{N}_0$ (for stressor load [0; 7] and mental health [0; 3]). Furthermore, the question wording has a clear count character (“In the past 7 days, on how many days did the previously mentioned situation occur?”; see Section 2) for stressor load. In combination with the data distribution, the $D_{KL}$ term favors mappings to the latent space where the lower end of the value range is close to zero. We assured positive correlations with the sum scores (as you would do with more conventional latent variables such as factor scores) to facilitate the interpretation of the ODE parameters (see Section 3.2 and Figure 1 in the Supporting Information S1). The Poisson distribution, however, does not perfectly match the observed data since we found moderate levels of overdispersion for some items. Yet, the Poisson distribution avoids an additional parameter, which might be difficult to determine, by coupling the expected value and the dispersion. In the final layer, we used a ReLU activation function to strictly pass nonnegative values to $\lambda_i$. We used the $\tanh$ activation functions in the middle layers.

Applying nonlinear VAEs for dimensionality reduction invites concerns about the validity of the lower dimensional representations. We conduct a correlation analysis comparing the means of the VAEs with conventional methods for reducing the number of dimensions and find strong correlations ($r > 0.85$; see Figure 1 in the Supporting Information S1). Accordingly, the VAEs—as a novel way of finding distributions in the latent space—bear sufficient similarity to other dimensionality reduction techniques and can be interpreted similarly to conventional approaches.

### 3.2 ODEs to model the trajectories of sl and mh in the latent space

We use ODEs to couple the latent representations of mental health problems (mh) and stressor load (sl) over time and allow for separate intra-individual parameter sets for periods between the resilience factor testing batteries (B0, B1), providing potential predictors of the individual dynamics. The exact design of such an ODE system is a crucial modeling decision since it governs how each component changes and requires domain expertise. We were able to model a slightly more complicated system of ODEs than Köber et al. (2022) due to the increased number of subjects and observations. Specifically, we use

$$\frac{df_{z_{mh}^i}}{dt} = \eta_{i,1,sp} \times z_{sl}^{mh} + \eta_{i,2,sp} \times z_{sl}^i,$$

$$\frac{df_{z_{sl}^i}}{dt} = \eta_{i,3,sp} \times z_{sl}^{sl} + \eta_{i,4,sp} \times z_{mh}^{sl},$$

where changes in $z_{mh}^{sl}$, the latent trajectory value of mental health problems (mh) of respondent $i$ at $t$ with $i \in \{1, ..., N\}$ and $t \in \{1, ..., T\}$, and $z_{sl}^{sl}$, the latent trajectory value of stressors exposure (sl), are driven by their own current value. Additionally, $z_{mh}^{sl}$ is allowed to change in response to $z_{sl}^{sl}$ and vice versa. Subscript $sp$ in $\eta_{i,1,sp}$ indicates that each individual set of parameters differs intra-individually between sp1 and sp2.

Negative values for $\eta_{i,1,sp}$ and $\eta_{i,3,sp}$ effectively realize system-inherent damping (Boker et al., 2010), where high sl and mh values are more quickly driven back to low values in the latent space. Thus, a high negative $\eta_{i,1,sp}$, in particular, reflects good recovery from mental health problems (since the majority of observations are mapped into the positive valued latent space or close to zero). This can be understood as one facet of stressor reactivity, where possible surges in mental health problems tend to be short-lasting. Positive values for $\eta_{i,2,sp}$ realize the adverse effect of sl on mh. A low positive $\eta_{i,2,sp}$ value thus reflects the low responsivity of mh to sl in the first place, as the second element in our equation system that, overall, describes an individual’s stressor reactivity. $\eta_{i,4,sp}$ allows for the opposite direction and expresses that people with high mh report more sl in the future. This is deemed plausible, both because mental health problems are stressors in their own right that can induce further adverse reactions and because mentally ill persons may generate, or be confronted with, more conflicts or other types of adversities (Gerin et al., 2019). However, key parameters for the interpretation of an individual’s resilience status are $\eta_{i,1,sp}$ (recovery) and $\eta_{i,2,sp}$ (reactivity).

We acknowledge that stressor load does not necessarily develop smoothly in continuous time but can change abruptly. This reflects the notion that stressor load is only partly driven by the continuous properties of the ODE system (i.e., damping and mental health). But mainly reflects exogenous forces, that is, the sudden occurrence or absence of stressors that lead to abrupt changes in the stressor load. Such a combination of continuous and discrete dynamics has been coined hybrid dynamical systems elsewhere (Van Der Schaft & Schumacher, 2000). It offers a wide range of practical applications.
TABLE 1  Mean squared errors of the out-of-sample prediction task. The last observation in the second subperiod was masked during training and predicted afterward. The ODEnet (our approach, highlighted in bold) has the lowest MSE of all other methods (i.e., a linear mixed model (LMM, see specification in Equation 8), the mean of all previous observations, last observation carried forward (LOCF)).

| Latent variable                    | ODEnet | LMM (tuned) | LMM | Mean | LOCF |
|------------------------------------|--------|-------------|-----|------|------|
| Mean of VAE (μ_i^mh)              | 0.393  | 0.417       | 0.438 | 0.43  | 0.514 |
| Mean scores of mental health       | 0.195  | 0.198       | 0.202 | 0.2   | 0.218 |

(see Jia & Benson, 2019, and the references therein). Technically, we realize this by updating the integrator of stressor load f_z_{isl} to the latent value z_{isl} of the actual observation x_{isl} at the precise point in (study) time.

The benefits of such an ODE system compared to discrete-time models like regression are crucial for analyzing the data from the MARP study. Most importantly, differential equations take all available information at the precise time into account. Thereby, irregular sampling intervals and entirely (or partly) missing observations are dealt with by the properties of our dynamical system (assuming noninformative missingness).

3.3 Finding individual and potentially changing ODE parameters with the ODEnet

We parameterize our system of ODEs with the ODEnet, that is, a three-layer feed-forward neural net with a comparatively small number of trainable coefficients. The benefit of having another neural network here stems from its seamless combination with differential equation solvers (Rackauckas et al., 2019). We utilize this flexibility by incorporating jumps and several regularization terms. Most notably, we penalize intra-individual differences of the ODE parameters between subperiods.

Compared to Köber et al. (2022), the critical advancement of our method is the extension of the ODEnet to learn potentially changing ODE parameters \eta_{i,ssp}. The key idea is depicted in Figure 1. The ODEnet is trained by minimizing the Loss_{ODE}(\tau, \eta_{i,ssp}) and provides a two-element vector of ODE parameters, each of length \mathcal{S} \in \{1, \ldots, 4\}, for both subperiods

\eta_{i,ssp} = (\eta_{i,ssp1}, \eta_{i,ssp2}) = \text{ODEnet}_{\tau}(x_i^{mh}, x_i^{sl})

as well as the initial conditions for both trajectories for each subperiod (in total six values per subperiod) with the observed items x_{i}^{mh} and x_{i}^{sl} as inputs and the trainable parameters \tau. Learning the initial conditions (ICs) is a step forward compared to Köber et al. (2022) because we do not regard the first observation of mental health as an exceptionally reliable measurement anymore. The previous fixation of the ICs to the mapped values was a problematic assumption because ICs govern the course of the trajectory; instead, it is left to gradient descent to find ICs that allow learning the trajectory optimally.

The ODEnet internally calls the EncoderNet (see Equations 1 and 2) which maps the observed values into the latent space. Subsequently, we calculate several summary statistics of the intra-individual subperiod in the latent space (e.g., integrals, first and last observations, and differences; see Table 1 in Köber et al. 2022). It was left to gradient descent to find a good combination of these summary statistics to minimize Loss_{ODE}(\tau, \eta_{i,ssp}). All summary statistics are required to be computable with only two observations from mh and sl (which can be reported at any point in time, not necessarily during the same observation). The main purpose of the ODEnet is to minimize the sum of the squared difference of the trajectory f_{z_i}(t, \eta_{i,ssp}) and the mean of the latent space distribution \mu_i at the precise point in study time t depending on period-specific individual ODE parameters \eta_{i,ssp}. Accordingly, the training objective of the ODEnet is

Loss_{ODE}(\tau, \eta_{i,ssp}) = \sum_{i=1}^{N} \sum_{t=1}^{T} (f_{z_i}(t, \eta_{i,ssp}) - \mu_{z_i})^2 + \lambda_{sp} \times \sum_{s=1}^{4} (\eta_{i,ssp1} - \eta_{i,ssp2})^2 + \lambda_{ODEp} \times \sum_{s=1}^{4} \eta_{i,ssp} + \lambda_{ODEnet} \times \sum_{i} \tau.

(7)
Importantly, the period-specific individual parameters sets within $\eta_{i,sp}$ are tied together by penalizing the squared difference of each ODE parameter in the two individualized parameter sets $\eta_{i,sp1}$ and $\eta_{i,sp2}$. This bond of ODE parameters across the subperiods reflects the dependence of the dynamics between the intra-individual subperiods. Furthermore, this tying helps to prevent overfitting of the ODE parameters to a certain subperiod with, for example, only a few unusual measurements. The hyperparameter $\lambda_{sp} \in \mathbb{R}_{>0}$ allows to tune the strength of this connection. We additionally penalize both the sum of the ODE parameters with $\lambda_{ODEP}$ as well as the weights of the ODEnet with $\lambda_{ODEnet}$ to avoid overfitting. The concrete values of the hyperparameter $\lambda_{sp}$ can be decided based on subject-matter considerations, for example, by imposing stronger similarity between intra-individual subperiod dynamics for subject areas which are known for evolving slowly, or by optimization criteria, for example, cross-validated mean squared error (MSE) as usual in lasso analyses (Hastie et al., 2015). Please note, the differences of the ICs are not penalized across subperiods, although the ODEnet also provides them.

We use a ReLU activation function in the middle layer and no transformation in the final layer. Using no activation function in the last layer does allow the ODEnet to find the parameters of the dynamical system freely; put differently, it enables the ODEnet to provide the full numerical range of possible ODE parameters. Learning the ICs overcomes the limitation of Köber et al. (2022), who regarded the first observation as ground truth.

To increase training stability, we initialized the ODEnet with very small weights. Flexible dynamical modeling is provided by DifferentialEquations.jl (Rackauckas & Nie, 2017) and differentiable jointly with neural nets via DiffEqFlux.jl (Rackauckas et al., 2019). To deal with unit nonresponse, Loss $\text{Loss}(\theta, \phi; x, \lambda_{VAE})$ and Loss $\text{ODE}(\tau, \eta)$ are only evaluated at actual measurement time points. All neural networks were trained with Flux.jl (Innes et al., 2018) and the Adam optimizer (Kingma & Ba, 2014). We used this website (LeNail, 2019) as a basis for drawing the ODEnet in Figure 1.

3.4 | A two-stage lasso approach to repeated baseline updates

The MARP study design schedules the laboratory battery time points (B0, B1, …) every 1.75 years, where comprehensive investigations are conducted, and numerous potential resilience predictors will be measured again (see Section 2 for more information). Given time and money constraints, one might ask which predictors are worth updating the most. The underlying assumption here is that there is a selective “aging” of predictors—affecting some more than others—making their updates more valuable from a prediction point of view.

We assess the capabilities of the incoming predictors to improve the prediction beyond the older (but potentially still relevant) subject information with a variant of the lasso tailored to this data situation. The particularity of these data is that an extensive (and expensive) battery of fMRI data, behavioral data, and questionnaires are repeated at regular intervals. In the meantime, several samples of key resilience variables (mental health problems and stressor load) are drawn. To be able to learn about the usefulness of the incoming predictors, we add an additional weight vector $w$ to the standard lasso training criterion. The optimal parameters for the incoming predictors are

$$\hat{\beta}_{sp2} = \arg\max_{\beta} \left\{ \frac{1}{2} |y - X \times \beta|^2 + \lambda \sum_{j=1}^{p} w \times \beta \right\}.$$  

This approach has been already suggested to the cross-sectional data settings before (Zou, 2006) and can be implemented simply using the established R package glmnet (Friedman et al., 2010); this circumstance fosters reusability of the approach, also independently of the estimation of dynamics with the ODEnet as suggested above. In the first step of our two-stage approach, the lasso is trained only with predictor data of the initial baseline observation (B0) to predict the learned dynamics of the first subperiod. Since no previous knowledge is assumed, the penalty factor is set to its default ($w = 1$) for all potential predictors. In the second step, both the initial and latest baseline information serve as potential predictors of the dynamics of the second subperiod. We use the previous lasso analysis and penalize the initial data from B0 with $-\log(VIF)$ while we penalize incoming predictors uniformly with the default.

We use the variable inclusion frequencies (VIF) because single lasso analyses might be unstable regarding the selected variables. For this reason, Wallisch et al. (2021) suggest using resampling to determine model stability (see also Heinze et al., 2018, and Sauerbrei et al., 2015). More precisely, we resampled a proportion of $m = 0.8$ and repeated the lasso analysis 1,000 times, using sixfold cross-validation to determine the strength of the lasso penalty once.

More detailed, assuming that each baseline observation provides the same number of variables $v$, and there are less individuals who met the minimum requirements $N_{sp1, lasso}$ is usually larger than this number in the second sub-
FIGURE 2 Exemplary comparison of the dynamical systems of two respondents with high (top row) or low (bottom row) intra-individual parameter difference trained with different hyperparameters (columns). The dots represent the means of the latent space distributions—learned by the VAEs—separately for mental health problems and stressor load. The lines show the solutions of the ODE system for each subperiod; we indicate the ODE solution of the second subperiod (sp2) with dotted lines. Subplot A shows the effectiveness of the intra-individual deviation penalty and compares the results of an ODEnet trained with low (\(\lambda_{sp} = 0.0001\)) and high (\(\lambda_{sp} = 0.05\)) hyperparameter choices. The top row shows a respondent with considerable change in the parameters, where a similar stressor load leads to high mental health problems (the last two observations). The increase in mental health problems is visibly mitigated, although still present, when the ODEnet is trained with \(\lambda_{sp} = 0.05\) (right column of subplot A). The bottom row shows a respondent with a low change in ODE parameters. Accordingly, increasing \(\lambda_{sp}\) (from left to right) does not affect the intra-individual comparison strongly. Subplot B shows the solutions of the ODE system when the ODEnet is trained in prediction mode, that is, the last observation is omitted from the training data (indicated by a nearly transparent last observation) and a moderate \(\lambda_{sp} = 0.0155\). We use this hyperparameter configuration for the rest of the Results section.

period \(N_{sp2,lasso}\). The \(v \times n_{sp1}\) matrix \(X_{sp1}\) of potential resilience factors is accompanied by the weight vector \(w_{sp1} = (w_{sp1,1}, \ldots, w_{sp1,v}) = (1, \ldots, 1)\) when predicting an indicator of the dynamics of the first subperiod \(y_{sp1}\). For the second subperiod, however, we stack both baseline information on top of each other, accordingly, \(X_{sp2}\) is \(p \times N_{sp2,lasso}\), where \(p = 2 \times v\) (in case \(v_{sp1} = v_{sp2}\)). The weight vector of the second subperiod can be expressed as

\[
w_{sp2} = \begin{cases} 
-\log(VIF(X_r)), & \text{for } r \leq \frac{p}{2} \\
1, & \text{for } r > \frac{p}{2}
\end{cases}
\]

with \(r \in \{1, \ldots, p\}\).

4 RESULTS

In this section, we start with showing two exemplary respondents and contrast the effects of different intra-individual difference penalization (\(\lambda_{sp}\)) choices. Afterward, we demonstrate the suggested procedure to derive prediction targets and report the VIF of potential predictors. Lastly, we evaluate the out-of-sample prediction performance, omitting the last observation while training in prediction mode (see Figure 2B). We used the same hyperparameters (documented in the code provided as Supporting Information S2) to train the ODEnet throughout this section, except for Figure 2A, with new weight initializations on the respective datasets.
Figure 3: Variable inclusion frequencies (VIF; in %) of the lasso analysis with several prediction targets. This heat map provides insights into which battery measures (x-axis) predict the targets (y-axis) derived for this longitudinal focus. The lower rows of the y-axis depict the two subperiods or, more precisely, $\eta_i,2,sp1$ and $\eta_i,2,sp2$ from the ODEs (see Equation 5), which captures the individual gradient of mental health in response to stressor load in the dynamical system. Importantly, sp1 was predicted solely with B0 (initial lab visit at the beginning of the study). The second subperiod sp2 was predicted with both B0 and B1 (the latter is the first repetition of the initial lab visit's testing battery around 1.75 years later), albeit with unequal penalty weights. This plot indicates that there are four frequently picked (i.e., VIF $\geq 0.9$) predictive variables in B0 (see lowest row). Due to their high VIF, these variables are penalized with zero/very small $w_{sp2}$ (see the Methods section for more details). Accordingly, they are likely included in the prediction of sp2 dynamics as well. Therefore, in the middle row, we show only the B1 predictors. Regarding potential resilience factors in the second testing battery B1, there is a mixed picture. While updates of frequently picked variables are chosen rather seldom (see middle row), some other predictors—chosen less frequently in sp1—are indicated as valuable information for predicting sp2 dynamics. The upper row of Figure 3 shows how well the difference measure can be predicted with B0 information. There is a visible overlap in the selected parameters to the sp1 and sp2 lasso analysis, which we interpret as evidence for the suitability of both approaches to quantify resilience dynamics. We find one resilience factor with VIF $\geq 0.6$ throughout all prediction targets. We deliberately do not show or discuss the constructs that we found to predict the individual dynamics since we focus on methods development here; publications with in-depth analyses of the constructs with this data are in preparation by the subject matter experts in our consortium.

4.1 Exemplary comparisons of individual dynamical systems with differing subperiod deviation penalty terms

Figure 2 exemplary compares the dynamical systems of two individuals (rows) trained with different hyperparameters (columns). In subplot A, we visualize the effectiveness of increasing the intra-individual difference penalization term $\lambda_{sp}$ (columns). In subplot B, we show the solutions of our ODEnet, trained in prediction mode (masking the last observation). As in Figure 1, the y-axis shows the position in the latent space and the x-axis shows time. The expected values of the latent value distributions learned by the VAEs are expressed as dots. The trajectories, governed by the ODEs, are depicted as lines. Mental health problems (mh) are shown in blue, while stressor load (sl) is red; dotted lines indicate the dynamics in the second subperiod (sp2).

4.2 Predicting changes in resilience with updated battery measures

We investigate which battery parameters predict changes in resilience in two related ways. First, we extend the logic of the artificial stress test to our perspective on changing dynamics. In short, the artificial stress test utilizes the individualized ODE parameters for each subperiod to predict mental health in response to considerable stressor load after predefined time points. With that, the artificial stress test provides us with a continuous prediction target (see Köber et al., 2022, for a more detailed explanation and visualization of the artificial stress test). We enhance and adapt the artificial stress test here by sampling from the dimensions in the latent space and predicting the difference between both subperiods. We show the inclusion frequencies for this measure in the first row of Figure 3, where we also discuss the findings. More detailed, we introduce a considerable stressor load; which is reflected in an initial condition $IC_{sl} = 2$ and sample the initial conditions of mental health $IC_{mh}$ several times from $q_\phi(z^{mh}_{i,1} | x^{mh}_{i,1})$ (see Equations 3 and 4), that is, the latent distribution gave the first observation $x^{mh}_{i,1}$. Then, we plug in the individually learned subperiod-specific ODE parameters $\eta_{i,sp1}$ and $\eta_{i,sp2}$ (see Equation 5) to solve the ODE system for each subperiod. Lastly, we pick the difference of the mean of all sampled mental health trajectories at $t = 1$ as the prediction target. This extension of the artificial stress test signals which predictors indicate changing resilience patterns in the course of the study with information available already at the beginning of the study.
Second, we investigate the question of which parts of the repeated predictor battery are particularly worth updating. This question emerges in many longitudinal situations, usually when potential predictor data are gathered more than once but less often than the information of key interest (in our case, mental health and stressor load). Therefore, we predict the ODE parameter $\eta_{i,2,sp}$ directly—which captures how the level of stressor load influences the gradient of mental health problems—with a two-staged lasso approach (described in the Methods section) and discuss the results in Figure 3. Please note that while only information of B0 is included in the analysis of sp1, variables of B0 and B1 are included in the analysis of sp2. However, due to the very low penalization weight $w$ of the frequently selected variables (and reversely the very high weight of the nonselected variables), they are almost certainly (not) selected, which renders their visualization pointless. Since variable selection approaches are known for being unstable, we choose VIF (Wallisch et al., 2021) for both as the decisive criterion for our two-staged procedure determining the weights and for visualizing the results.

4.3 Performance evaluation and comparison to simpler methods

In the following, we compare the prediction performance of the ODEnet to simpler competitors. Therefore, we trained the ODEnet in prediction mode and forecast the last successful mental health observation as visualized in Figure 2 subplot B. Masking the last entire observation results in a slightly smaller dataset with $N_{sp2,\text{train}} = 104$ subjects (compared to $N_{sp2} = 110$). As described in Section 2, we use all subjects meeting the minimum longitudinal requirements for training the ODEnet. We evaluate the performance with the MSE and compare it to other methods (described below). The ODEnet (our approach) has the lowest MSE in this comparison (see Table 1). We expect this gap to widen with more training data becoming available as the MARP study continues.

We train the ODEnet to predict the mean scores with the same hyperparameters as the ODEnet for the means of the VAE ($\mu_i^{mh}$). This additional prediction performance evaluation demonstrates the modularity of our approach, that is, the two components (dimensionality reduction and trajectory estimation) can be used separately if desirable (as demonstrated in our Jupyter Notebook in the GitHub repository).

An autoregressive linear mixed model (LMM) is a more sophisticated competitor of our ODEnet. We modeled mental health of individual $i$ at $t$ ($mh_{i,t}$) with

$$ mh_{i,t} = \beta_{0,i} + \beta_{1,i} \times mh_{i,t-1,i} + \beta_{2,i} \times sl_{i,t-1,i} + \beta_{3,i} \times t_{i} + \beta_{4,i} \times sp_{i} + e_{i}, $$

(8)

where $\beta_{0,i} = \gamma_{00} + u_{0,i}$ as random intercept, $\beta_{1,i} = \gamma_{10} + u_{1,i}$, $\beta_{2,i} = \gamma_{20} + u_{2,i}$, $\beta_{3,i} = \gamma_{30} + u_{3,i}$, and $\beta_{4,i} = \gamma_{40} + u_{4,i}$ as random slopes. The individual variability $u_{0,i}$, $u_{1,i}$, $u_{2,i}$, $u_{3,i}$, and $u_{4,i}$ is assumed to be distributed normally with $u_i \sim N(0, \sigma^2_u)$ around the fixed effects. This flexible regression model had a relatively poor out-of-sample prediction performance. Therefore, we optimized it by removing some random effects ($\beta_{1,i}$, $\beta_{2,i}$, and $\beta_{4,i}$; LMM (tuned)). We further compare the prediction performance of our approach to the mean of each subject’s time series (mean), and the last observation carried forward method (LOCF).

5 DISCUSSION

Recently, there have been several proposals for deep dynamic modeling, where differential equations are used on latent representations obtained by VAEs. Motivated by an application from psychological resilience research, we removed the assumption of constant dynamics throughout the whole study period, implicit in these approaches. Instead, we showed how individual-level dynamic models could be fitted for two individual subperiods and introduced a dependency between by a penalization term, which made our approach feasible also for a dataset with a relatively small number of individuals.

We derive a suitable prediction target that captures changes between subperiods, and we suggest a two-stage lasso approach for identifying valuable predictor updates of subsequent lab visits. Our variable selection efforts regarding potentially changing dynamics allow identifying promising predictors in repeatedly assessed lab batteries as well as pruning the repeated batteries, thus potentially saving time and money. We find that four resilience factors in the initial testing battery are particularly promising for predicting potentially changing dynamics as well as an ODE parameter capturing stessor reactivity in the first subperiod. While we pick one variable update from the second lab visit in the second subperiod again, we could identify a small set of variables that we could recommend to update. This suggests that a dramatically reduced...
battery would be sufficient for this prediction task. We argue that we could not have identified these characteristics in case they would not reflect a true underlying structure.

The ODEnet excels at predicting out-of-sample observations in the latent space mapped by the VAEs compared to more straightforward alternatives, including an autoregressive linear mixed model. The ODEnet also excels in the prediction performance task, training it to predict another latent variable (the mean score) when trained with the same hyperparameters. We provide Julia code and an exemplary Jupyter Notebook on GitHub where we demonstrate our approach with simulated data together with a flexible data structure allowing for modular use of the two components (if required).

To deal with a relatively small number of respondents, our model made efficient use of the vast majority of training data to learn the lower dimensional representations, trajectories, and predictors. In particular, we provided all measurement time points separately as observations to the VAE and did not attempt to fit more complex neural network architectures with time structure, such as recurrent neural networks, as time structure is already covered by the ODEs. Such dynamic models at the heart of neural networks are known to reduce sample size requirements in deep learning algorithms drastically (Rackauckas et al., 2020).

Nevertheless, we had to take several additional countermeasures to reduce the data hunger of deep learning methods, such as coupling the expected value and dispersion parameter by assuming a Poisson distribution for the observed data. Our insights and approaches regarding the feasibility of deep learning approaches when facing moderate sample sizes may also be more generally helpful in other studies that want to use similarly sophisticated approaches but are in doubt of whether their sample size is sufficient.

Here we concentrated on modeling intra-individual changes. Therefore, we modeled the expected values of the latent distributions with ODEs. There are numerous tempting extensions of this approach, especially regarding stochasticity. For example, instead of fitting the means with an ODE, one might force the decoder to learn smoothed representations by providing the samples of the latent distributions as inputs to the decoder and training all networks simultaneously (as discussed in Köber et al., 2022 and Hackenberg et al., 2022). Another tempting direction for future research might be replacing the ODEs with another temporal model, for example, with a system of stochastic differential equations (SDEs). While Innes et al. (2019), Li, Wong, et al. (2020), and Jia & Benson (2019) showed the technical feasibility of combining SDEs with gradient descent, SDEs remain to be embedded into the latent space and parameterized with neural networks.

This article focuses on learning the potentially changing individualized latent space dynamics between laboratory battery time points (B0, B1, ...). Therefore, we did not investigate the reconstructions of the latent space predictions, which would have helped us to analyze the consequences of our dimensionality reduction and temporal modeling on the observed-data-level predictions. Furthermore, sacrificing the focus on learning the individualized dynamics between predefined laboratory battery time points would allow for learning the temporal segments flexibly. Inferring the segment-specific parameters efficiently in a multiple-shooting manner (Iakovlev et al., 2022) and combining it with existing ideas to identify couplable segments (Kamalabad & Grzegorczyk, 2021) could result in another powerful approach. We have to leave this for future work.

ACKNOWLEDGMENTS

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 777084 (DynaMORE project).

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

RK receives advisory honoraria from JoyVentures, Herzlia, Israel. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study might be available from the LIR, Mainz. Restrictions apply to the availability of these data, which were used under license for this study.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section. This article has earned an open data badge “Reproducible Research” for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.
REFERENCES

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association, 112*(518), 859–877.

Boker, S. M., Montpetit, M. A., Hunter, M. D., & Bergeman, C. S. (2010). Modeling resilience with differential equations. In *Individual pathways of change: Statistical models for analyzing learning and development* (pp. 183–206). APA.

Bonanno, G. A., Westphal, M., & Mancini, A. D. (2011). Resilience to loss and potential trauma. *Annual Review of Clinical Psychology, 7*, 511–535.

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods, 22*(3), 409.

Chen, M., Chow, S.-M., & Hunter, M. D. (2018). *Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F.* (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods, 22*(3), 409.

Chen, T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. In *Advances in Neural Information Processing Systems* (pp. 6572–6583). Curran Associates.

Chmítorz, A., Kurth, K., Mey, L. K., Wenzel, M., Lieb, K., Tüscer, O., Kubiak, T., & Kalisch, R. (2020). Assessment of microstresses in adults: Questionnaire development and ecological validation of the Mainz inventory of microstresses. *JMIR Mental Health, 7*(2), e14566.

Driver, C. C., & Voelkle, M. C. (2018). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods, 23*(4), 774.

Esteve, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. https://www.jstatsoft.org/v33/i01/

Gerin, M. I., Viding, E., Pingault, J.-B., Puetz, V. B., Knodt, A. R., Radtke, S. R., Brigidi, B. D., Swartz, J. R., Hariri, A. R., & McCrory, E. J. (2019). Heightened amygdala reactivity and increased stress generation predict internalizing symptoms in adults following childhood maltreatment. *Journal of Child Psychology and Psychiatry, 60*(7), 752–761.

Girin, L., Leglaive, S., Bie, X., Diard, J., Hueber, T., & Alameda-Pineda, X. (2021). Dynamical Variational autoencoders: A comprehensive review. *Foundations and Trends in Machine Learning, 15*, 1–175. http://dx.doi.org/10.1561/2200000008

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine, 27*(1), 191–197.

Hackenberg, M., Grodd, M., Kreutz, C., Fischer, M., Esins, J., Grabenhenrich, L., Karagiannidis, C., & Binder, H. (2021). Using differentiable programming for flexible statistical modeling. *The American Statistician, 76*(3), 270–279. https://doi.org/10.1080/00031305.2021.2002189

Hackenberg, M., Harms, P., Pfaffenlehner, M., Pechmann, A., Kirscher, J., Schmidt, T., & Binder, H. (2022). Deep dynamic modeling with just two time points: Can we still allow for individual trajectories? *Biometrical Journal, 64*, 1426–1445. https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.202000366

Hahn, S. E., & Smith, C. S. (1999). Daily hassles and chronic stressors: Conceptual and measurement issues. *Stress Medicine, 15*(2), 89–101.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC.

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection–a review and recommendations for the practicing statistician. *Biometrical Journal, 60*(3), 431–449.

Hong, Z., & Lian, H. (2012). Time-varying coefficient estimation in differential equation models with noisy time-varying covariates. *Journal of Multivariate Analysis, 103*(1), 58–67.

Iakovlev, V., Yildiz, C., Heinonen, M., & Lähdesmäki, H. (2022). Latent neural ODEs with sparse bayesian multiple shooting. arXiv. https://doi.org/10.48550/arXiv.2210.03466

Inan, G., & Wang, L. (2017). PGEF: An R package for analysis of longitudinal data with high-dimensional covariates. *The R Journal, 9*(1), 393–402. https://doi.org/10.32614/RJ-2017-030

Innes, M., Edelman, A., Fischer, K., Rackauckas, C., Saba, E., Shah, V. B., & Tebbutt, W. (2019). A differentiable programming system to bridge machine learning and scientific computing. CoRR. http://arxiv.org/abs/1907.07587

Innes, M., Saba, E., Fischer, K., Gandhi, D., Rudilosso, M. C., Joy, N. M., Karmali, T., Pal, A., & Shah, V. (2018). *Fashionable modelling with flux*. CoRR. http://arxiv.org/abs/1811.01457

Jia, J., & Benson, A. R. (2019). Neural jump stochastic differential equations. *Advances in Neural Information Processing Systems, 32.*

Kalisch, R., Baker, D. G., Basten, U., Boks, M. P., Bonanno, G. A., Brummelman, E., Chmítorz, A., Fernández, G., Fiebach, C. J., Galatzer-Levy, I., Geuze, E., Groppa, S., Helmreich, I., Hendler, T., Hermans, E., Jovanovic, T., Kubiak, T., Lieb, K., Lutz, B., … Kleim, B. (2017). The resilience framework as a strategy to combat stress-related disorders. *Nature Human Behaviour, 1*(11), 784.

Kalisch, R., Cramer, A. O., Binder, H., Fritz, J., Leertouwer, I., Lunansky, G., Meyer, B., Timmer, J., Veer, I. M., & Van Harmelen, A.-L. (2019). Deconstructing and reconstructing resilience: A dynamic network approach. *Perspectives on Psychological Science, 14*(5), 765–777.

Kalisch, R., Köber, G., Binder, H., Ahrens, K. F., Basten, U., Chmítorz, A., Choi, K. W., Fiebach, C. J., Goldbach, N., Neumann, R. J., Kampa, M., Kollmann, B., Lieb, K., Plichta, M. M., Reif, A., Schick, A., Sebastian, A., Walter, H., Wessa, M., … Engen, H. (2021). The frequent stressor and mental health monitoring-paradigm: A proposal for the operationalization and measurement of resilience and the identification of...
resilience processes in longitudinal observational studies. Frontiers in Psychology, 12, 3377. https://www.frontiersin.org/article/10.3389/fpsyg.2021.710493
Kalisch, R., Müller, M. B., & Tüscher, O. (2015). A conceptual framework for the neurobiological study of resilience. Behavioral and Brain Sciences, 38, e92.
Kamalabad, M. S., & Grzegorczyk, M. (2021). A new bayesian piecewise linear regression model for dynamic network reconstruction. BMC Bioinformatics, 22(2), 1–24.
Kidger, P., Morrill, J., Foster, J., & Lyons, T. (2020). Neural controlled differential equations for irregular time series, 33.
Kim, C.-J., & Nelson, C. R. (1999). State-space models with REGIME Switching: Classical and Gibbs-sampling approaches with applications. MIT Press.
Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv. http://arxiv.org/abs/1412.6980
Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. arXiv. http://arxiv.org/abs/1312.6114
Kingma, D. P., & Welling, M. (2019). An introduction to variational autoencoders. Foundations and Trends® in Machine Learning, 12(4), 307–392. https://doi.org/10.1561/2200000056
Köber, G., Pooseh, S., Engen, H., Chmitorz, A., Kampa, M., Schick, A., Sebastian, A., Tüscher, O., Wessa, M., Yuen, K. S. L., Walter, H., Kalisch, R., Timmer, J., & Binder, H. (2022). Individualizing deep dynamic models for psychological resilience data. Scientific Reports, 12(1), 8061. https://doi.org/10.1038/s41598-022-11650-6
LeNail, A. (2019). Nn-svg: Publication-ready neural network architecture schematics. Journal of Open Source Software, 4(33), 747.
Li, L., Yan, J., Wang, H., & Jin, Y. (2020). Anomaly detection of time series with smoothness-inducing sequential variational auto-encoder. IEEE Transactions on Neural Networks and Learning Systems, 32(3), 1177–1191.
Li, X., Wong, T.-K. L., Chen, R. T. Q., & Duvenaud, D. (2020). Scalable gradients for stochastic differential equations. In S. Chiappa & R. Calandra (Eds.) Proceedings of the Twenty-Third International Conference on Artificial Intelligence and Statistics, vol. 108 of Proceedings of Machine Learning Research, (pp. 3870–3882). PMLR. https://proceedings.mlr.press/v108/li20i.html
Meng, L., Zhang, J., Zhang, X., & Feng, G. (2021). Bayesian estimation of time-varying parameters in ordinary differential equation models with noisy time-varying covariates. Communications in Statistics - Simulation and Computation, 50(3), 708–723. https://doi.org/10.1080/03610918.2019.1565584
Montpetit, M. A., Bergeman, C. S., Deboeck, P. R., Tiberio, S. S., & Boker, S. M. (2010). Resilience-as-process: Negative affect, stress, and coupled dynamical systems. Psychology and Aging, 25(3), 631–640.
Nazábal, A., Olmos, P. M., Ghahramani, Z., & Valera, I. (2020). Handling incomplete heterogeneous data using VAEs. Pattern Recognition, 107, 107501. https://www.sciencedirect.com/science/article/pii/S0031320320303046
Ou, L., Hunter, M. D., & Chow, S.-M. (2019). What’s for dysn: A package for linear and nonlinear dynamic modeling in R. The R Journal, 11, 1–20.
Putter, H., & van Houwelingen, H. C. (2017). Understanding landmarking and its relation with time-dependent Cox regression. Statistics in Biosciences, 9(2), 489–503.
Rackauckas, C., Innes, M., Ma, Y., Bettencourt, J., White, L., & Dixit, V. (2019). DiffEqFlux.jl – A Julia library for neural differential equations. CoRR. http://arxiv.org/abs/1902.02376
Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., & Ramadhan, A. (2020). Universal differential equations for scientific machine learning. arXiv. https://arxiv.org/abs/2001.04385
Rackauckas, C., & Nie, Q. (2017). DifferentialEquations.jl – A performant and feature-rich ecosystem for solving differential equations in Julia. Journal of Open Research Software, 5, 15. https://doi.org/10.5334/jors.151
Refisch, L., Lorenz, F., Riedlinger, T., Taubenböck, H., Fischer, M., Grabenhenrich, L., Wolkewitz, M., Binder, H., & Kreutz, C. (2022). Data-driven prediction of COVID-19 cases in Germany for decision making. BMC Medical Research Methodology, 22(1), 1–13.
Rizopoulos, D. (2012). Joint models for longitudinal and time-to-event data: With applications in R. CRC Press.
Rubanova, Y., Chen, R. T. Q., & Duvenaud, D. (2019). Latent ODEs for irregularly-sampled time series, 33.
Sauerbrei, W., Buchholz, A., Boulesteix, A.-L., & Binder, H. (2015). On stability issues in deriving multivariable regression models. Biometrical Journal, 57(4), 531–555.
Schelldorfer, J., Bühlmann, P., & de Geer, S. v. (2011). Estimation for high-dimensional linear mixed-effects models using ℓ1-penalization. Scandinavian Journal of Statistics, 38(2), 197–214. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9469.2011.00740.x
Schelldorfer, J., Meier, L., & Bühlmann, P. (2014). GLMMLasso: An algorithm for high-dimensional generalized linear mixed models using ℓ1-penalization. Journal of Computational and Graphical Statistics, 23(2), 460–477.
Schmidtma, I., Elsäßer, A., Weinmann, A., & Binder, H. (2014). Coupled variable selection for regression modeling of complex treatment patterns in a clinical cancer registry. Statistics in Medicine, 33(30), 5358–5370.
Serido, J., Almeida, D. M., & Wethington, E. (2004). Chronic stressors and daily hassles: Unique and interactive relationships with psychological distress. Journal of Health and Social Behavior, 45(1), 17–33.
Trepner, M., Salas-Bastos, A., Hess, M., Lenz, S., Vogel, T., & Binder, H. (2021). Synthetic single cell RNA sequencing data from small pilot studies using deep generative models. Scientific Reports, 11(1), 1–11.
Vahdat, A., & Kautz, J. (2020). NVAE: A deep hierarchical variational autoencoder. Advances in Neural Information Processing Systems, 33, –.
Wallisch, C., Dunkler, D., Rauch, G., de Bin, R., & Heinze, G. (2021). Selection of variables for multivariable models: Opportunities and limitations in quantifying model stability by resampling. *Statistics in Medicine, 40*(2), 369–381. https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8779

Wang, L., Chen, G., & Li, H. (2007). Group SCAD regression analysis for microarray time course gene expression data. *Bioinformatics, 23*(12), 1486–1494.

Wang, L., Zhou, J., & Qu, A. (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics, 68*(2), 353–360. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2011.01678.x

Zöller, D., Schmidtmann, I., Weinmann, A., Gerds, T. A., & Binder, H. (2016). Stagewise pseudo-value regression for time-varying effects on the cumulative incidence. *Statistics in Medicine, 35*(7), 1144–1158.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association, 101*(476), 1418–1429.

**SUPPORTING INFORMATION**
Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Köber, G., Kalisch, R., Puhlmann, L. M. C., Chmitorz, A., Schick, A., & Binder, H. (2023). Deep learning and differential equations for modeling changes in individual-level latent dynamics between observation periods. *Biometrical Journal, 65*, 2100381. https://doi.org/10.1002/bimj.202100381