Bayesian Online Prediction of Change Points (Supplementary Material)

We provide detailed descriptions of the algorithms used within the proposed framework together with improved visualizations of the plots for the sake of clarity. Note that the presented plots can be reproduced using Python and following the instructions given in the software repository[1] (README file).

```
Parameters : Hazard function H(·).
             Observation model prior hyperparameters Θ_prior.
for each new observation y_t do
    for r_t = 0, ..., R do
        π_t(r_t) = p(y_t | Θ_{t-1}^{r_t})                      ▷ Compute p(y_t | r_t, Y^{r_t}) using sufficient statistics
    for r_t = 1, ..., R do
        α_t(r_t) = (1 − H(r_t − 1)) ∗ π_t(r_t) ∗ γ_{t-1}(r_t − 1)   ▷ Compute growth probabilities
    α_t(0) = 0
    for r_t = 0, ..., R do
        α_t(0) = α_t(0) + H(r_t) ∗ π_t(0) ∗ γ_{t-1}(r_t)         ▷ Compute change point probabilities
    Θ_t^0 = Θ_prior
    for r_t = 1, ..., R do
        Update Θ_t^{r_t} based on Θ_{t-1}^{r_t − 1} and y_t       ▷ Update observation model sufficient statistics
    e_t = 0
    for r_t = 0, ..., R do
        e_t = e_t + α_t(r_t)                                     ▷ e_t = p(y_t | Y_{1:t-1}) normalizes α_t(r_t)
    output γ_t(r_t) = α_t(r_t)/e_t                               ▷ denotes the run length posterior γ_t(r_t) = p(r_t | Y_{1:t})
```

Algorithm 1: Bayesian Online Change Point Detection (BOCPD). This algorithm has complexity $\mathcal{O}(D)$ per online update. D denotes the maximum total duration. Note that it is parameterized in terms of the maximum run length R instead of D ($R+1=D$).

[1]Source code and data are available at a public repository. Download link: `https://github.com/DiegoAE/BOSD`

Parameters : Hazard function $H(\cdot)$.
Run length posterior $\gamma_t(r_t)$ computed in Algorithm 1.

for $r_t = 0, \dots, R$ **do**
  $p = 1$
  for $l_t = 0, \dots, R$ **do**
    if $r_t + l_t <= R$ **then**
      $g(l_t, r_t) = p * H(r_t + l_t)$ ▷ Compute $p(l_t|r_t)$ offline
      $p = p * (1 - H(r_t + l_t))$

for $t = 0, \dots, T$ **do**
  for $l_t = 0, \dots, R$ **do**
    $w_t(l_t) = 0$
    for $r_t = 0, \dots, R$ **do**
      $w_t(l_t) = w_t(l_t) + g(l_t, r_t) * \gamma_t(r_t)$ ▷ Recall that $\gamma(r_t) = p(r_t|\mathbf{Y}_{1:t})$
  outputs $w_t(l_t)$ ▷ denotes the posterior $p(l_t|\mathbf{Y}_{1:t})$

Algorithm 2: Residual time prediction for BOCPD. Note that given the run length posterior $p(r_t|\mathbf{Y}_{1:t})$ for all time steps the algorithm is independent of the actual observations. The computational complexity per online update is $\mathcal{O}(D^2)$. We have presented the run length inference (Algorithm 1) and the residual time inference (Algorithm 2) as separate algorithms for the sake of clarity; however, we can easily combine them within the BOCPD framework to obtain a fully online algorithm (assuming $p(l_t|r_t)$ is precomputed).

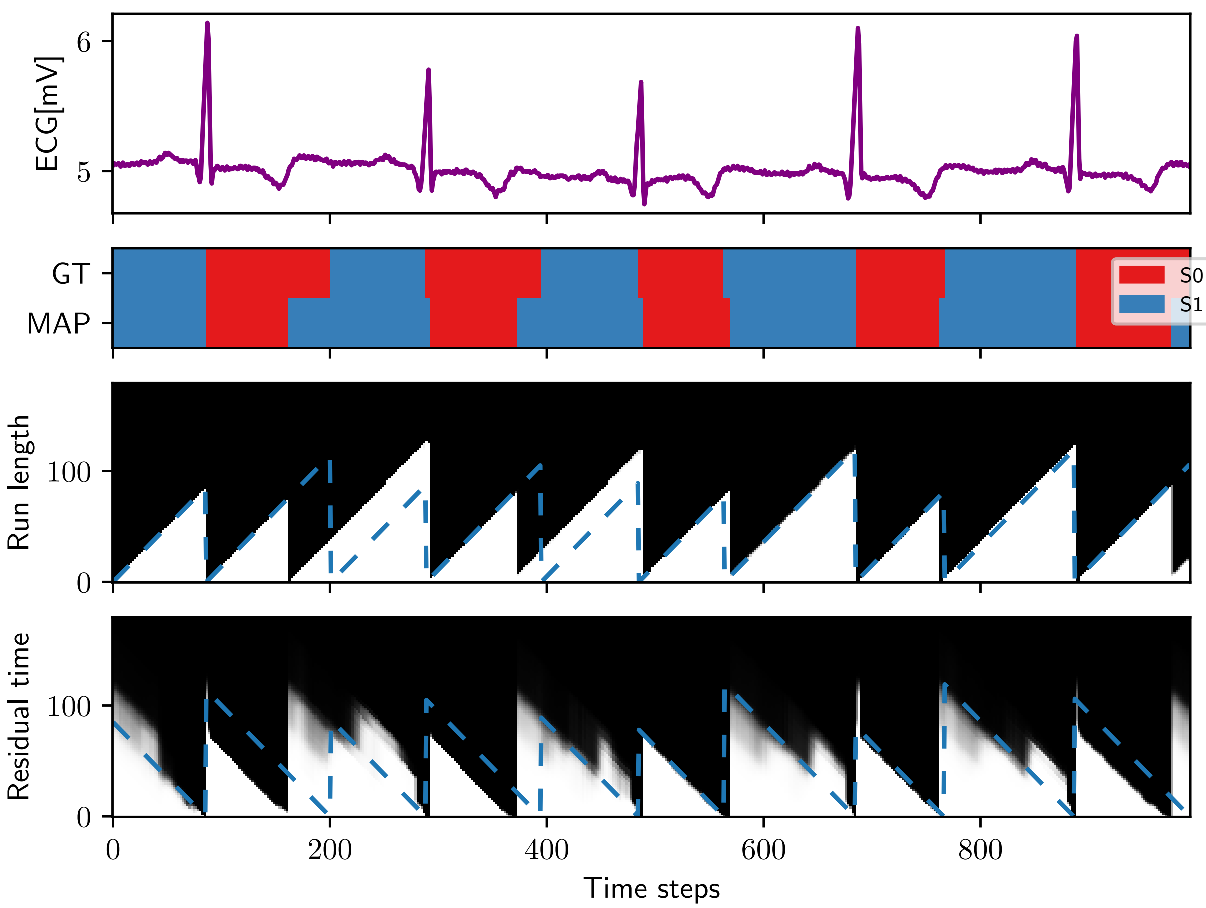

Figure 1: ECG segmentation

Parameters : Duration matrix $\mathbf{D}(\cdot,\cdot)$ for each hidden state z_t.
Transition matrix $\mathbf{A}(\cdot,\cdot)$
Observation model prior hyperparameters $\Theta_{z_t.prior}$.

for *each new observation* $\mathbf{y}_t$ **do**
  for $z_t = 1, \ldots, K$ **do**
    for $d_t = 1, \ldots, D$ **do**
      for $r_t = 0, \ldots, d_t - 1$ **do**
        $\pi_t(r_t, d_t, z_t) = p(\mathbf{y}_t | \Theta_{t-1}^{r_t, d_t, z_t})$ ▷ Compute $p(\mathbf{y}_t | r_t, d_t, z_t, \mathbf{Y}^{r_t})$
  for $z_t = 1, \ldots, K$ **do**
    for $d_t = 1, \ldots, D$ **do**
      for $r_t = 0, \ldots, d_t - 1$ **do**
        $\alpha_t(r_t, d_t, z_t) = \pi_t(r_t, d_t, z_t) * \gamma_{t-1}(r_t - 1, d_t, z_t)$ ▷ Growth probabilities
  for $z_{t-1} = 1, \ldots, K$ **do**
    $\eta_t(z_{t-1}) = 0$
    for $d_{t-1} = 1, \ldots, D$ **do**
      $\eta_t(z_{t-1}) = \eta_t(z_{t-1}) + \gamma_{t-1}(d_{t-1} - 1, d_{t-1}, z_{t-1})$
    for $z_t = 1, \ldots, K$ **do**
      $\beta_t(z_t) = \beta_t(z_t) + \mathbf{A}(z_{t-1}, z_t) * \eta_t(z_{t-1})$ ▷ Change point probabilities
  for $z_t = 0, \ldots, K$ **do**
    for $d_t = 1, \ldots, D$ **do**
      $\alpha_t(0, d_t, z_t) = \mathbf{D}(z_d, d_t) * \pi_t(0, d_t, z_t) * \beta_t(z_t)$ ▷ Duration likelihood at a CP
  Update $\Theta_t^{r_t, d_t, z_t}$ based on $\Theta_{t-1}^{r_t - 1, d_t, z_t}$ and $\mathbf{y}_t$ ▷ Update sufficient statistics
  $e_t = 0$
  for $z_t = 1, \ldots, K$ **do**
    for $d_t = 1, \ldots, D$ **do**
      for $r_t = 0, \ldots, d_t - 1$ **do**
        $e_t = e_t + \alpha_t(r_t, d_t, z_t)$ ▷ $e_t = p(\mathbf{y}_t | \mathbf{Y}_{1:t-1})$ normalizes $\alpha_t(r_t, d_t, z_t)$
  output $\gamma_t(r_t, d_t, z_t) = \alpha_t(r_t, d_t, z_t) / e_t$ ▷ denotes the posterior $p(r_t, d_t, z_t | \mathbf{Y}_{1:t})$

Algorithm 3: Bayesian Online Segment Detection. This algorithm has complexity $\mathcal{O}(K^2 + KD^2)$, where K denotes the number of hidden states. We leave as an open question whether it is possible to achieve a better complexity on D under certain conditions. As in the original BOCPD formulation we take advantage of observation models whose likelihood can be computed incrementally through a set of sufficient statistics (i.e., exponential family likelihoods). For arbitrary observation models we get $\mathcal{O}(D^2)$ complexity for BOCPD and $\mathcal{O}(K^2 + KD^3)$ for BOSD.

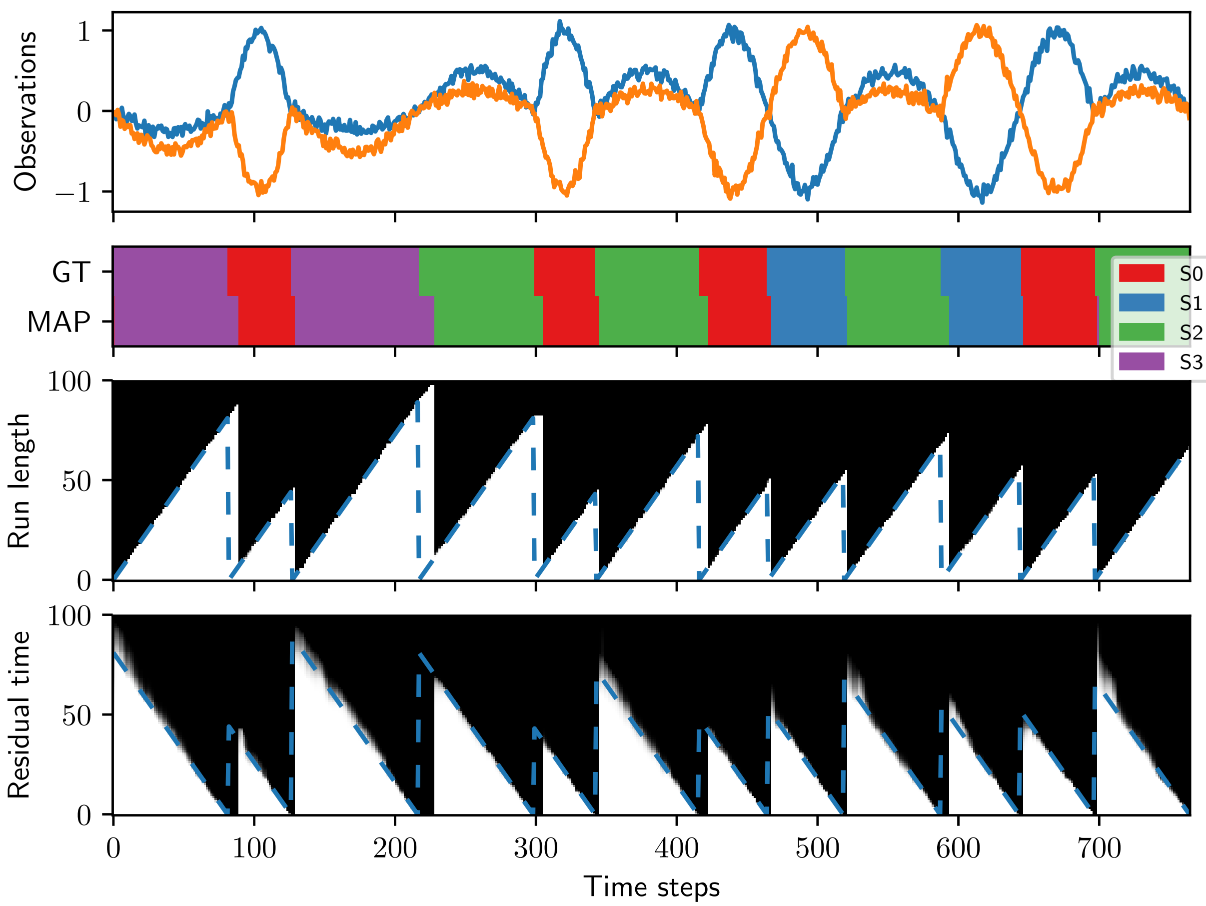

Figure 2: Synthetic experiment

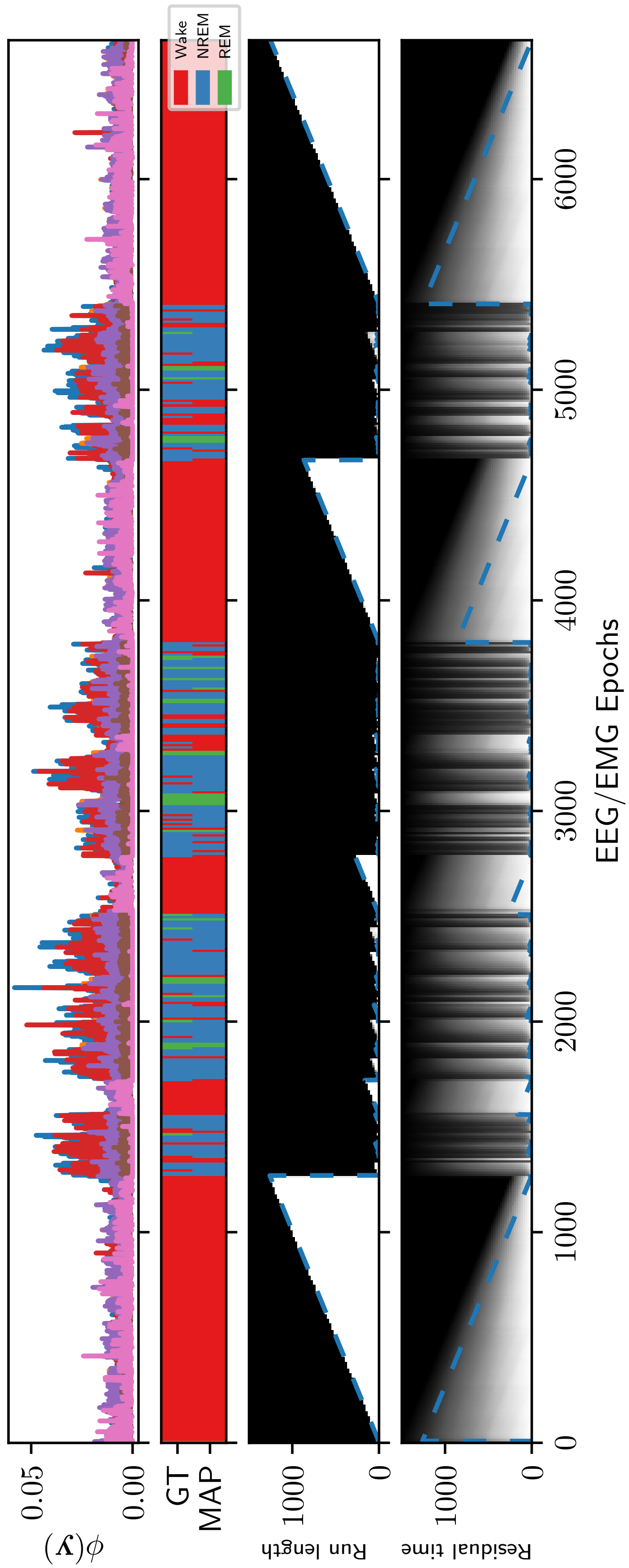

Figure 3: Mice sleep staging through Bayesian online segment detection