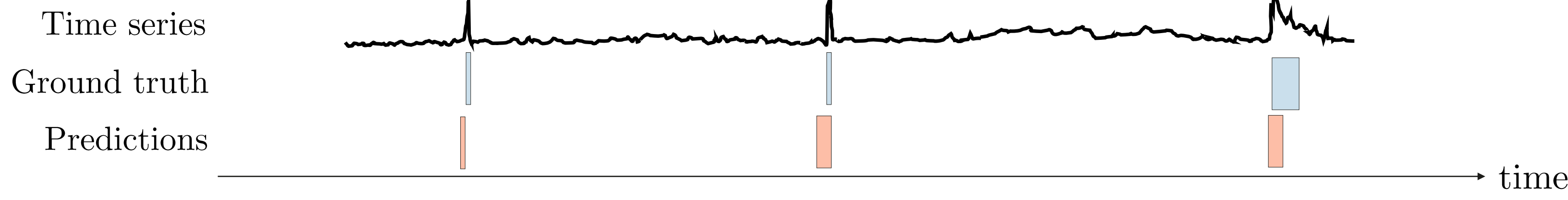


Goal

We define a new pair of precision/recall metrics for evaluating time series anomaly detection tasks: the *affiliation metrics*.

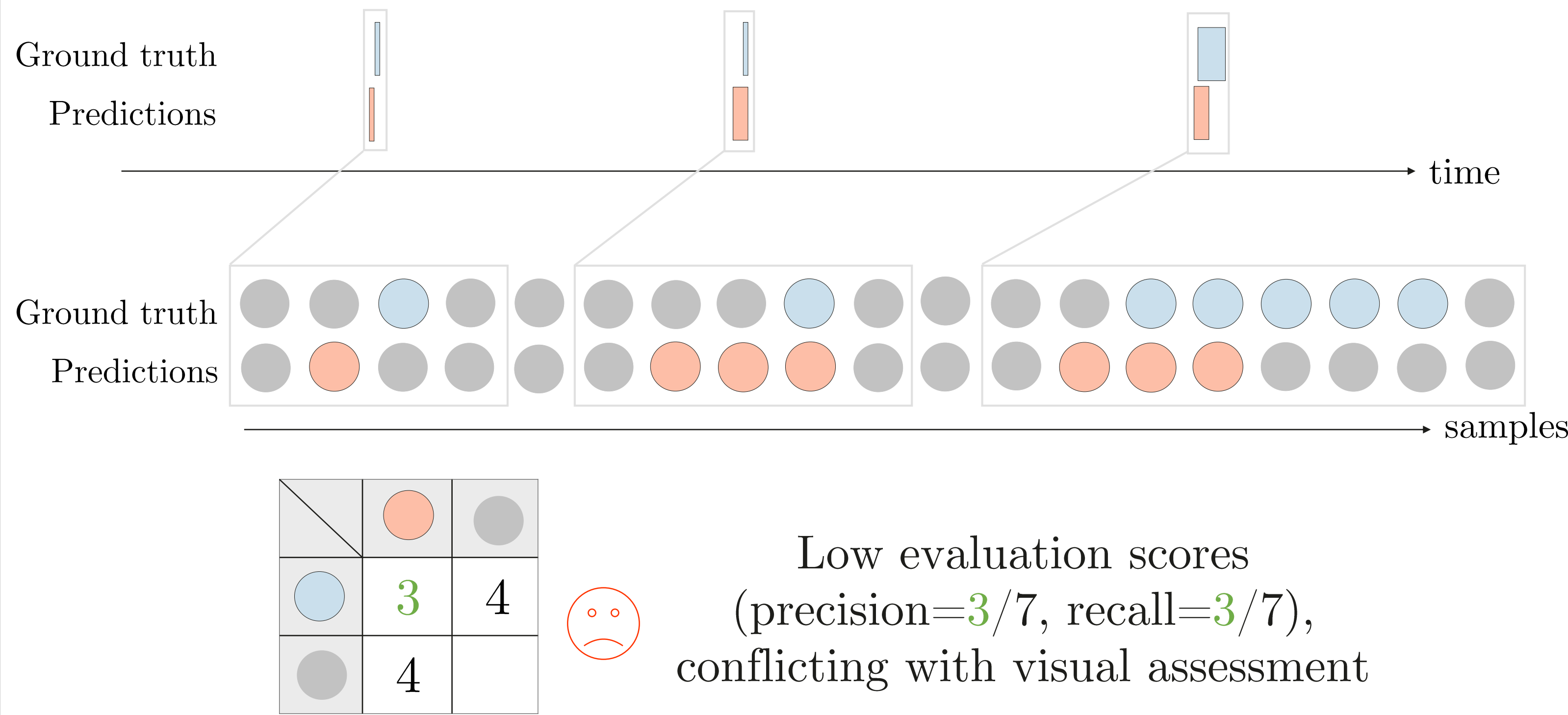


Performance evaluation

? How good are the predictions?

Motivations

1. Sample-based precision/recall cannot handle time series well



- A Unaware of temporal adjacency
- B Unaware of the event duration

2. Recent range-based metrics [1, 2] are easily gamed by adversary predictions.

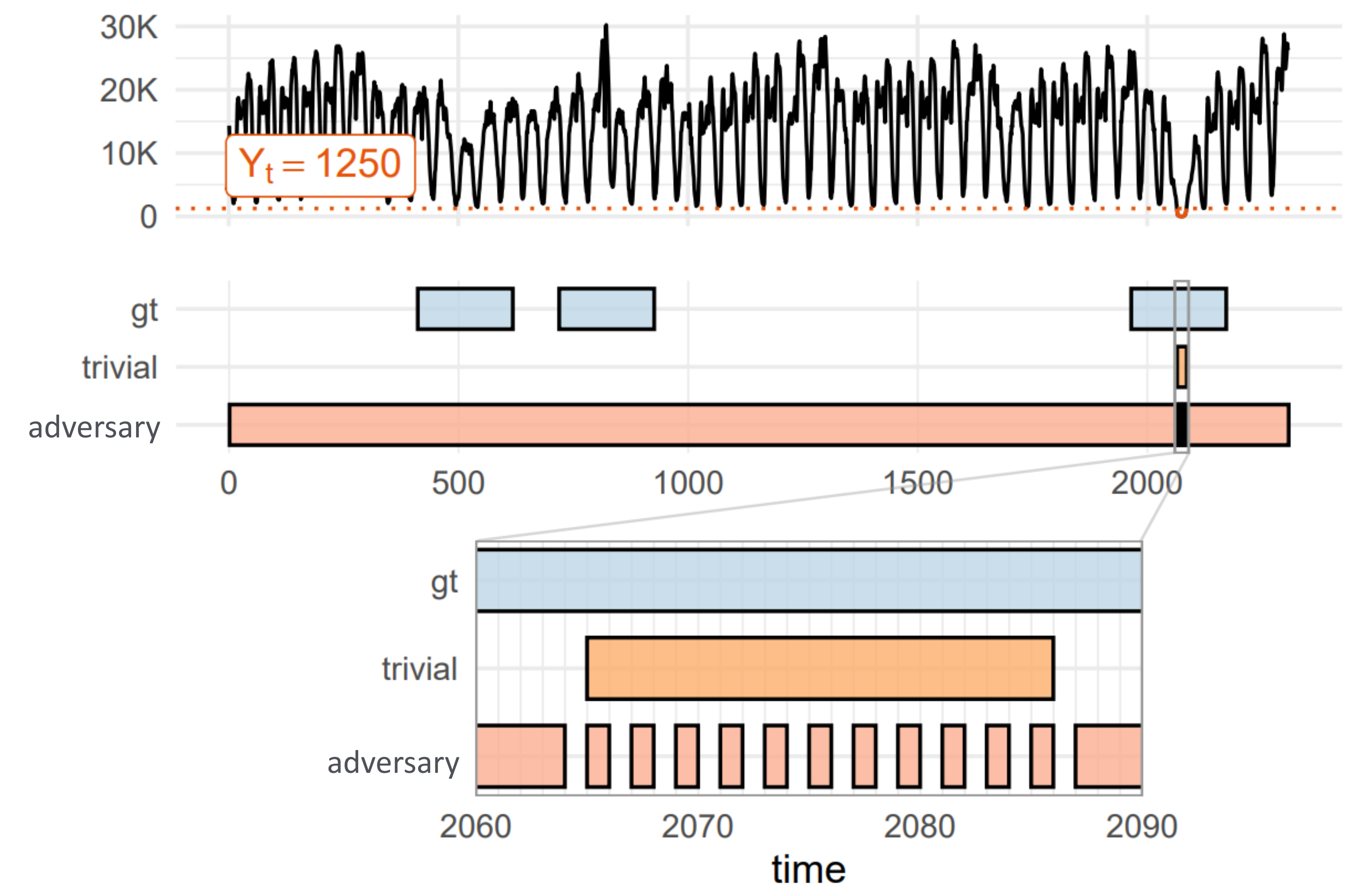
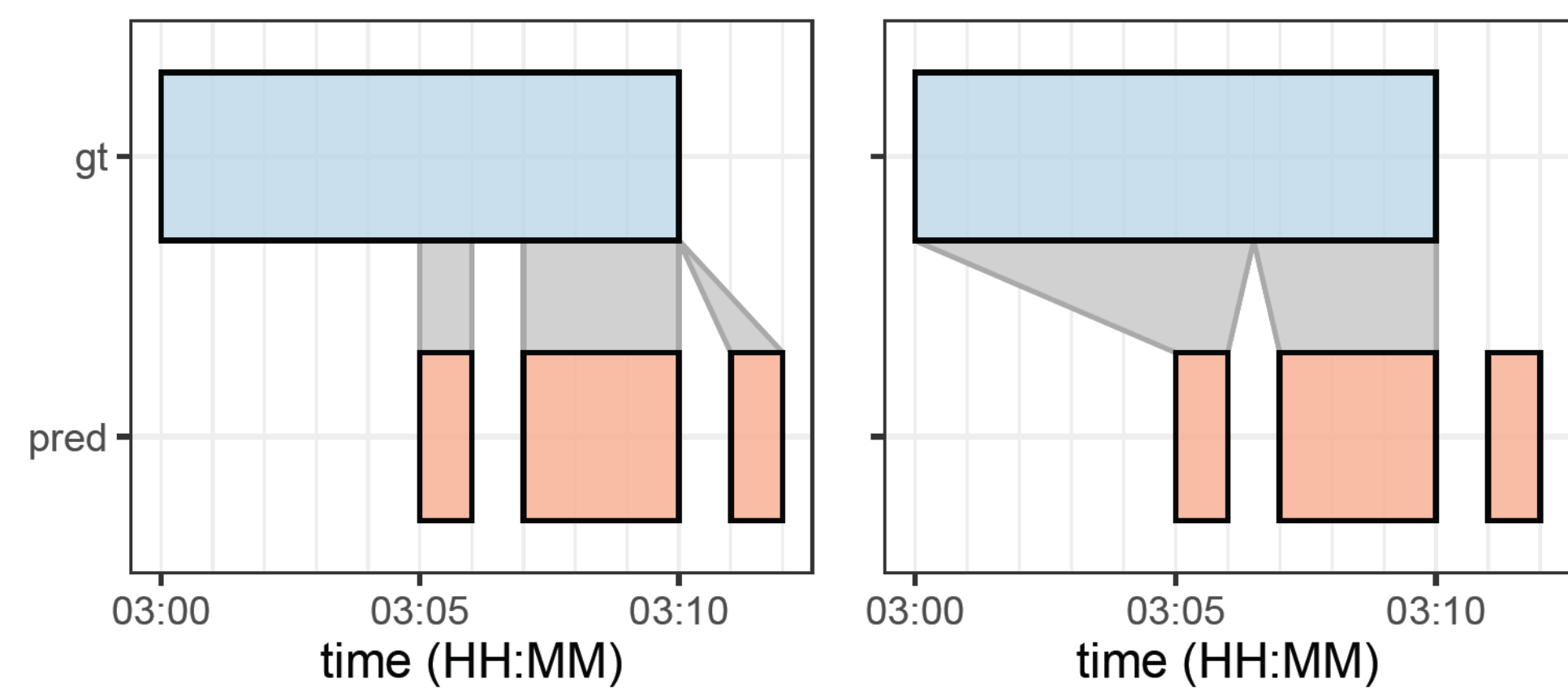


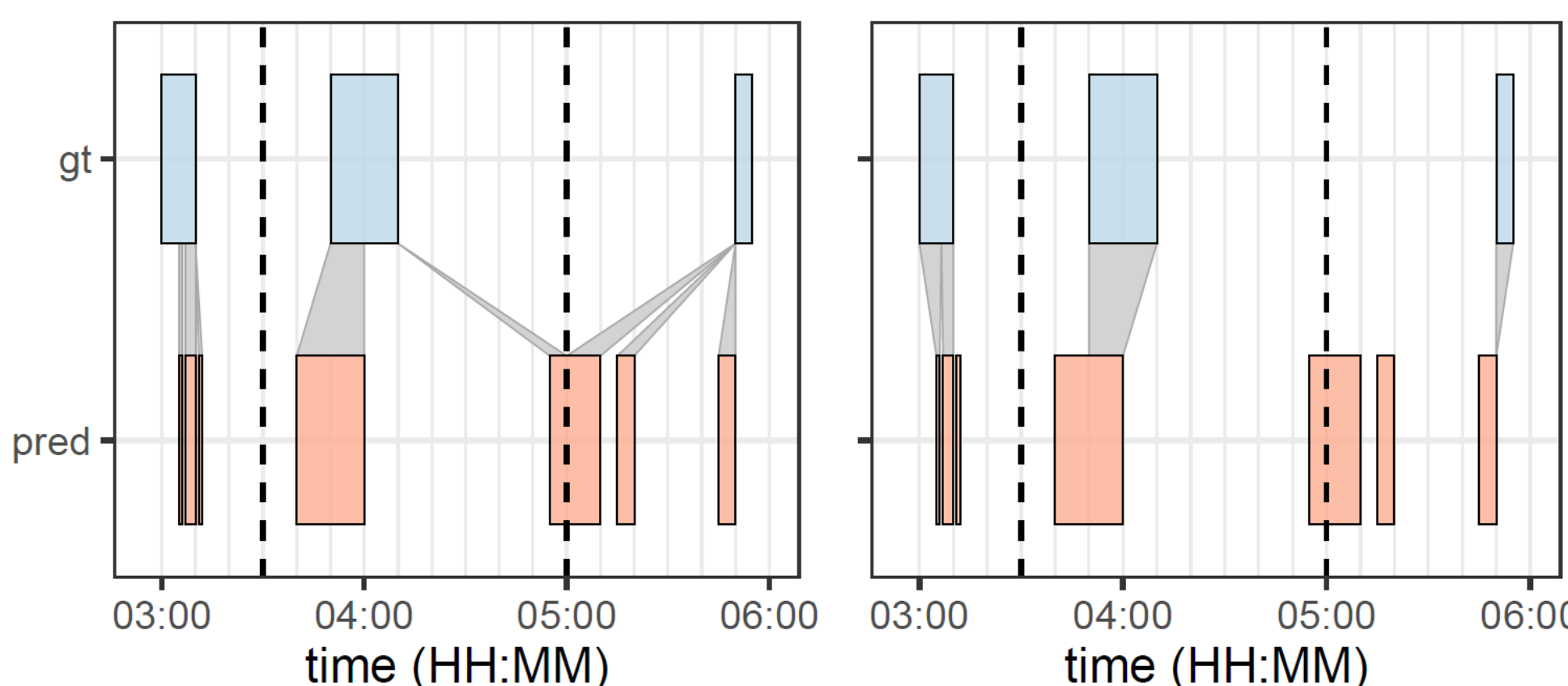
Figure 3: Construction of the adversary predictions

Concept 1: Proximity



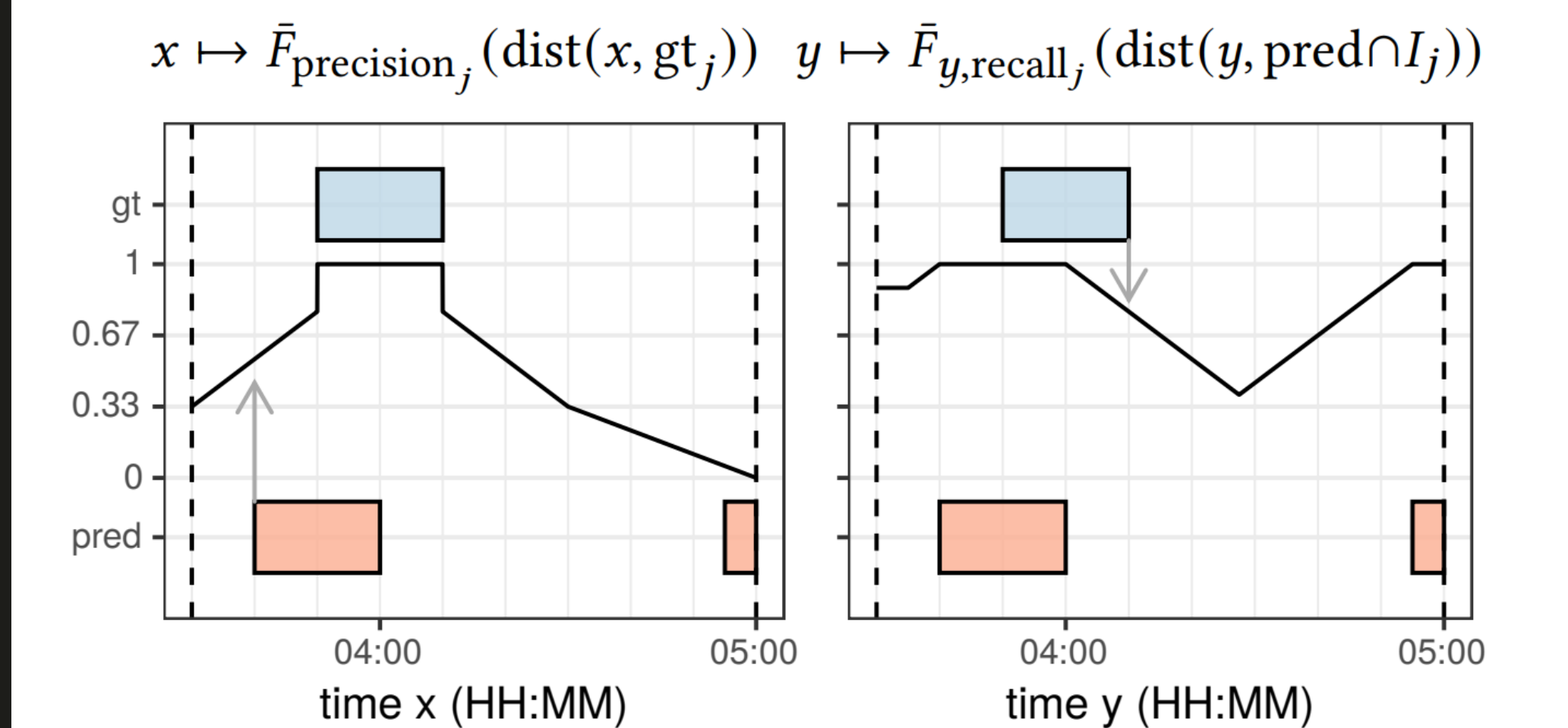
(a) *Average distance between sets:* example of the directed distance computed from predicted events to ground truth (left) and from ground truth to predicted events (right).

Concept 2: Locality



(b) *Local affiliation to the closest ground truth event:* example resulting in zones delimited by the dashed lines. The zones are similar for both directions: precision (left), and recall (right).

Concept 3: Probability



(c) *Comparison against random sampling:* example for converting each predicted sample to a precision score (left) and each ground truth sample to a recall score (right).

And more...

Locally interpretable and visualizable (Sec. 4.3)...

Algorithm	Mean of 35 events	Ev. 1	Ev. 2
iForest	0.52/0.84/0.64	0.37/0.53/0.44	1.00/0.91/0.95
seq2seq	0.86/0.79/0.83	0.96/1.00/0.98	0.86/1.00/0.93

Expected precision and recall given a single random prediction. The expected precision and recall are given by:

$$P_{\text{precision}} = \frac{1}{2} + \frac{p^2}{2}, \quad P_{\text{recall}} = \frac{1}{2}. \quad (8)$$

This property confirms that scores around 1/2 corresponds to random detector (cf. Appendix C.2 for proof). In this case, the classical precision/recall would give a precision of p and a recall of 0 assuming a large number of samples.

...Theoretical bounds on the score (Sec. 4.4)

Summary

- Parameter-free,
- Retain a physical meaning as linked to a time,
- Locally interpretable and visualizable (Sec. 4.3),
- Robust against adversary predictions (Sec. 4.2),
- Existence of statistical bounds of the score (Sec. 4.4),
- Aware of temporal adjacency (A) and of event duration (B).

References

- [1] N. Tatbul et al. 2018. Precision and recall for time series. NeurIPS.
- [2] W. Hwang et al. 2019. Time-series aware precision and recall for anomaly detection. In ACM CIKM