Learning Time Series Detection Models from Temporally Imprecise Labels: Supplementary Material

1 Model Derivations

1.1 Gradient Derivations

In this section, we present the complete derivations for the gradient equations presented in section 3 of the main paper. The gradient of the log marginal likelihood $\log p_\psi(z|\boldsymbol{x}, \boldsymbol{t})$ with respect to θ is given by

$$
\begin{aligned}
\nabla_\theta \log p_\psi(z|\boldsymbol{x}, \boldsymbol{t}) &= \frac{1}{p_\psi(z|\boldsymbol{x}, \boldsymbol{t})} \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in O} p_\phi(z|\boldsymbol{o}, \boldsymbol{t}) p_\pi(\boldsymbol{o}|\boldsymbol{y}) \nabla_\theta p_\theta(\boldsymbol{y}|\boldsymbol{x}) \\
&= \frac{1}{p_\psi(z|\boldsymbol{x}, \boldsymbol{t})} \sum_{\boldsymbol{y}\in\mathcal{Y}} p_{\phi,\pi}(z|\boldsymbol{y}, \boldsymbol{t}) p_\theta(\boldsymbol{y}|\boldsymbol{x}) \nabla_\theta \log p_\theta(\boldsymbol{y}|\boldsymbol{x}) \\
&= \frac{1}{p_\psi(z|\boldsymbol{x}, \boldsymbol{t})} \sum_{\boldsymbol{y}\in\mathcal{Y}} p_{\phi,\pi}(z|\boldsymbol{y}, \boldsymbol{t}) p_\theta(\boldsymbol{y}|\boldsymbol{x}) \sum_{i=1}^{L} \nabla_\theta \ell_\theta(y_i|\boldsymbol{x}_i) \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}} \frac{p_\psi(z, \boldsymbol{y}|\boldsymbol{x}, \boldsymbol{t})}{p_\psi(z|\boldsymbol{x}, \boldsymbol{t})} \sum_{i=1}^{L} \nabla_\theta \ell_\theta(y_i|\boldsymbol{x}_i) \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}} p_\psi(\boldsymbol{y}|z, \boldsymbol{x}, \boldsymbol{t}) \sum_{i=1}^{L} \nabla_\theta \ell_\theta(y_i|\boldsymbol{x}_i) \\
&= \sum_{i=1}^{L} \sum_{y_i=0}^{1} p_\psi(y_i|z, \boldsymbol{x}_i, \boldsymbol{t}) \nabla_\theta \ell_\theta(y_i|\boldsymbol{x}_i) \\
&= \sum_{i=1}^{L} \mathbb{E}_{p_\psi(y_i|z, \boldsymbol{x}, \boldsymbol{t})} \nabla_\theta \ell_\theta(y_i|\boldsymbol{x}_i).
\end{aligned}
$$

The gradient of the marginal likelihood $\log p_\psi(z|\boldsymbol{x}, \boldsymbol{t})$ with respect to ϕ is given by

$$\begin{aligned}
\nabla_\phi \log p_\psi(z|\boldsymbol{x}, \boldsymbol{t}) &= \frac{1}{p_\psi(z|\boldsymbol{x}, \boldsymbol{t})} \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in O} p_\theta(\boldsymbol{y}|\boldsymbol{x}) p_\pi(\boldsymbol{o}|\boldsymbol{y}) \nabla_\phi p_\phi(z|\boldsymbol{y}, \boldsymbol{t}) \\
&= \frac{1}{p_\psi(z|\boldsymbol{x}, \boldsymbol{t})} \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in O} p_\phi(z|\boldsymbol{y}, \boldsymbol{t}) p_\theta(\boldsymbol{y}|\boldsymbol{x}) p_\pi(\boldsymbol{o}|\boldsymbol{y}) \sum_{i=1}^{L} \sum_{l=1}^{M} [o_i > 0] \cdot w(i, l) \nabla_\phi \ell_\phi(z_l|t_i) \\
&= \sum_{\boldsymbol{o}\in O} p_\psi(\boldsymbol{o}|z, \boldsymbol{x}, \boldsymbol{t}) \sum_{i=1}^{L} \sum_{l=1}^{M} [o_i > 0] \cdot w(i, l) \nabla_\phi \ell_\phi(z_l|t_i) \\
&= \sum_{i=1}^{L} \sum_{l=1}^{M} p_\psi\left([o_i > 0] w(i, l) | z, \boldsymbol{x}, \boldsymbol{t}\right) \nabla_\phi \ell_\phi(z_l|t_i) \\
&= \sum_{l=1}^{M} \mathbb{E}_{p_\psi(w(i,l)|z,\boldsymbol{x},\boldsymbol{t})} \nabla_\phi \ell_\phi(z_l|t_i).
\end{aligned}$$

The gradient of the marginal likelihood $\log p_\psi(z|\boldsymbol{x}, \boldsymbol{t})$ with respect to π is given by

$$\begin{aligned}
\nabla_\pi \log p_\psi(z|\boldsymbol{x}, \boldsymbol{t}) &= \frac{1}{p_\psi(z|\boldsymbol{x}, \boldsymbol{t})} \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in O} p_\phi(z|\boldsymbol{o}, \boldsymbol{t}) p_\theta(\boldsymbol{y}|\boldsymbol{x}) \nabla_\pi p_\pi(\boldsymbol{o}|\boldsymbol{y}) \\
&= \frac{1}{p_\psi(z|\boldsymbol{x}, \boldsymbol{t})} \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in O} p_{\phi,\pi}(z|\boldsymbol{o}, \boldsymbol{t}) p_\theta(\boldsymbol{y}|\boldsymbol{x}) p_\pi(\boldsymbol{o}|\boldsymbol{y}) \sum_{i=1}^{L} \nabla_\pi \ell_\pi(o_i|y_i) \\
&= \sum_{i=1}^{L} \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in O} p_\psi(\boldsymbol{o}, \boldsymbol{y}|z, \boldsymbol{x}, \boldsymbol{t}) \nabla_\pi \ell_\pi(o_i|y_i) \\
&= \sum_{i=1}^{L} \mathbb{E}_{p_\psi(o_i,y_i|z,\boldsymbol{x},\boldsymbol{t})} \nabla_\pi \ell_\pi(o_i|y_i).
\end{aligned}$$

1.2 Posterior Marginal Derivations

This section presents the derivation of the posterior marginals $p_\psi(y_i|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{t})$, $p_\psi(w(i,l)|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{t})$, and $p_\psi(o_i, y_i|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{t})$ from section 3.4 of the main paper, given the dynamic programs $a(i,l)$ and $b(i,l)$.

$$\begin{aligned}
p_\psi(w(i,l)|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{t}) &= \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in\mathcal{O}} p_\psi(\boldsymbol{y}, \boldsymbol{o}|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{t}) \cdot w(i,l) \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in\mathcal{O}} \frac{p_\psi(\boldsymbol{y}, \boldsymbol{o}, \boldsymbol{z}|\boldsymbol{x}, \boldsymbol{t})}{p_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{t})} \cdot w(i,l) \\
&= \sum_{c=1}^{M} \sum_{j=l-c}^{l} \frac{p_\psi(\boldsymbol{z}_{1:j-1}|\boldsymbol{x}_{1:i-1}, \boldsymbol{t}_{1:i-1}) p_\psi(\boldsymbol{z}_{j+c+1:M}|\boldsymbol{x}_{i+1:L}, \boldsymbol{t}_{i+1:L}) p_\psi(\boldsymbol{z}_{j:j+c}, o_i = c|\boldsymbol{x}_i, t_i)}{p_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{t})} \\
&= \sum_{c=1}^{M} \sum_{j=l-c}^{l} \frac{p_\psi(\boldsymbol{z}_{1:j-1}|\boldsymbol{x}_{1:i-1}, \boldsymbol{t}_{1:i-1}) p_\psi(\boldsymbol{z}_{j+c+1:M}|\boldsymbol{x}_{i+1:L}, \boldsymbol{t}_{i+1:L}) \prod_{k=j}^{j+c} p_\phi(z_k|t_i) \cdot \sum_{y_i} p_\pi(o_i|y_i) p_\theta(y_i|\boldsymbol{x}_i)}{p_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{t})} \\
&= \sum_{c=1}^{M} \sum_{j=l-c}^{l} \frac{a(i-1, j-1) b(i+1, j+c+1) \prod_{k=j}^{j+c} p_\phi(z_k|t_i) \cdot \sum_{y_i} p_\pi(o_i|y_i) p_\theta(y_i|\boldsymbol{x}_i)}{a(L, M)}
\end{aligned}$$

$$\begin{aligned}
p_\psi(o_i = c, y_i = v|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{t}) &= \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in\mathcal{O}} p_\psi(\boldsymbol{y}, \boldsymbol{o}|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{t}) [o_i = c][y_i = v] \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}} \sum_{\boldsymbol{o}\in\mathcal{O}} \frac{p_\psi(\boldsymbol{y}, \boldsymbol{o}, \boldsymbol{z}|\boldsymbol{x}, \boldsymbol{t})}{p_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{t})} [o_i = c][y_i = v] \\
&= \sum_{j=1}^{M-c} \frac{p_\psi(\boldsymbol{z}_{1:j-1}|\boldsymbol{x}_{1:i-1}, \boldsymbol{t}_{1:i-1}) p_\psi(\boldsymbol{z}_{j+c+1:M}|\boldsymbol{x}_{i+1:L}, \boldsymbol{t}_{i+1:L}) p_\psi(\boldsymbol{z}_{j:j+c}, o_i = c, y_i = v|\boldsymbol{x}_i, t_i)}{p_\psi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{t})} \\
&= \sum_{j=1}^{M-c} \frac{a(i-1, j-1) b(i+1, j+c+1) \prod_{k=j}^{j+c} p_\phi(z_k|t_i) \cdot p_\pi(o_i|y_i) p_\theta(y_i|\boldsymbol{x}_i)}{a(L, M)}
\end{aligned}$$

$$\begin{aligned}
p_\psi(y_i = v|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{t}) &= \sum_{c=0}^{M} p_\psi(o_i = c, y_i = v|\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{t}) \\
&= \sum_{c=0}^{M} \sum_{j=1}^{M-c} \frac{a(i-1, j-1) b(i+1, j+c+1) \prod_{k=j}^{j+c} p_\phi(z_k|t_i) \cdot p_\pi(o_i|y_i) p_\theta(y_i|\boldsymbol{x}_i)}{a(L, M)}
\end{aligned}$$

2 Empirical Noise Distribution

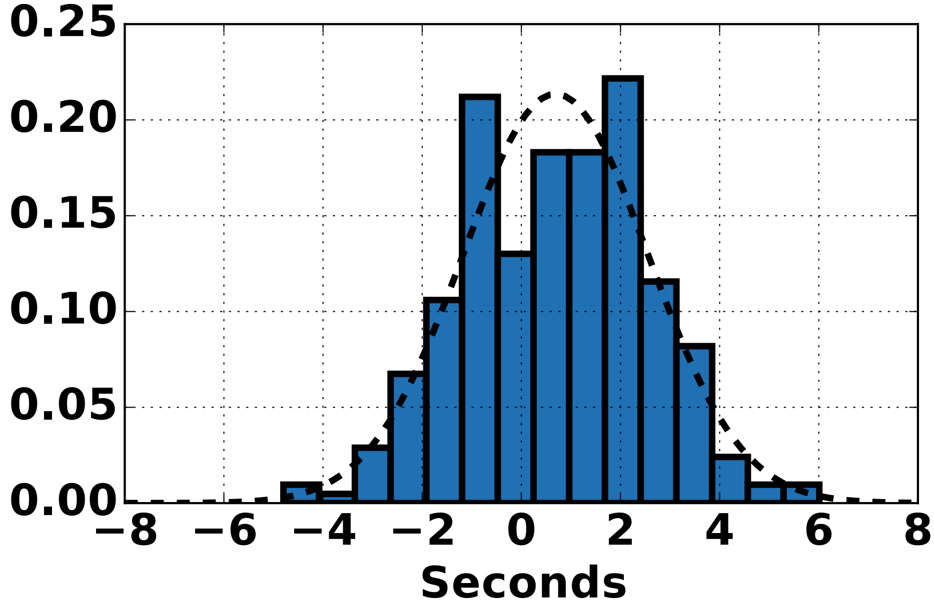
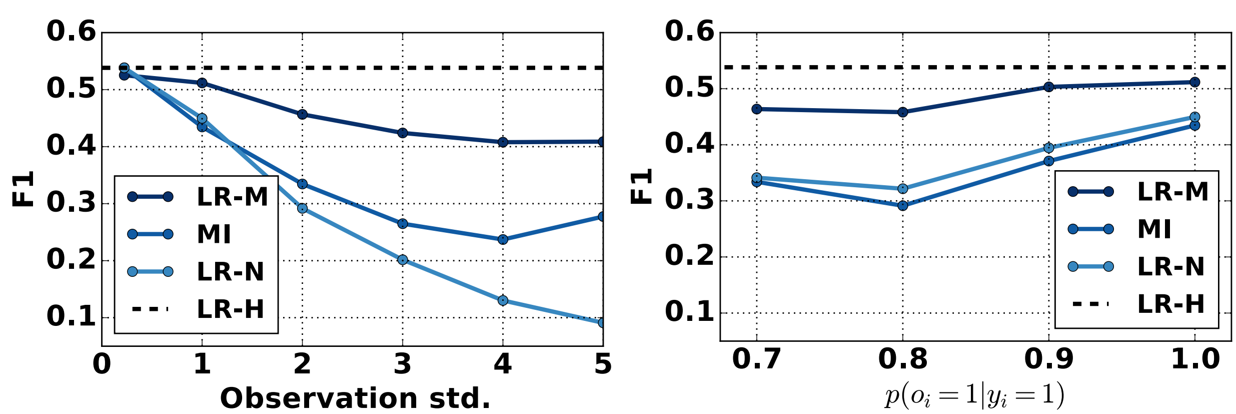


Figure 1: The marginal distribution of the difference between the hand-aligned and observed time stamps for positive instances in the puffMarker data. The dashed lined shows a Normal distribution fit to these data.

3 mPuff Results



(a) (b)

Figure 2: Figures (a) and (b) show the prediction performance for all logistic regression based models when varied amounts of synthetic noise are added to the hand aligned labels of the mPuff dataset. In figure (a) we vary the standard deviation of the timestamp noise and in figure (b) we vary the probability of not observing a positive instance.