A Personalized Predictive Framework for Multivariate Clinical Time Series via Adaptive Model Selection

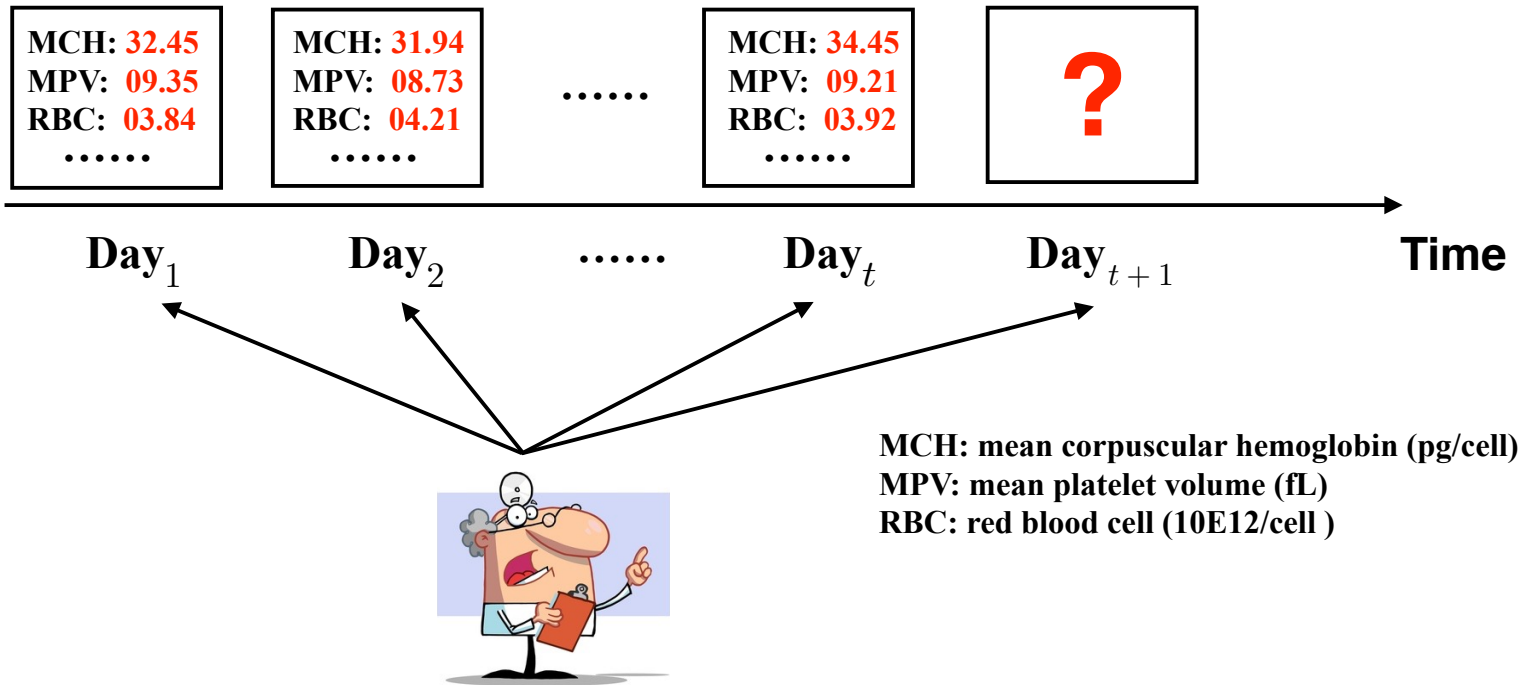
Zitao Liu[1] Milos Hauskrecht[2]

[1]Pinterest Lab Researcher
[2]University of Pittsburgh

zitaoliu@pinterest.com

Personalized Prediction Problem

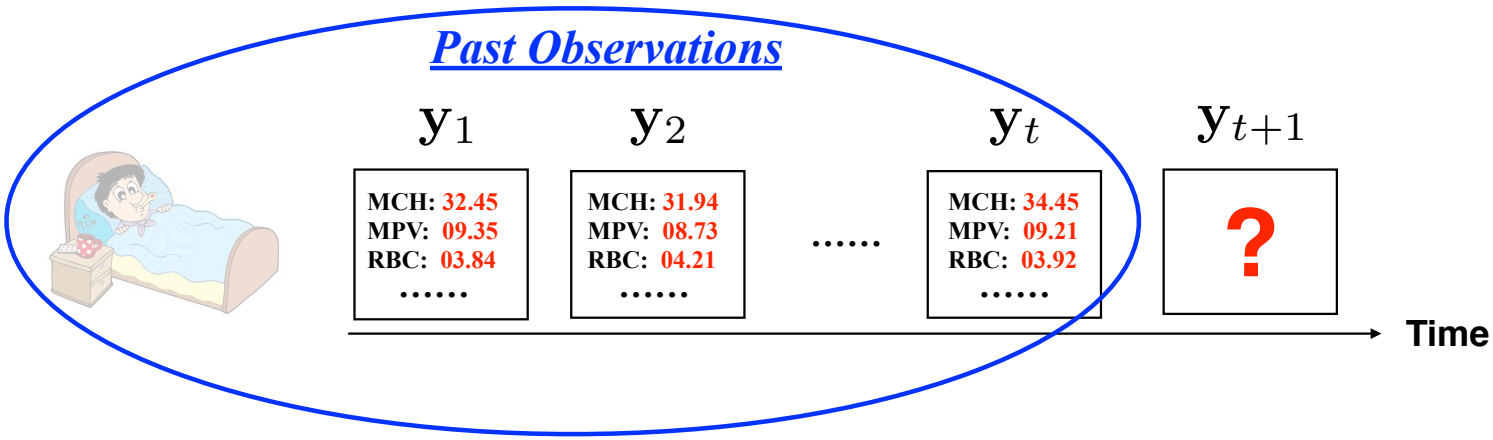
Clinical time series forecasting for **each individual patient**.



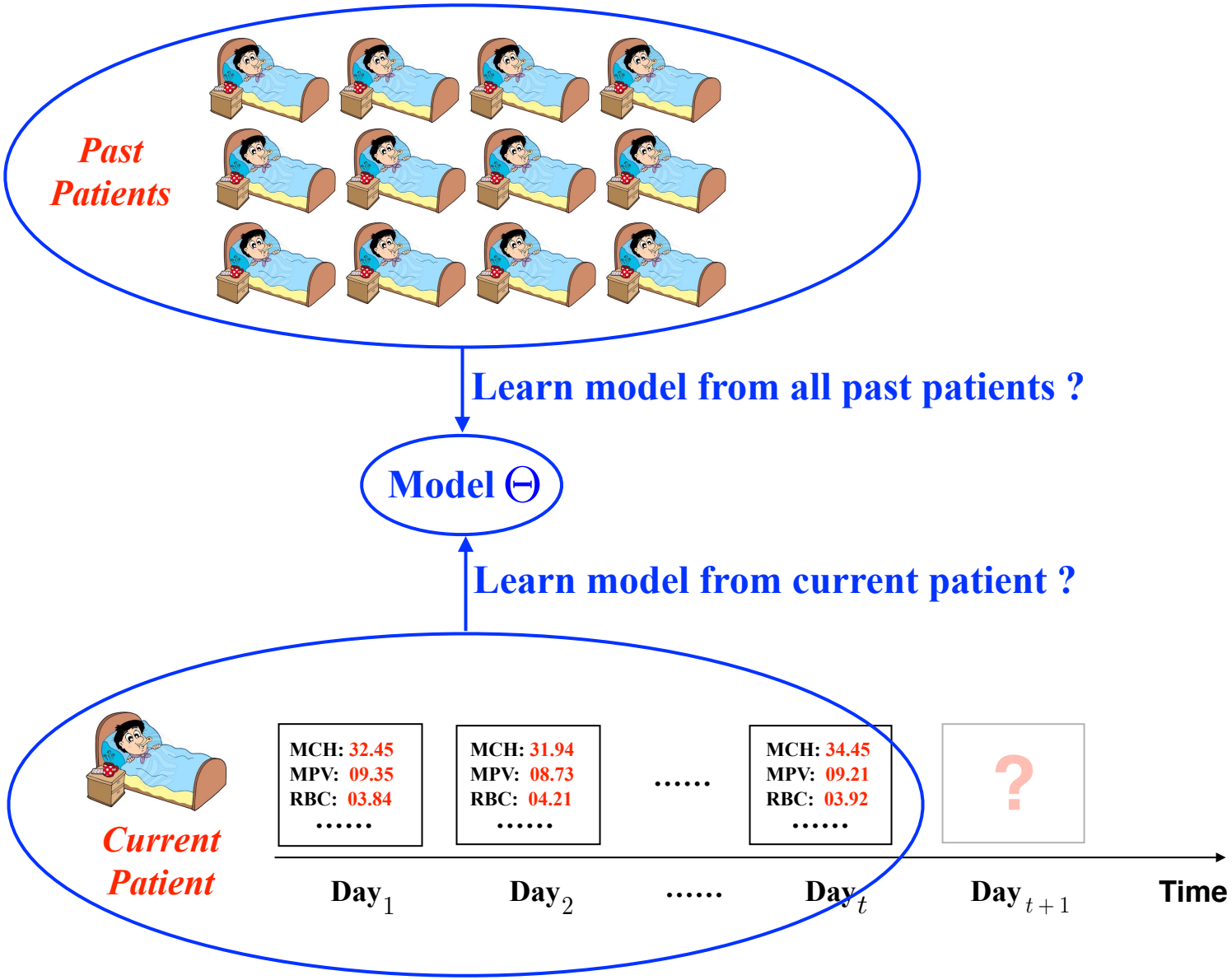
MCH: mean corpuscular hemoglobin (pg/cell)
MPV: mean platelet volume (fL)
RBC: red blood cell (10E12/cell)

How to learn a good forecasting model(Θ) from the past?

Question: How to learn a good personalized forecasting model(Θ) from past observations?



How to learn a good forecasting model(Θ) from the past?



Two Existing Approaches

- **Population-based models**

  Problem: patient-specific variations are typically large and population-based models are unable to support accurate predictions for each individual patient.

- **Instance/patient-specific models**

  Problem: time series observed for one patient may be too short to learn a high-quality patient-specific model.

Goal

Build a personalized predictive model that relies on the data for **past patient population** but can adapt quickly to new observations made for **the target patient**.

Building Personalized Predictive Models

- **Subpopulation models**
- **Model adaptation**
- **Adaptive model selection**

Building Personalized Predictive Models

- **Subpopulation models**: learned from a selected collection of similar examples out of the entire population.

  Disadvantages:
  - Finding "neighbors" is difficult when **the sequence is very short**.
  - Intensive neighbor searching process has to be redone once new observations arrive
  - A subpopulation may still be very large and exhibit a lots of patient-specific variations.

Building Personalized Predictive Models

- **Subpopulation models**
- **Model adaptation**: adjusting the population-based model to fit better the specific instance.

  Disadvantages:
  - Designing and deriving adaptation is difficult and varies from model to model. (For example, deriving exact posterior in Bayesian setting.)
  - A larger number of instance-specific features or observations are required to conduct adaptation.

Building Personalized Predictive Models

- **Subpopulation models**
- **Model adaptation**
- **Adaptive model selection**: combining a pool of predictive models which are built either from the entire population or a subpopulation of instances.

  Disadvantages:
  - Error feedback over longer periods of time are required to optimize the combination weights.
  - Recent errors are smoothed out by the errors made in the early stage of the process.

Our Approaches

We develop a personalized clinical time series prediction frameworks that better mimic patient specific temporal behaviors and variations.

- **Personalized prediction via adaptive model selection**: selecting the appropriate model at each time stamp for each patient.

Personalized Prediction via Adaptive Model Selection

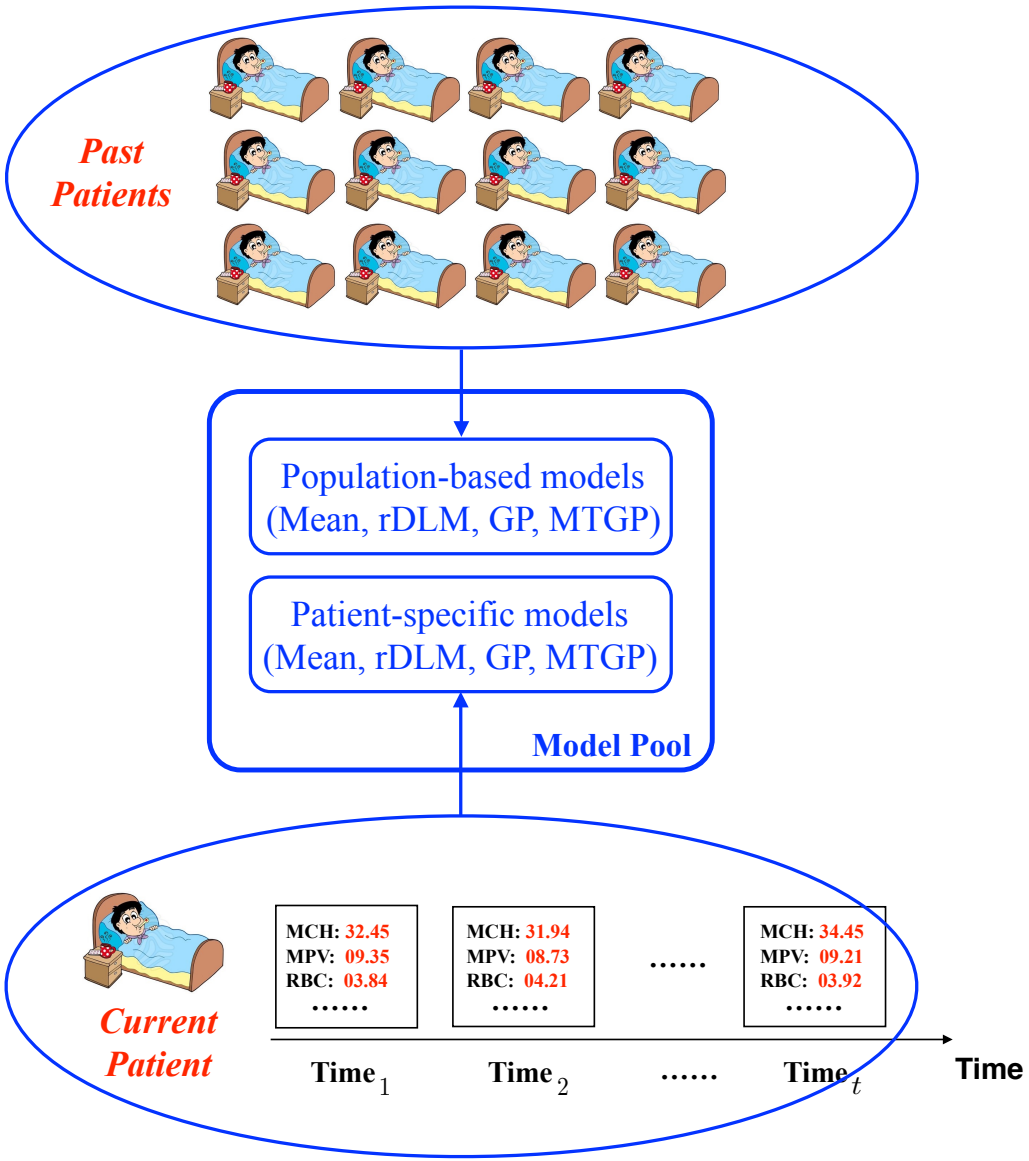
Intuition: predictions at different times may be driven by the different types of prediction models.

Our Approach: online model switching.

Learning: Build a pool of population-based and patient-specific prediction models.

Prediction: Select the appropriate model by the weighted sum of prediction errors (or deviations) of each model on past patient's data.

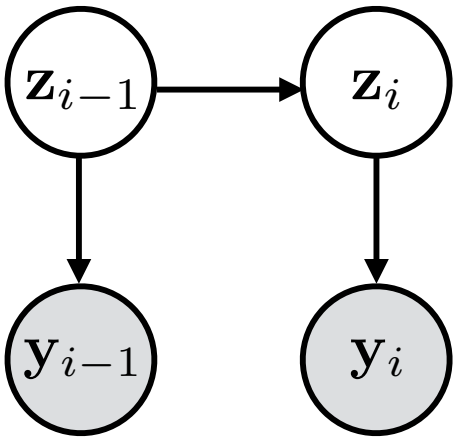
Learn A Pool of Models



Dynamic Linear Model (DLM)

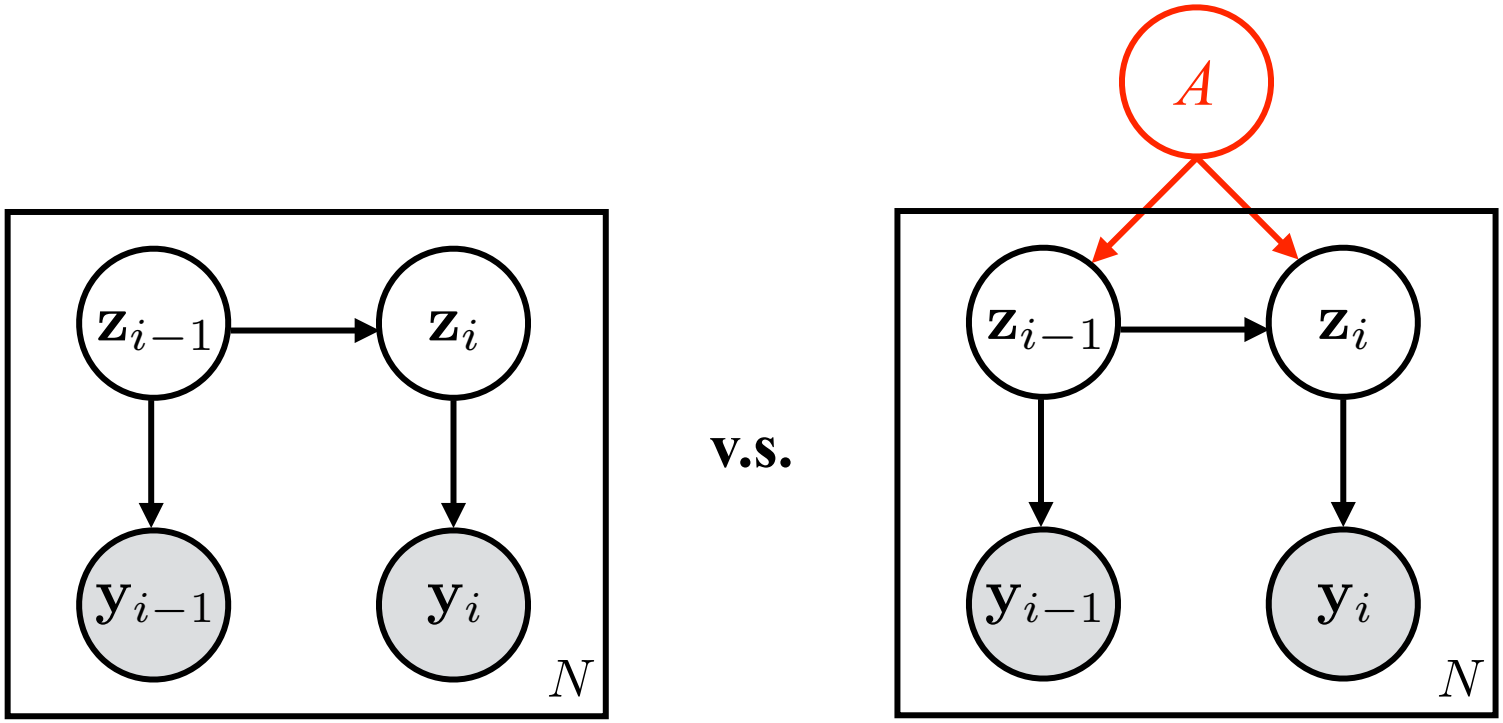
Dynamic Linear Model

$$\mathbf{y}_i = C\mathbf{z}_i + \zeta_i; \qquad \mathbf{z}_i = A\mathbf{z}_{i-1} + \epsilon_i$$
$$\epsilon_i \sim \mathcal{N}(0, Q), \zeta_i \sim \mathcal{N}(0, R) \text{ and } z_1 \sim \mathcal{N}(\xi, \Psi)$$



- $\{\mathbf{y}_i\}$ time series of observations.
- $\{\mathbf{z}_i\}$ hidden states driving the dynamics.
- Parameters $\Lambda = \{A, C, Q, R, \xi, \Psi\}$.

Regularized Dynamic Linear Model (rDLM)



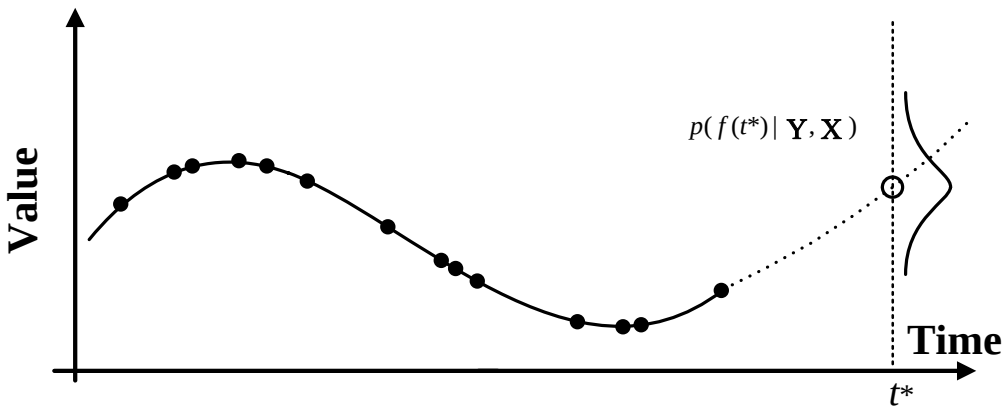
DLM **v.s.** **regularized DLM**

Achieve low-rank property by introducing priors on A!

Gaussian Processes (GP)

GP can be used to calculate the posterior distribution $p(f(\mathbf{X}^*)|(\mathbf{X},\mathbf{Y}))$ of f values for inputs $\mathbf{X}^*$, given a set of observed values $\mathbf{Y}$ for $\mathbf{X}$.

$$f(\mathbf{X}^*)|(\mathbf{X},\mathbf{Y}) \sim \mathcal{N}(m(\mathbf{X}^*|(\mathbf{X},\mathbf{Y})), K^G(\mathbf{X}^*|(\mathbf{X},\mathbf{Y})))$$

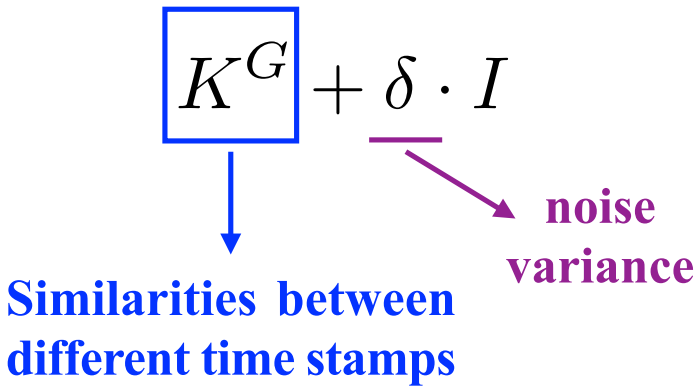


Advantage:

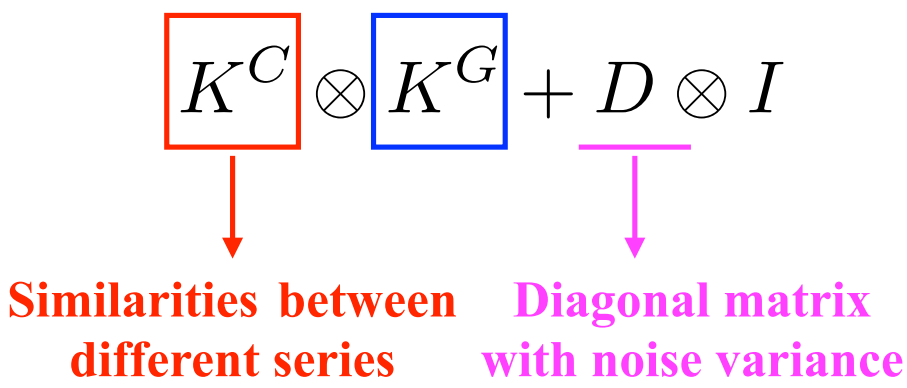
- Non-parametric
- Continuous time
- Capture short-term variability

Multi-task Gaussian Process (MTGP)

The MTGP [Boni 07] is an extension of GP. The covariance matrix of MTGP:



GP



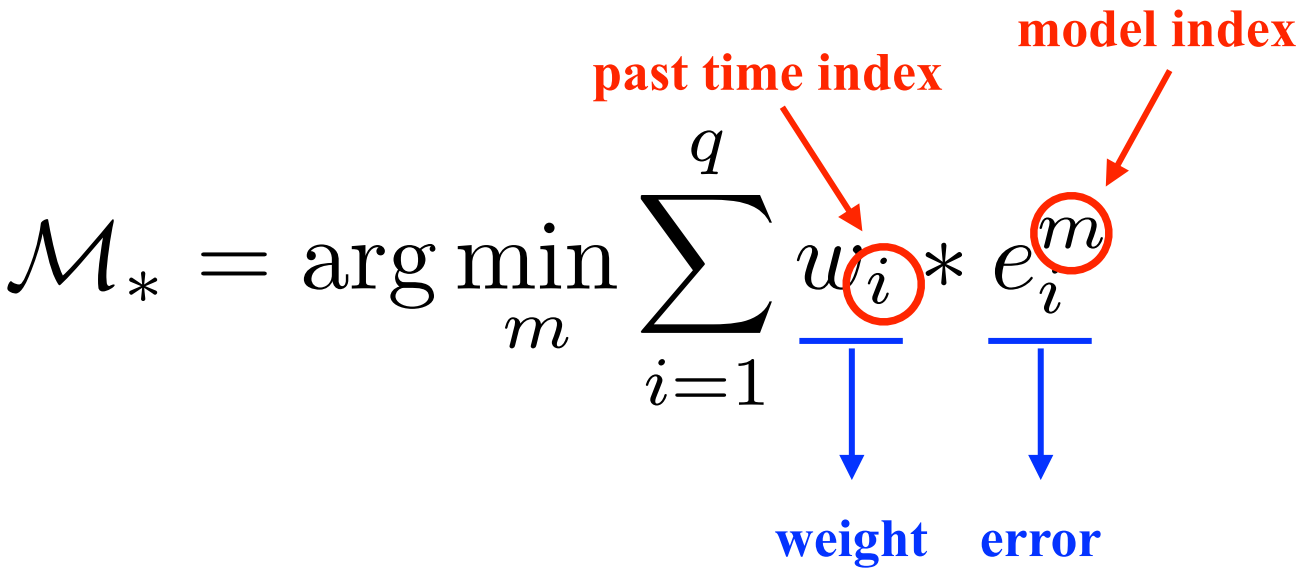
MTGP

Online Model Switching

We develop a novel online model switching strategy, i.e., "weighted Follow-the-Leader" (wFTL) which

- quickly adapts to specific patient given short observations.
- is optimized for recent performance.

The model being pick at time t^* is selected by $\mathcal{M}_* = \arg\min_m \sum_{i=1}^{q} w_i * e_i^m$.



Error Weighting

"weighted Follow-the-Leader" (wFTL): $\mathcal{M}_* = \arg\min_m \sum_{i=1}^{q} w_i * e_i^m$

Intuition: errors made far away should be less penalized compared to the most recent errors.

- **Square exponential kernel:** $K_{se}(t_i, t^*) = \exp\left(-\frac{(t_i - t^*)^2}{\gamma}\right)$
- **Mean reverting kernel:** $K_{mr}(t_i, t^*) = \exp\left(-\frac{|t_i - t^*|}{\gamma}\right)$

where t_i is the all the past time stamps, $i = 1, 2, \cdots q$ and γ is the bandwidth parameter.

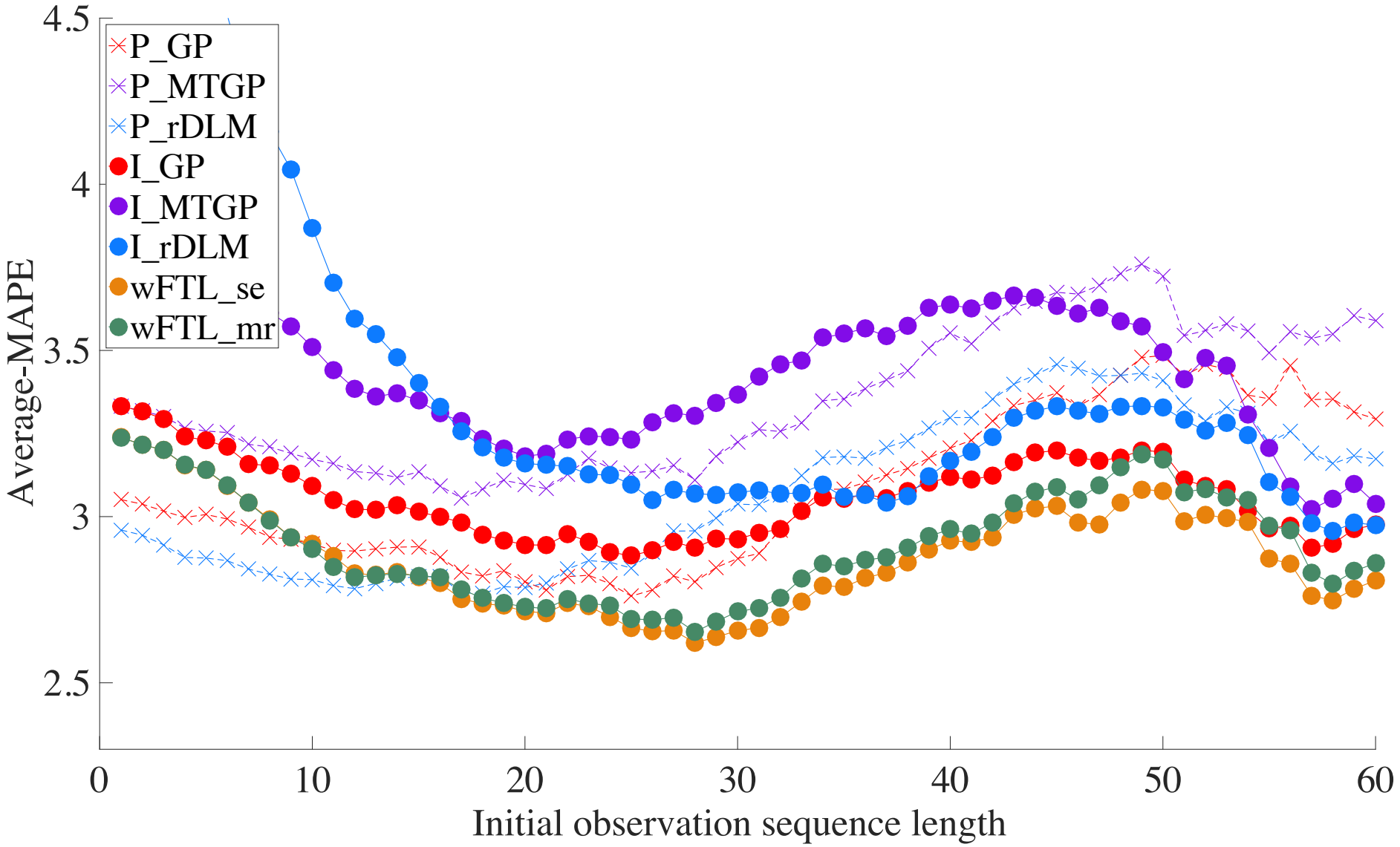
Experiment

- Clinical data: 500 patients who had their *Complete Blood Count* (CBC) tests in PCP database [Haus 10].
- Evaluation metric: Mean Absolute Percentage Error (MAPE)

$$\text{Average-MAPE} = \frac{\sum_{l=1}^{N}\sum_{j=1}^{n}\sum_{i=1}^{T_l}|1-\hat{y}_{j,i}^{l}/y_{j,i}^{l}|}{n\sum_{l=1}^{N}T_l} \times 100\%$$

- Prediction task: per (patient(l), time stamp(t))

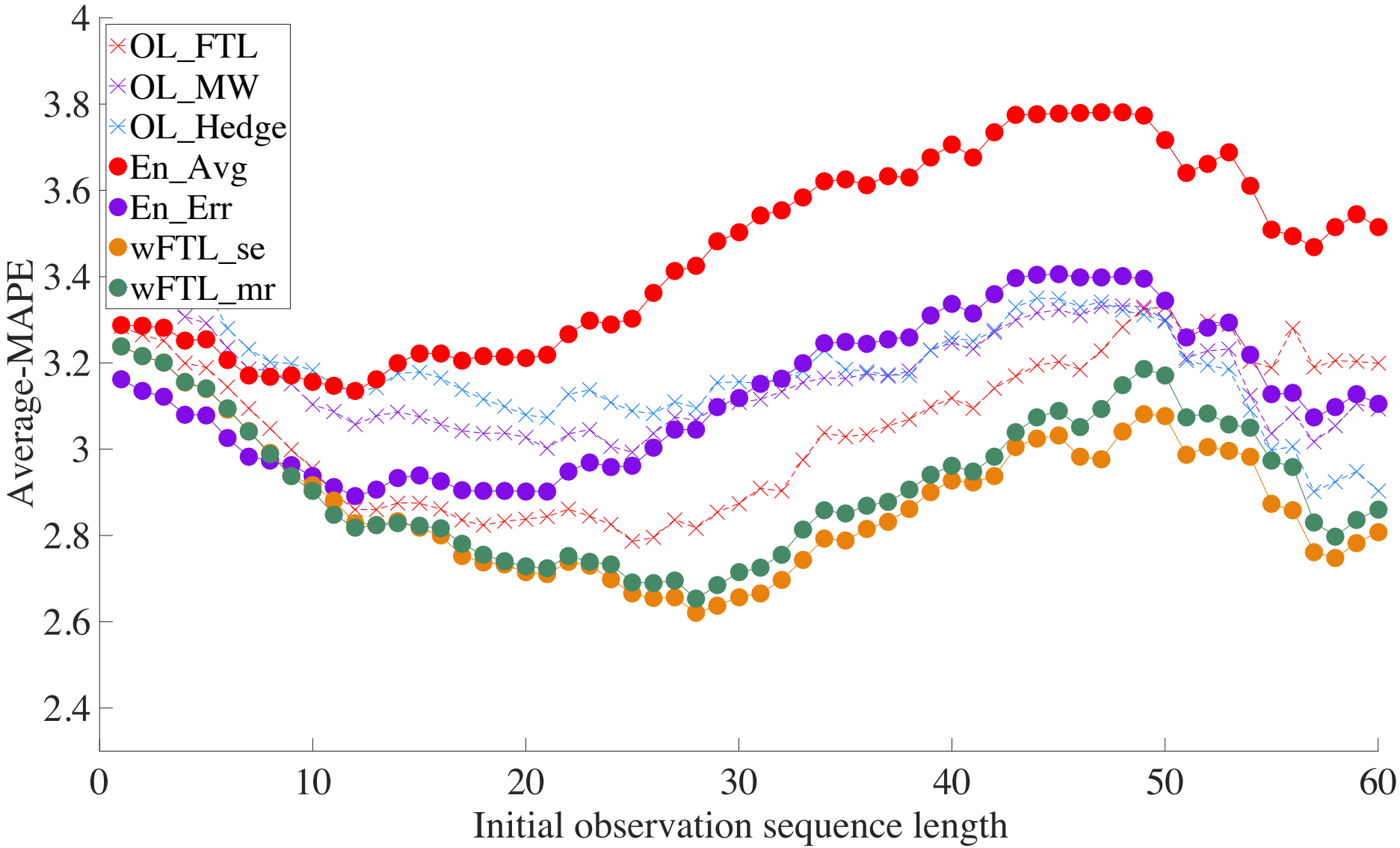
Population-based V.S. Patient-specific Models



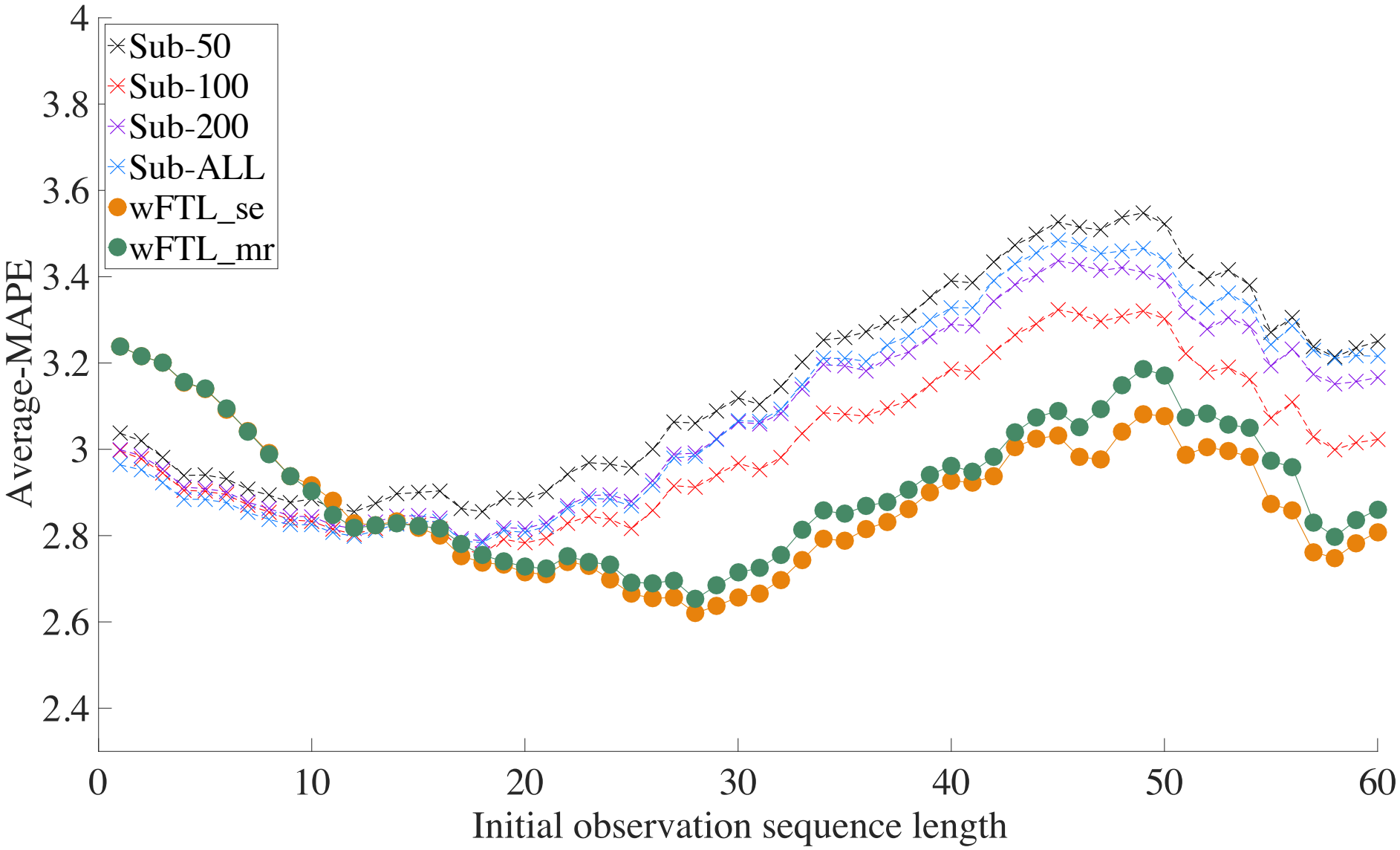
Baselines

- Subpopulation methods
  - *Sub-k*: For each patient at each time stamp, top k similar patients are selected and are used to train the rDLM model.
- Model adaptation
  - *rDLM+reGP* [Liu 15]
  - *rDLM+reMTGP* [Liu 15]
- Adaptive model selection
  - Ensemble methods
    - *En_Avg*: uniformly averaging
    - *En_Err*: exponential weight method
  - Online learning
    - *OL_FTL*: Follow-the-Leader
    - *OL_MW*: Multiplicative weights algorithm $w_m^+ = w_m(1 - \eta e_m)$
    - *OL_Hedge*: Hedge algorithm $w_m^+ = w_m \exp(-\eta e_m)$

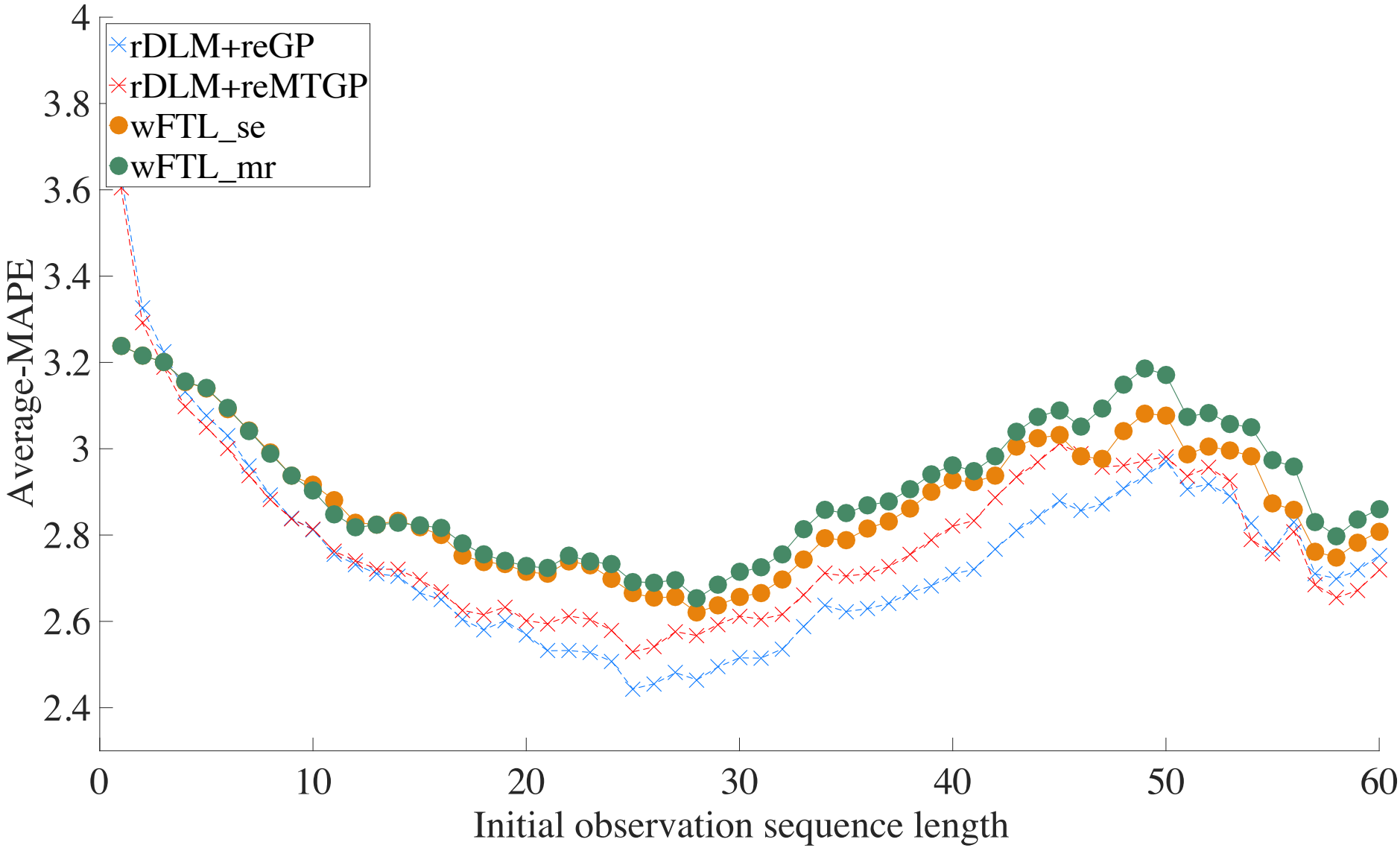
Comparison of Results for Ensemble and Online Methods



Comparison of Results for Subpopulation Methods



Comparison of Results for Model Adaptation Approaches



Summary

- Build a pool of population-based and patient-specific models.
- Develop a novel online model switching strategy, i.e., "weighted Follow-the-Leader"
- Experiment with real-world clinical data.

Reference I

E. V. Bonilla, K. M. Chai, and C. Williams.
"Multi-task Gaussian process prediction".
In: *Advances in Neural Information Processing Systems*, pp. 153–160, Whistler, B.C., Canada, 2007.

M. Hauskrecht, M. Valko, I. Batal, G. Clermont, S. Visweswaran, and G. Cooper.
"Conditional outlier detection for clinical alerting".
In: *AMIA Annual Symposium Proceedings*, pp. 286–290, Washington DC, USA, 2010.

Z. Liu and M. Hauskrecht.
"A regularized linear dynamical system framework for multivariate time series analysis".
In: *The Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp. 1798–1804, Austin, Texas, USA, 2015.

Thank You
Q & A