Learning Adaptive Forecasting Models from Irregularly Sampled Multivariate Clinical Data - Supplemental Material

Zitao Liu and **Milos Hauskrecht**
Computer Science Department, University of Pittsburgh, 210 South Bouquet St., Pittsburgh, PA, 15260 USA

A1. Kalman Filter Algorithm

For each patient *l*, we applied the DVI techniques on $\mathbf{Y}^l$ to get the regularly sampled time series $\tilde{\mathbf{Y}}^l$. All the following operations are performed on the discretized data $\tilde{\mathbf{Y}}^l$ and we will repeat the Kalman filter algorithm on every patient sequence independently. For the sake of notational brevity, we omit the explicit sample index (*l*) and the tilde sign $\tilde{(\cdot)}$ in the rest of Section A1.

Let $\mathbf{Y}_{i:j}$ be the multivariate sequence segment from the *i*th observations to the *j*th observations. Let denote $\hat{\mathbf{z}}_{t|T} \equiv \mathbb{E}[\mathbf{z}_t|\mathbf{Y}]$, $M_{t|T} \equiv \mathbb{E}[\mathbf{z}_t\mathbf{z}_t^\top|\mathbf{Y}]$, $M_{t,t-1|T} \equiv \mathbb{E}[\mathbf{z}_t\mathbf{z}_{t-1}^\top|\mathbf{Y}]$, $P_{t|T} = \mathbb{VAR}[\mathbf{z}_t|\mathbf{Y}]$, and $P_{t,t-1|T} = \mathbb{VAR}[\mathbf{z}_t\mathbf{z}_{t-1}^\top|\mathbf{Y}]$. Let $\hat{\mathbf{z}}_{t|t-1}$ be the *priori* estimation of $\mathbb{E}[\mathbf{z}_t|\mathbf{Y}_{1:t-1}]$, $\hat{\mathbf{z}}_{t-1|t-1}$ be the *posteriori* estimation of $\mathbb{E}[\mathbf{z}_{t-1}|\mathbf{Y}_{1:t-1}]$, $P_{t|t-1}$ be the *priori* estimate error covariance of $\mathbb{E}[(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1})(\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1})^\top]$ and $P_{t-1|t-1}$ be the *posteriori* estimate error covariance of $\mathbb{E}[(\mathbf{z}_{t-1} - \hat{\mathbf{z}}_{t-1|t-1})(\mathbf{z}_{t-1} - \hat{\mathbf{z}}_{t-1|t-1})^\top]$.

Kalman filter algorithm is used to infer the expectations at current time stamp (t) given the current observations ($\mathbf{Y}_{1:t}$), which is summarized in Algorithm 1.

Algorithm 1 Kalman filter algorithm for LDS

INPUT:
- MTS data $\mathbf{Y}$.
- Current step LDS parameters: $\Omega = \{A, C, Q, R, \boldsymbol{\xi}, \Psi\}$.

PROCEDURE:
```
// Initialize the iteration
ẑ_{1|1} = ξ and P_{1|1} = Ψ.
// Start the iteration
for t = 2 → T do
    // Time Update:
    ẑ_{t|t-1} = A ẑ_{t-1|t-1}
    P_{t|t-1} = A P_{t-1|t-1} A^⊤ + Q
    // Measure Update:
    K_t = P_{t|t-1} C^⊤ (C P_{t|t-1} C^⊤ + R)^{-1}
    ẑ_{t|t} = ẑ_{t|t-1} + K_t (y_t - C ẑ_{t|t-1})
    P_{t|t} = P_{t|t-1} - K_t C P_{t|t-1}
end for
```
OUTPUT: $\{\hat{\mathbf{z}}_{t|t-1}\}_{t=2}^T$, $\{\hat{\mathbf{z}}_{t|t}\}_{t=1}^T$, $\{P_{t|t}\}_{t=1}^T$, $\{P_{t|t-1}\}_{t=2}^T$ and $\{K_t\}_{t=1}^T$.

A2. Backward Algorithm in EM - E step

Here, we use the same notation defined in Section A1. Backward algorithm is used to infer the expectations at current time stamp (t) given the entire observations ($\mathbf{Y}$), which is summarized in Algorithm 2.

Algorithm 2 EM: E-step backward algorithm for LDS

INPUT:
- Output from Kalman filter algorithm: $\{\hat{\mathbf{z}}_{t|t-1}\}_{t=2}^T$, $\{\hat{\mathbf{z}}_{t|t}\}_{t=1}^T$, $\{P_{t|t}\}_{t=1}^T$, $\{P_{t|t-1}\}_{t=2}^T$ and $\{K_t\}_{t=1}^T$. Kalman filter algorithm is presented in Algorithm 1 in Section A1.
- Current step LDS parameters: $\Omega = \{A, C, Q, R, \boldsymbol{\xi}, \Psi\}$.

PROCEDURE:
```
// Initialize the iteration
M_{T|T} = P_{T|T} + ẑ_{T|T} ẑ_{T|T}^⊤
J_{T-1} = P_{T-1|T-1} A^⊤ (P_{T|T-1})^{-1}
P_{T-1|T} = P_{T-1|T-1} + J_{T-1} (P_{T|T} - P_{T|T-1}) J_{T-1}^⊤
ẑ_{T-1|T} = ẑ_{T-1|T-1} + J_{T-1} (ẑ_{T|T} - A ẑ_{T-1|T-1})
P_{T,T-1|T} = (I - K_T C) A P_{T-1|T-1}
M_{T,T-1|T} = P_{T,T-1|T} + ẑ_{T|T} ẑ_{T-1|T}^⊤
// Start the iteration
for t = T-1 → 1 do
    M_{t|T} = P_{t|T} + ẑ_{t|T} ẑ_{t|T}^⊤
    J_{t-1} = P_{t-1|t-1} A^⊤ (P_{t|t-1})^{-1}
    P_{t,t-1|T} = P_{t|t} J_{t-1}^⊤ + J_t (P_{t+1,t|T} - A P_{t|t}) J_{t-1}^⊤
    M_{t,t-1|T} = P_{t,t-1|T} + ẑ_{t|T} ẑ_{t-1|T}^⊤
    ẑ_{t-1|T} = ẑ_{t-1|t-1} + J_{t-1} (ẑ_{t|T} - A ẑ_{t-1|t-1})
    P_{t-1|T} = P_{t-1|t-1} + J_{t-1} (P_{t|T} - P_{t|t-1}) J_{t-1}^⊤
end for
```
OUTPUT: $\{\hat{\mathbf{z}}_{t-1|T}\}_{t=1}^T$, $\{M_{t|T}\}_{t=1}^T$ and $\{M_{t,t-1|T}\}_{t=1}^T$.

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

A3. Update Rules in EM - M step

The updated rules of $\Omega = \{A, C, Q, R, \boldsymbol{\xi}, \Psi\}$ are shown in eq.(1) - eq.(6).

$$
\begin{aligned}
A^{(k+1)} &= \Big(\sum_{l=1}^{N}\sum_{t=2}^{\tilde{T}_l}\mathbb{E}[\mathbf{z}_t^l(\mathbf{z}_{t-1}^l)^\top|\tilde{\mathbf{Y}}^l]\Big) \\
&\cdot \Big(\sum_{l=1}^{N}\sum_{t=2}^{\tilde{T}_l}\mathbb{E}[\mathbf{z}_{t-1}^l(\mathbf{z}_{t-1}^l)^\top|\tilde{\mathbf{Y}}^l]\Big)^{-1} \qquad (1)\\
C^{(k+1)} &= \Big(\sum_{l=1}^{N}\sum_{t=1}^{\tilde{T}_l}\tilde{\mathbf{y}}_t^l(\mathbb{E}[\mathbf{z}_t^l|\tilde{\mathbf{Y}}^l])^\top\Big) \\
&\cdot \Big(\sum_{l=1}^{N}\sum_{t=1}^{\tilde{T}_l}\mathbb{E}[\mathbf{z}_t^l(\mathbf{z}_t^l)^\top|\tilde{\mathbf{Y}}^l]\Big)^{-1} \qquad (2)\\
R^{(k+1)} &= \frac{1}{\sum_{l=1}^{N}\tilde{T}_l}\Big(\sum_{l=1}^{N}\sum_{t=1}^{\tilde{T}_l}\tilde{\mathbf{y}}_t^l(\tilde{\mathbf{y}}_t^l)^\top \\
&- C^{(k+1)}\sum_{l=1}^{N}\sum_{t=1}^{\tilde{T}_l}\mathbb{E}[\mathbf{z}_t^l|\tilde{\mathbf{Y}}^l](\tilde{\mathbf{y}}_t^l)^\top\Big) \qquad (3)\\
Q^{(k+1)} &= \frac{1}{\sum_{l=1}^{N}\tilde{T}_l - N}\Big(\sum_{l=1}^{N}\sum_{t=2}^{\tilde{T}_l}\mathbb{E}[\mathbf{z}_t^l(\mathbf{z}_t^l)^\top|\tilde{\mathbf{Y}}^l] \\
&- A^{(k+1)}\sum_{l=1}^{N}\sum_{t=2}^{\tilde{T}_l}\mathbb{E}[\mathbf{z}_{t-1}^l(\mathbf{z}_t^l)^\top|\tilde{\mathbf{Y}}^l]\Big) \qquad (4)\\
\boldsymbol{\xi}^{(k+1)} &= \sum_{l=1}^{N}\mathbb{E}[\mathbf{z}_1^l|\tilde{\mathbf{Y}}^l] \qquad (5)\\
\Psi^{(k+1)} &= \sum_{l=1}^{N}\mathbb{E}[\mathbf{z}_1^l(\mathbf{z}_1^l)^\top|\tilde{\mathbf{Y}}^l] - \sum_{l=1}^{N}\mathbb{E}[\mathbf{z}_1^l|\tilde{\mathbf{Y}}^l](\mathbb{E}[\mathbf{z}_1^l|\tilde{\mathbf{Y}}^l])^\top \\
& \qquad (6)
\end{aligned}
$$

A4. More Qualitative Results

More qualitative results are shown in Figure 1 to Figure 12.

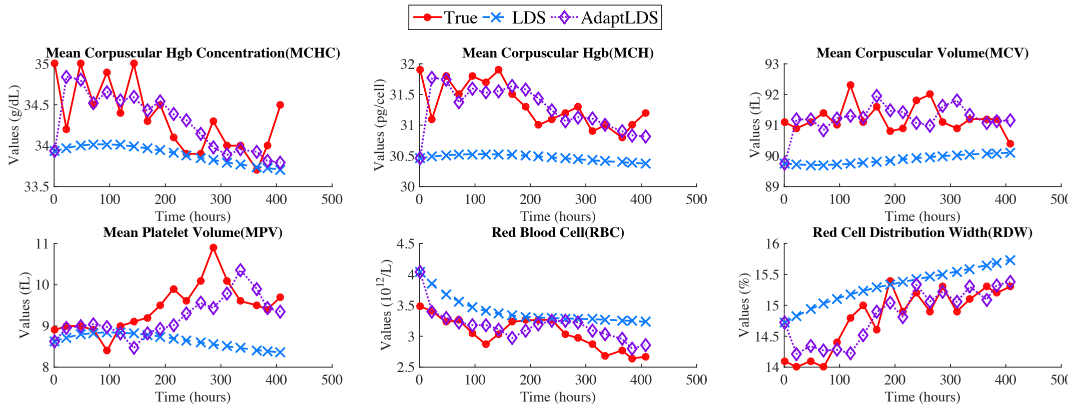

Figure 1: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

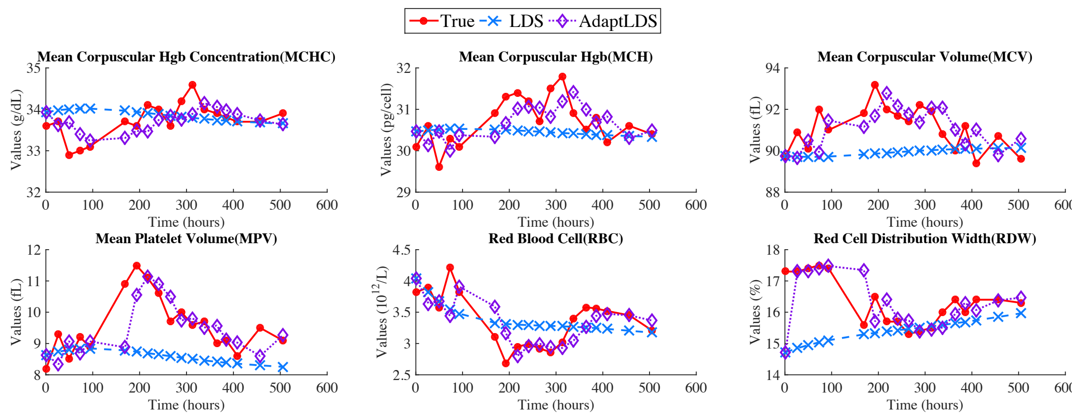

Figure 2: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

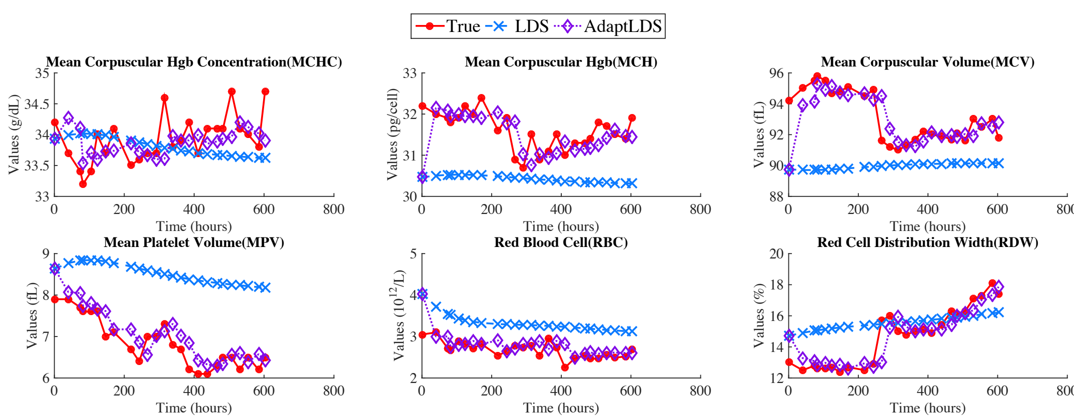

Figure 3: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

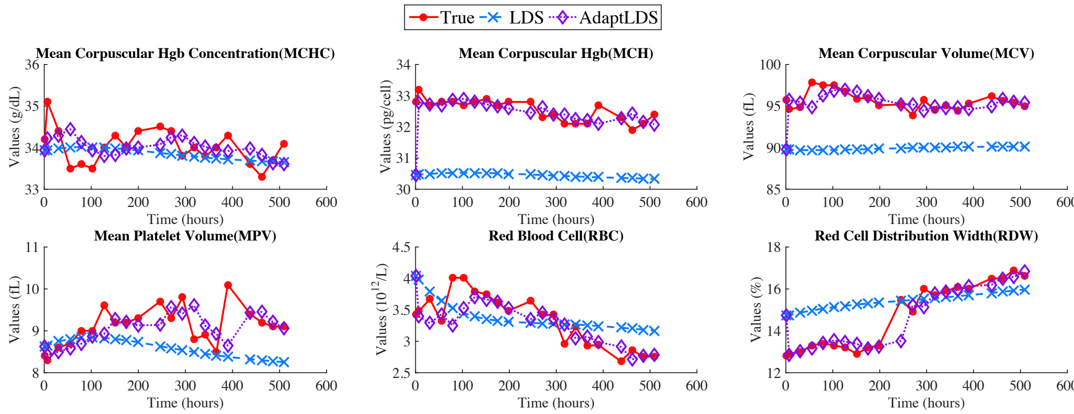

Figure 4: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

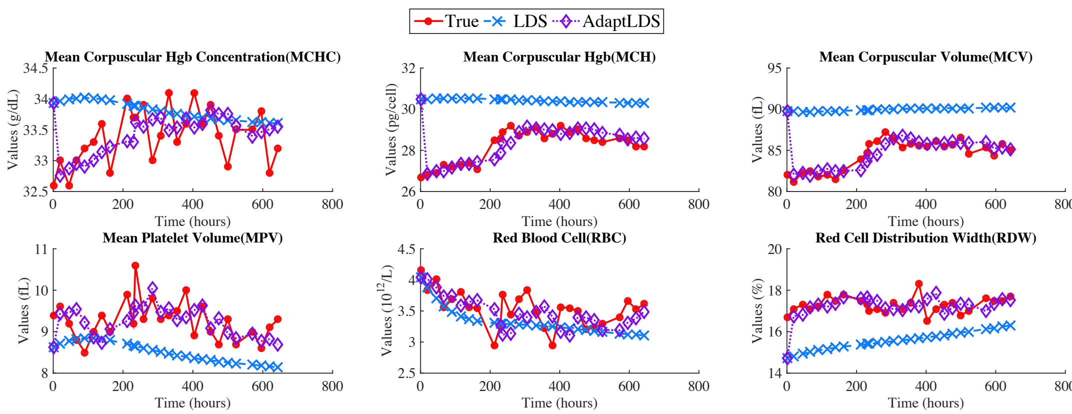

Figure 5: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

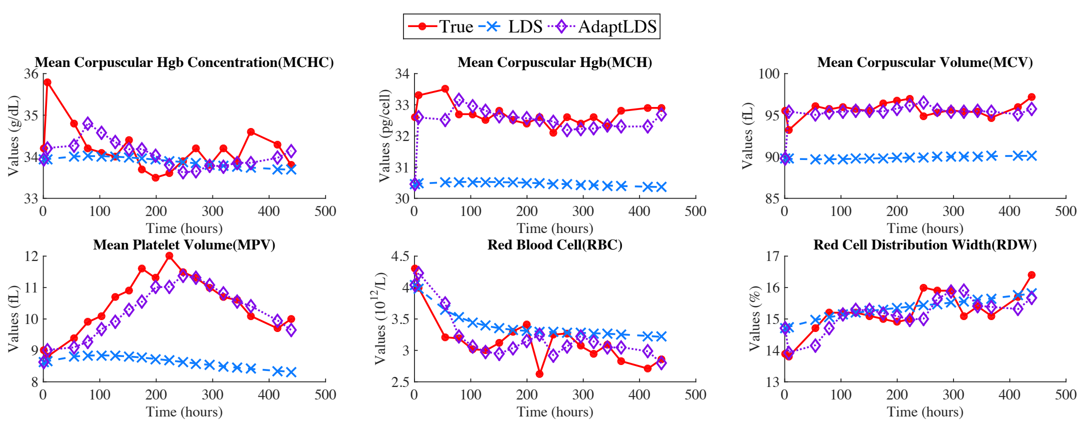

Figure 6: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

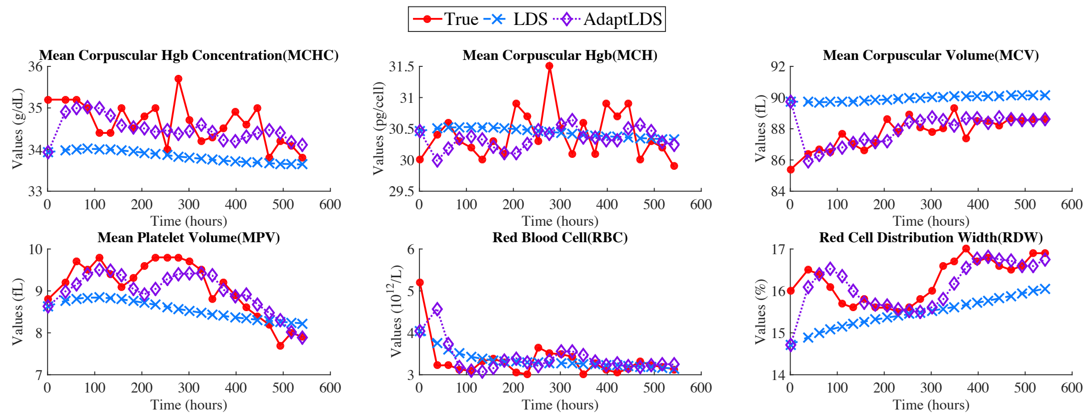

Figure 7: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

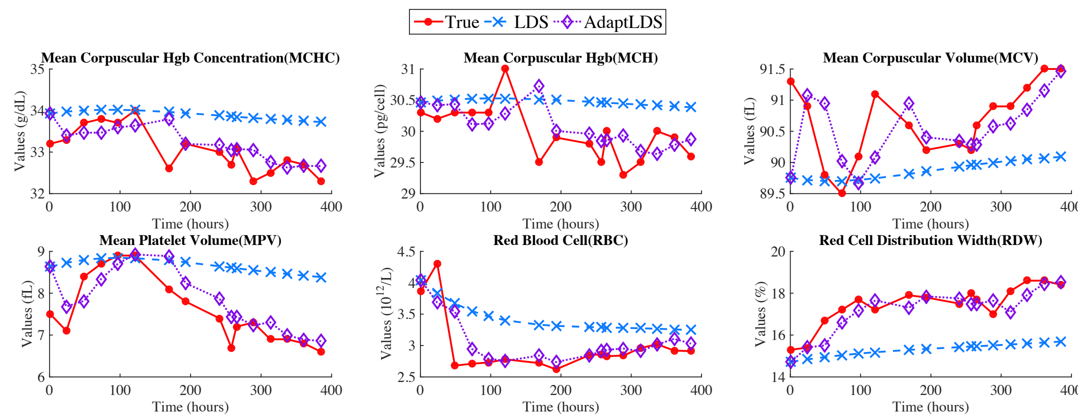

Figure 8: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

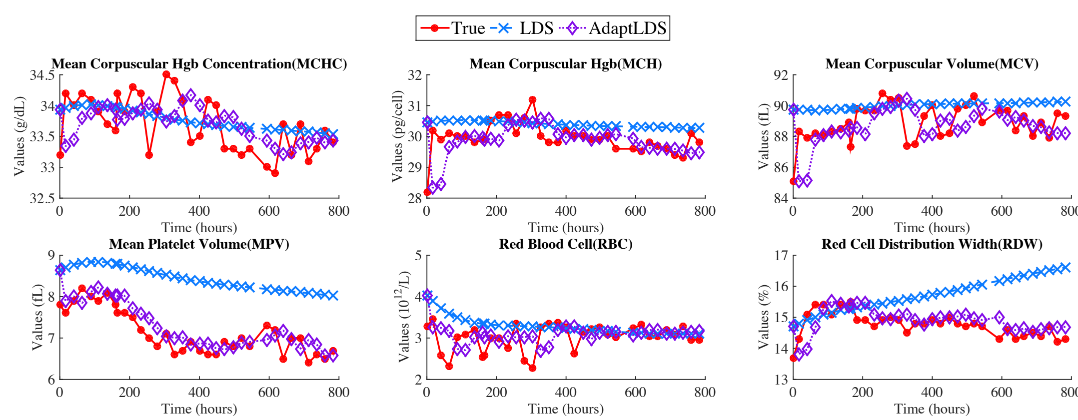

Figure 9: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

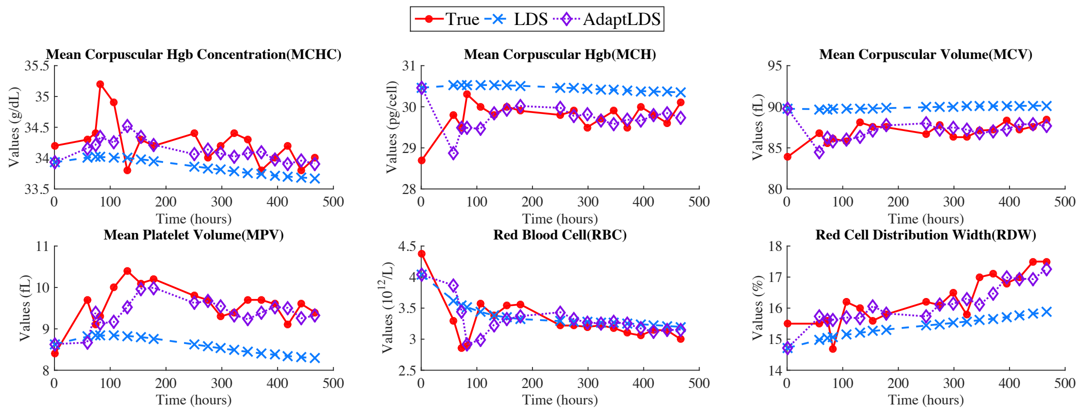

Figure 10: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

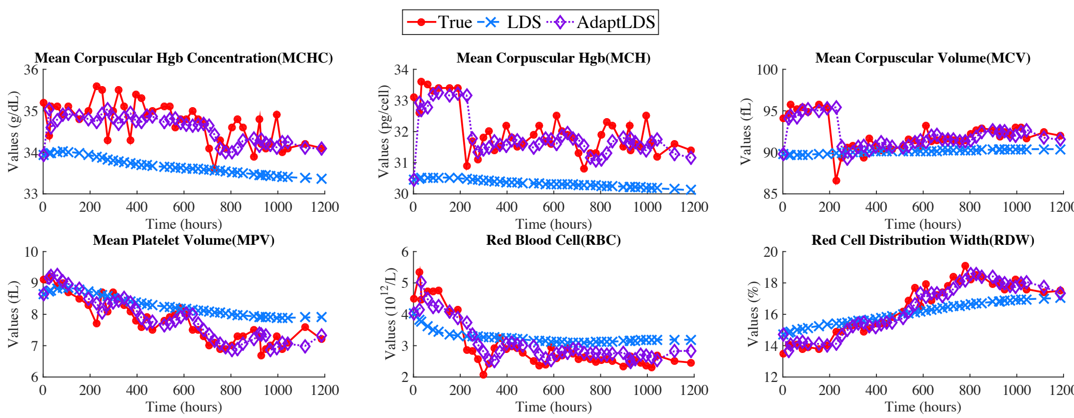

Figure 11: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.

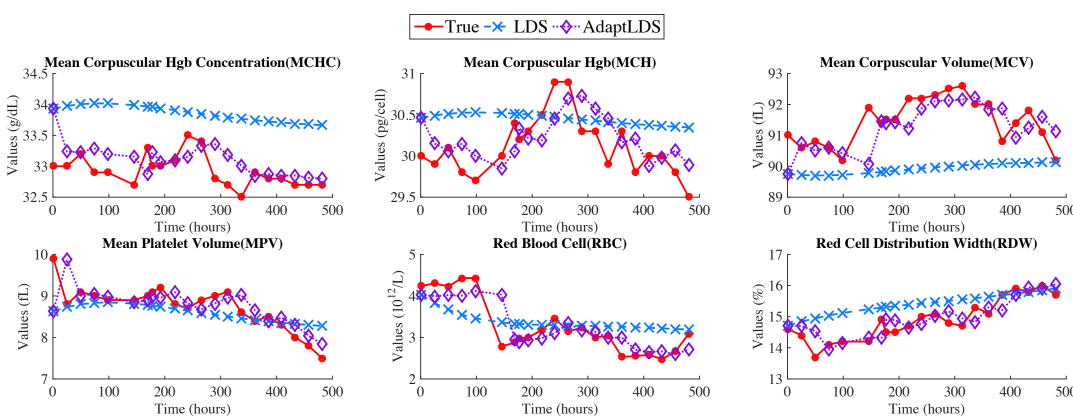

Figure 12: Clinical MTS predictions for one patient. The population based LDS model is trained on 400 patient sequences.