

*Connecting Black-Box Machine Learning
Models to Benchmark Dynamical Systems
Models*

Sy-Miin Chow

Pennsylvania State University

Talk presented at the QuantDev Brownbag

Sept 9, 2026

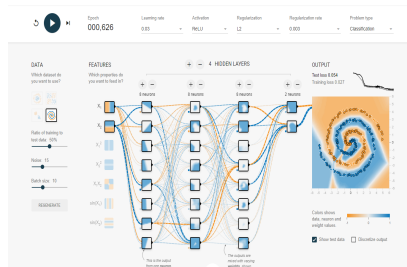
What are machine learning (ML) models?

... **learned mathematical or computational functions** that map inputs to outputs, representations, or generated samples by fitting parameters or structure to data in ways that capture **generalizable patterns beyond the training examples** (Murphy, 2012)”

What are machine learning (ML) models?

... **learned mathematical or computational functions** that map inputs to outputs, representations, or generated samples by fitting parameters or structure to data in ways that capture **generalizable patterns beyond the training examples** (Murphy, 2012)”

- Deep learning models are a specific class of ML models.
- So are “conventional” longitudinal models such as growth curve models, differential/difference equation models, etc.



Screenshot of TensorFlow playground by Daniel Smilkov and Shan Carter.

What Can We Accomplish with ML and Intensive Longitudinal Data (ILD)?

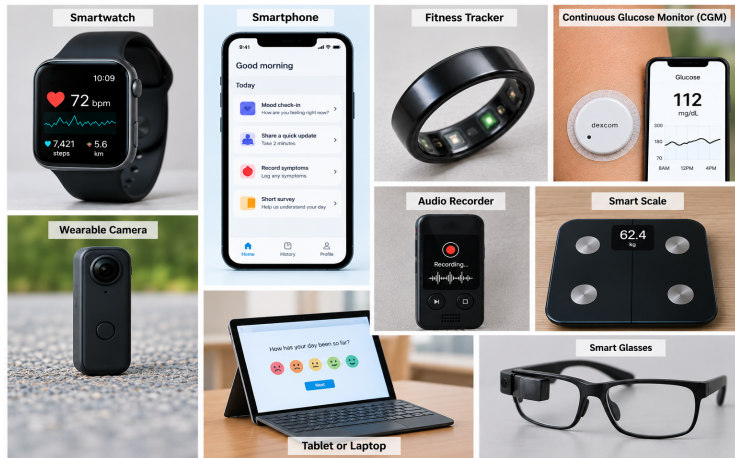


Figure created with assistance from ChatGPT (GPT-5.6 Sol, OpenAI)

Intensive Longitudinal Health Behavior Network (ILHBN)

**Activity
Exercise**



**Substance
Abuse**

**Suicidal
thoughts &
behaviors**



Sleep

Smoking



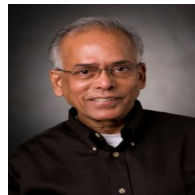
**Weight
Control**

Chow et al. (2023). The ILHBN: challenges, opportunities, and solutions from harmonizing data under heterogeneous study designs, target populations, and measurement protocols. *Translational Behavioral Medicine*. NIH grant #U24AA027684

Outline

- Dynamic Feature Engineering and Selection Using XGBoost
 - Device non-wear
 - Multiple time scales
- Variational autoencoders for fitting stochastic differential equation models (SDEs) as a way to balance data fit and priors from benchmark dynamical systems models
 - What if reduction of test forecast/classification errors and recovery of interpretable modeling parameters are conflicting goals?
- Other challenges/opportunities
 - Uncertainty quantification
 - Separating and pinpointing time scales for adaptations/interventions
 - Missing data

Using Dynamical Systems-Inspired Summary Measures from ILD to Predict Device Non-Wear



Das, J. N. et al. (2025). Performance evaluation of a machine learning-based methodology using dynamical features to detect nonwear intervals in actigraphy data in a free-living setting. *Sleep Health*.

Dynamic ILD Features to Predict Non-Wear

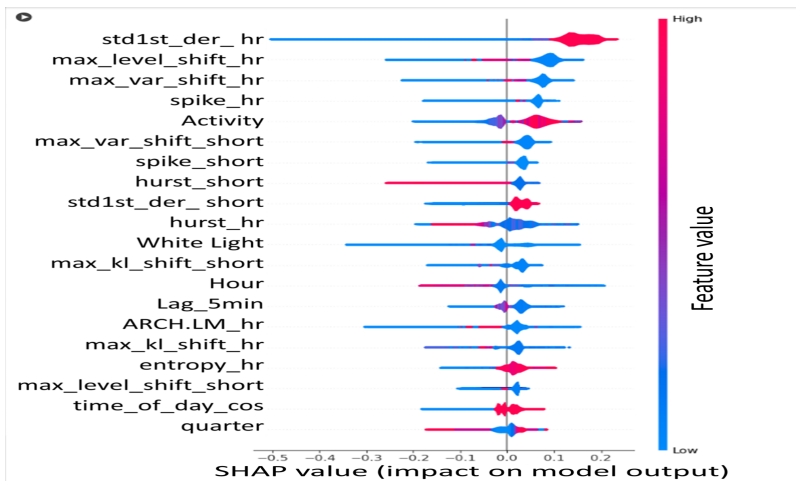


Figure 3 from Das et al. (2025). Short = 20-minute windows; hr = hourly windows; 853 employed adults over 1 week; 597 used for training. Non-wear epoch varied from 2 to 19,970 epochs (i.e., 1 minute to 7 days for 30-second epochs).

Ontological Challenges in Using Dynamic Summary Measures or Computational “Phenotypes” from ILD

Concepts/Measures	Time Scale	Selected References
iSD	Macro, micro	
MSSD	Micro	Jahng et al., 2008
Extremeness (min, max)	Not specific	
Probability of acute change	Micro	Jahng et al., 2008
@ Emotional reactivity	Meso, macro	
@ Transitions & durations		
from state-space grids	Micro	Hollenstein (2015)
@ Dynamic flexibility	Micro	Hollenstein (2015)
@ Inertia	Micro (secs, hrs)	Kuppens et al. (2010, 2015)
@Centralizing tendency	Micro (hours)	Oravecz et al. (2011)

Table: Selected examples of overlapping dynamic summary measures of affect dynamics from ILD; iSD = intraindividual standard deviation; MSSD = mean squared successive difference. (Chow, 2026, R & R)

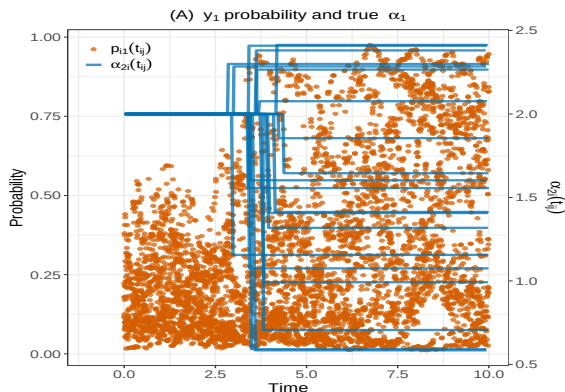
Ontological Challenges (Contd.)

Concepts/Measures	Time Scale	Selected References
** Emotion context insensitivity	Meso	Rottenberg (2005)
** Reactive flexibility	Meso	Hollenstein (2015)
** Robustness	Meso, macro (years)	Ong & Ram, 2016
** Cross-context lability/rigidity	Meso	Ram & Gerstorf (2009)
xx Lability as rapidity of emotional ups and downs	Micro	Chow, Ram et al. (2005)

Table: Selected examples of overlapping dynamic summary measures of affect dynamics from ILD (Chow, 2026, R & R)

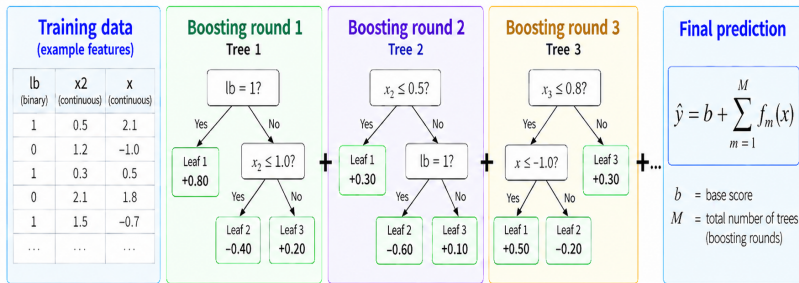
Motivating Illustrative Data: Multi-Temporal Ornstein-Uhlenbeck (OU) Model

$$\begin{aligned}
 dz_i(t) &= \mathbf{B}_i [\mathbf{z}_i(t) - \boldsymbol{\alpha}_i(t)] dt + \mathbf{G} d\mathbf{W}_i(t), \\
 \boldsymbol{\alpha}_i(t) &= \boldsymbol{\alpha}_{\text{base}} + \Delta_{l_b} + \Gamma_{l_b} \mathbf{x}_{3i}, \quad \mathbf{B}_i = \begin{bmatrix} b_{11} & b_{12i} = f_1(\mathbf{x}_{2i}) \\ b_{21i} = f_2(\mathbf{x}_{2i}) & b_{22} \end{bmatrix} \\
 y_{1i}(t_{ij}) &\sim \text{Bernoulli}(p_{1i}(t_{ij})) \quad y_{2i}(t_{ij}) \sim N(z_{2i}(t_{ij}), \sigma_\epsilon^2) \quad (1)
 \end{aligned}$$



Extreme Gradient Boosting (XGBoost; Chen & Guestrin, 2016)

XGBoost builds an ensemble of regression or classification trees, with each new tree trained to improve (i.e., boost) the predictive performance of the existing ensemble.



Additive boosting: trees are added sequentially

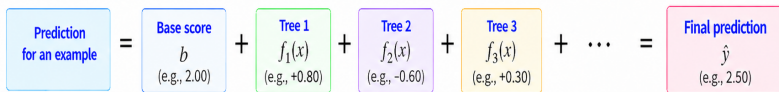


Figure created with assistance from ChatGPT (GPT-5.6 Sol, OpenAI)

Person-Level Dynamic Features

Feature	Measure	Predictor for		
		SD _{y₁}	SD _{y₂}	r _{y₁y₂}
<i>y₁ dynamics (binary)</i>				
Y1_acf1	Lag-1 autocorrelation		✓	✓
Y1_entropy	Shannon entropy		✓	✓
Y1_LV	Lempel–Ziv complexity		✓	✓
<i>y₂ dynamics (continuous)</i>				
Y2_acf1	Lag-1 autocorrelation	✓		✓
Y2_LV	Lempel–Ziv complexity	✓		✓
<i>Cross-process coupling</i>				
cor_Y2lag_Y1	y ₂ → y ₁ lagged correlation	✓		
cor_Y1lag_Y2	y ₁ → y ₂ lagged correlation		✓	
<i>Person-level covariates</i>				
X2i, X3i, X4i	Time-invariant covariates	✓	✓	✓

Note. Shannon entropy (Shannon, 1948) summarizes binary response balance. LV denotes length-normalized Lempel–Ziv complexity (Kaspar & Schuster, 1987; Lempel & Ziv, 1976); y_2 was median-binarized for LV. Features were computed on an equally spaced grid with missing values retained; correlations used only available pairs.

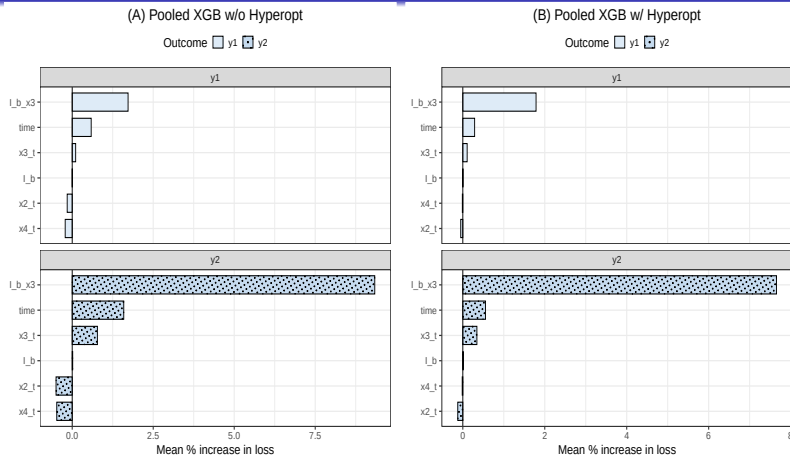
Hyperparameter Tuning

Table: XGBoost Hyperparameters Tuned with tree-structured Parzen estimator (TPE) in Hyperopt (Bergstra et al., 2011, 2015)

Hyperparameter	Search space	Untuned	Pooled	Fast: residCorr
max_depth	$Q(2, 5, 1)$	3	2	2
n_estimators	$Q(50, 350, 50)$	300	100	300
learning_rate	$\text{Log-U}(.01, .10)$	.050	.016	.014
min_child_weight	$\text{Log-U}(1, 20)$	1.00 ^a	2.84	3.25
subsample	$U(.60, 1)$	.80	.75	.66
colsample_bytree	$U(.60, 1)$	.80	.86	.65
reg_lambda	$\text{Log-U}(10^{-2}, 10^b)$	1.00	.033	.060
<i>Fixed parameters</i>				
gamma	—	0	0	0
reg_alpha	—	0	0	0

Note. U = uniform; Log-U = log-uniform; $Q(a, b, c)$ = uniform over $[a, b]$, quantized in increments of c . Continuous outcomes were standardized before XGBoost fitting. Fast-process hyperparameters were tuned separately by outcome; only residCorr is shown. ^aXGBoost default. ^b $b = 0$ for pooled and $b = 1$ for fast models.

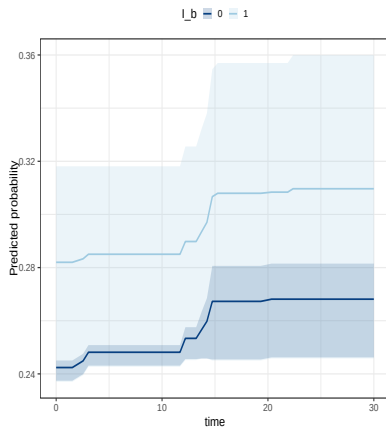
Permutation Importance (100 trials) from Pooled Person-Time XGBoost Analysis



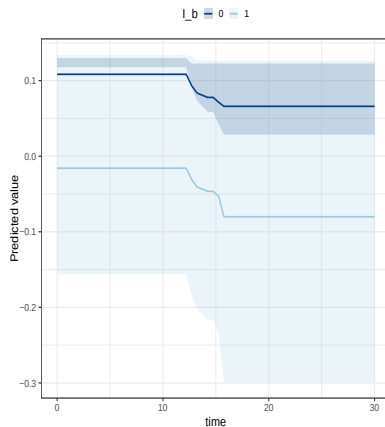
Average % increase in test loss across 100 subject-level 70:30 train/test splits, with each predictor permuted 10 times per trial. Out-of-sample McFadden's pseudo- R^2 and R^2 for y_1 and $y_2 = 0.02$ and 0.05 , respectively, in the tuned model, and 0.00 and -0.01 , respectively, for the untuned model).

Effect of Contextual Predictor, I_b , Showed up as Nonlinear Effect of Time

(C) Pooled XGB: predicted y1 over time



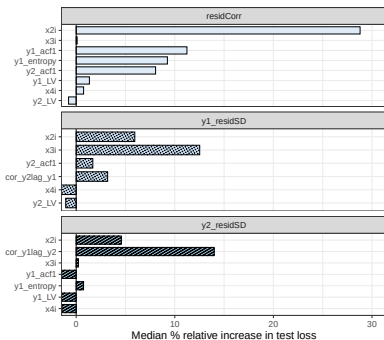
(D) Pooled XGB: predicted y2 over time



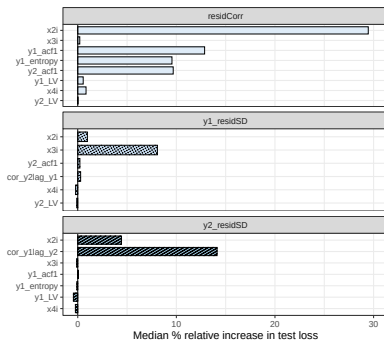
Sequential (Slow-Fast Process) XGBoost Analysis

$$\begin{aligned} \text{Detrended: } g\{E(y_{ki}(t_{ij}))\} &= (\beta_0^{(k)} + u_{0i}^{(k)}) + (\beta_1^{(k)} + u_{1i}^{(k)})I_{bi}(t_{ij}) + \beta_2^{(k)}x_{2i}(t_{ij}) \\ &+ \beta_3^{(k)}x_{3i}(t_{ij}) + \beta_4^{(k)}x_{4i}(t_{ij}) + \beta_5^{(k)}I_{bi}(t_{ij})x_{3i}(t_{ij}) + \mathbf{s}(t_{ij})^\top \beta_t^{(k)}, \end{aligned} \quad (2)$$

(A) Between-person XGB w/o Hyperopt

Outcome residCorr y1_residSD y2_residSD

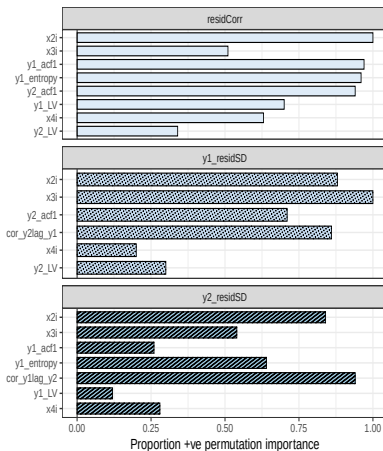
(B) Between-person XGB w/ Hyperopt

Outcome residCorr y1_residSD y2_residSD

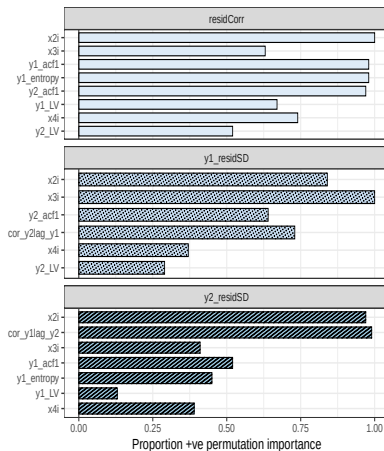
Median % increase in test loss across 100 subject-level 70:30 train/test splits, with each predictor permuted 10 times per trial. Out-of-sample $R^2 = 0.29$ for residCorr, 0.05 for y1_residSD, and 0.10 for y2_residSD, respectively, in the tuned model; and 0.20, -0.08, and -0.08, respectively, in the untuned model).

Sequential (Slow-Fast Process) XGBoost Analysis

(C) Between-person XGB w/o Hyperopt

Outcome  residCorr  y1_residSD  y2_residSD

(D) Between-person XGB w/ Hyperopt

Outcome  residCorr  y1_residSD  y2_residSD

Proportion of trials with increase in test loss across 100 subject-level 70:30 train/test splits, with each predictor permuted 10 times per trial.

Interim Summary of Key Insights

- Part of the effect of the phase indicator, I_b , on baseline shifts was bypassed as a nonlinear effect of time.
- Successful detection of important between-person covariates of fast dynamics and corresponding cross-process couplings requires separation of the slow- and fast temporal processes.
- Some dynamic summary measures overlap completely (e.g., residual SDs and RMSSD, depending on time scales; residual entropy and SD for binary DV)
- Hyperparameter tuning using Hyperopt (Bergstra, Yamins, & Cox, 2013) in the tuned analysis was better able to distinguish true from spurious covariates and improved R^2 and pseudo- R^2 .
- Separate tuning for each of the residual outcomes as opposed to "one-size-fits-all" hyperparameter tuning facilitated identification of meaningful features.

Variational autoencoder for stochastic differential equations
(VAE-SDE)

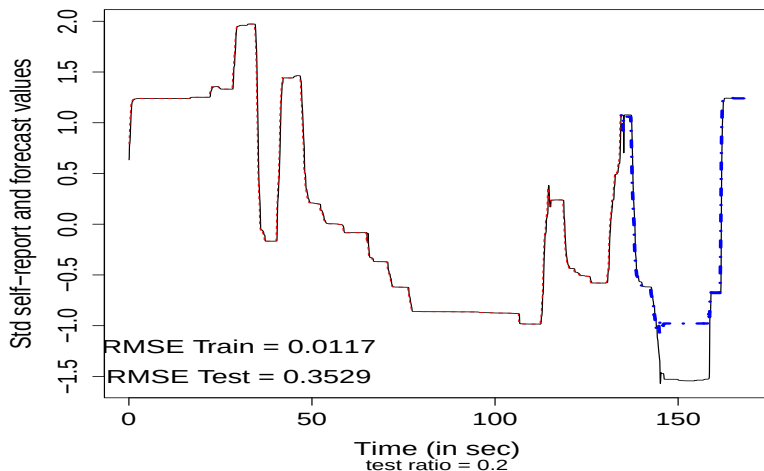
(Chow, 2026, R & R)

ML in Psychometrics and Related Disciplines

- Advanced applications (e.g., using recurrent neural nets to classify mobile intoxication and forecast depression; Arizmendi, Urban, & Gates, 2023; Y. W. Cho, Elavsky, & Chow, 2026, under review)
- As computational engine to fit parametric latent variable models, such as item response theory (IRT) models (Converse, Curi, Oliveira, & Templin, 2021; Curi, Converse, Hajewski, & Oliveira, 2019; Urban, 2021)
- Hybrid models integrating ML with parametric, interpretable models:
 - Deep learning generalized structured component analysis (G. Cho & Hwang, 2023)
 - Other IRT + deep learning variations (Cheng et al., 2019; Tsutsumi, Kinoshita, & Ueno, 2021)
 - Tree based methods (e.g., random forest) + IRT to address the cold start problem in adaptive learning systems (Pliakos et al., 2019)
 - Functional data analysis as a way to fit interpretable ODE models (Ramsay, Hooker, Campbell, & Cao, 2007)

ML Forecast Within an Individual Given Historical Data

Forecasts with wavelet random forest; ID = 4, LPA_Slide



Using WaveletRF package in R (Paul, Yeasin, & Paul, 2022; Rezaali et al., 2021)

Variational Autoencoder (VAE) for Fitting IRT Models

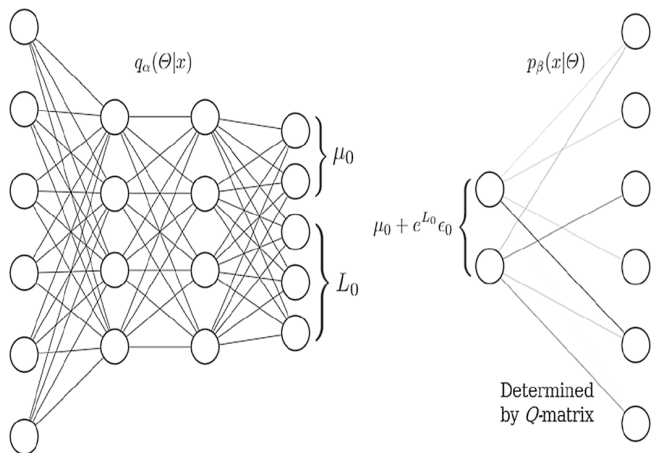
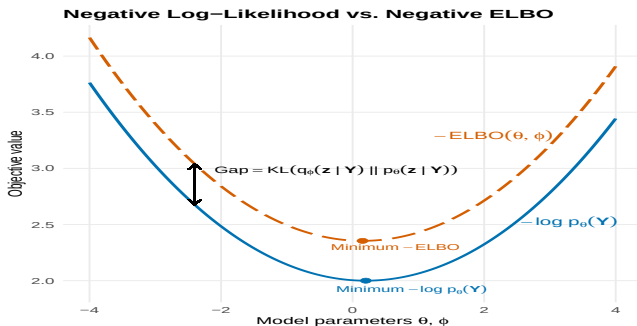


Figure 1 from Converse et al. (2021). Architecture for fitting a multidimensional logistic 2-parameter model with VAE (ML2P-VAE) for two correlated latent traits and six input items

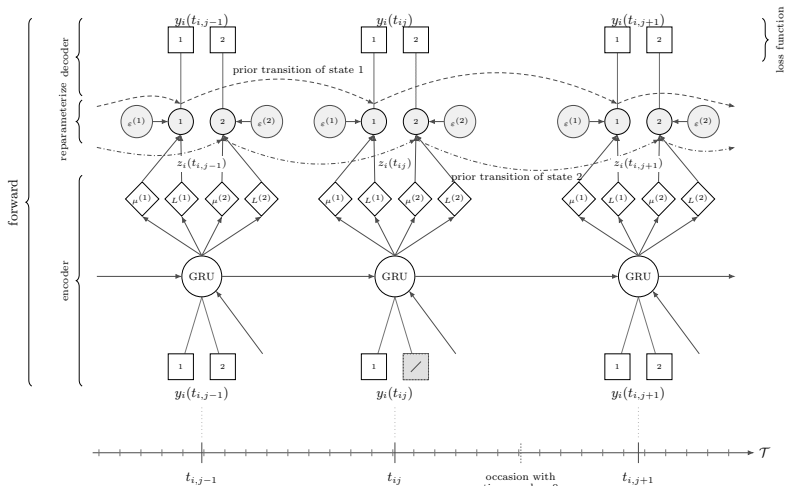
VAE for stochastic differential equations (VAE-SDE)

$$\mathcal{L}_i = \underbrace{-\mathbb{E}_{q_\phi(\mathbf{z}_i, \mathcal{T}_i | \mathbf{Y}_i, \mathcal{T}_i, \mathbf{M}_i, \mathcal{T}_i, \mathbf{X}_i, \mathcal{T}_i)} [\log p_\theta(\mathbf{Y}_i, \mathcal{T}_i | \mathbf{z}_i, \mathcal{T}_i)]}_{\mathcal{L}_{\text{rec}, i}} + \underbrace{\lambda_{\text{KL}} \text{KL} [q_\phi(\mathbf{z}_i, \mathcal{T}_i | \mathbf{Y}_i, \mathcal{T}_i, \mathbf{M}_i, \mathcal{T}_i, \mathbf{X}_i, \mathcal{T}_i) \parallel p_\theta(\mathbf{z}_i, \mathcal{T}_i | \mathbf{X}_i, \mathcal{T}_i)]}_{\lambda_{\text{KL}} \mathcal{L}_{\text{KL}, i}}, \quad (3)$$



Conceptual illustration of the negative evidence lower bound (ELBO) as an upper bound on the true marginal negative log-likelihood (Chow, 2026, R. & R).

States-Based VAE-SDE (Chow, 2026, R & R)



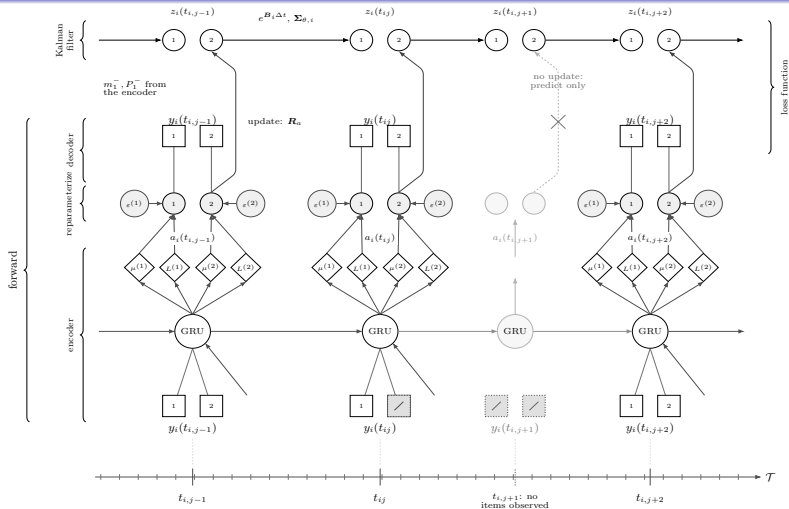
GRU = gated recurrent unit; squares = observed variables; diamonds = deterministic computational outputs; shared circles = stochastic quantities

Encoder: A GRU reads observed DVs y , t_{ij} , missingness indicators

Reparameterize: $\mathbf{z}_{i,1:T_i}^{\mathcal{T},(s)} = \mu^{\mathcal{T},(s)} + L^{\mathcal{T},(s)} \epsilon^{\mathcal{T},(s)}$; $\epsilon^{\mathcal{T},(s)} \sim \mathcal{N}(0, I_2)$

Decoder and loss: Map latent samples to observed DVs to compute loss function

Neural Continuous-Discrete State Space Models – Filtering-Based VAE-SDE (Ansari, Heng, Lim, & Soh, 2023)



Auxiliary variables. Ansari et al. (2023) used deep learning to approximate auxiliary variables $a_i(t_{ij})$, with $a_i(t_{ij}) \mid z_i(t_{ij}) \sim N(\mathbf{H}z_i(t_{ij}), \mathbf{R}_a)$ and $\mathbf{H} = \mathbf{I}$. The dynamic states in $z_i(t_{ij})$ are not sampled but integrated out using continuous-discrete Kalman filtering. $\mathbf{R}_a$ is a fixed covariance matrix.

Hyperparameter Tuning in VAE-SDE

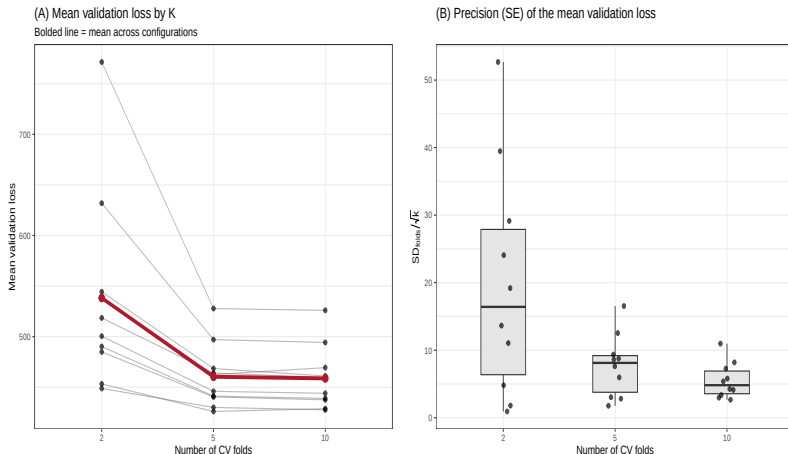
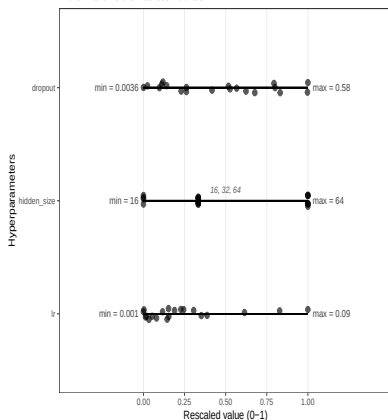


Figure: (A) Mean validation losses averaged across 10 Hyperopt trials evaluated with the exact same sequence of possible hyperparameter values; (B) precision of the average validation loss values, obtained as $\frac{SD_{folds}}{\sqrt{k}}$, where SD_{folds} is the standard deviation across the k folds within each Hyperopt trial.

Hyperparameter Tuning in VAE-SDE (Contd.)

(C) Hyperparameter values evaluated
in the final 5-fold CV across 20 trials



(D) Sensitivity of mean validation loss by hyperparameters
20 Hyperopt trials; panels ordered by |Spearman rho| with the objective

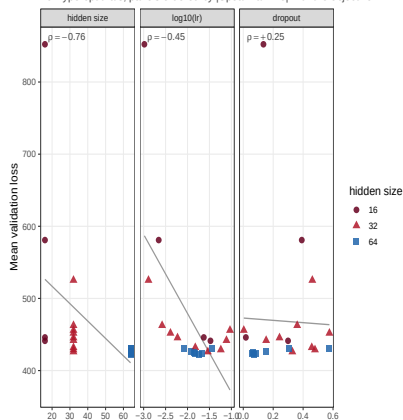


Figure: (C) The range of hyperparameter values evaluated by Hyperopt across 20 trials with 5-fold CV; (D) variations in mean validation loss values against different hyperparameters.

SDE parameter Recovery and Forecast Test Loss Minimization as Conflicting Goals

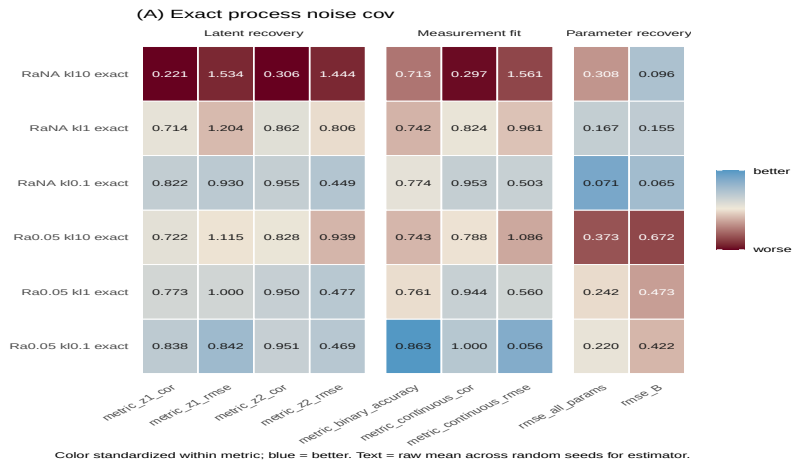


Figure: In the row labels: $RaNAkl/\lambda$ = states-based VAE-SDE with $\lambda_{KL} = \lambda$; $Ra0.05\lambda$ = filtering-based VAE-SDE with $R_a = 0.05I_2$ and $\lambda_{KL} = \lambda$.

SDE parameter Recovery and Forecast Test Loss Minimization as Conflicting Goals (Euler Approx)

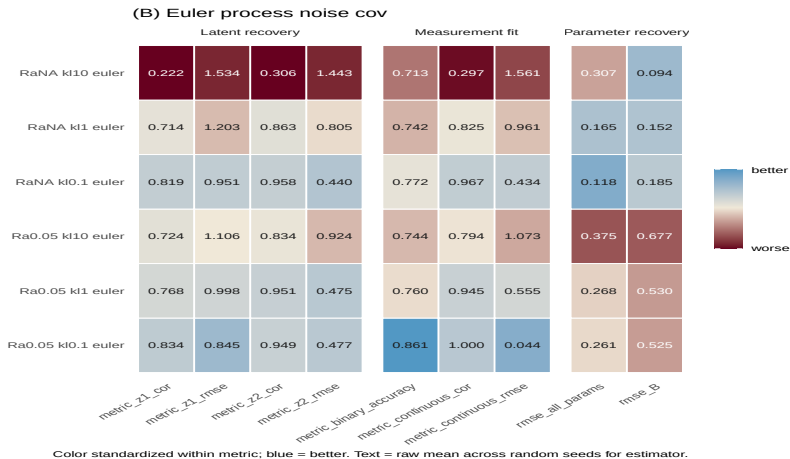


Figure: In the row labels: $R_a N a k / \lambda$ = states-based VAE-SDE with $\lambda_{KL} = \lambda$; $R_a 0.05 \lambda$ = filtering-based VAE-SDE with $R_a = 0.05 I_2$ and $\lambda_{KL} = \lambda$.

Summary of Key Insights

- Evaluation of the sensitivity of optimization results across different hyperparameter configurations via K -fold validation is critical.
- Consideration and variations of hyperparameters may have to be done in stages (e.g., λ_{KL} weights $\rightarrow$ hidden layers $\rightarrow$ hidden dimension).
- Inspection of the number of CV folds needed to ensure stability of results is helpful.
- Multiple performance measures may have to be considered. Different performance measures may prefer different models and deep learning architectures.
- Different estimators may have different respective strengths.

Summary of Key Insights

What is your key goal in using ML with ILD?

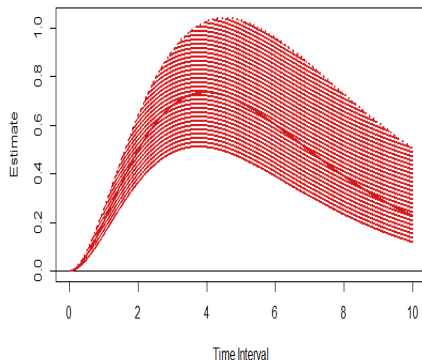


cTMed on CRAN

Effect sizes (standardized slopes), standard error estimates, graphical tool, and parallel computing

```
mcstd <- -MCMedStd(phi =
phi, sigma = sigma, vcovtheta =
vcov, deltat = deltat, from =
"x", to = "y", med = "m", R =
20000L, ncores = parallel ::
detectCores())plot(mcstd)
```

95% CI for the Indirect Effect (Monte Carlo Method)





Simple Trend Models May be Sufficient — But Don't Ignore Trends

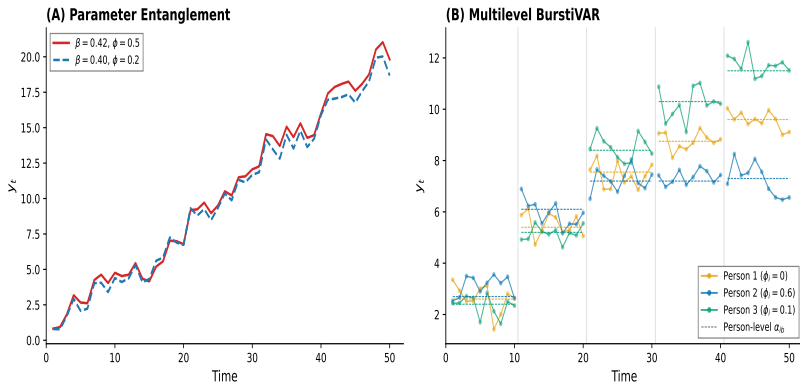
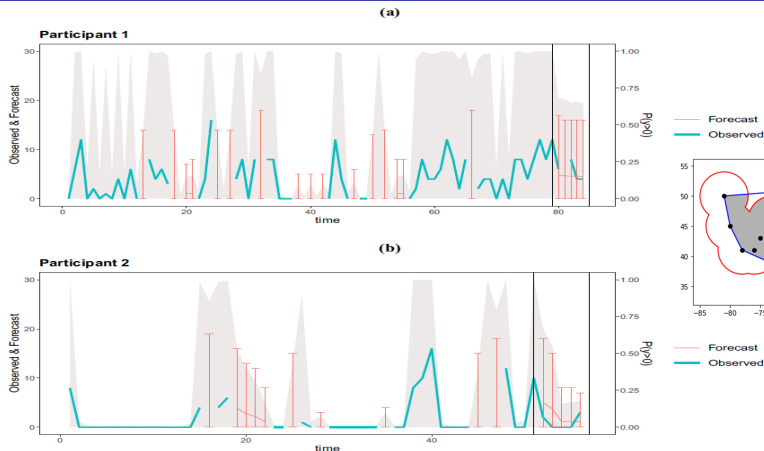


Figure from Xiong, Pesigan, Oravec, Lanza, and Chow (2026, under review).



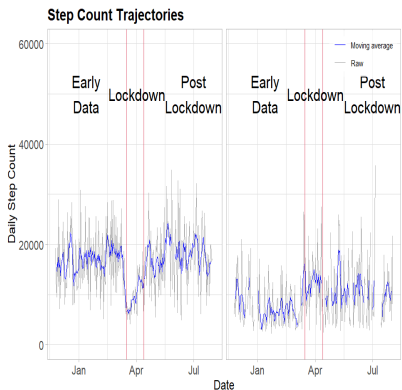
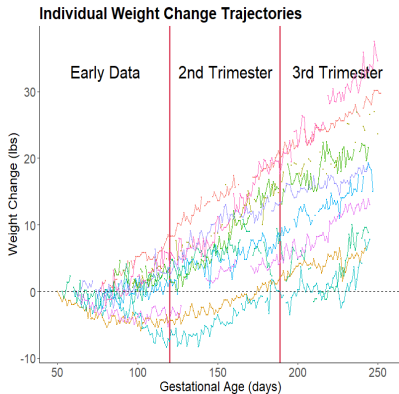
Forecasting Substance Use with Spatial Data and Bayesian Zero-Inflated Regime-Switching Model



Observed alcohol use (green thick lines); imputed/forecast alcohol use (red thin lines); uncertainty (red error bars); and estimated/predictive probability of drinking alcohol (shaded area). Higher LO(drinking regime $\rightarrow$ zero-inflated regime) on weeks with more time spent with twin siblings than usual (Li et al., 2022).



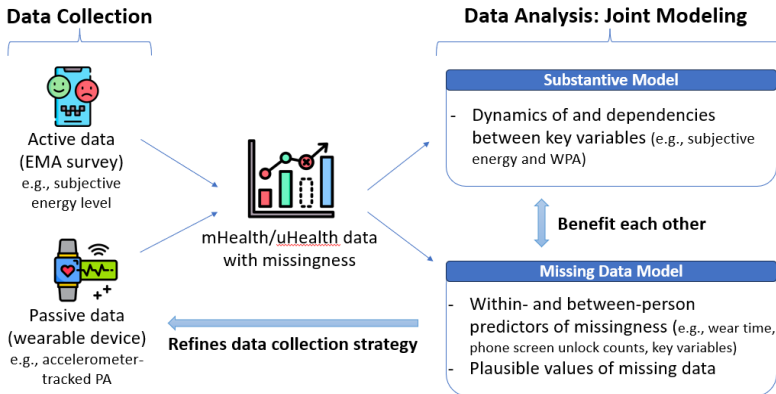
Timely Adaptations: Practical Considerations



Cho et al. (2025). When Deep Learning Outperforms Conventional Models in Long-Term Forecasting from Early Intensive Longitudinal Data. Dissertation and student poster presented at the 2025 SMEP Meeting.



Incorporation of Missing Data Model - Multilevel Decomposition Likely Helpful



WPA = Physical activity as measured with wearable device

Figure taken from Y. W. Cho et al. (2025). *JMIR mHealth and uHealth*.

References I

- Ansari, A. F., Heng, A., Lim, A., & Soh, H. (2023, 23–29 Jul). Neural continuous-discrete state space models for irregularly-sampled time series. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, & J. Scarlett (Eds.), *Proceedings of the 40th international conference on machine learning* (Vol. 202, pp. 926–951). PMLR. Retrieved from <https://proceedings.mlr.press/v202/ansari23a.html>
- Arizmendi, C. J., Urban, C. J., & Gates, K. M. (2023). Deep-learning methods for mobile sensing. *Mobile Sensing in Psychology: Methods and Applications*, 432.
- Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 24.

References II

- Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., & Cox, D. D. (2015). Hyperopt: A Python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, 8(1), 014008. doi: 10.1088/1749-4699/8/1/014008
- Bergstra, J., Yamins, D., & Cox, D. D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. *Journal of Machine Learning Research*, 13(Oct), 281–305.

References III

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (p. 785 794). New York, NY, USA: Association for Computing Machinery. Retrieved from <https://doi.org/10.1145/2939672.2939785> doi: 10.1145/2939672.2939785
- Cheng, S., Liu, Q., Chen, E., Huang, Z., Huang, Z., Chen, Y., . . . Hu, G. (2019). Dirt: Deep learning enhanced item response theory for cognitive diagnosis. In *Proceedings of the 28th acm international conference on information and knowledge management* (pp. 2397–2400).

References IV

- Cho, G., & Hwang, H. (2023, 07). Deep learning generalized structured component analysis: An interpretable artificial neural network model with composite indexes. *Structural Equation Modeling A Multidisciplinary Journal*.
- Cho, Y. W., Chow, S., Li, J., Wang, W.-L., Wang, S., Ji, L., . . . Dunton, G. F. (2025). Within-and between-individual compliance in mobile health: Joint modeling approach to nonrandom missingness in an intensive longitudinal observational study. *JMIR mHealth and uHealth*, 13, e65350.
- Cho, Y. W., Elavsky, S., & Chow, S.-M. (2026, under review). Sequential deep diagnosis framework (SDDF): Using deep learning and latent class modeling to diagnose pandemic-related physical activity declines. *Multivariate Behavioral Research*.

References V

- Chow, S. (2026, R & R). It's about time – intensive longitudinal modeling with machine learning methods. *Multivariate Behavioral Research*.
- Chow, S., Nahum-Shani, I., Baker, J. T., Spruijt-Metz, D., Allen, N. B., Auerbach, R. P., ... others (2023). The ilhbn: challenges, opportunities, and solutions from harmonizing data under heterogeneous study designs, target populations, and measurement protocols. *Translational Behavioral Medicine*, 13(1), 7–16.
- Converse, G., Curi, M., Oliveira, S., & Templin, J. (2021). Estimation of multidimensional item response theory models with correlated latent variables using variational autoencoders. *Machine Learning*, 110, 1463–1480. doi: 10.1007/s10994-021-06005-7

References VI

- Curi, M., Converse, G., Hajewski, J., & Oliveira, S. (2019). Interpretable variational autoencoders for cognitive models. In *2019 international joint conference on neural networks (ijcnn)* (pp. 1–8). doi: 10.1109/IJCNN.2019.8852333
- Kaspar, F., & Schuster, H. G. (1987). Easily calculable measure for the complexity of spatiotemporal patterns. *Physical Review A*, 36(2), 842–848. doi: 10.1103/PhysRevA.36.842
- Lempel, A., & Ziv, J. (1976). On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22, 21–27.

References VII

- Li, Y., Oravecz, Z., Zhou, S., Bodovski, Y., Barnett, I. J., Chi, G., ... Chow, S.-M. (2022). Bayesian forecasting with a regime-switching zero-inflated multilevel poisson regression model: An application to adolescent alcohol use with spatial covariates. *Psychometrika*, 87, 376-402. doi: 10.1007/s11336-021-09831-9
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
- Paul, R. K., Yeasin, M., & Paul, M. R. K. (2022). Package waveletrf.
- Pliakos, K., Joo, S.-H., Park, J. Y., Cornillie, F., Vens, C., & Van den Noortgate, W. (2019). Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education*, 137, 91-103.

References VIII

- Ramsay, J. O., Hooker, G., Campbell, D., & Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach (with discussion). *Journal of Royal Statistical Society: Series B*, 69(5), 741-796. doi: 10.1111/j.1467-9868.2007.00610.x
- Rezaali, M., Fouladi-Fard, R., Mojarad, H., Sorooshian, A., Mahdinia, M., & Mirzaei, N. (2021). A wavelet-based random forest approach for indoor btex spatiotemporal modeling and health risk assessment. *Environmental Science and Pollution Research*, 28, 22522–22535.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-432.
- Tsutsumi, E., Kinoshita, R., & Ueno, M. (2021). Deep item response theory as a novel test theory based on deep learning. *Electronics*, 10(9), 1020.

References IX

- Urban, C. J. (2021). *Machine learning-based estimation and goodness-of-fit for large-scale confirmatory item factor analysis* (Unpublished master's thesis). The University of North Carolina at Chapel Hill.
- Xiong, X., Pesigan, I. J. A., Oravecz, Z., Lanza, S. T., & Chow, S.-M. (2026, under review). Burstivar for capturing multilevel vector autoregressive processes with trends in measurement burst designs. *Psychological Methods*.